New Geometry of Parametric Statistical Models

Jun Zhang[(✉)] and Gabriel Khan

University of Michigan, Ann Arbor, MI 48109, USA
junz@umich.edu

Abstract. We provide an alternative differential geometric framework of the manifold $\mathbb{M}$ of parametric statistical models. While adopting the Fisher-Rao metric as the Riemannian metric g on $\mathbb{M}$, we treat the original parameterization of the statistical model as affine coordinate chart on the manifold endowed with a flat connection, instead of using a pair of torsion-free affine connections with generally non-vanishing curvature. We then construct its g-conjugate connection which, while necessarily curvature-free, carries torsion in general. So instead of associating a statistical structure to $\mathbb{M}$, we construct a statistical manifold admitting torsion (SMAT). We show that $\mathbb{M}$ is dually flat if and only if torsion of the conjugate connection vanishes.

Keywords: Torsion · Weitzenböck connection · Hessian manifold

1 Introduction

Recall that in the now-classic information geometry, a parametric family of density functions, $p(\cdot|x)$, called a *parametric statistical model*, is the association $x \mapsto p(\cdot|x)$ of a point $x = [x^1, \cdots, x^n]$ in a connected open subset of $\mathbb{R}^n$ to p, such that x serves as a local coordinate chart of $p \in \mathbb{M}$ [Ama85, AN00]. The Fisher-Rao metric and the α-connections are given by

$$g_{ij}(x) = \int_\Omega d\omega \left\{ p(\omega|x) \frac{\partial \log p(\omega|x)}{\partial x^i} \frac{\partial \log p(\omega|x)}{\partial x^j} \right\};$$
$$\Gamma^{(\alpha)}_{ij,k}(x) = \int_\Omega d\omega \frac{\partial p(\omega|x)}{\partial x^k} \left(\frac{1-\alpha}{2} \frac{\partial \log p(\omega|x)}{\partial x^i} \frac{\partial \log p(\omega|x)}{\partial x^j} + \frac{\partial^2 \log p(\omega|x)}{\partial x^i \partial x^j} \right).$$

The α- and $(-\alpha)$-connection are conjugate to each other with respect to the Fisher-Rao metric g. Note that all α-connections are torsion-free; yet generally they have non-zero curvatures, with curvature of $(\pm\alpha)$-connections equal but opposite sign of each other. When the curvatures of (± 1)-connections vanish, g takes the form of a Hessian metric. It is important to keep in mind that each member of the α-connection is Codazzi-coupled to the Fisher-Rao metric g.

The project is supported by DARPA/ARO Grant W911NF-16-1-0383 ("Information Geometry: Geometrization of Science of Information", PI: Zhang).

F. Nielsen and F. Barbaresco (Eds.): GSI 2019, LNCS 11712, pp. 288–296, 2019.
https://doi.org/10.1007/978-3-030-26980-7_30

In this paper, we take a different perspective about the manifold $\mathbb{M}$ of parametric statistical models $p(\cdot|x)$. We take the parameter x to be the local coordinates of a parallelizable manifold after trivialization of its tangent bundle $T\mathbb{M}$, i.e, x is taken to be the affine coordinates of a flat connection ∇^* on $\mathbb{M}$. We continue to take Fisher-Rao metric g as the Riemannian metric on $\mathbb{M}$. Denote ∇ to be the g-conjugate of this flat connection ∇^*. Though ∇ is necessarily curvature-free, in general, ∇ will not be torsion-free. This connection is adapted to the g-conjugate frame, and we call it "pseudo-Weitzenböck connection." In the literature, a manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ for which ∇^* is flat is called a "statistical manifold admitting torsion" or SMAT [Kur07,HM11], and ∇ and g are coupled by

$$(\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y) = g(T^\nabla(Z, X), Y).$$

Below, we actually describe parametric statistical model as SMAT by constructing *the* biorthogonal frame B based on g being the Fisher-Rao metric. ∇ is torsion-free, and hence becomes "flat", if and only if g is Hessian. For more details including proofs, see [ZK19].

2 Theoretical Foundation

2.1 *g*-Conjugate Connection

We recall that given any connection ∇ and an arbitrary Riemannian metric g, the g-conjugate connection ∇^* is defined as the (unique) connection that jointly preserves g with ∇:

$$Z(g(X, Y)) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y), \tag{1}$$

where X, Y, Z are all vector fields on $\mathbb{M}$.

The curvature and torsion of conjugate connections ∇ and ∇^* are related:

(i) their curvature tensors R^∇, R^{∇^*} satisfy

$$g(R^\nabla(Z, W)X, Y) + g(R^{\nabla^*}(Z, W)Y, X) \equiv 0; \tag{2}$$

(ii) their torsion tensors T^∇, T^{∇^*} satisfy

$$g(T^{\nabla^*}(Z, X) - T^\nabla(Z, X), Y) \equiv (\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y). \tag{3}$$

A consequence of (2) is that if ∇ is curvature-free, then so is ∇^*. The consequence of (3) is that ∇ and ∇^* carry the same amount of torsion if and only if

$$(\nabla_Z g)(X, Y) = (\nabla_X g)(Z, Y),$$

which is known as the "Codazzi coupling" of (g, ∇). It is easily verified that (g, ∇) is Codazzi-coupled if and only if (g, ∇^*) is Codazzi-coupled. Both (2) and (3) are well-known facts in information geometry. A connection is called *flat* when it is both curvature-free and torsion-free. A manifold is called *dually flat* when it carries two flat connections ∇ and ∇^* that form a conjugate pair with respect to the (necessarily) Hessian metric constructed from either ∇ or ∇^*.

2.2 Connection Adapted to a Frame

Let us start by defining a local frame on a parallelizable manifold $\mathbb{M}$. A frame $\mathfrak{B} = \{\mathtt{b}_1, \cdots, \mathtt{b}_n\}$ with $n = \dim(\mathbb{M})$ is a collection of n locally linearly independent vector fields $\{\mathtt{b}_i\}_{i=1}^n$ on $\mathbb{M}$. Under local coordinate system $x = \{x^i\}_{i=1}^n$, the expression of a frame $\mathfrak{B}$ is $\mathtt{b}_i = B_i^j \partial_{x^j}$, where ∂_{x^j} is the shorthand for $\partial/\partial x^j$, and B_i^j is an $n \times n$ matrix, assumed to be of full rank and hence invertible:

$$(B^{-1})_l^i\, B_j^l = \delta_j^i = B_l^i\, (B^{-1})_j^l.$$

Here, and in the rest of the paper, B^{-1} denotes the matrix inverse of B_j^i, and Einstein summation notation is in effect.

When the B-matrix is taken to be the Jacobian matrix of coordinate transform: $x \longrightarrow y$

$$(B^{-1})_j^\alpha = \frac{\partial y^\alpha}{\partial x^j} \longleftrightarrow B_\alpha^j = \frac{\partial x^j}{\partial y^\alpha}, \tag{4}$$

then the frame $\{\mathtt{b}_i\}_{i=1}^n$ forms a coordinate frame:

$$\mathtt{b}_i = \frac{\partial x^\alpha}{\partial y^i}\frac{\partial}{\partial x^\alpha} = \frac{\partial}{\partial y^i} := \partial_{y^i}.$$

The necessary and sufficient condition for (4) is

$$\partial_{x^i}(B^{-1})_j^\alpha = \partial_{x^j}(B^{-1})_i^\alpha. \tag{5}$$

Necessity is obvious. As for sufficiency, note that when Eq. 5 is satisfied, then for each α there exists a function $y^\alpha = y^\alpha(x)$ such that

$$(B^{-1})_j^\alpha = \frac{\partial y^\alpha}{\partial x^j}.$$

Definition 1 (Adapted connection). *Given any frame $\mathfrak{B}$, the adapted connection $\nabla^{\mathfrak{B}}$ is defined by $\nabla^{\mathfrak{B}} = B\,\partial(B^{-1})$ or in component forms:*

$$\Gamma_{k\alpha}^\beta = B_j^\beta(\partial_{x^\alpha}(B^{-1})_k^j) = -(B^{-1})_k^j(\partial_{x^\alpha} B_j^\beta). \tag{6}$$

$\nabla^{\mathfrak{B}}$ as constructed is known as the "connection of parallelization" [BG80], since they always exist on a parallelizable manifold after trivialization of its tangent bundle with a global frame $\mathfrak{B}$. The following is well-known [BG80, p.223].

Proposition 1. *Given a frame $\mathfrak{B} = \{\mathtt{b}_1, \cdots, \mathtt{b}_n\}$, then*

(i) $\nabla_{\mathtt{b}_i}^{\mathfrak{B}} \mathtt{b}_j \equiv 0, \quad \forall i, j;$
(ii) $R^{\nabla^{\mathfrak{B}}} = 0$ *;*
(iii) $T^{\nabla^{\mathfrak{B}}} = 0$ *iff $\mathfrak{B}$ is a coordinate frame, i.e.,* $[\mathtt{b}_i, \mathtt{b}_j] = 0$.

Definition 2 (g-Biorthogonal frame). *Given any frame $\mathfrak{B} = \{\mathtt{b}_i\}_{i=1}^n$, the g-biorthogonal frame is defined as the (unique) frame $\mathfrak{B}^\star = \{\mathtt{b}_i^\star\}_{i=1}^n$ that is biorthogonal with respect to the given g:*

$$g(\mathtt{b}_i, \mathtt{b}_j^\star) \equiv \delta_{ij}.$$

We have the following nice property

Theorem 3 *[ZK19, Theorem 10]***.** *With respect to any Riemannian metric g, the g-conjugation of a connection induced by a frame $\mathfrak{B}$ equals the connection indued by the g-biorthogonal frame $\mathfrak{B}^\star$:*

$$\left(\nabla^{\mathfrak{B}}\right)^* = \nabla^{(\mathfrak{B}^\star)}.$$

Historically, an affine connection adapted to an orthonormal frame is called the *Weitzenböck connection*, and has been used in theoretical physics to describe an alternative theory to Einstein's general relativity. Here this construction is extended to an arbitrary frame, and hence the terminology "pseudo-Weitzenböck connections." Theorem 3 shows that the notion of biorthogonal frames is compatible with the notion of conjugate connections when the pair of connections are both adapted connections.

2.3 Dually Flat Versus Partially-Flat Manifolds

Recall that the Hessian operator (second derivative) on a function Φ on a manifold is a bilinear form sometimes denoted as $(\nabla d\Phi)(X,Y)$. Operating on the coordinate base $(X = \partial_{x^i}, Y = \partial_{x^j})$ it takes the form

$$Hess_\nabla(\Phi)(\partial_{x^i}, \partial_{x^j}) = \frac{\partial^2 \Phi}{\partial x^i \partial x^j} - \Gamma_{ij}^k \frac{\partial \Phi}{\partial x^k}.$$

Torsion-freeness of ∇ is reflected as $\Gamma_{ij}^k = \Gamma_{ji}^k$. When ∇ is further curvature-free (and hence ∇ is flat), $\Gamma_{ij}^k = 0$ using x as affine coordinates, so that

$$Hess_\nabla(\Phi)(\partial_{x^i}, \partial_{x^j}) = \frac{\partial^2 \Phi}{\partial x^i \partial x^j}.$$

It is established in Zhang and Khan [ZK19] that

Proposition 2 *[ZK19, Theorem 3]***.** *Given a torsion-free connection ∇ and a smooth function Φ on a manifold, then $(\nabla, Hess_\nabla(\Phi))$ is Codazzi coupled iff $d\Phi(R^\nabla) = 0$.*

A consequence is that any flat connection ∇ is always Codazzi coupled to $Hess_\nabla(\Phi)$, as [Shi07] observed. Denote ∇^* the conjugate connection with respect to the symmetric bilinear form $Hess_\nabla(\Phi)$ induced from a flat ∇. Then ∇^* is also flat (both curvature- and torsion-free). Assuming Φ is convex, then we have the standard Hessian manifold with

$$g = Hess_\nabla(\Phi) = Hess^*_\nabla(\Phi^*),$$

where Φ^* is the convex conjugate function of Φ.

The above analysis also tells us that given $(\mathbb{M}, g, \nabla)$ with a flat connection ∇, then whenever $g \neq Hess_\nabla(\Phi)$, then (∇, g) is in general *not* a Codazzi pair, as [Shi07] pointed out, so the g-conjugate connection ∇^* is *not* torsion-free. This is the situation of the so-called "partially-flat" manifold [Hen17]. Next, we apply this concept to the manifold of parametric statistical models.

3 Parametric Statistical Models as Partially-Flat Geometry

3.1 Riemannian Manifold of Parametric Statistical Models

Take $x = [x_1, \cdots, x_n]$, the parameter of a parametric statistical model $p(\cdot|x)$, to be the affine coordinates on a parallelizable manifold with flat connection ∇^*, i.e., the Christoffel symbol Γ^{*i}_{jk} vanishes. Writing out the equation of conjugate connections ∇, ∇^* under this coordinate chart

$$\frac{\partial g_{ij}}{\partial x^k} = g_{lj}\Gamma^l_{ki} + g_{il}\Gamma^{*l}_{kj} = g_{lj}\Gamma^l_{ki} \,.$$

Therefore, the pseudo-Weitzenböck connection ∇ of the parametric statistical model is (written as its Christoffel symbol Γ^i_{jk})

$$\Gamma^j_{ki} = g^{jl}\frac{\partial g_{il}}{\partial x^k} \,, \tag{7}$$

with g^{ij} denoting the elements of the matrix inverse of g, the Fisher-Rao metric. It is well-known [BG80] that such a connection is always curvature-free., but carries torsion

$$T^j_{ik} = g^{jl}\left(\frac{\partial g_{kl}}{\partial x^i} - \frac{\partial g_{il}}{\partial x^k}\right).$$

In general, $T \neq 0$, unless $\frac{\partial g_{il}}{\partial x^k}$ is totally symmetric, i.e., g is Hessian. Otherwise, from any connection ∇ with torsion T^∇, we can construct a torsion-free connection $\nabla - \frac{1}{2}T^\nabla$; in the present case,

$$\Gamma^j_{ki} - \frac{1}{2}T^j_{ki} = \frac{g^{jl}}{2}\left(\frac{\partial g_{kl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^k}\right)$$

is always torson-free, and differs from the Levi-Civita connection of g by $\frac{1}{2}g^{jl}\partial_{x^l} g_{ik}$.

Even though Γ^j_{ki} given by (7) may carry torsion, its geodesic equation

$$\frac{d^2 x^j}{ds^2} + g^{jl}\frac{\partial g_{il}}{\partial x^k}\frac{dx^i}{ds}\frac{dx^k}{ds} = 0$$

or equivalently

$$\frac{d}{ds}\left(g_{ij}\frac{dx^j}{ds}\right) = 0$$

still yields the same solution as given by

$$g_{ij}\frac{dx^j}{ds} = const, \qquad i = 1, 2, \cdots, n.$$

Torsion of Γ^j_{ki} is *not* captured in the geodesic curves themselves; it describes the "screw" component of the motion with axis of rotation precisely the tangent direction of the curve. When two connections differ only by torsion, then their associated geodesic equations are the same, since the anti-symmetric part of Γ is canceled after summation with $\frac{dx^i}{ds}\frac{dx^k}{ds}$.

The associated frame, which we call "canonical frame" of $\mathbb{M}$ and denote by upper script $\{\mathbf{b}^i\}_{i=1}^n$, is

$$\mathfrak{B} = \{\mathbf{b}^i\}_{i=1}^n = \{g^{ij}\partial_{x^j}\,,\ i = 1, 2, \cdots, n\}\,.$$

This frame is nothing but the "natural gradient" vector popularly known to the machine learning community after Amari [Ama98].

3.2 Pre-contrast Function and α-Connections

Just as a statistical structure may be induced by a contrast function, a SMAT may be induced by a pre-contrast function ρ [HM11] which, in the partially-flat case, has a canonical expression [Hen17]. We show that

Proposition 3. *The canonical pre-contrast function* $\mathbb{M} \times T\mathbb{M} \to \mathbb{R}$ *is*

$$\rho(\partial_{x^i}, x, x') = -g(\partial_{x^i}, (x'^j - x^j)\partial_{x^j}) = (x^j - x'^j)\, g_{ij}(x).$$

This can be seen from

$$\begin{aligned} -\left.\frac{\partial \rho}{\partial x'^j}\right|_{x'=x} &= g_{ij}(x), \\ -\left.\frac{\partial^2 \rho}{\partial x'^k \partial x'^j}\right|_{x'=x} &= 0, \\ -\left.\frac{\partial \rho}{\partial x^k \partial x'^j}\right|_{x'=x} &= \frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ki,j}. \end{aligned}$$

where the canonical connection Γ carries torsion, $\Gamma_{ki,j} \neq \Gamma_{ik,j}$, in general.

The family of α-connections, $\widetilde{\nabla}^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^* = \frac{1+\alpha}{2}\nabla$ all carry torsion (except $\alpha = -1$)

$$\begin{aligned} \widetilde{\Gamma}^{(\alpha)}_{ki,j}(x) = \frac{1+\alpha}{2}\Bigg(&\int_\Omega d\omega\,\left\{\frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^i}\frac{\partial p(\omega|x)}{\partial x^j} + \frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^j}\frac{\partial p(\omega|x)}{\partial x^i}\right\} \\ &+ \int_\Omega d\omega\, p(\omega|x)\,\frac{\partial \log p(\omega|x)}{\partial x^i}\frac{\partial \log p(\omega|x)}{\partial x^j}\frac{\partial \log p(\omega|x)}{\partial x^k}\Bigg), \end{aligned}$$

with torsion given by

$$\widetilde{T}^{(\alpha)j}_{ik} = \frac{1+\alpha}{2}\, g^{jl}\int_\Omega d\omega\,\left\{\frac{\partial^2 \log p(\omega|x)}{\partial x^i \partial x^l}\frac{\partial p(\omega|x)}{\partial x^k} - \frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^l}\frac{\partial p(\omega|x)}{\partial x^i}\right\}.$$

3.3 Univariate Normal Distribution: An Example

We consider the univariate normal family on the real line ($-\infty < \omega < \infty$)

$$\mathcal{N}(\omega|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\omega-\mu)^2}{2\sigma^2}\right)$$

with parameters $m = \sqrt{2}\,\mu$ and σ (the factor of $\sqrt{2}$ is for later convenience). Reparametrizing, it is possible to consider $\mathcal{N}(\omega|\mu,\sigma)$ as an exponential family with natural coordinates $x = (x^1, x^2)$ and expectation coordinates $u = (u_1, u_2)$:

$$\begin{aligned} x^1 &= \frac{\mu}{\sigma^2}, \qquad & x^2 &= -\frac{1}{2\sigma^2}; \\ u_1 &= \mu, \qquad & u_2 &= \mu^2 + \sigma^2. \end{aligned}$$

When treating x (or u) as affine coordinates for the dually flat connections, the Fisher-Rao metric g becomes the Hessian metric with potential Φ

$$\Phi(x) = -\frac{x^1 \cdot x^1}{4x^2} + \frac{1}{2}\log\left(-\frac{\pi}{x^2}\right).$$

As the mean μ and variance σ parameters of the univariate normal model are intrinsically meaningful in statistics, it is desirable to treat (μ, σ) as affine coordinates for some flat connection. As such, if we consider the coordinate frame

$$\left\{\frac{\partial}{\partial m}, \frac{\partial}{\partial \sigma}\right\},$$

its biorthogonal frame with respect to the Fisher-Rao metric

$$g = \frac{2}{\sigma^2}(dm^2 + d\sigma^2)$$

is

$$\left\{\frac{\sigma^2}{2}\frac{\partial}{\partial m}, \frac{\sigma^2}{2}\frac{\partial}{\partial \sigma}\right\}.$$

By computing the Lie bracket of the biorthogonal frame, we find that

$$\left[\frac{\sigma^2}{2}\frac{\partial}{\partial m}, \frac{\sigma^2}{2}\frac{\partial}{\partial \sigma}\right] = -\frac{\sigma^3}{2}\frac{\partial}{\partial m}.$$

This is the torsion of the pseudo-Weitzenböck connection adapted to the g-biorthogonal frame. It is not a coordinate frame and the torsion of the g-conjugate connection is non-zero.

Note that the Fisher-Rao metric, when expressed in the (m, σ)-coordinates, is *not* Hessian. The pseudo-Weitzenböck connection derived above has geodesics which are reparametrizations of straight lines in the upper half-plane. This fact does not hold in general, but turns out in the present case because the Fisher-Rao metric, though not Hessian, is so simple for our choice of parametrization.

To summarize, we have constructed a presentation of the univariate normal family (as a manifold of upper half-plane), not as a manifold of dual flatness (Hessian manifold) in the conventionally-adopted natural and expectation coordinates, but as a partially-flat statistical manifold admitting torsion (SMAT) in the original (m, σ)-coordinates.

4 Discussions

Classical information geometry involves statistical manifolds, with two equivalent definitions as follows:

(i) Lauritzen's [Lau87] viewpoint: $(\mathbb{M}, g, \nabla, \nabla^*)$ where the pair of g-conjugated connections ∇ and ∇^* are both torsion-free;
(ii) Kurose's [Kur90] viewpoint: $(\mathbb{M}, g, \nabla)$ where ∇ is torsion-free and Codazzi-coupled to g.

With application to parametric statistical models, the Riemannian metric is the Fisher-Rao metric and the pair of conjugate connections are the (± 1)-connections, generated by divergence (contrast) functions. These are "canonical" objects once the parametric statistical model $p(\cdot|x)$ is specified, canonical because they are unique second- and third-order invariants for parametric statistical models (see [Dow18] and [AJVLS15]). Here we provide another "canonical" construction of a parametric statistical model as a parallelizable manifold with a "partially-flat" geometry [Hen17] under which both conjugate connections are curvature-free. A partially-flat structure (of a parallelizable manifold) is a slightest relaxation to the dually flat Hessian structure, by allowing one of the connections (say, ∇^*) to be torsion-free. The metric g need not be Hessian, nor is the flat connection required to be Codazzi coupled to g. In other words, our construction of this manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ is such that ∇^* is flat and ∇ is curvature-free but usually carries torsion, while g is still the Fisher-Rao metric. This is a special case of a statistical manifold admitting torsion (SMAT, [Kur07]) that can be generated by "pre-contrast functions" [HM11]. Compared to statistical manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ à la Lauritzen, our alternative approach selects a pair of connections both of which are, instead of torsion-free, curvature-free. Compared to statistical manifold $(\mathbb{M}, g, \nabla)$ à la Kurose, our alternative approach selects a connection that is, instead of Codazzi-coupled, SMAT-coupled to g. The switch of emphasis from curvature to torsion may lead to interesting reformulation of information geometry.

References

[AJVLS15] Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields **162**(1–2), 327–364 (2015)

[Ama85] Amari, S.: Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics. Springer, New York (1985). https://doi.org/10.1007/978-1-4612-5056-2

[Ama98] Amari, S.-I.: Natural gradient works efficiently in learning. Neural Comput. **10**(2), 251–276 (1998)

[AN00] Amari, S., Nagaoka, H.: Methods of Information Geometry, vol. 191. American Mathematical Soc. (2000)

[BG80] Bishop, R.L., Samuel, I.: Tensor Analysis on Manifolds. Courier Corporation, Goldberg (1980)

[Dow18] Dowty, J.G.: Chentsov's theorem for exponential families. Inf. Geom. **1**(1), 117–135 (2018)

[Hen17] Henmi, M.: Statistical manifolds admitting torsion, pre-contrast functions and estimating functions. In: Nielsen, F., Barbaresco, F. (eds.) GSI 2017. LNCS, vol. 10589, pp. 153–161. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68445-1_18

[HM11] Henmi, M., Matsuzoe, H.: Geometry of pre-contrast functions and non-conservative estimating functions. In: AIP Conference Proceedings, vol. 1340, pp. 32–41. AIP (2011)

[Kur90] Kurose, T.: Dual connections and affine geometry. Math. Z. **203**(1), 115–121 (1990)

[Kur07] Kurose, T.: Statistical manifolds admitting torsion. Geometry and Something (2007)

[Lau87] Lauritzen, S.L.: Statistical manifolds. Differ. Geom. Stat. Infer. **10**, 163–216 (1987)

[Shi07] Shima, H.: The Geometry of Hessian Structures. World Scientific, Singapore (2007)

[ZK19] Zhang, J., Khan, G.: From Hessian to Weitzenböck: manifolds with torsion-carrying connections. Inf. Geom. (2019)