連続最適化および関連分野に関する夏季学校
リーマン多様体上の最適化理論とその周辺　演習問題の解答例

2021 年 8 月 24 日–8 月 25 日

京都大学　佐藤 寛之

問題 1.

(1) 密着空間において，任意の点列は任意の点に収束することを証明せよ．

(2) ハウスドルフ空間において，収束する点列の極限は一意であることを証明せよ．

(3) 2 個以上の要素をもつ密着空間はハウスドルフ空間でないことを証明せよ．

(4) 非加算無限個の要素をもつ離散空間は第 2 可算空間でないことを証明せよ．

(5) 滑らかな多様体 M, N, Q と滑らかな写像 $F\colon N \to Q, G\colon M \to N$ について，任意の点 $p \in M$ に対して $\mathrm{D}(F \circ G)(p) = \mathrm{D}F(G(p)) \circ \mathrm{D}G(p)$ が成り立つことを証明せよ．

(6) リーマン多様体上の滑らかな実数値関数の勾配は一意に定まることを証明せよ．

解答例

(1) 密着空間 X において，点列 $\{x_n\}$ と点 x を任意にとる．密着位相の定義から，x を含む開集合は X のみである．一方，任意の n に対して $x_n \in X$ であることから，$\{x_n\}$ は x に収束する．

(2) ハウスドルフ空間 X において，点列 $\{x_n\}$ が $x, y \in X$ に収束するとし，$x \neq y$ であると仮定する．ハウスドルフ空間の定義から，$x \in U, y \in V, U \cap V = \emptyset$ なる X の開集合 U, V が存在する．ここで，$\{x_n\}$ は x, y に収束するから，十分大きい自然数 N が存在して，$n \geq N$ ならば $x_n \in U$ かつ $x_n \in V$ となるが，これは $U \cap V = \emptyset$ に矛盾する．よって，$x = y$ である．

(3) 密着空間 X が 2 個以上の要素をもつとき，$x, y \in X$ $(x \neq y)$ が存在する．ところが，密着位相の定義により，x を含む開集合も y を含む開集合も X のみであり，明らかに $X \cap X \neq \emptyset$ であるから，X はハウスドルフ空間でない．

(4) 離散空間 $(X, \mathscr{O})$ の任意の要素 x について，1 点集合 $\{x\}$ は離散位相の定義から X の開集合である．よって，$\mathscr{O}$ の任意の開基 $\mathscr{B}$ は各 $x \in X$ について $\{x\}$ を含み，X が非可算無限個の要素を持つことから，$\mathscr{B}$ は非加算無限個の X の部分集合を含む．したがって $(X, \mathscr{O})$ は第二可算空間でない．

(5) まず，一般に，任意の $\xi \in T_xN$ について，$\gamma(0) = x, \dot{\gamma}(0) = \xi$ となる N 上の曲線 γ と Q 上の曲線 $\gamma_F := F \circ \gamma$ を選ぶと，すべての $g \in \mathfrak{F}_{F(x)}(Q)$ について

$$(\mathrm{D}F(x)[\xi])g = \dot{\gamma}_F(0)g = \left.\frac{d}{dt}(g \circ \gamma_F)(t)\right|_{t=0} = \left.\frac{d}{dt}((g \circ F) \circ \gamma)(t)\right|_{t=0} = \dot{\gamma}(0)(g \circ F) = \xi(g \circ F)$$

となる．これを用いると，任意の $\eta \in T_pM$ に対して，すべての $h \in \mathfrak{F}_{F(G(p))}(Q)$ について

$$(\mathrm{D}(F \circ G)(p)[\eta])h = \eta(h \circ (F \circ G)) = \eta((h \circ F) \circ G) = (\mathrm{D}G(p)[\eta])(h \circ F) = (\mathrm{D}F(G(p))[\mathrm{D}G(p)[\eta]])h$$

が成り立つ．これより $\mathrm{D}(F \circ G)(p)[\eta] = \mathrm{D}F(G(p))[\mathrm{D}G(p)[\eta]]$ となり，さらに $\eta \in T_pM$ は任意だから，$\mathrm{D}(F \circ G)(p) = \mathrm{D}F(G(p)) \circ \mathrm{D}G(p)$ が示せた．

(6) リーマン多様体 (M, g) 上の任意の点 x において，$\xi_1, \xi_2 \in T_xM$ が，$\operatorname{grad} f(x)$ が満たすべき条件を満たすとする．すなわち，すべての $\eta \in T_xM$ に対して $g_x(\xi_1, \eta) = g_x(\xi_2, \eta) = 0$ が成り立つとする．このとき $g_x(\xi_1 - \xi_2, \eta) = 0$ となり，T_xM において $\xi_1 - \xi_2$ は任意の $\eta \in T_xM$ と直交するから，$\xi_1 - \xi_2 = 0$, すなわち $\xi_1 = \xi_2$ が成り立つ．よって，$\operatorname{grad} f(x)$ は一意に定まる．

問題 2. $\mathbb{R}^n$ にリーマン計量

$$\bar{g}_x(\xi,\eta) = \xi^T G_x \eta, \quad \xi,\eta \in T_x\mathbb{R}^n \simeq \mathbb{R}^n,\ x \in \mathbb{R}^n$$

を定める．ここで，任意の $x \in \mathbb{R}^n$ に対して G_x は n 次正定値対称行列で，$x \mapsto G_x$ は滑らかである．球面 $S^{n-1} = \{x \in \mathbb{R}^n \mid x^T x = 1\} \subset \mathbb{R}^n$ に誘導計量 g を定め，$(\mathbb{R}^n, \bar{g})$ のリーマン部分多様体とする．

(1) S^{n-1} 上の点 x における接空間 $T_xS^{n-1} = \{\xi \in \mathbb{R}^n \mid x^T\xi = 0\}$ への直交射影 $P_x\colon \mathbb{R}^n \to T_xS^{n-1}$ を求めよ．すなわち，任意の $d \in \mathbb{R}^n$ に対して，$P_x(d) \in T_xS^{n-1}$ かつ，$\bar{g}_x(d - P_x(d), \xi) = 0$ がすべての $\xi \in T_xS^{n-1}$ に対して成り立つような P_x を求めよ．

(2) $A \in \mathrm{Sym}(n)$ とし，実数値関数 $\bar{f}\colon \mathbb{R}^n \to \mathbb{R}$ を $x \in \mathbb{R}^n$ に対して $\bar{f}(x) := x^T A x$ と定める．ユークリッド勾配 $\nabla \bar{f}(x) = (\partial \bar{f}(x)/\partial x_i) \in \mathbb{R}^n$ について，(1) の P_x による射影 $P_x(\nabla \bar{f}(x))$ を求めよ．

(3) (2) の $\bar{f}$ を S^{n-1} に制限した $f := \bar{f}|_{S^{n-1}}\colon S^{n-1} \to \mathbb{R}$ の，S^{n-1} 上の点 x における勾配 $\operatorname{grad} f(x)$ を求めよ．これが (2) の結果と一般には一致しないのはなぜか？

解答例

(1) すべての $\xi \in T_xS^{n-1}$ に対して，$0 = \bar{g}_x(d - P_x(d), \xi) = (d - P_x(d))^T G_x \xi = (G_x(d - P_x(d)))^T \xi$ が成り立つから，ある $\alpha \in \mathbb{R}$ が存在して $G_x(d - P_x(d)) = \alpha x$ と書けて，$P_x(d) = d - \alpha G_x^{-1} x$ となる．一方，$P_x(d) \in T_xS^{n-1}$ から $x^T P_x(d) = 0$ となるので，$x^T d - \alpha x^T G_x^{-1} x = 0$ より，$\alpha = (x^T d)/(x^T G_x^{-1} x)$ と書ける．よって，

$$P_x(d) = d - \frac{x^T d}{x^T G_x^{-1} x} G_x^{-1} x = \left(I_n - \frac{G_x^{-1} x x^T}{x^T G_x^{-1} x}\right) d$$

である．

(2) $\nabla \bar{f}(x) = 2Ax$ より，

$$P_x(\nabla \bar{f}(x)) = 2\left(I_n - \frac{G_x^{-1} x x^T}{x^T G_x^{-1} x}\right) A x$$

である．

(3) リーマン多様体 $(\mathbb{R}^n, \bar{g})$ 上の点 x における $\bar{f}$ の勾配 $\operatorname{grad} \bar{f}(x)$ は，定義から任意の $\xi \in \mathbb{R}^n$ に対して $\mathrm{D}\bar{f}(x)[\xi] = \bar{g}_x(\operatorname{grad} \bar{f}(x), \xi)$, すなわち $2x^T A \xi = (\operatorname{grad} \bar{f}(x))^T G_x \xi$ を満たす．ゆえに $G_x \operatorname{grad} \bar{f}(x) = 2Ax$ となり，$\operatorname{grad} \bar{f}(x) = 2G_x^{-1} A x$ である．よって，求める勾配はこれを T_xS^{n-1} に直交射影することで計算でき，

$$\operatorname{grad} f(x) = P_x(\operatorname{grad} \bar{f}(x)) = 2\left(I_n - \frac{G_x^{-1} x x^T}{x^T G_x^{-1} x}\right) G_x^{-1} A x$$

となる．なお，(2) の結果は，標準内積から定まるリーマン多様体 $\mathbb{R}^n$ での勾配（ユークリッド勾配）$\nabla \bar{f}(x)$ を射影したものである．今回の (S^{n-1}, g) はそのような $\mathbb{R}^n$ のリーマン部分多様体ではないから，$\nabla \bar{f}(x)$ の直交射影は $\operatorname{grad} f(x)$ とは一般には一致しない．

問題 3. $1 \le p \le n$ とする．シュティーフェル多様体 $\mathrm{St}(p,n)$ 上の点 X における接空間 $T_X\,\mathrm{St}(p,n)$ は

$$T_X\,\mathrm{St}(p,n) = \{\xi \in \mathbb{R}^{n\times p} \mid X^T\xi + \xi^T X = 0\} \tag{\#}$$

と表される．

(1) $X_\perp^T X_\perp = I_{n-p}$ かつ $X^T X_\perp = 0$ となるような $X_\perp \in \mathbb{R}^{n\times(n-p)}$ が存在することを証明せよ．

(1) における $X_\perp$ は一意的ではないが，以下では $X_\perp$ を一つ固定する．

(2) 任意の $A \in \mathbb{R}^{n\times p}$ はある $B \in \mathbb{R}^{p\times p}, C \in \mathbb{R}^{(n-p)\times p}$ を用いて $A = XB + X_\perp C$ と書けることを証明せよ．

(3) 行列 $B \in \mathbb{R}^{p\times p}, C \in \mathbb{R}^{(n-p)\times p}$ を用いて表される行列 $\xi = XB + X_\perp C$ について，$\xi \in T_X\,\mathrm{St}(p,n)$ となるための必要十分条件を B, C を用いて表し，$\dim T_X\,\mathrm{St}(p,n) = p(p-1)/2 + p(n-p)$ を証明せよ．

(4) $\mathbb{R}^{n\times p}$ に標準内積によるリーマン計量

$$\bar{g}_X(\xi,\eta) = \mathrm{tr}(\xi^T\eta), \quad \xi,\eta \in T_X\mathbb{R}^{n\times p} \simeq \mathbb{R}^{n\times p},\ X \in \mathbb{R}^{n\times p}$$

を定める．シュティーフェル多様体 $\mathrm{St}(p,n)$ に誘導計量 g を定め，$(\mathbb{R}^{n\times p}, \bar{g})$ のリーマン部分多様体とする．このとき，接空間 $T_X\,\mathrm{St}(p,n)$ への直交射影 P_X を，(#) または (3) の結果を用いて求めよ．すなわち，任意の $Y \in \mathbb{R}^{n\times p}$ に対して $P_X(Y)$ を求めよ．

解答例

(1) X の列ベクトル $x_1, x_2, \ldots, x_p \in \mathbb{R}^n$ は $\mathbb{R}^n$ の標準内積に関して正規かつ互いに直交するから，ある $x_{p+1}, x_{p+2}, \ldots, x_n \in \mathbb{R}^n$ が存在して，$\{x_1, x_2, \ldots, x_n\}$ が $\mathbb{R}^n$ の正規直交基底となる．このとき，$X_\perp = (x_{p+1}, x_{p+2}, \ldots, x_n) \in \mathbb{R}^{n\times(n-p)}$ とおくと，X と $X_\perp$ を並べた行列 $(X, X_\perp)$ は n 次直交行列であるから，$I_n = (X, X_\perp)^T(X, X_\perp) = \begin{pmatrix} X^TX & X^TX_\perp \\ X_\perp^TX & X_\perp X_\perp \end{pmatrix}$ となり，$X_\perp^T X_\perp = I_{n-p}$ かつ $X^T X_\perp = 0$ が成り立つ．

(2) $\{x_1, x_2, \ldots, x_n\}$ は $\mathbb{R}^n$ の基底だから，任意のベクトル $a \in \mathbb{R}^n$ は $x_1, x_2, \ldots, x_n$ の線形結合として表せるので，ある $b \in \mathbb{R}^p, c \in \mathbb{R}^{n-p}$ を用いて $a = Xb + X_\perp c$ と書ける．よって，A の第 i 列ベクトル a_i に対して，$b_i \in \mathbb{R}^p, c_i \in \mathbb{R}^{n-p}$ が存在して $a_i = Xb_i + X_\perp c_i$ となる．そこで，$B = (b_1, b_2, \ldots, b_p) \in \mathbb{R}^{p\times p}$, $C = (c_1, c_2, \ldots, c_p) \in \mathbb{R}^{(n-p)\times p}$ とおけば，$A = XB + X_\perp C$ となる．

(3) $\xi \in T_X\,\mathrm{St}(p,n) \iff X^T(XB + X_\perp C) + (XB + X_\perp C)^T X = 0 \iff B + B^T = 0 \iff B \in \mathrm{Skew}(p)$. これより，$T_X\,\mathrm{St}(p,n) = \{XB + X_\perp C \mid B \in \mathrm{Skew}(p),\ C \in \mathbb{R}^{(n-p)\times p}\}$ が成り立つ．したがって，$\dim T_X\,\mathrm{St}(p,n) = \dim \mathrm{Skew}(p) + \dim \mathbb{R}^{(n-p)\times p} = p(p-1)/2 + p(n-p)$ となる．

(4) $Y \in \mathbb{R}^{n\times p}$ について $P_X(Y)$ は，任意の $\xi = XB + X_\perp C \in T_X\,\mathrm{St}(p,n)$（$B \in \mathrm{Skew}(p)$）に対して，

$$0 = \bar{g}_X(Y - P_X(Y), \xi) = \mathrm{tr}((Y - P_X(Y))^T(XB + X_\perp C))$$

を満たす．(2) より，ある $D \in \mathbb{R}^{p\times p}, E \in \mathbb{R}^{(n-p)\times p}$ に対して $Y - P_X(Y) = XD + X_\perp E$ と書けるから，

$$0 = \mathrm{tr}((XD + X_\perp E)^T(XB + X_\perp C)) = \mathrm{tr}(D^TB + E^TC)$$

である．これが任意の $B \in \mathrm{Skew}(p), C \in \mathbb{R}^{(n-p)\times p}$ について成り立つから，$D \in \mathrm{Sym}(p)$ かつ $E = 0$ となる．したがって，$P_X(Y) = Y - XD$ かつ $D \in \mathrm{Sym}(p)$ であり，一方，$P_X(Y) \in T_X\,\mathrm{St}(p,n)$ より $X^TP_X(Y) + P_X(Y)^TX = 0$ であるから，$X^T(Y - XD) + (Y - XD)^TX = 0$ となる．これを D について解くと，$D = (X^TY + Y^TX)/2 = \mathrm{sym}(X^TY)$ と求まるから，

$$P_X(Y) = Y - X\,\mathrm{sym}(X^TY)$$

となる．

問題 4. $1 \le p \le n$ とする．シュティーフェル多様体 $\mathrm{St}(p,n)$ 上のレトラクションとして，

$$R_X^{\mathrm{QR}}(\eta) = \mathrm{qf}(X+\eta), \quad \eta \in T_X \mathrm{St}(p,n),\ X \in \mathrm{St}(p,n) \tag{*}$$

が知られている．ここで，$\mathrm{qf}(\cdot)$ は，フルランク行列 $Y \in \mathbb{R}_p^{n\times p}$ に対して，QR 分解の Q 成分を返す．すなわち，

$$Y = QR, \quad Q \in \mathrm{St}(p,n) \text{ かつ，} R \in \mathbb{R}^{p\times p} \text{ は対角成分がすべて正の上三角行列} \tag{**}$$

のとき，$\mathrm{qf}(Y) = Q$ である．式 (*) で定義される R^{QR} が実際に $\mathrm{St}(p,n)$ 上の well-defined なレトラクションであることを確認しよう．

(1) $Q_1, Q_2 \in \mathrm{St}(p,n)$ がある正則行列 $G \in \mathbb{R}^{p\times p}$ について $Q_2 = Q_1 G$ を満たすとき，$G \in \mathcal{O}(p)$ であることを証明せよ．

(2) フルランク行列 $Y \in \mathbb{R}_p^{n\times p}$ に対して，QR 分解 (**) の Q, R は一意に定まることを証明せよ．なお，そのような Q, R の組が少なくとも一つ存在することは，Gram–Schmidt の直交化法により保証される．

(3) $X \in \mathrm{St}(p,n), \eta \in T_X \mathrm{St}(p,n)$ のとき，$X+\eta \in \mathbb{R}^{n\times p}$ がフルランクであること，すなわち $\mathrm{rank}(X+\eta) = p$ であることを証明せよ．

(4) 写像 $\mathrm{qf}\colon \mathbb{R}_p^{n\times p} \to \mathrm{St}(p,n)$ について，$Y \in \mathbb{R}_p^{n\times p}, Z \in \mathbb{R}^{n\times p}$ に対して

$$\mathrm{D}\,\mathrm{qf}(Y)[Z] = Q\rho_{\mathrm{skew}}(Q^T Z R^{-1}) + (I_n - QQ^T)ZR^{-1}$$

が成り立つ（証明はたとえば [Sato 2021, Prop. 3.4] を参照）．ここで，$Q = \mathrm{qf}(Y) \in \mathrm{St}(p,n), R = Q^T Y$（つまり $Y = QR$ は Y の QR 分解 (**)）であり，正方行列 $A = (a_{ij})$ に対して $\rho_{\mathrm{skew}}(A) = (r_{ij})$ は狭義下三角部分が A の狭義下三角部分と一致するような n 次反対称行列である．すなわち，

$$r_{ij} = \begin{cases} a_{ij} & (i > j \text{ のとき}), \\ 0 & (i = j \text{ のとき}), \\ -a_{ji} & (i < j \text{ のとき}). \end{cases}$$

これを用いて，(*) で定義される R^{QR} が $\mathrm{St}(p,n)$ 上のレトラクションであることを証明せよ．

解答例

(1) $I_p = Q_2^T Q_2 = (Q_1 G)^T (Q_1 G) = G^T (Q_1^T Q_1) G = G^T G$ より，$G \in \mathcal{O}(p)$ である．

(2) $Y = Q_1 R_1 = Q_2 R_2$ なる QR 分解を考えると，R_1, R_2 は正則だから $Q_2 = Q_1 (R_1 R_2^{-1})$ で，(1) より $R_1 R_2^{-1} \in \mathcal{O}(p)$ となる．また，R_2 は上三角行列だから R_2^{-1} も，したがって $R_1 R_2^{-1}$ も上三角行列である．対角成分がすべて正の上三角な直交行列は，第 1 列が $(1,0,\ldots,0)^T$, 第 2 列が $(0,1,0,\ldots,0)^T, \ldots$ と定まり単位行列となるので，$R_1 R_2^{-1} = I_p$ となって $R_1 = R_2$ である．これより $Q_1 = Q_2$ もわかる．

(3) $(X+\eta)^T (X+\eta) = X^T X + (X^T \eta + \eta^T X) + \eta^T \eta = I_p + \eta^T \eta$ は正定値対称行列であるから，p 次正則行列である．よって，$\mathrm{rank}(X+\eta) = \mathrm{rank}((X+\eta)^T (X+\eta)) = p$ が成り立つ．

(4) 任意の $X \in \mathrm{St}(p,n)$ の QR 分解は $X = XI_p$ であり $\mathrm{qf}(X) = X$ となるから，$R^{\mathrm{QR}}(0) = \mathrm{qf}(X) = X$ が成り立つ．また，任意の $\eta \in T_X \mathrm{St}(p,n)$ に対して，$X^T \eta \in \mathrm{Skew}(p)$ より $\rho_{\mathrm{skew}}(X^T \eta) = X^T \eta$ に注意すると，

$$\mathrm{D}R_X^{\mathrm{QR}}(0)[\eta] = \mathrm{D}\,\mathrm{qf}(X)[\eta] = X\rho_{\mathrm{skew}}(X^T \eta I_p^{-1}) + (I_n - XX^T)\eta I_p^{-1} = XX^T \eta + (I_n - XX^T)\eta = \eta$$

が成り立つ．これで，R^{QR} はレトラクションの定義における 2 つの条件を満たすことが証明できた．

問題 5. M, N を微分同相な多様体とし，$\Phi\colon M \to N$ を微分同相写像とする．すなわち，Φ は全単射で，Φ も Φ^{-1} も滑らかであるとする．R^N が N 上のレトラクションであるとき，

$$R_x^M(\eta) := \Phi^{-1}(R_{\Phi(x)}^N(\mathrm{D}\,\Phi(x)[\eta])), \quad \eta \in T_xM,\ x \in M$$

により定義される R^M は M 上のレトラクションであることを証明せよ．

解答例

- まず，任意の $y \in N$ について $R_y^N(0) = y$ が成り立つことに注意して，任意の $x \in M$ に対して $R_x^M(0) = x$ を示す．微分 $\mathrm{D}\,\Phi(x)$ が線形写像であることから，$\mathrm{D}\,\Phi(x)[0] = 0$ となることに注意すると，$R_x^M(0) = \Phi^{-1}(R_{\Phi(x)}^N(\mathrm{D}\,\Phi(x)[0])) = \Phi^{-1}(R_{\Phi(x)}^N(0)) = \Phi^{-1}(\Phi(x)) = 0$ が確かに成り立つ．
- 次に，任意の $y \in N$ について $\mathrm{D}\,R_y^N(0) = \mathrm{id}_{T_yN}$ が成り立つことに注意して，任意の $x \in M$ に対して $\mathrm{D}\,R_x^M(0) = \mathrm{id}_{T_xM}$ を示す．合成写像の微分についての公式（問題 1. の (5)）を繰り返し用いると，任意の $\eta \in T_xM$ に対して，

$$\begin{aligned}
\mathrm{D}\,R_x^M(0)[\eta] &= \mathrm{D}\,\Phi^{-1}(R_{\Phi(x)}^N(\mathrm{D}\,\Phi(x)[0]))[\mathrm{D}\,R_{\Phi(x)}^N(\mathrm{D}\,\Phi(x)[0])[\mathrm{D}\,\Phi(x)[\eta]]] \\
&= \mathrm{D}\,\Phi^{-1}(R_{\Phi(x)}^N(0))[\mathrm{D}\,R_{\Phi(x)}^N(0)[\mathrm{D}\,\Phi(x)[\eta]]] \\
&= \mathrm{D}\,\Phi^{-1}(\Phi(x))[\mathrm{D}\,\Phi(x)[\eta]] \\
&= \mathrm{D}\,(\Phi^{-1} \circ \Phi)(x)[\eta] \\
&= \mathrm{D}\,\mathrm{id}(x)[\eta] \\
&= \eta
\end{aligned}$$

  が確かに成り立つ．なお，最後の等号は，任意の $f \in \mathfrak{F}_x(M)$ に対して $(\mathrm{D}\,\mathrm{id}(x)[\eta])f = \eta(f \circ \mathrm{id}) = \eta f$ であることから従う．

以上より，R^M はレトラクションの定義における 2 つの条件を満たすことが証明できた．

問題 6. リーマン多様体上の最適化問題に対するアルゴリズムを適用し，結果を観察せよ．用いるプログラミング言語・扱うリーマン多様体・その上の目的関数・適用するアルゴリズムは自由に選んで良い．講義中に紹介したライブラリなどの使い方を調べて利用することを想定している．

解答例

省略するが，たとえば MATLAB でのツールボックス Manopt (`https://www.manopt.org`) については，具体的な計算例が `https://www.manopt.org/tutorial.html` にて詳細に解説されている．