

ルジャンドル特異点と情報幾何学

中島 直道 (北海道大学大学院情報科学院, JSPS 特別研究員 DC2)*

1 はじめに

情報幾何学では、統計モデルや機械学習モデルのパラメータ空間を、フィッシャー情報行列をリーマン計量とするリーマン多様体とみなし、さらにその上にある双対性を満たす二種類のアファイン接続を考える。統計多様体とは、この双対的な構造を微分幾何学的な観点から定式化した概念である。

情報幾何学は、このような空間構造を通してパラメータ空間を扱うことにより、統計学や機械学習、最適化問題等への統一的な幾何学的解釈をもたらし、様々な解析を可能とする [2, 3].

特に情報幾何学における主要な空間概念として、甘利・長岡により導入された双対平坦多様体がある [2]. 双対平坦多様体は、アファイン接続が平坦である統計多様体であり、これはアファイン微分幾何学におけるヘッセ多様体と同一の概念である (志摩 [17]). すなわち、双対平坦多様体では、局所アファイン座標系においてリーマン計量が局所凸ポテンシャル関数のヘッセ行列で与えられる。さらに、二種類の平坦アファイン接続それぞれに付随する局所凸ポテンシャルは互いにルジャンドル双対性により結びついている。

双対平坦多様体は実応用の観点からも注目されている一方で、深層学習を含め実応用で現れる空間ではしばしば計量が退化する [19]. すなわち、局所ポテンシャル関数は非凸で変曲的あるいは多価になり得る。このような場合には、厳密には、双対平坦多様体の理論が適用できない。そこで著者らは、計量の退化性を許容する双対平坦多様体の理論の一般化として概ヘッセ多様体の理論を接触幾何学と特異点理論の観点から建設した [15]. 理論刷新の核は、波面、すなわち非凸ポテンシャルあるいは多価ポテンシャルのグラフに依然として働くルジャンドル双対性である。

本稿では [15, 14] に基づき、情報幾何学の概説を行い、概ヘッセ多様体の理論について、統計多様体の幾何学との関連を踏まえて紹介する。統計多様体の幾何学は、コントラスト関数の理論と密接に関連する。コントラスト関数は、多様体上のある条件を満たす非対称な“距離 (二乗) 関数”であり、統計多様体上にさまざまなテンソルを導入する。これらのテンソルは統計多様体の幾何学を特徴づけ、その関連について江口は徹底的に調べている [9]. 特に 4 次テンソルに基づく統計多様体の幾何学は、概ヘッセ多様体の幾何学との関連が示唆されるものである。

本稿を通して、写像や多様体は全て滑らかなものを考え、太字は列ベクトルを表すものとする、e.g., $\boldsymbol{x} = (x_1, \dots, x_n)^T$.

* 〒060-0814 札幌市北区北 14 条西 9 丁目
e-mail: nakajima-n@ist.hokudai.ac.jp

2 情報幾何学の概観

本節では情報幾何学に登場する諸概念についての定義を整理しながら、双対平坦多様体の理論についての紹介を行う。

M を n 次元多様体, h をその上の擬リーマン計量 (非退化な対称 $(0, 2)$ テンソル場) とする. ∇ を接束 TM の振率がゼロであるアファイン接続とする. (M, h, ∇) が統計多様体であるとは, $(0, 3)$ テンソル場 $C := \nabla h$ が全対称であるとき, すなわち任意のベクトル場 $X, Y, Z \in \mathfrak{X}(M)$ に対して次が成り立つときをいう:

$$(\nabla_X h)(Y, Z) = (\nabla_Y h)(X, Z).$$

このときテンソル C を Amari-Chentsov の 3 次テンソルと呼ぶ [2]. 統計多様体 (M, h, ∇) に対し, TM の接続である ∇ の双対接続 ∇^* を次式で定義する:

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla_X^* Z).$$

ここで, X, Y, Z は M 上の任意のベクトル場である. このとき, 双対接続 ∇^* もまた振率が消えている. 上記の統計多様体の定義と同値な定義として Lauritzen によるものがある [11]. すなわち, 擬リーマン多様体 (M, h) 上に対称な $(0, 3)$ テンソル場 C が与えられているとき, 次式によりアファイン接続 ∇, ∇^* を定義すると (M, h, ∇) は統計多様体となる:

$$h(\nabla_X Y, Z) = h(\bar{\nabla}_X Y, Z) - \frac{1}{2}C(X, Y, Z), \quad h(\nabla_X^* Y, Z) = h(\bar{\nabla}_X Y, Z) + \frac{1}{2}C(X, Y, Z).$$

ここで, $\bar{\nabla}$ は h のレビ・チビタ接続である. いずれの定義であっても統計多様体の幾何学を考える上では (h, C, ∇, ∇^*) を考えることが重要である.

統計多様体 (M, h, ∇) において, ∇ が平坦であることと, その双対接続 ∇^* が平坦であることが同値となる. このような統計多様体 (M, h, ∇, ∇^*) を双対平坦多様体 [2] あるいはヘッセ多様体 [17] と呼び, (h, ∇, ∇^*) を M 上の双対平坦構造と呼ぶ. 双対平坦多様体 (M, h, ∇, ∇^*) は次の重要な性質をもつ.

命題 2.1 M の各点 p の周りにおいて ∇ に関するアファイン座標系 (∇ -アファイン座標系) $\mathbf{x} = (x_1, \dots, x_n)$ とその上の関数 $z = f(\mathbf{x})$ が存在して, 次が成り立つ:

- (1) 任意の $1 \leq i, j \leq n$ に対し, $h(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$;
- (2) $p_i := \frac{\partial f}{\partial x_i}$ ($1 \leq i \leq n$) で定義される $\mathbf{p} = (p_1, \dots, p_n)$ は ∇^* -アファイン座標系である ;
- (3) $h(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial p_j}) = \delta_{ij}$;
- (4) ∇^* -アファイン座標系 \mathbf{p} 上の関数 $z' = \varphi(\mathbf{p})$ が存在して, $z + z' - \mathbf{x}^T \mathbf{p} = 0$ を満たす ;
- (5) $C(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}, \frac{\partial}{\partial x_k}) = \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}$.

上記の命題 2.1 における ∇^* -アファイン座標系 \mathbf{p} は ∇ -アファイン座標系 \mathbf{x} の双対座標と呼ばれ、関数 $f(\mathbf{x})$ はポテンシャル関数、 $\varphi(\mathbf{p})$ は f の双対ポテンシャル関数あるいはルジャンドル変換と呼ばれる。

双対平坦多様体 (M, h, ∇, ∇^*) では、局所的にポテンシャル関数 $f(\mathbf{x})$ と双対ポテンシャル関数 $\varphi(\mathbf{p})$ が存在するが、これらを用いて定義される

$$D : M \times M \rightarrow \mathbb{R}, \quad D(p, q) = f(\mathbf{x}(p)) + \varphi(\mathbf{p}(q)) - \mathbf{x}(p)^T \mathbf{p}(q)$$

をブレグマンダイバージェンスという（厳密には、 M の対角線集合の近傍で定義される）。上式の定義はアファイン座標系 \mathbf{x}, \mathbf{p} やポテンシャル関数 f, φ の取り方にはよらないことに注意する。情報幾何学の応用を考える上でブレグマンダイバージェンスは重要な役割を果たすものである [2]。

3 接触幾何学と概ヘッセ多様体の理論

本節では、2 節で述べた双対平坦多様体の理論の一般化である概ヘッセ多様体の理論の紹介を行う。証明等の詳細は [15] を参照されたい。

N を $2n+1$ 次元多様体、 ξ をその上の超平面場とする。 (N, ξ) が接触多様体であるとは、超平面場 ξ が局所的に $\theta \wedge (d\theta)^n \neq 0$ なる 1 次微分形式 θ の核で表されるときをいう。このとき、 θ を（局所）接触形式という。 $2n+1$ 次元接触多様体 (N, ξ) に対して、 n 次元部分多様体 $L \subset N$ がルジャンドル部分多様体であるとは、各点 $p \in N$ において $T_p L \subset \xi_p$ なることをいう。

$\mathbb{R}^{2n+1} (= T^*\mathbb{R}^n \times \mathbb{R})$ に対して、 $(\mathbf{x}, \mathbf{p}, z)$ でその標準的な座標系を表すものとする。ここで、 \mathbf{x} と \mathbf{p} はそれぞれ $T^*\mathbb{R}^n$ の底空間とファイバーの座標系を表し、これら二つの空間を \mathbb{R}_x^n と \mathbb{R}_p^n と書き、区別することにする。 \mathbb{R}^{2n+1} 上の 1 次微分形式 θ を

$$\theta := z - \mathbf{p}^T d\mathbf{x} = z - \sum_{i=1}^n p_i dx_i$$

で定める。このとき、 $(\mathbb{R}^{2n+1}, \theta)$ は接触多様体であり、この接触多様体を標準接触多様体と呼ぶ。

標準接触多様体 $(\mathbb{R}^{2n+1}, \theta)$ に対して、ファイバー束

$$\pi : \mathbb{R}^{2n+1} \rightarrow \mathbb{R}_x^n \times \mathbb{R}_z, \quad (\mathbf{x}, \mathbf{p}, z) \mapsto (\mathbf{x}, z)$$

はルジャンドルファイブレーション（各ファイバーがルジャンドル部分多様体である）を与える。変換 $\mathcal{L} : \mathbb{R}^{2n+1} \rightarrow \mathbb{R}^{2n+1}$,

$$\mathcal{L}(\mathbf{x}, \mathbf{p}, z) = (\mathbf{x}', \mathbf{p}', z') = (\mathbf{p}, \mathbf{x}, \mathbf{x}^T \mathbf{p} - z)$$

を考えると、これは微分同相写像であり接触超平面場を保つ。ファイバー方向への射影

$$\pi' := \pi \circ \mathcal{L} : \mathbb{R}^{2n+1} \rightarrow \mathbb{R}_p^n \times \mathbb{R}_{z'}, \quad (\mathbf{x}, \mathbf{p}, z) \mapsto (\mathbf{p}, \mathbf{x}^T \mathbf{p} - z)$$

もまたルジャンドルファイブレーションとなる。このとき、二つのルジャンドルファイブレーションからなる次の図式を標準接触多様体 $(\mathbb{R}^{2n+1}, \theta)$ に対するダブルファイブレーション構造と呼ぶ。

$$\mathbb{R}_x^n \times \mathbb{R}_z \xleftarrow{\pi} \mathbb{R}^{2n+1} \xrightarrow{\pi'} \mathbb{R}_p^n \times \mathbb{R}_{z'}$$

$L \subset \mathbb{R}^{2n+1}$ をルジャンドル部分多様体とする。 L に対し、ルジャンドル写像

$$\pi^e := \pi \circ \iota : L \rightarrow \mathbb{R}_x^n \times \mathbb{R}_z, \quad \pi^m := \pi' \circ \iota : L \rightarrow \mathbb{R}_p^n \times \mathbb{R}_{z'}$$

をそれぞれ e/m -ルジャンドル写像という。ここで、 $\iota : L \rightarrow \mathbb{R}^{2n+1}$ は包含写像である。

定義 3.1 ([15]) e/m -ルジャンドル写像 π^e と π^m に対して、

$$W_e(L) := \pi^e(L) \subset \mathbb{R}_x^n \times \mathbb{R}_z, \quad W_m(L) := \pi^m(L) \subset \mathbb{R}_p^n \times \mathbb{R}_{z'}$$

をそれぞれ L に付随する e/m -波面と呼ぶ。

L 上のベクトル束 $E (= E_L)$ を

$$E := \{ (p, w) \in L \times (\mathbb{R}_x^n \times \mathbb{R}_z) \mid dz_p(w) - \mathbf{p}(p)^T d\mathbf{x}_p(w) = 0 \}$$

で定める。 L はルジャンドル部分多様体であることから、 $d\pi^e(T_p L) \subset E_p$ が各点 $p \in L$ において成り立つ。ここで E_p は p におけるファイバーである。このことから、束写像

$$\Phi : TL \rightarrow E, \quad v_p \mapsto d\pi_p^e(v_p)$$

が意味を持つことに注意する。

$\tilde{\nabla}$ を $\mathbb{R}_x^n \times \mathbb{R}_z$ 上の平坦アファイン接続とし、 $\psi_p : \mathbb{R}_x^n \times \mathbb{R}_z \rightarrow E_p$ を z -軸に沿った線形射影とする。このとき、

$$\nabla_X^E \eta(p) := \psi_p \circ \tilde{\nabla}_X \eta(p)$$

により E 上のアファイン接続 ∇^E を定義する。ここで、 X は L 上のベクトル場であり η は E の切断である。

命題 3.2 ([15]) 接続 ∇^E は平坦であり、任意の L 上のベクトル場 X, Y に対し

$$\nabla_X^E(\Phi(Y)) - \nabla_Y^E(\Phi(X)) = \Phi([X, Y])$$

が成り立つ。

(E, Φ, ∇^E) を e -波面 $W_e(L)$ に付随する連接接束と呼ぶ (cf. [16])。同様にして、 m -波面 $W_m(L)$ に対しても連接接束 $(E', \Phi', \nabla^{E'})$ を定める。

これらベクトル束 E, E' は L 上で定義されたものであるが、 \mathbb{R}^{2n+1} 上で定義され得るものである。超平面場 ξ は直和分解

$$\xi_p = \ker d\pi_p' \oplus \ker d\pi_p \simeq E_p \oplus E'_p \simeq \mathbb{R}_x^n \oplus \mathbb{R}_p^n$$

をもつことに注意する. このとき, ξ 上にシンプレクティック形式 ω と (n, n) 型の擬リーマン計量 τ が

$$\omega := \sum_{i=1}^n dx_i \wedge dp_i, \quad \tau := \sum_{i=1}^n dx_i dp_i = \frac{1}{2} \sum_{i=1}^n (dx_i \otimes dp_i + dp_i \otimes dx_i)$$

により自然に定義される.

定義 3.3 ([15]) L 上の概ヘッセ計量 h を τ の引き戻しにより定義する:

$$h(Y, Z) := \tau(\iota_* Y, \iota_* Z) \quad (Y, Z \in TL).$$

ここで, $\iota_* = \Phi \oplus \Phi' : TL \hookrightarrow \xi = E \oplus E'$ は包含写像である.

一般に, ルジャンドル部分多様体 $L \subset \mathbb{R}^{2n+1}$ は局所的に母関数 $g(\mathbf{x}_I, \mathbf{p}_J)$ を用いて

$$L = \{(\mathbf{x}_I, \mathbf{x}_J, \mathbf{p}_I, \mathbf{p}_J, z) \in \mathbb{R}^{2n+1} \mid \mathbf{p}_I = \frac{\partial g}{\partial \mathbf{x}_I}, \mathbf{x}_J = -\frac{\partial g}{\partial \mathbf{p}_J}, z = \mathbf{p}_J^T \mathbf{x}_J + g(\mathbf{x}_I, \mathbf{p}_J)\}$$

のように表される [5]. ここで, $I \sqcup J = \{1, \dots, n\}$ は分割であり, $(\mathbf{x}_I, \mathbf{p}_J)$ は L の局所座標系である. また, $\frac{\partial g}{\partial \mathbf{x}_I}$ は列ベクトル $(\frac{\partial g}{\partial x_i})_{i \in I}^T$ を表すものとする.

命題 3.4 ([15]) $g(\mathbf{x}_I, \mathbf{p}_J)$ を L の母関数とする. このとき,

$$h = \sum_{i,k \in I} \frac{\partial^2 g}{\partial x_i \partial x_k} dx_i dx_k - \sum_{j,l \in J} \frac{\partial^2 g}{\partial p_j \partial p_l} dp_j dp_l.$$

$\mathbb{R}_x^n \times \mathbb{R}_z$ と $\mathbb{R}_p^n \times \mathbb{R}_{z'}$ 上のアファイン変換

$$F(\mathbf{x}, z) = (A\mathbf{x} + \mathbf{b}, z + \mathbf{c}^T \mathbf{x} + d), \quad F^*(\mathbf{p}, z') = (A'\mathbf{p} + \mathbf{b}', z' + \mathbf{c}'^T \mathbf{p} + d')$$

はアファインルジャンドル同値 $\mathcal{L}_F : \mathbb{R}^{2n+1} \rightarrow \mathbb{R}^{2n+1}$,

$$\mathcal{L}_F(\mathbf{x}, \mathbf{p}, z) = (A\mathbf{x} + \mathbf{b}, A'\mathbf{p} + \mathbf{b}', z + \mathbf{c}^T \mathbf{x} + d)$$

を定める. ここで, A は正則行列で, $A' = (A^T)^{-1}$, $\mathbf{b}' = A'\mathbf{c}$, $\mathbf{b} = A\mathbf{c}$, $d' = \mathbf{b}'^T \mathbf{b} - d$ である. この変換 \mathcal{L}_F はダブルファイブレーション構造と接触構造を保つ (従って ω と τ も保つ).

二つのルジャンドル部分多様体 L_1, L_2 がアファインルジャンドル同値 \mathcal{L}_F により同一視されているとする: $\mathcal{L}_F(L_1) = L_2$. このとき, \mathcal{L}_F は概ヘッセ計量を保ち, 接続接束間の同型 $E_{L_1} \simeq E_{L_2}$, $E'_{L_1} \simeq E'_{L_2}$ を自然に引き起こす. この同型はまた, 接続接束上の平坦アファイン接続を自然に同一視する.

定義 3.5 ([15]) 概ヘッセ多様体 $(M, \mathcal{U} = \{L_\alpha\})$ とは, ルジャンドル部分多様体 $L_\alpha \subset \mathbb{R}^{2n+1}$ をアファインルジャンドル同値により張り合わせてできたものである. このとき, 各 L_α の構造から, M 上に退化し得る $(0, 2)$ 型テンソル h と接続接束 (E, Φ, ∇^E) , $(E', \Phi', \nabla^{E'})$ が well-defined に定まる. 各 L_α を M の局所モデルと呼ぶ.

以上の構成により概ヘッセ多様体 M を含む接触多様体が自然に定義されることに注意する.

$(M, h, (E, \Phi, \nabla^E), (E', \Phi', \nabla^{E'}))$ を概ヘッセ多様体とする. M 上の任意のベクトル場 Y, Z に対して,

$$(\eta, \eta') := (\Phi \oplus \Phi')(Y), (\zeta, \zeta') := (\Phi \oplus \Phi')(Z), \tau(\eta, \zeta') := \tau(\eta \oplus 0, 0 \oplus \zeta')$$

とおく.

定義 3.6 ([15]) 概ヘッセ多様体 M に対して, 標準 3 次テンソル C を次で定義する:

$$C(X, Y, Z) := \tau(\eta, \nabla_X^{E'} \zeta') + \tau(\zeta, \nabla_X^{E'} \eta') - \tau(\nabla_X^E \eta, \zeta') - \tau(\nabla_X^E \zeta, \eta').$$

ここで, X, Y, Z は M 上のベクトル場である.

命題 3.7 ([15]) 標準 3 次テンソル C は局所的に母関数 $g(\mathbf{x}_I, \mathbf{p}_J)$ の 3 次導関数である: 任意の k, l, m に対し

$$C(\partial_k, \partial_l, \partial_m) = \partial_k \partial_l \partial_m g.$$

ここで, $\partial_k := \frac{\partial}{\partial x_k}$ ($k \in I$) または $\partial_k := \frac{\partial}{\partial p_k}$ ($k \in J$) である. 従って, C は対称である.

注意 3.8 概ヘッセ計量 h がいたるところ非退化であるとき, M 上には従来の双対平坦構造が自然に復元される. 実際このとき Φ, Φ' は同型写像であり, $\Phi(\nabla_X Y) = \nabla_X^E \Phi(Y)$, $\Phi'(\nabla_X^* Y) = \nabla_X^{E'} \Phi'(Y)$ により TM 上の接続 ∇, ∇^* が一意的に定まる. ここで, X, Y は M 上の任意のベクトル場である. さらにこのとき,

$$\begin{aligned} \tau(\nabla_X^E \eta, \zeta') &= \tau(\nabla_X^E \Phi(Y), \Phi'(Z)) = \tau(\Phi(\nabla_X Y), \Phi'(Z)) = \frac{1}{2} h(\nabla_X Y, Z), \\ \tau(\eta, \nabla_X^{E'} \zeta') &= \tau(\Phi(Y), \nabla_X^{E'} \Phi'(Z)) = \tau(\Phi(Y), \Phi'(\nabla_X^* Z)) = \frac{1}{2} h(Y, \nabla_X^* Z). \end{aligned}$$

である. このため Amari-Chentsov テンソルも自然に復元される:

$$\begin{aligned} C(X, Y, Z) &= \frac{1}{2}(h(\nabla_X^* Y, Z) - h(\nabla_X Y, Z)) + \frac{1}{2}(h(\nabla_X^* Z, Y) - h(\nabla_X Z, Y)) \\ &= (\nabla h)(X, Y, Z). \end{aligned}$$

双対平坦多様体上のブレグマンダイバージェンスは概ヘッセ多様体上の正準ダイバージェンスとして以下のように一般化される. $L \subset \mathbb{R}^{2n+1}$ をルジャンドル部分多様体とする. $p \in L$ について, \mathbb{R}^{2n+1} の座標系 $(\mathbf{x}, \mathbf{p}, z)$ を用いて

$$p = (\mathbf{x}(p), \mathbf{p}(p), z(p)) \in \mathbb{R}^{2n+1}, \quad z'(p) = \mathbf{x}(p)^T \mathbf{p}(p) - z(p) \in \mathbb{R}$$

と書くことにする. L 上の正準ダイバージェンス $\mathcal{D}_L : L \times L \rightarrow \mathbb{R}$ を

$$\mathcal{D}_L(p, q) = z(p) + z'(q) - \mathbf{x}(p)^T \mathbf{p}(q)$$

で定義する. 概ヘッセ計量が非退化であるとき, \mathcal{D}_L はブレグマンダイバージェンスに他ならない. \mathcal{D}_L はアファインルジャンドル同値の下で不変である [15]. このため, 自然に概ヘッセ多様体 $(M, \mathcal{U} = \{L_\alpha\})$ 上に正準ダイバージェンス \mathcal{D}_M が定義される.

4 コントラスト関数の理論と概ヘッセ多様体の幾何学

本節では、江口によるコントラスト関数の理論を通して、統計多様体の幾何学と概ヘッセ多様体の幾何学の関連について述べる。

多様体 M 上の関数 $\rho: M \times M \rightarrow \mathbb{R}$ とベクトル場 $X_1, \dots, X_k, Y_1, \dots, Y_l$ を用いて、関数 $\rho[X_1 \cdots X_k | Y_1 \cdots Y_l]: M \rightarrow \mathbb{R}$,

$$\rho[X_1 \cdots X_k | Y_1 \cdots Y_l](r) = (X_1)_p \cdots (X_k)_p (Y_1)_q \cdots (Y_l)_q (\rho(p, q))|_{p=q=r}$$

を定義する。関数 $\rho: M \times M \rightarrow \mathbb{R}$ がコントラスト関数であるとは、任意の $r \in M$ と M 上のベクトル場 X, Y に対して次を満たすときをいう：

- (i) $\rho[-|-](r) = \rho(r, r) = 0$,
- (ii) $\rho[X|-](r) = \rho[-|X](r) = 0$,
- (iii) $-\rho[X|Y]$ は M 上の擬リーマン計量である。

一般に、 M 上にコントラスト関数 $\rho: M \times M \rightarrow \mathbb{R}$ が与えられたとき、

$$h(X, Y) := -\rho[X|Y], \quad C(X, Y, Z) := -\rho[Z|XY] + \rho[XY|Z] \quad (1)$$

により M 上に擬リーマン計量 h と対称 3 次テンソル C が導入され、 (M, h, C) は統計多様体になる。このとき同時に、

$$h(\nabla_X Y, Z) = -\rho[XY|Z], \quad h(\nabla_X^* Y, Z) = -\rho[Z|XY]$$

により、アファイン接続 ∇ とその双対接続 ∇^* も導入されることに注意する。江口はこのようなコントラスト関数から導入される統計多様体の幾何学を徹底的に調べている [9]。また、統計多様体 (M, h, C) に対し、上式 (1) によって統計多様体の構造を復元するコントラスト関数が存在することが知られている [12]。

(1, 1) テンソル場 B, B^* をコントラスト関数の 4 階微分を用いて次で定義する：

$$h(B(X, Y)Z, W) = -\rho[XYZ|W] - h(\nabla_X \nabla_Y Z, W), \quad (2)$$

$$h(Z, B^*(X, Y)W) = -\rho[Z|XYW] - h(Z, \nabla_X^* \nabla_Y^* W). \quad (3)$$

これらテンソルは江口により [9] で与えられたものであり、また、[13] においては (双対) バートレットテンソルと呼ばれ、 h や ∇ , ∇^* の曲率との関連が調べられている。

2 節で扱ったブレグマンダイバージェンスは双対平坦多様体上のコントラスト関数であり、その構造を復元する。

命題 4.1 双対平坦多様体 (M, h, ∇, ∇^*) と 3 次テンソル $C = \nabla h$ に対し、ブレグマンダイバージェンス $\mathcal{D}: M \times M \rightarrow \mathbb{R}$ は次を満たす：

$$h(X, Y) = -\mathcal{D}[X|Y], \quad C(X, Y, Z) = -\mathcal{D}[Z|XY] + \mathcal{D}[XY|Z].$$

命題 4.2 ([9]) 双対平坦多様体 (M, h, ∇, ∇^*) において、ブレグマンダイバージェンス $\mathcal{D} : M \times M \rightarrow \mathbb{R}$ から定義されるテンソル B, B^* は共に消える。すなわち、

$$\begin{aligned} h(\nabla_X \nabla_Y Z, W) &= -\mathcal{D}[XYZ|W], \\ h(Z, \nabla_X^* \nabla_Y^* W) &= -\mathcal{D}[Z|XYW]. \end{aligned}$$

概ヘッセ多様体 $(M, h, (E, \nabla^E, \Phi), (E', \nabla^{E'}, \Phi'))$ における正準ダイバージェンス $\mathcal{D}_M : M \times M \rightarrow \mathbb{R}$ は、コントラスト関数の条件式 (i), (ii) を満たし、(iii) は一般には満たさない。このような関数を弱コントラスト関数と呼ぶ [15]。正準ダイバージェンスは概ヘッセ多様体の構造を復元する。

定理 4.3 ([15]) 概ヘッセ多様体 $(M, h, (E, \nabla^E, \Phi), (E', \nabla^{E'}, \Phi'))$ と標準3次テンソル C に対し、正準ダイバージェンス $\mathcal{D}_M : M \times M \rightarrow \mathbb{R}$ は次を満たす：

$$h(X, Y) = -\mathcal{D}_M[X|Y], \quad C(X, Y, Z) = -\mathcal{D}_M[Z|XY] + \mathcal{D}_M[XY|Z].$$

概ヘッセ多様体上では、概ヘッセ計量の非退化性は一般には保証されないため式 (2) や (3) のようにしてパートレットテンソルを定義することはできないが、特に次の定理の式 (5) は双対平坦多様体におけるブレグマンダイバージェンスとパートレットテンソルの関係を反映したものである。

定理 4.4 ([14]) X, Y, Z, W を M 上のベクトル場とする。このとき、次が成り立つ：

$$\tau(\Phi(Z), \nabla_Y^{E'} \Phi'(W)) = -\frac{1}{2} \mathcal{D}_M[Z|YW], \quad (4)$$

$$\tau(\Phi(Z), \nabla_X^{E'} \nabla_Y^{E'} \Phi'(W)) = -\frac{1}{2} \mathcal{D}_M[Z|XYW]. \quad (5)$$

注意 4.5 定理 4.4 の式 (4), (5) に現れる $\tau(\Phi(Z), \nabla_Y^{E'} \Phi'(W))$ や $\tau(\Phi(Z), \nabla_X^{E'} \nabla_Y^{E'} \Phi'(W))$ は、概ヘッセ多様体における e/m -波面のある特異点を判定する量になっている。詳細は [14] を参照されたい。

5 機械学習との関連と今後の展望

本節では概ヘッセ多様体の理論と機械学習との関連、および今後の展望について述べる。情報幾何学の典型的な応用として、統計的推定の幾何学的解釈や em -アルゴリズムがある [2]。

$M = \{p(\mathbf{x}|\theta)\}_{\theta \in U}$, $p(\mathbf{x}|\theta) = \exp(\mathbf{x}^T \theta - \psi(\theta))$ を指数型分布族とする。ここで $\mathbf{x} \in \mathbb{R}^n$ は確率変数 (測度は $d\mu$) であり、 $\theta \in U$ はパラメータである ($U \subset \mathbb{R}^n$ は開集合)。 M をパラメータ空間 U と同一視することにより指数型分布族を可微分多様体とみなす。このとき、 M 上のフィッシャー・ラオ計量 h と、座標系 θ をアファイン座標系とする接続 ∇ を考え、正規化関数 $\psi(\theta)$ をポテンシャル関数とみることにより (M, h, ∇, ∇^*) は双対平坦多様体になる [2]。ここで、 ∇^* は ∇ の双対接続である。さらにブレグマンダイバー

ジェンス $\mathcal{D} : M \times M \rightarrow \mathbb{R}$ は情報理論で知られている KL ダイバージェンスとなる：

$$\mathcal{D}_{KL}(p, q) = KL[q, p] := \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} d\mu.$$

統計モデル $S := \{p(\mathbf{x}|\theta(\mathbf{u}))\}_{\mathbf{u} \in V} \subset M$ が曲指数型分布族、すなわち S が M の部分多様体であるとする。ここで $V \subset \mathbb{R}^m$ は開集合で、 $\mathbf{u} \in V$ はパラメータである。統計的推定では、観測データ $\{\mathbf{x}_i\}_{i=1}^N$ が定める観測点 $\hat{p} \in M$ に対し、 \hat{p} の最も良い近似 $p(\mathbf{x}|\theta(\mathbf{u}_0)) \in S$ を求めることを考える。統計的推定の一つである最尤推定とは、観測データ $\{\mathbf{x}_i\}_{i=1}^N$ から定義される尤度関数 $l(\mathbf{u}) := \sum_{i=1}^N \log p(\mathbf{x}_i|\theta(\mathbf{u}))$ を最大にするパラメータ \mathbf{u}_0 を求めるものであり、この \mathbf{u}_0 は最尤推定値と呼ばれる。情報幾何学は最尤推定の幾何学的解釈を与える。すなわち、最尤推定値 \mathbf{u}_0 は関数 $\mathcal{D}_{KL}(-, \hat{p}) : S \rightarrow \mathbb{R}$ を最小にするものであり、また、 \hat{p} から S への ∇^* -測地線に関する直交射影により与えられる。

一般に最尤方程式 $\frac{\partial}{\partial \mathbf{u}} l(\mathbf{u}) = 0$ の解として最尤推定値を求めることが多い一方で、最尤方程式は複数の解を持つことがあり、その個数は観測点 $\hat{p} \in M$ の位置に応じて変化する(分岐現象)。最尤方程式の複数解が与える推定への影響については実験的に調べられているが [18]、統一的な理論的枠組みは存在しない。

関数 $F : S \times M \rightarrow \mathbb{R}$ を

$$F(q, p) = \mathcal{D}_{KL}(\iota(q), p)$$

で定義する。ここで、 $\iota : S \rightarrow M$ は包含写像である。このとき、 F を大域的な母関数族とみなすことで得られる $T^*M \times \mathbb{R}$ 内のルジャンドル部分多様体 L_S [5] は概ヘッセ多様体の典型例であり、 L_S に付随するコースティクスにより解の分岐現象を特徴づけられ得る。

この枠組みは、*em*-アルゴリズムにも有効である。*em*-アルゴリズムとは、二つの部分多様体 $S, D \subset M$ に対し、 ∇ -測地線および ∇^* -測地線に関する直交射影を S と D の間で交互に繰り返すことにより、 $S \times D$ 上におけるダイバージェンスの最小化を図るアルゴリズムである。このアルゴリズムは統計学における EM-アルゴリズム [8] の幾何学的解釈を与えるものであり、実際、これらアルゴリズムは適切な仮定の下で一致する [1, 7]。この場合、二つの部分多様体 S, D それぞれに付随する概ヘッセ多様体を通じて EM/*em*-アルゴリズムの挙動が記述され得る。

また、多くの統計モデルはそのフィッシャー情報行列が退化する特異統計モデルであり [19]、深層学習等に現れる多層ニューラルネットワーク (MLP) はその典型例である。回帰の設定においては、しばしば MLP の出力に一定の分散を持つガウスノイズを加えることで、MLP を統計モデルとみなす [2]。このとき、重みパラメータからなるパラメータ空間 M 上では 3 次テンソル C が消え、 M は退化したフィッシャー・ラオ計量 h を持つ自己双対なリーマン多様体 $(M, h, \nabla = \nabla^*)$ とみなされる。この観点に基いた深層学習の解析として、従来の逆伝播学習法 (back-propagation method) に替わり、甘利により自然勾配学習法 (natural gradient method) が提案されている [4]。

しかしながら、上記の枠組みではガウスノイズの分散は固定されている一方で、より

一般に分散をパラメータに含めるほうが理論上は自然であろう [6, 10]. この場合, パラメータ空間上では 3 次テンソルはゼロにはならず, 真に特異な計量を許す双対平坦構造 (すなわち概ヘッセ構造) からの解析が必要となる. これは今後の課題である.

参考文献

- [1] S. Amari, *Information geometry of the EM and em algorithms for neural networks*, Neural Networks, **8** (1995), 1379–1408.
- [2] S. Amari, *Information Geometry and Its Application*, Applied Math. Sci., 194, Springer (2016)
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, A.M.S., Oxford Univ. Press (2000).
- [4] S. Amari and T. Ozeki, *Differential and Algebraic Geometry of Multilayer Perceptrons*, IEICE Trans., **84** (2001), 31–38.
- [5] V.I. Arnol'd et al., *Singularities of Differentiable Maps I*, Monographs in Math. 82, Birkhäuser (1986).
- [6] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).
- [7] I. Csizsár and G. Tusnády, *Information Geometry and Alternating Minimization Procedures*, Statistics and decisions, Supplementary Issue, No. 1 (1984), 205–237.
- [8] A. P. Dempster, N. M. Laird and D. B. Rubin, *Maximum Likelihood from Incomplete Data Via the EM Algorithm*, J. R. Stat. Soc. Ser. B. Stat. Methodol., **39** (1977), 1–38.
- [9] S. Eguchi, *Geometry of minimum contrast*, Hiroshima Math. J., **22** (1992), 631–647.
- [10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press (2016).
- [11] S. L. Lauritzen, *Statistical manifolds*, IMS Lecture Notes - Monograph Series (1987), 96–163.
- [12] T. Matsumoto, *Any statistical manifold has a contrast function – On the C^3 -functions taking the minimum at the diagonal of the product manifold*, Hiroshima Math. J., **23** (1993), 327–332.
- [13] H. Matsuzoe, *Geometry of contrast functions and conformal geometry*, Hiroshima Math. J., **29** (1999), 175–191.
- [14] N. Nakajima, *Local normal forms of em-wavefronts in affine flat coordinates*, preprint, arXiv:2204.13288 (2022).
- [15] N. Nakajima and T. Ohmoto, *The dually flat structure for singular models*, Info. Geom., **4** (2021), 31–64.
- [16] K. Saji, M. Umehara and K. Yamada, *The geometry of fronts*, Ann. of Math., **169** (2009), 491–529.
- [17] H. Shima, *The geometry of Hessian Structures*, World Scientific (2007).
- [18] C. Small and W. Jinfang, *Numerical Methods for Nonlinear Estimating Equations*, Oxford Univ. Press (2003).
- [19] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Cambridge Univ. Press (2008).