パラメータ空間の特異性と拡張ピタゴラスの定理

北海道大学大学院情報科学学院　中島　直道
Naomichi Nakajima
Graduate School of Information Science and Technology,
Hokkaido University

1　はじめに

情報幾何学では，統計学や機械学習におけるパラメータ空間をフィッシャー行列をリーマン計量とするリーマン多様体とみなすことで様々な解析を行う．特に甘利・長岡による双対平坦構造は情報幾何学における最も重要な概念であり，この空間構造を通して統計学や機械学習，最適化問題等に統一的な幾何学的解釈が与えられる [2, 1]．多様体 M 上の双対平坦構造 (h, ∇, ∇^*) とは，擬リーマン計量 h とある種の双対性を満たす平坦接続の組 (∇, ∇^*) からなるものであり，この構造を有する多様体 (M, h, ∇, ∇^*) を双対平坦多様体と呼ぶ．双対平坦多様体の持つ最も顕著な特徴は，計量 h が局所アファイン座標系上のポテンシャル関数のヘッセ行列として表現されることであり，この特徴を持つ多様体はアファイン微分幾何学ではヘッセ多様体と呼ばれる（志摩 [16, 17]）．

情報幾何学の応用は双対平坦多様体 (M, h, ∇, ∇^*) 上で成立している拡張ピタゴラスの定理と射影定理によって支えられている．これらの定理は，双対平坦多様体上の "距離関数" であるダイバージェンスと ∇ および ∇^* の測地線に関する定理であり，ダイバージェンスによる定量的な観点と測地線による定性的な観点を繋ぐ重要なものである．しかしながら，深層学習を含む多くの実応用の場面では，フィッシャー行列が退化し双対平坦構造が定義できない [18]．そこで著者らは，接触幾何学と特異点理論の観点から退化した計量を許容する双対平坦多様体として概ヘッセ多様体を導入し，甘利・長岡による双対平坦多様体の理論の一般化をおこなった [13]．特に概ヘッセ多様体上では測地線やその直交性，ダイバージェンスの概念が一般化され，拡張ピタゴラスの定理と射影定理が適切に定式化される．すなわち，我々の特異的設定においてもこれらの定理が依然として成立することが確かめられる．本稿では，始めに具体例を交えながら双対平坦多様体の理論を概観し，その後概ヘッセ多様体の理論の紹介を行う．本稿の主たる部分は著者らによる [13] の概説であり，特に拡張ピタゴラスの定理と射影定理について焦点を当てたものである．証明や詳細な議論は [13] を参

照されたい．

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.，$\boldsymbol{x} = (x_1, \cdots, x_n)^T$.

本研究は，JSPS 科研費 JP22J10499 の助成を受けたものである．

2　双対平坦多様体の理論

(M, h) を n 次元擬リーマン多様体とし，∇ を TM 上の捩れのないアファイン接続とする．$(0, 3)$-テンソル $T := \nabla h$ が全対称であるとき，(M, h, ∇) を統計多様体といい，このときテンソル T を Amari-Chentsov テンソルという [1, 2]．次式で定義される TM 上の接続 ∇^* は ∇ の h に関する双対接続と呼ばれる：$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla^*_X Z)$. ここで X, Y, Z は M 上の任意のベクトル場である．一方で，統計多様体に対する Lauritzen による同値な定義として擬リーマン多様体 (M, h) 上に対称 $(0, 3)$-テンソル T を与えるものがある [11].

(M, h, ∇) が統計多様体であるとき，∇ が平坦である（捩率および曲率が消える）ことは，∇^* が平坦であることと必要十分になる [12]．このような事実を受けて，双対平坦多様体を次で定義する．

定義 2.1　統計多様体 (M, h, ∇, ∇^*) が双対平坦多様体であるとは，∇ が平坦であるときをいう．また，このとき (h, ∇, ∇^*) を M 上の双対平坦構造という．

双対平坦多様体の概念はアファイン微分幾何学におけるヘッセ多様体の概念と同じものである（志摩 [17]）．

双対平坦多様体 (M, h, ∇, ∇^*) の持つ顕著な特徴は，計量 h が局所アファイン座標系においてポテンシャル関数のヘッセ行列として表現されることである．すなわち，次のような性質が成り立つ．

命題 2.2（e.g., [17]）　$p \in M$ とし，$\boldsymbol{x} = (x_1, \cdots, x_n)$ を p の周りの ∇ に関するアファイン座標系とする．$\partial_i := \frac{\partial}{\partial x_i}$ $(1 \leq i \leq n)$ とおく．このとき，アファイン座標系 $\boldsymbol{x}$ 上のある滑らかな関数 $f(\boldsymbol{x})$ が存在して以下が成り立つ：

(1)　$h(\partial_i, \partial_j) = \partial_i \partial_j f$.

(2)　任意の $1 \leq i \leq n$ に対し，$p_i := \partial_i f$ とおくと，$\boldsymbol{p} = (p_1, \cdots, p_n)$ は ∇^* に関するアファイン座標系になる．この座標系を $\boldsymbol{x}$ の双対座標系と呼ぶ．

(3)　アファイン座標系 $\boldsymbol{x}$ と $\boldsymbol{p}$ に対し，$h(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial p_j}) = \delta_{ij}$.

(4)　双対座標系 $\boldsymbol{p}$ 上のある滑らかな関数 φ が存在して，$f(\boldsymbol{x}) + \varphi(\boldsymbol{p}) - \boldsymbol{x}^T \boldsymbol{p} = 0$ を満たす．

(5)　Amari-Chentsov テンソル $T = \nabla h$ に対し，$T(\partial_i, \partial_j, \partial_k) = \partial_i \partial_j \partial_k f$.

上記命題の (4) における φ は f のルジャンドル変換，または双対ポテンシャル関数と呼ばれる．

情報幾何学の応用を考える上で，次のダイバージェンス $\mathcal{D}: M \times M \to \mathbb{R}$ は重要な役割を果たす：

$$\mathcal{D}(p,q) = f(\boldsymbol{x}(p)) + \varphi(\boldsymbol{p}(q)) - \boldsymbol{x}(p)^T \boldsymbol{p}(q).$$

ここで，$f(\boldsymbol{x})$ と $\varphi(\boldsymbol{p})$ は双対平坦多様体 M のポテンシャル関数とその双対ポテンシャル関数である（厳密には上の定義は M の対角集合のある近傍上で意味を持つ）．ダイバージェンスは一般には非対称であり距離関数の公理を満たさないが，双対平坦多様体上の二点の差を測る幾何的量としての役割を持ち，拡張ピタゴラスの定理および射影定理と密接に関わる重要なものである．

$U \subset \mathbb{R}^n$ を開集合とし，パラメータ $\theta = (\theta_1, \cdots, \theta_n) \in U$ を持つ統計モデル $M = \{p(\boldsymbol{x}|\theta)\}_{\theta \in U}$ に導入される幾何構造を考える．情報幾何学ではしばしば M 上に次で定義されるフィッシャー・ラオ計量 $h = [h_{ij}]$ と対称 $(0,3)$-テンソル $T = (T_{ijk})$ を考える：

$$h_{ij} = \mathbb{E}[(\partial_i \log p)(\partial_j \log p)] = \int (\partial_i \log p)(\partial_j \log p) p \, d\boldsymbol{x}, \tag{1}$$

$$T_{ijk} = \mathbb{E}[(\partial_i \log p)(\partial_j \log p)(\partial_k \log p)] = \int (\partial_i \log p)(\partial_j \log p)(\partial_k \log p) p \, d\boldsymbol{x}. \tag{2}$$

ここで $\partial_i := \frac{\partial}{\partial \theta_i}$ であり，フィッシャー・ラオ計量 h は半正定値であることに注意する．統計モデル M が座標系 θ を持つ可微分多様体とみなせるとき，(M,h,T) は統計多様体としての構造をもつ．一方で，あるクラスに属する統計モデル M に対し，その上に勝手なリーマン計量 h と対称 $(0,3)$-テンソル T を与えたとき，統計学に動機を持つある種の不変性（統計的不変性などと呼ばれる [2]）を統計多様体 (M,h,T) に課すと，h と T は上記の形に限られることが知られている [8, 6]．このような理由から情報幾何学ではしばしば統計モデルを上記のテンソル (1) および (2) が備わった統計多様体とみなす．

例 2.3（指数型分布族 [2]） $\mathbb{R}^n$ 上の確率密度関数 $p(\boldsymbol{x}|\theta) = \exp\{\boldsymbol{x}^T\theta - \psi(\theta)\}$（$U \subset \mathbb{R}^n$ は開集合）からなる族 $M := \{p(\boldsymbol{x}|\theta)\}_{\theta \in U}$ を指数型分布族という．ここで $\boldsymbol{x} = (x_1, \cdots, x_n)$ は確率変数（その測度を $d\mu$ とする），$\theta = (\theta_1, \cdots, \theta_n)$ はパラメータ，$\psi(\theta) = \log \int \exp\{\boldsymbol{x}^T\theta\} d\mu(\boldsymbol{x})$ は正規化関数である．M をパラメータ空間 U と同一視し，M を可微分多様体とみなす．$\partial_i := \frac{\partial}{\partial \theta_i}$ とおく．このとき M は $\psi(\theta)$ をポテンシャル関数とする双対平坦構造を持つ．実際，θ に対し期待値

$$\eta_i := \mathbb{E}[x_i|\theta] = \int x_i p(\boldsymbol{x}|\theta) d\mu(\boldsymbol{x}) = \partial_i \psi(\theta)$$

と分散共分散行列（これはフィッシャー行列であり正定値である）

$$h_{ij}(\theta) := \mathbb{E}[(x_i - \eta_i)(x_j - \eta_j)|\theta] = \mathbb{E}[(\partial_i \log p)(\partial_j \log p)|\theta] = \partial_i \partial_j \psi(\theta)$$

を考えると，$\eta=(\eta_1,\cdots,\eta_n)$ は計量 $h=[h_{ij}]$ に関する双対座標系を与える．すなわち，θ と η をアファイン座標系とする接続 ∇，∇^* を考えると，(M,h,∇,∇^*) は双対平坦多様体である．このとき Amari-Chentsov テンソルは式 (2) によるテンソルと一致していることが確かめられる．

例 2.4（ニューラルネットワーク (cf. [2])）　フィッシャー・ラオ計量が退化する例としてニューラルネットワークから定義される統計モデルを考える．$f:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$, $f(\boldsymbol{x},\boldsymbol{w})=f_w(\boldsymbol{x})$ は，$\boldsymbol{x}$ を入力，$\boldsymbol{w}$ を重みパラメータとするニューラルネットワークであるとする（$\boldsymbol{x}$ および $\boldsymbol{w}$ に関する非線形関数）．ニューラルネットワークの出力 $f_w(\boldsymbol{x})$ を平均 0，分散 1 のガウスノイズで摂動して得られる確率変数 y を考える：

$$p(y|\boldsymbol{x},\boldsymbol{w})=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(y-f_w(\boldsymbol{x}))^2\right\}.$$

統計モデル $M:=\{p(y|\boldsymbol{x},\boldsymbol{w})\}_{\boldsymbol{w}}$ に対し，フィッシャー行列 $F(\boldsymbol{w})=[F_{ij}(\boldsymbol{w})]$ と対称 $(0,3)$-テンソル $T(\boldsymbol{w})=(T_{ijk}(\boldsymbol{w}))$ は次のように計算される：

$$F_{ij}(\boldsymbol{w})=\frac{\partial f_w(\boldsymbol{x})}{\partial w_i}\frac{\partial f_w(\boldsymbol{x})}{\partial w_j},\quad T_{ijk}(\boldsymbol{w})=0.$$

従って，$\frac{\partial f_{\boldsymbol{w}}(\boldsymbol{x})}{\partial w_i}=0$ なるときにはフィッシャー計量は退化することがわかる．また，$T=0$ であることから M は双対的な構造を持たず，このとき M は自己双対であると呼ばれる [2]．このような構造に基づいた深層学習の解析として，Amari は自然勾配法を提案している [3]．

しかしながら，上記の枠組みでは分散を 1 に固定している一方で，分散 $s>0$ もパラメータに含める方が理論的には自然であろう（e.g., [7, 9]）．この場合，$(0,3)$-テンソル T は恒等的にゼロにはならないことが確かめられ，このとき M は双対的な構造を持つ．このような場合には真に特異点を許容する双対平坦構造，すなわち我々の導入する概ヘッセ構造による解析が必要となるであろう．

3　接触幾何学と概ヘッセ多様体の理論

我々の概ヘッセ多様体を導入するために，接触幾何学における用語の整理を行う（cf. [4, 5, 10]）．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，局所的に超平面場 ξ が $\theta\wedge(d\theta)^n\neq0$ なる 1 次微分形式 θ の核で表されるときをいう．このとき超平面場 ξ を N 上の接触構造，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，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，$\dim L=n$ であり N の各点 p において $T_pL\subset\xi_p$ なるときをいう．

$\mathbb{R}^{2n+1}=T^*\mathbb{R}^n\times\mathbb{R}$ に対し，標準的な座標系を $(\boldsymbol{x},\boldsymbol{p},z)$ とし，$T^*\mathbb{R}^n$ の底空間を $\mathbb{R}^n_{\boldsymbol{x}}$，ファイバーを $\mathbb{R}^n_{\boldsymbol{p}}$ と書き表すことにする．ここで，$\boldsymbol{x}$ は $\mathbb{R}^n_{\boldsymbol{x}}$ の座標系であり，$\boldsymbol{p}$ は $\mathbb{R}^n_{\boldsymbol{p}}$ の座

標系である．このとき，$\mathbb{R}^{2n+1}$ は

$$\theta = dz - \boldsymbol{p}^T d\boldsymbol{x} = dz - \sum_{i=1}^{n} p_i dx_i$$

を接触形式とする接触多様体になる．この接触多様体 $(\mathbb{R}^{2n+1}, \xi = \ker\theta)$ を標準接触多様体と呼ぶ.

次で定義される $\mathcal{L} : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$ は標準接触多様体上の接触同型写像である：

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

ここで，標準接触多様体上に定まる二つのルジャンドルファイブレーションを考える：

$$\begin{aligned} &\pi : \mathbb{R}^{2n+1} \to \mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}, \ \ (\boldsymbol{x}, \boldsymbol{p}, z) \mapsto (\boldsymbol{x}, z), \\ &\pi' := \pi \circ \mathcal{L} : \mathbb{R}^{2n+1} \to \mathbb{R}^n_{\boldsymbol{p}} \times \mathbb{R}, \ \ (\boldsymbol{x}, \boldsymbol{p}, z) \mapsto (\boldsymbol{p}, \boldsymbol{x}^T \boldsymbol{p} - z). \end{aligned}$$

このとき次の図式を標準接触多様体 $(\mathbb{R}^{2n+1}, \xi)$ 上のダブルファイブレーション構造と呼ぶ：

$$\mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}_z \xleftarrow{\ \pi\ } \mathbb{R}^{2n+1} \xrightarrow{\ \pi'\ } \mathbb{R}^n_{\boldsymbol{p}} \times \mathbb{R}_{z'}$$

ルジャンドル部分多様体 $L \subset \mathbb{R}^{2n+1}$ に対し,

$$\pi^e := \pi \circ \iota : L \to \mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}_z, \quad \pi^m := \pi' \circ \iota : L \to \mathbb{R}^n_{\boldsymbol{p}} \times \mathbb{R}_{z'}$$

をそれぞれ e/m-ルジャンドル写像と呼び，また，z 座標と z' 座標を無視して得られる写像 $\pi^e_1 : L \to \mathbb{R}^n_{\boldsymbol{x}}$, $\pi^m_1 : L \to \mathbb{R}^n_{\boldsymbol{p}}$ をそれぞれ e/m-ラグランジュ写像と呼ぶ．ここで $\iota : L \hookrightarrow \mathbb{R}^{2n+1}$ は包含写像である．また，e/m-ルジャンドル写像によるルジャンドル部分多様体の像

$$W_e(L) := \pi^e(L) \subset \mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}_z, \quad W_m(L) := \pi^m(L) \subset \mathbb{R}^n_{\boldsymbol{p}} \times \mathbb{R}_{z'}$$

をそれぞれ e/m-波面と呼ぶ.

L 上のベクトル束 $E\,(= E_L)$ を次で定める：

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

E_p を $p \in L$ におけるファイバーとする．L はルジャンドル部分多様体であるため，$d\pi^e(T_pL) \subset E_p$ である．従って，束写像

$$\Phi : TL \to E, \quad v_p \mapsto d\pi^e_p(v_p)$$

が意味をもつ．さらに，E 上のアファイン接続 ∇^E を次のように定める：$\widetilde{\nabla}$ を $\mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}$ 上の平坦アファイン接続，$\psi_p : \mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}_z \to E_p$ を z-軸に沿った線形射影とするとき，

$$\nabla^E_X \eta(p) := \psi_p \circ \widetilde{\nabla}_X \eta(p).$$

ここで X は L 上のベクトル場，η は E の切断である.

命題 3.1（[13]）　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

$$\nabla^E_X(\Phi(Y)) - \nabla^E_Y(\Phi(X)) = \Phi([X, Y])$$

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 [14, 15]．m-波面に対しても同様に連接接束 $(E', \Phi', \nabla^{E'})$ を定義する．

ベクトル束 E, E' は L 上で定められたものであるが，構成は $\mathbb{R}^{2n+1}$ 上で意味を持つものである．従って，それらは $\mathbb{R}^{2n+1}$ 上で定義され得る．このとき，超平面場 ξ は次の直和分解をもつことに注意する：

$$\xi_p = \ker d\pi'_p \oplus \ker d\pi_p \simeq E_p \oplus E'_p \simeq \mathbb{R}^n_{\boldsymbol{x}} \oplus \mathbb{R}^n_{\boldsymbol{p}}.$$

この分解により，ξ 上には (n, n) 型の擬リーマン計量

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

が自然に定まる．

定義 3.2（[13]）　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．

$\mathbb{R}^n_{\boldsymbol{x}} \times \mathbb{R}$ と $\mathbb{R}^n_{\boldsymbol{p}} \times \mathbb{R}$ 上のアファイン変換

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

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

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

を導く [13]．ここで，A は正則行列で，$A' = (A^T)^{-1}$, $\boldsymbol{b}' = A'\boldsymbol{c}$, $\boldsymbol{b} = A\boldsymbol{c}'$, $d' = \boldsymbol{b}'^T\boldsymbol{b} - d$ である．二つのルジャンドル部分多様体 $L_1, L_2 \subset \mathbb{R}^{2n+1}$ がアファインルジャンドル同値 $\mathcal{L}_F$ により移り合うとき，その上の概ヘッセ計量は保たれ，また，連接接束間の同型 $E_{L_1} \simeq E_{L_2}$, $E'_{L_1} \simeq E'_{L_2}$ が得られる．これによりそれぞれの連接接束上の平坦接続が同一視される．

定義 3.3（[13]）　ルジャンドル部分多様体の族 $\mathcal{U} = \{L_\alpha\}$, $L_\alpha \subset \mathbb{R}^{2n+1}$ とアファインルジャンドル同値を基にした張り合わせによる構成で得られる多様体 $(M, \mathcal{U})$ を概ヘッセ多様体と呼ぶ．このとき，各 L_α の構造から，M 上に退化し得る対称 $(0, 2)$-テンソル h と連接接束 (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．

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

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

定義 3.4（[13]） 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：

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

ここで，X, Y, Z は M 上のベクトル場である．

概ヘッセ計量 h がいたるところで非退化であるとき，概ヘッセ多様体は双対平坦多様体になり，標準 3 次テンソルは Amari-Chentsov の 3 次テンソルに一致することが確かめられる [13].

4 ダイバージェンスと拡張ピタゴラスの定理

ルジャンドル部分多様体 $L \subset \mathbb{R}^{2n+1}$ に対し，L 上の正準ダイバージェンス $\mathcal{D}_L : L \times L \to \mathbb{R}$ を次で定義する：

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

ここで，$\mathbb{R}^{2n+1}$ の標準座標系 $(\boldsymbol{x}, \boldsymbol{p}, z)$ を用いて $p, q \in L \subset \mathbb{R}^{2n+1}$ を表しており，$z' = \boldsymbol{x}^T \boldsymbol{p} - z$ である．L 上で定義された正準ダイバージェンスはアファインルジャンドル同値で不変になることが確かめられる [13]．これより概ヘッセ多様体 $(M, h, (E, \Phi, \nabla^E)$, $(E', \Phi', \nabla^{E'}))$ 上に自然に正準ダイバージェンス $\mathcal{D}_M$ が定義される．

$c : I \to M$ を曲線とし，$\dot{c} : I \to TM$ を速度ベクトル場とする（$I \subset \mathbb{R}$ は区間）．

定義 4.1（[13]） 曲線 c が m-曲線であるとは，c ははめ込みであり，各 $t \in I$ に対し，$E'_{c(t)}$ 上のベクトル

$$\Phi' \circ \dot{c}(t),\ \ \nabla^{E'}_{\dot{c}}(\Phi' \circ \dot{c})(t),\ \ (\nabla^{E'}_{\dot{c}})^2(\Phi' \circ \dot{c})(t), \cdots$$

は同時には消えず，任意の二つが線形従属になるときをいう．e-曲線の概念も (E, Φ, ∇^E) を用いて同様に定義する．

局所モデル L_α における e-ラグランジュ写像 $\pi^e_1 : L_\alpha \to \mathbb{R}^n_{\boldsymbol{x}}$ による e-曲線 c_e の像 $\boldsymbol{x}(t) := \pi^e_1 \circ c_e(t)$ の軌跡は $\mathbb{R}^n_{\boldsymbol{x}}$ 上のある直線を定めることが簡単に確かめられる．m-曲線 c_m もまた同様に $\mathbb{R}^n_{\boldsymbol{p}}$ 上の直線を定める．それぞれの直線の方向ベクトルとして $\boldsymbol{e}, \boldsymbol{m}$ をとる．次の定義は局所モデルや方向ベクトルの取り方によらない．

定義 4.2（[13]） c_e を e-曲線，c_m を m-曲線とし，$\boldsymbol{e}$ および $\boldsymbol{m}$ を上記の通りとする．また，$S \subset M$ を部分多様体とする．c_e と c_m が $q \in L_\alpha$ において交わっているとき，これらが q で真に直交するとは，$\boldsymbol{e}^T\boldsymbol{m} = 0$ が成り立つときをいう．さらに，q で c_e と S が交わるとき，c_e と S が直交するとは，任意の $v \in T_qS$ に対し，$\boldsymbol{e}^T d\boldsymbol{p}(v) = 0$ が成り立つときをいう．m-曲線 c_m と S の直交性も同様にして定義する．

上記の直交性の定義は，q において概ヘッセ計量が非退化であるときは従来の意味の（計量による）直交性の定義と同じである．しかしながら q が特異点である場合これらの定義は一般には異なる．実際，e/m-曲線の q における速度ベクトルはゼロになる可能性があり，従って方向ベクトルはその速度ベクトルからは定まらない．

概ヘッセ多様体上の正準ダイバージェンスと e/m-曲線に対し，次の拡張ピタゴラスの定理と射影定理が成立する．

定理 4.3（拡張ピタゴラスの定理 [13]） 相異なる 3 点 $p, q, r \in M$ に対し，p と q は e-曲線 c_e で結ばれており，q と r は m-曲線 c_m で結ばれているとする．c_e と c_m が q で真に直交しているとき，

$$\mathcal{D}_M(p,q) + \mathcal{D}_M(q,r) = \mathcal{D}_M(p,r)$$

が成り立つ．

定理 4.4（射影定理 [13]） $S \subset M$ を部分多様体とし，$p \in M, q \in S$ とする．p と q は m-曲線 c_m により結ばれているとする．このとき，関数 $\mathcal{D}_M(\cdot, p) : S \to \mathbb{R}$ が q で極値をとることと，m-曲線 c_m が S と直交することは同値である．関数 $\mathcal{D}_M(p, \cdot)$ と e-曲線を考えても同様のことが成り立つ．

参考文献

[1] S. Amari and H. Nagaoka, *Methods of information geometry*, A.M.S., Oxford Univ. Press (2000).

[2] S. Amari, *Information Geometry and Its Application*, Applied Math. Sci., 194, Springer (2016).

[3] S. Amari and T. Ozeki, *Differential and Algebraic Geometry of Multilayer Perceptrons*, IEICE Trans., **84** (2001), 31–38.

[4] V.I. Arnol'd, *Mathematical Methods of Classical Mechanics*, 2nd Edition, Grad. Texts Math. **60**, Springer-Verlag (1989).

[5] V.I. Arnol'd, S.M. Gusein-Zade and A.N. Varchenko, *Singularities of Differentiable Maps I*, Monographs in Math. **82**, Birkhäuser (1986).

[6] N. Ay, J. Jost, H. Lê and L. Schwachhöfer, *Information geometry and sufficient statistics*, Probab. Theory Related Fields, **162** (2015), 327–364.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).

[8] N.N. Chentsov, *Statistical Decision Rules and Optimal Inference*, AMS. (1982). (Originally published in Russian, Nauka, Moscow, 1972).

[9] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).

[10] 泉屋周一，石川剛郎，応用特異点論，共立出版 (1998).

[11] S. L. Lauritzen, *Statistical manifolds*, IMS Lecture Notes - Monograph Series, (1987), 96–163.

[12] H. Matsuzoe, *Statistical manifolds and affine differential geometry*, Advanced Stud. Pure Math. **57** (2010), 303–321.

[13] N. Nakajima and T. Ohmoto, *The dually flat structure for singular models*, Inf. Geom., **4** (2021), 31–64.

[14] K. Saji, M. Umehara, K. Yamada, *The geometry of fronts*, Ann. of Math., **169** (2009), 491–529.

[15] 佐治健太郎，梅原雅顕，山田光太郎，特異点をもつ曲線と曲面の微分幾何学，丸善出版 (2017).

[16] H. Shima, *The geometry of Hessian Structures*, World Scientific (2007).

[17] 志磨裕彦，ヘッセ幾何学，裳華房 (2001).

[18] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, Cambridge Univ. Press (2008).