

V ポテンシャルから導かれる 正定値対称行列空間の幾何とその応用

小原 敦美¹ 江口 真透²

1 はじめに

正定値対称行列のなす空間はリーマン対称空間あるいはジョルダン代数の例として古くから幾何的、代数的に調べられ(例えば [9, 11, 16, 32, 38]), その結果は統計科学, 物理学, 凸最適化などの多くの分野にも広く応用されている.

この空間のリーマン幾何, さらに情報幾何 [1, 30, 2] やヘッセ幾何 [34] では, 正定値対称行列 P に対して定義された次の関数 $\varphi(P) = -k \log \det P$, $k > 0$ が中心的な役割を果たす. 例えばその二階微分 (ヘッセ行列), 三階微分からリーマン計量, 複数のアファイン接続といった幾何学量 (概念) が定義され, 代数的な構造と協調した豊かな結果を生み出す. このように定義された幾何構造の大きな特徴の一つは, この空間に作用する自己同型群, すなわち合同変換群に対する幾何構造の不変性である [32, 38]. この群不変性あるいは空間の等質性は, 多変量解析 [22] や半正定値計画 (Semi-Definite Program) [25, 39] と呼ばれる数理計画などの応用分野でも活用され重要な働きを示す. この空間に関する特に情報幾何を意識した応用研究として, 例えば安定行列の構造 [27], 行列平均 [26], SDP の計算複雑度と曲率の関係 [14], 準ニュートン法の更新公式の拡張 [13] をあげておく.

本稿では第一に, 上記のような幾何学量の源, すなわちポテンシャルである関数 φ の変形を考え, そこから得られる情報幾何を考察する. ここで考えるのは $s > 0$ で定義された滑らかな関数 $V(s)$ で決まる $\varphi^{(V)} = V(\det P)$ の形のポテンシャルである. われわれはこれを V -ポテンシャルと呼ぶことにする. 上記の標準的な $\varphi = \varphi^{(-\log)}$ な場合との重要な違いとして, 一般の V -ポテンシャルから導かれる幾何構造は特殊線形群による合同変換でのみ不変となる結果が得られる.

上のポテンシャル変形の応用として, 本稿では第二に非ガウス型の多変量分布族の情報幾何を考察する. U をある実数区間上で正の導関数を持つ滑らかな凸関数として固定し, 正定値対称行列 P で指定される次のような形の密度関数

$$f(x, P) = u\left(-\frac{1}{2}x^T P x - c_U(\det P)\right), \quad u(s) = \frac{dU(s)}{ds},$$

を考え, U -モデル [4, 23] とよぶ. ただし, $c_U(\det P)$ は正規化定数である. この非指数型

¹福井大学大学院工学研究科 ohara@fuee.u-fukui.ac.jp

²統計数理研究所 eguchi@ism.ac.jp

の密度関数をもつ U -モデルに対して, KL ダイバージェンスの対応物として U -ダイバージェンス [4, 23] が定義でき, この U -モデル上にも自然な情報幾何を導入される.

各密度関数 f は P で指定されているので, このように定められた幾何は正定値対称行列上の情報幾何とも見なすことができる. そこで, この幾何と先に述べた V -ポテンシャルが定める情報幾何との対応関係を調べる.

なお, このような一般化ガウス分布とも呼ばれる確率分布に関わる統計科学や統計力学への応用・進展は, [10, 3, 24, 36] も参照されたい.

本稿の構成は以下の通りである: 2 節で情報幾何及び統計多様体に関して, 必要な事柄をまとめる. 3 節で V -ポテンシャルから求まるリーマン計量とその正定値性の条件を導く. 4 節では双対接続を計算し, 3 節と合わせて得られた情報幾何構造が一般には特殊線形群の合同変換のみで不変であることを見る. ただし, べき型ポテンシャルの場合は, 統計的な推論で基本的な性質は一般化線形群での合同変換でも不変となる. 5 節で, V -ポテンシャルが定める情報幾何について議論する. 特に行列式の値が一定である超曲面が定曲率性などの良い性質を持ち, この超曲面に関して拡張された形のダイバージェンス分解定理が成立することを示す. 6 節では, U -ダイバージェンスから U -モデルを通して正定値対称行列空間に導入された情報幾何と V -ポテンシャルから定められた情報幾何の対応関係を与える. 7 節で, 特に q -ガウス分布族 [24, 36] の情報幾何を定義するために V を計算し, この分布族の情報幾何構造が 4 節の意味で一般化線形群でも不変であることを見る.

なお, 本稿は既発表論文 [28, 29] の概説であり, 詳細や証明についてはこれらを参照されたい.

2 準備: 情報幾何と統計多様体

ここでは, 情報幾何 [2] (あるいはヘッセ幾何 [34]) に現れる統計多様体 [20, 18, 19] の双対的な幾何構造に関する基本的な結果を, 本稿で用いる事柄に限定して簡単にまとめる.

多様体 M 上のベクトル場全体を $\mathcal{X}(M)$ と表す. M で定義された振れのないアファイン接続 ∇ と (擬) リーマン計量 g の組 (M, ∇, g) が統計 (あるいは *Codazzi*) 多様体とは, 振れのない別のアファイン接続 ∇^* が存在し次の関係

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (1)$$

が任意の $X, Y, Z \in \mathcal{X}(M)$ で成り立つことである. このとき, ∇ と ∇^* は g に関して互いに双対であるといい, (g, ∇, ∇^*) を M 上の双対構造と呼ぶ.

ある実定数 k に対し統計多様体 (M, ∇, g) が定曲率 k であるとは, ∇ の曲率テンソル R

が

$$R(X, Y)Z = k\{g(Y, Z)X - g(X, Z)Y\}.$$

を満たすことである。 $k=0$ のとき、定曲率統計多様体は平坦というが、このとき自動的に ∇^* の曲率テンソル R^* もゼロになるので双対平坦とも呼ぶ [2].

ある実定数 α に対し二つの統計多様体 (M, ∇, g) と (M, ∇', g') が α -共形同値とは、 M 上の関数 ϕ が存在し次の関係が成り立つことである：

$$\begin{aligned} g'(X, Y) &= e^\phi g(X, Y) \\ g(\nabla'_X Y, Z) &= g(\nabla_X Y, Z) - \frac{1+\alpha}{2} d\phi(Z)g(X, Y) \\ &\quad + \frac{1-\alpha}{2} \{d\phi(X)g(Y, Z) + d\phi(Y)g(X, Z)\}. \end{aligned}$$

(M, ∇, g) と (M, ∇', g') が α -共形同値であることと (M, ∇^*, g) と (M, ∇'^*, g') が $-\alpha$ -共形同値であることは必要十分である [19]. 統計多様体 (M, ∇, g) は、それがある平坦な別の統計多様体に α -共形同値なとき、 α -共形平坦と呼ばれる。

$\{x^1, \dots, x^n\}$ を \mathbf{R}^n のある基底ベクトルに関するアファイン座標系とし、 ∇ をその標準的なアファイン接続、すなわち $\nabla_{\partial/\partial x^i} \partial/\partial x^j = 0$ を満たすものとする。ある領域 $\Omega \subset \mathbf{R}^n$ 上の関数 φ のヘッセ行列から定まる対称テンソル $g^{(\varphi)} = \sum (\partial^2 \varphi / \partial x^i \partial x^j) dx^i dx^j$ が非退化なとき、 $g^{(\varphi)}$ を (擬) リーマン計量、 ∇ をアファイン接続と見なして $(\Omega, \nabla, g^{(\varphi)})$ をヘッセ領域と呼ぶ。ヘッセ領域 $(\Omega, \nabla, g^{(\varphi)})$ は平坦な統計多様体であり、逆に平坦統計多様体は局所的にヘッセ領域でもある [2, 34].

\mathbf{R}^n の元 x とその双対空間 \mathbf{R}_n^* の元 y^* に対しそれらのペアリングを $\langle y^*, x \rangle$ と記し、 $X \in T_p \Omega$ と $X \in \mathbf{R}^n$ を同一視することで、 $p \in \Omega$ での勾配写像 $\text{grad} \varphi$ を $\langle \text{grad} \varphi, X \rangle = d\varphi(X)$ と定める。 $g^{(\varphi)}$ は非退化なので、 $p = (x^1, \dots, x^n)$ での勾配写像は、双対座標系 $\{x_1^*, \dots, x_n^*\}$ で

$$\text{grad} \varphi : x \mapsto x^* = (x_1^*, \dots, x_n^*), \quad x_i^* = \partial \varphi / \partial x^i$$

と表され、局所的に可逆である。よって、共役関数 φ^*

$$\varphi^*(x^*) = \langle x^*, (\text{grad} \varphi)^{-1}(x^*) \rangle - \varphi((\text{grad} \varphi)^{-1}(x^*)).$$

を局所的に定義できる。 (x, φ) から (x^*, φ^*) への写像をルジャンドル (Legendre) 変換という。また、 φ と φ^* はポテンシャルと呼ばれる。 $\{x_1^*, \dots, x_n^*\}$ を Ω の別な座標と見なすと、ヘッセ領域 $(\Omega, \nabla, g^{(\varphi)})$ に対し、 ∇ の双対接続 ${}^* \nabla^{(\varphi)}$ は $\{x_1^*, \dots, x_n^*\}$ の平坦接続として特徴付けられる。さらに、 $g^{(\varphi^*)} = \sum (\partial^2 \varphi^* / \partial x_i^* \partial x_j^*) dx_i^* dx_j^*$ は局所的に $g^{(\varphi)}$ に一致し、 $(\Omega, {}^* \nabla^{(\varphi)}, g^{(\varphi)} = g^{(\varphi^*)})$ も (局所的に) ヘッセ領域である [2].

φ の共役関数 φ^* が Ω で大域的に定義できるとき, $(\Omega, \nabla, g^{(\varphi)})$ の標準的ダイバージェンス $D^{(\varphi)}$

$$D^{(\varphi)}(p, q) = \varphi(x(p)) + \varphi^*(x^*(q)) - \langle x^*(q), x(p) \rangle$$

が二点 $p, q \in \Omega$ に対して得られる. ここで, $x(\cdot)$ と $x^*(\cdot)$ はそれぞれ点 \cdot の x -座標, x^* -座標を表す. $g^{(\varphi)}$ が正定値で Ω が凸ならば

$$\text{i) } D^{(\varphi)}(p, q) \geq 0, \quad \text{ii) } D^{(\varphi)}(p, q) = 0 \Leftrightarrow p = q \quad (2)$$

である. なぜなら $\sup_{p \in \Omega} \{\langle x^*(q), x(p) \rangle - \varphi(x(p))\}$ の最大値 $\varphi^*(x^*(q))$ は, $x(p) = (\text{grad} \varphi)^{-1}(x^*(q))$ で達成されるからである.

逆に任意の二点 $p, q \in M$ に対し (2) の二条件を満たす関数 $D : M \times M \rightarrow \mathbf{R}$ から半正定値な計量 $g^{(D)}$ と二つのアフライン接続 $\nabla^{(D)}, * \nabla^{(D)}$

$$\begin{aligned} g^{(D)}(X, Y) &= -D(X|Y), \\ g^{(D)}(\nabla_X^{(D)} Y, Z) &= -D(XY|Z), \quad g^{(D)}(* \nabla_X^{(D)} Y, Z) = -D(Z|XY), \end{aligned} \quad (3)$$

を $X, Y, Z \in \mathcal{X}(M)$ に対して定める. ただし, 右辺の記号は $X_i, Y_j \in \mathcal{X}(M)$ に対して

$$D(X_1 \cdots X_n | Y_1 \cdots Y_m)(p) = (X_1)_p \cdots (X_n)_p (Y_1)_q \cdots (Y_m)_q D(p, q)|_{p=q}$$

を意味する. もし $g^{(D)}$ が正定値ならば, D は M 上のダイバージェンスまたはコントラスト関数と呼ばれる. 次の結果は 6 節で用いられる:

Proposition 1 ([4, 2]) D が M 上のダイバージェンスならば, $(M, \nabla^{(D)}, g^{(D)})$ は統計多様体で, その双対接続は $* \nabla^{(D)}$ である.

3 V-ポテンシャルとリーマン計量

$n \times n$ の実対称行列の空間 $Sym(n, \mathbf{R})$ を考え, $X, Y \in Sym(n, \mathbf{R})$ に対する内積を $\langle X|Y \rangle = \text{tr}(XY)$ とする. 双対空間 $Sym(n, \mathbf{R})^*$ の元 Y^* を同じ記号を用いて $Sym(n, \mathbf{R})$ の元 Y^* で $\langle Y^*, X \rangle = \langle Y^*|X \rangle$ により同一視する. $\{E_1, \dots, E_{n(n+1)/2}\}$ を $Sym(n, \mathbf{R})$ の基底行列とし, この基底行列に関するアフライン座標系を $\{x^1, \dots, x^{n(n+1)/2}\}$, 標準的な平坦アフライン接続を ∇ とする. $PD(n, \mathbf{R})$ で $n \times n$ 正定値対称行列の集合がなす凸錐を表す.

$\mathcal{X}(PD(n, \mathbf{R}))$ と $T_P PD(n, \mathbf{R})$ で, それぞれ $PD(n, \mathbf{R})$ 上の接ベクトル場全体, $P \in PD(n, \mathbf{R})$ での接ベクトル空間を表す. E_i と自然基底 $(\partial/\partial x^i)_P$ を同一視することで, 接ベクトル $X_P \in T_P PD(n, \mathbf{R})$ を $X \in Sym(n, \mathbf{R})$ と表す. 同様に, $Sym(n, \mathbf{R})$ に値を持つ

$PD(n, \mathbf{R})$ 上の滑らかな関数 X は, E_i を定数関数と見なして $\partial/\partial x^i$ と同一視することで, $X \in \mathcal{X}(PD(n, \mathbf{R}))$ と考えることができる.

τ_G で G による合同変換, すなわち, $\tau_G X = GXG^T$ を表すことにする. τ_G の微分を τ_{G*} と表す. G が正則なら, τ_G は $PD(n, \mathbf{R})$ に推移的に作用する自己同型群の元である.

3節と4節で, あるクラスのポテンシャル関数からヘッセ領域として導入される $PD(n, \mathbf{R})$ 上の双対平坦構造について考える.

Definition 1 $V(s)$ を正実数 $s > 0$ 上の滑らかな関数とする. 次のように定義される $PD(n, \mathbf{R})$ 上の関数

$$\varphi^{(V)}(P) = V(\det P) \quad (4)$$

を V -ポテンシャルと呼ぶ.

$s > 0$ 上の関数 $\nu_i(s), i = 1, 2, 3$ を

$$\nu_i(s) = \frac{d\nu_{i-1}(s)}{ds} s, \quad i = 1, 2, 3, \quad \text{ただし } \nu_0(s) = V(s)$$

と定義し, 以降 $V(s)$ は次の二つの条件を満たすと仮定する:

$$\text{i) } \nu_1(s) < 0 \quad (s > 0), \quad \text{ii) } \beta^{(V)}(s) = \frac{\nu_2(s)}{\nu_1(s)} < \frac{1}{n} \quad (s > 0). \quad (5)$$

この2条件は, $\varphi^{(V)}(P)$ のヘッセ行列が $PD(n, \mathbf{R})$ で正定値であるための必要十分条件となることが示される. 最初の条件 $\nu_1(s) < 0 (s > 0)$ から, $V(s)$ は $s > 0$ で単調減少である. (5) を満たす $V(s)$ の重要な例は, $-\log s$ や $c_1 + c_2 s^\beta$ (ただし, 実定数 c_1, c_2, β は $c_2 \beta < 0$ と $\beta < 1/n$ を満たす) である. また別な例 $V(s) = c \log(cs + 1) - \log s$ (ただし $0 \leq c < 1$) は, ダイバージェンスの観点から準ニュートン法の更新則を導出する際に現れる [13].

行列式の微分に関する関係式 $\text{grad det } P = (\det P)P^{-1}$ を用いて, 勾配写像 $\text{grad}\varphi^{(V)}$ と1形式 $d\varphi^{(V)}$ がそれぞれ次のように求まる:

$$\text{grad}\varphi^{(V)} : P \mapsto P^* = \nu_1(\det P)P^{-1}, \quad (6)$$

$$d\varphi^{(V)} : X \mapsto d\varphi^{(V)}(X) = \nu_1(\det P) \text{tr}(P^{-1}X). \quad (7)$$

また $P \in PD(n, \mathbf{R})$ でポテンシャル $\varphi^{(V)}$ のヘッセ行列が定めるリーマン計量 $g_P^{(V)}$ は次のようになる:

$$\begin{aligned} g_P^{(V)}(X, Y) &= d(d\varphi^{(V)}(X))(Y) \\ &= -\nu_1(\det P) \text{tr}(P^{-1}XP^{-1}Y) + \nu_2(\det P) \text{tr}(P^{-1}X) \text{tr}(P^{-1}Y). \end{aligned}$$

Theorem 1 $g^{(V)}$ が $PD(n, \mathbf{R})$ で正定値である必要十分条件は (5) が成立することである。

(証明は [29] を参照)

Remark 1 $g^{(V)}$ は $SL(n, \mathbf{R})$ -不変, すなわち任意の $G \in SL(n, \mathbf{R})$ に対して, $g_{P'}^{(V)}(X', Y') = g_P^{(V)}(X, Y)$ (ただし, $P' = \tau_G P, X' = \tau_{G*} X, Y' = \tau_{G*} Y$) が成り立つ。

$\varphi^{(V)}$ の共役関数 $\varphi^{(V)*}$ は

$$\varphi^{(V)*}(P^*) = \sup_P \{ \langle P^*, P \rangle - \varphi^{(V)}(P) \} \quad (8)$$

である。極値条件

$$P^* = \text{grad} \varphi^{(V)}(P) = \nu_1(\det P) P^{-1}$$

と $g^{(V)}$ の正定値性より $\text{grad} \varphi^{(V)}$ が可逆となるので, $\varphi^{(V)*}$ は P を用いて次のように表せる:

$$\varphi^{(V)*}(P^*) = n \nu_1(\det P) - \varphi^{(V)}(P). \quad (9)$$

したがって, $(PD(n, \mathbf{R}), \nabla, g^{(V)})$ の正準ダイバージェンス $D^{(V)}$ は次のようになる:

$$\begin{aligned} D^{(V)}(P, Q) &= \varphi^{(V)}(P) + \varphi^{(V)*}(Q^*) - \langle Q^*, P \rangle \\ &= V(\det P) - V(\det Q) + \langle Q^*, Q - P \rangle. \end{aligned} \quad (10)$$

4 V-ポテンシャルから導かれる双対接続

∇ を $Sym(n, \mathbf{R})$ の正準な平坦接続とする。 $g^{(V)}$ をリーマン計量とする $PD(n, \mathbf{R})$ の双対平坦構造を見るために 2 節で与えた双対接続 $*\nabla^{(V)}$ を求める。

次式を満たす $PD(n, \mathbf{R})$ 上の滑らかな曲線 $\gamma = \{P_t \mid -\epsilon < t < \epsilon\}$ を考える:

$$(P_t)_{t=0} = P \in PD(n, \mathbf{R}), \quad \left(\frac{dP_t}{dt} \right)_{t=0} = X \in T_P PD(n, \mathbf{R}).$$

Lemma 1 勾配写像の微分 $\text{grad} \varphi^{(V)}$ は次式で与えられる:

$$(\text{grad} \varphi^{(V)})_* : X \mapsto \nu_2(\det P) \text{tr}(P^{-1} X) P^{-1} - \nu_1(\det P) P^{-1} X P^{-1}.$$

Proof ルジャンドル変換 $P_t^* = \text{grad} \varphi^{(V)}(P_t)$ を曲線 γ に沿って微分すると

$$\begin{aligned} \frac{dP_t^*}{dt} &= \frac{d\nu_1(\det P_t)}{dt} P_t^{-1} - \nu_1(\det P_t) P_t^{-1} \frac{dP_t}{dt} P_t^{-1} \\ &= \nu_1'(\det P_t) \frac{d \det P_t}{dt} P_t^{-1} - \nu_1(\det P_t) P_t^{-1} \frac{dP_t}{dt} P_t^{-1} \\ &= \nu_1'(\det P_t) \det P_t \text{tr} \left(P_t^{-1} \frac{dP_t}{dt} \right) P_t^{-1} - \nu_1(\det P_t) P_t^{-1} \frac{dP_t}{dt} P_t^{-1}, \end{aligned}$$

ただし $\nu_1' = d\nu_1/ds$. 両辺に $t = 0$ を代入すると題意が従う。 \square

Theorem 2 π_t を接続 $*\nabla^{(V)}$ の平行移動作用素とする. . 接ベクトル $Y (= \pi_0(Y)) \in T_P PD(n, \mathbf{R})$ の曲線 γ に沿った平行移動 $\pi_t(Y)$ は次式を満たす:

$$\left(\frac{d\pi_t(Y)}{dt} \right)_{t=0} = XP^{-1}Y + YP^{-1}X + \Phi(X, Y, P) + \Phi^\perp(X, Y, P),$$

ただし

$$\Phi(X, Y, P) = \frac{\nu_2(s) \operatorname{tr}(P^{-1}X)}{\nu_1(s)} Y + \frac{\nu_2(s) \operatorname{tr}(P^{-1}Y)}{\nu_1(s)} X, \quad (11)$$

$$\Phi^\perp(X, Y, P) = \rho P, \quad (12)$$

$$\rho = \frac{\{\nu_3(s)\nu_1(s) - 2\nu_2^2(s)\} \operatorname{tr}(P^{-1}X) \operatorname{tr}(P^{-1}Y) + \nu_2(s)\nu_1(s) \operatorname{tr}(P^{-1}XP^{-1}Y)}{\nu_1(s)\{\nu_1(s) - n\nu_2(s)\}}$$

で, $s = \det P$ である.

(証明は [29] を参照)

Remark 2 ρ の分母 $\nu_1(s)\{\nu_1(s) - n\nu_2(s)\}$ は, (5) の仮定より常に正である.

Corollary 1 3節の最初に述べた同一視で, E_i を接ベクトル場の自然基底 $\partial/\partial x^i$ を表すことにすると, 双対接続 $*\nabla^{(V)}$ の P での共変微分は以下のように表される:

$$\left(*\nabla_{\frac{\partial}{\partial x^i}}^{(V)} \frac{\partial}{\partial x^j} \right)_P = -E_i P^{-1} E_j - E_j P^{-1} E_i - \Phi(E_i, E_j, P) - \Phi^\perp(E_i, E_j, P),$$

Proof) Theorem 2 と共変微分の定義, すなわち

$$\left(*\nabla_{\frac{\partial}{\partial x^i}}^{(V)} \frac{\partial}{\partial x^j} \right)_P = \left(\frac{d\pi_{-t}((\partial/\partial x^j)_{P_t})}{dt} \right)_{t=0}, \quad P_t \in PD(n, \mathbf{R}), P = P_0.$$

から得られる. □

Remark 3 系 1 から, 双対接続 $*\nabla^{(V)}$ は一般に $SL(n, \mathbf{R})$ -不変, すなわち任意の $G \in SL(n, \mathbf{R})$ に対して

$$\tau_{G*}(\nabla_X Y)_P = (\nabla_{X'} Y')_{P'}, \quad \tau_{G*} \left(*\nabla_X^{(V)} Y \right)_P = \left(*\nabla_{X'}^{(V)} Y' \right)_{P'}$$

が成立する (∇ は $GL(n, \mathbf{R})$ -不変). ただし $P' = \tau_G P, X' = \tau_{G*} X, Y' = \tau_{G*} Y$ である. しかし, $V(s) = c_1 + c_2 s^\beta$ (c_1, c_2, β は実定数) で定まるべき型ポテンシャル $\varphi^{(V)}$ から導かれる $*\nabla^{(V)}$ は $GL(n, \mathbf{R})$ -不変となる. 加えてこのとき, $g^{(V)}$ 自身は $GL(n, \mathbf{R})$ -不変ではないが, $g^{(V)}$ に関する“直交性”は $GL(n, \mathbf{R})$ -不変となることに注意しよう. したがって, 様々な応用で非常に重要となるダイバージェンスに関するピタゴラスの定理や ∇ -射影と

* $\nabla^{(V)}$ -射影の概念 [1, 2] はべき型ポテンシャルに関しては $GL(n, \mathbf{R})$ -不変である。逆に、この意味の $GL(n, \mathbf{R})$ -不変性があるのは、 $-k \log s$, $k > 0$ を除いてべき型ポテンシャルのみである [28]

次のような正規化 $V(1) = 0$, $\nu_1(1) = -1$, すなわち $V(s) = (1 - s^\beta)/\beta$ を施したべき型ポテンシャル $\varphi^{(V)}$ を特に β ポテンシャルと呼ぶ。この場合, $\nu_1(s) = -s^\beta$, $\nu_2(s) = -\beta s^\beta$, $\nu_3(s) = -\beta^2 s^\beta$, $\beta^{(V)}(s) = \beta$ となる。特に $\beta \rightarrow 0$ とすると $V(s) = -\log s$ となり, 対数特性関数 $\varphi^{(-\log)}(P) = -\log \det P$ から導かれる $PD(n, \mathbf{R})$ 上の標準的な双対平坦構造 [30] が回復される。べき型のポテンシャルに関する詳細な議論は [7] も参照されたい。

5 行列式が一定値の超曲面

正定数 k に対して $V(s) = -k \log s$ としたとき, 対応する V -ポテンシャルから定義される双対幾何構造は様々な意味で標準的 [2, 30] で, 特に $G \in GL(n, \mathbf{R})$ の自己同型群 τ_G の作用に対し不変であった。一方, 前節の議論で (5) を満たす一般的な $V(s)$ に対しては, $PD(n, \mathbf{R})$ の双対幾何構造 $(g^{(V)}, \nabla, * \nabla^{(V)})$ は, この性質は失われ $G \in SL(n, \mathbf{R})$ なる τ_G の作用に対してのみ不変であった。

ここでは, V -ポテンシャルの定める $SL(n, \mathbf{R})$ -不変な双対幾何に関する結果の内, 興味深いと思われるものをいくつか紹介する。証明やその他の結果については [29] を参照されたい。

5.1 葉層構造

実パラメータ $s > 0$ で指定される超曲面 \mathcal{L}_s を定義し, $PD(n, \mathbf{R})$ の葉層構造を考える:

$$PD(n, \mathbf{R}) = \bigcup_{s>0} \mathcal{L}_s, \quad \mathcal{L}_s = \{P | P > 0, \det P = s\}.$$

\mathcal{L}_s の P での接ベクトル空間は

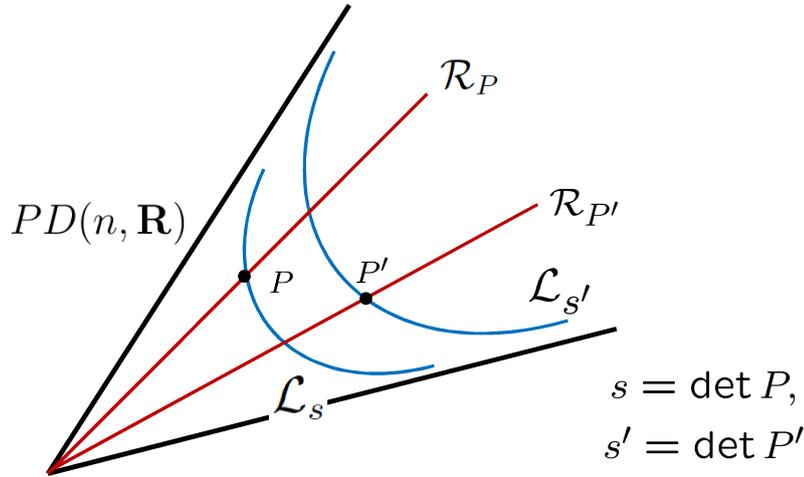
$$T_P \mathcal{L}_s = \{X | \operatorname{tr}(P^{-1}X) = 0, X = X^T\} = \{P^{1/2}Y P^{1/2} | \operatorname{tr} Y = 0, Y = Y^T\}.$$

のようになる。また原点から伸びる $P \in PD(n, \mathbf{R})$ を通る半直線を $\mathcal{R}_P = \{Q | Q = \lambda P, 0 < \lambda \in \mathbf{R}\}$ と表すと

$$PD(n, \mathbf{R}) = \bigcup_{P \in \mathcal{L}_s} \mathcal{R}_P.$$

である (Figure 1 参照)。

Proposition 2 [29] $P \in \mathcal{L}_s$ とする。(5) を満たす任意の V に対し, \mathcal{L}_s と \mathcal{R}_P は P で $g^{(V)}$ に関して直交する。

Figure 1: $PD(n, \mathbf{R})$ の葉層構造

5.2 部分統計多様体 \mathcal{L}_s の幾何

次に, $PD(n, \mathbf{R})$ の双対平坦構造 $(g^{(V)}, \nabla, * \nabla^{(V)})$ が超曲面 \mathcal{L}_s に誘導するリーマン計量を $\tilde{g}^{(V)}$, 互いに双対な接続を $\tilde{\nabla}, * \tilde{\nabla}^{(V)}$ と表し, 統計多様体 $(\mathcal{L}_s, \tilde{\nabla}, \tilde{g}^{(V)})$ の幾何を考察する. V に依存する $\tilde{g}^{(V)}$ と $* \tilde{\nabla}^{(V)}$ は, 次のように計算される [29]:

$$\tilde{g}^{(V)} = -\nu_1(s) \tilde{g}^{(-\log)}, \quad * \tilde{\nabla}^{(V)} = * \tilde{\nabla}^{(-\log)}$$

このように $\tilde{g}^{(V)}$ に関する $\tilde{\nabla}$ の双対接続 $* \tilde{\nabla}^{(V)}$ は, 実は \mathcal{L}_s 上では V の選び方によらず共通となるので, 以後これを単に $\tilde{\nabla}^*$ と表す.

さて, 統計多様体 $(\mathcal{L}_s, \tilde{\nabla}, \tilde{g}^{(V)})$ は以下のような幾何構造をもつ:

Theorem 3 [29]

- i) 統計多様体 $(\mathcal{L}_s, \tilde{\nabla}, \tilde{g}^{(V)})$ は ± 1 -共形平坦である.
- ii) 統計多様体 $(\mathcal{L}_s, \tilde{\nabla}, \tilde{g}^{(V)})$ は負の定曲率 $k_s = 1/(\nu_1(s)n)$ である.

情報幾何では, 双対平坦な多様体においてピタゴラスの定理が成立することはよく知られている. この拡張として, 定曲率な統計多様体での“ピタゴラスの定理”[19]と1-共形平坦な統計部分多様体への射影に関する結果[37]がある.

Theorem 3により, この二つを統合することができて \mathcal{L}_s に関して次のようなダイバージェンスの分解に関する結果が得られる:

Proposition 3 [29] $P \in PD(n, \mathbf{R})$ で $R, S \in \mathcal{L}_s$ とする. \mathcal{L}_s と \mathcal{R}_P の交点を Q とするとき, Q と R を結ぶ $\tilde{\nabla}$ -測地線 $\tilde{\gamma}$ と R と S を結ぶ $\tilde{\nabla}^*$ -測地線 $\tilde{\gamma}^*$ が R で $\tilde{g}^{(V)}$ に関して直交するならば (Figure 2 参照), 次の関係式が成立する:

$$D^{(V)}(P, S) = D^{(V)}(P, R) + \kappa D^{(V)}(R, S), \quad \kappa = \lambda \{1 - k_s D^{(V)}(Q, R)\}, \quad (13)$$

ただし $\lambda > 0$ は $Q = \lambda P$ で定義される定数である.

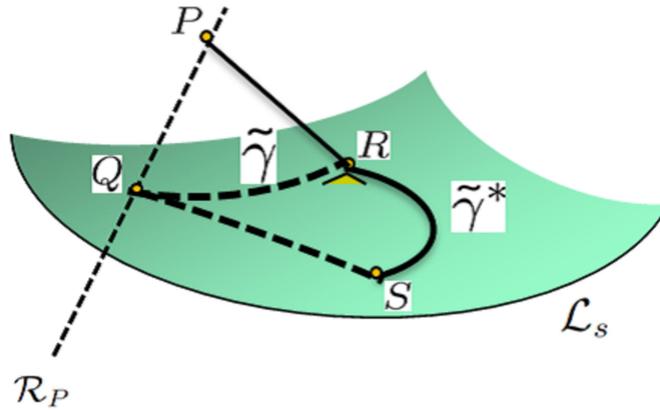


Figure 2: Proposition 3 の状況下でのダイバージェンスの分解

$P = Q$ とした時と $Q = R$ とした時が, それぞれ [19] の [37] の一般的な結果に対応する.

Remark 4 *i)* $k_s < 0$ であるので右辺の二項は常に非負である. したがって, もし \mathcal{L}_s の中に $\tilde{\nabla}^*$ -自己平行な部分多様体 \mathcal{N} を考えたとき, P から \mathcal{N} へのダイバージェンス $D^{(V)}$ を最小とする点 $\hat{P} \in \mathcal{N}$ は, P と \hat{P} を結ぶ $\tilde{\nabla}$ -測地線の \hat{P} での接ベクトルと $T_{\hat{P}}\mathcal{N}$ の直交性という局所的な最適性条件から唯一に定まる.

ii) $\mathcal{D}^{(V)}$ に関する Proposition 3 と双対な結果も成り立つ.

6 U-モデル上の双対幾何と正定値対称行列

本節では, 非指数型分布も含むあるクラスの変量統計モデルとその双対平坦構造を考え, V -ポテンシャルが誘導する正定値対称行列の双対平坦構造との対応関係を導く.

指数型分布族における統計的推論においては, KL (Kullback-Leibler) ダイバージェンスに基礎を置く最尤推定の理論が整備されている. 有効性などの最尤法の理論的な利点

を保ちながら推定結果のロバスト性を改良する方法として、一般化されたダイバージェンスの最小化によるロバスト推定法が、パターン認識、機械学習、主成分分析などの分野で提案されている [4, 6, 21, 12, 35, 23]. そのような代表的な例として、二つの確率密度関数 f, g に対し β -ダイバージェンス

$$D_\beta(f, g) = \int \frac{g(x)^{\beta+1} - f(x)^{\beta+1}}{\beta+1} - \frac{f(x)\{g(x)^\beta - f(x)^\beta\}}{\beta} dx \quad (14)$$

が知られている. β -ダイバージェンスは β が 0 に近づいたとき、KL ダイバージェンスに漸近する一方、 β が 1 に近づくと、 L^2 -距離の自乗に漸近することが分かる. したがって、 β が 0 に近づくと有効性が増す一方、 β が 1 に近づくとロバスト性が増加する [33, 5]. この意味で、有効性とロバスト性のトレードオフとして 0 と 1 の間で適当な β を調整することができる. β -ダイバージェンスは Tsallis エントロピー [36] に関係している.

このようなダイバージェンスの例をより一般化して考えてみよう

Definition 2 [4] \mathbf{R} またはそのある半無限区間を考える. そこで定義された関数 $U(s)$ を正の導関数 $u(s) = U'(s)$ をもつ凸関数、 ξ を u の逆関数とする. \mathbf{R}^n で定義された $f(x)$ と $g(x)$ に対し、次の積分

$$D_U(f, g) = \int U(\xi(g)) - U(\xi(f)) - [\xi(g) - \xi(f)] f dx$$

が存在して収束するとき、これを U -ダイバージェンスと呼ぶ.

$\xi_f = \xi(f)$, $\xi_g = \xi(g)$ とおいたとき被積分関数 $U(\xi_g) - [U(\xi_f) + u(\xi_f)(\xi_g - \xi_f)]$ が凸関数とその支持超平面の差になるので、常に $D_U(f, g) \geq 0$ であり、 $D_U(f, g) = 0$ となるのは $f = g$ の時に限ることがわかる. U -ダイバージェンスは Bregman ダイバージェンスの双対的な表現と見なせる [31] が、この表現の方が Bregman ダイバージェンスに比べて経験データに基づく統計的推論において有用な表現であることが知られている [4, 23]. また、 $U(s) = \frac{1}{\beta+1}(1 + \beta s)^{(\beta+1)/\beta}$, $s > -1/\beta$, とおけば、対応する U -ダイバージェンスは (14) で定義した β -ダイバージェンスである.

ある多様体 \mathcal{M} の元によってパラメトライズされた関数の族に対して、その関数族の上で定義された U -ダイバージェンスを考えれば、それを \mathcal{M} 上で定義されたダイバージェンスと見なせるので命題 1 を通して \mathcal{M} に双対平坦構造を誘導できる.

ここでは、 $P \in PD(n, \mathbf{R})$ で指定されるある多変量確率分布関数の族を考える. この族は U -ダイバージェンスが誘導する双対幾何構造に対して自然な統計モデルとなっている [4].

Definition 3 [4] U と u を Definition 2 で与えられた関数とする. 次のような密度関数

$$\mathcal{M}_U = \left\{ f(x, P) = u \left(-\frac{1}{2} x^T P x - c_U(\det P) \right) \mid P \in PD(n, \mathbf{R}) \right\},$$

で定義された多変量分布族を, U -ダイバージェンスに付随した (平均 0 の) U -モデルと呼ぶ. ただし, f の右辺が定義されないときは, $f(x, P) = 0$ とする. また P で定まる定数 $c_U(\det P)$ は, 次の条件

$$\int f(x, P) dx = (\det P)^{-\frac{1}{2}} \int u \left(-\frac{1}{2} y^T y - c_U(\det P) \right) dy = 1,$$

を, 積分収束のもとでみたす規格化定数である.

もし f の台が有界な場合, ρ を $u(-\rho/2 - c_U(\det P)) = 0$ を満たす定数とすると

$$\int u \left(-\frac{1}{2} y^T y - c_U(\det P) \right) dy = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \int_0^\rho u \left(-\frac{r}{2} - c_U(\det P) \right) r^{\frac{n}{2}-1} dr$$

と計算されるので, 規格化定数は

$$c_U(\det P) = \Gamma_{\frac{n}{2}, u}^{-1} \left(\frac{\Gamma(\frac{n}{2})(\det P)^{\frac{1}{2}}}{\pi^{\frac{n}{2}}} \right),$$

で与えられる. ただし, $\Gamma_{\frac{n}{2}, u}^{-1}$ は次のように定義される関数 $\Gamma_{a, u}$ の逆関数である:

$$\Gamma_{a, u}(c) = \int_0^{-2c-2\xi(0)} u \left(-\frac{r}{2} - c \right) r^{a-1} dr.$$

同様な議論は非有界な台を持つ密度関数の場合に対しても成り立つ. 後に q -ガウス分布族 (β -モデル) に対する計算例を示す.

なお, 関数 $u(s)$ がある種の自己相似性を持つ場合, 対応する U -モデルの密度関数 f は次のように通常の楕円分布族の形 [8, 22], すなわち $\det P$ に依存した定数 c_f を用いて

$$f(x, P) = c_f(\det P)^{\frac{1}{2}} u \left(-\frac{1}{2} x^T P x \right),$$

の形で書ける [28] ことを注意する. q -ガウス分布族はこのような統計モデルの例である.

命題 1 により上記の U -ダイバージェンス D_U は U -モデル \mathcal{M}_U を通してそのパラメータ空間 $PD(n, \mathbf{R})$ に双対構造を誘導する. この双対構造と 3 節と 4 節で議論した V -ポテンシャルが $PD(n, \mathbf{R})$ に直接定義する双対構造の関係を導こう. すなわち, 二つの構造の変形パラメータ (関数) である V と U の関係は以下のように与えられる:

Theorem 4 V -ポテンシャル $\varphi^{(V)}$ を定義する関数 V が, U -ダイバージェンスを定義する関数 U を用いて

$$V(s) = s^{-\frac{1}{2}} \int U \left(-\frac{1}{2} x^T x - c_U(s) \right) dx + c_U(s), \quad s > 0. \quad (15)$$

と表されたとする. この V が (5) を満たすとき, $PD(n, \mathbf{R})$ の双対構造 $(g^{(V)}, \nabla, * \nabla^{(V)})$ と U -ダイバージェンスが $PD(n, \mathbf{R})$ に定義する双対構造は一致する.

(証明は [29] を参照)

この定理から, U に対応する V が得られれば M_U の双対幾何構造を導出するために D_U に伴う積分計算を経ることなく, (15) で定まる V -ポテンシャルから 3 節, 4 節の公式を通して直接計算できる.

7 β -ダイバージェンスが q -ガウス分布族に誘導する双対構造と V -ポテンシャル

この節では, q -ガウス分布と呼ばれる多変量分布の族を考え, β -ダイバージェンスに付随する双対幾何構造を誘導するために対応する V を求める. その結果, 得られた幾何構造は Remark 3 で述べた意味での $GL(n; \mathbf{R})$ -不変性を示すことがわかる. この節の詳細は [28] を参照されたい.

β を $\beta \neq 0$ かつ $\beta \neq -1$ であるような実パラメータとし, 関数 U を次のように与える:

$$U(s) = \begin{cases} \frac{1}{\beta+1} (1+\beta s)^{(\beta+1)/\beta}, & s \in I_\beta = \{s \in \mathbf{R} | 1+\beta s > 0\}, \\ 0, & \text{otherwise.} \end{cases}$$

U の導関数 u は

$$u(s) = \begin{cases} \frac{dU(s)}{ds} = (1+\beta s)^{1/\beta}, & s \in I_\beta = \{s \in \mathbf{R} | 1+\beta s > 0\}, \\ 0, & \text{otherwise} \end{cases}$$

となり, その逆関数 ξ は I_β 上で次のようになる:

$$\xi(t) = \frac{t^\beta - 1}{\beta}, \quad t > 0.$$

$\beta > 0$ の場合 $s > -1/\beta$ で, $\beta < 0$ の場合 $s < -1/\beta$ で, U は凸かつ u は正である. β が 0 に近づくと u と ξ はそれぞれ通常の指数関数と対数関数に漸近する. これらの関数は, $q = 1 + \beta$ というパラメータ q を用いて q -指数関数と q -対数関数と呼ばれている [36, 24].

パラメータ β を固定し, q -指数関数 u を用いて $P \in PD(n, \mathbf{R})$ で指定される次のような多変量分布の密度関数 f を考える:

$$f(x, P) = u\left(-\frac{1}{2}x^T P x - c_\beta(\det P)\right).$$

ここで $c_\beta(\det P)$ は正規化定数である. この密度関数 f をもつ分布は (平均 0 の) q -ガウス分布と呼ばれる [36, 24]. ここでは, このような f の族を M_β で表し, q -ガウス分布族あるいは β -モデルと呼ぶ.

\mathcal{M}_β 上の U -ダイバージェンスは (14) の β -ダイバージェンス D_β であり, D_β が \mathcal{M}_β に誘導する双対幾何構造は, Theorem 4 の V が次のように陽に計算できるので, $PD(n, \mathbf{R})$ 上で V -ポテンシャルから直接決定できる:

Theorem 5 β -ダイバージェンスから誘導される \mathcal{M}_β の双対幾何構造は, 次のような関数 $V(s)$, $s > 0$ による V -ポテンシャルが定める $PD(n, \mathbf{R})$ 上の双対幾何構造 $(g^{(V)}, \nabla, * \nabla^{(V)})$ で特徴付けられる:

$$V(s) = \begin{cases} \frac{1}{\beta} + c^+ s^{1/(2n_\beta)}, & \beta > 0, \\ \frac{1}{\beta} + c^- s^{1/(2n_\beta)}, & -\frac{2}{n+2} < \beta < 0 \end{cases}$$

ただし, $n_\beta = n/2 + 1/\beta$ で c^\pm は β と n に依存したある定数である ([28] 参照). また, この V に対して (5) の条件は, それぞれの場合につき

$$\begin{aligned} \text{i) } \frac{c^+}{2n_\beta} > 0, \text{ ii) } \frac{1}{2n_\beta} < \frac{1}{n}, \quad \beta > 0, \\ \text{i) } \frac{c^-}{2n_\beta} < 0, \text{ ii) } \frac{1}{2n_\beta} < \frac{1}{n}, \quad -\frac{2}{n+2} < \beta < 0 \end{aligned}$$

となるが, これらは常に満たされる.

(証明は [28] を参照)

この結果は, \mathcal{M}_β の双対平坦構造が Remark 3 のべき型ポテンシャル $\varphi^{(V)}$ から定まり, そこで記した意味での $GL(n; \mathbf{R})$ -不変性を持つことも示している.

References

- [1] S. Amari, *Differential-geometrical methods in statistics*, Lecture notes in statist., 28, Springer, New York (1985).
- [2] S. Amari and H. Nagaoka, *Methods of information geometry*, AMS&OUP (2000).
- [3] A. P. David, The geometry of proper scoring rules, *Ann. Inst. Stat.*, **59**, 77–93 (2007).
- [4] S. Eguchi, Information geometry and statistical pattern recognition, *Sugaku Expositions*, Amer. Math. Soc., **19**, 197–216 (2006) (originally *Sūgaku*, **56**, 380–399 (2004) in Japanese).
- [5] S. Eguchi, Information divergence geometry and the application to statistical machine learning, F. Emmert-Streib and M. Dehmer (eds.) *Information Theory and Statistical Learning*, 309–332, Springer (2008).

- [6] S. Eguchi and J. Copas, A class of logistic-type discriminant functions, *Biometrika* **89**, no. 1, 1–22 (2002).
- [7] S. Eguchi, O. Komori and S. Kato, Projective power entropy and maximum Tsallis entropy distributions, *Entropy*, **13**, 1746–1764 (2011).
- [8] K. T. Fang, S. Kotz and K. W. Ng, *Symmetric multivariate and related distributions*, Chapman and Hall, London (1990).
- [9] J. Faraut and A. Korányi, *Analysis on symmetric cones*, Oxford Univ. Press, New York (1994).
- [10] P. D. Grunwald and A. P. David, Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory, *Ann. Stat.*, **32**, 1367–1433 (2004).
- [11] S. Helgason, *Differential geometry and symmetric spaces*, Academic Press, New York (1962).
- [12] I. Higuchi and S. Eguchi, Robust principal component analysis with adaptive selection for Tuning Parameters, *J. Machine Learning Research* **5**, 453–471 (2004).
- [13] T. Kanamori and A. Ohara, A Bregman extension of quasi-Newton updates I: An Information Geometrical framework, *Optimization Methods and Software*, **28**, No. 1, 96–123 (2013).
- [14] S. Kakihara, A. Ohara and T. Tsuchiya, Information geometry and interior-point algorithms in semidefinite programs and symmetric cone programs, *Journal of Optimization Theory and Applications*, **157**, 3, 749–780 (2013).
- [15] R. E. Kass and P. W. Vos, *Geometrical foundations of asymptotic inference*,
- [16] M. Koecher, *The Minnesota notes on Jordan algebras and their applications*, Springer, Berlin (1999).
- [17] S. Kullback, *Information theory and statistics*, John Wiley and Sons, Inc., New York, (1959).
- [18] T. Kurose, Dual connections and affine geometry, *Math. Z.* **203**, no. 1, 115–121 (1990).

- [19] T. Kurose, On the divergences of 1-conformally flat statistical manifolds, *Tohoku Math. J.* **46**, no. 3, 427–433 (1994).
- [20] S. Lauritzen, Statistical Manifolds, In: S.-I. Amari *et al.*, *Differential geometry in statistical inference*, Inst. Math. Statist., Hayward, CA, 1987.
- [21] M. Minami and S. Eguchi, Robust blind source separation by beta-divergence, *Neural Computation* **14**, 1859–1886 (2002).
- [22] R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley, New York (1982).
- [23] N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, Information geometry of U-boost and Bregman divergence, *Neural Computation* **16**, 1437–1481 (2004).
- [24] Naudts, J. *Generalized thermostatics*, Springer, **2010**.
- [25] Y. E. Nesterov and M. J. Todd, Primal-dual interior-point methods for self-scaled cones, *SIAM J. Optim.* **8**, 324–364 (1998).
- [26] A. Ohara, Geodesics for dual connections and means on symmetric cones, *Integral Equations and Operator Theory*, **50**, 537–548 (2004).
- [27] A. Ohara and S. Amari, Differential Geometric Structures of Stable State Feedback Systems with Dual Connections, *Kybernetika*, **30**, No.4, 369–386 (1994).
- [28] A. Ohara and S. Eguchi, Group Invariance of Information Geometry on q -Gaussian Distributions Induced by Beta-Divergence, *Entropy*. **15**, 4732–4747 (2013).
- [29] A. Ohara and S. Eguchi, Geometry on Positive Definite Matrices Deformed by V-Potentials and Its Submanifold Structure, *F. Nielsen Eds. Geometric Theory of Information*, Chapter 2, Springer (2014).
- [30] A. Ohara, N. Suda and S. Amari, Dualistic differential geometry of positive definite matrices and its applications to related problems, *Linear Algebra Appl.* **247**, 31–53 (1996).
- [31] A. Ohara and T. Wada, Information geometry of q -Gaussian densities and behaviors of solutions to related diffusion equations, *J. Phys. A: Math. Theor.*, **43**, 035002 (18pp.) (2010).

- [32] O. S. Rothaus, Domains of positivity, *Abh. Math. Sem. Univ. Hamburg* **24**, 189–235 (1960).
- [33] D. W. Scott, Parametric statistical modeling by minimum integrated square error, *Technometrics* **43**, 274–285 (2001).
- [34] H. Shima, The geometry of Hessian structures, World Scientific, Singapore (2007).
- [35] T. Takenouchi and S. Eguchi, Robustifying AdaBoost by adding the naive error rate, *Neural Computation*, **16**, no. 4, 767–787 (2004).
- [36] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics*, Springer, New York NY, USA (2009).
- [37] K. Uohashi, A. Ohara and T. Fujii, Foliations and divergences of flat statistical manifolds, *Hiroshima Math. J.* **30**, no. 3, 403–414 (2000).
- [38] E. B. Vinberg, The theory of convex homogeneous cones, *Trans. Moscow Math. Soc.* **12**, 340–430 (1963).
- [39] H. Wolkowicz, *et al.* eds., *Handbook of semidefinite programming*, Kluwer Acad. Publ., Boston, MA (2000).