『基礎演習 線形代数』補充問題[1]). なお本書でも用いているように，この演習書に対応する教科書『線形代数講義』を『教科書』の略称で引用します．

p.15 への追加問題

問 2.1.6 2 次の正方行列 A で $A^2 = O$) となるようなもの (冪零行列) の一般形を決定せよ．また，このような二つの行列が積について可換となるための条件を求めよ．

解答 $A = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$ と置けば，

$$A^2 = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{pmatrix} x_{11}^2 + x_{12}x_{21} & x_{11}x_{12} + x_{12}x_{22} \\ x_{21}x_{11} + x_{22}x_{12} & x_{21}x_{12} + x_{22}^2 \end{pmatrix} = O$$

から，条件として

$$x_{11}^2 + x_{12}x_{21} = 0 \ldots ①, \qquad x_{11}x_{12} + x_{12}x_{22} = x_{12}(x_{11} + x_{22}) = 0 \ldots ②,$$
$$x_{11}x_{21} + x_{21}x_{22} = x_{21}(x_{11} + x_{22}) = 0 \ldots ③, \qquad x_{12}x_{21} + x_{22}^2 = 0 \ldots ④.$$

が得られる．②, ③ から $x_{11} + x_{22} = 0 \ldots$ ⑤a か $x_{12} = x_{21} = 0$ ⑤b かのいずれかが成り立つ．他方，①, ④ から $x_{11}^2 = x_{22}^2$ ⑥ である．よって前者 ⑤a の場合はこれ以上条件は出てこず，$x_{12} = a, x_{21} = -b$ とパラメータを導入すれば，$x_{11} = \sqrt{ab}, x_{22} = -\sqrt{ab}$ あるいはその逆で，

$$A = \begin{pmatrix} \pm\sqrt{ab} & -b \\ a & \mp\sqrt{ab} \end{pmatrix} \tag{H.1}$$

という，2 パラメータの解が得られる．(実の行列となるためには，a, b は同符号でなければならない．) 後者からは零行列が得られ，前者の $a = b = 0$ という特別な場合に含まれる．なお，$a = 0$, あるいは $b = 0$ のどちらかだけを仮定すると，(H.1) はそれぞれ

$$\begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix}, \qquad \text{あるいは} \qquad \begin{pmatrix} 0 & -b \\ 0 & 0 \end{pmatrix} \tag{H.2}$$

となり，よく知られた冪零行列の形となる．

今，

$$A = \begin{pmatrix} \sqrt{ab} & -b \\ a & -\sqrt{ab} \end{pmatrix}, \qquad B = \begin{pmatrix} \sqrt{a'b'} & -b' \\ a' & -\sqrt{a'b'} \end{pmatrix}$$

を二つの冪零行列とせよ．(1,1) 成分の符号を正に決め打ちしても可換性の探求で一般性を失わないことは，もしこれが負なら全体に -1 を掛け, パラメータ a, b の符号を同時に反転させれば，上の形にできることから分かる．すると

$$AB = \begin{pmatrix} \sqrt{aba'b'} - a'b & b'\sqrt{ab} + b\sqrt{a'b'} \\ a\sqrt{a'b'} - a'\sqrt{ab} & -ab' + \sqrt{aba'b'} \end{pmatrix}$$

もし $AB = BA$ なら，$(AB)^2 = ABAB = AABB = OO = O$ より，AB も冪零，従って上で求めた形をしていなければならず，特に対角成分の和は 0 でなければならない．すなわち，

$$0 = \sqrt{aba'b'} - a'b - ab' + \sqrt{aba'b'} = 2\sqrt{aba'b'} - a'b - ab', \quad \therefore \quad 2\sqrt{aba'b'} = a'b + ab'.$$

最後の式の両辺を 2 乗すると

$$4aba'b' = a'^2b^2 + 2aba'b' + a^2b'^2 = 0, \quad \therefore \quad 0 = a'^2b^2 - 2aba'b' + a^2b'^2 = (a'b - ab')^2.$$

[1] この文書はサイエンス社から出ているこの書名の本 (以下『本書』と略称) に対する著者からの読者サービスです．ここにはスペースの制約で割愛した問題や出版後に気づいた “含めておいた方が良かった類題” などを集めて解答とともに掲げておきます．この文書には出版社からの制約は何もありませんが，著作権は出版されている本書の延長という扱いでお願いします (©金子 晃

従って $a'b = ab'$ が得られる．これは，もし $a \neq 0$ なら，$b = b' = 0$ か，あるいは $b \neq 0$ で $\dfrac{a'}{a} = \dfrac{b'}{b}$, すなわちある定数 k が存在して $a' = ka, b' = kb$ となっていることを意味する．前者の場合は A も B も $\begin{pmatrix} 0 & 0 \\ a & 0 \end{pmatrix}$ 型の冪零行列となり，これらの積は零行列となるので可換性は明らかに成り立つ．後者の場合は, 平方する前の条件式から $2\sqrt{k^2abab} = 2kab$, 従って $k \geq 0$ なので $\sqrt{k^2ab} = k\sqrt{ab}$ とでき

$$B = \begin{pmatrix} k\sqrt{ab} & -kb \\ ka & -k\sqrt{ab} \end{pmatrix} = k\begin{pmatrix} \sqrt{ab} & -b \\ a & -\sqrt{ab} \end{pmatrix} = kA$$

となる．よって積の可換性は自明である．

なお，最初の仮定で除外した $a = 0$ の場合は，$b \neq 0$ なら $a' = 0$ となって A も B も (H.2) の後者の型となる．この場合は積が零行列となり明らかに可換だが，これらの形の行列同士は, 一方が他方のスカラー倍になっている．B が零行列のときは A が何で有っても積は可換だが，これは A のスカラー 0 倍とみなせる．よって以上をまとめると，二つの 2 次冪零行列が可換なのは，一方が他方のスカラー倍のとき，かつそのときに限られるということができる．

p.23 への追加問題

問 2.3.8 任意の $m \times n$ 型行列 A について，次を示せ：

(1) n 次元数ベクトル $\boldsymbol{x}$ について $A\boldsymbol{x} = \mathbf{0}_m \iff A^TA\boldsymbol{x} = \mathbf{0}_n$. (零ベクトルの次元を添え字で表している.)

(2) $\operatorname{rank} A^TA = \operatorname{rank} A$.

解答 (1) $\Longrightarrow$ の向きは自明．逆に，$A^TA\boldsymbol{x} = \mathbf{0}_n$ とすると，$\boldsymbol{x}^TA^TA\boldsymbol{x} = 0$, すなわち $(A\boldsymbol{x})^TA\boldsymbol{x} = 0$. この左辺はトレースの定義により数ベクトル $A\boldsymbol{x}$ の成分の 2 乗和なので, これより $A\boldsymbol{x}$ の成分はすべて 0, すなわち $A\boldsymbol{x} = \mathbf{0}_m$ となる． □

(2) $r = \operatorname{rank} A$ と置く．一般公式 $\operatorname{rank} AB \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}$ (例題 2.3-2 (1)) より $\operatorname{rank} A^TA \leq r$ は明らか. 階数の定義により $A\boldsymbol{x}_1, \ldots, A\boldsymbol{x}_r$ という形の 1 次独立なベクトルが存在するが, このとき $A^TA\boldsymbol{x}_1, \ldots, A^TA\boldsymbol{x}_r$ も 1 次独立となる．実際，もし $\sum_{i=1}^r c_iA^TA\boldsymbol{x}_i = \mathbf{0}_m$ となる非自明な係数 $(c_1, \ldots, c_r)$ が存在したとすると，$A^TA\sum_{i=1}^r c_i\boldsymbol{x}_i = \mathbf{0}_n$. 従って (1) により $A\sum_{i=1}^r c_i\boldsymbol{x}_i = \mathbf{0}_m$, すなわち $\sum_{i=1}^r c_iA\boldsymbol{x}_i = \mathbf{0}_m$ となるが, これは仮定に反する．よって $\operatorname{rank} A^TA \geq r$ も示された． □

これらの主張は行列 A の特異値分解 (例えば本書 §8.5) の知識からは自明となる．

p.160 への追加問題

問 7.1.5 n 次実正方行列の空間において対称行列 A と歪対称行列 B は例題 7.1-1 (1) の内積で直交していること，従って問題 3.3.5 に与えた直和分解は直交分解であることを示せ.

解答 トレースは転置しても，また積の順序を入れ替えても変わらないので

$$(A, B) := \operatorname{tr}(A^TB) = \operatorname{tr}(AB) = \operatorname{tr}((AB)^T) = \operatorname{tr}(B^TA^T) = \operatorname{tr}(-BA) = -\operatorname{tr}(BA) = -\operatorname{tr}(AB) = -(A, B)$$

よって $(A, B) = 0$ となる． □

p.176 への追加問題

問 7.3.10 n 次正方行列 A の各成分が t の関数で，A^{-1} が存在するとき，$\dfrac{d}{dt}A^{-1}$ を計算せよ．

解答 合成関数の微分公式より $\dfrac{dA}{dt}$ と $\dfrac{d}{dA}A^{-1}$ で答が書けるはずであるが，非可換性のため，単にこれらの積という訳にはいかない．間違えないためには，最初から微分の定義に従って計算するのが良い．すなわち，A の各成分に含まれる t を $t + \delta t$ で置き換えたものの逆行列 $A(t + \delta t)^{-1}$ の δt に関する 1 次の項の係数を取り出せばよい．$A(t)$ の各成分を t で微分したものを $A'(t)$ と略記すると，$A(t + \delta t) = A(t) + A'(t)\delta t + o(\delta t^2)$. よって例題 7.3-4 の計算を利用すると，結局同例題の結果 $-A^{-1}\delta AA^{-1}$ において $\delta A = A'(t)\delta t$ と置いたものが得られる．故にスカラー δt を外に出せば，その係数として $-A(t)^{-1}A'(t)A(t)^{-1}$ が得られ，これが求める微分である． □

p.181 への追加問題

問 7.4.7 n 次正方行列 A の各成分が t の関数であるとき，次のような t に関する導関数を計算せよ．

(1) $\dfrac{d}{dt}|A|$,　　　　　　　(2) $\dfrac{d}{dt}\log|A|$.

解答 (1) これはスカラー関数のスカラーによる微分なので，普通の線形代数の教科書にもよく載っている．なのでまずそのような初等的計算で解いてみよう．$A = A(t) = (\boldsymbol{a}_1(t), \dots, \boldsymbol{a}_n(t))$ を A の列による表示とすれば，

$$\boldsymbol{a}_j(t+\delta t) = \boldsymbol{a}_j(t) + \boldsymbol{a}'_j(t)\delta t + o(\delta t)$$

となる．ここに $\boldsymbol{a}'_j(t)$ は各成分を t で微分して得られるベクトルである．行列式の列に関する多重線形性により

$$\det(\boldsymbol{a}_1(t+\delta t), \dots, \boldsymbol{a}_n(t+\delta t)) = \det(\boldsymbol{a}_1(t), \dots, \boldsymbol{a}_n(t)) + \sum_{j=1}^{n} \det(\boldsymbol{a}_1(t), \dots, \overset{j}{\boldsymbol{a}'_j}(t), \dots, \boldsymbol{a}_n(t))\delta t + o(\delta t)$$

となるので，求める微分 $\dfrac{d}{dt}|A|$ はこの δt の係数で $\sum_{j=1}^{n} \det(\boldsymbol{a}_1(t), \dots, \overset{j}{\boldsymbol{a}'_j}(t), \dots, \boldsymbol{a}_n(t))$ となる．行に関する多重線形性を用いれば，行に関する同様の表現が得られる．なお普通の線形代数の教科書に載っている演習問題としての解答は，$\det A$ の展開式 $\sum_{j_1,\dots,j_n} \mathrm{sgn}\,(j_1,\dots,j_n) a_{1j_1} \cdots a_{nj_n}$ の各項を積の微分で展開したものを上のような行列式の和にまとめ直すというものである．

別解 例題 7.4-4 で与えられている $\dfrac{d}{dA}|A| = \widetilde{A}^T$ という公式を利用した解法を示す．この公式を導くのはそう簡単ではなかったので，これは両者の関係を明らかにするための解法である．この公式の意味は，この微分が

$$\delta A \mapsto \mathrm{tr}\{(\frac{d}{dt}|A|)^T \delta A\} = \mathrm{tr}(\widetilde{A}\delta A)$$

という線形写像になるということであった．よってこれに $\delta A = A'(t)\delta t$ を代入すると，

$$= \mathrm{tr}(\widetilde{A}A'(t)\delta t) = \mathrm{tr}(\widetilde{A}A'(t))\delta t$$

よって微分の答はこの δt の係数であり，これは更に行列の積を展開することにより

$$= \sum_{j=1}^{n} (\widetilde{a_{j1}}, \dots, \widetilde{a_{jn}}) \begin{pmatrix} a'_{j1}(t) \\ \vdots \\ a'_{jn}(t) \end{pmatrix} = \sum_{j=1}^{n} \det(\boldsymbol{a}_1, \dots, \overset{j}{\boldsymbol{a}'_j}(t), \dots, \boldsymbol{a}_n)$$

と書き直されて，最初の結果と一致する．　□

(2) 同様に，

$$\begin{aligned}\log|A(t+\delta t)| &= \log|A(t) + A'(t)\delta t + o(\delta t)| = \log|A(t)(E + A(t)^{-1}A'(t)\delta t + o(\delta t)| \\ &= \log|A(t)| + \log|E + A(t)^{-1}A'(t)\delta t + o(\delta t)|\end{aligned}$$

ここで，

$$|E + A(t)^{-1}A'(t)\delta t + o(\delta t)| = |\exp(A(t)^{-1}A'(t)\delta t) + o(\delta t)| = \exp\mathrm{tr}(A(t)^{-1}A'(t)\delta t) + o(\delta t)$$

となることを用いると，δt の係数は $\mathrm{tr}(A(t)^{-1}A'(t))$ となり，これが微分の答である．

別解 (1) の答を利用すると

$$\frac{d}{dt}\log|A| = \frac{1}{|A|}\frac{d}{dt}|A| = \frac{1}{|A|}\mathrm{tr}(\widetilde{A}A'(t)) = \mathrm{tr}(\frac{1}{|A|}\widetilde{A}A'(t)) = \mathrm{tr}(A^{-1}A'(t)).$$

ここで最後の等号は $\widetilde{A}A = |A|E$ から導かれる逆行列の余因子行列による表現公式による．　□

p.194 への追加問題
問 8.1.4 (重回帰分析) $\boldsymbol{y}$ を実 m 次元ベクトル，A を $m\times n$ 型の実行列とするとき，

$$f(\boldsymbol{x}) = \|\boldsymbol{y}-A\boldsymbol{x}\|^2 = (\boldsymbol{y}-A\boldsymbol{x})^T(\boldsymbol{y}-A\boldsymbol{x})$$

の値を最小にするような n 次元実ベクトル $\boldsymbol{x}$ を特定せよ．ただし
(1) A^TA が正則のとき
(2) A^TA が正則でないとき
のそれぞれに分けて論ぜよ.

解答　(1)

$$f(\boldsymbol{x}) = \boldsymbol{x}^TA^TA\boldsymbol{x} - \boldsymbol{y}^TA\boldsymbol{x} - \boldsymbol{x}^TA^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y}$$

において，スカラーは転置しても変わらないので，右辺の第 2 項と第 3 項は一致し，まとめて $-2\boldsymbol{y}^TA\boldsymbol{x}$ となる．問 6.2.3 によれば，半正定値対称行列 A^TA に対して $B^2 = A^TA$ を満たすような半正定値対称行列 B が取れる. (実際, A^TA を対角化する直交行列 P を取り，$P^TA^TAP = \Lambda$ をその対角化とするとき，Λ の対角成分の非負平方根を取ったものを $\sqrt{\Lambda}$ とすれば, $B = P\sqrt{\Lambda}P^T$ がその答であった.）今，A^TA, 従って B が正則, すなわち 0 を固有値として持たなければ, 独立変数を $\boldsymbol{z} = B\boldsymbol{x}$, すなわち $\boldsymbol{x} = B^{-1}\boldsymbol{z}$ に変換すると

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{z}^T\boldsymbol{z} - 2\boldsymbol{y}^TAB^{-1}\boldsymbol{z} + \boldsymbol{y}^T\boldsymbol{y} \\ &= (\boldsymbol{z} - B^{-1}A^T\boldsymbol{y})^T(\boldsymbol{z} - B^{-1}A^T\boldsymbol{y}) - \boldsymbol{y}^TAB^{-1}B^{-1}A^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y} \\ &= (\boldsymbol{z} - B^{-1}A^T\boldsymbol{y})^T(\boldsymbol{z} - B^{-1}A^T\boldsymbol{y}) - \boldsymbol{y}^TA(A^TA)^{-1}A^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y} \end{aligned}$$

よって $f(\boldsymbol{x})$ は $\boldsymbol{z} = B^{-1}A^T\boldsymbol{y}$, すなわち $\boldsymbol{x} = B^{-2}A^T\boldsymbol{y} = (A^TA)^{-1}A^T\boldsymbol{y}$ のときに最小値 $-\boldsymbol{y}^TA(A^TA)^{-1}A^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{y}^T\{I - A(A^TA)^{-1}A^T\}\boldsymbol{y}$ を取る.

別解　$\boldsymbol{z} = \boldsymbol{y} - A\boldsymbol{x}$ と置けば, 問題 7.4.1 の簡単な場合として

$$\frac{\partial \boldsymbol{y}^T\boldsymbol{y}}{\partial \boldsymbol{y}} = 2\boldsymbol{y} \quad \text{(工学的に縦ベクトルに直した表現)}$$

よって $\dfrac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{y} - A\boldsymbol{x}) = -A^T$ (工学的に転置をとったもの) と合わせて合成関数の微分公式から

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = -A^T\{2(\boldsymbol{y} - A\boldsymbol{x})\} = 2A^TA\boldsymbol{x} - 2A^T\boldsymbol{y} \tag{H.3}$$

となる. (同じ結果は最初の解答で 2 次式を展開したものを項別に微分しても得られる.) よって $\boldsymbol{x} = (A^TA)^{-1}A^T\boldsymbol{y}$ で勾配ベクトルが零となる．これは極値の必要条件であるが, 与えられた問題は最小値のみを持つことが形から明らかなので, これが求める最小値を取る点と確定する．最小値はこれを元の関数に代入してみれば先の解答と同じ値が得られる．□

(2) A^TA が零固有値を持つときは, $\boldsymbol{z} = B\boldsymbol{x}$ は変数変換にならないので, 初等的にやるためには A^TA の平方根を求めた計算まで立ち戻るのが分かりやすい．そこで, $P^TA^TAP = \Lambda$ を再びその対角化とし，Λ の下 $n-r$ 個に零固有値が並んでいるものとすれば, まずまず $\boldsymbol{x} = P\boldsymbol{z}$, すなわち $\boldsymbol{z} = P^T\boldsymbol{x}$ と変換して

$$f(\boldsymbol{x}) = \boldsymbol{z}^TP^TA^TAP\boldsymbol{z} - 2\boldsymbol{y}^TAP\boldsymbol{z} + \boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{z}^T\Lambda\boldsymbol{z} - 2\boldsymbol{y}^TAP\boldsymbol{z} + \boldsymbol{y}^T\boldsymbol{y}.$$

このとき最後の辺の第 1 項は $z_1,\dots,z_r$ の正値 2 次形式 $\lambda_1z_1^2 + \cdots + \lambda_rz_r^2$ となっているが, 第 2 項の $\boldsymbol{z}$ の 1 次式はやはり $z_1,\dots,z_r$ のみを含む．実際, もし $z_{r+1},\dots,z_n$ のいずれかが含まれていれば, $z_1 = \cdots = z_r = 0$ と置いたときにこれら残った変数を適当に動かせば, 第 2 項は負のいくらでも小さい値を取り得ることになって, 元の関数が半正定値であったことに矛盾する．よって, 今 $\boldsymbol{z}' = (z_1,\dots,z_r)^T$ と置き, 残りの座標に対応する $n-r$ 次の零ベクトルを $\boldsymbol{0}''$ で表し，Λ の左上の正則な r 次の小行列を Λ' と書けば, 与えられた関数は

$$\boldsymbol{z}'^T\Lambda'\boldsymbol{z}' - 2\boldsymbol{y}^TAP\begin{pmatrix}\boldsymbol{z}' \\ \boldsymbol{0}''\end{pmatrix} + \boldsymbol{y}^T\boldsymbol{y}$$

と書け, これが最小となるときに最小値を取る．このとき $z_{r+1},\dots,z_n$ は何でもよい．すなわち最小値はある線形部分多様体の上で取られる．この部分集合を具体的に表すため，更に $\boldsymbol{w}'=\sqrt{\Lambda'}\boldsymbol{z}'$ と変換すれば，上は

$$
\begin{aligned}
&=\boldsymbol{w}'^T\boldsymbol{w}'-2\boldsymbol{y}^TAP\begin{pmatrix}\sqrt{\Lambda'}^{-1}\\ O_{21}\end{pmatrix}\boldsymbol{w}'+\boldsymbol{y}^T\boldsymbol{y}\\
&=\Big(\boldsymbol{w}'-(\sqrt{\Lambda'}^{-1},O_{12})P^TA^T\boldsymbol{y}\Big)^T\Big(\boldsymbol{w}'-(\sqrt{\Lambda'}^{-1},O_{12})P^TA^T\boldsymbol{y}\Big)-\boldsymbol{y}^TAP\begin{pmatrix}\sqrt{\Lambda'}^{-1}\\ O_{21}\end{pmatrix}(\sqrt{\Lambda'}^{-1},O_{12})P^TA^T\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{y}\\
&=\Big\|\boldsymbol{w}'-(\sqrt{\Lambda'}^{-1},O_{12})P^TA^T\boldsymbol{y}\Big\|^2-\boldsymbol{y}^TAP\begin{pmatrix}\Lambda'^{-1}&O_{12}\\ O_{21}&O''\end{pmatrix}P^TA^T\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{y}
\end{aligned}
$$

となる．ここで O_{12},O_{21},O'' はそれぞれ適当なサイズの零行列とした．よって最小値は $\boldsymbol{w}'=(\sqrt{\Lambda'}^{-1},O''^T)P^TA^T\boldsymbol{y}$, 従って $\boldsymbol{z}'=(\Lambda'^{-1},O''^T)P^TA^T\boldsymbol{y}$ のときであるが, これは $\begin{pmatrix}\boldsymbol{z}'\\ \boldsymbol{0}''\end{pmatrix}=\begin{pmatrix}\Lambda'^{-1}&O_{12}\\ O_{21}&O''\end{pmatrix}P^TA^T\boldsymbol{y}$ と同値であり，この両辺に左から $P^TA^TAP=\begin{pmatrix}\Lambda'&O_{12}\\ O_{21}&O''\end{pmatrix}$ を掛けると

$$
P^TA^TAP\boldsymbol{z}=\begin{pmatrix}\Lambda'&O_{12}\\ O_{21}&O''\end{pmatrix}\begin{pmatrix}\boldsymbol{z}'\\ \boldsymbol{z}''\end{pmatrix}=\begin{pmatrix}\Lambda'&O_{12}\\ O_{21}&O''\end{pmatrix}\begin{pmatrix}\boldsymbol{z}'\\ \boldsymbol{0}''\end{pmatrix}=\begin{pmatrix}I'&O_{12}\\ O_{21}&O''\end{pmatrix}P^TA^T\boldsymbol{y}
$$

となる．最左辺は $P^TA^TA\boldsymbol{x}$ に等しい．他方，右辺も階数 r のベクトルであり, $\operatorname{rank}A^TA=r$ ということから $\operatorname{rank}A^T=\operatorname{rank}A=r$ となる (補充問題の問 2.3.8 (2))．すると行列 $\begin{pmatrix}I'&O_{12}\\ O_{21}&O''\end{pmatrix}$ を $\begin{pmatrix}I'&O_{12}\\ O_{21}&I''\end{pmatrix}$ に変えても右辺は変わらない．(実際，これにより追加されるベクトルは非零成分が下 $n-r$ 個にしか存在せず，元の像である非零成分が上 r 個にしか無いベクトルとは非自明な 1 次関係式を持たないので，もし出てくれば像の次元すなわち階数が増えてしまう.) よって，上の関係式は結局 $P^TA^TAP\boldsymbol{z}=P^TA^T\boldsymbol{y}$, すなわち

$$
A^TA\boldsymbol{x}=A^T\boldsymbol{y}
$$

と同値になる．これが最小値を達成する r 次元の線形部分多様体, すなわち，線形部分空間を平行移動した図形となる．

この場合，偏微分を用いる別解は (1) の別解と計算は全く同じになる．違いは (H.3) を 0 と置いて得られた条件式が $\boldsymbol{x}$ について具体的に解けないというだけである． □

通常の回帰分析は 2 次元のデータ $\{(x_i,y_i);i=1,\dots,n\}$ の平面での分布に，最も良く当てはまる直線を引くもので，その答は『教科書』の p.199 の でも紹介したように，

$$
f(m,b):=\sum_{i=1}^n|mx_i+b-y_i|^2
$$

を最小にするような傾き m と y-切片 b を求めるものです．上の問題は重回帰分析と言いながら，この拡張になっているようには見えませんが，この式を

$$
f(m,b)=\left\|\begin{pmatrix}x_1&1\\ x_2&1\\ \vdots&\\ x_n&1\end{pmatrix}\begin{pmatrix}m\\ b\end{pmatrix}-\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}\right\|^2
$$

と書き直してみると，この $n\times 2$ 型行列を与えられた A とする重回帰分析の形になっていることが分かります．このとき

$$
A^TA=\begin{pmatrix}x_1^2+\cdots+x_n^2&x_1+\cdots+x_n\\ x_1+\cdots+x_n&1\end{pmatrix}
$$

となり，従って $\det(A^TA)=x_1^2+\cdots+x_n^2-(x_1+\cdots+x_n)^2$ となりますが，この量は回帰分析で求めた m や b の表現の分母に現れます．この量は一般には $\boldsymbol{R}^n$ の無数の点で 0 となり得るのですが, 普通に回帰分析で

使われるようなデータは $\forall x_i \geq 0$ の場合がほとんどで，そのとき $\det(A^TA) = 0$ となる条件は，適当なスカラー倍の調節で

$$x_1^2 + \cdots + x_n^2 = x_1 + \cdots + x_n = 1$$

の場合に帰着されます．すると $x_i \geq 0$ なら $x_i^2 \leq x_i$ であり，$0 < x_i < 1$ のときは $x_i^2 < x_i$ となるので, 上で最初の等号が成り立つのは $\forall x_i$ が 0 か 1 のどちらかである場合に限られますが，二つ目の等号から一つだけが 1 で他はすべて 0 となります．このような特殊なデータに回帰分析を適用することは有り得ないので，普通は安心して割り算をして m, b を求めている訳です．

問 8.1.5 (リッジ回帰) $\boldsymbol{y}$ を実 m 次元ベクトル，A を $m \times n$ 型の実行列，λ を正の定数とするとき，次の関数を最小にするような n 次元実ベクトル $\boldsymbol{x}$ を求めよ．

$$f(\boldsymbol{x}) = \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|^2.$$

ここで付け加えられた項は正則化項と呼ばれ，通常の重回帰分析で求められる $\boldsymbol{x}$ がデータ $\boldsymbol{y}$ に含まれる異常値のせいで大きくずれてしまうのを抑制する効果を持つ．

解答

$$f(\boldsymbol{x}) = (\boldsymbol{y} - A\boldsymbol{x})^T(\boldsymbol{y} - A\boldsymbol{x}) + \lambda\boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{x}^T(A^TA + \lambda E)\boldsymbol{x} - \boldsymbol{y}^TA\boldsymbol{x} - \boldsymbol{x}^TA^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y}$$

よって前問における $\boldsymbol{x}$ の 2 次の項の係数行列が A^TA から $A^TA + \lambda E$ に置き換わっただけであり，付加項のお陰でこれは常に正定値となるので，正則な平方根 $B = \sqrt{A^TA + \lambda E}$ が取れ，同問の解答の変形, すなわち $\boldsymbol{z} = B\boldsymbol{x}, \boldsymbol{x} = B^{-1}\boldsymbol{z}$ として

$$f(\boldsymbol{x}) = (\boldsymbol{z} - B^{-1}A^T\boldsymbol{y})^T(\boldsymbol{z} - B^{-1}A^T\boldsymbol{y}) - \boldsymbol{y}^TAB^{-2}A^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y}$$

が通用する．よって，f は $\boldsymbol{z} = B^{-1}A^T\boldsymbol{y}$, すなわち $\boldsymbol{x} = B^{-2}A^T\boldsymbol{y} = (A^TA + \lambda E)^{-1}A^T\boldsymbol{y}$ において最小値 $-\boldsymbol{y}^TA(A^TA + \lambda E)^{-1}A^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{y}$ を取る．微分計算によっても全く同じ結論が得られる． □