

# Statistical Manifolds Admitting Torsion, Pre-contrast Functions and Estimating Functions

Masayuki Henmi<sup>(✉)</sup>

The Institute of Statistical Mathematics,  
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan  
henmi@ism.ac.jp

**Abstract.** It is well-known that a contrast function defined on a product manifold  $M \times M$  induces a Riemannian metric and a pair of dual torsion-free affine connections on the manifold  $M$ . This geometrical structure is called a statistical manifold and plays a central role in information geometry. Recently, the notion of pre-contrast function has been introduced and shown to induce a similar differential geometrical structure on  $M$ , but one of the two dual affine connections is not necessarily torsion-free. This structure is called a statistical manifold admitting torsion. This paper summarizes such previous results including the fact that an estimating function on a parametric statistical model naturally defines a pre-contrast function to induce a statistical manifold admitting torsion and provides some new insights on this geometrical structure. That is, we show that the canonical pre-contrast function can be defined on a partially flat space, which is a flat manifold with respect to only one of the dual connections, and discuss a generalized projection theorem in terms of the canonical pre-contrast function.

**Keywords:** Statistical manifold · Torsion · Contrast function · Estimating function · Standardization · Godambe information matrix · Information geometry

## 1 Introduction

In information geometry, a central role is played by a statistical manifold, which is a Riemannian manifold with a pair of two dual torsion-free affine connections. This geometrical structure is induced from an asymmetric (squared) distance-like smooth function called a contrast function by taking its second and third derivatives [1, 2]. The Kullback-Leibler divergence on a regular parametric statistical model is a typical example of contrast functions and its induced geometrical objects are the Fisher metric, the exponential and mixture connections. The structure determined by these objects play an important role in the geometry of statistical inference, as is widely known [3, 4].

---

M. Henmi—This work was supported by JSPS KAKENHI Grant Number 15K00064.

A statistical manifold admitting torsion (SMAT) is a Riemannian manifold with a pair of two dual affine connections, where only one of them must be torsion-free but the other is necessarily *not*. This geometrical structure naturally appears in a quantum statistical model (i.e. a set of density matrices representing quantum states) [3] and the notion of SMAT was originally introduced to study such a geometrical structure from a mathematical point of view [5]. A pre-contrast function was subsequently introduced as a generalization for the first derivative of a contrast function and it was shown that an pre-contrast function induces a SMAT by taking its first and second derivatives [6].

Henmi and Matsuzoe [7] showed that a SMAT also appears in “classical” statistics through an estimating function. More precisely, an estimating function naturally defines a pre-contrast function on a parametric statistical model and a SMAT is induced from it.

This paper summarizes such previous results and provides some new insights for this geometrical structure. That is, we show that the canonical pre-contrast function can be defined on a partially flat space, which is a SMAT where only one of its dual connections is flat, and discuss a generalized projection theorem in a partially flat space. This theorem relates orthogonal projection of the geodesic with respect to the flat connection to the canonical pre-contrast function.

## 2 Statistical Manifolds and Contrast Functions

In this paper, we assume that all geometrical objects on differentiable manifolds are smooth and restrict our attention to Riemannian manifolds, although the most of the concepts can be defined for semi-Riemannian manifolds.

Let  $(M, g)$  be a Riemannian manifold and  $\nabla$  be an affine connection on  $M$ . The *dual connection*  $\nabla^*$  of  $\nabla$  with respect to  $g$  is defined by

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M))$$

where  $\mathcal{X}(M)$  is the set of all vector fields on  $M$ .

For a affine connection  $\nabla$  on  $M$ , its curvature tensor field  $R$  and torsion tensor field  $T$  are defined by the following equations as usual:

$$\begin{aligned} R(X, Y)Z &:= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)), \\ T(X, Y) &:= \nabla_X Y - \nabla_Y X - [X, Y] \quad (\forall X, \forall Y \in \mathcal{X}(M)). \end{aligned}$$

It is said that an affine connection  $\nabla$  is *torsion-free* if  $T = 0$ . Note that for a torsion-free affine connection  $\nabla$ ,  $\nabla^* = \nabla$  implies that  $\nabla$  is the Levi-Civita connection with respect to  $g$ . Let  $R^*$  and  $T^*$  be the curvature and torsion tensor fields of  $\nabla^*$ , respectively. It is easy to see that  $R = 0$  always implies  $R^* = 0$ , but  $T = 0$  does not necessarily implies  $T^* = 0$ .

Let  $\nabla$  be a torsion-free affine connection on a Riemannian manifold  $(M, g)$ . Following [8], we say that  $(M, g, \nabla)$  is a *statistical manifold* if and only if  $\nabla g$  is a symmetric  $(0, 3)$ -tensor field, that is

$$(\nabla_X g)(Y, Z) = (\nabla_Y g)(X, Z) \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)). \quad (1)$$

This condition is equivalent to  $T^* = 0$  under the condition that  $\nabla$  is a torsion-free. If  $(M, g, \nabla)$  is a statistical manifold, so is  $(M, g, \nabla^*)$  and it is called the *dual statistical manifold* of  $(M, g, \nabla)$ . Since  $\nabla$  and  $\nabla^*$  are both torsion-free for a statistical manifold  $(M, g, \nabla)$ ,  $R = 0$  implies that  $\nabla$  and  $\nabla^*$  are both flat. In this case,  $(M, g, \nabla, \nabla^*)$  is called a *dually flat space*.

Let  $\phi$  be a real-valued function on the direct product  $M \times M$  of a manifold  $M$  and  $X_1, \dots, X_i, Y_1, \dots, Y_j$  be vector fields on  $M$ . The functions  $\phi[X_1, \dots, X_i|Y_1, \dots, Y_j]$ ,  $\phi[X_1, \dots, X_i| ]$  and  $\phi[ |Y_1, \dots, Y_j]$  on  $M$  are defined by the equations

$$\phi[X_1, \dots, X_i|Y_1, \dots, Y_j](r) := (X_1)_p \cdots (X_i)_p (Y_1)_q \cdots (Y_j)_q \phi(p, q)|_{p=r, q=r}, \tag{2}$$

$$\phi[X_1, \dots, X_i| ](r) := (X_1)_p \cdots (X_i)_p \phi(p, r)|_{p=r}, \tag{3}$$

$$\phi[ |Y_1, \dots, Y_j](r) := (Y_1)_q \cdots (Y_j)_q \phi(r, q)|_{q=r} \tag{4}$$

for any  $r \in M$ , respectively [1]. Using these notations, a *contrast function*  $\phi$  is defined to be a real-valued function which satisfies the following conditions on  $M$  [1,2]:

- (a)  $\phi(p, p) = 0 \quad (\forall p \in M)$
- (b)  $\phi[X| ] = \phi[ |X] = 0 \quad (\forall X \in \mathcal{X}(M))$
- (c)  $g(X, Y) := -\phi[X|Y] \quad (\forall X, \forall Y \in \mathcal{X}(M))$  is a Riemannian metric on  $M$ .

Note that these conditions imply that in some neighborhood of the diagonal set  $\{(r, r)|r \in M\}$  in  $M \times M$ ,

$$\phi(p, q) \geq 0, \quad \phi(p, q) = 0 \iff p = q.$$

Although a contrast function is not necessarily symmetric, this inequality means that a contrast function measures some discrepancy between two points on  $M$  (at least locally). For a given contrast function  $\phi$ , the two affine connections  $\nabla$  and  $\nabla^*$  are defined by

$$g(\nabla_X Y, Z) = -\phi[XY|Z], \quad g(Y, \nabla_X^* Z) = -\phi[Y|XZ] \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)).$$

In this case,  $\nabla$  and  $\nabla^*$  are both torsion-free and dual to each other with respect to  $g$ , which means that both of  $(M, g, \nabla)$  and  $(M, g, \nabla^*)$  are statistical manifolds. In particular,  $(M, g, \nabla)$  is called the statistical manifold induced from the contrast function  $\phi$ .

Now we briefly mention a typical example of contrast functions. Let  $S = \{p(\mathbf{x}; \boldsymbol{\theta}) \mid \boldsymbol{\theta} = (\theta^1, \dots, \theta^d) \in \Theta \subset \mathbf{R}^d\}$  be a regular parametric statistical model, which is a set of probability density functions with respect to a dominating measure  $\nu$  on a sample space  $\mathcal{X}$ . Each element is indexed by a parameter (vector)  $\boldsymbol{\theta}$  in an open subset  $\Theta$  of  $\mathbf{R}^d$  and the set  $S$  satisfies some regularity conditions, under which  $S$  can be seen as a differentiable manifold. The Kullback-Leibler divergence of the two density functions  $p_1(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_1)$  and  $p_2(\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\theta}_2)$  in  $S$  is defined to be

$$\phi_{KL}(p_1, p_2) := \int_{\mathcal{X}} p_2(\mathbf{x}) \log \frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} \nu(d\mathbf{x}).$$

It is easy to see that the Kullback-Leibler divergence satisfies the conditions (a), (b) and (c), and so it is a contrast function on  $S$ . Its induced Riemannian metric and dual connections are Fisher metric, the exponential and mixture connections, respectively, and given as follows:

$$\begin{aligned}
 g_{jk}(\boldsymbol{\theta}) &:= g(\partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{s^j(\mathbf{x}, \boldsymbol{\theta})s^k(\mathbf{x}, \boldsymbol{\theta})\}, \\
 \begin{cases} \Gamma_{ij,k}(\boldsymbol{\theta}) &:= g(\nabla_{\partial_i}\partial_j, \partial_k) = E_{\boldsymbol{\theta}}[\{\partial_i s^j(\mathbf{x}, \boldsymbol{\theta})\}s^k(\mathbf{x}, \boldsymbol{\theta})] \\ \Gamma_{ik,j}^*(\boldsymbol{\theta}) &:= g(\partial_j, \nabla_{\partial_i}^*\partial_k) = \int_{\mathcal{X}} s^j(\mathbf{x}, \boldsymbol{\theta})\partial_i\partial_k p(\mathbf{x}; \boldsymbol{\theta})\nu(d\mathbf{x}) \end{cases} ,
 \end{aligned}$$

where  $E_{\boldsymbol{\theta}}$  indicates that the expectation is taken with respect to  $p(\mathbf{x}; \boldsymbol{\theta})$ ,  $\partial_i = \frac{\partial}{\partial\theta^i}$  and  $s^i(\mathbf{x}; \boldsymbol{\theta}) = \partial_i \log p(\mathbf{x}; \boldsymbol{\theta})$  ( $i = 1, \dots, d$ ). As is widely known, this geometrical structure plays the most fundamental and important role in the differential geometry of statistical inference [3, 4].

### 3 Statistical Manifolds Admitting Torsion and Pre-contrast Functions

A statistical manifold admitting torsion is an abstract notion for the geometrical structure where only one of the dual connections is allowed to have torsion, which naturally appears in a quantum statistical model [3]. The definition is obtained by generalizing (1) in the definition of statistical manifold as follows [5].

Let  $(M, g)$  be a Riemannian manifold and  $\nabla$  be an affine connection on  $M$ . We say that  $(M, g, \nabla)$  is a *statistical manifold admitting torsion* (SMAT for short) if and only if

$$(\nabla_X g)(Y, Z) - (\nabla_Y g)(X, Z) = -g(T(X, Y), Z) \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)). \quad (5)$$

This condition is equivalent to  $T^* = 0$  in the case where  $\nabla$  possibly has torsion. Note that the condition (5) reduces to (1) if  $\nabla$  is torsion-free and that  $(M, g, \nabla^*)$  is not necessarily a statistical manifold although  $\nabla^*$  is torsion-free. It should be also noted that  $(M, g, \nabla^*)$  is a SMAT whenever a torsion-free affine connection  $\nabla$  is given on a Riemannian manifold  $(M, g)$ .

For a SMAT  $(M, g, \nabla)$ ,  $R = 0$  does not necessarily imply that  $\nabla$  is flat, but it implies that  $\nabla^*$  is flat since  $R^* = 0$  and  $T^* = 0$ . In this case, we call  $(M, g, \nabla, \nabla^*)$  a *partially flat space*.

Let  $\rho$  be a real-valued function on the direct product  $TM \times M$  of a manifold  $M$  and its tangent bundle  $TM$ , and  $X_1, \dots, X_i, Y_1, \dots, Y_j, Z$  be vector fields on  $M$ . The function  $\rho[X_1, \dots, X_i Z | Y_1, \dots, Y_j]$  is defined by

$$\rho[X_1, \dots, X_i Z | Y_1, \dots, Y_j](r) := (X_1)_p \cdots (X_i)_p (Y_1)_q \cdots (Y_j)_q \rho(Z_p, q)|_{p=r, q=r}$$

for any  $r \in M$ . Note that the role of  $Z$  is different from vector fields in the notation of (2). The functions  $\rho[X_1, \dots, X_i Z]$  and  $\rho[|Y_1, \dots, Y_j]$  are also defined in the similar way to (3) and (4).

We say that  $\phi$  is a *pre-contrast function* on  $M$  if and only if the following conditions are satisfied [6, 7]:

- (a)  $\rho(f_1 X_1 + f_2 X_2, q) = 0$  ( $\forall f_i \in C^\infty(M), \forall X_i \in \mathcal{X}(M)$  ( $i = 1, 2$ ),  $\forall q \in M$ )
- (b)  $\rho[X | ] = 0$  ( $\forall X \in \mathcal{X}(M)$ ) i.e.  $\rho(X_p, p) = 0$  ( $\forall p \in M$ )
- (c)  $g(X, Y) := -\rho[X|Y]$  ( $\forall X, \forall Y \in \mathcal{X}(M)$ ) is a Riemannian metric on  $M$ .

Note that for any contrast function  $\phi$ , the function  $\rho_\phi$  which is defined by  $\rho_\phi(X_p, q) := X_p \phi(p, q)$  ( $\forall p, \forall q \in M, \forall X_p \in T_p(M)$ ) is a pre-contrast function on  $M$ . The notion of pre-contrast function is obtained by taking the fundamental properties of the first derivative of a contrast function as axioms. For a given pre-contrast function, two affine connections  $\nabla$  and  $\nabla^*$  are defined by the following equations in the same way as a contrast function:

$$g(\nabla_X Y, Z) = -\rho[XY|Z], \quad g(Y, \nabla_X^* Z) = -\rho[Y|XZ] \quad (\forall X, \forall Y, \forall Z \in \mathcal{X}(M)).$$

In this case,  $\nabla$  and  $\nabla^*$  are dual to each other with respect to  $g$  and  $\nabla^*$  is torsion-free. However, the affine connection  $\nabla$  possibly has torsion. This means that  $(M, g, \nabla)$  is a SMAT and it is called the SMAT induced from the pre-contrast function  $\rho$ .

## 4 Generalized Projection Theorem in Partially Flat Spaces

In a dually flat space  $(M, g, \nabla, \nabla^*)$ , it is well-known that the canonical contrast functions (called  $\nabla$ - and  $\nabla^*$ -divergences) are naturally defined, and the Pythagorean theorem and the projection theorem are stated in terms of the  $\nabla$  and  $\nabla^*$  geodesics and the canonical contrast functions [3, 4]. In a partially flat space  $(M, g, \nabla, \nabla^*)$ , where  $R = R^* = 0$  and  $T^* = 0$ , a pre-contrast function which seems to be canonical can be defined and a projection theorem holds on the “canonical” pre-contrast function and the  $\nabla^*$ -geodesic.

**Proposition 1 (Canonical Pre-contrast Functions).** *Let  $(M, g, \nabla, \nabla^*)$  be a partially flat space (i.e.  $(M, g, \nabla)$  is a SMAT with  $R = R^* = 0$  and  $T^* = 0$ ) and  $(U, \eta_i)$  be an affine coordinate neighborhood with respect to  $\nabla^*$  in  $M$ . The function  $\rho$  on  $TU \times U$  defined by the following equation is a pre-contrast function on  $U$  which induces the SMAT  $(U, g, \nabla)$ :*

$$\rho(Z_p, q) := -g_p(Z_p, \dot{\gamma}^*(0)) \quad (\forall p, \forall q \in U, \forall Z_p \in T_p(U)), \quad (6)$$

where  $\gamma^* : [0, 1] \rightarrow U$  is the  $\nabla^*$ -geodesic such that  $\gamma^*(0) = p, \gamma^*(1) = q$  and  $\dot{\gamma}^*(0)$  is the tangent vector of  $\gamma^*$  on  $p$ .

*Proof.* For the function  $\rho$  defined as (6), the condition (a) in the definition of pre-contrast functions follows from the bilinearity of the inner product  $g_p$ . The condition (b) immediately follows from  $\dot{\gamma}^*(0) = 0$  when  $p = q$ . By calculating the derivatives of  $\rho$  with the affine coordinate system  $(\eta_i)$ , it can be shown that the condition (c) holds and that the induced Riemannian metric and dual affine connections coincide with the original  $g, \nabla$  and  $\nabla^*$ .  $\square$

In particular, if  $(M, g, \nabla, \nabla^*)$  is a dually flat space, the pre-contrast function  $\rho$  defined in (6) coincides with the directional derivative of  $\nabla^*$ -divergence  $\phi^*(\cdot, q)$  with respect to  $Z_p$  (cf. [9, 10]). Hence, the definition of (6) seems to be natural one and we call the function  $\rho$  in (6) the *canonical pre-contrast function* in a partially flat space  $(U, g, \nabla, \nabla^*)$ .

From the definition of the canonical pre-contrast function, we can immediately obtain the following theorem.

**Corollary 1 (Generalized Projection Theorem).** *Let  $U$  be an affine coordinate neighborhood and  $\rho$  be the canonical pre-contrast function defined in Proposition 1. For any submanifold  $N$  in  $U$ , the following conditions are equivalent:*

- (i) *The  $\nabla^*$ -geodesic starting at  $q \in U$  is perpendicular to  $N$  at  $p \in N$*
- (ii)  *$\rho(Z_p, q) = 0$  for any  $Z_p$  in  $T_p(N)$ .*

In the case where  $(U, g, \nabla, \nabla^*)$  is a dually flat space, the projection theorem states that the minimum of the  $\nabla^*$ -divergence  $\phi^*(\cdot, q) : N \rightarrow \mathbf{R}$  should attain at the point  $p \in N$  where the  $\nabla^*$ -geodesic starting at  $q$  is perpendicular to  $N$ . It immediately follows from the generalized projection theorem, since the directional derivative of  $\phi^*(\cdot, q)$  is the canonical pre-contrast function.

## 5 Statistical Manifolds Admitting Torsion Induced from Estimating Functions

As we mentioned in Introduction, a SMAT naturally appears through estimating functions in a “classical” statistical model as well as in a quantum statistical model. In this section, we briefly explain how a SMAT is induced on  $S$  from an estimating function. See [7] for more details including a concrete example.

Let  $S = \{p(\mathbf{x}; \boldsymbol{\theta}) \mid \boldsymbol{\theta} = (\theta^1, \dots, \theta^d) \in \Theta \subset \mathbf{R}^d\}$  be a regular parametric statistical model. An estimating function on  $S$ , which we consider here, is a  $\mathbf{R}^d$ -valued function  $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$  satisfying the following conditions:

$$E_{\boldsymbol{\theta}}\{\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})\} = \mathbf{0}, \quad E_{\boldsymbol{\theta}}\{\|\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})\|^2\} < \infty, \quad \det \left[ E_{\boldsymbol{\theta}} \left\{ \frac{\partial \mathbf{u}}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \right\} \right] \neq 0 \quad (\forall \boldsymbol{\theta} \in \Theta).$$

The first condition is called the unbiasedness of estimating functions, which is important to ensure the consistency of the estimator obtained from an estimating function. Let  $X_1, \dots, X_n$  be a random sample from an unknown probability distribution  $p(\mathbf{x}; \boldsymbol{\theta}_0)$  in  $S$ . The estimator  $\hat{\boldsymbol{\theta}}$  for  $\boldsymbol{\theta}_0$ , which is obtained as a solution to the estimating equation  $\sum_{i=1}^n \mathbf{u}(\mathbf{X}_i, \boldsymbol{\theta}) = \mathbf{0}$ , is called an M-estimator. The M-estimator  $\hat{\boldsymbol{\theta}}$  has the consistency  $\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0$  (in probability as  $n \rightarrow \infty$ ) and the asymptotic normality  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \rightarrow N(\mathbf{0}, \text{Avar}(\hat{\boldsymbol{\theta}}))$  (in distribution as  $n \rightarrow \infty$ ) under some additional regularity conditions [11], where  $\text{Avar}(\hat{\boldsymbol{\theta}})$  is an asymptotic variance-covariance matrix of  $\hat{\boldsymbol{\theta}}$  and is given by

$\text{Avar}(\hat{\theta}) = \{A(\theta_0)\}^{-1}B(\theta_0)\{A(\theta_0)\}^{-T}$  with  $A(\theta) := E_{\theta} \{(\partial \mathbf{u}/\partial \theta)(\mathbf{x}, \theta)\}$  and  $B(\theta) := E_{\theta} \{\mathbf{u}(\mathbf{x}, \theta)\mathbf{u}(\mathbf{x}, \theta)^T\}$ .

In order to induce the structure of SMAT on  $S$  from an estimating function, we consider the notion of *standardization* of estimating functions. For an estimating function  $\mathbf{u}(\mathbf{x}, \theta)$ , its standardization (or *standardized estimating function*) is defined by

$$\mathbf{u}_*(\mathbf{x}, \theta) := E_{\theta} \{s(\mathbf{x}, \theta)\mathbf{u}(\mathbf{x}, \theta)^T\} [E_{\theta} \{\mathbf{u}(\mathbf{x}, \theta)\mathbf{u}(\mathbf{x}, \theta)^T\}]^{-1} \mathbf{u}(\mathbf{x}, \theta),$$

where  $s(\mathbf{x}, \theta) = (\partial/\partial \theta) \log p(\mathbf{x}; \theta)$  is the score function [12]. Geometrically, the  $i$ th component of the standardized estimating function  $\mathbf{u}_*(\mathbf{x}, \theta)$  is the orthogonal projection of the  $i$ th component of the score function  $s(\mathbf{x}, \theta)$  onto the linear space spanned by all components of the estimating function  $\mathbf{u}(\mathbf{x}, \theta)$  in the Hilbert space

$$\mathcal{H}_{\theta} := \{a(\mathbf{x}) \mid E_{\theta}\{a(\mathbf{x})\} = 0, E_{\theta}\{a(\mathbf{x})^2\} < \infty\}$$

with the inner product  $\langle a(\mathbf{x}), b(\mathbf{x}) \rangle_{\theta} := E_{\theta}\{a(\mathbf{x})b(\mathbf{x})\}$  ( $\forall a(\mathbf{x}), \forall b(\mathbf{x}) \in \mathcal{H}_{\theta}$ ). In terms of the standardization, the asymptotic variance-covariance matrix can be rewritten as  $\text{Avar}(\hat{\theta}) = \{G(\theta_0)\}^{-1}$ , where  $G(\theta) := E_{\theta} \{\mathbf{u}_*(\mathbf{x}, \theta)\mathbf{u}_*(\mathbf{x}, \theta)^T\}$ . The matrix  $G(\theta)$  is called a Godambe information matrix [13], which is a generalization of the Fisher information matrix.

As we have seen in Sect. 2, the Kullback-Leibler divergence  $\phi_{KL}$  is a contrast function on  $S$ . Hence, the first derivative of  $\phi_{KL}$  is a pre-contrast function on  $S$  and given by

$$\rho_{KL}((\partial_j)_{p_1}, p_2) := (\partial_j)_{p_1} \phi_{KL}(p_1, p_2) = - \int_{\mathcal{X}} s^j(\mathbf{x}, \theta_1) p(\mathbf{x}; \theta_2) \nu(d\mathbf{x})$$

for any two probability distributions  $p_1(\mathbf{x}) = p(\mathbf{x}; \theta_1)$ ,  $p_2(\mathbf{x}) = p(\mathbf{x}; \theta_2)$  in  $S$  and  $j = 1, \dots, d$ . This observation leads to the following proposition.

**Proposition 2 (Pre-contrast Functions from Estimating Functions).** *For an estimating function  $\mathbf{u}(\mathbf{x}, \theta)$  on the parametric model  $S$ , a pre-contrast function  $\rho_{\mathbf{u}} : TS \times S \rightarrow \mathbf{R}$  is defined by*

$$\rho_{\mathbf{u}}((\partial_j)_{p_1}, p_2) := - \int_{\mathcal{X}} u_*^j(\mathbf{x}, \theta_1) p(\mathbf{x}; \theta_2) \nu(d\mathbf{x})$$

*for any two probability distributions  $p_1(\mathbf{x}) = p(\mathbf{x}; \theta_1)$ ,  $p_2(\mathbf{x}) = p(\mathbf{x}; \theta_2)$  in  $S$  and  $j = 1, \dots, d$ , where  $u_*^j(\mathbf{x}, \theta)$  is the  $j$ th component of the standardization  $\mathbf{u}_*(\mathbf{x}, \theta)$  of  $\mathbf{u}(\mathbf{x}, \theta)$ .*

The use of the standardization  $\mathbf{u}_*(\mathbf{x}, \theta)$  instead of  $\mathbf{u}(\mathbf{x}, \theta)$  ensures that the definition of the function  $\rho_{\mathbf{u}}$  does not depend on the choice of coordinate system (parameter) of  $S$ . In fact, for a coordinate transformation (parameter transformation)  $\boldsymbol{\eta} = \Phi(\theta)$ , the estimating function  $\mathbf{u}(\mathbf{x}, \theta)$  is changed into  $\mathbf{v}(\mathbf{x}, \boldsymbol{\eta}) = \mathbf{u}(\mathbf{x}, \Phi^{-1}(\boldsymbol{\eta}))$  and we have  $\mathbf{v}_*(\mathbf{x}, \boldsymbol{\eta}) = (\partial \theta / \partial \boldsymbol{\eta})^T \mathbf{u}_*(\mathbf{x}, \theta)$ . The proof of

Proposition 2 is straightforward. In particular, the condition (b) in the definition of pre-contrast function follows from the unbiasedness of the (standardized) estimating function. The Riemannian metric  $g$ , dual connections  $\nabla$  and  $\nabla^*$  induced from the pre-contrast function  $\rho_{\mathbf{u}}$  are given as follows:

$$g_{jk}(\boldsymbol{\theta}) := g(\partial_j, \partial_k) = E_{\boldsymbol{\theta}}\{u_*^j(\mathbf{x}, \boldsymbol{\theta})u_*^k(\mathbf{x}, \boldsymbol{\theta})\} = G(\boldsymbol{\theta})_{jk},$$

$$\begin{cases} \Gamma_{ij,k}(\boldsymbol{\theta}) := g(\nabla_{\partial_i}\partial_j, \partial_k) = E_{\boldsymbol{\theta}}[\{\partial_i u_*^j(\mathbf{x}, \boldsymbol{\theta})\}s^k(\mathbf{x}, \boldsymbol{\theta})] \\ \Gamma_{ik,j}^*(\boldsymbol{\theta}) := g(\partial_j, \nabla_{\partial_i}^*\partial_k) = \int_{\mathcal{X}} u_*^j(\mathbf{x}, \boldsymbol{\theta})\partial_i\partial_k p(\mathbf{x}; \boldsymbol{\theta})\nu(d\mathbf{x}) \end{cases},$$

where  $G(\boldsymbol{\theta})_{jk}$  is the  $(j, k)$  component of the Godambe information matrix  $G(\boldsymbol{\theta})$ . Note that  $\nabla^*$  is always torsion-free since  $\Gamma_{ik,j}^* = \Gamma_{ki,j}^*$ , whereas  $\nabla$  is not necessarily torsion-free unless  $\mathbf{u}_*(\mathbf{x}, \boldsymbol{\theta})$  is integrable with respect to  $\boldsymbol{\theta}$ .

Henmi and Matsuzoe [7] discussed the quasi score function in [14], which is a well-known example of non-integrable estimating functions. They showed that one of the induced affine connections actually has torsion and the other connection is flat, that is, a partially flat space is induced. The pre-contrast function defined from the estimating function coincides with the canonical pre-contrast function and the generalized projection theorem can be applied. However, its statistical meaning has not been clarified yet. Although it is expected that the SMAT induced from an estimating function has something to do with statistical inference based on the estimating function, the clarification on it is a future problem.

## References

1. Eguchi, S.: Geometry of minimum contrast. *Hiroshima Math. J.* **22**, 631–647 (1992)
2. Matsuzoe, H.: Geometry of contrast functions and conformal geometry. *Hiroshima Math. J.* **29**, 175–191 (1999)
3. Amari, S., Nagaoka, H.: *Method of Information Geometry*. Amer. Math. Soc., Providence, Oxford University Press, Oxford (2000)
4. Amari, S.: *Information Geometry and Its Applications*. AMS, vol. 194. Springer, Tokyo (2016). doi:[10.1007/978-4-431-55978-8](https://doi.org/10.1007/978-4-431-55978-8)
5. Kurose, T.: *Statistical manifolds admitting torsion*. Geometry and Something, Fukuoka University (2007)
6. Matsuzoe, H.: *Statistical manifolds admitting torsion and pre-contrast functions*. Information Geometry and Its Related Fields, Osaka City University (2010)
7. Henmi, M., Matsuzoe, H.: Geometry of pre-contrast functions and non-conservative estimating functions. In: *AIP Conference Proceedings*, vol. 1340, pp. 32–41 (2011)
8. Kurose, T.: On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J.* **46**, 427–433 (1994)
9. Henmi, M., Kobayashi, R.: Hooke’s law in statistical manifolds and divergences. *Nagoya Math. J.* **159**, 1–24 (2000)
10. Ay, N., Amari, S.: A novel approach to canonical divergences within information geometry. *Entropy* **17**, 8111–8129 (2015)
11. van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge University Press, Cambridge (2000)



12. Heyde, C.C.: Quasi-Likelihood and Its Application. Springer, New York (1997). doi:[10.1007/b98823](https://doi.org/10.1007/b98823)
13. Godambe, V.: An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208–1211 (1960)
14. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman and Hall, Boca Raton (1989)