New Metric and Connections in Statistical Manifolds*

Rui F. Vigelis† David C. de Souza‡ Charles C. Cavalcante§

November 5, 2015

Abstract

We define a metric and a family of α-connections in statistical manifolds, based on φ-divergence, which emerges in the framework of φ-families of probability distributions. This metric and α-connections generalize the Fisher information metric and Amari's α-connections. We also investigate the parallel transport associated with the α-connection for $\alpha = 1$.

1 Introduction

In the framework of φ-families of probability distributions [11], the authors introduced a divergence $\mathcal{D}_\varphi(\cdot\,\|\,\cdot)$ between probabilities distributions, called *φ-divergence*, that generalizes the Kullback–Leibler divergence. Based on $\mathcal{D}_\varphi(\cdot\,\|\,\cdot)$ we can define a new metric and connections in statistical manifolds. The definition of metrics or connections in statistical manifolds is a common subject in the literature [2, 3, 7]. In our approach, the metric and α-connections are intrinsically related to φ-families. Moreover, they can be recognized as a generalization of the Fisher information metric and Amari's α-connections [1, 4].

Statistical manifolds are equipped with the Fisher information metric, which is given in terms of the derivative of $l(t;\theta) = \log p(t;\theta)$. Another metric can be defined if the logarithm $\log(\cdot)$ is replaced by the inverse of a φ-function $\varphi(\cdot)$ [11]. Instead of $l(t;\theta) = \log p(t;\theta)$, we can consider $f(t;\theta) = \varphi^{-1}(p(t;\theta))$. The manifold equipped with

*This work was partially funded by CNPq (Proc. 309055/2014-8).

†Computer Engineering, Campus Sobral, Federal University of Ceará, Sobral-CE, Brazil, rfvigelis@ufc.br.

‡Federal Institute of Ceará, Fortaleza-CE, Brazil, davidcs@ifce.edu.br.

§Wireless Telecommunication Research Group, Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza-CE, Brazil, charles@ufc.br.

this metric, which coincides with the metric derived from $\mathcal{D}_\varphi(\cdot \,\|\, \cdot)$, is called a *generalized statistical manifold*.

Using the φ-divergence $\mathcal{D}_\varphi(\cdot \,\|\, \cdot)$, we can define a pair of mutually dual connections $D^{(1)}$ and $D^{(-1)}$, and then a family of α-connections $D^{(\alpha)}$. The connections $D^{(1)}$ and $D^{(-1)}$ corresponds to the exponential and mixture connections in classical information geometry. For example, in parametric φ-families, whose definition is found in Section 2.1, the connection $D^{(1)}$ is flat (i.e, its torsion tensor T and curvature tensor R vanish identically). As a consequence, a parametric φ-family admits a parametrization in which the Christoffel symbols $\Gamma^{(-1)}_{ijk}$ associated with $D^{(-1)}$ vanish identically. In addition, parametric φ-families are examples of Hessian manifolds [8].

The rest of the paper is organized as follows. In Section 2, we define the generalized statistical manifolds. Section 2.1 deals with parametric φ-families of probability distribution. In Section 3, α-connections are introduced. The parallel transport associated with $D^{(1)}$ is investigated in Section 3.1.

2 Generalized Statistical Manifolds

In this section, we provide a definition of generalized statistical manifolds. We begin with the definition of φ-functions. Let (T, Σ, μ) be a measure space. In the case $T = \mathbb{R}$ (or T is a discrete set), the measure μ is considered to be the Lebesgue measure (or the counting measure). A function $\varphi\colon \mathbb{R} \to (0, \infty)$ is said to be a *φ-function* if the following conditions are satisfied:

(a1) $\varphi(\cdot)$ is convex,

(a2) $\lim_{u\to-\infty} \varphi(u) = 0$ and $\lim_{u\to\infty} \varphi(u) = \infty$.

Moreover, we assume that a measurable function $u_0\colon T \to (0, \infty)$ can be found such that, for each measurable function $c\colon T \to \mathbb{R}$ such that $\varphi(c(t)) > 0$ and $\int_T \varphi(c(t))d\mu = 1$, we have

(a3) $\displaystyle\int_T \varphi(c(t) + \lambda u_0(t))d\mu < \infty$, for all $\lambda > 0$.

The exponential function and the Kaniadakis' κ-exponential function [6] satisfy conditions (a1)–(a3) [11]. For $q \neq 1$, the q-exponential function $\exp_q(\cdot)$ [9] is not a φ-function, since its image is $[0, \infty)$. Notice that if the set T is finite, condition (a3) is always satisfied. Condition (a3) is indispensable in the definition of non-parametric families of probability distributions [11].

A generalized statistical manifold is a family of probability distributions $\mathcal{P} = \{p(t;\theta) : \theta \in \Theta\}$, which is defined to be contained in

$$\mathcal{P}_\mu = \left\{ p \in L^0 : p > 0 \text{ and } \int_T p d\mu = 1 \right\},$$

where L^0 denotes the set of all real-valued, measurable functions on T, with equality μ-a.e. Each $p_\theta(t) := p(t;\theta)$ is given in terms of parameters $\theta = (\theta^1, \dots, \theta^n) \in \Theta \subseteq \mathbb{R}^n$ by a one-to-one mapping. The family $\mathcal{P}$ is called a *generalized statistical manifold* if the following conditions are satisfied:

(P1) Θ is a domain (an open and connected set) in $\mathbb{R}^n$.

(P2) $p(t;\theta)$ is a differentiable function with respect to θ.

(P3) The operations of integration with respect to μ and differentiation with respect to θ^i commute.

(P4) The matrix $g = (g_{ij})$, which is defined by

$$g_{ij} = -E'_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i \partial\theta^j}\Big], \tag{1}$$

is positive definite at each $\theta \in \Theta$, where $f_\theta(t) = f(t;\theta) = \varphi^{-1}(p(t;\theta))$ and

$$E'_\theta[\cdot] = \frac{\int_T (\cdot)\varphi'(f_\theta) d\mu}{\int_T u_0 \varphi'(f_\theta) d\mu}.$$

Notice that expression (1) reduces to the Fisher information matrix in the case that φ coincides with the exponential function and $u_0 = 1$. Moreover, the right-hand side of (1) is invariant under reparametrization. The matrix (g_{ij}) can also be expressed as

$$g_{ij} = E''_\theta\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^j}\Big], \tag{2}$$

where

$$E''_\theta[\cdot] = \frac{\int_T (\cdot)\varphi''(f_\theta) d\mu}{\int_T u_0 \varphi'(f_\theta) d\mu}.$$

Because the operations of integration with respect to μ and differentiation with respect to θ^i are commutative, we have

$$0 = \frac{\partial}{\partial\theta^i}\int_T p_\theta d\mu = \int_T \frac{\partial}{\partial\theta^i}\varphi(f_\theta) d\mu = \int_T \frac{\partial f_\theta}{\partial\theta^i}\varphi'(f_\theta) d\mu, \tag{3}$$

and

$$0 = \int_T \frac{\partial^2 f_\theta}{\partial\theta^i \partial\theta^j} \varphi'(f_\theta) d\mu + \int_T \frac{\partial f_\theta}{\partial\theta^i} \frac{\partial f_\theta}{\partial\theta^j} \varphi''(f_\theta) d\mu. \tag{4}$$

Thus expression (2) follows from (4). In addition, expression (3) implies

$$E'_\theta\Big[\frac{\partial f_\theta}{\partial\theta^i}\Big] = 0. \tag{5}$$

A consequence of (2) is the correspondence between the functions $\partial f_\theta / \partial\theta^i$ and the basis vectors $\partial/\partial\theta^i$. The inner product of vectors

$$X = \sum_i a^i \frac{\partial}{\partial\theta^i} \qquad \text{and} \qquad Y = \sum_i b^j \frac{\partial}{\partial\theta^j}$$

can be written as

$$g(X,Y) = \sum_{i,j} g_{ij} a^i b^j = \sum_{i,j} E''_\theta\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^j}\Big] a^i b^j = E''_\theta[\widetilde{X}\widetilde{Y}], \tag{6}$$

where

$$\widetilde{X} = \sum_i a^i \frac{\partial f_\theta}{\partial\theta^i} \qquad \text{and} \qquad \widetilde{Y} = \sum_i b^j \frac{\partial f_\theta}{\partial\theta^j}.$$

As a result, the tangent space $T_{p_\theta}\mathcal{P}$ can be identified with $\widetilde{T}_{p_\theta}\mathcal{P}$, which is defined as the vector space spanned by $\partial f_\theta/\partial\theta^i$, equipped with the inner product $\langle \widetilde{X}, \widetilde{Y}\rangle_\theta = E''_\theta[\widetilde{X}\widetilde{Y}]$. By (5), if a vector $\widetilde{X}$ belongs to $\widetilde{T}_{p_\theta}\mathcal{P}$, then $E'_\theta[\widetilde{X}] = 0$. Independent of the definition of (g_{ij}), the expression in the right-hand side of (6) always defines a semi-inner product in $\widetilde{T}_{p_\theta}\mathcal{P}$.

2.1 Parametric φ-Families of Probability Distribution

Let $c\colon T \to \mathbb{R}$ be a measurable function such that $p := \varphi(c)$ is probability density in $\mathcal{P}_\mu$. We take any measurable functions $u_1, \dots u_n\colon T \to \mathbb{R}$ satisfying the following conditions:

(i) $\int_T u_i \varphi'(c) d\mu = 0$, and

(ii) there exists $\varepsilon > 0$ such that

$$\int_T \varphi(c + \lambda u_i) d\mu < \infty, \qquad \text{for all } \lambda \in (-\varepsilon, \varepsilon).$$

Define $\Theta \subseteq \mathbb{R}^n$ as the set of all vectors $\theta = (\theta^i) \in \mathbb{R}^n$ such that

$$\int_T \varphi\Big(c + \lambda \sum_{k=1}^n \theta^i u_i\Big) d\mu < \infty, \qquad \text{for some } \lambda > 1.$$

The elements of the *parametric φ-family* $\mathcal{F}_p = \{p(t;\theta) : \theta \in \Theta\}$ centered at $p = \varphi(c)$ are given by the one-to-one mapping

$$p(t;\theta) := \varphi\biggl(c(t) + \sum_{i=1}^{n} \theta^i u_i(t) - \psi(\theta) u_0(t)\biggr), \qquad \text{for each } \theta = (\theta^i) \in \Theta. \tag{7}$$

where the *normalizing function* $\psi\colon \Theta \to [0,\infty)$ is introduced so that expression (7) defines a probability distribution in $\mathcal{P}_\mu$.

Condition (ii) is always satisfied if the set T is finite. It can be shown that the normalizing function ψ is also convex (and the set Θ is open and convex). Under conditions (i)–(ii), the family $\mathcal{F}_p$ is a submanifold of a non-parametric φ-family. For the non-parametric case, we refer to [11, 10].

By the equalities

$$\frac{\partial f_\theta}{\partial \theta^i} = u_i(t) - \frac{\partial \psi}{\partial \theta^i}, \qquad -\frac{\partial^2 f_\theta}{\partial \theta^i \partial \theta^j} = -\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j},$$

we get

$$g_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}.$$

In other words, the matrix (g_{ij}) is the Hessian of the normalizing function ψ.

For $\varphi(\cdot) = \exp(\cdot)$ and $u_0 = 1$, expression (7) defines a parametric exponential family of probability distributions $\mathcal{E}_p$. In exponential families, the normalizing function is recognized as the Kullback–Leibler divergence between $p(t)$ and $p(t;\theta)$. Using this result, we can define the φ-divergence $\mathcal{D}_\varphi(\cdot \parallel \cdot)$, which generalizes the Kullback–Leibler divergence $\mathcal{D}_{\mathrm{KL}}(\cdot \parallel \cdot)$.

By (7) we can write

$$\psi(\theta) u_0(t) = \sum_{i=1}^{n} \theta^i u_i(t) + \varphi^{-1}(p(t)) - \varphi^{-1}(p(t;\theta)).$$

From condition (i), this equation yields

$$\psi(\theta) \int_T u_0 \varphi'(c) d\mu = \int_T [\varphi^{-1}(p) - \varphi^{-1}(p_\theta)] \varphi'(c) d\mu.$$

In view of $\varphi'(c) = 1/(\varphi^{-1})'(p)$, we get

$$\psi(\theta) = \frac{\displaystyle\int_T \frac{\varphi^{-1}(p) - \varphi^{-1}(p_\theta)}{(\varphi^{-1})'(p)} d\mu}{\displaystyle\int_T \frac{u_0}{(\varphi^{-1})'(p)} d\mu} =: \mathcal{D}_\varphi(p \parallel p_\theta), \tag{8}$$

which defines the φ-divergence $\mathcal{D}_\varphi(p \,\|\, p_\theta)$. Clearly, expression (8) can be used to extend the definition of $\mathcal{D}_\varphi(\cdot \,\|\, \cdot)$ to any probability distributions p and q in $\mathcal{P}_\mu$.

3 α-Connections

We use the φ-divergence $\mathcal{D}_\varphi(\cdot \,\|\, \cdot)$ to define a pair of mutually dual connection in generalized statistical manifolds. Let $\mathcal{D}\colon M \times M \to [0,\infty)$ be a non-negative, differentiable function defined on a smooth manifold M, such that

$$\mathcal{D}(p \,\|\, q) = 0 \quad \text{if and only if} \quad p = q. \tag{9}$$

The function $\mathcal{D}(\cdot \,\|\, \cdot)$ is called a *divergence* if the matrix (g_{ij}), whose entries are given by

$$g_{ij}(p) = -\Big[\Big(\frac{\partial}{\partial\theta^i}\Big)_p\Big(\frac{\partial}{\partial\theta^j}\Big)_q \mathcal{D}(p \,\|\, q)\Big]_{q=p}, \tag{10}$$

is positive definite for each $p \in M$. Hence a divergence $\mathcal{D}(\cdot \,\|\, \cdot)$ defines a metric in M. A divergence $\mathcal{D}(\cdot \,\|\, \cdot)$ also induces a pair of mutually dual connections D and D^*, whose Christoffel symbols are given by

$$\Gamma_{ijk} = -\Big[\Big(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\Big)_p\Big(\frac{\partial}{\partial\theta^k}\Big)_q \mathcal{D}(p \,\|\, q)\Big]_{q=p} \tag{11}$$

and

$$\Gamma^*_{ijk} = -\Big[\Big(\frac{\partial}{\partial\theta^k}\Big)_p\Big(\frac{\partial^2}{\partial\theta^i\partial\theta^j}\Big)_q \mathcal{D}(p \,\|\, q)\Big]_{q=p}, \tag{12}$$

respectively. By a simple computation, we get

$$\frac{\partial g_{jk}}{\partial\theta^i} = \Gamma_{ijk} + \Gamma^*_{ikj},$$

showing that D and D^* are mutually dual.

In Section 2.1, the φ-divergence between two probability distributions p and q in $\mathcal{P}_\mu$ was defined as

$$\mathcal{D}_\varphi(p \,\|\, q) := \frac{\displaystyle\int_T \frac{\varphi^{-1}(p) - \varphi^{-1}(q)}{(\varphi^{-1})'(p)} d\mu}{\displaystyle\int_T \frac{u_0}{(\varphi^{-1})'(p)} d\mu}. \tag{13}$$

Because φ is convex, it follows that $\mathcal{D}_\varphi(p \,\|\, q) \geq 0$ for all $p, q \in \mathcal{P}_\mu$. In addition, if we assume that $\varphi(\cdot)$ is strictly convex, then $\mathcal{D}_\varphi(p \,\|\, q) = 0$ if and only if $p = q$. In a generalized statistical manifold $\mathcal{P} = \{p(t;\theta) : \theta \in \Theta\}$, the metric derived from the divergence $\mathcal{D}(q \,\|\, p) := \mathcal{D}_\varphi(p \,\|\, q)$ coincides with (1). Expressing the φ-divergence $\mathcal{D}_\varphi(\cdot \,\|\, \cdot)$

between p_θ and p_ϑ as

$$\mathcal{D}(p_\theta \parallel p_\vartheta) = E'_\vartheta[(f_\vartheta - f_\theta)],$$

after some manipulation, we get

$$\begin{aligned} g_{ij} &= -\Big[\Big(\frac{\partial}{\partial\theta^i}\Big)_p\Big(\frac{\partial}{\partial\theta^j}\Big)_q \mathcal{D}(p \parallel q)\Big]_{q=p} \\ &= -E'_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\Big]. \end{aligned}$$

As a consequence, expression (13) defines a divergence on statistical manifolds.

Let $D^{(1)}$ and $D^{(-1)}$ denote the pair of dual connections derived from $\mathcal{D}_\varphi(\cdot \parallel \cdot)$. By (11) and (12), the Christoffel symbols $\Gamma^{(1)}_{ijk}$ and $\Gamma^{(-1)}_{ijk}$ are given by

$$\Gamma^{(1)}_{ijk} = E''_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big] - E'_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\Big]E''_\theta\Big[u_0\frac{\partial f_\theta}{\partial\theta^k}\Big] \tag{14}$$

and

$$\begin{aligned} \Gamma^{(-1)}_{ijk} &= E''_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big] + E'''_\theta\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big] \\ &\quad - E''_\theta\Big[\frac{\partial f_\theta}{\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big]E''_\theta\Big[u_0\frac{\partial f_\theta}{\partial\theta^i}\Big] - E''_\theta\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^k}\Big]E''_\theta\Big[u_0\frac{\partial f_\theta}{\partial\theta^j}\Big], \end{aligned} \tag{15}$$

where

$$E'''_\theta[\cdot] = \frac{\int_T (\cdot)\varphi'''(f_\theta)d\mu}{\int_T u_0\varphi'(f_\theta)d\mu}.$$

Notice that in parametric φ-families, the Christoffel symbols $\Gamma^{(1)}_{ijk}$ vanish identically. Thus, in these families, the connection $D^{(1)}$ is flat.

Using the pair of mutually dual connections $D^{(1)}$ and $D^{(-1)}$, we can specify a family of α-connections $D^{(\alpha)}$ in generalized statistical manifolds. The Christoffel symbol of $D^{(\alpha)}$ is defined by

$$\Gamma^{(\alpha)}_{ijk} = \frac{1+\alpha}{2}\Gamma^{(1)}_{ijk} + \frac{1-\alpha}{2}\Gamma^{(-1)}_{ijk}. \tag{16}$$

The connections $D^{(\alpha)}$ and $D^{(-\alpha)}$ are mutually dual, since

$$\frac{\partial g_{jk}}{\partial\theta^i} = \Gamma^{(\alpha)}_{ijk} + \Gamma^{(-\alpha)}_{ikj}.$$

For $\alpha = 0$, the connection $D^{(0)}$, which is clearly self-dual, corresponds to the Levi–Civita connection ∇. One can show that $\Gamma^{(0)}_{ijk}$ can be derived from the expression defining the

Christoffel symbols of ∇ in terms of the metric:

$$\Gamma_{ijk} = \sum_m \Gamma_{ij}^m g_{mk} = \frac{1}{2}\Big(\frac{\partial g_{ki}}{\partial\theta^j} + \frac{\partial g_{kj}}{\partial\theta^i} - \frac{\partial g_{ij}}{\partial\theta^k}\Big).$$

The connection $D^{(\alpha)}$ can be equivalently defined by

$$\Gamma_{ijk}^{(\alpha)} = \Gamma_{ijk}^{(0)} - \alpha T_{ijk},$$

where

$$\begin{aligned} T_{ijk} = \frac{1}{2}E_\theta'''\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big] - \frac{1}{2}E_\theta''\Big[\frac{\partial f_\theta}{\partial\theta^k}\frac{\partial f_\theta}{\partial\theta^i}\Big]E_\theta''\Big[u_0\frac{\partial f_\theta}{\partial\theta^j}\Big] \\ - \frac{1}{2}E_\theta''\Big[\frac{\partial f_\theta}{\partial\theta^k}\frac{\partial f_\theta}{\partial\theta^j}\Big]E_\theta''\Big[u_0\frac{\partial f_\theta}{\partial\theta^i}\Big] - \frac{1}{2}E_\theta''\Big[\frac{\partial f_\theta}{\partial\theta^i}\frac{\partial f_\theta}{\partial\theta^j}\Big]E_\theta''\Big[u_0\frac{\partial f_\theta}{\partial\theta^k}\Big]. \quad (17) \end{aligned}$$

In the case that φ is the exponential function and $u_0 = 1$, equations (14), (15), (16) and (17) reduce to the classical expressions for statistical manifolds.

3.1 Parallel Transport

Let $\gamma\colon I \to M$ be a smooth curve in a smooth manifold M, with a connection D. A vector field V along γ is said to be *parallel* if $D_{d/dt}V(t) = 0$ for all $t \in I$. Take any tangent vector V_0 at $\gamma(t_0)$, for some $t_0 \in I$. Then there exists a unique vector field V along γ, called the *parallel transport* of V_0 along γ, such that $V(t_0) = V_0$.

A connection D can be recovered from the parallel transport. Fix any smooth vectors fields X and Y. Given $p \in M$, define $\gamma\colon I \to M$ to be an integral curve of X passing through p. In other words, $\gamma(t_0) = p$ and $\frac{d\gamma}{dt} = X(\gamma(t))$. Let $P_{\gamma,t_0,t}\colon T_{\gamma(t_0)}M \to T_{\gamma(t)}M$ denote the parallel transport of a vector along γ from t_0 to t. Then we have

$$(D_X Y)(p) = \frac{d}{dt}P_{\gamma,t_0,t}^{-1}(Y(c(t))\Big|_{t=t_0}.$$

For details, we refer to [5].

To avoid some technicalities, we assume that the set T is finite. In this case, we can consider a generalized statistical manifold $\mathcal{P} = \{p(t;\theta) : \theta \in \Theta\}$ for which $\mathcal{P} = \mathcal{P}_\mu$. The connection $D^{(1)}$ can be derived from the parallel transport

$$P_{q,p}\colon \widetilde{T}_q\mathcal{P} \to \widetilde{T}_p\mathcal{P}$$

given by

$$\widetilde{X} \mapsto \widetilde{X} - E_\theta'[\widetilde{X}]u_0,$$

where $p = p_\theta$. Recall that the tangent space $T_p\mathcal{P}$ can be identified with $\widetilde{T}_p\mathcal{P}$, the vector space spanned by the functions $\partial f_\theta/\partial\theta^i$, equipped with the inner product $\langle \widetilde{X}, \widetilde{Y}\rangle = E''_\theta[\widetilde{X}\widetilde{Y}]$, where $p = p_\theta$. We remark that $P_{q,p}$ does not depend on the curve joining q and p. As a result, the connection $D^{(1)}$ is flat. Denote by $\gamma(t)$ the coordinate curve given locally by $\theta(t) = (\theta^1, \ldots, \theta^i + t, \ldots, \theta^n)$. Observing that $P^{-1}_{\gamma(0),\gamma(t)}$ maps the vector $\frac{\partial f_\theta}{\partial\theta^j}(t)$ to

$$\frac{\partial f_\theta}{\partial\theta^j}(t) - E'_{\theta(0)}\Big[\frac{\partial f_\theta}{\partial\theta^j}(t)\Big]u_0,$$

we define the connection

$$\begin{aligned}
\widetilde{D}_{\partial f_\theta/\partial\theta_i}\frac{\partial f_\theta}{\partial\theta_j} &= \frac{d}{dt}P^{-1}_{\gamma(0),\gamma(t)}\Big(\frac{\partial f_\theta}{\partial\theta_j}(\gamma(t))\Big)\Big|_{t=0} \\
&= \frac{d}{dt}\Big(\frac{\partial f_{\theta(t)}}{\partial\theta^j} - E'_{\theta(0)}\Big[\frac{\partial f_{\theta(t)}}{\partial\theta^j}\Big]u_0\Big)\Big|_{t=0} \\
&= \frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j} - E'_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\Big]u_0.
\end{aligned}$$

Let us denote by D the connection corresponding to $\widetilde{D}$, which acts on smooth vector fields in $T_p\mathcal{P}$. By this identification, we have

$$\begin{aligned}
g\Big(D_{\partial/\partial\theta_i}\frac{\partial}{\partial\theta_j}, \frac{\partial}{\partial\theta_k}\Big) &= \Big\langle \widetilde{D}_{\partial f_\theta/\partial\theta_i}\frac{\partial f_\theta}{\partial\theta_j}, \frac{\partial f_\theta}{\partial\theta_k}\Big\rangle \\
&= E''_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\frac{\partial f_\theta}{\partial\theta^k}\Big] - E'_\theta\Big[\frac{\partial^2 f_\theta}{\partial\theta^i\partial\theta^j}\Big]E''_\theta\Big[u_0\frac{\partial f_\theta}{\partial\theta^k}\Big] \\
&= \Gamma^{(1)}_{ijk},
\end{aligned}$$

showing that $D = D^{(1)}$.

References

[1] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000. Translated from the 1993 Japanese original by Daishi Harada.

[2] Shun-ichi Amari and Atsumi Ohara. Geometry of q-exponential family of probability distributions. *Entropy*, 13(6):1170–1185, 2011.

[3] Shun-ichi Amari, Atsumi Ohara, and Hiroshi Matsuzoe. Geometry of deformed exponential families: invariant, dually-flat and conformal geometries. *Phys. A*, 391(18):4308–4319, 2012.

[4] Ovidiu Calin and Constantin Udrişte. *Geometric Modeling in Probability and Statistics*. Springer, 2014.

[5] Mafredo P. do Carmo. *Riemannian Geometry*. Birkhäuser, 14th edition, 2013.

[6] G. Kaniadakis. Statistical mechanics in the context of special relativity. *Phys. Rev. E (3)*, 66(5):056125, 17, 2002.

[7] Hiroshi Matsuzoe. Hessian structures on deformed exponential families and their conformal structures. *Differential Geom. Appl.*, 35(suppl.):323–333, 2014.

[8] Hirohiko Shima. *The geometry of Hessian structures*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[9] Constantino Tsallis. What are the numbers that experiments provide? *Quimica Nova*, 17(6):468–471, 1994.

[10] Rui F. Vigelis and Charles C. Cavalcante. The Δ_2-condition and φ-families of probability distributions. In *Geometric science of information*, volume 8085 of *Lecture Notes in Comput. Sci.*, pages 729–736. Springer, Heidelberg, 2013.

[11] Rui F. Vigelis and Charles C. Cavalcante. On φ-families of probability distributions. *J. Theoret. Probab.*, 26(3):870–884, 2013.