信号処理 100 本ノック 解答

Yuma Kinoshita

1　線形代数の基礎

問 1　連立方程式

(i) 拡大係数行列 $\tilde{\mathrm{A}} = (\mathrm{A}|b)$ に対し，拡大係数行列の左側が以下のような行列 (対角に相当する成分のみ 1 を含み，他の成分がすべて 0 である行列) となるよう，行基本変形を施す.

$$\begin{pmatrix} 1 & & & & & & 0 \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{pmatrix} \tag{1}$$

この結果得られる拡大係数行列の右側が元の方程式の解に相当する．よって解 x は,

$$x = \begin{pmatrix} 6 \\ 2 \\ 3 \end{pmatrix} \tag{2}$$

である.

(ii)
- (i) と同様に拡大係数行列 $\tilde{\mathrm{A}} = (\mathrm{A}|b)$ を変形すると以下を得る (変形を途中でやめていることに注意).

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right) \tag{3}$$

  このことから，$\mathrm{rank}(\mathrm{A}) \neq \mathrm{rank}(\tilde{\mathrm{A}})$ であり，$Ax - b = 0$ は解を持たない.

- $\|\mathrm{A}x - b\|$ を最小とする x は，$\|\mathrm{A}x - b\|^2$ を最小とする x と等しい．よって，$\|\mathrm{A}x - b\|^2$ を最小とする x を考える．$\|\mathrm{A}x - b\|^2$ は二次式だから,

$$\frac{\partial}{\partial x}\|\mathrm{A}x - b\|^2 = 0 \tag{4}$$

  を満たす x を求めればよい．上式を解くと,

$$x = (\mathrm{A}^\top \mathrm{A})^{-1}\mathrm{A}^\top b \tag{5}$$

  を得る．したがって,

$$x = \begin{pmatrix} 1 \\ \frac{2}{3} \end{pmatrix}. \tag{6}$$

別解 1 $\|\mathrm{A}x - b\|^2$ を x の各要素について平方完成して最小値を求める

別解 2 $\|\mathrm{A}x - b\|^2$ を x の各要素について微分して,

$$\frac{\partial}{\partial x_1}\|\mathrm{A}x - b\|^2 = \frac{\partial}{\partial x_2}\|\mathrm{A}x - b\|^2 = 0 \tag{7}$$

を解き，最小値を求める.

(iii)● (i) と同様に拡大係数行列 $\tilde{\mathrm{A}} = (\mathrm{A}|b)$ を変形すると以下を得る．

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{8}$$

よって，解 x は，任意の $t \in \mathbb{R}$ を用いて

$$x = t\begin{pmatrix} -2 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ -1 \\ 0 \end{pmatrix} \tag{9}$$

として与えられる．

● $\|x\|$ を最小とする x は，$\|x\|^2$ を最小とする x と等しい．よって，$\|x\|^2$ を最小とする x を考える．$\|x\|^2$ は二次式だから，

$$\frac{d}{dt}\|x\|^2 = 0 \tag{10}$$

を満たす x を求めればよい．上式を解くと，

$$t = -\frac{2}{7} \tag{11}$$

を得る．したがって，

$$x = -\frac{1}{7}\begin{pmatrix} 3 \\ -5 \\ 9 \\ 2 \end{pmatrix}. \tag{12}$$

別解 1 $\|\mathrm{A}x - b\|^2$ を t について平方完成して最小値を求める

問 2　線形変換

(i) $\boldsymbol{y} = a\boldsymbol{x} + \boldsymbol{x}_{\perp}$ より，$\boldsymbol{x}^{\top}\boldsymbol{y} = \boldsymbol{x}^{\top}(a\boldsymbol{x} + \boldsymbol{x}_{\perp}) = a\|\boldsymbol{x}\|^2$．$a$ について整理すると，

$$a = \frac{1}{\|\boldsymbol{x}\|^2}\boldsymbol{x}^{\top}\boldsymbol{y} \tag{13}$$

を得る．

(ii) $\mathrm{A}\boldsymbol{y} = a\boldsymbol{x}$ に (i) の結果を代入することで，

$$\mathrm{A} = \frac{1}{\|\boldsymbol{x}\|^2}\boldsymbol{x}\boldsymbol{x}^{\top} \tag{14}$$

を得る．

問 3　行列のランク

(i) a

(ii) b

(iii) c

(iv) d

(v) e

問 4　行列式と逆行列の存在条件

(i)

$$\det(\mathrm{V}) = \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ 0 & x_2(x_2 - x_1) & x_3(x_3 - x_1) & x_4(x_4 - x_1) \\ 0 & x_2^2(x_2 - x_1) & x_3^2(x_3 - x_1) & x_4^2(x_4 - x_1) \end{vmatrix} \tag{15}$$

$$= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ x_2(x_2 - x_1) & x_3(x_3 - x_1) & x_4(x_4 - x_1) \\ x_2^2(x_2 - x_1) & x_3^2(x_3 - x_1) & x_4^2(x_4 - x_1) \end{vmatrix} \tag{16}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1) \begin{vmatrix} 1 & 1 & 1 \\ x_2 & x_3 & x_4 \\ x_2^2 & x_3^2 & x_4^2 \end{vmatrix} \tag{17}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1) \begin{vmatrix} 1 & 1 & 1 \\ 0 & x_3 - x_2 & x_4 - x_2 \\ 0 & x_3(x_3 - x_2) & x_4(x_3 - x_2) \end{vmatrix} \tag{18}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)(x_3 - x_2)(x_4 - x_2) \begin{vmatrix} 1 & 1 \\ x_3 & x_4 \end{vmatrix} \tag{19}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)(x_3 - x_2)(x_4 - x_2)(x_4 - x_3) \tag{20}$$

(ii) 行列 V が逆行列を持つための必要十分条件は，$\det(\mathrm{V}) \neq 0$ であることである．

$$\det(\mathrm{V}) = (x_2 - x_1)(x_3 - x_1)(x_4 - x_1)(x_3 - x_2)(x_4 - x_2)(x_4 - x_3) \neq 0 \tag{21}$$

より，$i, j \in \{1, 2, 3, 4\}$ かつ $i \neq j$ であるすべての (i, j) について $x_i \neq x_j$ を満たすとき，行列 V は逆行列を持つ．

問 5　回転行列

$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ とする.

(i) $\boldsymbol{e}'_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$, $\boldsymbol{e}'_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$

(ii)

$$\begin{pmatrix} \boldsymbol{e}'_1 & \boldsymbol{e}'_2 \end{pmatrix} = \mathrm{R}_\theta \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{pmatrix} \tag{22}$$

より,

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \mathrm{R}_\theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{23}$$

したがって,

$$\mathrm{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{24}$$

(iii)

$$\begin{aligned} \boldsymbol{\nu} &= \mathrm{R}_\phi \mathrm{R}_\theta \boldsymbol{e}_1 && (25) \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \boldsymbol{e}_1 && (26) \\ &= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi \end{pmatrix} && (27) \end{aligned}$$

(iv)

$$\mathrm{R}_{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{28}$$

であり,

$$\mathrm{R}_\theta \mathrm{R}_{-\theta} = \mathrm{R}_\theta \mathrm{R}_\theta = \mathrm{I} \tag{29}$$

（I は単位行列）より，$\mathrm{R}_\theta^{-1} = \mathrm{R}_{-\theta}$.

問 6　エルミート行列

(i) 任意の行列 $\mathrm{M} \in \mathbb{C}^{n\times m}$ について，$\mathrm{MM^H}$ および $\mathrm{M^HM}$ がエルミート行列となることを証明する．$(\mathrm{M\tilde{M}})^{\mathrm{H}} = \tilde{\mathrm{M}}^{\mathrm{H}}\mathrm{M^H}$ より，$(\mathrm{MM^H})^{\mathrm{H}} = (\mathrm{M^H})^{\mathrm{H}}\mathrm{M^H} = \mathrm{MM^H}$ であるため，$\mathrm{MM^H}$ はエルミート行列．同様に，$\mathrm{M^HM}$ もエルミート行列.

(ii) 任意のエルミート行列 $\mathrm{A} \in \mathbb{C}^{n\times n}$ について，すべての固有値が実数となることを証明する．定義より，行列 A の固有値 λ と固有ベクトル $\boldsymbol{v}$ は，$\mathrm{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ を満たす．よって，

$$vecsymv^{\mathrm{H}}\mathrm{A}\boldsymbol{v} = vecsymv^{\mathrm{H}}\lambda\boldsymbol{v} = \lambda vecsymv^{\mathrm{H}}\boldsymbol{v}. \tag{30}$$

また，

$$vecsymv^{\mathrm{H}}\mathrm{A^H}\boldsymbol{v} = (\mathrm{A}vecsymv)^{\mathrm{H}}\boldsymbol{v} = (\lambda vecsymv)^{\mathrm{H}}\boldsymbol{v} = \overline{\lambda}vecsymv^{\mathrm{H}}\boldsymbol{v}. \tag{31}$$

したがって，$\lambda = \overline{\lambda}$ でなければならないため，λ は実数.

(iii) 任意のエルミート行列 $\mathrm{A} \in \mathbb{C}^{n\times n}$ について，A があるユニタリ行列 U で対角化されることを証明する．

問 7　特異値分解

(i) $\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}}$ より，

$$\mathrm{A}\mathrm{A}^{\mathrm{H}} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}}(\mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}})^{\mathrm{H}} \tag{32}$$

$$= \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}}\mathrm{V}\Sigma^{\mathrm{H}}\mathrm{U}^{\mathrm{H}}. \tag{33}$$

ここで，V はユニタリ行列だから，$\mathrm{V}^{\mathrm{H}}\mathrm{V} = \mathrm{I}$ (I は単位行列)．よって，

$$\mathrm{A}\mathrm{A}^{\mathrm{H}} = \mathrm{U}\Sigma\Sigma^{\mathrm{H}}\mathrm{U}^{\mathrm{H}}. \tag{34}$$

$\Sigma\Sigma^{\mathrm{H}}$ は対角行列となるため，$\Sigma\Sigma^{\mathrm{H}} = \Lambda$ とおけば，

$$\mathrm{A}\mathrm{A}^{\mathrm{H}} = \mathrm{U}\Lambda\mathrm{U}^{\mathrm{H}} \tag{35}$$

であり，これは，$\mathrm{A}\mathrm{A}^{\mathrm{H}}$ の固有値分解である．よって U は，$\mathrm{A}\mathrm{A}^{\mathrm{H}}$ の（ノルムが 1 に正規化された）固有ベクトルを並べた行列として与えられる．

(ii) 同様に，$\Sigma^{\mathrm{H}}\Sigma = \Lambda'$ とおけば，

$$\mathrm{A}^{\mathrm{H}}\mathrm{A} = \mathrm{V}\Lambda'\mathrm{V}^{\mathrm{H}} \tag{36}$$

であり，これは，$\mathrm{A}^{\mathrm{H}}\mathrm{A}$ の固有値分解である．よって V は，$\mathrm{A}^{\mathrm{H}}\mathrm{A}$ の（ノルムが 1 に正規化された）固有ベクトルを並べた行列として与えられる．

(iii) $\Sigma\Sigma^{\mathrm{H}} = \Lambda$ であり，Λ の (i, i) 要素は Σ の (i, i) 要素の自乗である．また，Λ は $\mathrm{A}\mathrm{A}^{\mathrm{H}}$ の固有値を対角に並べた行列である．したがって，Σ は，$\mathrm{A}\mathrm{A}^{\mathrm{H}}$ の固有値の平方根を対角並べた行列となる．

(iv) A が縦長行列であるため，$\mathrm{A}\mathrm{A}^{\mathrm{H}}$ の固有値分解より $\mathrm{A}^{\mathrm{H}}\mathrm{A}$ の固有値分解のほうが簡単である．$\mathrm{A}^{\mathrm{H}}\mathrm{A}$ の固有値分解のほうが簡単である．

$$\mathrm{A}^{\mathrm{H}}\mathrm{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \tag{37}$$

$$= \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \tag{38}$$

より，固有値は $2, 3$ であり，対応する固有ベクトルはそれぞれ $(a, 0)^{\top}, (0, b)^{\top}$ である（ただし，a, b は任意のスカラー）．よって，$a = b = 1$ とすれば，

$$\mathrm{V} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{pmatrix} \tag{39}$$

を得る．また，$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}}$ より，$\mathrm{U} = \mathrm{A}\mathrm{V}^{\mathrm{H}}$

問 8　Moore-Penrose の擬似逆行列

方程式 $\mathrm{A}\boldsymbol{x}=\boldsymbol{b}$ は，A が正則行列であれば，その逆行列を用いて $\boldsymbol{x}=\mathrm{A}^{-1}\boldsymbol{b}$ として解くことができる．任意の行列 A については，その擬似逆行列を用いて $\boldsymbol{x}=\mathrm{A}^{\dagger}\boldsymbol{b}$ として一つの（近似）解を得ることができる．その解は，

- 方程式 $\mathrm{A}\boldsymbol{x}=\boldsymbol{b}$ が決定系のときには，その解 $\boldsymbol{x}$
- 方程式 $\mathrm{A}\boldsymbol{x}=\boldsymbol{b}$ が優決定系のときには，$\|\mathrm{A}\boldsymbol{x}-\boldsymbol{b}\|$ が最小となる $\boldsymbol{x}$
- 方程式 $\mathrm{A}\boldsymbol{x}=\boldsymbol{b}$ が劣決定系のときには，$\mathrm{A}\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{0}$ の解 $\boldsymbol{x}$ のうち，$\|\boldsymbol{x}\|$ が最小となる $\boldsymbol{x}$

に一致する．

(i)

$$\mathrm{A}^{\dagger}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3}\end{pmatrix} \tag{40}$$

$$\mathrm{A}^{\dagger}\boldsymbol{b}=\begin{pmatrix}1\\ \frac{2}{3}\end{pmatrix} \tag{41}$$

$$\tag{42}$$

よって，問 1(ii) で計算した $\boldsymbol{x}$ と等しい．

(ii)

$$\mathrm{A}^{\dagger}=\frac{1}{735}\begin{pmatrix}23 & -26 & 143 & 9\\ 97 & 146 & -68 & 6\\ -43 & -239 & 212 & 111\\ -8 & 41 & 142 & -99\end{pmatrix} \tag{43}$$

$$\mathrm{A}^{\dagger}\boldsymbol{b}=\frac{1}{7}\begin{pmatrix}-3\\ 5\\ -9\\ -2\end{pmatrix} \tag{44}$$

$$\tag{45}$$

よって，問 1(iii) で計算した $\boldsymbol{x}$ と等しい．

問 9　離散畳み込みと DFT

(i)

$$\mathrm{L}=\begin{pmatrix}a_0 & 0 & 0 & 0\\ a_1 & a_0 & 0 & 0\\ a_2 & a_1 & a_0 & 0\\ a_3 & a_2 & a_1 & a_0\end{pmatrix}. \tag{46}$$

線形畳み込みを行列表現するとテプリッツ行列となる．

(ii)

$$\mathrm{C}=\begin{pmatrix}a_0 & a_3 & a_2 & a_1\\ a_1 & a_0 & a_3 & a_2\\ a_2 & a_1 & a_0 & a_3\\ a_3 & a_2 & a_1 & a_0\end{pmatrix}. \tag{47}$$

巡回畳み込みを行列表現すると巡回行列となる．

(iii)

$$\begin{aligned}\mathrm{FCF}^{-1}&=\frac{1}{4}\mathrm{FCF}^{\mathrm{H}} &(48)\\ &=\frac{1}{4}\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & j & -1 & -j\\ 1 & -1 & 1 & -1\\ 1 & -j & -1 & j\end{pmatrix}\begin{pmatrix}a_0 & a_3 & a_2 & a_1\\ a_1 & a_0 & a_3 & a_2\\ a_2 & a_1 & a_0 & a_3\\ a_3 & a_2 & a_1 & a_0\end{pmatrix}\cdot\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & -j & -1 & j\\ 1 & -1 & 1 & -1\\ 1 & j & -1 & -j\end{pmatrix} &(49)\\ &=\begin{pmatrix}a_0+a_1+a_2+a_3 & 0 & 0 & 0\\ 0 & a_0+a_1j-a_2-a_3j & 0 & 0\\ 0 & 0 & a_0-a_1+a_2-a_3 & 0\\ 0 & 0 & 0 & a_0-a_1j-a_2+a_3j\end{pmatrix} &(50)\end{aligned}$$

より対角行列となる．ちなみに，F は離散フーリエ変換を施す DFT 行列である．任意の巡回行列は，その要素に関わらず DFT 行列により対角化される．すなわち，巡回行列は，DFT 行列の列ベクトルを固有ベクトルとして持つ．

(iv)

$$\boldsymbol{Y}=\mathrm{F}\boldsymbol{y}=\mathrm{FC}\boldsymbol{x}=\mathrm{FCF}^{-1}\mathrm{F}\boldsymbol{x}=\mathrm{FCF}^{-1}\boldsymbol{X} \tag{51}$$

より，$\boldsymbol{Y}=\mathrm{FCF}^{-1}\boldsymbol{X}$

(v) (iii) より，FCF^{-1} は対角行列である．この行列の対角要素を D_n とおく，すなわち，

$$D_1=a_0+a_1+a_2+a_3 \tag{52}$$
$$D_2=a_0+a_1j-a_2-a_3j \tag{53}$$
$$D_3=a_0-a_1+a_2-a_3 \tag{54}$$
$$D_4=a_0-a_1j-a_2+a_3j \tag{55}$$

とすれば，(iv) より，$Y_n=D_nX_n$ が成り立つ．このことは，時間領域の畳み込みが周波数領域の掛け算で表されることの線形代数的な理解を与える．ちなみに，$\boldsymbol{D}=(D_1,D_2,D_3,D_4),\boldsymbol{d}=(a_0,a_1,a_2,a_3)$ とおけば，$\boldsymbol{D}=\mathrm{F}\boldsymbol{d}$ として与えられる．

問 10　深層学習？

$f_n(\boldsymbol{x}) = \boldsymbol{x}$ より，$f(\boldsymbol{x})$ は

$$f(\boldsymbol{x}) = \mathrm{W}_k(\mathrm{W}_{k-1}(\cdots\mathrm{W}_2(\mathrm{W}_1\boldsymbol{x} + \boldsymbol{b}_1) + \boldsymbol{b}_2)\cdots) + \boldsymbol{b}_{k-1}) + \boldsymbol{b}_k \tag{56}$$

となる．かっこを展開して整理すると

$$f(\boldsymbol{x}) = \mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_2\mathrm{W}_1\boldsymbol{x} + \mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_2\boldsymbol{b}_1 + \mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_3\boldsymbol{b}_2 + \cdots + \mathrm{W}_k\boldsymbol{b}_{k-1} + \boldsymbol{b}_k \tag{57}$$

が得られる．ここで, $\mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_2\mathrm{W}_1$ および $\mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_2\boldsymbol{b}_1 + \mathrm{W}_k\mathrm{W}_{k-1}\cdots\mathrm{W}_3\boldsymbol{b}_2 + \cdots + \mathrm{W}_k\boldsymbol{b}_{k-1} + \boldsymbol{b}_k$ をそれぞれ，新たに W，$\boldsymbol{b}$ とおくと，$f(\boldsymbol{x}) = \mathrm{W}\boldsymbol{x} + \boldsymbol{b}$ となる．これは，$f_1(\boldsymbol{x}) = \boldsymbol{x}$ である 1 層のニューラルネットワークである．すなわち，$f_n(\boldsymbol{x}) = \boldsymbol{x}$ のとき，k 層のニューラルネットワークと等価な 1 層のニューラルネットワークが存在する．

2　微積分

問 1　最小二乗法の閉形式解

(i)

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}) &= (\mathrm{A}\boldsymbol{x}-\boldsymbol{b})^\top(\mathrm{A}\boldsymbol{x}-\boldsymbol{b}) &(58)\\ &= (\boldsymbol{x}^\top\mathrm{A}^\top-\boldsymbol{b}^\top)(\mathrm{A}\boldsymbol{x}-\boldsymbol{b}) &(59)\\ &= \boldsymbol{x}^\top\mathrm{A}^\top\mathrm{A}\boldsymbol{x}-\boldsymbol{x}^\top\mathrm{A}^\top\boldsymbol{b}-\boldsymbol{b}^\top\mathrm{A}\boldsymbol{x}-\boldsymbol{b}^\top\boldsymbol{b} &(60)\\ &= \boldsymbol{x}^\top\mathrm{A}^\top\mathrm{A}\boldsymbol{x}-2\boldsymbol{x}^\top\mathrm{A}^\top\boldsymbol{b}-\boldsymbol{b}^\top\boldsymbol{b} &(61)\end{aligned}$$

より，

$$\frac{\partial}{\partial\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}) = 2\mathrm{A}^\top\mathrm{A}\boldsymbol{x}-2\mathrm{A}^\top\boldsymbol{b} \tag{62}$$

(ii) $\mathcal{L}(\boldsymbol{x})$ を最小にする $\boldsymbol{x}$ は,

$$\frac{\partial}{\partial\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}) = \boldsymbol{0} \tag{63}$$

を満たす．したがって，$\boldsymbol{x}$ が満たすべき条件は,

$$\mathrm{A}^\top\mathrm{A}\boldsymbol{x} = \mathrm{A}^\top\boldsymbol{b} \tag{64}$$

（正規方程式）

(iii) $\mathrm{A}^\top\mathrm{A}$ が正則であることから,

$$\boldsymbol{x} = (\mathrm{A}^\top\mathrm{A})^{-1}\mathrm{A}^\top\boldsymbol{b} \tag{65}$$

問 2　最急降下法

最急降下法のプログラムを以下に示す.

Listing 1: Gradient descent

```
import numpy as np


def func1(x):
  return (1/6)*(x**6) - (3/5)*(x**5) - (x**4) + 4*(x**3)

def diff1(x):
  return (x**5) - 3*(x**4) - 4*(x**3) + 12*(x**2)

def gradient_descent(func, diff, x_init, learning_rate, max_iter):
  x = x_init
  for i in range(max_iter):
    x = x - alpha * diff(x)
  return x

# (1)
x_init = 1/2
alpha = 0.2
x = gradient_descent(func=func1, diff=diff1, x_init=x_init, learning_rate=alpha, max_iter=20)
print("f"x:{x}, f:{func1(x)}, f\':{diff1(x)}")
```

(i) x:0.012574881004397218, f:7.928552292318071e-06, f':0.0018895031440419867

(ii) x:3.0026363348708207, f:2.7001568945390773, f':0.11921980789784925

(iii) x:-1.998144027244153, f:-18.133195794962337, f':0.14807862625745827

(iv) x:3.0, f:2.700000000000017, f':0.0

(v) 増減表を書くと, $x = 3$ のとき極小値, $x = 2$ のとき極大値, $x = 0$ のとき変曲点, $x = -2$ のとき最小値をとることがわかる. 最急降下法を用いた場合には, 初期値 x_0 とパラメータ α の値によって得られる解が変化する. α が十分小さい時には, 得られる解は, 傾きが 0 である点のうち初期値 x_0 から勾配の負の方向で最も近いものになる. ただし, α が小さい時には収束が遅くなる上, 変曲点の近くでトラップされてしまうことがある. 一方, α を大きくした場合には, (iv) のように, 最小値をとる点から近い点を初期値として与えた場合にも, 別の点へ収束してしまうことがある. さらに, α が大きい場合には, 解が発散してしまうこともある (α をより大きくして実験するとわかる).