「機械学習のための数学」（飯塚秀明著　コロナ社）
章末問題解答例

1章

【1.1】(1) 表 1 と表 2　(2) 表 3 と表 4

【1.2】(1) 表 5　(2) 表 6

表 1: $P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R)$

P	Q	R	$Q \vee R$	$P \wedge Q$	$P \wedge R$	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	T	T	T	T	T
T	T	F	T	T	F	T	T
T	F	T	T	F	T	T	T
T	F	F	F	F	F	F	F
F	T	T	T	F	F	F	F
F	T	F	T	F	F	F	F
F	F	T	T	F	F	F	F
F	F	F	F	F	F	F	F

表 2: $P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R)$

P	Q	R	$Q \wedge R$	$P \vee Q$	$P \vee R$	$P \vee (Q \wedge R)$	$(P \vee Q) \wedge (P \vee R)$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	T	F	F	F
F	F	T	F	F	T	F	F
F	F	F	F	F	F	F	F

表 3: $\neg(P \vee Q) = \neg P \wedge \neg Q$

P	Q	$\neg P$	$\neg Q$	$P \vee Q$	$\neg(P \vee Q)$	$\neg P \wedge \neg Q$
T	T	F	F	T	F	F
T	F	F	T	T	F	F
F	T	T	F	T	F	F
F	F	T	T	F	T	T

表 4: $\neg(P \wedge Q) = \neg P \vee \neg Q$

P	Q	$\neg P$	$\neg Q$	$P \wedge Q$	$\neg(P \wedge Q)$	$\neg P \vee \neg Q$
T	T	F	F	T	F	F
T	F	F	T	F	T	T
F	T	T	F	F	T	T
F	F	T	T	F	T	T

表 5: $P \wedge (P \to Q) \Rightarrow Q$

P	Q	$P \to Q$	$P \wedge (P \to Q)$
T	**T**	T	**T**
T	F	F	F
F	T	T	F
F	F	T	F

表 6: $(P \to Q) \wedge (Q \to R) \Rightarrow (P \to R)$

P	Q	R	$P \to Q$	$Q \to R$	$(P \to Q) \wedge (Q \to R)$	$P \to R$
T	T	T	T	T	**T**	**T**
T	T	F	T	F	F	F
T	F	T	F	T	F	T
T	F	F	F	T	F	F
F	T	T	T	T	**T**	**T**
F	T	F	T	F	F	T
F	F	T	T	T	**T**	**T**
F	F	F	T	T	**T**	**T**

2章

【2.1】 (1) $x \in A \cap (B \cup C) \Leftrightarrow x \in A \wedge x \in B \cup C \Leftrightarrow x \in A \wedge (x \in B \vee x \in C) \Leftrightarrow (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C) \Leftrightarrow (x \in A \cap B) \vee (x \in A \cap C) \Leftrightarrow x \in (A \cap B) \cup (A \cap C)$

$x \in A \cup (B \cap C) \Leftrightarrow x \in A \vee x \in B \cap C \Leftrightarrow x \in A \vee (x \in B \wedge x \in C) \Leftrightarrow (x \in A \vee x \in B) \wedge (x \in A \vee x \in C) \Leftrightarrow (x \in A \cup B) \wedge (x \in A \cup C) \Leftrightarrow x \in (A \cup B) \cap (A \cup C)$

(2) $A = A \cap (A \cup B) \underset{[\text{命題 } 2.5(1)]}{=} (A \cap A) \cup (A \cap B) = A \cup (A \cap B)$

【2.2】 (1) $A \cap (B \backslash C) \underset{[\text{式 } (2.5)]}{=} A \cap (B \cap C^{\mathrm{c}}) = A \cap B \cap C^{\mathrm{c}}$ となる。一方で，

$$\begin{aligned}
(A \cap B) \backslash (A \cap C) &\underset{[\text{式 } (2.5)]}{=} (A \cap B) \cap (A \cap C)^{\mathrm{c}} \\
&\underset{[\text{命題 } 2.4]}{=} (A \cap B) \cap (A^{\mathrm{c}} \cup C^{\mathrm{c}}) \\
&\underset{[\text{命題 } 2.5(1)]}{=} (A \cap B \cap A^{\mathrm{c}}) \cup (A \cap B \cap C^{\mathrm{c}}) \\
&\overset{[\text{命題 } 2.2(2)]}{\underset{[\text{命題 } 2.3(2)]}{=}} \emptyset \cup (A \cap B \cap C^{\mathrm{c}}) \\
&\underset{[\text{命題 } 2.1(1)]}{=} A \cap B \cap C^{\mathrm{c}}
\end{aligned}$$

から，$A \cap (B \backslash C) = (A \cap B) \backslash (A \cap C)$ が成り立つ。

(2) $(A \cup B) \backslash C \underset{[\text{式 } (2.5)]}{=} (A \cup B) \cap C^{\mathrm{c}} \underset{[\text{命題 } 2.5(1)]}{=} (A \cap C^{\mathrm{c}}) \cup (B \cap C^{\mathrm{c}}) \underset{[\text{式 } (2.5)]}{=} (A \backslash C) \cup (B \backslash C)$

$(A \backslash C) \cap (B \backslash C) \underset{[\text{式 } (2.5)]}{=} (A \cap C^{\mathrm{c}}) \cap (B \cap C^{\mathrm{c}}) \underset{[\text{命題 } 2.2(3),(5)]}{=} (A \cap B) \cap C^{\mathrm{c}} \underset{[\text{式 } (2.5)]}{=} (A \cap B) \backslash C$

(3) $C \backslash (A \cup B) \underset{[\text{式 } (2.5)]}{=} C \cap (A \cup B)^{\mathrm{c}} \underset{[\text{命題 } 2.4]}{=} C \cap (A^{\mathrm{c}} \cap B^{\mathrm{c}}) \underset{[\text{命題 } 2.2(3),(5)]}{=} (C \cap A^{\mathrm{c}}) \cap (C \cap B^{\mathrm{c}}) \underset{[\text{式 } (2.5)]}{=} (C \backslash A) \cap (C \backslash B)$

$C \backslash (A \cap B) \underset{[\text{式 } (2.5)]}{=} C \cap (A \cap B)^{\mathrm{c}} \underset{[\text{命題 } 2.4]}{=} C \cap (A^{\mathrm{c}} \cup B^{\mathrm{c}}) \underset{[\text{命題 } 2.5(1)]}{=} (C \cap A^{\mathrm{c}}) \cup (C \cap B^{\mathrm{c}}) \underset{[\text{式 } (2.5)]}{=} (C \backslash A) \cup (C \backslash B)$

3 章

【3.1】任意の $k \in \mathbb{N}$ に対して，

$$x - y = (x - x_k) + (x_k - y_k) + (y_k - y) \underset{[x_k \le y_k]}{\le} (x - x_k) + (y_k - y)$$

が成り立つことと三角不等式から，

$$\forall k(x - y \le \underbrace{|x_k - x| + |y_k - y|}_{z_k}) \tag{1}$$

を得る。$(x_k \to x) \wedge (y_k \to y)$ から，任意の $\varepsilon > 0$ に対して，

$$\exists k_0 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \ge k_0 \Rightarrow |x_k - x| < \varepsilon)$$
$$\exists k_1 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \ge k_1 \Rightarrow |y_k - y| < \varepsilon)$$

が成り立つ。$k_2 = \max\{k_0, k_1\}$ とすると，

$$\begin{aligned}\exists k_2 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \ge k_2 &\Rightarrow |x_k - x| < \varepsilon \wedge |y_k - y| < \varepsilon \\ &\Rightarrow z_k = |x_k - x| + |y_k - y| < 2\varepsilon)\end{aligned} \tag{2}$$

つまり，$z_k \to 0$ である。背理法（1.4.2 項）を用いて，$x \le y$ を示す。いま，$\neg(x \le y) = (x > y)$ が成り立つとすると，式 (1) から，$\forall k(0 < x - y \le z_k)$，つまり，命題 3.4 から，$z_k \nrightarrow 0$ となる。これは，$z_k \to 0$ に反する。

【3.2】$|z_k - x_k| \underset{[x_k \le z_k]}{=} z_k - x_k \underset{[\text{三角不等式}]}{\le} |z_k - x| + |x_k - x|$ と式 (2) を得るための同様の議論から，$z_k - x_k \to 0$ を得る。さらに，

$$\begin{aligned}|y_k - x| = |y_k - z_k + z_k - x| &\underset{[\text{三角不等式}]}{\overset{[y_k \le z_k]}{\le}} (z_k - y_k) + |z_k - x| \\ &\underset{[x_k \le y_k]}{\le} \underbrace{(z_k - x_k) + |z_k - x|}_{w_k}\end{aligned}$$

を得る。$(z_k - x_k \to 0) \wedge (z_k \to x)$ から，$w_k \to 0$ を満たすので（式 (2) 参照），

$$\forall \varepsilon > 0\ \exists k_0 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \ge k_0 \Rightarrow w_k < \varepsilon \Rightarrow |y_k - x| < \varepsilon)$$

つまり，$y_k \to x$ となる。

【3.3】$x_k \to x$，$y_k \to y$ とする。

(1) 式 (3.13) から，$0 \le ||x_k| - |x|| \le |x_k - x| \to 0$ なので，はさみうちの原理から，$|x_k| \to |x|$ を得る。

(2) $|(x_k + y_k) - (x + y)| = |(x_k - x) + (y_k - y)| \underset{[\text{三角不等式}]}{\le} |x_k - x| + |y_k - y| \to 0$ とはさみうちの原理から，$x_k + y_k \to x + y$ を得る。

(3) 定理 3.2(1) から，$y_k \to y$ ならば，

$$\exists C > 0\ \forall k \in \mathbb{N}\ (|y_k| \le C) \tag{3}$$

が成り立つことを利用して，

$$\begin{aligned}|x_k y_k - xy| &= |(x_k - x)y_k + x(y_k - y)| \\ &\underset{[\text{三角不等式}]}{\le} |(x_k - x)y_k| + |x(y_k - y)| \\ &= |x_k - x||y_k| + |x||y_k - y| \\ &\underset{[\text{式 (3)}]}{\le} C|x_k - x| + |x||y_k - y| \to 0\end{aligned}$$

を得る。よって，$x_k y_k \to xy$ である。

(4) $y_k \to y \neq 0$ から，$\varepsilon = |y|/2 > 0$ に対して，ある $k_0 \in \mathbb{N}$ が存在して，$k \geq k_0$ に対して，

$$|y_k - y| < \frac{|y|}{2} \underset{[\text{式 (3.13)}]}{\Rightarrow} -\frac{|y|}{2} < |y_k| - |y| < \frac{|y|}{2} \Rightarrow \frac{|y|}{2} < |y_k|$$

が成り立つ。$k \geq k_0$ に対して，

$$\begin{aligned}\left|\frac{x_k}{y_k} - \frac{x}{y}\right| &= \frac{|yx_k - xy_k|}{|y_k||y|} \\ &< \frac{2}{|y|^2}|yx_k - xy_k| \\ &\underset{[\text{三角不等式}]}{\leq} \frac{2}{|y|^2}(|y||x_k - x| + |x||y_k - y|) \underset{【3.3】(3)}{\to} 0\end{aligned}$$

から，$x_k/y_k \to x/y$ となる。

【3.4】 $\overline{x}_k = \sup_{n\geq k} x_n \to \overline{\lim}_{k\to+\infty} x_k$，$\underline{x}_k = \inf_{n\geq k} x_n \to \underline{\lim}_{k\to+\infty} x_k$ が成り立つ（定義 3.5）。

(1) $k \in \mathbb{N}$ とする。このとき，$\forall n(n \geq k \Rightarrow x_n \leq y_n)$ なので，$\forall n(n \geq k \Rightarrow x_n \leq y_n \leq \overline{y}_k = \sup_{m\geq k} y_m)$，つまり，$x_n \leq \overline{y}_k$ を満たす。定義 3.2 から，$\overline{x}_k \leq \overline{y}_k$ となる。よって，【3.1】から，$\overline{\lim}_{k\to+\infty} x_k \leq \overline{\lim}_{k\to+\infty} y_k$ を得る。同様の議論から，$\underline{\lim}_{k\to+\infty} x_k \leq \underline{\lim}_{k\to+\infty} y_k$ を得る。

(2) 命題 3.2(1) から得られる

$$\sup_{n\geq k}(x_n + y_n) \leq \sup_{n\geq k} x_n + \sup_{n\geq k} y_n = \overline{x}_k + \overline{y}_k$$

と【3.1】から，$\overline{\lim}_{k\to+\infty}(x_k + y_k) \leq \overline{\lim}_{k\to+\infty} x_k + \overline{\lim}_{k\to+\infty} y_k$ を得る。

(3) 命題 3.2(1) から得られる

$$\inf_{n\geq k}(x_n + y_n) \geq \inf_{n\geq k} x_n + \inf_{n\geq k} y_n = \underline{x}_k + \underline{y}_k$$

と【3.1】から，$\underline{\lim}_{k\to+\infty}(x_k + y_k) \geq \underline{\lim}_{k\to+\infty} x_k + \underline{\lim}_{k\to+\infty} y_k$ を得る。

(4) 【3.4】(2) と $x_k \to x$ から，

$$\overline{\lim_{k\to+\infty}}(x_k + y_k) \leq \overline{\lim_{k\to+\infty}} x_k + \overline{\lim_{k\to+\infty}} y_k = x + \overline{\lim_{k\to+\infty}} y_k$$

が成り立つ。

$$\begin{aligned}\overline{\lim_{k\to+\infty}} y_k &= \overline{\lim_{k\to+\infty}}(x_k + y_k - x_k) \\ &\underset{【3.4】(2)}{\leq} \overline{\lim_{k\to+\infty}}(x_k + y_k) + \overline{\lim_{k\to+\infty}}(-x_k) \\ &\underset{【3.3】(3)}{=} \overline{\lim_{k\to+\infty}}(x_k + y_k) - x\end{aligned}$$

から，

$$x + \overline{\lim_{k\to+\infty}} y_k \leq \overline{\lim_{k\to+\infty}}(x_k + y_k)$$

も成り立つ。以上のことから，$\overline{\lim}_{k\to+\infty}(x_k + y_k) = x + \overline{\lim}_{k\to+\infty} y_k$ となる。

(5) 【3.4】(3) と $x_k \to x$ から，

$$\underline{\lim}_{k\to+\infty}(x_k + y_k) \geq \underline{\lim}_{k\to+\infty} x_k + \underline{\lim}_{k\to+\infty} y_k = x + \underline{\lim}_{k\to+\infty} y_k$$

が成り立つ。

$$\underline{\lim}_{k\to+\infty} y_k = \underline{\lim}_{k\to+\infty}(x_k + y_k - x_k)$$

$$
\begin{aligned}
&\underset{【3.4】(3)}{\geq} \varliminf_{k\to+\infty}(x_k+y_k)+\varliminf_{k\to+\infty}(-x_k)\\
&\underset{【3.3】(3)}{=} \varliminf_{k\to+\infty}(x_k+y_k)-x
\end{aligned}
$$

から，

$$
x+\varliminf_{k\to+\infty} y_k \geq \varliminf_{k\to+\infty}(x_k+y_k)
$$

も成り立つ。以上のことから，$\varliminf_{k\to+\infty}(x_k+y_k)=x+\varliminf_{k\to+\infty} y_k$ となる。

【3.5】 (1) $\langle \boldsymbol{x},\boldsymbol{y}\rangle_2=\sum_{i=1}^d x_iy_i$ が (I1)–(I4) の性質を満たすことを以下で確認する。

(I1) $\langle \boldsymbol{x},\boldsymbol{x}\rangle_2=\sum_{i=1}^d x_i^2\geq 0$ が成り立つ。$\langle \boldsymbol{x},\boldsymbol{x}\rangle_2=\sum_{i=1}^d x_i^2=0 \Leftrightarrow \forall i(x_i^2=0)\Leftrightarrow \forall i(x_i=0)\Leftrightarrow \boldsymbol{x}=\boldsymbol{0}$

(I2) $\langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}\rangle_2=\sum_{i=1}^d(x_i+y_i)z_i=\sum_{i=1}^d x_iz_i+\sum_{i=1}^d y_iz_i=\langle \boldsymbol{x},\boldsymbol{z}\rangle_2+\langle \boldsymbol{y},\boldsymbol{z}\rangle_2$

(I3) $\langle \alpha\boldsymbol{x},\boldsymbol{y}\rangle_2=\sum_{i=1}^d(\alpha x_i)y_i=\alpha\sum_{i=1}^d x_iy_i=\alpha\langle \boldsymbol{x},\boldsymbol{y}\rangle_2$

(I4) $\langle \boldsymbol{x},\boldsymbol{y}\rangle_2=\sum_{i=1}^d x_iy_i=\sum_{i=1}^d y_ix_i=\langle \boldsymbol{y},\boldsymbol{x}\rangle_2$

(2) $\langle \boldsymbol{x},\boldsymbol{y}\rangle_H=\langle \boldsymbol{x},H\boldsymbol{y}\rangle_2$ が (I1)–(I4) の性質を満たすことを以下で確認する（正定値行列 H の性質については 4.1.3 項を参照）。

(I1) $H\in\mathbb{S}^d_{++}$ から，$\langle \boldsymbol{x},\boldsymbol{x}\rangle_H=\boldsymbol{x}^\top H\boldsymbol{x}\geq 0$ が成り立つ。$H=(H^{1/2})^2$ と $\langle\cdot,\cdot\rangle_2$ が (I1) を満たすことから，

$$
\langle \boldsymbol{x},\boldsymbol{x}\rangle_H=\left\langle H^{\frac{1}{2}}\boldsymbol{x},H^{\frac{1}{2}}\boldsymbol{x}\right\rangle_2=0\Leftrightarrow H^{\frac{1}{2}}\boldsymbol{x}=\boldsymbol{0}
$$

を得る。$H^{1/2}\in\mathbb{S}^d_{++}$ は逆行列をもつので，$H^{1/2}\boldsymbol{x}=\boldsymbol{0}\Leftrightarrow\boldsymbol{x}=\boldsymbol{0}$ を得る。

(I2) $H=(H^{1/2})^2$ と $\langle\cdot,\cdot\rangle_2$ が (I2) を満たすことを利用して，

$$
\begin{aligned}
\langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{z}\rangle_H&=\left\langle H^{\frac{1}{2}}(\boldsymbol{x}+\boldsymbol{y}),H^{\frac{1}{2}}\boldsymbol{z}\right\rangle_2\\
&=\left\langle H^{\frac{1}{2}}\boldsymbol{x}+H^{\frac{1}{2}}\boldsymbol{y},H^{\frac{1}{2}}\boldsymbol{z}\right\rangle_2\\
&=\left\langle H^{\frac{1}{2}}\boldsymbol{x},H^{\frac{1}{2}}\boldsymbol{z}\right\rangle_2+\left\langle H^{\frac{1}{2}}\boldsymbol{y},H^{\frac{1}{2}}\boldsymbol{z}\right\rangle_2\\
&=\langle \boldsymbol{x},H\boldsymbol{z}\rangle_2+\langle \boldsymbol{y},H\boldsymbol{z}\rangle_2\\
&=\langle \boldsymbol{x},\boldsymbol{z}\rangle_H+\langle \boldsymbol{y},\boldsymbol{z}\rangle_H
\end{aligned}
$$

を得る。

(I3) $\langle\cdot,\cdot\rangle_2$ が (I3) を満たすことから，

$$
\begin{aligned}
\langle \alpha\boldsymbol{x},\boldsymbol{y}\rangle_H&=\left\langle H^{\frac{1}{2}}(\alpha\boldsymbol{x}),H^{\frac{1}{2}}\boldsymbol{y}\right\rangle_2=\left\langle \alpha H^{\frac{1}{2}}\boldsymbol{x},H^{\frac{1}{2}}\boldsymbol{y}\right\rangle_2\\
&=\alpha\left\langle H^{\frac{1}{2}}\boldsymbol{x},H^{\frac{1}{2}}\boldsymbol{y}\right\rangle_2=\alpha\left\langle \boldsymbol{x},H\boldsymbol{y}\right\rangle_2=\alpha\langle \boldsymbol{x},\boldsymbol{y}\rangle_H
\end{aligned}
$$

となる。

(I4) $\langle\cdot,\cdot\rangle_2$ が (I4) を満たすことから，

$$
\begin{aligned}
\langle \boldsymbol{x},\boldsymbol{y}\rangle_H&=\langle \boldsymbol{x},H\boldsymbol{y}\rangle_2=\langle H\boldsymbol{y},\boldsymbol{x}\rangle_2=(H\boldsymbol{y})^\top\boldsymbol{x}\\
&=(\boldsymbol{y}^\top H)\boldsymbol{x}=\boldsymbol{y}^\top(H\boldsymbol{x})=\langle \boldsymbol{y},H\boldsymbol{x}\rangle_2=\langle \boldsymbol{y},\boldsymbol{x}\rangle_H
\end{aligned}
$$

を得る（転置 $^\top$ の性質については命題 4.2 を参照）。

【3.6】 $\|\boldsymbol{x}\|^2=\langle \boldsymbol{x},\boldsymbol{x}\rangle\underset{(\mathrm{I1})}{\geq}0$ に注意する。

(1) $\langle\cdot,\cdot\rangle$ が (I1)–(I4) を満たすことを利用して，

$$
\|\boldsymbol{x}+\boldsymbol{y}\|^2=\langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}\rangle
$$

$$
\begin{aligned}
&\underset{(\mathrm{I2})}{=} \langle \boldsymbol{x}, \boldsymbol{x}+\boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle \\
&\underset{(\mathrm{I4})}{=} \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{y}\rangle \\
&\underset{(\mathrm{I2})}{=} \langle \boldsymbol{x}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&= \|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\
&\underset{(\mathrm{I4})}{=} \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2
\end{aligned}
$$

から，

$$
\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \tag{4}
$$

を得る。式 (4) の $\boldsymbol{y}$ に $-\boldsymbol{y}$ を代入すると，

$$
\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, -\boldsymbol{y}\rangle + \|-\boldsymbol{y}\|^2
$$

となる。さらに，

$$
\langle \boldsymbol{x}, -\boldsymbol{y}\rangle \underset{(\mathrm{I4})}{=} \langle -\boldsymbol{y}, \boldsymbol{x}\rangle \underset{(\mathrm{I3})}{=} -\langle \boldsymbol{y}, \boldsymbol{x}\rangle \underset{(\mathrm{I4})}{=} -\langle \boldsymbol{x}, \boldsymbol{y}\rangle
$$

であり，また，任意の $\alpha \in \mathbb{R}$ に対して，

$$
\begin{aligned}
\|\alpha \boldsymbol{y}\|^2 = \langle \alpha\boldsymbol{y}, \alpha\boldsymbol{y}\rangle &\underset{(\mathrm{I3})}{=} \alpha\langle \boldsymbol{y}, \alpha\boldsymbol{y}\rangle \underset{(\mathrm{I4})}{=} \alpha\langle \alpha\boldsymbol{y}, \boldsymbol{y}\rangle \underset{(\mathrm{I3})}{=} \alpha^2\langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&= \alpha^2\|\boldsymbol{y}\|^2
\end{aligned} \tag{5}
$$

が成り立つので，$\|-\boldsymbol{y}\|^2 = (-1)^2\|\boldsymbol{y}\|^2 = \|\boldsymbol{y}\|^2$ である。よって，

$$
\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \tag{6}
$$

となる。

(2) 式 (6) から得られる等式

$$
2\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 \tag{7}
$$

を利用して，

$$
\begin{aligned}
\|\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}\|^2 &\underset{[\text{式 }(4)]}{=} \|\alpha\boldsymbol{x}\|^2 + 2\langle \alpha\boldsymbol{x}, (1-\alpha)\boldsymbol{y}\rangle + \|(1-\alpha)\boldsymbol{y}\|^2 \\
&\overset{(\mathrm{I3}),\,(\mathrm{I4})}{\underset{[\text{式 }(5)]}{=}} \alpha^2\|\boldsymbol{x}\|^2 + 2\alpha(1-\alpha)\langle \boldsymbol{x}, \boldsymbol{y}\rangle + (1-\alpha)^2\|\boldsymbol{y}\|^2 \\
&\underset{[\text{式 }(7)]}{=} \alpha^2\|\boldsymbol{x}\|^2 + \alpha(1-\alpha)\left\{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2\right\} + (1-\alpha)^2\|\boldsymbol{y}\|^2 \\
&= \alpha\|\boldsymbol{x}\|^2 + (1-\alpha)\|\boldsymbol{y}\|^2 - \alpha(1-\alpha)\|\boldsymbol{x}-\boldsymbol{y}\|^2
\end{aligned}
$$

が成り立つ。

(3) $\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2$ と $\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2$ を足し合わせることで，$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$ を得る。

(4) 関数 $F\colon \mathbb{R} \to \mathbb{R}_+$ を，任意の $t \in \mathbb{R}$ に対して，

$$
\begin{aligned}
F(t) = \|t\boldsymbol{x} + \boldsymbol{y}\|^2 &\underset{[\text{式 }(4)]}{=} \|t\boldsymbol{x}\|^2 + 2\langle t\boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2 \\
&\overset{[\text{式 }(5)]}{\underset{(\mathrm{I3})}{=}} \|\boldsymbol{x}\|^2 t^2 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle t + \|\boldsymbol{y}\|^2 \geq 0
\end{aligned}
$$

と定義する。$F(t) = 0$ の判別式から，

$$
\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2 - \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \leq 0 \Leftrightarrow (\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{x}\|\|\boldsymbol{y}\|)(\langle \boldsymbol{x}, \boldsymbol{y}\rangle - \|\boldsymbol{x}\|\|\boldsymbol{y}\|) \leq 0
$$

$$
\begin{aligned}
&\Rightarrow (\langle \boldsymbol{x},\boldsymbol{y}\rangle + \|\boldsymbol{x}\|\|\boldsymbol{y}\| \le 0 \wedge \langle \boldsymbol{x},\boldsymbol{y}\rangle - \|\boldsymbol{x}\|\|\boldsymbol{y}\| \ge 0)\\
&\quad \vee (\langle \boldsymbol{x},\boldsymbol{y}\rangle + \|\boldsymbol{x}\|\|\boldsymbol{y}\| \ge 0 \wedge \langle \boldsymbol{x},\boldsymbol{y}\rangle - \|\boldsymbol{x}\|\|\boldsymbol{y}\| \le 0)\\
&\Rightarrow (\|\boldsymbol{x}\|\|\boldsymbol{y}\| = 0) \vee (|\langle \boldsymbol{x},\boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|)\\
&\Rightarrow \underbrace{(\boldsymbol{x}=\boldsymbol{0}) \vee (\boldsymbol{y}=\boldsymbol{0})}_{\neg(\boldsymbol{x}\neq\boldsymbol{0}\wedge\boldsymbol{y}\neq\boldsymbol{0})} \vee (|\langle \boldsymbol{x},\boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|)
\end{aligned}
$$

を得る。よって，$\forall\boldsymbol{x}\forall\boldsymbol{y}(|\langle \boldsymbol{x},\boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|)$ となる。

(5) $\|\boldsymbol{x}\pm\boldsymbol{y}\|^2 \underset{(\mathrm{I1})}{\le} \|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 \underset{[\text{中線定理}]}{=} 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$

【3.7】 $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i|$ が (N1)–(N3) を満たすことを以下で確認する。

(N1) $\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i| \ge 0$ が成り立つ。$\|\boldsymbol{x}\|_1 = \sum_{i=1}^d |x_i| = 0 \Leftrightarrow \forall i(|x_i| = 0) \Leftrightarrow \forall i(x_i = 0) \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

(N2) $\|\alpha\boldsymbol{x}\|_1 = \sum_{i=1}^d |\alpha x_i| = |\alpha| \sum_{i=1}^d |x_i| = |\alpha|\|\boldsymbol{x}\|_1$

(N3) $\|\boldsymbol{x}+\boldsymbol{y}\|_1 = \sum_{i=1}^d |x_i + y_i| \underset{[\text{三角不等式}]}{\le} \sum_{i=1}^d (|x_i| + |y_i|) = \sum_{i=1}^d |x_i| + \sum_{i=1}^d |y_i| = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1$

$\|\boldsymbol{x}\|_\infty = \max\{|x_i| \colon i = 1, 2, \cdots, d\}$ が (N1)–(N3) を満たすことを以下で確認する。ここで，$[d] = \{1, 2, \cdots, d\}$ としたとき，

$$\|\boldsymbol{x}\|_\infty = \sup_{i\in[d]} |x_i|$$

が成り立つことに注意する。

(N1) $\|\boldsymbol{x}\|_\infty = \sup_{i\in[d]} |x_i| \ge 0$ が成り立つ。$\|\boldsymbol{x}\|_\infty = \sup_{i\in[d]} |x_i| = 0 \Leftrightarrow \forall i(|x_i| = 0) \Leftrightarrow \forall i(x_i = 0) \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$ が成り立つ。

(N2) $\|\alpha\boldsymbol{x}\|_\infty = \sup_{i\in[d]} |\alpha x_i| = \sup_{i\in[d]} |\alpha||x_i| \underset{[\text{命題 3.2(2)}]}{=} |\alpha| \sup_{i\in[d]} |x_i| = |\alpha|\|\boldsymbol{x}\|_\infty$

(N3) $\|\boldsymbol{x}+\boldsymbol{y}\|_\infty = \sup_{i\in[d]} |x_i + y_i| \underset{[\text{三角不等式}]}{\le} \sup_{i\in[d]}(|x_i| + |y_i|) \underset{[\text{命題 3.2(1)}]}{\le} \sup_{i\in[d]} |x_i| + \sup_{i\in[d]} |y_i| = \|\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty$

【3.8】 $p \in (1, +\infty)$ とする。$\|\boldsymbol{x}\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$ が (N1)–(N2) を満たすことを以下で確認する。

(N1) $\|\boldsymbol{x}\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p} \ge 0$ が成り立つ。$\|\boldsymbol{x}\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p} = 0 \Leftrightarrow \forall i(|x_i|^p = 0) \Leftrightarrow \forall i(x_i = 0) \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$

(N2) $\|\alpha\boldsymbol{x}\|_p = (\sum_{i=1}^d |\alpha x_i|^p)^{1/p} = (\sum_{i=1}^d |\alpha|^p|x_i|^p)^{1/p} = (|\alpha|^p \sum_{i=1}^d |x_i|^p)^{1/p} = |\alpha|(\sum_{i=1}^d |x_i|^p)^{1/p} = |\alpha|\|\boldsymbol{x}\|_p$

(1) $a = 0 \vee b = 0$ のときの成立は明らかである。$a \neq 0 \wedge b \neq 0$，つまり，$a > 0 \wedge b > 0$ とする。$c > 0$ とし，$x > 0$ に対して，$y = \log x$ とする。このとき，$cy = c\log x = \log x^c \Leftrightarrow x^c = e^{cy}$ を満たす。よって，

$$
\begin{aligned}
x = a \wedge c = \frac{1}{p} &\Rightarrow a^{\frac{1}{p}} = e^{\frac{1}{p}\log a}\\
x = b \wedge c = \frac{1}{q} &\Rightarrow b^{\frac{1}{q}} = e^{\frac{1}{q}\log b}
\end{aligned}
$$

と $e^x = \exp x$ $(x > 0)$ の凸性（例 3.5(4) 参照）から，

$$
\begin{aligned}
a^{\frac{1}{p}} b^{\frac{1}{q}} &= e^{\frac{1}{p}\log a} e^{\frac{1}{q}\log b} = \exp\left(\frac{1}{p}\log a + \frac{1}{q}\log b\right)\\
&\le \frac{1}{p}\exp(\log a) + \frac{1}{q}\exp(\log b) = \frac{a}{p} + \frac{b}{q}
\end{aligned} \tag{8}
$$

が成り立つ。

(2) $A=\sum_{i=1}^d a_i^p$ とし，$B=\sum_{i=1}^d b_i^q$ とする。$A=0 \vee B=0 \Leftrightarrow \forall i(a_i=0) \vee \forall i(b_i=0)$ なので，$A=0 \vee B=0$ のときの成立は明らかである。$A \neq 0 \wedge B \neq 0$ とする。任意の $i \in \{1,2,\cdots,d\}$ に対して，

$$\frac{a_i b_i}{A^{\frac{1}{p}} B^{\frac{1}{q}}} = \left(\frac{a_i^p}{A}\right)^{\frac{1}{p}} \left(\frac{b_i^q}{B}\right)^{\frac{1}{q}} \underset{[\text{式 } (8)]}{\leq} \frac{a_i^p}{pA} + \frac{b_i^q}{qB}$$

から，

$$\frac{1}{A^{\frac{1}{p}} B^{\frac{1}{q}}} \sum_{i=1}^d a_i b_i \leq \frac{1}{pA} \sum_{i=1}^d a_i^p + \frac{1}{qB} \sum_{i=1}^d b_i^q = \frac{1}{p} + \frac{1}{q} = 1$$

を得る。これは，

$$\sum_{i=1}^d a_i b_i \leq A^{\frac{1}{p}} B^{\frac{1}{q}} = \left(\sum_{i=1}^d a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^d b_i^q\right)^{\frac{1}{q}} \tag{9}$$

を意味する。

(3) $1/p+1/q=1 \Rightarrow q(p-1)=p+q-q=p$ に注意をして，

$$\begin{aligned}
\sum_{i=1}^d (a_i+b_i)^p &= \sum_{i=1}^d (a_i+b_i)(a_i+b_i)^{p-1} \\
&= \sum_{i=1}^d a_i(a_i+b_i)^{p-1} + \sum_{i=1}^d b_i(a_i+b_i)^{p-1} \\
&\underset{[\text{式 } (9)]}{\leq} \left(\sum_{i=1}^d a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^d (a_i+b_i)^{q(p-1)}\right)^{\frac{1}{q}} \\
&\quad + \left(\sum_{i=1}^d b_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^d (a_i+b_i)^{q(p-1)}\right)^{\frac{1}{q}} \\
&= \left\{\left(\sum_{i=1}^d a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^d b_i^p\right)^{\frac{1}{q}}\right\} \left(\sum_{i=1}^d (a_i+b_i)^p\right)^{\frac{1}{q}}
\end{aligned}$$

と $1-1/q=1/p$ から，

$$\left(\sum_{i=1}^d (a_i+b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^d a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^d b_i^p\right)^{\frac{1}{p}} \tag{10}$$

が成り立つ。$\|\boldsymbol{x}\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$ が (N3) を満たすことを以下で確認する。

(N3) $\|\boldsymbol{x}+\boldsymbol{y}\|_p = (\sum_{i=1}^d |x_i+y_i|^p)^{1/p} \underset{[\text{式 } (10)]}{\leq} (\sum_{i=1}^d |x_i|^p)^{1/p} + (\sum_{i=1}^d |y_i|^p)^{1/p} = \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p$

【3.9】 $\boldsymbol{x}_k = (x_{k,1}, x_{k,2}, \cdots x_{k,d})^\top \in \mathbb{R}^d$ とし，$\boldsymbol{x}^* = (x_1^*, x_2^*, \cdots, x_d^*)^\top \in \mathbb{R}^d$ とする。

(1) $\boldsymbol{x}_k \to \boldsymbol{x}^* \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall i(x_{k,i} \to x_i^*) \underset{[\text{定理 } 3.3(1)]}{\Rightarrow} \forall i[\forall (x_{k_j,i}) \subset (x_{k,i})\ (x_{k_j,i} \to x_i^*)] \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall (\boldsymbol{x}_{k_j}) \subset (\boldsymbol{x}_k)\ (\boldsymbol{x}_{k_j} \to \boldsymbol{x}^*)$

(2) $\boldsymbol{x}_k \to \boldsymbol{x}^* \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall i(x_{k,i} \to x_i^*) \underset{[\text{定理 } 3.3(2)]}{\Leftrightarrow} \forall i[\forall (x_{k_j,i}) \subset (x_{k,i})\ \exists (x_{k_{j_l},i}) \subset (x_{k_j,i})\ (x_{k_{j_l},i} \to x_i^*)] \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall (\boldsymbol{x}_{k_j}) \subset (\boldsymbol{x}_k)\ \exists (\boldsymbol{x}_{k_{j_l}}) \subset (\boldsymbol{x}_{k_j})\ (\boldsymbol{x}_{k_{j_l}} \to \boldsymbol{x}^*)$

(3) $\boldsymbol{x}_k \to \boldsymbol{x}^* \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall i(x_{k,i} \to x_i^*) \underset{[\text{定理 } 3.3(3)]}{\Leftrightarrow} \forall i[\forall (x_{k_j,i}) \subset (x_{k,i})$ $((x_{k_j,i})$ が収束 $\Rightarrow \lim_{j\to+\infty} x_{k_j,i} = x_i^*)] \underset{[\text{式 } (3.21)]}{\Leftrightarrow} \forall (\boldsymbol{x}_{k_j}) \subset (\boldsymbol{x}_k)$ $((\boldsymbol{x}_{k_j})$ が収束 $\Rightarrow \lim_{j\to+\infty} \boldsymbol{x}_{k_j} = \boldsymbol{x}^*)$

【3.10】 $x_k = \|\boldsymbol{x}_k - \boldsymbol{x}\|$ とする。$x_k \to 0 \Leftrightarrow \forall \varepsilon > 0\ \exists k_0 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \geq k_0 \Rightarrow x_k < \varepsilon \Rightarrow x_k^2 < \varepsilon^2)$ から，$x_k^2 \to 0$ が成り立つ。逆に，$x_k^2 \to 0 \Leftrightarrow \forall \varepsilon > 0\ \exists k_1 \in \mathbb{N}\ \forall k \in \mathbb{N}\ (k \geq k_1 \Rightarrow x_k^2 < \varepsilon \Rightarrow x_k < \sqrt{\varepsilon})$ から，$x_k \to 0$ が成り立つ。

4章

【4.1】 (1) $A \bullet B = \mathrm{Tr}(BA^\top)$ が (a)–(d) を満たすことを以下で確認する。

(a) $A \bullet A = \mathrm{Tr}(AA^\top) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \geq 0$ が成り立つ。$A \bullet A = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = 0 \Leftrightarrow \forall i \forall j (a_{ij}^2 = 0) \Leftrightarrow \forall i \forall j (a_{ij} = 0) \Leftrightarrow A = O$

(b) $(A+B) \bullet C = \mathrm{Tr}(C(A+B)^\top) \underset{[命題\ 4.2(2)]}{=} \mathrm{Tr}(C(A^\top + B^\top)) \underset{[命題\ 4.1(2)]}{=} \mathrm{Tr}(CA^\top + CB^\top) \underset{[命題\ 4.5(1)]}{=} \mathrm{Tr}(CA^\top) + \mathrm{Tr}(CB^\top) = A \bullet C + B \bullet C$

(c) $(\alpha A) \bullet B = \mathrm{Tr}(B(\alpha A)^\top) \underset{[命題\ 4.2(3)]}{=} \mathrm{Tr}(B(\alpha A^\top)) \underset{[命題\ 4.1(1)]}{=} \mathrm{Tr}(\alpha(BA^\top)) \underset{[命題\ 4.5(2)]}{=} \alpha \mathrm{Tr}(BA^\top) = \alpha(A \bullet B)$

(d) $A \bullet B = \mathrm{Tr}(BA^\top) \underset{[命題\ 4.5(4)]}{=} \mathrm{Tr}((BA^\top)^\top) \overset{[命題\ 4.2(1)]}{\underset{[命題\ 4.2(4)]}{=}} \mathrm{Tr}(AB^\top) = B \bullet A$

(2) $\|A\|_{\mathrm{F}} = \sqrt{A \bullet A}$ が (a)–(c) を満たすことを以下で確認する。

(a) $\|A\|_{\mathrm{F}} = \sqrt{A \bullet A} = \sqrt{\mathrm{Tr}(AA^\top)} \geq 0$ が成り立つ。$\|A\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(AA^\top)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} = 0 \Leftrightarrow \forall i \forall j (a_{ij}^2 = 0) \Leftrightarrow \forall i \forall j (a_{ij} = 0) \Leftrightarrow A = O$

(b) $\cdot \bullet \cdot$ が内積の性質 (a)–(d) を満たすこと（上記参照）を利用して，$\|\alpha A\|_{\mathrm{F}}^2 = (\alpha A) \bullet (\alpha A) = \alpha(A \bullet (\alpha A)) = \alpha((\alpha A) \bullet A) = \alpha^2 (A \bullet A) = \alpha^2 \|A\|_{\mathrm{F}}^2$ から，$\|\alpha A\|_{\mathrm{F}} = |\alpha| \|A\|_{\mathrm{F}}$ が成り立つ。

(c) $\|A+B\|_{\mathrm{F}}^2 \underset{[命題\ 4.9(1)]}{=} \|A\|_{\mathrm{F}}^2 + 2A \bullet B + \|B\|_{\mathrm{F}}^2 \underset{[命題\ 4.9(4)]}{\leq} \|A\|_{\mathrm{F}}^2 + 2\|A\|_{\mathrm{F}} \|B\|_{\mathrm{F}} + \|B\|_{\mathrm{F}}^2 = (\|A\|_{\mathrm{F}} + \|B\|_{\mathrm{F}})^2$ から，$\|A+B\|_{\mathrm{F}} \leq \|A\|_{\mathrm{F}} + \|B\|_{\mathrm{F}}$ が成り立つ。

(3) $p \in \{1, 2\} \cup \{+\infty\}$ とする。$\|\boldsymbol{x}\|_p$ が $\mathbb{R}^n$ のノルムであることを利用して，$\|A\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A\boldsymbol{x}\|_p$ が (a)–(c) を満たすことを以下で確認する。

(a) $\|A\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A\boldsymbol{x}\|_p \geq 0$ が成り立つ。$\|A\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A\boldsymbol{x}\|_p = 0 \Rightarrow \forall \boldsymbol{x} (\|\boldsymbol{x}\|_p \leq 1 \Rightarrow \|A\boldsymbol{x}\|_p \leq 0 \Rightarrow A\boldsymbol{x} = \boldsymbol{0}) \Rightarrow A = O$ を得る。$\|O\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|O\boldsymbol{x}\|_p = 0$ から，$A = O \Rightarrow \|A\|_p = 0$ を満たす。

(b) $\|\alpha A\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|(\alpha A)\boldsymbol{x}\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} |\alpha| \|A\boldsymbol{x}\|_p \underset{[命題\ 3.2(2)]}{=} |\alpha| \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A\boldsymbol{x}\|_p = |\alpha| \|A\|_p$

(c) $\|A+B\|_p = \sup_{\|\boldsymbol{x}\|_p \leq 1} \|(A+B)\boldsymbol{x}\|_p \leq \sup_{\|\boldsymbol{x}\|_p \leq 1} (\|A\boldsymbol{x}\|_p + \|B\boldsymbol{x}\|_p) \underset{[命題\ 3.2(1)]}{\leq} \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A\boldsymbol{x}\|_p + \sup_{\|\boldsymbol{x}\|_p \leq 1} \|B\boldsymbol{x}\|_p = \|A\|_p + \|B\|_p$

(4) $A \in \mathbb{R}^{m \times n}$ に対して，$\|A\|_1 = \max_{j \in [n]} \sum_{i=1}^m |a_{ij}|$ を示す。ただし，$[n] = \{1, 2, \cdots, n\}$ とする。$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^\top \in \mathbb{R}^n$ に対して，

$$A\boldsymbol{x} = \left(\sum_{j=1}^n a_{1j} x_j, \sum_{j=1}^n a_{2j} x_j, \cdots, \sum_{j=1}^n a_{mj} x_j \right)^\top \in \mathbb{R}^m$$

から，

$$\begin{aligned}
\|A\boldsymbol{x}\|_1 &\underset{[式\ (3.18)]}{=} \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij} x_j \right| \\
&\underset{[三角不等式]}{\leq} \sum_{i=1}^m \sum_{j=1}^n |a_{ij}| |x_j| \\
&= \sum_{j=1}^n |x_j| \sum_{i=1}^m |a_{ij}| \\
&\leq \sum_{j=1}^n |x_j| \max_{k \in [n]} \sum_{i=1}^m |a_{ik}| \\
&= \left(\max_{k \in [n]} \sum_{i=1}^m |a_{ik}| \right) \|\boldsymbol{x}\|_1
\end{aligned}$$

を得る。以上のことから，

$$\|A\|_1 \underset{[命題\ 4.10(2)]}{=} \max_{\|\boldsymbol{x}\|_1 = 1} \|A\boldsymbol{x}\|_1 \leq \max_{k \in [n]} \sum_{i=1}^m |a_{ik}| \tag{11}$$

を満たす。一方で，$\sum_{i=1}^{m} |a_{ik^*}| = \max_{k\in[n]} \sum_{i=1}^{m} |a_{ik}|$ となる $k^* \in [n]$ が存在するので，$\boldsymbol{x} = \boldsymbol{e}_{k^*}$ とすると，$\|\boldsymbol{x}\|_1 = \|\boldsymbol{e}_{k^*}\|_1 = 1$ と

$$A\boldsymbol{x} = A\boldsymbol{e}_{k^*} = (a_{1k^*}, a_{2k^*}, \cdots, a_{mk^*})^\top \in \mathbb{R}^m$$

から，

$$\|A\|_1 \underset{[\text{命題 } 4.10(2)]}{=} \max_{\|\boldsymbol{x}\|_1=1} \|A\boldsymbol{x}\|_1 \geq \|A\boldsymbol{e}_{k^*}\|_1 = \sum_{i=1}^{m} |a_{ik^*}| \tag{12}$$

を得る。式 (11) と式 (12) から，$\|A\|_1 = \max_{j\in[n]} \sum_{i=1}^{m} |a_{ij}|$ が成り立つ。

$A \in \mathbb{R}^{m\times n}$ に対して，$\|A\|_\infty = \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}|$ を示す。$A = O$ のときは，$\|A\|_\infty \underset{[\text{式 } (3.19)]}{=} \max_{\|\boldsymbol{x}\|_\infty=1} \|A\boldsymbol{x}\|_\infty = 0 = \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}|$ から成り立つ。以降では，$A \neq O$ とする。$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^\top \in \mathbb{R}^n$ に対して，

$$\begin{aligned}
\|A\boldsymbol{x}\|_\infty &\underset{[\text{式 } (3.19)]}{=} \max_{i\in[m]} \left| \sum_{j=1}^{n} a_{ij}x_j \right| \\
&\underset{[\text{三角不等式}]}{\leq} \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}||x_j| \\
&\leq \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}| \max_{k\in[n]} |x_k| \\
&= \left(\max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}| \right) \|\boldsymbol{x}\|_\infty
\end{aligned}$$

を得る。よって，

$$\|A\|_\infty \underset{[\text{命題 } 4.10(2)]}{=} \max_{\|\boldsymbol{x}\|_\infty=1} \|A\boldsymbol{x}\|_\infty \leq \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}| \tag{13}$$

を満たす。一方で，$\sum_{j=1}^{n} |a_{i^*j}| = \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}|$ となる $i^* \in [m]$ が存在する。$A \neq O$ に注意して，任意の $j \in [n]$ に対して，

$$u_j = \begin{cases} \frac{a_{i^*j}}{|a_{i^*j}|} & (a_{i^*j} \neq 0) \\ 1 & (a_{i^*j} = 0) \end{cases}$$

で定義される u_j を第 j 成分にもつベクトル $\boldsymbol{u} = (u_1, u_2, \cdots, u_n)^\top \in \mathbb{R}^n$ は，$\|\boldsymbol{u}\|_\infty = 1$，$a_{i^*j}u_j = |a_{i^*j}|$，および，

$$A\boldsymbol{u} = \left(\sum_{j=1}^{n} a_{1j}u_j, \sum_{j=1}^{n} a_{2j}u_j, \cdots, \sum_{j=1}^{n} a_{mj}u_j \right)^\top \in \mathbb{R}^m$$

を満たすことから，

$$\begin{aligned}
\|A\|_\infty &\underset{[\text{命題 } 4.10(2)]}{=} \max_{\|\boldsymbol{x}\|_\infty=1} \|A\boldsymbol{x}\|_\infty \geq \|A\boldsymbol{u}\|_\infty \\
&\underset{[\text{式 } (3.19)]}{=} \max_{i\in[m]} \left| \sum_{j=1}^{n} a_{ij}u_j \right| \geq \left| \sum_{j=1}^{n} a_{i^*j}u_j \right| = \sum_{j=1}^{n} |a_{i^*j}|
\end{aligned} \tag{14}$$

を得る。式 (13) と式 (14) から，$\|A\|_\infty = \max_{i\in[m]} \sum_{j=1}^{n} |a_{ij}|$ が成り立つ。

(5) $A \in \mathbb{O}^d \Leftrightarrow A^\top A = I \Rightarrow \mathrm{Det}(A)^2 \underset{[\text{命題 } 4.6(2)]}{=} \mathrm{Det}(A^\top)\mathrm{Det}(A) \underset{[\text{命題 } 4.6(3)]}{=} \mathrm{Det}(A^\top A) = 1$ から，$\mathrm{Det}(A) = \pm 1$ となる。

$A \in \mathbb{O}^d$ とする。$(A^\top)^\top A^\top \underset{[命題\ 4.2(1)]}{=} AA^\top \underset{[A\in\mathbb{O}^d]}{=} I$ から，$A^\top \in \mathbb{O}^d$ である。$(A^{-1})^\top A^{-1} \underset{[命題\ 4.4(3)]}{=} (A^\top)^{-1}A^{-1} \underset{[命題\ 4.4(2)]}{=} (AA^\top)^{-1} \underset{[A\in\mathbb{O}^d]}{=} I^{-1} = I$ から，$A^{-1} \in \mathbb{O}^d$ である。

$A, B \in \mathbb{O}^d \Rightarrow (AB)^\top(AB) \underset{[命題\ 4.2(4)]}{=} (B^\top A^\top)(AB) \underset{[命題\ 4.1(1)]}{=} B^\top(A^\top A)B \underset{[A\in\mathbb{O}^d]}{=} B^\top B \underset{[B\in\mathbb{O}^d]}{=} I$ から，$AB \in \mathbb{O}^d$ である。

(6) $\|AB\|_{\mathrm{F}} \leq \|A\|_{\mathrm{F}}\|B\|_{\mathrm{F}}$ を示す。$A = (a_{ij}) \in \mathbb{R}^{l\times m}$ と $B = (b_{jk}) \in \mathbb{R}^{m\times n}$ に対して，$AB = (\sum_{j=1}^m a_{ij}b_{jk}) \in \mathbb{R}^{l\times n}$ となる。任意の $i \in \{1, 2, \cdots, l\}$ に対して，$\boldsymbol{a}_i = (a_{i1}, a_{i2}, \cdots, a_{im})^\top \in \mathbb{R}^m$ とし，任意の $k \in \{1, 2, \cdots, n\}$ に対して，$\boldsymbol{b}_k = (b_{1k}, b_{2k}, \cdots, b_{mk})^\top \in \mathbb{R}^m$ とする。このとき，以下が成り立つ。

$$
\begin{aligned}
\|AB\|_{\mathrm{F}}^2 &\underset{[式\ (4.23)]}{=} \sum_{i=1}^{l}\sum_{k=1}^{n}\left(\sum_{j=1}^{m} a_{ij}b_{jk}\right)^2 \\
&\underset{[式\ (3.11)]}{=} \sum_{i=1}^{l}\sum_{k=1}^{n}\langle \boldsymbol{a}_i, \boldsymbol{b}_k\rangle_2^2 \\
&\underset{[命題\ 3.7(4)]}{\leq} \sum_{i=1}^{l}\sum_{k=1}^{n}\|\boldsymbol{a}_i\|_2^2\|\boldsymbol{b}_k\|_2^2 \\
&\underset{[式\ (3.15)]}{=} \sum_{i=1}^{l}\sum_{k=1}^{n}\left(\sum_{j=1}^{m} a_{ij}^2\right)\left(\sum_{j=1}^{m} b_{jk}^2\right) \\
&= \sum_{i=1}^{l}\left(\sum_{j=1}^{m} a_{ij}^2\right)\sum_{k=1}^{n}\left(\sum_{j=1}^{m} b_{jk}^2\right) \\
&\underset{[式\ (4.23)]}{=} \|A\|_{\mathrm{F}}^2\|B\|_{\mathrm{F}}^2
\end{aligned}
$$

$\|AB\|_p \leq \|A\|_p\|B\|_p$ を示す。作用素ノルムの定義 (4.27) から，$\forall \boldsymbol{y} \in \mathbb{R}^m(\|A\boldsymbol{y}\|_p \leq \|A\|_p\|\boldsymbol{y}\|_p)$，かつ，$\forall \boldsymbol{x} \in \mathbb{R}^n(\|B\boldsymbol{x}\|_p \leq \|B\|_p\|\boldsymbol{x}\|_p)$ なので，

$$
\|(AB)(\boldsymbol{x})\|_p = \|A(B\boldsymbol{x})\|_p \leq \|A\|_p\|B\boldsymbol{x}\|_p \leq \|A\|_p\|B\|_p\|\boldsymbol{x}\|_p
$$

つまり，

$$
\|AB\|_p = \sup_{\boldsymbol{x}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\|AB\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \|A\|_p\|B\|_p
$$

となる。

5章

【5.1】 $\boldsymbol{x} \in D_1 \cap D_2$ とする。このとき，f_1 と f_2 が $\boldsymbol{x}$ で微分可能なので，

$$\forall i \in \{1,2\} \quad \left(F_i(\boldsymbol{h}) = \frac{f_i(\boldsymbol{x}+\boldsymbol{h}) - f_i(\boldsymbol{x}) - \langle \nabla f_i(\boldsymbol{x}), \boldsymbol{h} \rangle}{\|\boldsymbol{h}\|} \underset{[\text{定義 } 5.2]}{\to} 0 \right)$$

となる。$(\alpha f_1 + \beta f_2)(\boldsymbol{x}) = \alpha f_1(\boldsymbol{x}) + \beta f_2(\boldsymbol{x})$ と内積の性質から，

$$\begin{aligned} &\frac{(\alpha f_1 + \beta f_2)(\boldsymbol{x}+\boldsymbol{h}) - (\alpha f_1 + \beta f_2)(\boldsymbol{x}) - \langle \alpha \nabla f_1(\boldsymbol{x}) + \beta \nabla f_2(\boldsymbol{x}), \boldsymbol{h} \rangle}{\|\boldsymbol{h}\|} \\ &\underset{(\mathrm{I3})}{\overset{(\mathrm{I2})}{=}} \alpha F_1(\boldsymbol{h}) + \beta F_2(\boldsymbol{h}) \to 0 \ (\boldsymbol{h} \to \boldsymbol{0}) \end{aligned} \tag{15}$$

を満たすので，$\alpha f_1 + \beta f_2$ は $\boldsymbol{x}$ で微分可能となり，$\nabla(\alpha f_1 + \beta f_2)(\boldsymbol{x})$ が存在する。さらに，式 (15) から，

$$\nabla(\alpha f_1 + \beta f_2)(\boldsymbol{x}) = \alpha \nabla f_1(\boldsymbol{x}) + \beta \nabla f_2(\boldsymbol{x}) = (\alpha \nabla f_1 + \beta \nabla f_2)(\boldsymbol{x})$$

が成り立つ。

【5.2】 $\boldsymbol{d} \neq \boldsymbol{0}$ とし，$g(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{d})$ とする。$\boldsymbol{x} + \alpha \boldsymbol{d} \in D$ となる $\alpha \geq 0$ に対して，

$$\begin{aligned} g'(\alpha) &= \lim_{\beta \downarrow 0} \frac{g(\alpha+\beta) - g(\alpha)}{\beta} \\ &= \lim_{\beta \downarrow 0} \frac{f(\boldsymbol{x} + (\alpha+\beta)\boldsymbol{d}) - f(\boldsymbol{x}+\alpha\boldsymbol{d})}{\beta} \\ &= \lim_{\beta \downarrow 0} \frac{f((\boldsymbol{x}+\alpha\boldsymbol{d}) + \beta\boldsymbol{d}) - f(\boldsymbol{x}+\alpha\boldsymbol{d})}{\beta} \\ &= f'(\boldsymbol{x}+\alpha\boldsymbol{d}; \boldsymbol{d}) \\ &\underset{[\text{式 } (5.6)]}{=} \langle \nabla f(\boldsymbol{x}+\alpha\boldsymbol{d}), \boldsymbol{d} \rangle \end{aligned}$$

となる。$\boldsymbol{d} = \boldsymbol{0}$ のとき，$g'(\alpha) = f'(\boldsymbol{x}+\alpha\boldsymbol{d}; \boldsymbol{d}) = 0 = \langle \nabla f(\boldsymbol{x}+\alpha\boldsymbol{d}), \boldsymbol{d} \rangle$ となる。

【5.3】 $\boldsymbol{x} \in D_1 \cap D_2$ とする。このとき，f_1 と f_2 が $\boldsymbol{x}$ で 2 回微分可能なので，

$$\forall i \in \{1,2\} \quad \left(F_i(\boldsymbol{h}) = \frac{\nabla f_i(\boldsymbol{x}+\boldsymbol{h}) - \nabla f_i(\boldsymbol{x}) - \nabla^2 f_i(\boldsymbol{x})\boldsymbol{h}}{\|\boldsymbol{h}\|} \underset{[\text{定義 } 5.6]}{\to} \boldsymbol{0} \right)$$

となる。$(\alpha \nabla f_1 + \beta \nabla f_2)(\boldsymbol{x}) = \alpha \nabla f_1(\boldsymbol{x}) + \beta \nabla f_2(\boldsymbol{x})$（命題 5.2；【5.1】参照）と行列の線形性から，

$$\begin{aligned} &\frac{(\alpha \nabla f_1 + \beta \nabla f_2)(\boldsymbol{x}+\boldsymbol{h}) - (\alpha \nabla f_1 + \beta \nabla f_2)(\boldsymbol{x})}{\|\boldsymbol{h}\|} - \frac{(\alpha \nabla^2 f_1(\boldsymbol{x}) + \beta \nabla^2 f_2(\boldsymbol{x}))\boldsymbol{h}}{\|\boldsymbol{h}\|} \\ &= \alpha F_1(\boldsymbol{h}) + \beta F_2(\boldsymbol{h}) \to \boldsymbol{0} \ (\boldsymbol{h} \to \boldsymbol{0}) \end{aligned}$$

を満たすので，$\alpha f_1 + \beta f_2$ は $\boldsymbol{x}$ で 2 回微分可能となり，$\nabla^2(\alpha f_1 + \beta f_2)(\boldsymbol{x})$ が存在する。さらに，

$$\nabla^2(\alpha f_1 + \beta f_2)(\boldsymbol{x}) = \alpha \nabla^2 f_1(\boldsymbol{x}) + \beta \nabla^2 f_2(\boldsymbol{x}) = (\alpha \nabla^2 f_1 + \beta \nabla^2 f_2)(\boldsymbol{x})$$

が成り立つ。

【5.4】 (1) $g_n = \sum_{k=1}^n \alpha_k f_k$ とする。任意の $\lambda \in [0,1]$ と任意の $\boldsymbol{x}, \boldsymbol{y} \in D$ に対して，

$$\begin{aligned} g_n(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \sum_{k=1}^n \alpha_k f_k(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &\underset{[f_k \text{は凸}]}{\leq} \sum_{k=1}^n \alpha_k \{\lambda f_k(\boldsymbol{x}) + (1-\lambda) f_k(\boldsymbol{y})\} \\ &\underset{[\Sigma\text{の線形性}]}{=} \lambda \sum_{k=1}^n \alpha_k f_k(\boldsymbol{x}) + (1-\lambda) \sum_{k=1}^n \alpha_k f_k(\boldsymbol{y}) \\ &= \lambda g_n(\boldsymbol{x}) + (1-\lambda) g_n(\boldsymbol{y}) \end{aligned}$$

から，g_n は凸となる。

(2) $f(\boldsymbol{x}) = \sup_{k\in\mathbb{N}} f_k(\boldsymbol{x})$ とする。任意の $\lambda \in [0,1]$ と任意の $\boldsymbol{x}, \boldsymbol{y} \in D$ に対して，

$$\begin{aligned} f(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= \sup_{k\in\mathbb{N}} f_k(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \\ &\underset{[f_k\text{は凸}]}{\le} \sup_{k\in\mathbb{N}}\{\lambda f_k(\boldsymbol{x}) + (1-\lambda) f_k(\boldsymbol{y})\} \\ &\underset{[\text{命題 } 3.2]}{\le} \lambda \sup_{k\in\mathbb{N}} f_k(\boldsymbol{x}) + (1-\lambda)\sup_{k\in\mathbb{N}} f_k(\boldsymbol{y}) \\ &= \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \end{aligned}$$

から，f は凸となる。

(3) $f(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$ とし，$(g \circ f)(\boldsymbol{x}) = g(A\boldsymbol{x} - \boldsymbol{b})$ とする。$A \in \mathbb{R}^{m\times n}$ の線形性と例 3.6(2) の議論から，任意の $\lambda \in [0,1]$ と任意の $\boldsymbol{x}, \boldsymbol{y} \in D$ に対して，

$$\begin{aligned} A(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) - \boldsymbol{b} &= \lambda A\boldsymbol{x} + (1-\lambda)A\boldsymbol{y} - (\lambda\boldsymbol{b} + (1-\lambda)\boldsymbol{b}) \\ &= \lambda(A\boldsymbol{x} - \boldsymbol{b}) + (1-\lambda)(A\boldsymbol{y} - \boldsymbol{b}) \end{aligned}$$

となるので，g の凸性から，

$$\begin{aligned} (g\circ f)(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= g(A(\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y}) - \boldsymbol{b}) \\ &= \lambda g(A\boldsymbol{x} - \boldsymbol{b}) + (1-\lambda) g(A\boldsymbol{y} - \boldsymbol{b}) \\ &= \lambda(g\circ f)(\boldsymbol{x}) + (1-\lambda)(g\circ f)(\boldsymbol{y}) \end{aligned}$$

を満たすので，$g \circ f$ は凸関数となる。

【5.5】 $\partial(\alpha f_1)(\boldsymbol{x}) = \alpha\partial f_1(\boldsymbol{x})$ を示す。

$$\begin{aligned} \partial(\alpha f_1)(\boldsymbol{x}) &= \{\boldsymbol{g} \in \mathbb{R}^d \colon \forall \boldsymbol{y}((\alpha f_1)(\boldsymbol{y}) \ge (\alpha f_1)(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x}\rangle)\} \\ &= \{\boldsymbol{g} \in \mathbb{R}^d \colon \forall \boldsymbol{y}(\alpha f_1(\boldsymbol{y}) \ge \alpha f_1(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x}\rangle)\} \\ &= \left\{\boldsymbol{g} \in \mathbb{R}^d \colon \forall \boldsymbol{y}\left(f_1(\boldsymbol{y}) \ge f_1(\boldsymbol{x}) + \left\langle \frac{\boldsymbol{g}}{\alpha}, \boldsymbol{y} - \boldsymbol{x}\right\rangle\right)\right\} \end{aligned}$$

から，

$$\boldsymbol{g} \in \partial(\alpha f_1)(\boldsymbol{x}) \Leftrightarrow \frac{\boldsymbol{g}}{\alpha} \in \partial f_1(\boldsymbol{x}) \Rightarrow \boldsymbol{g} = \alpha\frac{\boldsymbol{g}}{\alpha} \in \alpha\partial f_1(\boldsymbol{x})$$

つまり，$\partial(\alpha f_1)(\boldsymbol{x}) \subset \alpha\partial f_1(\boldsymbol{x})$ となる。逆に，

$$\begin{aligned} \boldsymbol{g} \in \alpha\partial f_1(\boldsymbol{x}) &\Rightarrow \exists\hat{\boldsymbol{g}} \in \partial f_1(\boldsymbol{x})\ (\boldsymbol{g} = \alpha\hat{\boldsymbol{g}}) \Rightarrow \hat{\boldsymbol{g}} = \frac{\boldsymbol{g}}{\alpha} \in \partial f_1(\boldsymbol{x}) \\ &\Leftrightarrow \boldsymbol{g} \in \partial(\alpha f_1)(\boldsymbol{x}) \end{aligned}$$

つまり，$\partial(\alpha f_1)(\boldsymbol{x}) \supset \alpha\partial f_1(\boldsymbol{x})$ が成り立つ。以上のことから，$\partial(\alpha f_1)(\boldsymbol{x}) = \alpha\partial f_1(\boldsymbol{x})$ を得る。

$\partial f_1(\boldsymbol{x}) + \partial f_2(\boldsymbol{x}) \subset \partial(f_1 + f_2)(\boldsymbol{x})$ を示す。

$$\boldsymbol{g} \in \partial f_1(\boldsymbol{x}) + \partial f_2(\boldsymbol{x}) \Leftrightarrow \exists\boldsymbol{g}_1 \in \partial f_1(\boldsymbol{x})\ \exists\boldsymbol{g}_2 \in \partial f_2(\boldsymbol{x})\ (\boldsymbol{g} = \boldsymbol{g}_1 + \boldsymbol{g}_2)$$

から，

$$\begin{aligned} &\forall \boldsymbol{y}(f_1(\boldsymbol{y}) \ge f_1(\boldsymbol{x}) + \langle \boldsymbol{g}_1, \boldsymbol{y} - \boldsymbol{x}\rangle \wedge f_2(\boldsymbol{y}) \ge f_2(\boldsymbol{x}) + \langle \boldsymbol{g}_2, \boldsymbol{y} - \boldsymbol{x}\rangle \\ &\Rightarrow (f_1 + f_2)(\boldsymbol{y}) \ge (f_1 + f_2)(\boldsymbol{x}) + \langle \boldsymbol{g}_1 + \boldsymbol{g}_2, \boldsymbol{y} - \boldsymbol{x}\rangle) \end{aligned}$$

つまり，$\boldsymbol{g} = \boldsymbol{g}_1 + \boldsymbol{g}_2 \in \partial(f_1 + f_2)(\boldsymbol{x})$ を得る。

6 章

【6.1】 $\mathrm{P}(A_1 \cup A_2) \underset{[命題\,6.1(4)]}{=} \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2)$ が成り立つことを利用して，

$$\begin{aligned}
\mathrm{P}(A_1 \cup A_2 \cup A_3) &= \mathrm{P}((A_1 \cup A_2) \cup A_3) \\
&= \mathrm{P}(A_1 \cup A_2) + \mathrm{P}(A_3) - \mathrm{P}((A_1 \cup A_2) \cap A_3) \\
&= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) - \mathrm{P}(A_1 \cap A_2) - \underbrace{\mathrm{P}((A_1 \cup A_2) \cap A_3)}_{p}
\end{aligned}$$

および，

$$\begin{aligned}
p &= \mathrm{P}((A_1 \cap A_3) \cup (A_2 \cap A_3)) \\
&= \mathrm{P}(A_1 \cap A_3) + \mathrm{P}(A_2 \cap A_3) - \mathrm{P}((A_1 \cap A_3) \cap (A_2 \cap A_3)) \\
&= \mathrm{P}(A_1 \cap A_3) + \mathrm{P}(A_2 \cap A_3) - \mathrm{P}(A_1 \cap A_2 \cap A_3)
\end{aligned}$$

から，

$$\begin{aligned}
\mathrm{P}(A_1 \cup A_2 \cup A_3) &= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) \\
&\quad - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_3)
\end{aligned} \tag{16}$$

が成り立つ。よって，

$$\begin{aligned}
\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4) &= \mathrm{P}((A_1 \cup A_2 \cup A_3) \cup A_4) \\
&= \mathrm{P}(A_1 \cup A_2 \cup A_3) + \mathrm{P}(A_4) - \mathrm{P}((A_1 \cup A_2 \cup A_3) \cap A_4) \\
&= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \mathrm{P}(A_4) \\
&\quad - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_3) \\
&\quad - \underbrace{\mathrm{P}((A_1 \cup A_2 \cup A_3) \cap A_4)}_{q}
\end{aligned}$$

および，式 (16) から得られる等式

$$\begin{aligned}
q &= \mathrm{P}((A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_3 \cap A_4)) \\
&= \mathrm{P}(A_1 \cap A_4) + \mathrm{P}(A_2 \cap A_4) + \mathrm{P}(A_3 \cap A_4) \\
&\quad - \mathrm{P}((A_1 \cap A_4) \cap (A_2 \cap A_4)) - \mathrm{P}((A_1 \cap A_4) \cap (A_3 \cap A_4)) \\
&\quad - \mathrm{P}((A_2 \cap A_4) \cap (A_3 \cap A_4)) \\
&\quad + \mathrm{P}((A_1 \cap A_4) \cap (A_2 \cap A_4) \cap (A_3 \cap A_4)) \\
&= \mathrm{P}(A_1 \cap A_4) + \mathrm{P}(A_2 \cap A_4) + \mathrm{P}(A_3 \cap A_4) \\
&\quad - \mathrm{P}(A_1 \cap A_4 \cap A_2) - \mathrm{P}(A_1 \cap A_4 \cap A_3) \\
&\quad - \mathrm{P}(A_2 \cap A_4 \cap A_3) + \mathrm{P}(A_1 \cap A_4 \cap A_2 \cap A_3)
\end{aligned}$$

により，以下を得る。

$$\begin{aligned}
\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4) &= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \mathrm{P}(A_4) \\
&\quad - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_3) \\
&\quad - \{\mathrm{P}(A_1 \cap A_4) + \mathrm{P}(A_2 \cap A_4) + \mathrm{P}(A_3 \cap A_4) \\
&\quad - \mathrm{P}(A_1 \cap A_4 \cap A_2) - \mathrm{P}(A_1 \cap A_4 \cap A_3) \\
&\quad - \mathrm{P}(A_2 \cap A_4 \cap A_3) + \mathrm{P}(A_1 \cap A_4 \cap A_2 \cap A_3)\} \\
&= \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) + \mathrm{P}(A_4) \\
&\quad - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_1 \cap A_4) - \mathrm{P}(A_2 \cap A_3)
\end{aligned}$$

$$
\begin{aligned}
&- \mathrm{P}(A_2 \cap A_4) - \mathrm{P}(A_3 \cap A_4) \\
&+ \mathrm{P}(A_1 \cap A_2 \cap A_3) + \mathrm{P}(A_1 \cap A_2 \cap A_4) + \mathrm{P}(A_1 \cap A_3 \cap A_4) \\
&+ \mathrm{P}(A_2 \cap A_3 \cap A_4) - \mathrm{P}(A_1 \cap A_2 \cap A_3 \cap A_4)
\end{aligned}
$$

【6.2】 数学的帰納法（1.4.4 項）を用いて証明する。乗法定理（命題 6.2(1)）から，

$$\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2|A_1)$$

を得るので，$n = 1$ のときの式 (6.5) が成り立つ。$n = k$ のときの式 (6.5)，つまり，

$$\mathrm{P}\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{j=1}^{k} \mathrm{P}\left(A_j \middle| \bigcap_{i=1}^{j-1} A_i\right)$$

が成り立つとする。このとき，

$$
\begin{aligned}
\mathrm{P}\left(\bigcap_{i=1}^{k+1} A_i\right) &= \mathrm{P}\left(\bigcap_{i=1}^{k} A_i \cap A_{k+1}\right) \\
&\underset{[\text{命題 } 6.2(1)]}{=} \mathrm{P}\left(\bigcap_{i=1}^{k} A_i\right) \mathrm{P}\left(A_{k+1} \middle| \bigcap_{i=1}^{k} A_i\right) \\
&\underset{[\text{前提}]}{=} \prod_{j=1}^{k} \mathrm{P}\left(A_j \middle| \bigcap_{i=1}^{j-1} A_i\right) \mathrm{P}\left(A_{k+1} \middle| \bigcap_{i=1}^{k} A_i\right) \\
&= \prod_{j=1}^{k+1} \mathrm{P}\left(A_j \middle| \bigcap_{i=1}^{j-1} A_i\right)
\end{aligned}
$$

により，$n = k + 1$ のときの式 (6.5) が成り立つ。以上のことから，式 (6.5) が任意の $n \in \mathbb{N}$ に対して成り立つ。

【6.3】 (1) $\mathbb{E}_X[X + c] = \mathbb{E}_X[X] + c$ を示す。X が離散型確率変数のとき，

$$
\begin{aligned}
\mathbb{E}_X[X + c] &\underset{[\text{式 } (6.20)]}{=} \sum_{i=1}^{+\infty} (x_i + c)\mathrm{P}(X = x_i) \\
&= \sum_{i=1}^{+\infty} x_i \mathrm{P}(X = x_i) + c \underbrace{\sum_{i=1}^{+\infty} \mathrm{P}(X = x_i)}_{=1 \because (6.7)} \\
&= \mathbb{E}_X[X] + c
\end{aligned}
$$

が成り立つ。X が連続型確率変数のとき，

$$
\begin{aligned}
\mathbb{E}_X[X + c] &\underset{[\text{式 } (6.20)]}{=} \int_{-\infty}^{+\infty} (x + c) f(x) \mathrm{d}x \\
&= \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x + c \underbrace{\int_{-\infty}^{+\infty} f(x) \mathrm{d}x}_{=1 \because (6.8)} \\
&= \mathbb{E}_X[X] + c
\end{aligned}
$$

が成り立つ。

$\mathbb{E}_X[cX] = c\mathbb{E}_X[X]$ を示す。同様の議論から，X が離散型確率変数のときは，

$$\mathbb{E}_X[cX] \underset{[\text{式 } (6.20)]}{=} c \sum_{i=1}^{+\infty} x_i \mathrm{P}(X = x_i) = c\mathbb{E}_X[X]$$

X が連続型確率変数のときは，

$$\mathbb{E}_X[cX] \underset{[\text{式 } (6.20)]}{=} c \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x = c\mathbb{E}_X[X]$$

が成り立つ。

(2) $\mathbb{V}_X[X+c] = \mathbb{V}_X[X]$ を示す。$\mathbb{E}_X[X+c] \underset{\text{【6.3】(1)}}{=} \mathbb{E}_X[X] + c$ が成り立つ。$\mu = \mathbb{E}_X[X]$ とする。このとき，以下が成り立つ。

$$\begin{aligned}
\mathrm{V}_X[X+c] &= \mathbb{E}_X\left[|(X+c) - \mathbb{E}_X[X+c]|^2\right] \\
&= \mathbb{E}_X\left[|(X+c) - (\mu + c)|^2\right] \\
&= \mathbb{E}_X\left[|X-\mu|^2\right] \\
&= \mathrm{V}_X[X]
\end{aligned}$$

$\mathbb{V}_X[cX] = c^2\mathbb{V}_X[X]$ を示す。$\mathbb{E}_X[cX] \underset{\text{【6.3】(1)}}{=} c\mathbb{E}_X[X]$ が成り立つ。$\mu = \mathbb{E}_X[X]$ とする。このとき，以下が成り立つ。

$$\begin{aligned}
\mathrm{V}_X[cX] &= \mathbb{E}_X\left[|cX - \mathbb{E}_X[cX]|^2\right] \\
&= \mathbb{E}_X\left[|cX - c\mu|^2\right] \\
&= \mathbb{E}_X\left[c^2|X-\mu|^2\right] \\
&= c^2\mathbb{E}_X\left[|X-\mu|^2\right] \\
&= c^2\mathrm{V}_X[X]
\end{aligned}$$

【6.4】 $\boldsymbol{v} \in \mathbb{R}^d$ とする。$\boldsymbol{\mu} = \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}]$ とする。$\mathrm{Cov}[\boldsymbol{X}] \underset{\text{[注意 6.4]}}{=} \mathbb{E}_{\boldsymbol{X}}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top]$ から，

$$\begin{aligned}
\langle \boldsymbol{v}, \mathrm{Cov}[\boldsymbol{X}]\boldsymbol{v}\rangle_2 &= \left\langle \boldsymbol{v}, \mathbb{E}_{\boldsymbol{X}}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\right]\boldsymbol{v}\right\rangle_2 \\
&\underset{\text{[命題 6.3(4)]}}{=} \mathbb{E}_{\boldsymbol{X}}\left[\left\langle \boldsymbol{v}, (\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{v}\right\rangle_2\right] \\
&\underset{\text{[命題 4.3(1)]}}{=} \mathbb{E}_{\boldsymbol{X}}\left[\left\langle (\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{v}, (\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{v}\right\rangle_2\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\left\|(\boldsymbol{X}-\boldsymbol{\mu})^\top\boldsymbol{v}\right\|_2^2\right] \\
&\geq 0
\end{aligned}$$

を得る。定理 4.2(1) から，$\mathrm{Cov}[\boldsymbol{X}] \in \mathbb{S}_+^d$ となる。

【6.5】 (1) $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)^\top$ とし，$\boldsymbol{\mu}_{\boldsymbol{X}} = (\mu_1, \mu_2, \cdots, \mu_d)^\top = \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}]$ とする。このとき，以下が成り立つ。

$$\begin{aligned}
\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}] &= \mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}\|_2^2\right] \\
&\underset{\text{[注意 6.4]}}{=} \mathbb{E}_{\boldsymbol{X}}\left[\sum_{i=1}^d (X_i - \mu_i)^2\right] \\
&\underset{\text{[命題 6.3(1)]}}{=} \sum_{i=1}^d \mathbb{E}_{X_i}\left[(X_i-\mu_i)^2\right] \\
&= \sum_{i=1}^d \mathbb{E}_{X_i}\left[(X_i - \mathbb{E}_{X_i}[X_i])^2\right] \\
&= \sum_{i=1}^d \mathbb{V}_{X_i}[X_i]
\end{aligned}$$

(2) $\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}]$，$\boldsymbol{\mu}_{\boldsymbol{Y}} = \mathbb{E}_{\boldsymbol{Y}}[\boldsymbol{Y}]$ とする。$\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}[\boldsymbol{X}+\boldsymbol{Y}] \underset{\text{[命題 6.3(3)]}}{=} \boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{\mu}_{\boldsymbol{Y}}$ から，

$$\begin{aligned}
\mathbb{V}_{(\boldsymbol{X},\boldsymbol{Y})}[\boldsymbol{X}+\boldsymbol{Y}] &= \mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\|(\boldsymbol{X}+\boldsymbol{Y}) - \mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}[\boldsymbol{X}+\boldsymbol{Y}]\|_2^2\right] \\
&= \mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\|(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}) + (\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}})\|_2^2\right] \\
&= \mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}\|_2^2 + \|\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}}\|_2^2 + 2\langle \boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}}\rangle_2\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}\|_2^2\right] + \mathbb{E}_{\boldsymbol{Y}}\left[\|\boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}}\|_2^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+2\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\langle \boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}}\rangle_2\right] \\
&= \mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}] + \mathbb{V}_{\boldsymbol{Y}}[\boldsymbol{Y}] + 2\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\langle \boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}, \boldsymbol{Y}-\boldsymbol{\mu}_{\boldsymbol{Y}}\rangle_2\right]
\end{aligned}
$$

が成り立つ。

(3) $\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}]$，$\boldsymbol{c} \in \mathbb{R}^d$ とする。$\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}] \underset{[\text{命題 } 6.3(3)]}{=} \boldsymbol{\mu}_{\boldsymbol{X}} + \boldsymbol{c}$ から，以下を得る。

$$
\begin{aligned}
\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}] &= \mathbb{E}_{\boldsymbol{X}}\left[\|(\boldsymbol{X}+\boldsymbol{c}) - \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}]\|_2^2\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}}\|_2^2\right] \\
&= \mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}]
\end{aligned}
$$

(4) $\boldsymbol{X}$ が離散型確率変数のときを考察する（$\boldsymbol{X}$ が連続型のときも以下と同様の議論から結論を得る）。このとき，

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{X}+\boldsymbol{Z}|\boldsymbol{y}] &= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \sum_{\boldsymbol{z}\in \mathrm{R}(\boldsymbol{Z})} (\boldsymbol{x}+\boldsymbol{z})\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}\cap \boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \sum_{\boldsymbol{z}\in \mathrm{R}(\boldsymbol{Z})} \boldsymbol{x}\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}\cap \boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y}) \\
&\quad + \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \sum_{\boldsymbol{z}\in \mathrm{R}(\boldsymbol{Z})} \boldsymbol{z}\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}\cap \boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \boldsymbol{x} \underbrace{\sum_{\boldsymbol{z}\in \mathrm{R}(\boldsymbol{Z})} \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}\cap \boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y})}_{\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y})\ \because(6.13)} \\
&\quad + \sum_{\boldsymbol{z}\in \mathrm{R}(\boldsymbol{Z})} \boldsymbol{z} \underbrace{\sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}\cap \boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y})}_{\mathrm{P}(\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{y})\ \because(6.13)} \\
&= \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}] + \mathbb{E}_{\boldsymbol{Z}}[\boldsymbol{Z}|\boldsymbol{y}]
\end{aligned}
$$

および，

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}|\boldsymbol{y}] &= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} (\boldsymbol{x}+\boldsymbol{c})\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y}) \\
&= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \boldsymbol{x}\mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y}) + \boldsymbol{c} \underbrace{\sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y})}_{=1} \\
&= \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}] + \boldsymbol{c}
\end{aligned}
$$

が成り立つ。

(5) $\boldsymbol{X}$ が離散型確率変数のときを考察する（$\boldsymbol{X}$ が連続型のときも以下と同様の議論から結論を得る）。$\boldsymbol{c} = (c_1, c_2, \cdots, c_d)^\top \in \mathbb{R}^d$ とし，$\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^\top \in \mathrm{R}(\boldsymbol{X})$ とする。このとき，

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{X}}[\langle \boldsymbol{c}, \boldsymbol{X}\rangle_2|\boldsymbol{y}] &= \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} \langle \boldsymbol{c}, \boldsymbol{x}\rangle_2 \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y}) \\
&= c_1 \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} x_1 \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y}) + \cdots + c_d \sum_{\boldsymbol{x}\in \mathrm{R}(\boldsymbol{X})} x_d \mathrm{P}(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{y}) \\
&= c_1 \sum_{x_1\in \mathrm{R}(X_1)} x_1 \mathrm{P}(X_1=x_1|\boldsymbol{y}) + \cdots + c_d \sum_{x_d\in \mathrm{R}(X_d)} x_d \mathrm{P}(X_d=x_d|\boldsymbol{y}) \\
&= c_1 \mathbb{E}_{X_1}[X_1|\boldsymbol{y}] + \cdots + c_d \mathbb{E}_{X_d}[X_d|\boldsymbol{y}] \\
&= \langle \boldsymbol{c}, \mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]\rangle_2
\end{aligned}
$$

が成り立つ。

(6) $\boldsymbol{X}=(X_1,X_2,\cdots,X_d)^\top$ とし，$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}=(\mu_{1|\boldsymbol{y}},\mu_{2|\boldsymbol{y}},\cdots,\mu_{d|\boldsymbol{y}})^\top=\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]$ とする。このとき，【6.5】(1) と同様の議論から，

$$\begin{aligned}\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]&=\mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\|_2^2|\boldsymbol{y}\right]\\&=\mathbb{E}_{\boldsymbol{X}}\left[\sum_{i=1}^d(X_i-\mu_{i|\boldsymbol{y}})^2\middle|\boldsymbol{y}\right]\\&=\sum_{i=1}^d\mathbb{E}_{X_i}\left[(X_i-\mu_{i|\boldsymbol{y}})^2|\boldsymbol{y}\right]\\&=\sum_{i=1}^d\mathbb{V}_{X_i}[X_i|\boldsymbol{y}]\end{aligned}$$

を得る。

(7) $\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}=\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]$，$\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}=\mathbb{E}_{\boldsymbol{Z}}[\boldsymbol{Z}|\boldsymbol{y}]$ とする。$\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}[\boldsymbol{X}+\boldsymbol{Z}|\boldsymbol{y}]\underset{[\text{命題 }6.5(1)]}{=}\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}+\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}$ と【6.5】(2) と同様の議論から，

$$\begin{aligned}\mathbb{V}_{(\boldsymbol{X},\boldsymbol{Z})}[\boldsymbol{X}+\boldsymbol{Z}|\boldsymbol{y}]&=\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}\left[\|(\boldsymbol{X}+\boldsymbol{Z})-\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}[\boldsymbol{X}+\boldsymbol{Z}|\boldsymbol{y}]\|_2^2\right]\\&=\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}\left[\|(\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}})+(\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}})\|_2^2\right]\\&=\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\|_2^2+\|\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}\|_2^2+2\langle\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}},\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}\rangle_2|\boldsymbol{y}\right]\\&=\mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\|_2^2|\boldsymbol{y}\right]+\mathbb{E}_{\boldsymbol{Z}}\left[\|\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}\|_2^2|\boldsymbol{y}\right]\\&\quad+2\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}\left[\langle\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}},\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}\rangle_2|\boldsymbol{y}\right]\\&=\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]+\mathbb{V}_{\boldsymbol{Z}}[\boldsymbol{Z}|\boldsymbol{y}]+2\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Z})}\left[\langle\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}},\boldsymbol{Z}-\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{y}}\rangle_2|\boldsymbol{y}\right]\end{aligned}$$

が成り立つ。

(8) $\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}=\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]$，$\boldsymbol{c}\in\mathbb{R}^d$ とする。$\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}|\boldsymbol{y}]\underset{[\text{命題 }6.5(1)]}{=}\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}+\boldsymbol{c}$ から，以下を得る。

$$\begin{aligned}\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}|\boldsymbol{y}]&=\mathbb{E}_{\boldsymbol{X}}\left[\|(\boldsymbol{X}+\boldsymbol{c})-\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}+\boldsymbol{c}|\boldsymbol{y}]\|_2^2|\boldsymbol{y}\right]\\&=\mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\|_2^2\right]\\&=\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]\end{aligned}$$

(9) $\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}=\mathbb{E}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]$，$\boldsymbol{c}\in\mathbb{R}^d$ とする。$\mathbb{E}_{\boldsymbol{X}}[\langle\boldsymbol{c},\boldsymbol{X}\rangle_2|\boldsymbol{y}]\underset{[\text{命題 }6.5(2)]}{=}\langle\boldsymbol{c},\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\rangle_2$ から，以下を得る。

$$\begin{aligned}\mathbb{V}_{\boldsymbol{X}}[\langle\boldsymbol{c},\boldsymbol{X}\rangle_2|\boldsymbol{y}]&=\mathbb{E}_{\boldsymbol{X}}\left[|\langle\boldsymbol{c},\boldsymbol{X}\rangle_2-\mathbb{E}_{\boldsymbol{X}}[\langle\boldsymbol{c},\boldsymbol{X}\rangle_2|\boldsymbol{y}]|^2|\boldsymbol{y}\right]\\&=\mathbb{E}_{\boldsymbol{X}}\left[|\langle\boldsymbol{c},\boldsymbol{X}\rangle_2-\langle\boldsymbol{c},\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\rangle_2|^2|\boldsymbol{y}\right]\\&=\mathbb{E}_{\boldsymbol{X}}\left[\langle\boldsymbol{c},\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\rangle_2^2|\boldsymbol{y}\right]\\&\leq\mathbb{E}_{\boldsymbol{X}}\left[\|\boldsymbol{c}\|_2^2\|\boldsymbol{X}-\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{y}}\|_2^2|\boldsymbol{y}\right]\\&=\|\boldsymbol{c}\|_2^2\mathbb{V}_{\boldsymbol{X}}[\boldsymbol{X}|\boldsymbol{y}]\end{aligned}$$

ただし，上記の不等式では，Cauchy–Schwarz の不等式（命題 3.7(4)）を利用した。

【6.6】 $\mathbb{E}_{(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n)}[(n/n-1)s_n^2]=\sigma^2$ を示すことができれば，$(n/n-1)s_n^2$ が σ^2 に対する不偏推定量となる。

$$\begin{aligned}\mathbb{E}\left[\frac{n}{n-1}s_n^2\right]&\underset{[\text{定義 }6.13]}{=}\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n\left\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_n\right\|_2^2\right]\\&\underset{[\text{命題 }6.3(2)]}{=}\frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^n\left\|\boldsymbol{X}_i-\overline{\boldsymbol{X}}_n\right\|_2^2\right]\\&=\frac{1}{n-1}\underbrace{\mathbb{E}\left[\sum_{i=1}^n\left\|(\boldsymbol{X}_i-\boldsymbol{\mu})-\left(\overline{\boldsymbol{X}}_n-\boldsymbol{\mu}\right)\right\|_2^2\right]}_{A_n}\end{aligned}$$

となる。$\overline{\boldsymbol{X}}_n = (1/n)\sum_{i=1}^{n} \boldsymbol{X}_i$ と命題 3.7(1) から，

$$\begin{aligned}
A_n &= \mathbb{E}\left[\sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2^2 + n\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2 - 2\sum_{i=1}^{n}\langle \boldsymbol{X}_i - \boldsymbol{\mu}, \overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\rangle_2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2^2 + n\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2 - 2n\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2^2 - n\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{X}_i}\left[\|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2^2\right] - n\mathbb{E}_{(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)}\left[\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2\right]
\end{aligned}$$

を得る。$\mathbb{V}_{\boldsymbol{X}_i}[\boldsymbol{X}_i] = \mathbb{E}_{\boldsymbol{X}_i}[\|\boldsymbol{X}_i - \mathbb{E}_{\boldsymbol{X}_i}[\boldsymbol{X}_i]\|_2^2] = \mathbb{E}_{\boldsymbol{X}_i}[\|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2^2] = \sigma^2$ が成り立つ。さらに，

$$\begin{aligned}
\frac{\sigma^2}{n} &\underset{[\text{命題 } 6.7]}{=} \mathbb{V}_{(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)}[\overline{\boldsymbol{X}}_n] \\
&= \mathbb{E}_{(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)}\left[\left\|\overline{\boldsymbol{X}}_n - \mathbb{E}_{(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)}[\overline{\boldsymbol{X}}_n]\right\|_2^2\right] \\
&\underset{[\text{命題 } 6.7]}{=} \mathbb{E}_{(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)}\left[\left\|\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}\right\|_2^2\right]
\end{aligned}$$

を満たすので，$A_n = (n-1)\sigma^2$，つまり，

$$\mathbb{E}\left[\frac{n}{n-1}s_n^2\right] = \frac{1}{n-1}A_n = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

を得る。