第D章　凸解析と最適化アルゴリズム

D.1　凸関数とその性質

この節では, 凸関数の定義をまず述べる. 凸関数のこの定義は幾何的なものであり, 凸関数がどのようなものであるかを直観的に理解するが用意である. しかし, 解析的な議論をするためには, この定義では不十分[1]である. 凸関数であるための必要十分条件が重要になってくる. まず, ① 閉区間上の凸関数はその内部で絶対連続であることを示す. さらに, 凸関数であるための必要十分条件を 2 つ述べる. ひとつは, ② 凸関数の平均変動は単調増加であることであり, もうひとつは, ③下から支える直線の存在することである. これらは, 凸関数を特徴付けていることに注意する.

定義 D.1. $a, b \in \mathbb{R}, a < b$ とする. 関数 $f : [a, b] \to \mathbb{R}$ が $x, y \in [a, b], 0 \le \lambda \le 1$ に対して

$$f\bigl(\lambda x + (1-\lambda)y\bigr) \le \lambda f(x) + (1-\lambda) f(y)$$

をみたすとき, 関数 f は**凸関数**であるという. とくに $x \ne y, 0 < \lambda < 1$ のとき

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

が成立する場合は, 関数 f は**狭義凸関数**であるという.

注意 D.2. $-f$ が凸関数 (狭義凸関数) のとき, 関数 f は**凹関数** (**狭義凹関数**) であるという. □

定理 D.3. 区間 $[a, b]$ 上の凸関数は開区間 (a, b) 上の各点で連続である.

Proof. まず, $\forall x \in [a, b]$ は

$$x = \lambda a + (1-\lambda) b \quad (0 \le \lambda \le 1)$$

と表せるから, 凸関数の定義より

$$f(x) \le \lambda f(a) + (1-\lambda) f(b) \le \max\{f(a),\, f(b)\} =: K$$

[1]たとえば, この定義の直接的に利用では Jensen の不等式の証明できない.

となることに注意する．よって, 凸関数 f は上に有界である．$\forall x_0 \in (a,\, b)$ とし, $f(x)$ は $x = x_0$ で連続であることを示すために関数 g を

$$g(t) := f(x_0 + t) - f(x_0) \qquad (a - x_0 < t < b - x_0)$$

と定める．このとき $g(t)$ は $(a - x_0,\, b - x_0) \ni \{0\}$ 上で凸関数であり, $g(0) = 0$ である．いま $\forall \epsilon > 0$ に対して正数 ϵ' を

$$0 < \epsilon' < \min\left\{\frac{\epsilon}{K + |f(x_0)|},\, 1\right\} \tag{D.1}$$

をみたすように選ぶ．さらに正数 ρ と δ を

$$0 < \rho < \min\{x_0 - a,\, b - x_0\}, \quad 0 < \delta < \rho\epsilon' \tag{D.2}$$

をみたすように定める．$0 < \epsilon' \leq 1$ だから

$$|t| < \delta \;\Rightarrow\; |t| < \rho$$

となる．したがって $|t| < \delta$ に対して, $g(t)$ は定義されている．しかも

$$|t| < \delta \;\Rightarrow\; \left|\frac{t}{\epsilon'}\right| < \rho \;\Rightarrow\; x_0 + \frac{t}{\epsilon'} \in (a,\, b)$$

となること[2]がわかる．よって, (D.3) から

$$\begin{aligned} g\left(\frac{t}{\epsilon'}\right) &= f\left(x_0 + \frac{t}{\epsilon'}\right) - f(x_0) \leq \left|f\left(x_0 + \frac{t}{\epsilon'}\right)\right| + |f(x_0)| \\ &\leq K + |f(x_0)| =: K_1 \end{aligned}$$

を得る．ところで

$$t = (1 - \epsilon') \times 0 + \epsilon'\left(\frac{t}{\epsilon'}\right)$$

と表されることと (D.1) から

$$g(t) \leq (1 - \epsilon')g(0) + \epsilon' g\left(\frac{t}{\epsilon'}\right) = \epsilon' g\left(\frac{t}{\epsilon'}\right) \leq \epsilon' K_1 < \epsilon \tag{D.3}$$

となることに注意をする．他方

$$0 = \frac{\epsilon'}{1 + \epsilon'}\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1 + \epsilon'} \cdot t$$

[2] $\left|\dfrac{t}{\epsilon'}\right| < \rho$ と (D.2) から $-x_0 + a < -\rho < \dfrac{t}{\epsilon'} < \rho < b - x_0$ となるので $a < \dfrac{t}{\epsilon'} + x_0 < b$ がわかる．

と表されるので

$$
\begin{aligned}
0 \le g(0) &\le \frac{\epsilon'}{1+\epsilon'} g\left(-\frac{t}{\epsilon'}\right) + \frac{1}{1+\epsilon'} g(t) \le \frac{\epsilon' K_1}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t) \\
&< \frac{\epsilon}{1+\epsilon'} + \frac{1}{1+\epsilon'} g(t)
\end{aligned}
$$

となり

$$
-\epsilon < g(t) \tag{D.4}
$$

を得る．よって (D.3) と (D.4) より

$$
|t| < \delta \ \Rightarrow\ |g(t)| < \epsilon
$$

である．$t = x - x_0$ と $g(t)$ の定義より

$$
|x - x_0| < \delta \ \Rightarrow\ |f(x) - f(x_0)| < \epsilon
$$

となり, $f(x)$ は点 x_0 で連続となる． □

定理 D.4. 関数 $\varphi : [a,\, b] \to \mathbb{R}$ が単調増加ならば

$$
f(x) := C + \int_a^x \varphi(t)\,\mathrm{d}t \qquad (a \le x \le b)
$$

は凸関数である．ただし C は定数である．

Proof. まず, $x,\, y \in [a,\, b]$, $x < y$, $0 \le \lambda \le 1$ のとき

$$
x \le \lambda x + (1-\lambda)y \le y
$$

であることに注意する．ここで, 関数 φ の単調性に注意すれば

$$
\int_{\lambda x+(1-\lambda)y}^{y} \varphi(t)\,\mathrm{d}t \ge \{y - \lambda x - (1-\lambda)y\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr), \tag{D.5}
$$

$$
\int_{x}^{\lambda x+(1-\lambda)y} \varphi(t)\,\mathrm{d}t \le \{\lambda x + (1-\lambda)y - x\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \tag{D.6}
$$

がわかる．f の凸性を示すために, 以下の式が非負であることがわかればよい．

$$
\begin{aligned}
&\lambda f(x) + (1-\lambda) f(y) - f\bigl(\lambda x + (1-\lambda)y\bigr) \\
&\quad = (1-\lambda)\bigl\{f(y) - f\bigl(\lambda x + (1-\lambda)y\bigr)\bigr\} - \lambda\bigl\{f\bigl(\lambda x + (1-\lambda)y\bigr) - f(x)\bigr\} \\
&\quad = (1-\lambda)\int_{\lambda x+(1-\lambda)y}^{y} \varphi(t)\,\mathrm{d}t - \lambda \int_{x}^{\lambda x+(1-\lambda)y} \varphi(t)\,\mathrm{d}t \\
&\quad \ge (1-\lambda)\{y - \lambda x - (1-\lambda)y\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \qquad (\because \text{(D.5)}) \\
&\qquad - \lambda\{\lambda x + (1-\lambda)y - x\}\varphi\bigl(\lambda x + (1-\lambda)y\bigr) \qquad (\because \text{(D.6)}) \\
&\quad = 0
\end{aligned}
$$

となる. 以上の議論から

$$\lambda f(x) + (1-\lambda)f(y) \geq f\bigl(\lambda x + (1-\lambda)y\bigr)$$

が示せた. □

定理 D.5. 関数 f が区間 $[a, b]$ 上の凸のとき, 開区間 (a, b) の各点 x で右微分係数 $\overset{\bullet}{f}_+(x)$ が存在し, 関数

$$\overset{\bullet}{f}_+\colon (a, b) \ni x \mapsto \overset{\bullet}{f}_+(x) \in \mathbb{R}$$

は単調増加である.

同様に左微分係数 $\overset{\bullet}{f}_-(x)$ が存在し, 関数

$$\overset{\bullet}{f}_-\colon (a, b) \ni \mapsto \overset{\bullet}{f}_-(x) \in \mathbb{R}$$

も単調増加である. さらに各点 $x \in (a, b)$ において

$$\overset{\bullet}{f}_-(x) \leq \overset{\bullet}{f}_+(x)$$

が成立する.

Proof. 任意の $x \in (a, b)$ に対して実数 η と ξ を

$$0 < \eta < \min\{x-a, b-x\}, \quad 0 < \xi < \eta$$

をみたすように選ぶ. このとき

$$0 < \frac{\xi}{\eta} < 1, \quad x + \xi = \frac{\xi}{\eta}(x+\eta) + \left(1 - \frac{\xi}{\eta}\right)x$$

と書けることに注意する. 関数 f は凸であることから

$$f(x+\xi) \leq \frac{\xi}{\eta}f(x+\eta) + \left(1-\frac{\xi}{\eta}\right)f(x) = \frac{\xi}{\eta}\bigl\{f(x+\eta) - f(x)\bigr\} + f(x)$$

となる. つまり $0 < \xi < \eta$ のとき

$$\frac{f(x+\xi) - f(x)}{\xi} \leq \frac{f(x+\eta) - f(x)}{\eta}$$

となり, $\xi \searrow 0$ とすれば

$$\overset{\bullet}{f}_+(x) \leq \frac{f(x+\eta) - f(x)}{\eta} \tag{D.7}$$

がわかる．同様に

$$x-\xi=\frac{\xi}{\eta}(x-\eta)+\left(1-\frac{\xi}{\eta}\right)x$$

と表せるから

$$f(x-\xi)\leq\frac{\xi}{\eta}f(x-\eta)+\left(1-\frac{\xi}{\eta}\right)f(x)=\frac{\xi}{\eta}\big\{f(x-\eta)-f(x)\big\}+f(x)$$

である．したがって $0<\xi<\eta$ のとき

$$\frac{f(x)-f(x-\eta)}{\eta}\leq\frac{f(x)-f(x-\xi)}{\xi}$$

となり, $\xi\searrow 0$ とすれば

$$\overset{\bullet}{f}_-(x)\geq\frac{f(x)-f(x-\eta)}{\eta} \tag{D.8}$$

を得る．さらに

$$f(x)=f\left(\frac{(x+\xi)+(x-\xi)}{2}\right)\leq\frac{f(x+\xi)+f(x-\xi)}{2}$$

となり

$$f(x)-f(x-\xi)\leq f(x+\xi)-f(x)$$

である．だから

$$\frac{f(x)-f(x-\xi)}{\xi}\leq\frac{f(x+\xi)-f(x)}{\xi}$$

ががわかる．ここで, $\xi\searrow 0$ とすると

$$\overset{\bullet}{f}_-(x)\leq\overset{\bullet}{f}_+(x) \tag{D.9}$$

がわかる．したがって, (D.7) – (D.9) を考慮すれば

$$\frac{f(x)-f(x-\eta)}{\eta}\leq\overset{\bullet}{f}_-(x)\leq\overset{\bullet}{f}_+(x)\leq\frac{f(x+\eta)-f(x)}{\eta} \tag{D.10}$$

がわかる．よって $\overset{\bullet}{f}_-(x)$ と $\overset{\bullet}{f}_+(x)$ は有界であることがわかる．

次に, 単調性を示す．そのために $a<x_1<x_2<b$ とする．(D.10) の

$$f_-(x)\leq f_+(x)\leq\frac{f(x+\eta)-f(x)}{\eta}$$

において, $x=x_1$, $x+\eta=x_2$ とおくと

$$f_-(x_1)\leq f_+(x_1)\leq\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

を得る．同様に (D.10) の

$$\frac{f(x)-f(x-\eta)}{\eta} \leq f_-(x) \leq f_+(x)$$

において, $x=x_2$, $x-\eta=x_1$ とおくと

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq f_-(x_2) \leq f_+(x_2)$$

を得る．これらの式を合わせると

$$\overset{\bullet}{f}_-(x_1) \leq \overset{\bullet}{f}_+(x_1) \leq \frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \overset{\bullet}{f}_-(x_2) \leq \overset{\bullet}{f}_+(x_2) \tag{D.11}$$

となる．よって

$$\overset{\bullet}{f}_-(x_1) \leq \overset{\bullet}{f}_-(x_2), \quad \overset{\bullet}{f}_+(x_1) \leq \overset{\bullet}{f}_+(x_2)$$

が得られる． □

定理 D.6. 区間 $(a,\, b)$ 上で定義された凸関数 f は, 単調増加で右連続関数 g と点 $c \in (a,\, b)$ によって

$$f(x) = f(c) + \int_c^x g(t)\,\mathrm{d}t \qquad (a < x < b)$$

と表せる．

Proof. $a < c < x < b$ とし, 閉区間 $[c,\, x]$ の分割を

$$c = x_0 < x_1 < \cdots < x_n = x$$

とする．このとき (D.11) より, $k = 1,\, 2,\, \ldots,\, n$ に対し

$$\overset{\bullet}{f}_-(x_{k-1}) \leq \overset{\bullet}{f}_+(x_{k-1}) \leq \frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}} \leq \overset{\bullet}{f}_-(x_k) \leq \overset{\bullet}{f}_+(x_k)$$

と書ける．これらに $x_k - x_{k-1}$ を掛ければ

$$\begin{aligned}\overset{\bullet}{f}_-(x_{k-1})(x_k-x_{k-1}) \leq \overset{\bullet}{f}_+(x_{k-1})(x_k-x_{k-1}) &\leq f(x_k)-f(x_{k-1}) \\ &\leq \overset{\bullet}{f}_-(x_k)(x_k-x_{k-1}) \leq \overset{\bullet}{f}_+(x_k)(x_k-x_{k-1})\end{aligned}$$

が成立し, 上の不等式の辺々について, 1 から n まで k について和をとると

$$\begin{aligned}\sum_{k=1}^n \overset{\bullet}{f}_-(x_{k-1})(x_k-x_{k-1}) \leq \sum_{k=1}^n \overset{\bullet}{f}_+(x_{k-1})(x_k-x_{k-1}) &\leq f(x)-f(c) \\ &\leq \sum_{k=1}^n \overset{\bullet}{f}_-(x_k)(x_k-x_{k-1}) \leq \sum_{k=1}^n \overset{\bullet}{f}_+(x_k)(x_k-x_{k-1})\end{aligned}$$

が成立する．$\overset{\bullet}{f}_-$ と $\overset{\bullet}{f}_+$ は単調増加だから Riemann 積分可能である．よって $\max_{1\le k\le n}\{x_k - x_{k-1}\} \to 0\,(n\to\infty)$ とすれば

$$
\begin{aligned}
\sum_{k=1}^{n} \overset{\bullet}{f}_-(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-(t)\,\mathrm{d}t\\
\sum_{k=1}^{n} \overset{\bullet}{f}_+(x_{k-1})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+(t)\,\mathrm{d}t\\
\sum_{k=1}^{n} \overset{\bullet}{f}_-(x_{k})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_-(t)\,\mathrm{d}t\\
\sum_{k=1}^{n} \overset{\bullet}{f}_+(x_{k})(x_k - x_{k-1}) &\to \int_c^x \overset{\bullet}{f}_+(t)\,\mathrm{d}t
\end{aligned}
$$

となり

$$
f(x) - f(c) = \int_c^x \overset{\bullet}{f}_-(t)\,\mathrm{d}t = \int_c^x \overset{\bullet}{f}_+(t)\,\mathrm{d}t
$$

がわかる．他方 $a < x < b$ のときも閉区間 $[x,\,c]$ について同様の論理を展開して

$$
f(c) - f(x) = \int_x^c \overset{\bullet}{f}_-(t)\,\mathrm{d}t = \int_x^c \overset{\bullet}{f}_+(t)\,\mathrm{d}t
$$

を示すことができる．したがって

$$
g(x) = \overset{\bullet}{f}_+(x+0)
$$

とおけば, g は右連続で単調増加で

$$
f(x) = f(c) + \int_c^x g(t)\,\mathrm{d}t
$$

となる． □

定理 D.7. 関数 f が区間 $(a,\,b)$ 上の凸であるための必要十分条件は, $\forall x_0 \in (a,\,b)$ について関数

$$
(a,\,b)\setminus\{x_0\} \ni x \mapsto \frac{f(x)-f(x_0)}{x-x_0} \in \mathbb{R} \tag{D.12}
$$

が単調増加となることである．

Proof. <u>必要性</u>: 関数 f が開区間 $(a,\,b)$ 上で凸ならば, 定理 D.6 により, 右連続な単調増加関数 g が存在して

$$
f(x) - f(x_0) = \int_{x_0}^x g(t)\,\mathrm{d}t
$$

と表せる．よって $x \neq x_0$ のとき

$$\frac{f(x)-f(x_0)}{x-x_0} = \frac{1}{x-x_0}\int_{x_0}^{x} g(t)\,\mathrm{d}t$$

は $(a,\,b)\setminus\{x_0\}$ 上で単調増加である．

十分性: (D.12) が成立したと仮定する．$\forall x,\,y \in (a,\,b),\, x<y,\, 0<\lambda<1$ に対し

$$x_0 = \lambda x + (1-\lambda)y$$

とおく．(D.12) より $x<x_0<y$ のとき

$$\frac{f(x_0)-f(x)}{x_0-x} \leq \frac{f(y)-f(x_0)}{y-x_0}.$$

この式に $x_0 = \lambda x + (1-\lambda)y$ を代入すると

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

を得る．上の不等式の辺々に $\lambda(1-\lambda)(y-x)$ を掛けると

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

を得る．これを整理すれば

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

がわかる．よって, $\lambda=0,\,1$ のときは自明だから, 関数 f は $(a,\,b)$ 上で凸であることがわかる． □

定義 D.8. 関数 f を $(a,\,b)$ 上の凸とする．点 $x_0 \in (a,\,b)$ において適当に実数 m を定めると $\forall x \in (a,\,b)$ に対して

$$S(x) := m(x-x_0)+f(x_0) \leq f(x) \tag{D.13}$$

が成立するとき, 直線 $y=S(x)$ を**点** x_0 **において** f **を下から支える直線**という．

定理 D.9. 関数 f が区間 $(a,\,b)$ 上の凸であるための必要十分条件は, $(a,\,b)$ の各点 x_0 において f を下から支える直線 $y=S(x)$ が存在することである．

Proof. 必要性: 関数 f は開区間 $(a,\,b)$ 上の凸で $x_0 \in (a,\,b)$ なので, 定理 D.6 より, ある右連続な単調増加関数 g が存在して

$$f(x) = f(x_0) + \int_{x_0}^{x} g(t)\,\mathrm{d}t$$

と書ける．このとき

$$f(x) \geq f(x_0) + g(x_0)\int_{x_0}^{x} \mathrm{d}t = f(x_0) + g(x_0)(x - x_0)$$

となることがわかる．したがって，直線

$$S(x) := f(x_0) + g(x_0)(x - x_0)$$

は点 x_0 で $f(x)$ を下から支えることがわかる．

十分性: $x,\, y \in (a,\, b)$, $x < y$, $0 \leq \lambda \leq 1$ とし

$$x_0 = \lambda x + (1-\lambda)y$$

とおく．仮定より x_0 において凸関数 f を下から支える直線 $S(x)$ が存在する．すると (D.13) から $S(x) \leq f(x)$ とである．さらに，S の線型性から

$$\begin{aligned} f\big(\lambda x + (1-\lambda)y\big) &= f(x_0) = S(x_0) = \lambda S(x) + (1-\lambda)S(y) \\ &\leq \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

となる．よって，関数 f は $(a,\, b)$ 上の凸である． □

D.2 凸集合とその性質

定義 D.10. C を $\mathbb{R}^n$ の空でない部分集合とする．部分集合 C は凸であるとは次の条件

$$\forall x,\, y \in C \text{ と } 0 \leq \lambda \leq 1 \ \Rightarrow\ \lambda x + (1-\lambda)y \in C$$

をみたすときをいう．

定義 D.11. $p \in \mathbb{N}$ とし，$\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とする．

(1) $\lambda_k \geq 0\,(k = 1,\, 2,\, \ldots,\, p)$ かつ $\sum_{k=1}^{p} \lambda_k = 1$ のとき

$$\sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k$$

を $\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p$ の**凸結合**という．

(2) $\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p$ によって**生成された凸多面体**を

$$\big\langle \boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p \big\rangle_{2,\,n} = \left\{ \sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k;\ \lambda_k \geq 0,\ \sum_{k=1}^{p} \lambda_k = 1 \right\} \subset \mathbb{R}^n$$

で定義する．

定理 D.12. (Carathéodory の定理) $p \in \mathbb{N}$ とし, $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とする. $\boldsymbol{y} \in \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p \rangle$ は以下のように表せる.

$$\boldsymbol{y} = \sum_{k=1}^{n+1} \lambda_k \boldsymbol{a}_{j_k},$$
$$1 \leq j_1 < j_2 < \cdots < j_{n+1} \leq p, \quad \lambda_k \geq 0, \quad \sum_{k=1}^{n+1} \lambda_k = 1.$$

Proof. $\boldsymbol{y} \in\in \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p \rangle =: A$ なので, ある $r \in \mathbb{N}$ と $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r \in A$ があって

$$\boldsymbol{y} = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i, \quad \lambda_i > 0\,(i = 1, 2, \ldots, r), \quad \sum_{i=1}^{r} \lambda_i = 1$$

と書ける. r は最小とする. すなわち, $r-1$ 個以下の A の点では $\boldsymbol{y}$ をそれらの凸結合として表せないとする. $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$ はアファィン従属と仮定する. すると, ある $a_i \in \mathbb{R}\,(i = 1, 2, \ldots, r)$ で $a_1 a_2 \cdots a_r \neq 0$ なるものが存在して

$$\sum_{i=1}^{r} a_i \boldsymbol{x}_i = \boldsymbol{0}_n \quad \text{かつ} \quad \sum_{i=1}^{r} a_i = 0$$

とできる. ただし, $\boldsymbol{0}_n$ は $\mathbb{R}^n$ の零ベクトルである. 集合 $\{a_1, a_2, \ldots, a_r\}$ の中の少なくとも 1 つは正である. 正であるもののなかで $\dfrac{\lambda_i}{a_i}$ が最小となるもの $a_i\,(i \in \{1, 2, \ldots, r\})$ を選ぶ. 一般性を失わずにそれを a_1 と書くことができる. すると

$$\boldsymbol{y} = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i - \frac{\lambda_1}{a_1} \underbrace{\sum_{i=1}^{r} a_i \boldsymbol{x}_i}_{=0} = \sum_{i=1}^{r} \left(\lambda_i - \frac{\lambda_1}{a_1} a_i \right) \boldsymbol{x}_i$$

と書ける. すると $a_i > 0\,(i \in \{2, 3, \ldots, r\})$ のとき, $\dfrac{\lambda_1}{a_1}$ の最小性から, $\dfrac{\lambda_i}{a_i} > \dfrac{\lambda_1}{a_1}$ となる. したがって

$$\lambda_i - \frac{\lambda_1}{a_1} a_i \geq \lambda_i - \frac{\lambda_i}{a_i} a_i = 0$$

となる. $a_i \leq 0\,(i \in \{2, 3, \ldots, r\})$ のときは, 明らかのに $\lambda_i - \dfrac{\lambda_1}{a_1} a_i > 0$ となる. 以上から任意の $i \in \{1, 2, \ldots, r\}$ に対して

$$\lambda_i - \frac{\lambda_1}{a_1} a_i \geq 0$$

がわかる. さらに $\left\{\lambda_i - \dfrac{\lambda_1}{a_1}a_i\right\}_{i=1,\,2,\,\ldots,\,r}$ の中の少なくとも 1 は零ではない. よって, r の最小性と矛盾する. よって, $x_1,\, x_2,\, \ldots, x_r$ はアフファィン独立となるので, $r \le k+1$ がわかる. □

補題 D.13. C を空でない $\mathbb{R}^n$ 凸閉部分集合とし, $\mathbf{0} \notin C$ とする. $\langle \cdot, \cdot \rangle_{2,n}$ を $\mathbb{R}^n$ の Euclid 内積とし, $|\cdot|_{2,n}$ を $\mathbb{R}^n$ の Euclid ノルムとする.

(1) $\boldsymbol{x}_0 \in C$ があって

$$|\boldsymbol{x}_0|_{2,\,n} = \min\big\{|\boldsymbol{x}|_{2,\,n};\, \boldsymbol{x} \in C\big\} > 0$$

となる.

(2) さらに

$$0 \le \langle \boldsymbol{x}_0,\, \boldsymbol{x} - \boldsymbol{x}_0\rangle_{2,\,n}\, (\forall \boldsymbol{x} \in C) \tag{D.14}$$

が成立する.

Proof.

$$\delta = \inf\big\{|\boldsymbol{x}|_{2,\,n};\, \boldsymbol{x} \in C \quad\big\} \ge 0$$

とおく. 点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset C$ を選んで

$$\delta = \lim_{k\to\infty} |\boldsymbol{x}_k|_{2,\,n}$$

となるようにする. $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は有界なので Bolzano-Weirstrass の定理より収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ が存在するので, この部分列の収束先を

$$\lim_{j\to\infty} \boldsymbol{x}_{k(j)} =: \boldsymbol{x}_0$$

とおく. 結果

$$|\boldsymbol{x}_0|_{2,\,n} = \lim_{j\to\infty} |\boldsymbol{x}_{k(j)}|_{2,\,n} = \delta$$

となる. 部分集合 C は閉なので $\boldsymbol{x}_0 \in C$ となる. $\mathbf{0}_n \notin C$ なので $\delta > 0$ がわかる.

(2) $\boldsymbol{x}_0 \in C$ は $|\boldsymbol{x}_0|_{2,\,n} = \delta$ をみたす唯一の点であることを示すために $\boldsymbol{x}_1 \in C$ で $|\boldsymbol{x}_1|_{2,\,n} = \delta$ とする. 部分集合 C の凸性から

$$\frac{1}{2}(\boldsymbol{x}_0 + \boldsymbol{x}_1) \in C$$

となる. したがって

$$\left|\frac{1}{2}(\boldsymbol{x}_0 + \boldsymbol{x}_1)\right|_{2,\,n} \ge \delta$$

である．しかし

$$4\delta^2 \le |\boldsymbol{x}_0+\boldsymbol{x}_1|_{2,n}^2 \le |\boldsymbol{x}_1-\boldsymbol{x}_0|_{2,n}^2+|\boldsymbol{x}_1+\boldsymbol{x}_0|_{2,n}^2 \le 2\bigl\{|\boldsymbol{x}_1|_{2,n}^2+|\boldsymbol{x}_0|_{2,n}^2\bigr\} \\ =4\delta^2$$

より

$$|\boldsymbol{x}_1-\boldsymbol{x}_0|_{2,n}^2=0 \Rightarrow \boldsymbol{x}_1=\boldsymbol{x}_0$$

である．(D.14) を示すために $\forall \boldsymbol{x}\in C$ を取る．このとき $(1-\lambda)\boldsymbol{x}_0+\lambda\boldsymbol{x}\in C\,(0\le\lambda\le 1)$ である．したがって

$$|\boldsymbol{x}_0|_{2,n}^2 \le |(1-\lambda)\boldsymbol{x}_0+\lambda\boldsymbol{x}|_{2,n}^2 = |\boldsymbol{x}_0|_{2,n}^2+\lambda^2|\boldsymbol{x}-\boldsymbol{x}_0|_{2,n}^2+2\lambda\langle\boldsymbol{x}_0,\,\boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,n}$$

である．結果

$$0\le\lambda|\boldsymbol{x}-\boldsymbol{x}_0|_{2,n}^2+2\langle\boldsymbol{x}_0,\,\boldsymbol{x}-\boldsymbol{x}_0\rangle_{2,n}$$

がわかる．ここで, $\lambda\to 0$ とすれば (D.14) がわかる． □

定義 D.14. S_1 と S_2 を $\mathbb{R}^n$ の空でない部分集合とする．$\boldsymbol{a}\in\mathbb{R}^n$, $b\in\mathbb{R}$ とする．

(1) 超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b$ は S_1 と S_2 を**分離**するとは次の条件をみたすときをいう．

$$\forall\boldsymbol{x}\in S_1 \Rightarrow \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\ge 0,$$
$$\forall\boldsymbol{x}\in S_2 \Rightarrow \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\le 0.$$

(2) 超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b$ は S_1 と S_2 を**厳密に分離**するとは次の条件をみたすときをいう．

$$\forall\boldsymbol{x}\in S_1 \Rightarrow \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b> 0,$$
$$\forall\boldsymbol{x}\in S_2 \Rightarrow \langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b< 0.$$

定理 D.15. $C\subset\mathbb{R}^n$ を空でない凸閉部分集合とし, $\boldsymbol{y}\notin C$ とする．このとき C と $\boldsymbol{y}$ を厳密に分離する超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b\,(\boldsymbol{a}\in\mathbb{R}^n,\,b\in\mathbb{R})$ が存在する．

Proof. 一般性を失うことなく $\boldsymbol{y}=\boldsymbol{0}$ としてよい．左の仮定から $\boldsymbol{0}\notin C$ に注意して補題 D.13 を用いると $\boldsymbol{x}_0\in C$ があって

$$0<\delta=|\boldsymbol{x}_0|_{2,n}=\min\{|\boldsymbol{x}|_{2,n};\,\boldsymbol{x}\in C\}$$

となる．ここで $\boldsymbol{z}=(1/2)\boldsymbol{x}_0$, $\boldsymbol{a}=\boldsymbol{x}_0/\delta$ とおく．超平面 $\langle\boldsymbol{a},\,\boldsymbol{x}-\boldsymbol{z}\rangle_{2,n}=0$ を考える．すなわち

$$\langle\boldsymbol{a},\,\boldsymbol{x}\rangle_{2,n}+b=0,\qquad b=-\langle\boldsymbol{a},\,\boldsymbol{z}\rangle_{2,n}$$

である. $\boldsymbol{y} = \boldsymbol{0}_n$ に対して

$$\langle \boldsymbol{a},\, \boldsymbol{y} \rangle_{2,n} + b = b = -\langle \boldsymbol{a},\, \boldsymbol{z} \rangle_{2,n} = -\left\langle \frac{\boldsymbol{x}_0}{\delta},\, \frac{\boldsymbol{x}_0}{2} \right\rangle_{2,n} = -\frac{|\boldsymbol{x}_0|_{2,n}^2}{2\delta} = -\frac{\delta}{2} < 0. \tag{D.15}$$

$\boldsymbol{x} \in C$ に対して (D.14) から

$$0 \leq \langle \boldsymbol{x}_0,\, \boldsymbol{x} - \boldsymbol{x}_0 \rangle_{2,n}$$

である. よって

$$0 \leq \langle \delta \boldsymbol{a},\, \boldsymbol{x} - 2\boldsymbol{z} \rangle_{2,n}$$

となる. 上の式の両辺を $\delta > 0$ で割れば

$$\begin{aligned} 0 &\leq \langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{z} \rangle_{2,n} - \langle \boldsymbol{a},\, \boldsymbol{z} \rangle_{2,n} = \langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{z} \rangle_{2,n} - \left\langle \frac{\boldsymbol{x}_0}{\delta},\, \frac{\boldsymbol{x}_0}{2} \right\rangle_{2,n} \\ &= \langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{z} \rangle_{2,n} - \frac{|\boldsymbol{x}_0|_{2,n}^2}{2\delta} \\ &= \langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{z} \rangle_{2,n} - \frac{\delta}{2} \end{aligned}$$

となる. よって

$$\langle \boldsymbol{a},\, \boldsymbol{x} - \boldsymbol{z} \rangle_{2,n} = \langle \boldsymbol{a},\, \boldsymbol{x} \rangle_{2,n} - \langle \boldsymbol{a},\, \boldsymbol{z} \rangle_{2,n} = \langle \boldsymbol{a},\, \boldsymbol{x} \rangle_{2,n} + b \geq \frac{\delta}{2} > 0 \tag{D.16}$$

となる. (D.15) と (D.16) より $\langle \boldsymbol{a},\, \boldsymbol{x} \rangle_{2,n} + b$ は $\boldsymbol{y} = \boldsymbol{0}_n$ と C を厳密に分離することが示せた. □

定義 D.16. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とする. 集合

$$C = \left\{ \sum_{j=1}^{p} \lambda_j \boldsymbol{a}_j;\, \lambda_j \geq 0 \, (j = 1,\, 2,\, \ldots,\, p) \right\}$$

を $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\}$ によって生成された**有限錐**という.

定義 D.17. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\}$ が線型独立のとき, このベクトルで生成された有限錐を**基本錐**という.

補題 D.18. $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ は有限錐 C を生成しているとする. $C_1,\, C_2,\, \ldots,\, C_q$ を $\{\boldsymbol{a}_1,\, \boldsymbol{a}_2,\, \ldots,\, \boldsymbol{a}_p\}$ の部分集合で生成されるすべての基本錐とする. このとき

$$C = \bigcup_{j=1}^{q} C_j$$

となる.

Proof. $C_j \subset C\,(j = 1, 2, \ldots, q)$ なので

$$\bigcup_{j=1}^{q} C_j \subset C$$

となる．いま $\boldsymbol{b} \in C = \{\mathbf{A}\boldsymbol{\beta};\, \boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top, \beta_j \geq 0\,(j = 1, 2, \ldots, p)\}$ を選ぶ．ただし $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p)$ は $n \times p$ 行列である．このとき

$$\mathbf{A}\boldsymbol{\beta} = \boldsymbol{b}, \quad \boldsymbol{\beta} \geq \mathbf{0}_p$$

なので, 基本解 $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \ldots, \beta_p^*)^\top \in \mathbb{R}^p$ が存在して

$$\mathbf{A}\boldsymbol{\beta}^* = \boldsymbol{b}$$

となることがわかる．すなわち $\exists m \in \mathbb{N}, (1 \leq m \leq n)$ があって, $1 \leq j_1 < \cdots < j_m \leq p$ は

$$\beta_{j_\ell} > 0\,(j = 1, 2, \ldots, m)$$

で

$$\{\boldsymbol{a}_{j_1}, \boldsymbol{a}_{j_2}, \ldots, \boldsymbol{a}_{j_m}\}$$

は線型独立である．よって $\boldsymbol{b}$ は $\{\boldsymbol{a}_{j_1}, \boldsymbol{a}_{j_2}, \ldots, \boldsymbol{a}_{j_m}\}$ で生成された基本錐に含まれるので

$$\boldsymbol{b} \in \cup_{j=1}^{q} C_j$$

となる．よって

$$C \subset \bigcup_{j=1}^{q} C_j$$

となる． □

定理 D.19. $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p\} \subset \mathbb{R}^n$ とし, C はこのベクトルで生成された有限錐とする．このとき C は凸かつ閉部分集合である．

Proof. 凸性: $\boldsymbol{c}_1, \boldsymbol{c}_2 \in C$, $0 \leq \lambda \leq 1$ とする．このとき $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^p$ が存在して

$$\boldsymbol{c}_1 = \mathbf{A}\boldsymbol{\beta}_1, \quad \boldsymbol{c}_2 = \mathbf{A}\boldsymbol{\beta}_2$$

となる．ただし $\mathbf{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p)$, $\boldsymbol{\beta}_1 \geq \mathbf{0}_p$, $\boldsymbol{\beta}_2 \geq \mathbf{0}_p$ である．したがって

$$(1-\lambda)\boldsymbol{c}_1 + \lambda\boldsymbol{c}_2 = \mathbf{A}\big((1-\lambda)\boldsymbol{\beta}_1 + \lambda\boldsymbol{\beta}_2\big)$$

となる．さらに $(1-\lambda)\boldsymbol{\beta}_1 + \lambda\boldsymbol{\beta}_2 \geq \mathbf{0}_p$ である．よって,

$$(1-\lambda)\boldsymbol{c}_1 + \lambda\boldsymbol{c}_2 \in C$$

より C は凸集合であることがわかる．

<u>C は閉集合</u>: $C_j\,(j=1,\,2,\,\ldots,\,q)$ を独立なベクトル

$$\{\boldsymbol{a}_{j_1},\,\boldsymbol{a}_{j_2},\,\ldots,\,\boldsymbol{a}_{j_\ell}\}\subset\{\boldsymbol{a}_1,\,\boldsymbol{a}_2,\,\ldots,\,\boldsymbol{a}_p\}$$

によって生成される基本錐とする. $\{\boldsymbol{c}_k\}_{k=1}^{\infty}$ は C_j の収束する点列とし, $\boldsymbol{c}_0\in\mathbb{R}^n$ に収束するとする.

$$\lim_{k\to\infty}\boldsymbol{c}_k=\boldsymbol{c}_0.$$

$\boldsymbol{c}_0\in C_j$ を示そう. $\{\boldsymbol{a}_{j_1},\,\boldsymbol{a}_{j_2},\,\ldots,\,\boldsymbol{a}_{j_\ell}\}$ は線型独立なので, 任意の $\boldsymbol{u}=(u_1,\,u_2,\,\ldots,\,u_\ell)^\top\in\mathbb{R}^\ell$ に対し

$$\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}=\boldsymbol{0}_n\ \Rightarrow\ \boldsymbol{u}=\boldsymbol{0}_n$$

となる. よって上の主張の対偶を取ると $\boldsymbol{u}\in\mathbb{R}^\ell$ で $|\boldsymbol{u}|_{2,\,\ell}=1$ なるものに対して

$$\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}\neq\boldsymbol{0}_\ell$$

となる. よって C_j は凸集合なので, 補題 D.18 より

$$\min\left\{\left|\sum_{k=1}^{\ell}u_k\boldsymbol{a}_{j_k}\right|_{2,\,\ell};\ |\boldsymbol{u}|_{2,\,\ell}=1\right\}=:\epsilon>0$$

がわかる. したがって ϵ の定義に注意すると $\boldsymbol{v}=(v_1,\,v_2,\,\ldots,\,v_\ell)^\top\in\mathbb{R}^\ell$ に対し

$$\left|\sum_{k=1}^{\ell}v_k\boldsymbol{a}_{j_k}\right|_{2,\,\ell}=|\boldsymbol{v}|_{2,\,\ell}\left|\sum_{j=1}^{\ell}\frac{v_k}{|\boldsymbol{v}|_{2\ \ell}}\boldsymbol{a}_{jk}\right|_{2,\,\ell}\geq\epsilon|\boldsymbol{v}|_{2,\,\ell}\tag{D.17}$$

となる. $\boldsymbol{c}_k\in C_j\,(k=1,\,2,\,\ldots)$ に対して

$$\boldsymbol{c}_k=\mathbf{A}\boldsymbol{\beta}_k$$

とおけば

$$\{j_1,\,j_2,\,\ldots,\,j_\ell\}=\{j\in\{1,\,2,\,\ldots,\,p\};\,\beta_j>0,\,\boldsymbol{\beta}_k=(\beta_1,\,\beta_2,\,\ldots,\,\beta_p)^\top\}.$$

このとき

$$\boldsymbol{c}_k=\sum_{k=1}^{\ell}\beta_{j_k}\boldsymbol{a}_{j_k}.$$

(D.17) より

$$\frac{1}{\epsilon}|\boldsymbol{c}_k|_{2,\,p}\geq|\boldsymbol{\beta}_k|_{2,\,p}\quad(k=1,\,2,\,\ldots)$$

となる. よって $\{\boldsymbol{c}_k\}_{k=1}^{\infty}$ は収束するので, $\{\boldsymbol{\beta}_k\}_{k=1}^{\infty}$ は有界. よって Bolzano-Weierstrass の定理より収束する部分列 $\{\boldsymbol{\beta}_{k_j}\}_{j=1}^{\infty}$ が取れ

$$\lim_{j\to\infty} \boldsymbol{\beta}_{k_j} =: \boldsymbol{\beta}_0$$

とする. このとき $\boldsymbol{\beta}_0 \geq \mathbf{0}$ で

$$\mathbf{A}\boldsymbol{\beta}_0 = \lim_{j\to\infty} \mathbf{A}\boldsymbol{\beta}_{k_j} = \lim_{j\to\infty} \boldsymbol{c}_{k_j} = \boldsymbol{c}_0$$

となる. $\boldsymbol{\beta}_0$ の j_k 成分 $(k = 1, 2, \ldots, \ell)$ 以外は零なので, $\boldsymbol{c}_0 \in C_j$ となる. よって C_j は閉集合である. だから

$$C = \bigcup_{j=1}^{q} C_j$$

も閉集合となることがわかる. □

定理 D.20. (Farkas の交代定理) $\mathbf{A}$ を $m \times n$ 行列とし, $\boldsymbol{b} \in \mathbb{R}^m$ とする. このとき次のいずれかが成立し, 同時には成立しない.

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geq \mathbf{0}_n$ は解 $\boldsymbol{x}^* \in \mathbb{R}^n$ を持つ.

(2) $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_m$, $\langle \boldsymbol{y}, \boldsymbol{b}\rangle_{2,m} < 0$ は解 $\boldsymbol{y}^* \in \mathbb{R}^m$ を持つ.

ただし $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^\top$, $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)^\top \in \mathbb{R}^n$ のとき

$$\boldsymbol{w} \geq \boldsymbol{z} \Leftrightarrow w_j \geq z_j \, (j = 1, 2, \ldots, n)$$

である.

Proof. $\boldsymbol{x}^* \in \mathbb{R}^m$ は (1) の解, $\boldsymbol{y}^* \in \mathbb{R}^n$ は (2) の解とする. このとき

$$0 \leq \langle \boldsymbol{x}^*, \mathbf{A}^\top \boldsymbol{y}^*\rangle_{2,n} = \langle \mathbf{A}\boldsymbol{x}^*, \boldsymbol{y}^*\rangle_{2,m} = \langle \boldsymbol{b}, \boldsymbol{y}\rangle_{2,m} < 0$$

となる. よって矛盾. (1) と (2) は同時に起こらないことがわかる.

(1) が起こらないとき, (2) が起こることを示す. 仮定より

$$\boldsymbol{b} \notin C := \{\mathbf{A}\boldsymbol{x}; \boldsymbol{x} \geq \mathbf{0}_n\}$$

となる. C は閉凸集合なので超平面が存在して C と $\boldsymbol{b}$ を厳密に分離する. すなわち, $\boldsymbol{a} \in \mathbb{R}^m$ と $c \in \mathbb{R}$ があって

$$\boldsymbol{z} \in C \Rightarrow \langle \boldsymbol{a}, \boldsymbol{z}\rangle_{2,m} + c > 0 \tag{D.18}$$

かつ

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{2,m} + c < 0. \tag{D.19}$$

$\boldsymbol{x} \geq \mathbf{0}_n,\ \lambda > 0$ とし

$$\boldsymbol{z} := \lambda \mathbf{A}\boldsymbol{x} = \mathbf{A}(\lambda \boldsymbol{x}) \in C$$

となる. (D.18) より

$$\langle \boldsymbol{a},\ \lambda \mathbf{A}\boldsymbol{x} \rangle_{2,n} + c > 0.$$

$\lambda > 0$ で割って $\lambda \to \infty$ とすれば

$$0 \leq \langle \boldsymbol{a},\ \mathbf{A}\boldsymbol{x} \rangle_{2,n} = \langle \mathbf{A}^\top \boldsymbol{a},\ \boldsymbol{x} \rangle_{2,n} \quad (\forall \boldsymbol{x} \geq \mathbf{0}_n)$$

となる. したがって

$$\mathbf{A}^\top \boldsymbol{a} \geq \mathbf{0}_n$$

となる. いま $\boldsymbol{y} = \boldsymbol{a}$ とおけば, $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ となる. (D.18) において $\boldsymbol{z} = \mathbf{0}_n \in C$ とおけば

$$c > 0$$

となる. さらに, (D.19) より

$$\langle \boldsymbol{b},\ \boldsymbol{y} \rangle_{2,m} = \langle \boldsymbol{b},\ \boldsymbol{a} \rangle_{2,m} < -c < 0$$

となる. よって $\langle \boldsymbol{b},\ \boldsymbol{y} \rangle_{2,n} < 0$ となる. よって, (1) が起こらなければ (2) が起こる. 対偶をとれば (2) が起こらなければ (1) が起こる. □

記法: $\mathbf{A}$ を $m \times n$ 行列とする.

$$\ker(\mathbf{A}) := \{\boldsymbol{x};\ \mathbf{A}\boldsymbol{x} = \mathbf{0}_m\} \subset \mathbb{R}^n,$$
$$\mathrm{range}(\mathbf{A}) := \{\mathbf{A}\boldsymbol{x};\ \boldsymbol{x} \in \mathbb{R}^m\} \subset \mathbb{R}^m.$$

$S \subset \mathbb{R}^m$ としたとき

$$S^\perp := \{\boldsymbol{x} \subset \mathbb{R}^m;\ \langle \boldsymbol{x},\ \boldsymbol{s} \rangle_{2,m} = 0\,(\forall \boldsymbol{s} \in S)\}$$

とする.

定理 D.21. $\mathbf{A}$ を $m \times n$ 行列, $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ とする. このとき次は同値である.

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ は解 $\boldsymbol{x}^*$ を持つ.

(2) $\boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp$ である.

Proof. $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ が解を持つための必要十分条件は

$$\widetilde{\mathbf{A}} = [\mathbf{A},\ -\mathbf{A}] \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \boldsymbol{b}, \quad \boldsymbol{u} \geq \mathbf{0}_n,\ \boldsymbol{v} \geq \mathbf{0}_n$$

は解を持つ．これは

$$\widetilde{\mathbf{A}} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \boldsymbol{b},\ \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \geq \mathbf{0}_{2n}$$

とした Farkas(1) である．よって

$$\text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」} \Leftrightarrow \text{「Farkas (1) が成立」} \tag{D.20}$$

となる．

一方 Farkas(2) が成立するとき

$$\widetilde{\mathbf{A}}^\top \boldsymbol{y} \geq \mathbf{0}_n, \quad \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,\,n} < 0$$

が解を持つ．これは

$$\begin{bmatrix} \mathbf{A}^\top \\ -\mathbf{A}^\top \end{bmatrix} \boldsymbol{y} \geq \begin{bmatrix} \mathbf{0}_n \\ \mathbf{0}_n \end{bmatrix}$$

である．よって $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ かつ $-\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$ より $\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n$ なので

$$\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n, \quad \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,\,n} < 0$$

は解をもつ．

以上の議論から

1. 「$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ は解 $\boldsymbol{x}^*$ を持つ」が成立する．
2. $\widetilde{\mathbf{A}}^\top \boldsymbol{y} \geq \mathbf{0}_n, \quad \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,\,n} < 0$ は解をもつ．

のいずれかが成立 (同時には起こらない) することがわかる．

さらに

$$\text{Farkas (2) が起こらない} \ \Leftrightarrow\ \boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp$$

である．なせならば $\boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp$ ならば

$$\mathbf{A}^\top \boldsymbol{y} = \mathbf{0}_n \quad \text{かつ} \quad \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,\,n} = 0$$

なので (2) は成立しない．逆に (2) が成立しないならば Farkas(1) が成立する．(D.20) より

$$\text{「Farkas (1) が起こる」} \Leftrightarrow \text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」}$$

なので $\boldsymbol{b} = \mathbf{A}\boldsymbol{x}^*$ と書ける．よって $\forall \boldsymbol{y} \in \ker\bigl(\mathbf{A}^\top\bigr)$ に対して

$$\langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,\,n} = \langle \boldsymbol{y},\, \mathbf{A}\boldsymbol{x}^* \rangle_{2,\,n} = \langle \mathbf{A}^\top \boldsymbol{y},\, \boldsymbol{x}^* \rangle_{2,\,n} = 0$$

から $\boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp$ がわかる.
　以上の議論から

$$\begin{aligned}
\text{「}\mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ は解 } \boldsymbol{x}^* \text{ を持つ」} &\Leftrightarrow \text{「Farkas(1) が成立」} \\
&\Leftrightarrow \text{Farkas(2) が成立しない」} \\
&\Leftrightarrow \boldsymbol{b} \in \ker\bigl(\mathbf{A}^\top\bigr)^\perp
\end{aligned}$$

より定理は証明された. □

D.3　共役凸関数と Young の不等式

定義 D.22. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ が凸であるとは, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ と $0 \leq \lambda \leq 1$ に対して

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

が成り立つときをいう.

注意 D.23. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ が凸であるための必要十分条件はそのエピグラフ

$$E = \left\{ \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix} : y \geq f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n \right\} \subset \mathbb{R}^{n+1}$$

が凸集合となることである.
　$m \in \mathbb{N}$ とする. 関数 f が凸のとき, 帰納法により $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ と $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ で $\sum_{k=1}^m \lambda_k = 1$ に対して

$$f\left(\sum_{k=1}^m \lambda_k \boldsymbol{x}_k\right) \leq \sum_{k=1}^m \lambda_k f(\boldsymbol{x}_k)$$

が成り立つ. □

定理 D.24. (1) 関数 $f : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とする. 関数 f が凸となるための必要十分条件は $\forall \boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{R}^n$ に対して

$$f(\boldsymbol{x}_0) + \bigl\langle \nabla f(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0 \bigr\rangle_{2,n} \leq f(\boldsymbol{x})$$

が成り立つことである. ただし

$$\begin{aligned}
\nabla f(\boldsymbol{x}) &= \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right)^\top, \\
\boldsymbol{x} &= (x_1, x_2, \ldots, x_n)^\top
\end{aligned}$$

である.
(2) 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は 2 回連続微分可能とする. 関数 f が凸であるための必要十分条件は $\forall\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$\nabla^2 f(\boldsymbol{x})\succeq\boldsymbol{0}_{n\times n}$$

が成り立つことである. ただし, $\nabla^2 f(\boldsymbol{x})\succeq\boldsymbol{0}_{n\times n}$ は $\forall\boldsymbol{y}=(y_1,\,y_2,\,\ldots,\,y_n)^\top\in\mathbb{R}^n$ に対して

$$\boldsymbol{y}^\top\nabla^2 f(\boldsymbol{x})\boldsymbol{y}=\sum_{i,\,j=1}^{n}\frac{\partial f^2}{\partial x_i\partial x_j}(\boldsymbol{x})y_iy_j\geq 0$$

が成り立つことである.

Proof. まず, 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸であるための必要十分条件は $\forall\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$ に対して, 関数

$$\phi(t):=f(\boldsymbol{x}+t\boldsymbol{y})\qquad(t\in\mathbb{R})\tag{D.21}$$

が凸であることである.
　まず, 関数 f は凸とし, 関数 ϕ は凸であることを示す. そのために, $\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$, $0\leq\lambda\leq 1$, $s,\,t\in\mathbb{R}$ を取り,

$$\boldsymbol{w}:=\boldsymbol{x}+s\boldsymbol{y},\quad\boldsymbol{z}:=\boldsymbol{x}+t\boldsymbol{y}$$

とおく. このとき

$$f\bigl(\lambda\boldsymbol{w}+(1-\lambda)\boldsymbol{z}\bigr)\leq\lambda f(\boldsymbol{w})+(1-\lambda)f(\boldsymbol{z})=\lambda\phi(s)+(1-\lambda)\phi(t)\tag{D.22}$$

となる. また

$$f\bigl(\lambda\boldsymbol{w}+(1-\lambda)\boldsymbol{z}\bigr)=f\bigl(\boldsymbol{x}+\{\lambda t+(1-\lambda)s\}\boldsymbol{y}\bigr)=\phi(\lambda t+(1-\lambda)s)\tag{D.23}$$

となる. よって, (D.22) と (D.23) を考慮すると ϕ は凸関数であることがわかる.
　逆を示す. $0\leq\lambda\leq 1,\boldsymbol{x},\,\boldsymbol{y}\in\mathbb{R}^n$ とする. (D.21) において, $\boldsymbol{x},\,\boldsymbol{y}$ は任意だったので, $\boldsymbol{y}$ と $\boldsymbol{x}-\boldsymbol{y}$ とする. すると

$$\phi(t)=f\bigl(\boldsymbol{y}+t(\boldsymbol{x}-\boldsymbol{y})\bigr)$$

とおく. 関数 ϕ は凸だから

$$\begin{aligned}f\bigl(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\bigr)&=f\bigl(\boldsymbol{y}+\lambda(\boldsymbol{x}-\boldsymbol{y})\bigr)=\phi(\lambda)\\&=\phi\bigl((1-\lambda)\times 0+\lambda\times 1\bigr)\\&\leq(1-\lambda)\phi(0)+\lambda\phi(1)\\&=(1-\lambda)f(\boldsymbol{y})+\lambda f(\boldsymbol{x})\end{aligned}$$

となる. よって $\forall \boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n$ に対して関数

$$\mathbb{R} \ni t \mapsto \phi(t) = f(\boldsymbol{x} + t\boldsymbol{y})$$

が凸関数ならば f も凸関数であることがわかる.
2. 凸関数 $\phi(t) = f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)$ に対して

$$\phi(t) + \overset{\bullet}{\phi}(t)(s-t) \le \phi(s) \quad (\forall s,\, t \in \mathbb{R})$$

である. ただし $\overset{\bullet}{\phi}(t) = \dfrac{d\phi}{dt} = \langle \nabla f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big),\, \boldsymbol{y}\rangle_{2,n}$ である. ここで $s = 1,\, t = 0$ とおくと

$$\phi(0) + \overset{\bullet}{\phi}(0) \le \phi(1) \ \Leftrightarrow\ f(\boldsymbol{x}_0) + \big\langle \nabla f(\boldsymbol{x}_0),\, \boldsymbol{y} \big\rangle_{2,\,n} \le f\big(\boldsymbol{x}_0 + \boldsymbol{y}\big)$$

となる. 上式に $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{x}_0$ を代入すると

$$f(\boldsymbol{x}_0) + \big\langle \nabla f(\boldsymbol{x}_0),\, \boldsymbol{x} - \boldsymbol{x}_0 \big\rangle_{2,\,n} \le f(\boldsymbol{x})$$

を得る.
3. 凸関数 $\phi(t) = f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)$ は

$$\overset{\bullet\bullet}{\phi}(t) \ge 0$$

をみたす. このとき

$$\overset{\bullet}{\phi}(t) = \big\langle \nabla f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big),\, \boldsymbol{y} \big\rangle_{2,\,n}, \quad \overset{\bullet\bullet}{\phi}(t) = \boldsymbol{y}^\top \nabla^2 f\big(\boldsymbol{x}_0 + t\boldsymbol{y}\big)\boldsymbol{y}$$

となる. ただし $\nabla^2 f(\boldsymbol{x}) = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,\, j=1,\, 2,\, \ldots,\, n}$ である. $t = 0$ とおくと

$$\boldsymbol{y}^\top \nabla^2 f(\boldsymbol{x}_0)\boldsymbol{y} \ge 0 \quad (\forall y \in \mathbb{R}^n)$$

がわかる. □

定義 D.25. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ はとする. 各 $\boldsymbol{x} \in \mathbb{R}^n$ に対して

$$\partial f(\boldsymbol{x}) = \big\{\boldsymbol{r} \in \mathbb{R}^n;\, f(\boldsymbol{x}) + \langle \boldsymbol{r},\, \boldsymbol{y} - \boldsymbol{x}\rangle_{2,n} \le f(\boldsymbol{y})\, (\forall \boldsymbol{y} \in \mathbb{R}^n)\big\}$$

とおく. この集合を点 $\boldsymbol{x}$ における凸関数 f の**劣微分**という.

注意 D.26. 凸関数 $f : \mathbb{R}^n \to \mathbb{R}$ は点 $\boldsymbol{x}$ で連続微分可能なとき

$$\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$$

となる. 一般に, $\partial f(\boldsymbol{x})$ は 1 つの元よりも多い元をもつことがある. □

例 D.27. (1) $n=1$, $f(x)=|x|$ とする．このとき

$$\partial f(x)=\begin{cases}\{1\} & (x>0)\\ [-1,\,1] & (x=0)\\ \{-1\} & (x<0)\end{cases}$$

となる．
(2) $n>1$, $f(\boldsymbol{x})=|\boldsymbol{x}|_{2,\,n}$ とする．このとき

$$\partial f(\boldsymbol{x})=\begin{cases}\left\{\dfrac{\boldsymbol{x}}{|\boldsymbol{x}|_{2,\,n}}\right\} & (\boldsymbol{x}\neq\boldsymbol{0}_n)\\ B(1) & (\boldsymbol{x}=\boldsymbol{0}_n)\end{cases}$$

となる．ただし $B(1):=\{\boldsymbol{x}\in\mathbb{R}^n;\,|\boldsymbol{x}|_{2,\,n}=1\}$ である．
　実際, $\forall\boldsymbol{r}\in\partial f(\boldsymbol{0}_n)$ は $\forall\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$|\boldsymbol{0}_n|_{2,\,n}+\langle\boldsymbol{r},\,\boldsymbol{x}-\boldsymbol{0}_n\rangle_{2,\,n}\leq|\boldsymbol{x}|_{2,\,n}$$

を意味する．$|\langle\boldsymbol{r},\,\boldsymbol{x}\rangle_{2,\,n}|\leq|\boldsymbol{r}|_{2,\,n}\times|\boldsymbol{x}|_{2,\,n}$ から

$$\langle\boldsymbol{r},\,\boldsymbol{x}\rangle_{2,\,n}\leq|\boldsymbol{x}|_{2,\,n}\ \Leftrightarrow\ |\boldsymbol{r}|_{2,\,n}\leq 1$$

となる．　□

定理 D.28. 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸とする．各 $\boldsymbol{x}\in\mathbb{R}^n$ に対して

$$\partial f(\boldsymbol{x})\,\text{は凸閉集合}$$

である．

Proof. 明らかに $\partial f(\boldsymbol{x})$ は凸集合である．これが閉集合であることを示すために

$$\{\boldsymbol{r}_k\}_{k=1}^{\infty}\subset\partial f(\boldsymbol{x})\quad\text{かつ}\quad\boldsymbol{r}_0=\lim_{k\to\infty}\boldsymbol{r}_k$$

を取る．$\boldsymbol{r}_k\in\partial f(\boldsymbol{x})\,(k=1,\,2,\,\ldots)$ だから $\forall\boldsymbol{y}\in\mathbb{R}^n$ に対して

$$f(\boldsymbol{x})+\langle\boldsymbol{r}_k,\,\boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n}\leq f(\boldsymbol{x})$$

となる．$k\to\infty$ とすれば, $\forall\boldsymbol{y}\in\mathbb{R}^n$ に対して

$$f(\boldsymbol{x})+\langle\boldsymbol{r}_0,\,\boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n}\leq f(\boldsymbol{y})$$

となるから $\boldsymbol{r}_0\in\partial f(\boldsymbol{x})$ となる．よって $\partial f(\boldsymbol{x})$ は閉集合となる．　□

定理 D.29. 関数 $f:\mathbb{R}^n\to\mathbb{R}$ は凸とする．このとき, 各 $\boldsymbol{x}\in\mathbb{R}^n$ に対して $\partial f(\boldsymbol{x})\neq\varnothing$ となる．

Proof. 凸関数 f のエピグラフ[3] E は凸閉集合なので正の整数 k に対して

$$\boldsymbol{e}_k := \begin{bmatrix} \boldsymbol{x} \\ f(\boldsymbol{x}) - \dfrac{1}{k} \end{bmatrix} \notin E$$

である. 分離定理 (定理 D.15) よりある超平面

$$\langle \boldsymbol{a}_k,\, \boldsymbol{x} \rangle_{2,\,n} + d_k \, (\boldsymbol{a}_k \in \mathbb{R}^{n+1},\, \mathrm{d}_k \in \mathbb{R})$$

は $\boldsymbol{e}_k$ と E を厳密に分離する.

$$\boldsymbol{a}_k = \begin{bmatrix} \boldsymbol{b}_k \\ c_k \end{bmatrix}, \quad \boldsymbol{b}_k \in \mathbb{R}^n, \quad c_k \in \mathbb{R}$$

と書くと

$$\langle \boldsymbol{a}_k,\, \boldsymbol{e}_k \rangle_{2,\,n} + d_k < 0, \tag{D.24}$$

$$\langle \boldsymbol{a}_k,\, \boldsymbol{z} \rangle_{2,\,n} + d_k > 0 \qquad (\forall \boldsymbol{z} \in E) \tag{D.25}$$

となる. (D.24) は

$$\langle \boldsymbol{b}_k,\, \boldsymbol{x} \rangle_{2,\,n} + c_k \left(f(\boldsymbol{x}) - \frac{1}{k} \right) + d_k < 0 \tag{D.26}$$

となる. $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{y}) \end{bmatrix} \in E$$

なので

$$\langle \boldsymbol{b}_k,\, \boldsymbol{y} \rangle_{2,\,n} + c_k f(\boldsymbol{y}) + d_k > 0 \tag{D.27}$$

となる. (D.26) を -1 倍して (D.27) に加わえると $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$\langle \boldsymbol{b}_k,\, \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\,n} + c_k \left(f(\boldsymbol{y}) - f(\boldsymbol{x}) + \frac{1}{k} \right) > 0 \tag{D.28}$$

となる. (D.28) に $\boldsymbol{y} = \boldsymbol{x}$ を代入すると

$$\frac{c_k}{k} > 0 \quad \Rightarrow c_k > 0$$

となる. したがって (D.28) は $\forall \boldsymbol{y} \in \mathbb{R}^n$ に対して

$$f(\boldsymbol{y}) + \frac{1}{k} \geq -\left\langle \frac{\boldsymbol{b}_k}{c_k},\, \boldsymbol{y} - \boldsymbol{x} \right\rangle_{2,\,n} + f(\boldsymbol{x}) \tag{D.29}$$

[3] epigraph の epi は「上の」という意味の接頭語.

を意味する．ここで

$$\boldsymbol{r}_k = -\frac{1}{c_k}\boldsymbol{b}_k$$

とおく．

次に $\{\boldsymbol{r}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ は有界であることを示す．$|\boldsymbol{r}_k|_{2,\,n} \neq 0$ ならば (D.29) において

$$\boldsymbol{y} = \boldsymbol{x} + \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}}$$

とおく[4]と

$$f\left(\boldsymbol{x} + \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}}\right) + \frac{1}{k} \geq |\boldsymbol{r}_k|_{2,\,n} + f(\boldsymbol{x})$$

となる．よって $k = 1, 2, \ldots$ に対して

$$|\boldsymbol{r}_k|_{2,\,n} \leq 1 + \max_{|\boldsymbol{u}|_{2,\,n}=1} \big|f(\boldsymbol{x}+\boldsymbol{u})\big| + f(\boldsymbol{x}) =: M$$

となる．よって $\{\boldsymbol{r}_k\}_{k=1}^{\infty}$ は有界であることがわかる．

Bolzano-Weirstrass の定理より収束する部分列 $\{\boldsymbol{r}_{k(j)}\}_{j=1}^{\infty}$ が取れる．

$$\boldsymbol{r}_0 := \lim_{j\to\infty} \boldsymbol{r}_{k(j)}$$

とおく．(D.29) において $k = k_j \overset{j\to\infty}{\longrightarrow} \infty$ とすれば

$$f(\boldsymbol{y}) \geq \langle \boldsymbol{r}_0,\, \boldsymbol{y}-\boldsymbol{x}\rangle_{2,\,n} + f(\boldsymbol{x}) \quad (\forall \boldsymbol{y} \in \mathbb{R}^n)$$

を得る．よって $\boldsymbol{r}_0 \in \partial f(\boldsymbol{x})$ がわかる．したがって $\partial f(\boldsymbol{x}) \neq \varnothing$ が示せた．
□

以降, この節では関数 $f : \mathbb{R}^n \to \mathbb{R}$ は凸で

$$\lim_{|\boldsymbol{x}|_{2,\,n}\to\infty} \frac{f(\boldsymbol{x})}{|\boldsymbol{x}|_{2,\,n}} = +\infty \tag{D.30}$$

をみたすとする．これを**超線型成長条件**という．

定義 D.30. $\boldsymbol{y} \in \mathbb{R}^n$ に対し

$$f^*(\boldsymbol{y}) = \max_{\boldsymbol{x}\in\mathbb{R}^n}\big\{\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x})\big\}$$

とする．f^* を f **の凸共役関数**という．

[4]定義から $|\boldsymbol{b}_k|^2 = c_k^2|\boldsymbol{r}_k|^2$ である．

$$\left\langle \frac{\boldsymbol{b}_k}{c_k},\, \frac{\boldsymbol{r}_k}{|\boldsymbol{r}_k|_{2,\,n}}\right\rangle_{2,\,n} = \left\langle \frac{\boldsymbol{b}_k}{c_k},\, -\frac{\boldsymbol{b}_k}{c_k|\boldsymbol{r}_k|_{2,\,n}}\right\rangle_{2,\,n} = -\frac{|\boldsymbol{b}_k|_{2,\,n}^2}{c_k^2|\boldsymbol{r}_k|_{2,\,n}} = -|\boldsymbol{r}_k|_{2,\,n}$$

になることに注意する．

注意 D.31. $n=1,\ f(x)=x^2/2\,(x\in\mathbb{R})$ とする．このとき $y\in\mathbb{R}$ に対し

$$f^*(y)=\max_x\left(xy-\frac{x^2}{2}\right)=\frac{y^2}{2}$$

となる．
(2) $n=1,\ 1<p<\infty,\ f(x)=|x|^p/p$ とする．このとき $y\in\mathbb{R}$ に対し

$$f^*(y)=\max_x\left(xy-\frac{|x|^p}{p}\right)=\frac{|y|^q}{q},\qquad \frac{1}{p}+\frac{1}{q}=1$$

となる．
$\because$) y を固定し

$$g(x)=xy-\frac{|x|^p}{p}$$

とする．すると $\dot{g}(x)=y-|x|^{p-1}\mathrm{sgn}(x)$ となる．ただし

$$\mathrm{sgn}(x)=\begin{cases}1 & (x>0)\\ -1 & (x<0)\\ 0 & (x=0)\end{cases}$$

とした．したがって

$$0=\dot{g}(x)\ \Leftrightarrow\ y=|x|^{p-1}\mathrm{sgn}(x)\ \Rightarrow\ x=|y|^{1/(p-1)}\mathrm{sgn}(y)$$

となる．結果

$$\begin{aligned}f^*(y)&=\left(|y|^{1/(p-1)}\mathrm{sgn}(y)\right)y-\frac{\left||y|^{1/(p-1)}\mathrm{sgn}(y)\right|^p}{p}=|y|^{\frac{1}{1-p}+1}-\frac{|y|^{\frac{p}{p-1}}}{p}\\&=\left(1-\frac{1}{p}\right)|y|^{\frac{p}{p-1}}=\frac{|y|^q}{q}\end{aligned}$$

がわかる．　□

補題 D.32. $\boldsymbol{x},\ \boldsymbol{y}\in\mathbb{R}^n$ に対し

$$\langle\boldsymbol{x},\ \boldsymbol{y}\rangle_{2,n}\le f(\boldsymbol{x})+f^*(\boldsymbol{y})\tag{D.31}$$

である．これを Fenchel-Young **双対**という．

Proof. 双対関数の定義 (定義 D.30) よりわかる．　□

定理 D.33. (1) $f^*:\mathbb{R}^n\to\mathbb{R}$ は凸関数である．
(2) $\displaystyle\lim_{|\boldsymbol{y}|_{2,n}\to\infty}\frac{|f^*(\boldsymbol{y})|_{2,n}}{|\boldsymbol{y}|_{2,n}}=+\infty$ となる．
(3) $f^{**}=f$ である．

Proof. 任意の $\boldsymbol{y}$, $\boldsymbol{w} \in \mathbb{R}^n$, $0 \leq \lambda \leq 1$ に対し

$$
\begin{aligned}
f^*\bigl(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{w}\bigr) &= \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&= \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\lambda \bigl\{ \langle \boldsymbol{y},\, \boldsymbol{x}[rangle_{2,\,n} - f(\boldsymbol{x}) \bigr\} \\
&\qquad + (1-\lambda) \bigl\{ \lambda \langle \boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \bigr\} \Bigr) \\
&\leq \lambda \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{y},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&\qquad + (1-\lambda) \max_{\boldsymbol{x} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{w},\, \boldsymbol{x} \rangle_{2,\,n} - f(\boldsymbol{x}) \Bigr) \\
&= \lambda f^*(\boldsymbol{y}) + (1-\lambda) f^*(\boldsymbol{w})
\end{aligned}
$$

がわかる. したがって 関数 f^* は凸である.
(2). 任意の $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$ に対して

$$
f(\boldsymbol{x}) + f^*(\boldsymbol{y}^*) \geq \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n}
$$

であった. ここで, $\boldsymbol{y} \neq \boldsymbol{0}$ と $\lambda > 0$ に対し

$$
\boldsymbol{x} = \frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}}
$$

とおく. このとき

$$
f^*(\boldsymbol{y}) \geq \left\langle \frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}},\, \boldsymbol{y} \right\rangle_{2,\,n} - f\left(\frac{\lambda \boldsymbol{y}}{|\boldsymbol{y}|_{2,\,n}} \right) \geq \lambda |\boldsymbol{y}|_{2,\,n} - \max_{|\boldsymbol{u}|_{2,\,n} = \lambda} f(\boldsymbol{u}).
$$

となることがわかる. したがって

$$
\frac{f^*(\boldsymbol{y})}{|\boldsymbol{y}|_{2,\,n}} \geq \lambda - \frac{1}{|\boldsymbol{y}|_{2,\,n}} \max_{|\boldsymbol{u}|_{2,\,n} = \lambda} f(\boldsymbol{u})
$$

となる. ここで

$$
\liminf_{|\boldsymbol{y}|_{2,\,n} \to \infty} \frac{f^*(\boldsymbol{y})}{|\boldsymbol{y}|_{2,\,n}} \geq \lambda > 0
$$

を得る. 最後に $\lambda \to \infty$ とすれば, (2) がわかる.
(3) まず, $f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \geq \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n}$ から

$$
f(\boldsymbol{x}) \geq \max_{\boldsymbol{y} \in \mathbb{R}^n} \Bigl(\langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n} - f^*(\boldsymbol{y}) \Bigr) = f^{**}(\boldsymbol{x}) \tag{D.32}
$$

となることに注意する. 逆に, 定理 D.29 から $\partial f(\boldsymbol{x}) \neq \varnothing$ であることがわかる. このことから $\boldsymbol{r} \in \partial f(\boldsymbol{x})$ を取る. すると $\boldsymbol{z} \in \mathbb{R}^n$ に対して

$$
f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{r},\, \boldsymbol{z} - \boldsymbol{x} \rangle_{2,\,n}
$$

となる. 上の式を書きかえると

$$\langle \boldsymbol{r},\, \boldsymbol{x}\rangle_{2,\,n} - f(\boldsymbol{x}) \geq \langle \boldsymbol{r},\, \boldsymbol{z}\rangle_{2,\,n} - f(\boldsymbol{z})$$

を得る. よって

$$\langle \boldsymbol{r},\, \boldsymbol{x}\rangle_{2,\,n} - f(\boldsymbol{x}) = \max_{\boldsymbol{z}\in\mathbb{R}^n}\Big(\langle \boldsymbol{r},\, \boldsymbol{z}\rangle_{2,\,n} - f(\boldsymbol{z})\Big) = f^*(\boldsymbol{r}) \tag{D.33}$$

がわかる. さらに, (D.33) から

$$f^{**}(\boldsymbol{x}) = \max_{\boldsymbol{y}\in\mathbb{R}^n}\Big(\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f^*(\boldsymbol{y})\Big) \geq \langle \boldsymbol{x},\, \boldsymbol{r}\rangle_{2,\,n} - f^*(\boldsymbol{r}) = f(\boldsymbol{x}) \tag{D.34}$$

がわかる. よって, (D.32) と (D.34) から $f^{**} = f$ がわかる. □

定理 D.34. $\boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n$ に対して次の条件は同値である.

(1) $\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} = f(\boldsymbol{x}) + f^*(\boldsymbol{y})$.

(2) $\boldsymbol{y} \in \partial f(\boldsymbol{x})$.

(3) $\boldsymbol{x} \in \partial f^*(\boldsymbol{y})$.

Proof. (1) $\Rightarrow$ (2) **の証明**. 任意の $\boldsymbol{z} \in \mathbb{R}^n$ に対し

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x}) = f^*(\boldsymbol{y}) \geq \langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,\,n} - f(\boldsymbol{z})$$

となる. 上の式の最左辺と最右辺から $\forall \boldsymbol{z} \in \mathbb{R}^n$ に対し

$$f(\boldsymbol{z}) \geq \langle \boldsymbol{y},\, \boldsymbol{z} - \boldsymbol{x}\rangle_{2,\,n} + f(\boldsymbol{x})$$

を得る. よって

$$\boldsymbol{y} \in \partial f(\boldsymbol{x})$$

が示せた.

(2) $\Rightarrow$ (1) **の証明**. $\boldsymbol{y} \in \partial f(\boldsymbol{x})$ のとき, $\forall \boldsymbol{z} \in \mathbb{R}^n$ に対し

$$f(\boldsymbol{z}) \geq \langle \boldsymbol{y},\, \boldsymbol{z} - \boldsymbol{x}\rangle_{2,\,n} + f(\boldsymbol{z}) \;\Rightarrow\; \langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x}) \geq \langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,\,n} - f(\boldsymbol{z})$$

となる. よって

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} - f(\boldsymbol{x}) \geq \max_{\boldsymbol{z}\in\mathbb{R}^n}\Big(\langle \boldsymbol{y},\, \boldsymbol{z}\rangle_{2,\,n} - f(\boldsymbol{z})\Big) = f^*(\boldsymbol{y}) \tag{D.35}$$

から

$$\langle \boldsymbol{x},\, \boldsymbol{y}\rangle_{2,\,n} \geq f(\boldsymbol{x}) + f^*(\boldsymbol{y})$$

を得る. Fenchel-Young の不等式 (D.31) から

$$\langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n} \leq f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \tag{D.36}$$

である. よって, (D.35) と (D.36) を合わせると

$$\langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n} = f(\boldsymbol{x}) + f^*(\boldsymbol{y})$$

が示せた.

(1) $\Leftrightarrow$ (3) も同様に示すことができる. □

D.4　非線形最適化と KKT 条件

$n,\, p \in \mathbb{N}$ とし, $f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とする. 次の記号を導入する.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix},$$

$$\nabla \boldsymbol{h} = \begin{bmatrix} \left(\nabla h_1\right)^\top \\ \left(\nabla h_2\right)^\top \\ \vdots \\ \left(\nabla h_p\right)^\top \end{bmatrix} = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_2} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \dfrac{\partial h_2}{\partial x_1} & \dfrac{\partial h_2}{\partial x_2} & \cdots & \dfrac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial h_p}{\partial x_1} & \dfrac{\partial h_p}{\partial x_2} & \cdots & \dfrac{\partial h_p}{\partial x_n} \end{bmatrix}.$$

この節では, 制約付く最適化問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to} \quad \boldsymbol{h} \leq \boldsymbol{0}_p \tag{D.37}$$

の解 $\boldsymbol{x}^*$ をみつけたい. ただし $\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}_p$ は $\forall j \in \{1,\, 2,\, \ldots,\, p\}$ に対して $h_j(\boldsymbol{x}) \leq 0$ が成立していることである.

$\boldsymbol{h}(\boldsymbol{x}) \leq \boldsymbol{0}_p$ をみたす点 $\boldsymbol{x}^*$ を**実行可能解**という. 実行可能解 $\boldsymbol{x}^*$ に対して $h_j(\boldsymbol{x}^*) = 0$ となる制約式を**有効制約式**という. 解 $\boldsymbol{x}^*$ に対する有効制約式の添え字集合を J と書く. すなわち

$$J = \{j \in \{1,\, 2,\, \ldots,\, p\};\, h_j(\boldsymbol{x}^*) = 0\}$$

である.

記号 関数 $\boldsymbol{r} : \mathbb{R} \to \mathbb{R}^n$ に対して

$$\boldsymbol{r} = o(t) \;\Leftrightarrow\; \lim_{t \to 0+} \frac{|\boldsymbol{r}(t)|_{2,\,n}}{t} = 0$$

と定める.

定義 D.35. $\boldsymbol{x}^*$ における制約 qualification 条件 (CQ 条件) が成立するとは

$$\left\langle \boldsymbol{y},\, \nabla h_j(\boldsymbol{x}) \right\rangle_{2,\,n} \le 0 \qquad (j \in J) \tag{D.38}$$

をみたす $\boldsymbol{y} \in \mathbb{R}^n$ にたいしてある正数 t_0 と連続曲線 $\{\boldsymbol{x}(t),\, 0 \le t < t_0\}$ があって

$$\boldsymbol{h}(\boldsymbol{x}) \le \boldsymbol{0}_p \qquad (0 \le t < t_0) \tag{D.39}$$

かつ

$$\boldsymbol{x}(t) = \boldsymbol{x}^* + t\boldsymbol{y} + o(t) \qquad (t \to 0+) \tag{D.40}$$

が成立するときをいう.

注意 D.36. 条件 (D.39) は $\boldsymbol{x}(t)\,(0 \le t < t_0)$ は実行可能解であることを保証している. (D.40) は $\boldsymbol{x}(t)$ の $t = 0$ における右微分係数ベクトル $\dot{\boldsymbol{x}}\,(0)$ が存在し

$$\dot{\boldsymbol{x}}\,(0) = \boldsymbol{y}$$

であることを保証している. □

定理 D.37. (Karush-Kuhn-Tucker(KKT) 条件) $\boldsymbol{x}^*$ は問題 (D.37) の解とし, $\boldsymbol{x}^*$ において CQ 条件は成立しているとする. このとき実数 $\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*$ が存在して

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.41}$$

が成立する. さらに $\boldsymbol{\lambda}^* := (\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*)^\top$ は

$$\boldsymbol{\lambda}^* \ge \boldsymbol{0}_p, \quad \left\langle \boldsymbol{\lambda}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} = 0 \tag{D.42}$$

をみたしている;

注意 D.38. 条件 (D.41) と (D.42) を合わせて Karush-Kuhn-Tuckeer 条件 (KKT 条件) という. (D.42) は

$$\nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda}^* = \boldsymbol{0}_n$$

とも表すことができる. $\boldsymbol{\lambda}^*$ は Lagrange 乗数という. $\left\langle \boldsymbol{\lambda}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,n} = 0$ を相補性条件という. 定理の証明からわかることであるが $\lambda_j^* = 0$ か $h_j(\boldsymbol{x}^*) = 0$ のいずれかが成立しているのでこのように呼んでいる. □

Proof. **定理** D.37 **の証明**. ベクトル $\boldsymbol{y} \in \mathbb{R}^n$ は (D.40) をみたしているとする．CQ 条件で保証されている $\boldsymbol{y}$ に対応する曲線を $\{\boldsymbol{x}(t);\, 0 \leq t < t_0\}$ とする．いま

$$\phi(t) := f(\boldsymbol{x}(t)) \tag{D.43}$$

とおく．このとき $\boldsymbol{x}^*$ は問題 (D.37) の解なので

$$\phi(0) = f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}(t)) = \phi(t) \quad (0 \leq t < t_0)$$

が成り立つ．したがって ϕ は $t = 0$ で最小値を取るので

$$\dot{\phi}\,(0) \geq 0$$

となる．さらに, (D.43) を t に関して微分して, $t = 0$ を代入すると

$$\dot{\phi}\,(0) = \langle \nabla f(\boldsymbol{x}^*),\, \dot{\boldsymbol{x}}\,(0) \rangle_{2,n} = \langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{y} \rangle_{2,n}$$

を得る．この式から (D.38) をみたす $\boldsymbol{y} \in \mathbb{R}^n$ に対し

$$\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{y} \rangle_{2,n} \geq 0$$

となるとなることがわかる．よって

$$\langle \boldsymbol{y},\, \nabla h_j(\boldsymbol{x}^*) \rangle_{2,n} \leq 0 \,(j \in J) \;\Rightarrow\; \langle \boldsymbol{y},\, \nabla f(\boldsymbol{x}^*) \rangle_{2,n} \geq 0 \tag{D.44}$$

を示せた．ここで Farkas の交代定理 (定理 D.20) を思い出す．以下の (1) か (2) のいずれかは成立する．

(1) $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, $\boldsymbol{x} \geq \mathbf{0}_n$ は解 $\boldsymbol{x}$ を持つ．

(2) $\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n$, $\langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,n} < 0$ は解を持つ．

さらに (1) と (2) は同時に成立しない．ただし, $\mathbf{A}$ は $n \times n$ 行列である．$J = \{j_1,\, j_2,\, \ldots,\, j_k\}$ と書き

$$\begin{aligned} \mathbf{A} &= -\bigl[\nabla h_{j_1}(\boldsymbol{x}^*),\, \nabla h_{j_2}(\boldsymbol{x}^*),\, \ldots,\, \nabla h_{j_k}(\boldsymbol{x}^*)\bigr], \\ \boldsymbol{b} &= \nabla f(\boldsymbol{x}^*) \end{aligned}$$

に対して Farkas の交代定理を適用する．(D.44) は

$$\mathbf{A}^\top \boldsymbol{y} \geq \mathbf{0}_n \;\Rightarrow\; \langle \boldsymbol{y},\, \boldsymbol{b} \rangle_{2,n} \geq 0$$

を主張しているので, Farkas 交代定理の (2) は成立しない．結果 Farkas 交代定理の (1) が成立する．つまり $\sigma_j \geq 0 \,(j \in J)$ が存在して

$$-\sum_{j \in J} \sigma_j \nabla h_j(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)$$

が成り立つ. そこで $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_p^*)^\top$ を

$$\lambda_j^* = \begin{cases} \sigma_j & (j \in J) \\ 0 & (j \notin J) \end{cases}$$

と定めると $j \in J$ のとき $h_j(\boldsymbol{x}^*) = 0$, $j \notin J$ のとき, $\lambda_j^* = 0$ なので

$$\langle \boldsymbol{\lambda}^*, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0$$

となる. さらに Frakas の交代定理の (1) から

$$\begin{aligned} &\boldsymbol{\lambda}^* \geq \mathbf{0}_p, \\ &\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}_n \end{aligned}$$

もわかる. □

注意 D.39. CQ 条件をみたす制約の例を上げておく

(1) $\{h_j\}_{j=1}^p$ は $\boldsymbol{x}$ の線型関数.

(2) $\{\nabla h_j(\boldsymbol{x}^*)\}_{j \in J}^p$ は $\mathbb{R}^n$ の線型独立なベクトル.

これらが CQ 条件をみたすことの証明は Evans の pp.115-117 を参照のこと. □

D.5 Lagrange 乗数についてのさらなる議論

$f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ 連続微分可能とし

$$\boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$$

とおく. この節では, 下記の最小化問題

$$\text{minimize} \quad f(\boldsymbol{x}) \qquad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}_m \text{ and } \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0}_p. \tag{D.45}$$

を考える.

定理 D.40. $\boldsymbol{x}^* \in \mathbb{R}^n$ は問題 (D.45) の解とする. このとき実数の組

$$\gamma^*, \quad \boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, \ldots, \mu_m^*)^\top, \quad \boldsymbol{\lambda}^* = (\lambda_1^*, \quad \lambda_2^*, \ldots, \lambda_p^*)^\top$$

が存在し, これはすべては 0 に同時になることはなく, さらに

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \sum_{k=1}^{m} \mu_k^* \nabla g_k(\boldsymbol{x}^*) + \sum_{j=1}^{p} \lambda_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}_n \tag{D.46}$$

と

$$\gamma^* \geq 0, \quad \boldsymbol{\lambda} \geq \mathbf{0}_p, \quad \langle \boldsymbol{\lambda}, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0 \tag{D.47}$$

をみたす.

注意 D.41. (1) (D.46) は

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*) \boldsymbol{\lambda}^* = \mathbf{0}_n$$

と書き直すことができる. ただし

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix},$$

$$\nabla \boldsymbol{h} = \begin{bmatrix} (\nabla h_1)^\top \\ (\nabla h_2)^\top \\ \vdots \\ (\nabla h_p)^\top \end{bmatrix} = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_2} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \dfrac{\partial h_2}{\partial x_1} & \dfrac{\partial h_2}{\partial x_2} & \cdots & \dfrac{\partial h_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial h_p}{\partial x_1} & \dfrac{\partial h_p}{\partial x_2} & \cdots & \dfrac{\partial h_p}{\partial x_n} \end{bmatrix},$$

$$\nabla \boldsymbol{g} = \begin{bmatrix} (\nabla g_1)^\top \\ (\nabla g_2)^\top \\ \vdots \\ (\nabla g_m)^\top \end{bmatrix} = \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} & \cdots & \dfrac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_m}{\partial x_1} & \dfrac{\partial g_m}{\partial x_2} & \cdots & \dfrac{\partial g_m}{\partial x_n} \end{bmatrix}$$

である.

(2) $\gamma^* = 1$ のとき (D.46) と (D.47) は

$$\nabla f(\boldsymbol{x}^*) + \nabla \boldsymbol{g}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*) \boldsymbol{\lambda}^* = \mathbf{0}_n \tag{D.48}$$

と

$$\boldsymbol{\lambda} \geq \mathbf{0}_p, \quad \langle \boldsymbol{\lambda}, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,p} = 0 \tag{D.49}$$

となり KKT 条件となる.

$\gamma^* \neq 0$ のときは γ^* または $-\gamma^*$ で割ることで (D.46) と (D.47) を KKT 条件に直すことができる.

$\gamma^* = 0$ のとき, 異常な乗数とよぶ. □

Proof. **(定理** D.40 **の証明)**. $k = 1, 2, \ldots$ に対して

$$F_k(\boldsymbol{x}) := f(\boldsymbol{x}) + \frac{k}{2}\left(|\boldsymbol{g}(\boldsymbol{x})|_{2,m}^2 + |\boldsymbol{h}^+(\boldsymbol{x})|_{2,p}^2\right)^2 + \frac{1}{2}|\boldsymbol{x} - \boldsymbol{x}^*|_{2,n}$$

とおく. ただし $\ell = 1, 2, \ldots, p$ に対し

$$h_\ell^+(\boldsymbol{x}) = \begin{cases} h_\ell(\boldsymbol{x}) & (h_\ell(\boldsymbol{x}) \geq 0 \text{ の場合}) \\ 0 & (h_\ell(\boldsymbol{x}) = 0 \text{ の場合}) \end{cases}$$

である. ここで $B = \{\boldsymbol{x} \in \mathbb{R}^n;\ |\boldsymbol{x} - \boldsymbol{x}^*|_{2,n} \leq 1\}$ とおく. すると B はコンパクトで F_k は連続関数なので, 点 $\boldsymbol{x}_k \in B$ があって

$$F_k(\boldsymbol{x}_k) = \min_{\boldsymbol{x} \in B} F_k(\boldsymbol{x})$$

が成り立つ. したがって

$$F_k(\boldsymbol{x}_k) \leq F_k(\boldsymbol{x}^*)$$

である. さらに $\boldsymbol{x}^*$ は問題 (D.45) の解なので, $\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}_m$ かつ $\boldsymbol{h}(\boldsymbol{x}^*) \leq \mathbf{0}_p$ となる. よって h_ℓ^+ の定義から $\boldsymbol{h}^+(\boldsymbol{x}^*) = \mathbf{0}_p$ となる. したがって

$$f(\boldsymbol{x}_k) + \frac{k}{2}\left(|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}^2 + |\boldsymbol{h}^+(\boldsymbol{x}_k)|_{2,p}^2\right) + \frac{1}{2}|\boldsymbol{x}_k - \boldsymbol{x}^*|_{2,n}^2 = F_k(\boldsymbol{x}_i) \tag{D.50}$$

$$\leq F_k(\boldsymbol{x}^*) = f(\boldsymbol{x}^*) + \frac{k}{2}\left(\underbrace{|\boldsymbol{g}(\boldsymbol{x}^*)|_{2,m}^2}_{=0} + \underbrace{|\boldsymbol{h}^+(\boldsymbol{x}^*)|_{2,p}^2}_{=0}\right) = f(\boldsymbol{x}^*) \tag{D.51}$$

が成り立つ. したがって, $\left\{\frac{k}{2}|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}\right\}_{k=1}^\infty$ と $\left\{\frac{k}{2}|\boldsymbol{h}^+(\boldsymbol{x}_k)|_{2,p}\right\}_{k=1}^\infty$ は有界なので

$$\lim_{k\to\infty} \boldsymbol{g}(\boldsymbol{x}_k) = \mathbf{0}_m, \quad \lim_{k\to\infty} \boldsymbol{h}^+(\boldsymbol{x}_k) = \mathbf{0}_p \tag{D.52}$$

となる．そうでなければ, $\{k/2\}_{k=1}^{\infty}$ は発散するので, $\left\{\frac{k}{2}|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}\right\}_{k=1}^{\infty}$ と $\left\{\frac{k}{2}|\boldsymbol{h}^{+}(\boldsymbol{x}_k)|_{2,p}\right\}_{k=1}^{\infty}$ にはならない．さらに Bolzano-Weirstrass の定理から $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset B$ から収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ を選ぶことができることがわかる．以上から

$$\lim_{j\to\infty} \boldsymbol{x}_{k(j)} = \overline{\boldsymbol{x}} \in B$$

と書くことにする．(D.51) は

$$f(\boldsymbol{x}_k) + \frac{1}{2}|\boldsymbol{x}_k - \boldsymbol{x}^*|_{2,n}^2 \leq f(\boldsymbol{x}^*)$$

となる．したがって

$$f(\overline{\boldsymbol{x}}) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n}^2 \leq f(\boldsymbol{x}^*) \tag{D.53}$$

となる．しかし (D.52) から

$$\boldsymbol{h}(\overline{\boldsymbol{x}}) = \mathbf{0}_p, \quad \text{かつ} \quad \boldsymbol{g}(\overline{\boldsymbol{x}}) = \mathbf{0}_m$$

となるので, $\overline{\boldsymbol{x}}$ も実行可能となる．したがって

$$f(\boldsymbol{x}^*) \leq f(\overline{\boldsymbol{x}}) \tag{D.54}$$

を得る．(D.53) と (D.54) を合わせると

$$f(\boldsymbol{x}^*) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} \leq f(\overline{\boldsymbol{x}}) + \frac{1}{2}|\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} \leq f(\boldsymbol{x}^*) \Leftrightarrow |\overline{\boldsymbol{x}} - \boldsymbol{x}^*|_{2,n} = 0$$

を得る．したがって

$$\boldsymbol{x}^* = \overline{\boldsymbol{x}}$$

がわかる．収束する部分列 $\{\boldsymbol{x}_{k(j)}\}_{j=1}^{\infty}$ はすべて同じ収束先をもつので

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^* \tag{D.55}$$

がわかる．

$\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は $\boldsymbol{x}^*$ に収束するので, k を十分大きくとれば $\boldsymbol{x}_k$ は B の境界から離れている．このことより関数

$$B \ni \boldsymbol{x} \mapsto F_k(\boldsymbol{x}) \in \mathbb{R}$$

は制約なしの局所解 $\boldsymbol{x}_k$ を持つ．よって

$$\begin{aligned}
\mathbf{0}_n &= \nabla F_k(\boldsymbol{x}_k) \\
&= \nabla f(\boldsymbol{x}_k) + k\Big(\big(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{g}(\boldsymbol{x}_k) + \big(\nabla \boldsymbol{h}^{+}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{h}^{+}(\boldsymbol{x}_k)\Big) + \boldsymbol{x}^* - \boldsymbol{x}_k \\
&= \nabla f(\boldsymbol{x}_k) + k\Big(\big(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{g}(\boldsymbol{x}_k) + \big(\nabla \boldsymbol{h}^{+}(\boldsymbol{x}_k)\big)^{\top}\boldsymbol{h}(\boldsymbol{x}_k)\Big) + \boldsymbol{x}^* - \boldsymbol{x}_k
\end{aligned} \tag{D.56}$$

となる．最後の等号は $h_\ell^+ \neq 0\,(\ell = 1, 2, \ldots, p)$ のとき $\dfrac{\partial}{\partial x_\ell} h_\ell^+ = \dfrac{\partial}{\partial x_\ell} h_\ell$ であることからわかる．つぎに定数

$$\gamma_k := \left(1 + k^2|\boldsymbol{g}(\boldsymbol{x}_k)|_{2,m}^2 + k^2|\boldsymbol{h}(\boldsymbol{x}_k)|_{2,p}^2\right)^{-1/2} > 0$$

を (D.56) の最右辺に掛けると

$$\mathbf{0}_n = \gamma_k \nabla f(\boldsymbol{x}_k) + \left(\nabla \boldsymbol{g}(\boldsymbol{x}_k)\right)^\top \boldsymbol{\mu}_k + \left(\nabla \boldsymbol{h}(\boldsymbol{x}_k)\right)^\top \boldsymbol{\lambda}_k + \gamma_k(\boldsymbol{x}_k - \boldsymbol{x}^*) \tag{D.57}$$

を得る．ただし

$$\boldsymbol{\mu}_k := \gamma_k k \boldsymbol{g}(\boldsymbol{x}_k), \quad \boldsymbol{\lambda}_k =: \gamma_k k \boldsymbol{h}^+(\boldsymbol{x}_k) \geq \mathbf{0}_p$$

である．記号の定義から

$$\gamma_k^2 + |\boldsymbol{\mu}_k|_{2,m}^2 + |\boldsymbol{\lambda}_k|_{2,p}^2 = 1$$

なので，$\{(\gamma_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k)\}_{k=1}^\infty$ は有界である．したがって，収束する部分列 $\{(\gamma_{k(j)}, \boldsymbol{\mu}_{k(j)}, \boldsymbol{\lambda}_{k(j)})\}_{j=1}^\infty$ を取ることができる．すなわち $j \to \infty$ のとき

$$\gamma_{k(j)} \to \gamma_0 \geq \in \mathbb{R}, \quad \boldsymbol{\mu}_{k(j)} \to \boldsymbol{\mu}_0 \in \mathbb{R}^m, \quad \boldsymbol{\lambda}_{k(j)} \to \boldsymbol{\lambda}_0 \in \mathbb{R}^p$$

と書くことができる．このとき

$$\gamma_0^2 + |\boldsymbol{\mu}_0|_{2,m}^2 + |\boldsymbol{\lambda}|_{2,p}^2 = 1$$

なので

$$(\gamma_0, , \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0) \neq (0, \mathbf{0}_m, \mathbf{0}_p)$$

である．(D.57) で $k = k(j) \overset{j\to\infty}{\infty}$ とすると (D.55) から (D.46) を得る．

次に $\boldsymbol{h}^*(\boldsymbol{x}^*) \geq \mathbf{0}_p$ と $\boldsymbol{\lambda}^* \geq \mathbf{0}_p$ に注意する．ある $\ell \in \{1, 2, \ldots, p\}$ に対して

$$h_\ell(\boldsymbol{x}^*) < 0$$

ならば十分大きな k に対して

$$h_\ell(\boldsymbol{x}_k) < 0$$

となる．よって $h_\ell^+(\boldsymbol{x}_k) = 0$ なので λ_k の第 ℓ 成分は 0 となるので $\lambda_\ell^* = 0$ となる．よって

$$\langle \boldsymbol{\lambda}^*, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,n} = 0$$

が成り立つ．よって (D.47) が示された． □

単調増加関数の不連続点は高々可算個であることの証明をのせる．

D.6 変分不等式

$C \subset \mathbb{R}^n$ を空でない凸部分集合とする. 基本的な問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to}\ \boldsymbol{x} \in C \tag{D.58}$$

の解 $\boldsymbol{x}^*$ をみつけることを考える.

定理 D.42. (i) 関数 $f: \mathbb{R}^n \to \mathbb{R}$ は連続微分可能で, $\boldsymbol{x}^* \in C$ は問題 (D.58) の解のとき

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0 \qquad (\boldsymbol{x} \in C) \tag{D.59}$$

が成立する.
(ii) 関数 f は凸とする. $\boldsymbol{x}^* \in C$ が (D.59) をみたすとき, $\boldsymbol{x}^*$ は問題 (D.58) の解である.

注意 D.43. (D.59) を**変分不等式**という. これは制約問題 (D.58) の 1 次変分の形になっている. □

定理 D.42 **の証明**: (i) $\boldsymbol{x} \in C$ を任意の点とする. このとき, C の凸から $0 \leq t \leq 1$ に対して

$$\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*) = t\boldsymbol{x} + (1-t)\boldsymbol{x}^* \in C$$

である. さらに $\boldsymbol{x}^*$ は問題 (D.58) の解なので, 関数

$$\phi(t) := f\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big) \qquad (0 \leq t \leq 1)$$

は $t = 0$ で最小となる. したがって

$$\overset{\bullet}{\phi}(0) = \left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{t=0} \geq 0$$

である. ただし $\overset{\bullet}{\phi}(0)$ は $t = 0$ における ϕ の右微分係数である. しかし

$$\overset{\bullet}{\phi}(t) = \left\langle \nabla f\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n}$$

なので

$$\overset{\bullet}{\phi}(0) = \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0$$

となる.
(ii) 関数 f は点 $\boldsymbol{x}^*$ で連続微分可能なので

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \qquad (\boldsymbol{x} \in C)$$

である. (D.59) より

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \quad \geq 0$$

なので

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) \qquad (\boldsymbol{x} \in C)$$

となる. よって $\boldsymbol{x}^*$ は問題 (D.58) の解である.

例 D.44. 関数 $f : \mathbb{R}^n \to \mathbb{R}$ は連続微分可能とし, $\mathbf{A} \in \mathrm{Mat}(p,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^p$ とする. 問題

$$\text{minimize}\, f(\boldsymbol{x}) \qquad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \tag{D.60}$$

を考える.

$\boldsymbol{x}^*$ は問題 (D.60) の解のとき (D.59) は

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0 \qquad (\forall \boldsymbol{x} \in \mathbb{R}^n;\, \mathbf{A}\boldsymbol{x} = \boldsymbol{b}) \tag{D.61}$$

である. $\boldsymbol{x} \in \mathbb{R}^n$ は

$$\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{w} \qquad (\boldsymbol{x},\, \boldsymbol{w} \in \ker(\mathbf{A}))$$

の形で書く. すると (D.61) から

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} = \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \geq 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

となる. $\boldsymbol{w}$ を $-\boldsymbol{w}$ に置き換えると

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} \leq 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

となるので

$$\left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{w} \right\rangle_{2,\,n} = 0 \qquad (\forall \boldsymbol{w} \in \ker(\mathbf{A}))$$

がわかる. したがって

$$\nabla f(\boldsymbol{x}^*) \in \big(\ker(\mathbf{A})\big)^{\perp} = \mathrm{range}(\mathbf{A}^{\top})$$

がわかる. したがって, $\nabla f(\boldsymbol{x}^*)$ が $\mathbf{A}^{\top}$ の像に含まれるので, ある $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ が存在して

$$\nabla f(\boldsymbol{x}^*) = \mathbf{A}^{\top}(-\boldsymbol{\lambda}^*) \;\Leftrightarrow\; \nabla f(\boldsymbol{x}^*) + \mathbf{A}^{\top}\boldsymbol{\lambda}^* = \mathbf{0}_n$$

と書ける. このとき $\boldsymbol{\lambda}^* = (\lambda_1^*,\, \lambda_2^*,\, \ldots,\, \lambda_p^*)^{\top}$ は制約 $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ の Lagrange 乗数となる. □

D.7　凸性と Lagrange 乗数

問題

$$\text{minimize } f(\boldsymbol{x}) \qquad \text{subject to } \boldsymbol{h} \leq \mathbf{0}_p \tag{D.62}$$

の解 $\boldsymbol{x}^*$ をみつけるときの Lagrange 乗数の存在を定理 D.37 は保証している． これは関数の凸性の仮定を用いていない． しかし多くの場合 qualification 条件を確認することが容易でない. この節では強い仮定

$$f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R} \text{ は凸関数} \tag{D.63}$$

を課す.

まず凸関数にたいして, KKT 条件は最適解の十分条件になることを示す.

定理 D.45. $f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$ は凸関数かつ連続微分可能とする. $\boldsymbol{x}^* \in \mathbb{R}^n$ は

$$\boldsymbol{h}(\boldsymbol{x}^*) \leq \mathbf{0}_p \tag{D.64}$$

をみたすとする. さらに, ある $\boldsymbol{\mu}^* = (\mu_1^*,\, \mu_2^*,\, \ldots,\, \mu_p^*)^\top \in \mathbb{R}^p$ が存在して KKT 条件

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}_n \tag{D.65}$$

$$\boldsymbol{\mu}^* \geq \mathbf{0}_p, \qquad \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*)\rangle_{2,\,p} \geq 0 \tag{D.66}$$

が成立するとする. このとき $\boldsymbol{x}^*$ は問題 (D.62) の解である.

Proof. $C := \{\boldsymbol{x} \in \mathbb{R}^n;\, \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0}\}$ は実行可能解の集合とする. 各 $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ は凸関数なので

$$C_j := \{\boldsymbol{x} \in \mathbb{R}^n;\, h_j(\boldsymbol{x}) \leq 0\}$$

は凸集合である. したがって

$$C = \bigcap_{j=1}^{p} C_j$$

も凸集合となる. (D.65) は

$$\begin{aligned} 0 &= \Big\langle \underbrace{\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*)}_{=\mathbf{0}_n},\, \boldsymbol{x} - \boldsymbol{x}^* \Big\rangle_{2,\,n} \\ &= \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n} + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \qquad (\forall \boldsymbol{x} \in \mathbb{R}^n) \end{aligned} \tag{D.67}$$

となる. $\boldsymbol{x} \in C$ である. C の定義と $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ の凸性と (D.64) より

$$h_j(\boldsymbol{x}^*) + \big\langle \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n} \le h_j(\boldsymbol{x}) \le 0 \tag{D.68}$$

がわかる. なぜならば, 関数 $\phi(t) := h_j\big(\boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*)\big)\,(0 \le t \le 1)$ を考える. すると関数 h_j の凸性から関数 ϕ も凸となる. したがって

$$h_j(0) + t\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} \le h_j(t) \tag{D.69}$$

を得る. さらに

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} = \big\langle \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n}$$

であることに注意して, (D.69) に $t = 1$ を代入すると

$$h_j(\boldsymbol{x}^*) + \big\langle \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n} = \phi(0) + \frac{\mathrm{d}}{\mathrm{d}t}\phi(t)\bigg|_{t=0} \le \phi(1) = h_j(\boldsymbol{x})$$

を得る. (D.68) の両辺を $\mu_j^*\,(\ge 0)$ 倍して和をとれば

$$0 \ge \underbrace{\big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \big\rangle_{2,\,p}}_{\ge 0 \quad \because (\mathrm{D.66})} + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n}$$

がわかる. したがって, 上式と (D.65) から

$$\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le 0 \Leftrightarrow \left\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \right\rangle_{2,\,n} \ge 0 \qquad (\forall \boldsymbol{x} \in C) \tag{D.70}$$

がわかる. 関数 f の凸性と (D.70) から

$$f(\boldsymbol{x}) \ge f(\boldsymbol{x}^*) + \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x} - \boldsymbol{x}^* \big\rangle_{2,\,n} \ge f(\boldsymbol{x}^*) \qquad (\forall \boldsymbol{x} \in C)$$

を得る. したがって $\boldsymbol{x}^*$ は問題 (D.62) の解であることが示せた. □

定義 D.46. 問題 (D.62) を考える.

$$\text{ある } \overline{\boldsymbol{x}} \in \mathbb{R}^n \text{ が存在して } \boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p \tag{D.71}$$

が成り立つとき, 問題 (D.62) に対する **Slater 条件**が成立するという. ただし, $\boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p$ は, $h_j(\overline{\boldsymbol{x}}) < 0\,(\forall j = 1,\, 2,\, \ldots,\, p)$ が成り立つことである.

定理 D.47. Slater 条件 (D.71) が成立するとする.

$$f,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$$

は凸関数かつ連続微分可能とし, $\boldsymbol{x}^*$ をは問題 (D.62) の解とする.
(i) このとき $\boldsymbol{\mu}^* \in \mathbb{R}^p$ が存在して KKT 条件 (D.65) と (D.66) が成り立つ.
(ii) さらに

$$\boldsymbol{x} \mapsto f(\boldsymbol{x}) + \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}) \rangle_{2,\,p}$$

は点 $\boldsymbol{x}^*$ で最小となる.

Proof. 定理 D.37 の John の formulation から, ある 非負実数の組 $\gamma^*,\, \mu_1^*,\, \mu_2^*,\, \ldots,\, \mu_p^*$ が存在 (すべてが同時に零になることはない.) して

$$\gamma^* \nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.72}$$

$$\gamma^* \geq 0, \qquad \boldsymbol{\mu}^* \geq \boldsymbol{0}_p, \qquad \langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \rangle_{2,\,p} \geq 0 \tag{D.73}$$

が成立する.
　まず

$$\gamma^* \neq 0 \tag{D.74}$$

を示す. そのために $\gamma^* = 0$ を仮定して矛盾を導く. すべては同時に零にならないので $\boldsymbol{\mu}^* \neq \boldsymbol{0}_p$ である. 一方, $\gamma^* = 0$ の仮定と (D.72) から

$$\sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n \tag{D.75}$$

となる. したがって, (D.75) と Slater 条件 D.71 から

$$\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} = 0 \quad \text{かつ} \quad \boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p \tag{D.76}$$

である. しかし $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ は凸関数なので

$$h_j(\boldsymbol{x}^*) + \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \leq h_j(\overline{\boldsymbol{x}})$$

が成り立つ. 上式の辺々と $\mu_j^*\,(j = 1,\, 2, \ldots,\, p)$ の積を取り, j について和を取ると

$$\begin{aligned}
\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} &= \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) = \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \left\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \\
&= \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} + \underbrace{\left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n}}_{=0 \quad \because (\text{D.76})} \\
&= \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p}
\end{aligned}$$

から

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} = \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} \tag{D.77}$$

を得る．よって, $\boldsymbol{\mu}^* \geq \boldsymbol{0}_p$ かつ $\boldsymbol{\mu}^* \neq \boldsymbol{0}_p$ で $\boldsymbol{h}(\overline{\boldsymbol{x}}) < \boldsymbol{0}_p$ $\big(\because$ (D.74)$\big)$ なので, (D.79) と合わせると

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} < 0$$

となる．これは (D.73) と矛盾する．背理法の仮定から $\gamma^* \neq 0$ となる．よって, このことと D.73) から $\gamma^* > 0$ が示せた.

次に, (D.72) を $\gamma^* > 0$ で割り

$$\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \frac{\mu_j^*}{\gamma^*} \nabla h_j(\boldsymbol{x}) = \boldsymbol{0}_n$$

を考える．$\widetilde{\mu}_j^* = \mu_j^*/\gamma^* \geq 0\,(j = 1,\, 2,\, \ldots,\, p)$ とおけば, (D.73) から

$$\begin{aligned} &\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}_n, \\ &\widetilde{\boldsymbol{\mu}} = (\widetilde{\mu}_1^*,\, \widetilde{\mu}_2^*,\, \ldots,\, \widetilde{\mu}_p^*) \geq \boldsymbol{0}_p, \\ &\left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} = 0 \end{aligned} \tag{D.78}$$

を得る．ただし, $\widetilde{\boldsymbol{\mu}}^* = (\widetilde{\mu}_1^*,\, \widetilde{\mu}_2^*,\, \ldots,\, \widetilde{\mu}_p^*)^\top$ とした．関数 $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ の凸性より

$$h_j(\overline{\boldsymbol{x}}) \geq h_j(\boldsymbol{x}^*) + \left\langle \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x}^* - \overline{\boldsymbol{x}} \right\rangle_{2,\,n}$$

を得る．上式の辺々に $\widetilde{\mu}_j$ を掛けて, j について和ををると

$$\left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} \geq \left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} + \left\langle \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}^*),\, \boldsymbol{x}^* - \overline{\boldsymbol{x}} \right\rangle_{2,\,p} \tag{D.79}$$

を得る．さらに f の凸性から

$$f(\overline{\boldsymbol{x}}) \geq f(\boldsymbol{x}^*) + \left\langle \nabla f(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \tag{D.80}$$

がわかる．(D.79) と (D.80) を合わせると

$$\begin{aligned} f(\overline{\boldsymbol{x}}) + \left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} &\geq f(\boldsymbol{x}^*) + \left\langle \nabla f(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} + \left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} \\ &\qquad + \left\langle \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \\ &= f(\overline{\boldsymbol{x}}) + \left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} \\ &\qquad + \left\langle \underbrace{\nabla f(\boldsymbol{x}^*) + \sum_{j=1}^{p} \widetilde{\mu}_j^* \nabla h_j(\boldsymbol{x}),}_{=\boldsymbol{0}_n \quad \because (\mathrm{D.78})} \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \\ &= f(\boldsymbol{x}^*) + \left\langle \widetilde{\boldsymbol{\mu}}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} \end{aligned}$$

を得る．よって (ii) は示された． □

D.8 凸双対性

D.8.1 双対問題

与えられた関数 $f,\, g_1,\, g_2,\, \ldots,\, g_m,\, h_1,\, h_2,\, \ldots,\, h_p : \mathbb{R}^n \to \mathbb{R}$ に対して

$$\boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}, \quad \boldsymbol{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$

と書くことにする．

問題

$$\text{minimize}\, f(\boldsymbol{x}) \quad \text{subject to} \;\; g(\boldsymbol{x}) = \mathbf{0}_m \text{ and } h(\boldsymbol{x}) \leq \mathbf{0}_p \tag{D.81}$$

の解を $\boldsymbol{x}^*$ をみつけたいとする．

定義 D.48. 問題 (D.81) に対応する Lagrange 問題 $L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu})\,(\boldsymbol{x} \in \mathbb{R}^n,\, \boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^p)$ を

$$L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \big\langle \boldsymbol{\lambda},\, \boldsymbol{g}(\boldsymbol{x}) \big\rangle_{2,\,m} + \big\langle \boldsymbol{\mu},\, \boldsymbol{h}(\boldsymbol{x}) \big\rangle_{2,\,p}$$

で定める．さらに

$$L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu})$$

と定める．

注意 D.49. $\boldsymbol{\lambda} = (\lambda_1,\, \lambda_2,\, \ldots,\, \lambda_m)^\top,\, \boldsymbol{\mu} = (\mu_1,\, \mu_2,\, \ldots,\, \mu_p)^\top$ と書いたとき

$$L(\boldsymbol{x},\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) + \sum_{j=1}^{p} \mu_j h_j(\boldsymbol{x})$$

とも書ける． □

例 D.50. 関数 f は凸とし，問題

$$\text{minimize}\, f(\boldsymbol{x}) \quad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{x} \geq \mathbf{0}_n \tag{D.82}$$

を考える．ただし $\mathrm{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^m$ とする．$\boldsymbol{\lambda} \in \mathbb{R}^m$ と $\boldsymbol{\mu} \leq \mathbf{0}_n\,(\boldsymbol{\mu} \in \mathbb{R}^n)$ を取り, $\boldsymbol{h}(\boldsymbol{x}) = -\boldsymbol{x}$ とおくと

$$\begin{aligned} L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \big\langle \boldsymbol{\lambda},\, \boldsymbol{g}(\boldsymbol{x}) \big\rangle_{2,\,m} + \big\langle \boldsymbol{\mu},\, \boldsymbol{h}(\boldsymbol{x}) \big\rangle_{2,\,n} \right\} \\ &= \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \big\langle \boldsymbol{\lambda},\, \mathrm{A}\boldsymbol{x} - \boldsymbol{b} \big\rangle_{2,\,m} - \big\langle \boldsymbol{\mu},\, \boldsymbol{x} \big\rangle_{2,\,n} \right\} \\ &= -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,m} + \inf_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \big\langle \mathrm{A}^\top \boldsymbol{\lambda} - \boldsymbol{\mu},\, \boldsymbol{x} \big\rangle_{2,\,n} \right\} \\ &= -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,m} - \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \big\langle \boldsymbol{\mu} - \mathrm{A}^\top \boldsymbol{\lambda},\, \boldsymbol{x} \big\rangle_{2,\,n} - f(\boldsymbol{x}) \right\} \end{aligned}$$

となる．したがって

$$L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) = -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,n} - f^*\big(\boldsymbol{\mu} - \mathrm{A}^\top \boldsymbol{\lambda}\big)$$

となる．ただし

$$f^*(\boldsymbol{x}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \langle \boldsymbol{x},\, \boldsymbol{y} \rangle_{2,\,n} - f(\boldsymbol{x}) \right\}$$

は凸関数 f の双対関数である． □

定義 D.51. 問題 (D.81) に対する双対問題は

$$\text{minimize}\, L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) \quad \text{subject to } \boldsymbol{\mu} \geq \mathbf{0}_p \tag{D.83}$$

の解 $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, $\boldsymbol{\mu}^* \in \mathbb{R}^p$ をみつけることである．

例 D.52. $f(\boldsymbol{x}) = \langle \boldsymbol{x},\, \boldsymbol{a} \rangle_{2,\,n}$ とする．ただし $\boldsymbol{a} \in \mathbb{R}^n$ は定数ベクトルである．線型計画法

$$\text{minimize } \langle \boldsymbol{x},\, \boldsymbol{a} \rangle_{2,\,n} \quad \text{subject to } \mathrm{A}\boldsymbol{x} = \boldsymbol{b},\, \boldsymbol{x} \geq \mathbf{0}_n$$

を考える．ただし $\mathrm{A} \in \mathrm{Mat}(m,\, n;\, \mathbb{R})$, $\boldsymbol{b} \in \mathbb{R}^m$ である．いま

$$f^*(\boldsymbol{y}) = \max_{\boldsymbol{x} \in \mathbb{R}^n} \big\{ \langle \boldsymbol{y} - \boldsymbol{a},\, \boldsymbol{x} \rangle_{2,\,n} \big\} = \begin{cases} 0 & (\boldsymbol{y} = \boldsymbol{a}) \\ \infty & (\text{その他の場合}) \end{cases}$$

と定める．したがって

$$\begin{aligned} L^*(\boldsymbol{\lambda},\, \boldsymbol{\mu}) &= -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,m} - f^*\big(\boldsymbol{\mu} - \mathrm{A}^\top \boldsymbol{\lambda}\big) \\ &= \begin{cases} -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,n} & (\boldsymbol{\mu} - \mathrm{A}^\top \boldsymbol{\lambda} = \boldsymbol{a}) \\ -\infty & (\text{その他の場合}) \end{cases} \end{aligned}$$

となる．このとき双対問題は

$$\text{minimize } -\langle \boldsymbol{\lambda},\, \boldsymbol{b} \rangle_{2,\,m} \quad \text{subject to } \boldsymbol{\mu} - \mathrm{A}^\top \boldsymbol{\lambda} = \boldsymbol{a} = \mathbf{0}_p \text{ and } \boldsymbol{\mu} \geq \mathbf{0}_p$$

となる． □

定理 D.53. (弱双対不等式) 不等式

$$\sup_{\boldsymbol{\mu}\in\mathbb{R}^p,\,\boldsymbol{\lambda}\in\mathbb{R}^m} L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le \inf_{\boldsymbol{x}\in\mathbb{R}^n;\,\boldsymbol{g}(\boldsymbol{x})=\mathbf{0}_m,\,\boldsymbol{h}(\boldsymbol{x})\le\mathbf{0}_p} f(\boldsymbol{x}) \tag{D.84}$$

が成立する．この不等式が厳密に成立するとき，**双対ギャップ**があるという．

Proof. $\boldsymbol{\mu} \ge \mathbf{0}_p$, $\boldsymbol{\lambda} \in \mathbb{R}^m$, , $\boldsymbol{g} = \mathbf{0}_m$, $\boldsymbol{h} \le \mathbf{0}_p$ とする．このとき

$$L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) = f(\boldsymbol{x}) + \underbrace{\langle\boldsymbol{\mu},\,\boldsymbol{h}(\boldsymbol{x})\rangle_{2,\,p}}_{\le 0} + \langle\boldsymbol{\lambda},\,\underbrace{\boldsymbol{g}(\boldsymbol{x})}_{=0}\rangle_{2,\,m} \le f(\boldsymbol{x})$$

となる．さらに

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu})$$

より

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \le f(\boldsymbol{x}).$$

よって (D.84) が示された． □

D.8.2 Slaker **条件** (再訪問)

$\mathbf{A} \in \mathrm{Mat}(m,\,n;\,\mathbb{R})$, $m \le n$, $\boldsymbol{b} \in \mathbb{R}^m$ とし

$$\boldsymbol{g}(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} - \boldsymbol{b}$$

とおく．さらに

$$\mathrm{rank}\,(\mathbf{A}) = m$$

とする．

主問題

$$\text{minimize } f(\boldsymbol{x}) \quad \text{subject to } \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{h}(\boldsymbol{x}) \le \mathbf{0}_p \tag{D.85}$$

の解 $\boldsymbol{x}^* \in \mathbb{R}^n$ をみつけた．対応する双対問題

$$\text{maximize } L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) \quad \text{subject to } \boldsymbol{\mu} \ge \mathbf{0}_p \tag{D.86}$$

の解 $\boldsymbol{\mu}^* \in \mathbb{R}^p$, $\boldsymbol{\lambda} \in \mathbb{R}^m$ をみつけたい．ただし

$$L^*(\boldsymbol{\lambda},\,\boldsymbol{\mu}) = \inf_{\boldsymbol{x}\in\mathbb{R}^n} L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}),$$
$$L(\boldsymbol{x},\,\boldsymbol{\lambda},\,\boldsymbol{\mu}) = f(\boldsymbol{x}) + \big\langle\boldsymbol{\lambda},\,\mathbf{A}\boldsymbol{x} + \boldsymbol{b}\big\rangle_{2,\,m} + \big\langle\boldsymbol{\mu},\,\boldsymbol{h}(\boldsymbol{x})\big\rangle_{2,\,p}$$

である．

定理 D.54. 関数 $h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ は凸かつ連続微分可能とする．さらに修正 Slaker 条件が成立するとする．

$$\text{ある } \overline{\boldsymbol{x}} \in \mathbb{R}^n \text{ があって } \boldsymbol{h}(\overline{\boldsymbol{x}}) \leq \mathbf{0}_p \text{ かつ } \mathbf{A}\overline{\boldsymbol{x}} = \boldsymbol{b} \tag{D.87}$$

をみたす．このとき以下が成り立つ．
(i) $\boldsymbol{x}^*$ は主問題 (D.85) の解のとき

$$\text{組 } (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \text{ は双対問題 (D.86) の解である．} \tag{D.88}$$

(ii) さらに

$$L^*(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*) \tag{D.89}$$

が成立する．したがって, 強双対性

$$\sup_{\boldsymbol{\mu}\in\mathbb{R}^p,\, \boldsymbol{\lambda}\in\mathbb{R}^m,\, \boldsymbol{\mu}\geq\mathbf{0}_p} L^*(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{g}(\boldsymbol{x})=\mathbf{0}_m,\, \boldsymbol{h}(\boldsymbol{x})\geq\mathbf{0}_p} f(\boldsymbol{x})$$

が成立する．

Proof. 定理 D.40 から $(\gamma^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \neq (0, \mathbf{0}_p, \mathbf{0}_m)$ があって F.John の条件をみたす．すなわち

$$\gamma^*\nabla f(\boldsymbol{x}^*) + \mathbf{A}^\top\boldsymbol{\lambda}^* + \nabla\boldsymbol{h}(\boldsymbol{x}^*)^\top\boldsymbol{\mu}^* = \mathbf{0}_n, \tag{D.90}$$
$$\gamma^* \geq 0,\, \boldsymbol{\mu} \geq \mathbf{0}_m,\, \boldsymbol{\lambda} \geq \mathbf{0}_p,\, \left\langle \boldsymbol{\mu}^*, \boldsymbol{h})(\boldsymbol{x}^*) \right\rangle_{2,n} = 0 \tag{D.91}$$

が成り立つ．
　まず背理法を用いて

$$\gamma^* \neq 0 \tag{D.92}$$

を示す．そのために $\gamma = 0$ と仮定すると (D.90) から

$$\mathbf{A}^\top\boldsymbol{\lambda}^* + \nabla\boldsymbol{h}(\boldsymbol{x}^*)^\top\boldsymbol{\mu}^* = \mathbf{0}_n \tag{D.93}$$

となる．さらに (D.87) をみたす $\overline{\boldsymbol{x}}$ を用いて $\boldsymbol{x}^* - \overline{\boldsymbol{x}}$ と上の式の内積をとると

$$\begin{aligned} 0 &= \left\langle \mathbf{A}^\top\boldsymbol{\lambda}^*, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} + \left\langle \nabla\boldsymbol{h}(\boldsymbol{x}^*)^\top\boldsymbol{\mu}^*, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,n} \\ &= \left\langle \boldsymbol{\lambda}^*, \mathbf{A}\overline{\boldsymbol{x}} - \mathbf{A}\boldsymbol{x}^* \right\rangle_{2,m} + \left\langle \boldsymbol{\mu}^*, \nabla\boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*) \right\rangle_{2,p} \end{aligned}$$

となる．しかし

$$\mathbf{A}\overline{\boldsymbol{x}} - \mathbf{A}\boldsymbol{x}^* = \boldsymbol{b} - \boldsymbol{b} = \mathbf{0}_m$$

なので

$$\left\langle \boldsymbol{\mu}^*,\, \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*)\right\rangle_{2,\,p} = 0 \tag{D.94}$$

を得る．背理法の仮定から $\gamma \neq 0$ なので KKT 条件は

$$\nabla \boldsymbol{h}(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^* = \mathbf{0}_n \tag{D.95}$$

となる．ただし，$\boldsymbol{\mu}^* = \gamma^{-1}\boldsymbol{\mu}$ と $\boldsymbol{\lambda}^* = \gamma^{-1}\boldsymbol{\lambda}$ とした．この等式が (D.92) を導くことを示そう．$f,\, h_j\,(j = 1,\, 2,\, \ldots,\, p)$ は凸関数であることと (D.94) を用いると

$$\begin{aligned}
f(\boldsymbol{x}^*) &= f(\boldsymbol{x}^*) + \big\langle \boldsymbol{\lambda}^*,\, \underbrace{\mathbf{A}\boldsymbol{x}^* - \boldsymbol{b}}_{=\mathbf{0}_m}\big\rangle_{2,\,m} + \underbrace{\big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*)\big\rangle_{2,\,p}}_{=0} \\
&\le \underbrace{f(\overline{\boldsymbol{x}}) + \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{x}^* - \overline{\boldsymbol{x}}\big\rangle_{2,\,n}}_{f\ \text{の凸性から}} + \big\langle \boldsymbol{\lambda}^*,\, (\mathbf{A}\boldsymbol{x}^* - \mathbf{A}\overline{\boldsymbol{x}}) + (\mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b})\big\rangle_{2,\,m} \\
&\quad + \underbrace{\big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) + \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\boldsymbol{x}^* - \overline{\boldsymbol{x}})\big\rangle_{2,\,p}}_{h_j\ \text{の凸性と}\ \boldsymbol{\mu}^* \ge \mathbf{0}_p} \\
&= f(\boldsymbol{x}) + \big\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b}\big\rangle_{2,\,m} \\
&\quad + \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}})\big\rangle_{2,\,p} + \big\langle \underbrace{\nabla f(\boldsymbol{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\mu}^*,\, \boldsymbol{x}^* - \overline{\boldsymbol{x}}}_{=\mathbf{0}_n \because (\text{D.95})}\big\rangle_{2,\,n} \\
&= f(\boldsymbol{x}) + \big\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\overline{\boldsymbol{x}} - \boldsymbol{b}\big\rangle_{2,\,p} + \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}})\big\rangle_{2,\,p}
\end{aligned}$$

となる．さらに $h_j\,(j = 1,\, 2,\, \ldots,\, p)$ の凸性から

$$h_j(\boldsymbol{x}^*) + \big\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^*\big\rangle_{2,\,n} \le h_j(\overline{\boldsymbol{x}})$$

を得る．上式の辺々に $\mu_j^*\,(j = 1,\, 2,\, \ldots,\, p)$ を掛けて，j について和をとると

$$\begin{aligned}
&\sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \sum_{j=1}^{p} \mu_j^* \big\langle \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^*\big\rangle_{2,\,n} \le \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) \\
&\Leftrightarrow\ \sum_{j=1}^{p} \mu_j^* h_j(\boldsymbol{x}^*) + \left\langle \sum_{j=1}^{p} \mu_j^* \nabla h_j(\boldsymbol{x}^*),\, \overline{\boldsymbol{x}} - \boldsymbol{x}^* \right\rangle_{2,\,n} \le \sum_{j=1}^{p} \mu_j^* h_j(\overline{\boldsymbol{x}}) \\
&\Leftrightarrow\ \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*)\big\rangle_{2,\,p} + \big\langle \big(\nabla \boldsymbol{h}(\boldsymbol{x}^*)\big)^\top \boldsymbol{\mu}^*,\, \overline{\boldsymbol{x}} - \boldsymbol{x}^*\big\rangle_{2,\,n} \le \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}})\big\rangle_{2,\,p} \\
&\Leftrightarrow\ \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*)\big\rangle_{2,\,p} + \underbrace{\big\langle \boldsymbol{\mu}^*,\, \nabla \boldsymbol{h}(\boldsymbol{x}^*)(\overline{\boldsymbol{x}} - \boldsymbol{x}^*)\big\rangle_{2,\,p}}_{=0 \quad \because (\text{D.94}))} \le \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}})\big\rangle_{2,\,p}
\end{aligned}$$

がわかる．したがって

$$\big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}^*(\boldsymbol{x}^*)\big\rangle_{2,\,p} \le \big\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}})\big\rangle_{2,\,p} \tag{D.96}$$

となる．しかし，$\boldsymbol{\mu}^* \neq \mathbf{0}_p$ のとき，修正 Slater 条件より

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\overline{\boldsymbol{x}}) \right\rangle_{2,\,p} < 0 \tag{D.97}$$

となる．(D.96) と (D.97) から

$$\left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}^*) \right\rangle_{2,\,p} < 0$$

となり，(D.91) と矛盾する．よって $\gamma^* \neq 0$ ならば $\boldsymbol{\mu} \neq \mathbf{0}_m$ とはならない．
次に $\boldsymbol{\mu} = \mathbf{0}_m$ の場合を考える．すると (D.95) から

$$\mathbf{A}\boldsymbol{\lambda}^* = \mathbf{0}_m$$

となる．$\mathbf{A}$ は非退化なので $\boldsymbol{\lambda}^* = \mathbf{0}_p$ となり，$(\lambda^*,\, \boldsymbol{\lambda},\, \boldsymbol{\mu}) = (0,\, \mathbf{0}_m,\, \mathbf{0}_p)$ となるので矛盾する．よって $\gamma^* \neq 0$ となる．したがって

$$f(\boldsymbol{x}^*) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \left\langle \boldsymbol{\lambda}^*,\, \mathbf{A}\boldsymbol{x} - \boldsymbol{b} \right\rangle_{2,\,m} + \left\langle \boldsymbol{\mu}^*,\, \boldsymbol{h}(\boldsymbol{x}) \right\rangle_{2,\,p} \right\} = L^*(\boldsymbol{\lambda}^*,\, \boldsymbol{\mu}^*)$$

となる．逆向きの不等式は定理 D.53 よりわかる．□

D.9　凸双対性（その２）

D.9.1　Fenchel 双対性

2 つの凸かつ下半連続関数[5]

$$f : \mathbb{R}^n \to (-\infty,\, \infty) \cup \{+\infty\}, \quad g : \mathbb{R}^m \to (-\infty,\, \infty) \cup \{+\infty\}$$

と $m \times n$ 行列 $\mathbf{A}$ が与えられたとする．

定義 D.55. 関数 f と g の**領域**をそれぞれ

$$\operatorname{dom} f = \{\boldsymbol{x} \in \mathbb{R}^n;\, f(\boldsymbol{x}) < \infty\}$$
$$\operatorname{dom} g = \{\boldsymbol{y} \in \mathbb{R}^m;\, g(\boldsymbol{y}) < \infty\}$$

で定義する．

[5]関数 f は点 $\boldsymbol{x}$ で下半（かはん）連続であるとは，任意の点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty} \subset \mathbb{R}^n$ で $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$ なるものに対して

$$\liminf_{k\to\infty} f(\boldsymbol{x}_k) = f(\boldsymbol{x})$$

をみたすときをいう．

仮定 D.56. 関数 f と g は $-\infty$ の値は取らず

$$\mathrm{dom}\ f \neq \varnothing, \quad \mathrm{dom}\ g \neq \varnothing$$

とする．このことが成り立つとき, これらの関数は proper であるという．

関数 f, g は凸なので, Young の不等式 (補題 D.32) を利用すると, 任意の $\boldsymbol{x} \in \mathbb{R}^n$ と $\boldsymbol{y} \in \mathbb{R}^m$ に対して

$$\begin{aligned} f(\boldsymbol{x}) + f^*(\mathbf{A}^\top \boldsymbol{y}) &\geq \big\langle \boldsymbol{x},\, \mathbf{A}^\top \boldsymbol{y} \big\rangle_{2,\,n}, \\ g\big(\mathbf{A}\boldsymbol{x}\big) + g^*(-\boldsymbol{y}) &\geq -\big\langle \mathbf{A}\boldsymbol{x},\, \boldsymbol{y} \big\rangle_{2,\,m} \end{aligned}$$

が成立するとする．

これらの不等式の辺々を加えると

$$f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x}\big) \geq -f\big(\mathbf{A}^\top \boldsymbol{y}\big) - g^*(-\boldsymbol{y}) \quad (\boldsymbol{x} \in \mathbb{R}^n,\, \boldsymbol{y} \in \mathbb{R}^m) \tag{D.98}$$

となる．このことを踏まえて以下の主問題と双対問題を導入することにする．

$$\text{minimize}\ \Big\{ f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x}\big) \Big\} \quad (\boldsymbol{x} \in \mathbb{R}^n) \tag{D.99}$$

と双対問題

$$\text{maximize}\ \Big\{ -f^*\big(\mathbf{A}^\top \boldsymbol{y}\big) - g^*(-\boldsymbol{y}) \Big\} \quad (\boldsymbol{y} \in \mathbb{R}^m). \tag{D.100}$$

このとき (D.98) は弱双対性を保証する．

以上の議論から以下の定理を得る．

定理 D.57. 不等式

$$\sup_{\boldsymbol{y} \in \mathbb{R}^m} \Big\{ -f^*\big(\mathbf{A}^\top \boldsymbol{y}\big) - g^*(-\boldsymbol{y}) \Big\} \leq \inf_{\boldsymbol{x} \in \mathbb{R}^n} \Big\{ f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x}\big) \Big\}$$

が成立する．

定義 D.58. 関数 $v : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ を

$$v(\boldsymbol{a}) := \inf_{\boldsymbol{x} \in \mathbb{R}^n} \big\{ f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) \big\} \quad (\boldsymbol{a} \in \mathbb{R}^m)$$

で定める．

補題 D.59. 関数 $v : \mathbb{R}^m \to (-\infty,\, \infty) \cup \{+\infty\}$ は凸かつ下半連続である．

Proof. 凸性の証明: 任意の $\boldsymbol{a}, \widetilde{\boldsymbol{a}} \in \mathbb{R}^m$ と $0 \le \lambda \le 1$ に対して, 関数 g の凸性から

$$g\big(\mathbf{A}\boldsymbol{x} + (\lambda \boldsymbol{a} + (1-\lambda)\widetilde{\boldsymbol{a}})\big) \le \lambda g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) + (1-\lambda) g\big(\mathbf{A}\boldsymbol{x} + \widetilde{\boldsymbol{a}}\big)$$

となる. 上の不等式と inf の性質から

$$\begin{aligned} v\big(\lambda \boldsymbol{a} + (1-\lambda)\widetilde{\boldsymbol{a}}\big) &\le \inf_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \mathbf{A}\boldsymbol{x} + \lambda g\big(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\big) + (1-\lambda) g\big(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\big) \Big\} \\ &= \inf_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \lambda\big(\mathbf{A}\boldsymbol{x} g\big(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\big)\big) + (1-\lambda)\big(\mathbf{A}\boldsymbol{x} + g\big(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\big)\big) \Big\} \\ &\le \lambda \inf_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \mathbf{A}\boldsymbol{x} + g\big(\mathbf{A}\boldsymbol{x}+\boldsymbol{a}\big) \Big\} \\ &\quad + (1-\lambda) \inf_{\boldsymbol{x}\in\mathbb{R}^n} \Big\{ \mathbf{A}\boldsymbol{x} + g\big(\mathbf{A}\boldsymbol{x}+\widetilde{\boldsymbol{a}}\big) \Big\} \\ &= \lambda v(\boldsymbol{a}) + (1-\lambda) v(\widetilde{\boldsymbol{a}}) \end{aligned}$$

を得る. よって, 関数 v は凸であることがわかる.
下半連続性の証明: 任意の $\boldsymbol{a} \in \mathbb{R}^m$ に対して, 任意の数列 $\{\boldsymbol{a}_k\}_{k=1}^{\infty} \subset \mathbb{R}^m$ で $\lim_{k\to\infty} \boldsymbol{a}_k = \boldsymbol{a}$ となるものを取る. このとき, 関数 g の下半連続性から

$$\begin{aligned} \liminf_{k\to\infty} v(\boldsymbol{a}_k) &= \liminf_{k\to\infty} \Big\{ f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}_k\big) \Big\} \\ &= f(\boldsymbol{x}) + \liminf_{k\to\infty} g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}_k\big) \\ &= f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) \quad (\because g \text{ は下半連続}) \end{aligned}$$

を得る. よって 関数 v も下半連続である. □

定義 D.60. 関数 f, g と行列 $\mathbf{A}$ の組が**双対条件**をみたすとは

$$\text{値関数 } v \text{ は原点の近くで } +\infty \text{ の値を取らない} \tag{D.101}$$

をみたすときをいう.

注意 D.61. 条件 (D.101) は次のように書きかえることができる. ある小さな正数 $\delta > 0$ があって $\boldsymbol{a} \in \mathbb{R}^m$ が $|\boldsymbol{a}|_{2,m} \le \delta$ をみたすとき, ある $\boldsymbol{y} \in \mathrm{dom}\ g$, $\boldsymbol{x} \in \mathrm{dom}\ f$ があって

$$\boldsymbol{a} = \boldsymbol{y} - \mathbf{A}\boldsymbol{x} \tag{D.102}$$

とかけることである. したがって

$$v(\boldsymbol{a}) \le f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) = f(\boldsymbol{x}) + g(\boldsymbol{y}) < \infty$$

となる. □

定理 D.62. $\boldsymbol{x}^* \in \mathbb{R}^n$ が主問題 (D.99) の解とし, 関数 $f,\, g$ と行列 $\mathbf{A}$ は双対条件 (D.101) をみたすとする. このとき以下が成立する.
(1) $\boldsymbol{y}^* \in \mathbb{R}^m$ が存在して

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) = -f^*\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)$$

をみたす.
(2) 強双対性

$$\inf_{\boldsymbol{x}\in\mathbb{R}^n}\Big\{f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x}\big)\Big\} = \sup_{\boldsymbol{y}\in\mathbb{R}^m}\Big\{-f^*\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)\Big\} \tag{D.103}$$

が成立する.

Proof. 注意 D.61 より, 双対条件 (D.101) は原点の近くで有限の値を取ることがわかる. よって, 関数 $v : \mathbb{R}^m \to \mathbb{R}$ は凸であることを合わせて考えると $\partial v(\mathbf{0}) \neq \varnothing$ となることがわかる. そこで

$$-\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)$$

を取る. このとき $\boldsymbol{x} \in \mathbb{R}^n$ と $\boldsymbol{a} \in \mathbb{R}^m$ に対して

$$\begin{aligned}
f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) &= v(\mathbf{0}_m) \le v(\boldsymbol{a}) + \langle \boldsymbol{a},\, \boldsymbol{y}^*\rangle_{2,\,m} \quad (\because -\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)) \\
&\le f(\boldsymbol{x}) + g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big) + \langle \boldsymbol{a},\, \boldsymbol{y}^*\rangle_{2,\,m} \\
&= -\Big(\big\langle \mathbf{A}^\top \boldsymbol{y}^*,\, \boldsymbol{x}\big\rangle_{2,\,n} - f(\boldsymbol{x})\Big) \\
&\qquad - \Big(-\big\langle \boldsymbol{y}^*,\, \mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big\rangle_{2,\,n} - g\big(\mathbf{A}\boldsymbol{x} + \boldsymbol{a}\big)\Big)
\end{aligned}$$

となる. $\boldsymbol{a}$ に関して infimum を取ると

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}\big) \le -\Big(\big\langle \mathbf{A}^\top \boldsymbol{y}^*,\, \boldsymbol{x}\big\rangle_{2,\,n} - f(\boldsymbol{x})\Big) - g^*(-\boldsymbol{y}^*)$$

となる. さらに $\boldsymbol{x}$ に関して infimum を取ると

$$f(\boldsymbol{x}^*) + g\big(\mathbf{A}\boldsymbol{x}^*\big) \le -f\big(\mathbf{A}^\top \boldsymbol{y}^*\big) - g^*(-\boldsymbol{y}^*)$$

を得る.
逆向きの不等式は定理 D.57 からわかる. □

D.10 半正定値計画法

記号 (1) $\mathrm{Sym}(n;\,\mathbb{R})$ を $n \times n$ 対称行列全体の成す線型部分空間とし

$$\mathrm{Sym}^{+}(n;\,\mathbb{R}) = \{\mathbf{A} \in \mathrm{Sym}(n;\,\mathbb{R});\ \mathbf{A} \succeq \mathbf{0}_{n\times n}\}$$

とする. ただし

$$\mathbf{A} \succeq \mathbf{0}_{n\times n} \ \Leftrightarrow\ \boldsymbol{x}^{\top}\mathbf{A}\boldsymbol{x} \geq 0\,(\forall \boldsymbol{x} \in \mathbb{R}^{n})$$

である.
(2) 行列 $\mathbf{A},\, \mathbf{B} \in \mathrm{Sym}(n;\,\mathbb{R})$ の内積を

$$\big\langle \mathbf{A},\, \mathbf{B} \big\rangle_{\mathrm{F}} = \sum_{i,\,j=1}^{n} a_{ij} b_{ij} \qquad (\mathbf{A} = (a_{ij}),\, \mathbf{B} = (b_{ij}))$$

で定める.
　与えられた $\mathbf{A}_k,\, \mathbf{C} \in \mathrm{Sym}(n;\,\mathbb{R})\,(k = 1,\, 2,\, \ldots,\, m)$ と $b_k \in \mathbb{R}\,(k = 1,\, 2,\, \ldots,\, k)$ に対して半正定値計画問題

$$\begin{aligned} &\text{minimize } \big\langle \mathbf{C},\, \mathbf{X} \big\rangle_{\mathrm{F}} \\ &\text{subject to } \big\langle \mathbf{A}_1,\, \mathbf{X} \big\rangle_{\mathrm{F}} = b_1,\, \big\langle \mathbf{A}_2,\, \mathbf{X} \big\rangle_{\mathrm{F}} = b_2,\, \ldots,\, \big\langle \mathbf{A}_m,\, \mathbf{X} \big\rangle_{\mathrm{F}} = b_m \\ &\qquad \mathbf{X} \succeq \mathbf{0}_{n\times n} \end{aligned} \tag{D.104}$$

を考える.
双対性. まず写像 $\boldsymbol{\alpha} : \mathrm{Sym}(n;\,\mathbb{R}) \to \mathbb{R}^m$ を

$$\boldsymbol{\alpha}(\mathbf{X}) = \big[\big\langle \mathbf{A}_1,\, \mathbf{X} \big\rangle_{\mathrm{F}},\, \big\langle \mathbf{A}_2,\, \mathbf{X} \big\rangle_{\mathrm{F}},\, \ldots,\, \big\langle \mathbf{A}_m,\, \mathbf{X} \big\rangle_{\mathrm{F}}\big]^{\top}$$

と定める. さらに写像 $\boldsymbol{\alpha}^{\top} : \mathbb{R}^m \to \mathrm{Sym}(n;\,\mathbb{R})$ を

$$\big\langle \boldsymbol{\alpha}^{\top}(\boldsymbol{y}),\, \mathbf{X} \big\rangle_{\mathrm{F}} = \big\langle \boldsymbol{y},\, \boldsymbol{\alpha}(\mathbf{X}) \big\rangle_{2,\,m} = \sum_{k=1}^{m} y_k \big\langle \mathbf{A}_k,\, \mathbf{X} \big\rangle_{\mathbf{F}} \qquad (\forall \mathbf{X} \in \mathrm{Sym}(n;\,\mathbb{R}^n))$$

で定める. 主問題 (D.104) を

$$\text{minimize } f(\mathbf{X}) + g\big(\boldsymbol{\alpha}(\mathbf{X})\big) \tag{D.105}$$

と書きかえる. ただし, $\boldsymbol{b} = (b_1,\, b_2,\, \ldots,\, b_m)^{\top}$ で

$$f(\mathbf{X}) = \begin{cases} \big\langle \mathbf{C},\, \mathbf{X} \big\rangle_{\mathrm{F}} & (\mathbf{X} \succeq \mathbf{0}) \\ +\infty & (\text{その他の場合}) \end{cases}; \tag{D.106}$$

$$g(\boldsymbol{y}) = \begin{cases} 0 & (\boldsymbol{y} = \boldsymbol{b}) \\ +\infty & (\text{その他の場合}) \end{cases} \tag{D.107}$$

である．このとき $\boldsymbol{z} \in \mathbb{R}^m$ に対し

$$g^*(\boldsymbol{z}) = \sup_{\boldsymbol{y}\in\mathbb{R}^m}\left\{\langle \boldsymbol{z},\, \boldsymbol{y}\rangle_{2,\,m} - g(\boldsymbol{y})\right\} = \langle \boldsymbol{z},\, \boldsymbol{b}\rangle_{2,\,m} \tag{D.108}$$

$$(\because\ \boldsymbol{y} \neq \boldsymbol{b} \text{ のとき } g(\boldsymbol{y}) = +\infty\ )$$

となる．さらに $\mathbf{W} \in \mathrm{Sym}(n;\,\mathbb{R})$ に対して

$$f^*(\mathbf{W}) = \sup_{\mathbf{X}\in\mathrm{Sym}(n;\,\mathbb{R})}\left\{\left\langle \mathbf{W},\, \mathbf{X}\right\rangle_{\mathrm{F}} - f(\mathbf{W})\right\} \tag{D.109}$$

$$= \sup_{\mathbf{X}\in\mathrm{Sym}(n;\,\mathbb{R})}\left\{\left\langle \mathbf{W}-\mathbf{C},\, \mathbf{X}\right\rangle_{\mathrm{F}}\right\} \tag{D.110}$$

$$= \begin{cases} 0 & (\mathbf{C} \preceq \mathbf{W}) \\ +\infty & (\text{その他の場合}) \end{cases} \tag{D.111}$$

$$(\because \mathbf{X} \succeq \mathbf{0}_n \text{ でなければ } f(\mathbf{X}) = +\infty)$$

となる．ただし

$$\mathbf{C} \preceq \mathbf{W} \;\Leftrightarrow\; \mathbf{W} - \mathbf{C} \succeq \mathbf{0}$$

である．さらに, 任意の $\mathbf{X} \in \mathrm{Sym}(n;\,\mathbb{R})$ と $\boldsymbol{y} = (y_1,\, y_2,\, \ldots,\, y_m)^\top \in \mathbb{R}^m$ に対し

$$\left\langle \boldsymbol{\alpha}^\top \boldsymbol{y},\, \mathbf{X}\right\rangle_{\mathrm{F}} = \left\langle \boldsymbol{y},\, \boldsymbol{\alpha}\mathbf{X}\right\rangle_{2,\,n} = \left\langle \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix},\, \begin{bmatrix} \left\langle \mathbf{A}_1,\, \mathbf{X}\right\rangle_{\mathrm{F}} \\ \left\langle \mathbf{A}_2,\, \mathbf{X}\right\rangle_{\mathrm{F}} \\ \vdots \\ \left\langle \mathbf{A}_m,\, \mathbf{X}\right\rangle_{\mathrm{F}} \end{bmatrix} \right\rangle_{2,\,m}$$

$$= \sum_{k=1}^{m} y_k \left\langle \mathbf{A}_k,\, \mathbf{X}\right\rangle_{\mathrm{F}} = \left\langle \sum_{k=1}^{m} y_k \mathbf{A}_k,\, \mathbf{X} \right\rangle_{\mathrm{F}}$$

から

$$\boldsymbol{\alpha}^\top \boldsymbol{y} = \sum_{k=1}^{m} y_k \mathbf{A}_k$$

がわかる．

Fenchel **双対問題**.

$$f^*(\boldsymbol{\alpha}^\top \boldsymbol{y}) - g^*(-\boldsymbol{y})$$

を最大化することは半正定値双対問題

$$\text{minimize } \left\langle \boldsymbol{y},\, \boldsymbol{b}\right\rangle_{2,\,m} \qquad \text{subject to } \sum_{k=1}^{m} y_k \mathbf{A}_k \preceq \mathbf{C} (\Leftrightarrow\ f^*(\boldsymbol{\alpha}^\top \boldsymbol{y}) = 0) \tag{D.112}$$

となる．これは線型計画法の双対問題のアプローチとなる．ただし制約は対称行列に対する不等式制約である．さらに主問題 (D.104) は対称行列 $\mathbf{X}$ が未知であり, 双対問題 (D.105) はベクトル $\boldsymbol{y} \in \mathbb{R}^m$ が未知である．

定理 D.63. 主問題 (D.104) の実行可能解 $\overline{\mathbf{X}}$ で

$$\overline{\mathbf{X}} \succeq \mathbf{0}_{n\times n} \tag{D.113}$$

をみたすものが存在したとする．さらに写像

$$\boldsymbol{\alpha} : \mathrm{Sym}(n;\, \mathbb{R}) \,\to\, \mathbb{R}^m$$

は全射とする．このとき主問題 (D.104) の解を $\mathbf{X}^*$ としたとき双対問題 (D.105) の実行解 $\boldsymbol{y}^*$ が存在し

$$\big\langle \mathbf{C},\, \mathbf{X}^* \big\rangle_{\mathrm{F}} = \big\langle \boldsymbol{b},\, \boldsymbol{y}^* \big\rangle_{2,\, m}$$

をみたす．

Proof. (D.112) を考慮すると，十分小さな正数 $\lambda > 0$ が存在して

$$\big\langle \mathbf{Y},\, \mathbf{Y} \big\rangle_{\mathrm{F}} \le \lambda^2 \;\Rightarrow\; \mathbf{X} = \overline{\mathbf{X}} + \mathbf{Y} \in \mathrm{Sym}^+(n;\, \mathbb{R})$$

とできる．写像

$$\boldsymbol{\alpha} : \mathrm{Sym}(n;\, \mathbb{R}) \,\to\, \mathbb{R}^m$$

は全射なので，ある正数 $\eta > 0$ が存在して

$$\boldsymbol{\alpha}(\mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\, 1)) \supset \mathbb{B}_{\mathbb{R}^m}(\mathbf{0}_m,\, \lambda) = \{\boldsymbol{y} \in \mathbb{R}^m;\, |\boldsymbol{y}|_{2,\, m} \le \lambda\}$$

とできる．ただし $\mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\, 1) = \{\mathbf{A} \in \mathrm{Sym}(n;\, \mathbb{R});\, \big\langle \mathbf{A},\, \mathbf{A} \big\rangle_{\mathrm{F}} \le 1\}$ である．よって，各 $\boldsymbol{a} \in \mathbb{B}_{\mathbb{R}^m}(\mathbf{0}_m,\, \lambda\eta)$ に対して $\mathbf{X} = \overline{\mathbf{X}} + \mathbf{Y} \in \mathrm{Sym}^+(n;\, \mathbb{R})$ かつ $\mathbf{Y} \in \mathbb{B}_{\mathrm{Sym}}(\mathbf{0}_m,\, \lambda)$ が存在して

$$\boldsymbol{\alpha}(\mathbf{X}) = \boldsymbol{\alpha}(\overline{\mathbf{X}}) + \boldsymbol{\alpha}(\mathbf{Y}) = \boldsymbol{b} - \boldsymbol{a}$$

となる．

関数 f, g の定義 (D.106) と (D.107) から

$$\mathrm{dom}\, f = \mathrm{Sym}^+(n;\, \mathbb{R}), \quad \mathrm{dom}\, g = \{\boldsymbol{b}\}$$

である．したがって $\lambda = \delta\eta$ かつ $|\boldsymbol{a}|_{2,\, m} \le \delta$ のとき (D.102) から

$$\boldsymbol{a} = \boldsymbol{b} - \boldsymbol{\alpha}(\mathbf{X})$$

と書ける．結局双対条件 (D.101) をみたす．すると値関数

$$v(\boldsymbol{a}) = \inf_{\mathbf{X} \in \mathrm{Sym}(n;\, \mathbb{R})} \big\{ f(\mathbf{X}) + g\big(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\big) \big\}$$

は原点近くで有限の値を取るので，$\partial v(\mathbf{0}_m) \neq \varnothing$ となる．そこで

$$-\boldsymbol{y}^* \in \partial v(\mathbf{0}_m)$$

を取る. このとき, 任意の $\mathbf{X} \in \mathrm{Sym}(n;\mathbb{R})$ と $\boldsymbol{a} \in \mathbb{R}^m$ に対して

$$\begin{aligned} f(\mathbf{X}^*) + g\big(\mathbf{A}\mathbf{X}\big) &= v(\mathbf{0}) \le v(\boldsymbol{a}) + \langle \boldsymbol{a},\, \boldsymbol{y}^* \rangle_{2,m} \\ &\le f(\mathbf{X}) + g\big(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\big) + \langle \boldsymbol{a},\, \boldsymbol{y}^* \rangle_{2,m} \\ &= -\Big(\big\langle \boldsymbol{\alpha}^\top \boldsymbol{y},\, \mathbf{X} \big\rangle_{\mathrm{F}} - f(\mathbf{X}) \Big) \\ &\quad\ - \Big(-\big\langle \boldsymbol{y},\, \boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a} \big\rangle_{2,m} - g\big(\boldsymbol{\alpha}(\mathbf{X}) + \boldsymbol{a}\big) \Big) \end{aligned}$$

を得る. 上の式の最右辺を $\boldsymbol{a}$ に関して infimum を取ると

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) \le -\Big(\big\langle \boldsymbol{\alpha}^\top(\boldsymbol{y}),\, \mathbf{X} \big\rangle_{\mathrm{F}} - f(\mathbf{X}) \Big) - g^*(-\boldsymbol{y}^*)$$

を得る. さらに上の式の右辺を $\mathbf{X}$ に関して infimum を取ると

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) \le -f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) - g^*(-\boldsymbol{y}^*)$$

を得る. 逆向きの不等式は常に成立するので,

$$f(\mathbf{X}^*) + g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) = -f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) - g^*(-\boldsymbol{y}^*) \tag{D.114}$$

を得る. (**??**) – (D.111) から

$$\begin{aligned} f(\mathbf{X}^*) &= \big\langle \mathbf{C},\, \mathbf{X}^* \big\rangle_{\mathrm{F}}, \\ g\big(\boldsymbol{\alpha}(\mathbf{X}^*)\big) &= 0 (\Leftrightarrow\ \boldsymbol{\alpha}(\mathbf{X}^*) = \boldsymbol{b}), \\ f^*\big(\boldsymbol{\alpha}^\top(\boldsymbol{y}^*)\big) &= 0\,(\Leftrightarrow\ \boldsymbol{\alpha}^\top(\boldsymbol{y}) \preceq \mathbf{C}), \\ g^*(-\boldsymbol{y}^*) &= -\langle \boldsymbol{y}^*,\, \boldsymbol{b} \rangle_{2,m} \end{aligned}$$

となるので, これらを (D.114) に代入すると

$$\big\langle \mathbf{C},\, \mathbf{X}^* \big\rangle_{\mathrm{F}} = \langle \boldsymbol{y}^*,\, \boldsymbol{b} \rangle_{2,m}$$

がわかる. □

D.11 最適化アルゴリズム

この節では, 3 つの基本的なアルゴリズムである最急降下法 (steepest-descent method), Newton 法, 共役勾配法 (conjugate-gradient method) について説明をする. 最初の 2 つのアルゴリズムに基づいて多くの最適化がなされている. 最急降下法は収束のスピードは遅く実用的ではないが, 簡単で信頼性が高い. Newton 法は非線型方程式の解をみつける一般的な方法である. しかし解の近傍に初期値がないと停留点に収束しない場合がある. 共役勾配法は凸 2 次計画法における最適化に用いると効果を発揮する.

D.11.1 多変数関数の微分

定義 D.64. $U \subset \mathbb{R}^d$ を空でない開部分集合とし, $f : U \to \mathbb{R}$ を関数とする. $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in U$ に対し, 極限

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) := \lim_{t \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_d) - f(\boldsymbol{x})}{t} \qquad (i = 1, 2, \ldots, \mathrm{d})$$

が存在するとき, $\dfrac{\partial f}{\partial x_i}(\boldsymbol{x})$ 点 $\boldsymbol{x}$ における x_i に関する f の**偏微分係数**という. すべての変数に関して偏微分係数が存在するとき, ベクトル

$$\nabla f(\boldsymbol{x}) := \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_d}(\boldsymbol{x}) \right)^\top$$

は f の**勾配** (gradient) という.

注意 D.65. $\boldsymbol{b} = (b_1, b_2, \ldots, b_d)^\top \in \mathbb{R}^d$ とし

$$\boldsymbol{e}_i := (\underbrace{0, 0, \ldots, 0}_{i-1}, 1, 0, \ldots, 0)^\top \qquad (i = 1, 2, \ldots, \mathrm{d})$$

とする. すると

$$\boldsymbol{b} = b_1 \boldsymbol{e}_1 + b_2 \boldsymbol{e}_2 + \cdots + b_d \boldsymbol{e}_d$$

となる. □

定義 D.66. 方向 $\boldsymbol{d} \in \mathbb{R}^d$ に沿った点 $\boldsymbol{x} \in U$ における f の**方向微分**を

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t \searrow 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$$

で定める. ただし $t > 0$ が 0 に近づくとき上の式の右辺の極限が存在するときに方向微分を定める.

明らかに $\alpha \geq 0$ に対して

$$\overset{\bullet}{f}(\boldsymbol{x}; \alpha\boldsymbol{d}) = \alpha \overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d})$$

となる. したがって $\overset{\bullet}{f}(\boldsymbol{x}; -\boldsymbol{d}) = -\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d})$ であるので

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = \lim_{t \to 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$$

となる. 実際

$$\overset{\bullet}{f}(\boldsymbol{x}; \boldsymbol{d}) = -\overset{\bullet}{f}(\boldsymbol{x}; -\boldsymbol{d}) = -\lim_{t \searrow 0} \frac{f(\boldsymbol{x} - t\boldsymbol{d}) - f(\boldsymbol{x})}{t} = \lim_{s \nearrow 0} \frac{f(\boldsymbol{x} + s\boldsymbol{d}) - \boldsymbol{x})}{s}$$

よりわかる.

定義 D.67. 関数 $f\colon U \to \mathbb{R}$ は点 $\boldsymbol{x} \in U$ で Gâteaux 微分可能であるとは, 任意の方向 $\boldsymbol{d} \in \mathbb{R}^d$ に対して, 方向微分 $\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d})$ は存在し, それが $\boldsymbol{d}$ の線型関数であるときをいう. f が U の任意の点 $\boldsymbol{x}$ で Gâtaeaux 微分可能のとき, f は U 上で Gâtaeux **微分可能**という.

注意 D.68. 関数 f が点 $\boldsymbol{x}$ において Gâteaux 微分可能とする. このとき, $\boldsymbol{d} = d_1\boldsymbol{e}_1 + d_2\boldsymbol{e}_2 + \cdots + d_d\boldsymbol{e}_d$ とおくと方向微分 $\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d})$ は $\boldsymbol{d}$ の線型関数なので

$$
\begin{aligned}
\overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{d}) &= \overset{\bullet}{f}(\boldsymbol{x};\, d_1\boldsymbol{e}_1 + d_2\boldsymbol{e}_2 + \cdots + d_d\boldsymbol{e}_d) \\
&= d_1 \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_1) + d_2 \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_2) + \cdots + d_d \overset{\bullet}{f}(\boldsymbol{x};\, \boldsymbol{e}_d) \\
&= \langle \boldsymbol{d},\, \nabla f(\boldsymbol{x}) \rangle_{2,\,\mathrm{d}}
\end{aligned}
$$

となる. ただし, $\langle \cdot\,, \cdot \rangle_{2,\,\mathrm{d}}$ は $\mathbb{R}^d$ 上の Euclid 内積である. □

定義 D.69. 関数 $f : U \to \mathbb{R}$ は点 $\boldsymbol{x}$ で Fréchet 微分可能であるとは, ある線型関数 $\ell : \mathbb{R}^d \to \mathbb{R}$ で $\ell(\boldsymbol{x}) = \langle \boldsymbol{\ell},\, \boldsymbol{x} \rangle_{2,\,\mathrm{d}}\, (\boldsymbol{\ell} \in \mathbb{R}^d)$ なるものが存在して

$$
\lim_{|\boldsymbol{h}|_{2,\,\mathrm{d}} \to 0} \frac{f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) - \langle \boldsymbol{\ell},\, \boldsymbol{h} \rangle_{2,\,\mathrm{d}}}{|\boldsymbol{h}|_{2,\,\mathrm{d}}} = 0 \tag{D.115}
$$

が成り立つときをいう. ただし, $|\cdot|_{2,\,\mathrm{d}} = \sqrt{\langle \cdot\,, \cdot \rangle_{2,\,\mathrm{d}}}$ である.

注意 D.70. ベクトル $o(\boldsymbol{h}) \in \mathbb{R}^d$ は

$$
\lim_{|\boldsymbol{h}|_{2,\,\mathrm{d}} \to 0} \frac{o(\boldsymbol{h})}{|\boldsymbol{h}|_{2,\,\mathrm{d}}} = 0
$$

をみたすものとする. この記号を用いると f が点 $\boldsymbol{x}$ で Fréchet 微分可能であることは

$$
f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \langle \boldsymbol{\ell},\, \boldsymbol{h} \rangle_{2,\,\mathrm{d}} + o(\boldsymbol{h}) \qquad (\boldsymbol{h} \to \boldsymbol{0}_d) \tag{D.116}
$$

と書けることがわかる.

$\boldsymbol{h} = t\boldsymbol{e}_i\, (i = 1, 2, \ldots, \mathrm{d})$ と $t \neq 0$ と取る. $\boldsymbol{\ell} = (\ell_1, \ell_2, \ldots, \ell_d)^\top$ としたとき, (D.115) は

$$
\lim_{t \to 0} \frac{f(\boldsymbol{x}+t\boldsymbol{e}_i) - f(\boldsymbol{x}) - t\ell_i}{t} = 0
$$

と書きなおすことができる. このとき (D.116) は

$$
f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}),\, \boldsymbol{h} \rangle_{2,\,\mathrm{d}} + o(\boldsymbol{h}) \qquad (\boldsymbol{h} \to \boldsymbol{0}_d)
$$

となる. □

定理 D.71. $U \subset \mathbb{R}^d$ は空でない開部分集合とし, 関数 $f : U \to \mathbb{R}$ は点 $\boldsymbol{x} \in U$ で Féchet 微分可能とする. このとき, 関数 f は点 $\boldsymbol{x}$ で Gâteaux 微分可能である.

Proof. Fréchet 微分と Gâteaux 微分の定義から明らか. □

定義 D.72. $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ を点とする. $[\boldsymbol{x}, \boldsymbol{y}]$ は $\boldsymbol{x}$ と $\boldsymbol{y}$ の間の線分とする. すなわち

$$[\boldsymbol{x}, \boldsymbol{y}] = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in \mathbb{R}^d; 0 \le t \le 1\}$$

である. 同様に, 両端点を含まない $\boldsymbol{x}$ と $\boldsymbol{y}$ の間の線分を

$$(\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in \mathbb{R}^d; 0 < t < 1\}$$

とする. さらに, $\boldsymbol{x} = \boldsymbol{y}$ のとき

$$[\boldsymbol{x}, \boldsymbol{y}] = (\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{x}\}$$

と定める.

補題 D.73. $U \subset \mathbb{R}^d$ を空でない開部分集合とし, 関数 $f : U \to \mathbb{R}$ は U 上で Gâteaux 微分可能とする. $\boldsymbol{x}, \boldsymbol{y} \in U$ は異なる点で $[\boldsymbol{x}, \boldsymbol{y}] \subset U$ とする. このとき, 点 $\boldsymbol{z} \in (\boldsymbol{x}, \boldsymbol{y})$ が存在して

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

とすることができる.

Proof. $\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}) \in U$ なる任意の $t \in \mathbb{R}$ に対し

$$h(t) := f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$$

とおく. f は Gâteaux 微分可能なので, $h(t)$ は微分可能で

$$\dot{h}(t) = \langle \nabla f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

となる. 1 変数実数値関数の中間値の定理から, ある $0 < \bar{t} < 1$ が存在して

$$h(1) - h(0) = \dot{h}(\bar{t}) = f(\boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})) = \langle \nabla f(\boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

と書ける. ここで, $\boldsymbol{z} := \boldsymbol{x} + \bar{t}(\boldsymbol{y} - \boldsymbol{x})$ とおく. $h(0) = f(\boldsymbol{x})$ かつ $h(1) = f(\boldsymbol{y})$ となるので

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) = \langle \nabla f(\boldsymbol{z}), \boldsymbol{y} - \boldsymbol{x} \rangle_{2,\mathrm{d}}$$

となる. □

定理 D.74. $U \subset \mathbb{R}^d$ を空でない開部分集合とし, $f : U \to \mathbb{R}$ を関数とする. 関数 f は点 $\boldsymbol{x}_0 \in U$ で Gâteaux 微分可能で, $\partial f / \partial x_i \, (i = 1, 2, \ldots, \mathrm{d})$ は点 $\boldsymbol{x}_0$ で連続微分可能ならば, f は点 $\boldsymbol{x}_0$ で Fréchet 微分可能である.

Proof. 補題 D.73 より任意の $\boldsymbol{h} \, (\neq \boldsymbol{0}_d) \in \mathbb{R}^d$ に対して，ある $\overline{\boldsymbol{x}} \in (\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{h})$ が存在し

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) - \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}} = \langle \nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}}$$

と書ける．さらに, f の勾配 ∇f は点 $\boldsymbol{x}_0$ で連続であるので

$$|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} \to 0 \quad (|\boldsymbol{h}|_{2,\mathrm{d}} \to 0)$$

となる．Cauchy-Schwarz の不等式とこのことから

$$\begin{aligned} \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}}}{|\boldsymbol{h}|_{2,\mathrm{d}}} &\leq \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} \frac{|\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} |\boldsymbol{h}|_{2,\mathrm{d}}}{|\boldsymbol{h}|_{2,\mathrm{d}}} \\ &= \lim_{|\boldsymbol{h}|_{2,\mathrm{d}} \to 0} |\nabla f(\overline{\boldsymbol{x}}) - \nabla f(\boldsymbol{x}_0)|_{2,\mathrm{d}} = 0 \end{aligned}$$

を得る．したがって

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) - \langle \nabla f(\boldsymbol{x}_0), \boldsymbol{h} \rangle_{2,\mathrm{d}} = o(\boldsymbol{h})$$

となり, f は点 $\boldsymbol{x}_0$ で Fréchet 微分可能であることが示せた. □

D.12　勾配降下法

D.12.1　降下方向

定義 D.75. $\boldsymbol{x} \in M$ は $\nabla f(\boldsymbol{x}) \neq \boldsymbol{0}_n$ とする．したがって点 $\boldsymbol{x}$ は f の**停留点** (critical point) ではない．f の**降下方向** (descent direction) は $\boldsymbol{d} \in \mathbb{R}^d \, (\neq \boldsymbol{0})$ で, ある $\bar{t} > 0$ が存在して

$$f(\boldsymbol{x} + t\boldsymbol{d}) < f(\boldsymbol{x}) \qquad (0 < \forall t < \bar{t})$$

をみたすときをいう．したがって十分小さなステップ幅 $t > 0$ に対して半直線 $\mathbb{R}_{++} := \{\boldsymbol{x} + t\boldsymbol{d};\, t > 0\}$ 上で f は厳密に減少している.

補題 D.76. $\boldsymbol{x} \in U$ は f の停留点ではないとし, $\boldsymbol{d} \in \mathbb{R}^d$ は $\boldsymbol{d} \neq \boldsymbol{0}_n$ とする．$\left\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \right\rangle_{2,\mathrm{d}} < 0$ のとき $\boldsymbol{d}$ は点 $\boldsymbol{x}$ における f の降下方向となる.

逆に $\boldsymbol{d}$ が点 $\boldsymbol{x}$ における降下方向としたとき $\left\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \right\rangle_{2,\mathrm{d}} \leq 0$ となる.

Proof. f は Gâteaux 微分可能なので

$$f(\boldsymbol{x}+t\boldsymbol{d}) = f(\boldsymbol{x}) + t\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}} + o(t) \tag{D.117}$$

となる. $\boldsymbol{d}$ は

$$\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}} < 9$$

をみたすので十分小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) < f(\boldsymbol{x})$$

となる.

逆に小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) < f(\boldsymbol{x})$$

のとき, $t \downarrow 0$ とすると

$$\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}} \leq 0$$

となる. □

D.12.2 最急降下方向

(D.117) から十分小さな $t>0$ に対して

$$f(\boldsymbol{x}+t\boldsymbol{d}) \approx f(\boldsymbol{x}) + t\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}}$$

と書ける. $t>0$ を固定する. $\boldsymbol{d} \in \mathbb{R}^d$ で $|\boldsymbol{d}|_{2,\mathrm{d}} = 1$ に対して $\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}}$ を最小にするものが f を最小にする方向である.

$\nabla f(\boldsymbol{x})$ と $\boldsymbol{d}$ の角度を θ とする. このとき

$$\big\langle \nabla f(\boldsymbol{x}),\, \boldsymbol{d} \big\rangle_{2,\mathrm{d}} = \big|\nabla f(\boldsymbol{x})\big|_{2,\mathrm{d}} \cdot |\boldsymbol{d}|_{2,\mathrm{d}} \cos\theta = \big|\nabla f(\boldsymbol{x})\big|_{2,\mathrm{d}} \cos\theta$$

となるので $\cos\theta = -1$ と取ると

$$\boldsymbol{d} := -\frac{\nabla f(\boldsymbol{x})}{\big|\nabla f(\boldsymbol{x})\big|_{2,\mathrm{d}}}$$

となる. このことより方向 $\boldsymbol{d} = -\nabla f(\boldsymbol{x})$ を点 $\boldsymbol{x}$ における f の**最急降下方向** (steepest descent direction) とよぶ.

D.12.3 ステップ幅の選択方法

$k = 1, 2, \ldots$ とする. $\boldsymbol{x}_k$ における降下方向 $\boldsymbol{d}_k$ が選択されると, ステップ幅 t_k を適切に選択して, 次の点 $\boldsymbol{x}_{k+1}$ を

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t\boldsymbol{d}_k \qquad (t_k > 0)$$

と定める. この方法を**直接降下法**と呼ぶ. t_k を適切に探索しないと $\boldsymbol{x}_k$ が停留点に収束する保証が得られないことがある.

代表的なステップ幅の選択方法である Armijo 法, Goldstein 法, Wolfe 法を説明する.

Armijo 法

$s > 0$, $0 < \beta < 1$, $0 < \sigma < 1$ を固定する. $i = 0, 1, 2, \ldots$ に対して

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \beta^i s\boldsymbol{d}_k) \geq -\sigma(\beta^i s)\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \tag{D.118}$$

をテストする.

$i = 0$ とすると

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + s\boldsymbol{d}_k) \geq -\sigma\big\langle \nabla f(\boldsymbol{x}),\, s\boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \tag{D.119}$$

をテストすることになる. 不等式 (D.119) が成り立てば

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + s\boldsymbol{d}_k$$

とする. 逆に, (D.119) が成立していなければ, ステップ幅が大きすぎるので, $i = 1$ の (D.118) をテストする. すなわち

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \beta s\boldsymbol{d}_k) \geq -\beta\sigma\big\langle \nabla f(\boldsymbol{x}_k),\, s\boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}$$

をテストする.

以上の操作を繰り返して, i をおおくしていくとある i で (D.118) が必ず成立する.

このことを背理法を用いて証明する. すなわち, (D.118) をみたす $i \in \mathbb{N}$ が存在しないと仮定して, 矛盾を導く. i が十分おおきいとき, $t := \beta^i s > 0$ は十分小さい. したがって (D.118) において

$$f(\boldsymbol{x}_k + t\boldsymbol{d}_k) < f(\boldsymbol{x}_k) + \sigma t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}$$

であったとする. ここで, 上式と

$$f(\boldsymbol{x}_k + t\boldsymbol{d}_k) = f(\boldsymbol{x}_k) + t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} + o(t)$$

を比較すると

$$(\sigma - 1)\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} + \frac{o(t)}{t} < 0$$

となる. $t \to 0$ とすると

$$(\sigma - 1)\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \leq 0$$

となり, $\sigma - 1 < 0$ なので

$$\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} > 0$$

ととなる. これは $\boldsymbol{d}_k$ が降下方向であることと矛盾することから (D.118) をみたす i が存在することがわかる.

$\mathbb{R}^d$ の点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は最急降下法により求められる点列とし, $\boldsymbol{d}_k$ を $\boldsymbol{x}_k$ における f の降下方向とする. $-\nabla f(\boldsymbol{x}_k)$ と $\boldsymbol{d}_k$ の角度を θ_k と書く. するとある $\epsilon > 0$ が存在して

$$\cos\theta_k := \frac{\big\langle -\nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}}{\big|\nabla f(\boldsymbol{x}_k)\big|_{2,\mathrm{d}}} \in (\epsilon,\, 1] \qquad |\boldsymbol{d}_k|_{2,\mathrm{d}} = 1 \tag{D.120}$$

となる.

Goldstein 法

$c \in (0,\, 1/2)$ を固定する. 降下方向 $\boldsymbol{d}_k\,(k = 1,\, 2,\, \ldots)$ に対して, ステップ幅 t_k を t に関する不等式

$$\begin{aligned} f(\boldsymbol{x}_k) + (1-c)t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} &\leq f(\boldsymbol{x}_k + t\boldsymbol{d}_k) \\ &\leq f(\boldsymbol{x}_k) + ct\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \end{aligned}$$

をみたすように取る. そして

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\boldsymbol{d}_k$$

と定める.

Wolfe 法

ふたつの定数 $0 < c_1 < c_2 < 1$ を固定する. 降下方向 $\boldsymbol{d}_k\,(k = 1,\, 2,\, \ldots)$ に対して, ステップ幅 t_k を t に関する不等式

$$\begin{aligned} f(\boldsymbol{x}_k + t\boldsymbol{d}_k) &\leq f(\boldsymbol{x}_k) + c_1 t\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}}, \\ \big\langle \nabla f(\boldsymbol{x}_k + t\boldsymbol{d}_k,\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} &\geq c_2\big\langle \nabla f(\boldsymbol{x}_k),\, \boldsymbol{d}_k \big\rangle_{2,\mathrm{d}} \end{aligned}$$

をみたすように取る. そして

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + t_k\boldsymbol{d}_k$$

と定める.

D.13　降下法アルゴリズムの収束

Armijo 法でステップ幅を定めた最急降下法の収束を保証する定理を述べる．Goldstein 法と Wolfe 法による降下法の収束は Güller (2010, pp. 367-368) を参照のこと．

定理 D.77. $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ を降下法によって生成された点列で

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

とする．ただし $\boldsymbol{d}_k$ は (D.118) をみたし, α_k はパラメータ s, β, σ に対する Armijo 条件をみたしているとする．$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^*$ としたとき $\boldsymbol{x}^*$ は f の停留点になる．すなわち

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}_d$$

をみたす．

Proof. 背理法で証明する．すなわち, $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}_d$ を仮定して矛盾を導く．必要であれば $\{\boldsymbol{d}_k\}_{k=1}^{\infty}$ の部分列を取ることで $\{\boldsymbol{d}_{k_i}\}_{i=1}^{\infty}$ はある点 $\boldsymbol{d}^* \in \{\boldsymbol{d} \in \mathbb{R}^d;\, |\boldsymbol{d}|_{2,\mathrm{d}} = 1\}$ に収束する．$|\boldsymbol{d}^*|_2 = 1$ である．i 段階において Armijo 条件が成立していたとする．すなわち

$$\begin{aligned} f(\boldsymbol{x}_{k_i}) - f(\boldsymbol{x}_{k_i} + \alpha_{k_i}\boldsymbol{d}_{k_i}) &\geq -\sigma\alpha_{k_i}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} \\ &= \sigma\alpha_{k_i}\big|\nabla f(\boldsymbol{x}_{k_i})\big|_{2,\mathrm{d}} \cos\theta_{k_i} \\ &\geq \epsilon\sigma\alpha_{k_i}\big|\nabla f(\boldsymbol{x}_{k_i})\big|_{2,\mathrm{d}} \end{aligned} \tag{D.121}$$

となる．ただし, ϵ は (D.120) で定めたものである．α_{k_i} で Armijo 条件が成立しているので, $i-1$ ステップでは Armijo 条件は成立していないことになる．このことから

$$f(\boldsymbol{x}_{k_i}) - f\left(\boldsymbol{x}_{k_i} + \frac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right) < -\sigma\frac{\alpha_{k_i}}{\beta}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} \tag{D.122}$$

が成立していることになる．$k \to \infty$ のとき, 点列 $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ は $\boldsymbol{x}^*$ に収束するので, その部分列も $\boldsymbol{x}^*$ に収束する．よって, $\boldsymbol{x}_{k_i} \xrightarrow{i\to\infty} \boldsymbol{x}^*$ で $\{f(\boldsymbol{x}_{k_i})\}_{i=1}^{\infty}$ は減少列なので, f の連続性から

$$f(\boldsymbol{x}_{k_i}) \searrow f(\boldsymbol{x}^*)$$

となる．(D.121) の左辺は $i \to \infty$ のとき 0 に収束する．なぜならば $f(\boldsymbol{x}_{k_i}) - f(\boldsymbol{x}_{k_{i+1}}) \xrightarrow{i\to\infty} 0$ からわかる．さらに $\alpha_{k_i} = |\boldsymbol{x}_{k_i} - \boldsymbol{x}_{k_{i+1}}|_2 \xrightarrow{i\to\infty} 0$ なので, (D.121) の他の項も $i \to \infty$ のときに 0 に収束する．$\boldsymbol{d}_k$ の定め方から (D.120) が成立するので,

$$\lim_{i\to\infty}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} = \big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^*\big\rangle_{2,\mathrm{d}} \quad \text{かつ} \quad \big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i}\big\rangle_{2,\mathrm{d}} < 0$$

である. さらに, 背理法の仮定から $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}_d$ なので

$$\big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^* \big\rangle_{2,\mathrm{d}} < 0, \quad \alpha_{k_i} = |\boldsymbol{x}_{k_i} - \boldsymbol{x}_{k_{i+1}}|_{2,\mathrm{d}} \to 0\,(i \to \infty) \qquad \text{(D.123)}$$

がわかる. i が十分大きいとき $\boldsymbol{x}_{k_i}$ においてバックトラックが必要となることがわかる.

一方, 中間値の定理から $\boldsymbol{z}_{k_i} \in \left(\boldsymbol{x}_{k_i},\, \boldsymbol{x}_{k_i} + \dfrac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right)$ が存在して

$$f(\boldsymbol{x}_{k_i}) - f\left(\boldsymbol{x}_{k_i} + \frac{\alpha_{k_i}}{\beta}\boldsymbol{d}_{k_i}\right) = -\frac{\alpha_{k_i}}{\beta}\big\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \big\rangle_{2,\mathrm{d}}$$

とできる. (D.122) と比較すると

$$\begin{aligned} &-\frac{\alpha_{k_i}}{\beta}\big\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \big\rangle_{2,\mathrm{d}} < -\sigma\frac{\alpha_{k_i}}{\beta}\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i} \big\rangle_{2,\mathrm{d}} \\ &\Leftrightarrow\ \big\langle \nabla f(\boldsymbol{z}_{k_i}),\, \boldsymbol{d}_{k_i} \big\rangle_{2,\mathrm{d}} > \sigma\big\langle \nabla f(\boldsymbol{x}_{k_i}),\, \boldsymbol{d}_{k_i} \big\rangle_{2,\mathrm{d}} \end{aligned}$$

を得る. $i \to \infty$ のとき $\lim \boldsymbol{z}_{k_i} = \boldsymbol{x}^*$ となり, $0 < \sigma < 1$ なので, 上の 2 番目の不等式より

$$\big\langle \nabla f(\boldsymbol{x}^*),\, \boldsymbol{d}^* \big\rangle_{2,\mathrm{d}} \geq 0$$

を得る. これは (D.123) と矛盾. よって $\nabla f(\boldsymbol{x}^*) = \mathbf{0}_d$ を示すことができたが示せた. □

D.14 章末注釈と参考文献