機械学習のための数学

博士（理学）　飯塚　秀明［著］

コロナ社

ま　え　が　き

人工知能 (AI, Artificial Intelligence) は，交通事故の削減や交通の効率化のための自動運転技術，医用画像の解析による病変や疾患の検出，株価予測や不正検知に挙げられるような交通，医療，金融を含むさまざまな分野で活躍している。会話型人工知能や動画生成人工知能の急速な技術革新により，人工知能が我々の生活においてより身近な存在になっている。

人工知能開発のための手法の一つに，機械学習 (machine learning) がある。機械学習は，与えられたデータからルールやパターンを学習し，それらを用いて未知のデータに対する予測や判断を行うことができる。例えば，自動運転技術に関する行動予測においては，過去の交通データを学習することで，他の車両や歩行者の行動を予測するのに機械学習が利用される。

機械学習［7 章］を理解するためには “数学” が必要である。数学の土台は論理［1 章］と集合［2 章］である。機械学習を議論するための土俵がユークリッド空間［3 章］である。機械学習で扱うユークリッド空間の次元数はとてつもなく大きい。例えば，カラー画像のデータセットからルールを学習するための機械学習に現れる空間の次元数は一千万を超える。このことから，一般の次元数を有するユークリッド空間［3.2 節］の性質を理解することが，機械学習を理解するための第一歩といえる。一方で，1 次元ユークリッド空間［3.1 節］，つまり，実数直線を軽んじてはいけない。なぜならば，機械学習などに関する数理的手法の性能は数値化して評価されることが多いからである。

4 次元ユークリッド空間のベクトルは 2 行 2 列の行列（2×2 行列）として表現することができる。この事実は一般のベクトルと行列に対しても成り立つので，ユークリッド空間［3 章］と行列に関する理論を展開する線形代数［4 章］には密接な関係がある［4.1 節］。最適化理論や機械学習を理解するうえで重要

な行列といえば対称行列［4.1.2 項］であろう。機械学習に現れる関数の 2 回微分係数［例 4.3(3)，例 4.6(3)］や確率変数の共分散［例 4.6(4)］は対称行列である。行列特有の距離（ノルム）の測り方［4.2 節］は最適化理論や機械学習のための手法の解析で役立つ。機械学習のための手法の多くは，対象の関数［7.1 節］の微分情報を利用し，その解析のために積分の概念を利用する［5 章］。ユークリッド空間［3.2 節］上で定義される関数の微分係数（勾配）［5.1 節］は，勾配降下法［7.2 節］や確率的勾配降下法［7.3 節］といった機械学習のための最適化手法を定義するうえで必要な概念である。さらに，微分可能な凸関数とその 2 回微分係数（対称行列）の関係［5.3.2 項］は微分積分と線形代数の架け橋となる結果を示す。確率・統計［6 章］は，データの確率分布の理解とその推定に必要な概念であり，確率変数の独立性，および，その期待値と分散の性質［6.3.2 項］を駆使することで，機械学習のための最適化手法の解析が可能となる［7.3 節］。機械学習のための最適化手法で利用する特有のパラメータ（ステップサイズとバッチサイズ）の設定による多峰性を有する関数の大域的最小化［7.4 節］は機械学習に限らず，最適化理論の観点から見ても興味深い。

本書の読者対象は理工系の大学 2 年生程度の知識をもつ方を想定している。本書が機械学習関連の論文を読み解こうとする方の一助を担うことができれば，著者にとって大きな喜びである。なお，章末問題の解答は，明治大学理工学部情報科学科数理最適化研究室が管理する GitHub Organization の Mathematical Optimization Lab. (`https://github.com/iiduka-researches`) からダウンロードすることができる。

本書の原稿に関して，適切な指摘と校正に協力いただいた明治大学数理最適化研究室の佐藤尚樹氏と大和田佳生氏に深く感謝する。また，本書の出版に関して，大変お世話になったコロナ社の皆様に深く御礼申し上げる。

最後に，理科系の道を志す我が息子たちにこの本を贈る。

2024 年 6 月　生田にて

飯塚秀明

目　　　　次

1.　論　　　　理

2. 集　合

3. ユークリッド空間

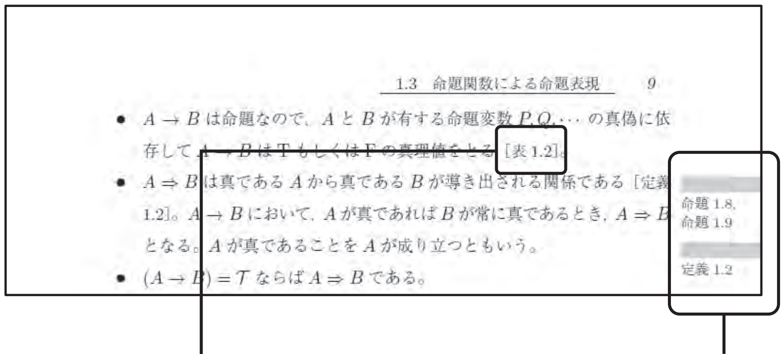
4. 線　形　代　数

5. 微　分　積　分

6.　確 率 ・ 統 計

7.　機械学習と最適化

本書の使い方


1.3　命題関数による命題表現　　9

- $A \to B$ は命題なので，A と B が有する命題変数 $P, Q, \cdots$ の真偽に依存して $A \to B$ は T もしくは F の真理値をとる［表 1.2］。
- $A \Rightarrow B$ は真である A から真である B が導き出される関係である［定義 1.2］。$A \to B$ において，A が真であれば B が常に真であるとき，$A \Rightarrow B$ となる。A が真であることを A が成り立つともいう。
- $(A \to B) = T$ ならば $A \Rightarrow B$ である。

命題 1.8,
命題 1.9
定義 1.2

見開きページ上にある「式」「命題」「定義」「定理」「例」「注意」を表しています。

この文章で参照しているものを表しています。
適宜参照しながら読み進めてください。

$\forall \boldsymbol{x} \in D\ \forall \boldsymbol{y} \in D\ \forall \boldsymbol{g}_{\boldsymbol{x}} \in \partial f(\boldsymbol{x})\ \forall \boldsymbol{g}_{\boldsymbol{y}} \in \partial f(\boldsymbol{y})$
$(\langle \boldsymbol{g}_{\boldsymbol{x}} - \boldsymbol{g}_{\boldsymbol{y}}, \boldsymbol{x} - \boldsymbol{y} \rangle \geqq c\|\boldsymbol{x} - \boldsymbol{y}\|^2)$

式 (5.47)

命題 5.18,
命題 5.19

5.3.2　微分可能凸関数とヘッセ行列：微分積分と線形代数の架け橋

機械学習に現れる関数は一般には高次元ユークリッド空間の部分集合を定義域にもつ非凸関数［7.1 節］であるため，これまでに詳解した凸関数の性質や解析を実問題に現れる最適化問題へ適用ができないと感じるかもしれない。しかしながら，この命題は真ではないことを示す［7.4.2 項，注意 7.5］。

定理 5.8

初めに，2 回連続的微分可能凸関数 f の必要十分条件がヘッセ行列 $\mathrm{H}_f(\boldsymbol{x})$ $(= \nabla^2 f(\boldsymbol{x}))$ の性質により表現できることを示す。定理 5.8 の証明は Taylor の定理［定理 5.6］を利用する。

定理 5.8　（2 回連続的微分可能凸関数の必要十分条件）　$c > 0$ とする。ユークリッド内積［式 (3.11)］から誘導されるユークリッドノルム［式 (3.15)］を有する d 次元ユークリッド空間 $\mathbb{R}^d$ の開凸集合 D で定義される C^2 級関数 $f\colon D \to \mathbb{R}$ に対して，(1)〜(3) が成り立つ。

(1)　[凸関数の必要十分条件] f が凸 $\Leftrightarrow \forall \boldsymbol{x} \in D\ (\mathrm{H}_f(\boldsymbol{x}) \in \mathbb{S}^d_+)$

(2)　[狭義凸関数の十分条件] $\forall \boldsymbol{x} \in D\ (\mathrm{H}_f(\boldsymbol{x}) \in \mathbb{S}^d_{++}) \Rightarrow f$ が狭義凸

(3)　[強凸関数の必要十分条件] f が c–強凸 $\Leftrightarrow \forall \boldsymbol{x} \in D\ (\mathrm{H}_f(\boldsymbol{x}) - cI \in \mathbb{S}^d_+) \underset{[定理\ 4.1(2)]}{\Leftrightarrow} \forall \boldsymbol{x} \in D\ (\lambda_{\min}(\mathrm{H}_f(\boldsymbol{x})) = c)$

証明　ヘッセ行列 $\mathrm{H}_f(\boldsymbol{x})$ は $\mathrm{H}_f(\boldsymbol{x}) \underset{[定理\ 5.3(2)]}{=} \nabla^2 f(\boldsymbol{x}) \underset{[定理\ 5.4]}{\in} \mathbb{S}^d$ を満たす。

(1)　$\Rightarrow$) $\boldsymbol{x} \in D \underset{[命題\ 3.12]}{\Rightarrow} \exists \delta > 0\ (N_2(\boldsymbol{x}; \delta) \subset D)$ から，$\boldsymbol{y}\ (\neq \boldsymbol{0}) \in \mathbb{R}^d \Rightarrow \forall \alpha \in (0, \delta/\|\boldsymbol{y}\|_2)\ (\boldsymbol{x} + \alpha \boldsymbol{y} \in N_2(\boldsymbol{x}; \delta) \subset D)$ を満たす。Taylor の定理［定理 5.6］から，$\tau \in (0,1)$ が存在して

$$f(\boldsymbol{x} + \alpha \boldsymbol{y}) = f(\boldsymbol{x}) + \alpha \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} \rangle_2 + \frac{\alpha^2}{2} \left\langle \boldsymbol{y}, \nabla^2 f(\boldsymbol{z}_\tau(\alpha)) \boldsymbol{y} \right\rangle_2 \qquad (5.47)$$

を満たす。ただし，$\boldsymbol{z}_\tau(\alpha) = \boldsymbol{x} + \tau \alpha \boldsymbol{y}$ は $\boldsymbol{z}_\tau(\alpha) \to \boldsymbol{x}\ (\alpha \downarrow 0)$ を満たす。

$$
\begin{aligned}
f \text{ が } D \text{ で凸} &\underset{[\text{命題 } 5.16(1)]}{\overset{[\text{定理 } 5.7(3)]}{\Rightarrow}} f(\boldsymbol{x}+\alpha\boldsymbol{y}) \geqq f(\boldsymbol{x}) + \alpha\langle\nabla f(\boldsymbol{x}), \boldsymbol{y}\rangle_2 \\
&\underset{[\text{式 } (5.47)]}{\Rightarrow} \left\langle \boldsymbol{y}, \nabla^2 f(\boldsymbol{z}_\tau(\alpha))\boldsymbol{y}\right\rangle_2 \geqq 0 \\
&\underset{[\text{定義 } 5.7]}{\overset{[\text{例 } 3.5(1)]}{\Rightarrow}} \left\langle \boldsymbol{y}, \nabla^2 f(\boldsymbol{x})\boldsymbol{y}\right\rangle_2 \geqq 0 \underset{[\text{定理 } 4.2(1)]}{\Leftrightarrow} \nabla^2 f(\boldsymbol{x}) \in \mathbb{S}_+^d
\end{aligned}
$$

$\Leftarrow$) $\boldsymbol{x}, \boldsymbol{y} \in D$ とする。Taylor の定理［定理 5.6］から，$\tau \in (0,1)$ が存在して，$\boldsymbol{z}_\tau = \boldsymbol{x} + \tau(\boldsymbol{y}-\boldsymbol{x}) = \tau\boldsymbol{y} + (1-\tau)\boldsymbol{x} \in D$ は以下を満たす。

$$
\begin{aligned}
f(\boldsymbol{y}) &\underset{[\text{定理 } 5.6]}{=} f(\boldsymbol{x}) + \langle\nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle_2 + \frac{1}{2}\left\langle \boldsymbol{y}-\boldsymbol{x}, \nabla^2 f(\boldsymbol{z}_\tau)(\boldsymbol{y}-\boldsymbol{x})\right\rangle_2 \\
&\underset{[\text{定理 } 4.2(1)]}{\geqq} f(\boldsymbol{x}) + \langle\nabla f(\boldsymbol{x}), \boldsymbol{y}-\boldsymbol{x}\rangle_2 \underset{[\text{命題 } 5.16(1)]}{\overset{[\text{定理 } 5.7(3)]}{\Rightarrow}} f \text{ が } D \text{ で凸}
\end{aligned}
$$

(2), (3)　定理 5.8(1) の証明方法と同様の議論（Taylor の定理を利用）と命題 5.18(1)，命題 5.19(1) を用いることで，定理 5.8(2), (3) を示すことができる（詳細については，文献13) の演習問題 2.16，2.17 解答例を参照されたい）。 □

定理 5.8 は，2 回連続的微分可能関数の凸性とヘッセ行列の正定値性［4.1.3 項］が同等の力を有することを示しており，まさに，微分積分と線形代数の架け橋を担っている。

注意 5.11：[1 変数凸関数の必要十分条件] 関数 $f\colon \mathbb{R} \to \mathbb{R}$ の（1 次）導関数 f' と 2 次導関数 f'' の情報を用いて

$$
\begin{aligned}
f \text{ が凸} &\Leftrightarrow f' \text{が単調増加（つまり}, x_1 \leqq x_2 \Rightarrow f'(x_1) \leqq f'(x_2)\text{）} \\
&\Leftrightarrow f'' \text{が非負値（つまり}, f''(x) \geqq 0\text{）}
\end{aligned}
$$

となる。f' の単調増加性は，定理 5.7(3) と命題 5.16(2) から得られる

$$
\langle\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle \geqq 0
$$

の一例（$(f'(x_2) - f'(x_1))(x_2 - x_1) \geqq 0$）であり，$f''$ の非負値性（半正定値性）は定理 5.8(1) の一例となる。例えば，1 変数 2 次凸関数 $f(x) = ax^2 + bx + c$（ただし，$a \geqq 0$）において，$f'(x) = 2ax + b$ は x に関して単調増加であり，$f''(x) = 2a$ の固有値 λ と固有ベクトル $y\ (\neq 0)$ は以下を満たす［式 (4.10)]。

$$
f''(x)y = \lambda y \Leftrightarrow 2ay = \lambda y \Rightarrow \lambda = 2a \geqq 0 \Leftrightarrow f''(x) \in \mathbb{S}_+^1
$$

注意 5.12：[狭義凸関数のヘッセ行列の非負固有値] 定理 5.8(2) の逆命題が成り立たないことを反例［1.4.3 項］を用いて示す。つまり，$\exists \boldsymbol{x}(\nabla^2 f(\boldsymbol{x}) \notin \mathbb{S}^d_{++})$ を満たす狭義凸関数 f が存在することを示す。$f(x) = x^4$ とすると，$f'(x) = 4x^3$，$f''(x) = 12x^2$ となる。$x_1 \neq x_2$ となる $x_1, x_2 \in \mathbb{R}$ に対して，$(f'(x_2) - f'(x_1))(x_2 - x_1) = 4(x_2 - x_1)^2(x_2^2 + x_1x_2 + x_1^2) > 0$ から，$f(x) = x^4$ は狭義凸である［命題 5.18(2)］。一方で，$\exists x(= 0)(f''(x) = 12x^2 = 0)$ から，$f''(0) \notin \mathbb{S}^1_{++}$ となる。

定理 5.8 を用いて，凸関数の例を挙げる。

例 5.6　（凸関数の例（その 2））

例 5.6

(1) 比例公平関数［例 5.4(1)］：$\forall i \in \{1, 2, \cdots, d\}$ $(w_i > 0)$ とする。$\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^\top \in \mathbb{R}^d_{++}$ に対して，比例公平関数は $g(\boldsymbol{x}) = \displaystyle\sum_{i=1}^d w_i \log x_i$ で定義される。このとき，$f = -g \colon \mathbb{R}^d_{++} \to \mathbb{R}$ は狭義凸関数である。

注意 5.11，注意 5.12

(2) 2 次関数［例 5.1(1)，例 5.2(1)］：$A \in \mathbb{S}^d$ とする。例 5.1(1) で定義される 2 次関数 $f(\boldsymbol{x}) = (1/2)\langle \boldsymbol{x}, A\boldsymbol{x}\rangle_2 + \langle \boldsymbol{b}, \boldsymbol{x}\rangle_2 + c$ の 2 次微分係数

$$\nabla^2 f(\boldsymbol{x}) \underset{[\text{例 } 5.2(1)]}{=} \frac{1}{2}(A + A^\top) \underset{[A \in \mathbb{S}^d]}{=} A$$

により，$A \in \mathbb{S}^d_+$ のとき，f は凸である（1 変数 2 次凸関数については，注意 5.11 を参照）。さらに，$A \in \mathbb{S}^d_{++}$ のとき，f は $\lambda_{\min}(A)$–強凸，かつ，$\lambda_{\max}(A)$–平滑となる。

(3) 連続写像の合成［例 5.1(2)，例 5.2(2)］：例 5.1(2) で定義される関数 $f(\boldsymbol{x}) = (1/2)\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2$ の 2 次微分係数

$$\nabla^2 f(\boldsymbol{x}) \underset{[\text{例 } 5.2(2)]}{=} A^\top A \underset{[\text{例 } 4.3(2)]}{\in} \mathbb{S}^n$$

により，$A^\top A \in \mathbb{S}^n_+$ のとき，f は凸であり，また，$A^\top A \in \mathbb{S}^n_{++}$ のとき，f は $\lambda_{\min}(A^\top A)$–強凸，かつ，$\lambda_{\max}(A^\top A)$–平滑となる。

証明

(1) f は $\mathbb{R}^d_{++}$ で C^2 級なので，$\forall \boldsymbol{x} \in \mathbb{R}^d_{++}$ $(\nabla^2 f(\boldsymbol{x}) \underset{[\text{定理 } 5.3(2)]}{=} \mathrm{H}_f(\boldsymbol{x}) =$

$\mathrm{diag}(w_i/x_i^2)) \underset{[例\ 4.6(2)]}{\Rightarrow} \forall \boldsymbol{x} \in \mathbb{R}^d_{++}\ (\nabla^2 f(\boldsymbol{x}) \in \mathbb{S}^d_{++}) \underset{[定理\ 5.8(2)]}{\Rightarrow} f$ は狭義凸

(2)　$\nabla^2 f(\boldsymbol{x}) = A \in \mathbb{S}^d_+$ のとき，定理 5.8(1) から，f は凸である。$\nabla^2 f(\boldsymbol{x}) = A \in \mathbb{S}^d_{++}$ のとき，$\lambda_{\min}(A) > 0$ に注意して

$$\forall \boldsymbol{x}(\lambda_{\min}(A)\|\boldsymbol{x}\|_2^2 \underset{[式\ (4.17)]}{\leqq} \langle \boldsymbol{x}, A\boldsymbol{x}\rangle_2 \Leftrightarrow \langle \boldsymbol{x}, (A - \lambda_{\min}(A)I)\boldsymbol{x}\rangle_2 \geqq 0)$$

$$\underset{[定理\ 4.2(1)]}{\Leftrightarrow} A - \lambda_{\min}(A)I \in \mathbb{S}^d_+ \underset{[定理\ 5.8(3)]}{\Leftrightarrow} f \text{ は } \lambda_{\min}(A)\text{–強凸}$$

となる。例 5.1(1) から，f の $\lambda_{\max}(A)$–平滑性が示される。

(3)　例 5.6(2) と同様の議論により証明ができる。　□

最小化問題の最小解［定義 3.14］について定義をする。

定義 5.12　（大域的最小解と局所的最小解）　ノルム空間 $\mathbb{R}^d$ で定義される関数 $f\colon \mathbb{R}^d \to \mathbb{R}$ に関する最小化問題を考察する†。このとき，f の値を最も小さくするベクトル $\boldsymbol{x}^\star$，つまり

$$\forall \boldsymbol{x} \in \mathbb{R}^d\ (f(\boldsymbol{x}^\star) \leqq f(\boldsymbol{x})) \Leftrightarrow \boldsymbol{x}^\star \in \underbrace{\left\{\boldsymbol{x}^\star \in \mathbb{R}^d \colon f(\boldsymbol{x}^\star) = \min_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x})\right\}}_{\mathrm{argmin}\, f(\boldsymbol{x})}$$

を満たす $\boldsymbol{x}^\star$ を最小化問題の**大域的最小解** (global minimum solution) という。ベクトル $\boldsymbol{x}_\star$ が最小化問題の**局所的最小解** (local minimum solution) であるとは，$\boldsymbol{x}_\star$ の局所的な範囲［例 3.4(1)］に限った場合，その範囲で f の値を最も小さくするベクトル $\boldsymbol{x}_\star$，つまり，以下を満たすときをいう。

$$\exists \delta > 0\ \forall \boldsymbol{x} \in \mathbb{R}^d\ (\boldsymbol{x} \in N(\boldsymbol{x}_\star; \delta) \Rightarrow f(\boldsymbol{x}_\star) \leqq f(\boldsymbol{x}))$$

定義 5.12 から，最小化問題の大域的最小解はその問題の局所的最小解となる。2 回連続的微分可能関数に関する最小化問題の局所的最小解となるための十分

† ある関数 $g\colon \mathbb{R}^d \to \mathbb{R}$ の最大化を考察する。このとき，$f = -g$ で定義される関数 f を最小化することと g を最大化することは同値である。よって，最適化においては最小化について考察すれば事が足りる。

条件を，Taylor の定理［定理 5.6］を利用することで定理 5.9 のように与えることができる。

式 (5.48)

定義 5.12

定理 5.9

定理 5.9 （局所的最小解のための十分条件） ユークリッド内積［式 (3.11)］から誘導されるユークリッドノルム［式 (3.15)］を有する d 次元ユークリッド空間 $\mathbb{R}^d$ で定義される C^2 級関数 f の最小化問題を考察する。$\boldsymbol{x}_\star \in \mathbb{R}^d$ とする。このとき

$$\nabla f(\boldsymbol{x}_\star) = \mathrm{D}_f(\boldsymbol{x}_\star) = \boldsymbol{0} \wedge \nabla^2 f(\boldsymbol{x}_\star) = \mathrm{H}_f(\boldsymbol{x}_\star) \in \mathbb{S}^d_{++}$$

ならば，$\boldsymbol{x}_\star$ は f に関する最小化問題の局所的最小解である。

証明 $\boldsymbol{d} \neq \boldsymbol{0}$ を満たす $\boldsymbol{d} \in \mathbb{R}^d$ を任意にとると，$\nabla^2 f(\boldsymbol{x}_\star) = \mathrm{H}_f(\boldsymbol{x}_\star) \in \mathbb{S}^d_{++} \underset{[\text{定理 } 4.4(1)]}{\Rightarrow} \varepsilon = \langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_\star)\boldsymbol{d}\rangle_2 > 0$ となる。$\nabla^2 f$ の連続性［定義 5.7］から，$F(\alpha) = \nabla^2 f(\boldsymbol{x}_\star + \alpha\boldsymbol{d})$ で定義される $F\colon \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ は連続である。内積の連続性［例 3.5(1)］から

$$\begin{aligned}
&\exists\delta > 0\ \forall\alpha \geqq 0\ \Bigl(\alpha \in [0,\delta) \\
&\quad \Rightarrow \bigl|\bigl\langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_\star + \alpha\boldsymbol{d})\boldsymbol{d}\bigr\rangle_2 - \bigl\langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_\star)\boldsymbol{d}\bigr\rangle_2\bigr| < \frac{\varepsilon}{2} \\
&\quad \Rightarrow \underbrace{\bigl\langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_\star + \alpha\boldsymbol{d})\boldsymbol{d}\bigr\rangle_2 > \frac{\varepsilon}{2}}_{\nabla^2 f(\boldsymbol{x}_\star + \alpha\boldsymbol{d}) \in \mathbb{S}^d_{++}}\Bigr)
\end{aligned} \tag{5.48}$$

となる。Taylor の定理［定理 5.6］から，$\forall\alpha \in [0,\delta)\ \exists\tau \in (0,1)$:

$$\begin{aligned}
f(\boldsymbol{x}_\star + \alpha\boldsymbol{d}) &= f(\boldsymbol{x}_\star) + \langle \underbrace{\nabla f(\boldsymbol{x}_\star)}_{=\boldsymbol{0}}, \boldsymbol{d}\rangle_2 \alpha + \frac{1}{2} \underbrace{\bigl\langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_\star + \tau\alpha\boldsymbol{d})\boldsymbol{d}\bigr\rangle_2}_{>0\ \because (5.48)} \alpha^2 \\
&\geqq f(\boldsymbol{x}_\star)
\end{aligned}$$

となる。$\|\boldsymbol{d}\|_2 \leqq 1$ を満たす $\boldsymbol{d}$ $(\neq \boldsymbol{0})$ に対しては，$\boldsymbol{x}_\star + \alpha\boldsymbol{d} \in N_2(\boldsymbol{x}_\star;\delta)$ となる。よって，α と $\boldsymbol{d}$ の任意性により，$\forall \boldsymbol{y} \in N_2(\boldsymbol{x}_\star;\delta)$ $(f(\boldsymbol{x}_\star) \leqq f(\boldsymbol{y}))$ を得る。 □

定理 5.9 から，$\boldsymbol{x}_\star$ が局所的最適解になるためには，勾配（(1 次）微分係数）$\nabla f(\boldsymbol{x}_\star)$ が零ベクトルになることを要求しているが，これは，1 変数微分可能

関数 f でいえば，x_* における f の接線の傾き $f'(x_*)$ が零となることなので直観的にも理解ができる。$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$ を満たす $\boldsymbol{x}$ は局所的最小解，もしくは，局所的最大解のいずれかであろう（**図 5.3** の $\boldsymbol{x}_\star$ と $\boldsymbol{z}$ を参照）。ここで，最小化問題の局所的最適解のためのもう一つの十分条件「2 次微分係数（ヘッセ行列）$\nabla^2 f(\boldsymbol{x}_\star)$ の正定値性」が重要な役割を演じる。具体的には

$$
\begin{aligned}
&\nabla^2 f(\boldsymbol{x}_\star) = \mathrm{H}_f(\boldsymbol{x}_\star) \in \mathbb{S}^d_{++} \\
&\underset{[\text{式 (5.48)}]}{\Rightarrow} \exists \delta > 0\ \forall \boldsymbol{y} \in N_2(\boldsymbol{x}_\star; \delta)\ (\nabla^2 f(\boldsymbol{y}) = \mathrm{H}_f(\boldsymbol{y}) \in \mathbb{S}^d_{++}) \\
&\underset{[\text{定理 5.8(2)}]}{\Rightarrow} f \text{ は } \boldsymbol{x}_\star \text{ の } \delta\text{–近傍 } N_2(\boldsymbol{x}_\star; \delta) \text{ で狭義凸} \\
&\underset{[\nabla f(\boldsymbol{x}_\star) = \boldsymbol{0}]}{\Rightarrow} \boldsymbol{x}_\star \text{ は最小化問題の局所的最小解}
\end{aligned}
$$

である。この結果により，局所的最小解の近傍においては関数の凸性が保証される（図 5.3 の $\boldsymbol{x}_\star$ と $\boldsymbol{x}^\star$ の近傍を参照）。

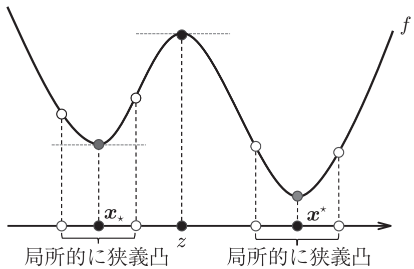


図 5.3　2 回連続的微分可能関数の局所的凸性

一般の非凸関数に関する最小化問題を考察するとき，対象とする非凸関数の最小解の近傍では凸関数になることから，最小解の周りでの解析をするうえで凸関数の性質や解析を適用することができる。例えば，ある最適化手法（例えば，勾配降下法［7.2 節］）により得られた近似解 $\boldsymbol{x}_\mathrm{A}$（勾配降下法を利用するとき，十分大きい反復回数 K に対して，$\boldsymbol{x}_\mathrm{A} = \boldsymbol{x}_{K+1} = \boldsymbol{x}_K - \alpha \nabla f(\boldsymbol{x}_K)$（ただし，$\boldsymbol{x}_0 \in \mathbb{R}^d$ とし，$\alpha > 0$ とする））が，十分小さい正数 $\varepsilon > 0$ に対して

$$\|\mathrm{D}_f(\boldsymbol{x}_\mathrm{A})\|_2 < \varepsilon \wedge \mathrm{H}_f(\boldsymbol{x}_\mathrm{A}) \in \mathbb{S}^d_{++} \tag{5.49}$$

式 (5.49)

を満たすならば，定理 5.9 により，$\boldsymbol{x}_\mathrm{A}$ は正しく局所的最小解 $\boldsymbol{x}_\star$ を近似していることがわかる。式 (5.49) の検証は，例えば，$\varepsilon = 10^{-3}$ として

$$\|\mathrm{D}_f(\boldsymbol{x}_\mathrm{A})\|_2^2 = \sum_{i=1}^{d}\left(\frac{\partial f(\boldsymbol{x}_\mathrm{A})}{\partial x_i}\right)^2 < (10^{-3})^2 \wedge \lambda_{\min}\left(\mathrm{H}_f(\boldsymbol{x}_\mathrm{A})\right) > 0$$

であることを計算機を用いて調べればよい。

章　末　問　題

【5.1】 命題 5.2（微分作用素 ∇ の線形性）を証明せよ。

【5.2】 命題 5.3（方向微分係数と導関数の関係）内の式 (5.7) を式 (5.6) を用いて証明せよ。

【5.3】 命題 5.6（微分作用素 ∇^2 の線形性）を証明せよ。

【5.4】 命題 5.10（凸関数の性質）を証明せよ。

【5.5】 命題 5.14（劣微分作用素 ∂ の加法性と正数同次性）の証明内にある $\forall\alpha > 0\ \forall\boldsymbol{x} \in D_1\ (\partial(\alpha f_1)(\boldsymbol{x}) = \alpha\partial f_1(\boldsymbol{x}))$ と $\forall\boldsymbol{x} \in D_1 \cap D_2\ (\partial(f_1 + f_2)(\boldsymbol{x}) \supset \partial f_1(\boldsymbol{x}) + \partial f_2(\boldsymbol{x}))$ が成り立つことを示せ。

索　　　引

—— 著 者 略 歴 ——

2005年　東京工業大学大学院情報理工学研究科博士後期課程修了（数理・計算科学専攻）
　　　　博士（理学）
　　　　東京工業大学大学院補佐員
2007年　日本学術振興会特別研究員（PD）
2008年　九州工業大学ネットワークデザイン研究センター准教授
2013年　明治大学理工学部准教授
2019年　明治大学理工学部教授
　　　　現在に至る

機械学習のための数学
Mathematics for Machine Learning

2024 年 8 月 26 日　初版第 1 刷発行　★

検印省略

著　者　飯塚 秀明（いいづか ひであき）
発 行 者　株式会社　コロナ社
　　　　代 表 者　牛来真也
印 刷 所　三美印刷株式会社
製 本 所　有限会社　愛千製本所

112–0011　東京都文京区千石 4–46–10
発 行 所　株式会社　コロナ社
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140–8–14844・電話(03)3941–3131(代)
ホームページ　https://www.coronasha.co.jp

ISBN 978–4–339–06132–1　C3041　Printed in Japan　（田中）