From Divergence Function to Information Geometry: Metric, Equiaffine, and Symplectic Structures

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

(In collaboration with Hiroshi Matsuzoe and Fubo Li.)

1 Divergence Functions

Divergence functons are fundamental objects in *Information Geometry*, the differential geometric study of the manifold of (parametric or non-parametric) probability distributions (see Amari, 1985; Amari and Nagaoka, 2000). They measure the directed (asymmetric) difference between two points on this manifold, where each point represents a probability function or a vector in the parametric space. This paper will review and report new results on how divergence functions are used to construct geometric structures on a manifold, namely the Riemannian structure on the tangent bundle and the symplectic structure on the cotangent bundle.

1.1 Definition and examples

Definition. A divergence function $\mathcal{D} : V \times V \to \mathbb{R}_{\geq 0}$ on a convex set $V \subseteq \mathbb{R}^n$ is a smooth function (differentiable up to third order) which satisfies

(i) $\mathcal{D}(x, y) \geq 0 \ \forall x, y \in V$ with equality holding if and only if $x = y$;

(ii) $\mathcal{D}_i(x, x) = \mathcal{D}_{,j}(x, x) = 0, \forall i, j \in \{1, 2, \cdot, n\}$;

(iii) $-\mathcal{D}_{i,j}(x, x)$ is positive definite.

Here $\mathcal{D}_i(x, y) = \partial_{x^i} D(x, y)$, $\mathcal{D}_{,i}(x, y) = \partial_{y^i}\mathcal{D}(x, y)$ denote partial derivatives with respect to the i-th component of the x-variable and of the y-variable, respectively, $\mathcal{D}_{i,j}(x, y) = \partial_{x^i}\partial_{y^j}\mathcal{D}(x, y)$ denotes the second-order mixed derivative, etc. The following are familiar examples.
Bregman divergence (Bregman, 1967). $\mathcal{B}_\Phi(x, y)$ is induced by a smooth and strictly convex function $\Phi : \mathbb{R}^n \to \mathbb{R}$:

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle x - y, \partial\Phi(y)\rangle \tag{1.1}$$

where $\partial\Phi = [\partial_1\Phi, \cdots, \partial_n\Phi]$ and $\langle\cdot,\cdot\rangle_n$ denotes the canonical pairing of a point/vector $x = [x^1, \cdots, x^n] \in V$ and a point/co-vector $u = [u_1, \cdots, u_n] \in \widetilde{V}$ (dual to V): $\langle x, u\rangle_n = \sum_{i=1}^n x^i u_i$.
Kullback-Leibler divergence (Kullback, 1957). $\mathcal{K}(x, y)$ plays a fundamental role in statistical inference:

$$\mathcal{K}(x, y) = \sum_i \left(y^i - x^i - x^i \log \frac{y^i}{x^i} \right) = \mathcal{K}^*(y, x) \ . \tag{1.2}$$

f-divergence (Csiszár, 1967). For a convex function $f_c : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $f_c(1) = 0, f_c'(1) = 0$:

$$\mathcal{F}_{f_c}(x, y) = \sum_i x^i \, f_c\left(\frac{y^i}{x^i}\right) \ . \tag{1.3}$$

Alpha divergence (Amari 1982; 1985) A parametric family $\mathcal{A}^{(\alpha)}$ of divergence functions :

$$\mathcal{A}^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2} \sum_i \left(\frac{1-\alpha}{2} x^i + \frac{1+\alpha}{2} y^i - (x^i)^{\frac{1-\alpha}{2}} (y^i)^{\frac{1+\alpha}{2}} \right) \ , \quad \alpha \in \mathbb{R} \ . \tag{1.4}$$

Note that $\lim_{\alpha \to 1} \mathcal{A}^{(\alpha)}(x, y) = \mathcal{K}(x, y), \lim_{\alpha \to -1} \mathcal{A}^{(\alpha)}(x, y) = \mathcal{K}(y, x)$.
Jenson Difference (Rao, 1987). A family of divergence functions (we add the $\frac{4}{1-\alpha^2}$ factor here)

$$\mathcal{E}^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2} \sum_i \left(\frac{1-\alpha}{2} x^i \log \frac{x^i}{\frac{1-\alpha}{2} x^i + \frac{1+\alpha}{2} y^i} + \frac{1+\alpha}{2} y^i \log \frac{y^i}{\frac{1-\alpha}{2} x^i + \frac{1+\alpha}{2} y^i} \right), \alpha \in \mathbb{R} \tag{1.5}$$

that also recovers the Kullback-Leibler divergence (1.2) by letting $\alpha \to \pm 1$.
Alpha-beta divergence (Zhang, 2004). A two-parameter family $\mathcal{D}^{\alpha,\beta}$ of divergence functions

$$\mathcal{D}^{(\alpha,\beta)}(x, y) = \frac{4}{1-\alpha^2} \frac{2}{1+\beta} \sum_i \left(\frac{1-\alpha}{2} x^i + \frac{1+\alpha}{2} y^i - \left(\frac{1-\alpha}{2} (x^i)^{\frac{1-\beta}{2}} + \frac{1+\alpha}{2} (y^i)^{\frac{1-\beta}{2}} \right)^{\frac{2}{1-\beta}} \right) \tag{1.6}$$

with $\alpha, \beta \in \mathbb{R}$, which reduces to (1.4) when $\alpha \to \pm 1$ or $\beta \to -1$, and to (1.5) when $\beta \to 1$.

1.2 Divergence Induced by Convex Functions

To allow unifying treatment of divergence functions and the geometrical structures they generate, Zhang (2004) proposed a general family of divergece functions based on convex analysis. A strictly convex (or simply "convex") function $\Phi : V \subseteq \mathbb{R}^n \to \mathbb{R}, x \mapsto \Phi(x)$ is defined on a non-empty convex set $S = \operatorname{int} \operatorname{dom}(\Phi) \subseteq \mathbb{R}^n$ such that for any two points $x \in S$, $y \in S$ and any real number $\alpha \in (-1, 1)$, the following is valid

$$\frac{1-\alpha}{2} \Phi(x) + \frac{1+\alpha}{2} \Phi(y) - \Phi\left(\frac{1-\alpha}{2} x + \frac{1+\alpha}{2} y \right) \geq 0 \tag{1.7}$$

The inequality sign is reversed when $|\alpha| > 1$) (with equality holding only when $x = y$). We assume Φ to be sufficiently smooth (differentiable up to fourth order). A family of functions on $V \times V$, as indexed by $\alpha \in \mathbb{R}$, were introduced by Zhang (2004) as divergence functions

$$\mathcal{D}_\Phi^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2} \left(\frac{1-\alpha}{2} \Phi(x) + \frac{1+\alpha}{2} \Phi(y) - \Phi\left(\frac{1-\alpha}{2} x + \frac{1+\alpha}{2} y \right) \right). \tag{1.8}$$

Clearly, $\mathcal{D}_\Phi^{(\alpha)}(x, y) = \mathcal{D}_\Phi^{(-\alpha)}(y, x)$, and $\mathcal{D}_\Phi^{(\pm 1)}(x, y)$ is defined by taking $\lim_{\alpha \to \pm 1}$:

$$\begin{aligned} \mathcal{D}_\Phi^{(1)}(x, y) &= \mathcal{D}_\Phi^{(-1)}(y, x) &= \mathcal{B}_\Phi(x, y), \\ \mathcal{D}_\Phi^{(-1)}(x, y) &= \mathcal{D}_\Phi^{(1)}(y, x) &= \mathcal{B}_\Phi(y, x). \end{aligned}$$

Lemma 1.1. *For a smooth function $\Phi : V \subseteq \mathbb{R}^n \to \mathbb{R}$, the following are equivalent:*

(i) Φ *is strictly convex;*

(ii) $\mathcal{D}^{(1)}_{\Phi}(x, y) \geq 0, \forall x, y \in V$*;*

(iii) $\mathcal{D}^{(-1)}_{\Phi}(x, y) \geq 0, \forall x, y \in V$*;*

(iv) $\mathcal{D}^{(\alpha)}_{\Phi}(x, y) \geq 0$ *for all* $|\alpha| < 1$ *and* $\forall x, y \in V$*;*

(v) $\mathcal{D}^{(\alpha)}_{\Phi}(x, y) \geq 0$ *for all* $|\alpha| > 1$ *and* $\forall x, y \in V$*.*

Denote the convex conjugate $\widetilde{\Phi} : \widetilde{V} \subseteq \mathbb{R}^n \to \mathbb{R}$ defined through the Legendre-Fenchel transform:

$$\widetilde{\Phi}(u) = \langle (\partial\Phi)^{-1}(u), u\rangle - \Phi((\partial\Phi)^{-1}(u)), \tag{1.9}$$

with $\widetilde{\widetilde{\Phi}} = \Phi$ and $(\partial\Phi) = (\partial\widetilde{\Phi})^{-1}$. The function $\widetilde{\Phi}$ is also convex, and satisfy the Fenchel inequaliy:

$$\Phi(x) + \widetilde{\Phi}(u) - \langle x, u\rangle \geq 0$$

for any $x \in V$, $u \in \widetilde{V}$, with equality holding if and only if (below "leftrightarrow" means "equivalently")

$$u = (\partial\Phi)(x) = (\partial\widetilde{\Phi})^{-1}(x) \longleftrightarrow x = (\partial\widetilde{\Phi})(u) = (\partial\Phi)^{-1}(u), \tag{1.10}$$

or, in component form,

$$u_i = \frac{\partial\Phi}{\partial x^i} \longleftrightarrow x^i = \frac{\partial\widetilde{\Phi}}{\partial u_i}\,. \tag{1.11}$$

Finally, introduce "canonical divergence" $\mathcal{A}_\Phi : V \times \widetilde{V} \to \mathbb{R}_{\geq 0}$ (and $\mathcal{A}_{\widetilde{\Phi}} : \widetilde{V} \times V \to \mathbb{R}_{\geq 0}$)

$$\mathcal{A}_\Phi(x, v) = \Phi(x) + \widetilde{\Phi}(v) - \langle x, v\rangle = \mathcal{A}_{\widetilde{\Phi}}(v, x).$$

They are related to the Bregman divergence (1.1) via

$$\mathcal{B}_\Phi(x, (\partial\Phi)^{-1}(v)) = \mathcal{A}_\Phi(x, v) = \mathcal{B}_{\widetilde{\Phi}}((\partial\widetilde{\Phi})(x), v).$$

2 Riemannian Structure Induced by Divergence Functions

2.1 Metric, conjugate connections, and equiaffine volumn forms

Let $\mathfrak{M}$ be a smooth manifold endowed with a Riemannian metric g:

$$g_{ij}(x) = g(\partial_i, \partial_j) \tag{2.1}$$

written in local coordinates $\partial_i \equiv \partial/\partial x^i$ of the tangent space $\mathcal{T}\mathfrak{M}_x$. An affine connection ∇, written out in its "contravariant" form Γ^l_{ij}, is

$$\nabla_{\partial_i}\partial_j = \sum_l \Gamma^l_{ij}\,\partial_l. \tag{2.2}$$

The torsion T of a connection ∇ is given by $T(\partial_i, \partial_j) = \nabla_{\partial_i}\partial_j - \nabla_{\partial_j}\partial_i = \sum_k T_{ij}^k \partial_k$, where

$$T_{ij}^k(x) = \Gamma_{ij}^k(x) - \Gamma_{ji}^k(x).$$

The curvature R of a connection ∇ is given by $R(\partial_i, \partial_j)\partial_k = (\nabla_{\partial_i}\nabla_{\partial_j} - \nabla_{\partial_j}\nabla_{\partial_i})\partial_k = \sum_l R_{kij}^l \partial_l$ with

$$R_{kij}^l(x) = \frac{\partial \Gamma_{jk}^l(x)}{\partial x^i} - \frac{\partial \Gamma_{ik}^l(x)}{\partial x^j} + \sum_m \Gamma_{im}^l(x)\Gamma_{jk}^m(x) - \sum_m \Gamma_{jm}^l(x)\Gamma_{ik}^m(x).$$

The covariant forms of Γ and R are

$$\Gamma_{ij,k} = g(\nabla_{\partial_i}\partial_j, \partial_k) = \sum_l g_{lk}\Gamma_{ij}^l, \quad R_{lkij} = \sum_m g_{lm}\, R_{kij}^m. \tag{2.3}$$

When Γ is torsion free, R_{lkij} is anti-symmetric when $i \leftrightarrow j$ or when $k \leftrightarrow l$, and symmetric when $(i,j) \leftrightarrow (l,k)$. The Ricci tensor Ric is

$$\mathrm{Ric}_{ij}(x) = \sum_k R_{ikj}^k(x) = \sum_{k,l} R_{likj} g^{kl} \ . \tag{2.4}$$

For torsion-free connections $\Gamma_{ij}^k = \Gamma_{ji}^k$, we can further consider a restrictive class, called "equiaffine" connections ∇, when the manifold $\mathfrak{M}$ (with such ∇) may admit uniquely a volume form Ω that is parallel under ∇. Here, a volume form is a non-degenerate, skew-symmetric multilinear map from $V\times, \cdots, \times V \to \mathbb{R}$, and "parallel" is in the sense that $(\partial_i\Omega)(\partial_1, \cdots, \partial_n) = 0$ where

$$(\partial_i\Omega)(\partial_1, \cdots, \partial_n) \equiv (\nabla_{\partial_i}\Omega)(\partial_1, \cdots, \partial_n) = \partial_i(\Omega(\partial_1, \cdots, \partial_n)) - \sum_{l=1}^n \Omega(\cdots, \nabla_{\partial_i}\partial_l, \cdots). \tag{2.5}$$

Applying (2.2), the equiaffine condition becomes

$$\partial_i(\Omega(\partial_1, \cdots, \partial_n)) = \sum_{l=1}^n \Omega\left(\cdots, \sum_{k=1}^n \Gamma_{il}^k \partial_k, \cdots\right) = \sum_{l,k=1}^n \Gamma_{il}^k\, \delta_k^l\, \Omega(\partial_1, \cdots, \partial_n) = \Omega(\partial_1, \cdots, \partial_n) \sum_{l=1}^n \Gamma_{il}^l \tag{2.6}$$

or

$$\sum_l \Gamma_{il}^l(x) = \frac{\partial \log \Omega(x)}{\partial x^i}. \tag{2.7}$$

For torsion-free connections, the equiaffine condition is equivalent to Ricci tensor being symmetric:

$$\mathrm{Ric}_{ij} - \mathrm{Ric}_{ji} = \frac{\partial}{\partial x^i}\left(\sum_l \Gamma_{jl}^l(x)\right) - \frac{\partial}{\partial x^j}\left(\sum_l \Gamma_{il}^l(x)\right) \ . \tag{2.8}$$

Eqivalently, $\sum_k R_{kij}^k = 0$. One can check that the Levi-Civita connection is always equiaffine.

Let us now revisit the concept of "conjugate connection". A connection ∇^* is said to be "conjugate" to ∇ with respect to g if

$$\partial_k g(\partial_i, \partial_j) = g(\nabla_{\partial_k}\partial_i, \partial_j) + g(\partial_i, \nabla^*_{\partial_k}\partial_j). \tag{2.9}$$

Clearly, $(\nabla^*)^* = \nabla$. Moreover, the Levi-Civita connection $\widehat{\nabla} = (\nabla + \nabla^*)/2$ is self-conjugate, with the parallel volume form given by

$$\widehat{\Omega}(x) = \sqrt{\det[g_{ij}(x)]} \longleftrightarrow \widehat{\Omega}(u) = \sqrt{\det[g^{ij}(u)]}.$$

Writing out (2.9) explicitly,

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ki,j} + \Gamma^*_{kj,i}, \tag{2.10}$$

where analogous to (2.2) and (2.3),

$$\nabla^*_{\partial_i} \partial_j = \sum_l \Gamma^{*l}_{ij}\, \partial_l$$

so that

$$\Gamma^*_{kj,i} = g(\nabla^*_{\partial_j}\partial_k, \partial_i) = \sum_l g_{il}\Gamma^{*l}_{kj}.$$

It can be shown that the curvatures R_{lkij} and R^*_{lkij} for ∇, ∇^* satisfy

$$R_{lkij} = R^*_{lkij}.$$

So, ∇ is flat if and only if ∇^* is flat. In this case, the manifold $\{\mathfrak{M}, g, \gamma, \gamma^*\}$ is said to be "dually flat"; it is called a "Hessian manifold" (Shima, 2001; Shima and Yagi, 1997). In general, when ∇, ∇^* are not dually flat but nevertheless are torsion-free (the torsion-freeness of one implies the torsion-freeness of the other), $\{\mathfrak{M}, g, \gamma, \gamma^*\}$ is called a "statistical manifold' (Lauritzen, 1987)'; they have been in affine differential geometry (Nomizu and Sasaki, 1994), quite seprately from the development of information geometry. Statistical manifolds admit a one-parameter family of affine connections $\nabla^{(\alpha)}$, called "α-connections" ($\alpha \in \mathbb{R}$) introduced by Amari (1985):

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2}\nabla + \frac{1-\alpha}{2}\nabla^*. \tag{2.11}$$

Obviously, $\nabla^{(0)} = \widehat{\nabla}$. When ∇, ∇^* are dually flat, $\Gamma^{(\alpha)}$ is called "α-transitively flat" (Uohashi, 2002). In such case, we call $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$ an "α-Hessian manifold". Conjugate connections which admit torsion has been recently studied by H. Matsuzoe.

2.2 Biorthogonal coordinates

Consider coordinate transform $x \mapsto u$,

$$\partial^i \equiv \frac{\partial}{\partial u_i} = \sum_l \frac{\partial x^l}{\partial u_i}\frac{\partial}{\partial x^l} = \sum_l F^{li}\partial_l$$

where the Jacobian matrix F is given by

$$F_{ij}(x) = \frac{\partial u_i}{\partial x^j}, \quad F^{ij}(u) = \frac{\partial x^i}{\partial u_j}, \quad \sum_l F_{il}F^{lj} = \delta^l_k \tag{2.12}$$

where δ^j_i is Kronecker delta (taking the value of 1 when $i = j$ and 0 otherwise). If the new coordinate system $u = [u_1, \cdots, u_n]$ (with components expressed by subscripts) is such that

$$F_{ij}(x) = g_{ij}(x), \tag{2.13}$$

then the x-coordinate system and the u-coordinate system are said to be "biorthogonal" to each other since, from the definition of metric tensor (2.1),

$$g(\partial_i, \partial^j) = g(\partial_i, \sum_l F^{lj}\partial_l) = \sum_l F^{lj} g(\partial_i, \partial_l) = \sum_l F^{lj} g_{il} = \delta_i^j.$$

In such a case, denote

$$g^{ij}(u) = g(\partial^i, \partial^j), \tag{2.14}$$

which equals $F^{ij}(u)$, the Jacobian of the inverse coordinate transform $u \mapsto x$. Also introduce the (contravariant version) of the affine connection ∇ under the u-coordinate system and denote it by an unconventional notation Γ_t^{rs} and Γ_t^{*rs} defined by

$$\nabla_{\partial^r}\partial^s = \sum_t \Gamma_t^{rs}\partial^t; \quad \nabla^*_{\partial^r}\partial^s = \sum_t \Gamma_t^{*rs}\partial^t.$$

The covariant version of the affine connections will be denoted by superscripted Γ and Γ^*

$$\Gamma^{ij,k}(u) = g(\nabla_{\partial^i}\partial^j, \partial^k), \quad \Gamma^{*ij,k}(u) = g(\nabla^*_{\partial^i}\partial^j, \partial^k). \tag{2.15}$$

Under biorthogonal coordinates, the component forms of the metric tensor g satisfy

$$\sum_k g^{ik}(u) g_{kj}(x) = \delta_j^i$$

while the pair of conjugate connections ∇, ∇^* satisfy

$$\Gamma^{*ts,r}(u) = -\sum_{i,j,k} g^{ir}(u) g^{js}(u) g^{kt}(u) \Gamma_{ij,k}(x) \tag{2.16}$$

and

$$\Gamma_r^{*\,ts}(u) = -\sum_j g^{js}(u) \Gamma_{jr}^t(x). \tag{2.17}$$

Next, we discuss the conditions under which biorthogonal coordinates exist on an arbitrary Riemannian manifold. From its definition (2.13), we have

Theorem 2.1. *A Riemannian manifold $\{\mathfrak{M}, g\}$ admits biorthogonal coordinates if and only if the derivative of the metric $\frac{\partial g_{ij}}{\partial x^k}$ is totally symmetric* [1]

$$\frac{\partial g_{ij}(x)}{\partial x^k} = \frac{\partial g_{ik}(x)}{\partial x^j}. \tag{2.18}$$

That (2.18) is satisfied for biorthogonal coordinates is evident by virtue of (2.12) and (2.13). Conversely, given (2.18), there must be n functions $u_i(x), i = 1, 2, \cdots, n$ such that

$$\frac{\partial u_i(x)}{\partial x^j} = g_{ij}(x) = g_{ji}(x) = \frac{\partial u_j(x)}{\partial x^i}.$$

The above identity, in turn, implies that there exist a function Φ such that $u_i = \partial_i \Phi$ and, by positive definiteness of g_{ij}, Φ would have to be a strictly convex function! In this case, the

[1] Note that $\frac{\partial g_{ij}}{\partial x^k} \equiv \partial_k(g(\partial_i, \partial_j)) \neq (\partial_k g)(\partial_i, \partial_j)$, the latter is necessarily totally symmetric whenever there exist a pair of torsion-free connections Γ, Γ^* that are conjugate with respect to g.

x- and u-variables satisfy (1.10), and the pair of convex functions, Φ and its conjugate $\widetilde{\Phi}$, are related to g_{ij} and g^{ij} by

$$g_{ij}(x) = \frac{\partial^2 \Phi(x)}{\partial x^i \partial x^j} \longleftrightarrow g^{ij}(u) = \frac{\partial^2 \widetilde{\Phi}(u)}{\partial u_i \partial u_j}.$$

When (2.18) is satisfied, we can define a pair of torsion-free connections by

$$\gamma_{ij,k}(x) = 0, \quad \gamma^*_{ij,k}(x) = \frac{\partial g_{ij}}{\partial x^k}$$

and show that they are conjugate with respect to g, that is, they satisfy (2.9). This is to say that we selected an affine connection γ such that x is its affine coordinate system. We can now express γ^* in u-coordinates (applying the standard formula for a change of coordinate)

$$\begin{aligned}
\gamma^{*rs,t}(u) &= \sum_{i,j,k} F^{ir}(u)F^{js}(u)F^{kt}(u)\gamma^*_{ij,k}(x) + \frac{\partial F^{ts}(u)}{\partial u_r} = \sum_{i,j,k} g^{ir}(u)g^{js}(u)\frac{\partial x^k}{\partial u_t}\frac{\partial g_{ij}(x)}{\partial x^k} + \frac{\partial g^{ts}(u)}{\partial u_r} \\
&= \sum_{i,j} g^{ir}(u)\left(-\frac{\partial g^{js}(u)}{\partial u_t} g_{ij}(x)\right) + \frac{\partial g^{ts}(u)}{\partial u_r} = -\sum_j \delta^r_j \frac{\partial g^{js}(u)}{\partial u_t} + \frac{\partial g^{ts}(u)}{\partial u_r} = 0.
\end{aligned}$$

This implies that u is an affine coordinate system with respect to γ^*. Therefore, biorthogonal coordinates are affine coordinates for a pair of dually-flat connections.

Let us now consider the parallel volume forms under biorthogonal coordinates. Contracting the indices t with r in (2.17), and invoking (2.7), we obtain

$$\frac{\partial \log \Omega^*(u)}{\partial u_s} + \sum_j \frac{\partial x^j}{\partial u_s}\frac{\partial \log \Omega(x)}{\partial x^j} = \frac{\partial \log \Omega^*(u)}{\partial u_s} + \frac{\partial \log \Omega(x(u))}{\partial u_s} = 0.$$

After integration,

$$\Omega^*(u)\,\Omega(x) = const. \tag{2.19}$$

Note that $\Omega(x) = \Omega(\partial_1, \cdots, \partial_n)$ and $\Omega^*(x) = \Omega^*(\partial_1, \cdots, \partial_n)$, as skew-symmetric multilinear maps, transform to $\Omega(u) = \Omega(\partial^1, \cdots, \partial^n)$ and $\Omega^*(u) = \Omega^*(\partial^1, \cdots, \partial^n)$ via

$$\Omega(x) = \det[F_{ij}(x)]\Omega(u) \longleftrightarrow \Omega^*(x) = \det[F^{ij}(u)]\Omega^*(u),$$

where $\det[F_{ij}(x)] = \det[g_{ij}(x)] = (\det[F^{ij}(u)])^{-1} = (\det[g^{ij}(u)])^{-1}$. Therefore,

$$\Omega(x)\,\Omega^*(x) = (\widehat{\Omega}(x))^2 = \det[g_{ij}(x)], \tag{2.20}$$

$$\Omega(u)\,\Omega^*(u) = (\widehat{\Omega}(u))^2 = \det[g^{ij}(u)]. \tag{2.21}$$

and

$$\Omega(u)\,\Omega^*(x) = const. \tag{2.22}$$

The relations (2.19) and (2.22) indicate that the volume forms of the pair of conjugate connections, when expressed in biorthogonal coordinates respectively, are inversely proportional to each other. The $\Gamma^{(\alpha)}$-parallel volume element $\Omega^{(\alpha)}$ can be shown to be given by

$$\Omega^{(\alpha)} = \Omega^{\frac{1+\alpha}{2}}(\Omega^*)^{\frac{1-\alpha}{2}}$$

in either x- or u-coordinates. Clearly,

$$\Omega^{(\alpha)}(x)\Omega^{(-\alpha)}(x) = \det[g_{ij}(x)] \longleftrightarrow \Omega^{(\alpha)}(u)\Omega^{(-\alpha)}(u) = \det[g^{ij}(u)].$$

2.3 α-Hessian Structure Induced from $\mathcal{D}_\Phi$-Divergence

It is well-known that a statistical structure $\{\mathfrak{M}, g, \Gamma, \Gamma^*\}$ can be induced from a divergence function.

Lemma 2.2 (Eguchi, 1983; 1992). *A divergence function induces a Riemannian metric g and a pair of torsion-free conjugate connections Γ, Γ^* given as*

$$\begin{aligned} g_{ij}(x) &= -\mathcal{D}_{i,j}(x,x); && (2.23)\\ \Gamma_{ij,k}(x) &= -\mathcal{D}_{ij,k}(x,x)\ ; && (2.24)\\ \Gamma^*_{ij,k}(x) &= -\mathcal{D}_{k,ij}(x,x)\ . && (2.25)\end{aligned}$$

It is easily verifiable that $g_{ij}, \Gamma_{ij,k}, \Gamma^*_{ij,k}$ as given above satisfy (2.10). Furthermore, under arbitrary coordinate transform, these quantities behave properly as desired. Applying (2.23)–(2.25) to the divergence function $\mathcal{D}^{(\alpha)}_\Phi(x,y)$ revealed an α-Hessian structure of $\mathfrak{M}$.

Theorem 2.3 (Zhang, 2004). *The manifold $\{\mathfrak{M}, g(x), \Gamma^{(\alpha)}(x), \Gamma^{(-\alpha)}(x)\}$*[2] *associated with $\mathcal{D}^{(\alpha)}_\Phi(x,y)$ is given by*

$$g_{ij}(x) = \Phi_{ij} \tag{2.26}$$

and

$$\Gamma^{(\alpha)}_{ij,k}(x) = \frac{1-\alpha}{2}\,\Phi_{ijk}, \quad \Gamma^{*(\alpha)}_{ij,k}(x) = \frac{1+\alpha}{2}\,\Phi_{ijk}. \tag{2.27}$$

Here, Φ_{ij}, Φ_{ijk} denote, respectively, second and third partial derivatives of $\Phi(x)$

$$\Phi_{ij} = \frac{\partial^2 \Phi(x)}{\partial x^i \partial x^j}, \quad \Phi_{ijk} = \frac{\partial^3 \Phi(x)}{\partial x^i \partial x^j \partial x^k}.$$

Note that such α-Hessian manifold is equipped with an α-independent metric and a family of α-transitively flat connections $\Gamma^{(\alpha)}$ (i.e., $\Gamma^{(\alpha)}$ satisfying (2.11) and $\Gamma^{(\pm 1)}$ are dually flat). From (2.27),

$$\Gamma^{*(\alpha)}_{ij,k} = \Gamma^{(-\alpha)}_{ij,k},$$

with the Levi-Civita connection given as:

$$\widehat{\Gamma}_{ij,k}(x) = \frac{1}{2}\Phi_{ijk}.$$

Straightforward calculation shows that:

Theorem 2.4 (Zhang, 2004; Zhang, 2007; Zhang and Matsuzoe, 2009). *For α-Hessian manifold $\{\mathfrak{M}, g(x), \Gamma^{(\alpha)}(x), \Gamma^{(-\alpha)}(x)\}$,*

(i) the curvature tensor of the α-connection is given by:

$$R^{(\alpha)}_{\mu\nu ij}(x) = \frac{1-\alpha^2}{4}\sum_{l,k}(\Phi_{il\nu}\Phi_{jk\mu} - \Phi_{il\mu}\Phi_{jk\nu})\Psi^{lk} \;=\; R^{*(\alpha)}_{ij\mu\nu}(x),$$

with Ψ^{ij} being the matrix inverse of Φ_{ij};

[2] The functional argument x (or u below) indicates that x-coordinate system (or u-coordinate system, resp) is being used. Under x (u, resp) local coordinates, g and Γ, in component forms, are expressed by lower (upper, resp) indices.

(ii) all α-connections are equiaffine, with the α-parallel volume forms (i.e., the volume forms that are parallel under α-connections) given by

$$\Omega^{(\alpha)}(x) = \det[\Phi_{ij}(x)]^{\frac{1-\alpha}{2}}\,.$$

As α-Hessian manifold admits biorthogonal coordinates, we can also express the volume form associated with $\Gamma^{(\alpha)}$ in u-coordinates as

$$\Omega^{(\alpha)}(u) = \det[\widetilde{\Phi}^{ij}(u)]^{\frac{1+\alpha}{2}}.$$

We remark that if two smooth, strictly convex functions $\Phi(x)$ and $\hat{\Phi}(x)$ are conformally related, i.e., if there exists some positive function $\sigma(x) > 0$ such that $\hat{\Phi}_{ij} = \sigma\Phi_{ij}$, then the curvatures of their respective α-connection can be shown to satisfy

$$\hat{R}^{(\alpha)}_{ij\mu\nu} = \sigma R^{(\alpha)}_{ij\mu\nu}\,. \tag{2.28}$$

Recall that the metric and conjugated connections in the forms (2.26) and (2.27) are induced from (1.8). Using the convex conjugate $\widetilde{\Phi} : \widetilde{V} \to \mathbb{R}$ given by (1.9), we introduce the following family of divergence functions $\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x, y) : V \times V \to \mathbb{R}_{\geq 0}$ defined by

$$\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x, y) \equiv \mathcal{D}^{(\alpha)}_{\widetilde{\Phi}}((\partial\Phi)(x), (\partial\Phi)(y)).$$

Explicitly written, this new family of divergence functions is

$$\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x, y) = \frac{4}{1-\alpha^2}\left(\frac{1-\alpha}{2}\,\widetilde{\Phi}(\partial\Phi(x)) + \frac{1+\alpha}{2}\,\widetilde{\Phi}(\partial\Phi(y))\,\widetilde{\Phi}\left(\frac{1-\alpha}{2}\,\partial\Phi(x) + \frac{1+\alpha}{2}\,\partial\Phi(y)\right)\right). \tag{2.29}$$

Straightforward calculation shows that an α-Hessian structure $\{\mathfrak{M}, g(x), \Gamma(x)^{(-\alpha)}, \Gamma(x)^{(\alpha)}\}$ is induced by $\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x, y)$, where $\Gamma^{(\mp\alpha)}$ are given by (2.27); that is, the pair of α-connections are themselves "conjugate" (in the sense of $\alpha \leftrightarrow -\alpha$) to those induced by $\mathcal{D}^{(\alpha)}_{\Phi}(x, y)$.

If, instead of choosing $x = [x^1, \cdots, x^n]$ as the local coordinates for the manifold $\mathfrak{M}$, we use its biorthogonal counterpart $u = [u_1, \cdots, u_n]$ to index points on $\mathfrak{M}$. Under this u-coordinate system, the divergence function $\mathcal{D}^{(\alpha)}_{\Phi}$ between the same two points on $\mathfrak{M}$ becomes

$$\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u, v) \equiv \mathcal{D}^{(\alpha)}_{\Phi}((\partial\widetilde{\Phi})(u), (\partial\widetilde{\Phi})(v)).$$

Explicitly written,

$$\begin{aligned}\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u, v) = \frac{4}{1-\alpha^2} \quad &\left(\frac{1-\alpha}{2}\,\Phi((\partial\Phi)^{-1}(u)) + \frac{1+\alpha}{2}\,\Phi((\partial\Phi)^{-1}(v))\right. \\ &\left. -\Phi\left(\frac{1-\alpha}{2}\,(\partial\Phi)^{-1}(u) + \frac{1+\alpha}{2}(\partial\Phi)^{-1}(v)\right)\right).\end{aligned}$$

Recall our notation (2.14) and (2.15), we have

Corollary 2.5. *The α-Hessian manifold $\{\mathfrak{M}, g(u), \Gamma^{(\alpha)}(u), \Gamma^{(-\alpha)}(u)\}$ associated with $\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u, v)$ is given by*

$$g^{ij}(u) = \widetilde{\Phi}^{ij}(u), \tag{2.30}$$

$$\Gamma^{(\alpha)ij,k}(u) = \frac{1+\alpha}{2}\,\widetilde{\Phi}^{ijk}, \quad \Gamma^{*(\alpha)ij,k}(u) = \frac{1-\alpha}{2}\,\widetilde{\Phi}_{ijk}, \tag{2.31}$$

Here, $\widetilde{\Phi}^{ij}$, $\widetilde{\Phi}^{ijk}$ denote, respectively, second and third partial derivatives of $\widetilde{\Phi}(u)$

$$\widetilde{\Phi}^{ij}(u) = \frac{\partial^2 \widetilde{\Phi}(u)}{\partial u_i \partial u_j}, \quad \widetilde{\Phi}^{ijk}(u) = \frac{\partial^3 \widetilde{\Phi}(u)}{\partial u_i \partial u_j \partial u_k}.$$

We remark that the same metric (2.30) and the same α-connections (2.31) are induced by $\mathcal{D}^{(-\alpha)}_{\widetilde{\Phi}}(u,v) \equiv \mathcal{D}^{(\alpha)}_{\widetilde{\Phi}}(v,u)$. An application of (2.16) gives rise to the following relations:

$$\begin{aligned}
\Gamma^{(\alpha)mn,l}(u) &= -\sum_{i,j,k} g^{im}(u) g^{jn}(u) g^{kl}(u) \Gamma^{(-\alpha)}_{ij,k}(x), \\
\Gamma^{*(\alpha)mn,l}(u) &= -\sum_{i,j,k} g^{im}(u) g^{jn}(u) g^{kl}(u) \Gamma^{(\alpha)}_{ij,k}(x), \\
R^{(\alpha)klmn}(u) &= \sum_{i,j,\mu,\nu} g^{ik}(u) g^{jl}(u) g^{\mu m}(u) g^{\nu n}(u) R^{(\alpha)}_{ij\mu\nu}(x).
\end{aligned}$$

When $\alpha = \pm 1$, $\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u,v)$, as well as $\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x,y)$ introduced earlier, take the form of Bregman divergence (1.1). In this case, the manifold is dually flat, with curvature tensor $R^{(\pm 1)}_{ij\mu\nu}(x) = R^{(\pm 1)klmn}(u) = 0$.

Zhang (2006) refers to $x \leftrightarrow u$ as the representational duality, and $\alpha \leftrightarrow -\alpha$ as the reference duality. Together, they reveal "reference-representation biduality", as shown by this section.

3 Symplectic Structue Induced by Divergence Functions

A divergence function is given as a bi-variable function on $\mathfrak{M}$ (of dimension n). We now view it as a (single-variable) function on $\mathfrak{M} \times \mathfrak{M}$ (of dimension $2n$) that assumes zero value along the diagonal $\Delta_{\mathfrak{M}} \subset \mathfrak{M} \times \mathfrak{M}$. We will use $\mathcal{D}$ to induce a both a symplectic form and a compatible metric on $\mathfrak{M} \times \mathfrak{M}$, which, when restricted to $\Delta_{\mathfrak{M}}$, is a Lagrange submanifold that carries a statistical structure.

Barndorff-Nielsen and Jupp (1997) associated a symplectic form on $\mathfrak{M} \times \mathfrak{M}$ with any divergence function $\mathcal{D}$ (referred to as "york"), defined as

$$\omega_{\mathcal{D}}(x,y) = \mathcal{D}_{i,j}(x,y) dx^i \wedge dy^j \tag{3.1}$$

(the comma separates the variable being in the first slot versus the second slot for differentiation). For example. Bregman divergence $\mathcal{B}_\Phi$ induces the symplectic form $\sum \Phi_{ij} dx^i \wedge dy^j$.

Fixing a particular y or a particular x in $\mathfrak{M} \times \mathfrak{M}$ results in two n-dimensional submanifolds of $\mathfrak{M} \times \mathfrak{M}$ that will be denoted, respectively, $\mathfrak{M}_x \equiv \mathfrak{M} \times \{y\}$ and $\mathfrak{M}_y \equiv \{x\} \times \mathfrak{M}$. Let us write out the canonical symplectic form ω_x on the cotangent bundle $\mathcal{T}^*\mathfrak{M}_x$ given by

$$\omega_x = dx^i \wedge d\xi^i.$$

Given a divergence function $\mathcal{D}$, we now define a map $L_{\mathcal{D}}$ from $\mathfrak{M} \times \mathfrak{M} \to \mathcal{T}^*\mathfrak{M}_x$, $(x,y) \mapsto (x,\xi)$ given by

$$L_{\mathcal{D}}: \quad (x,y) \mapsto (x, \mathcal{D}_i(x,y) dx^i).$$

It is easily to check that in a neighborhood $\mathcal{U}$ of the diagonal $\Delta_{\mathfrak{M}} \subset \mathfrak{M} \times \mathfrak{M}$, the map $L_{\mathcal{D}}$ is a diffeomorphism since the Jacobian matrix is

$$\begin{pmatrix} \delta_{ij} & \mathcal{D}_{ij} \\ 0 & \mathcal{D}_{i,j} \end{pmatrix}.$$

which is nondegenerate in such neighborhood $\mathcal{U}$ of the diagonal $\Delta_{\mathfrak{M}}$.

We can calculate the pullback of this symplectic form (defined on $\mathcal{T}^*\mathfrak{M}_x$) to $\mathfrak{M} \times \mathfrak{M}$:

$$L_{\mathcal{D}}^* \omega_x = L_{\mathcal{D}}^* (dx^i \wedge d\xi^i) = dx^i \wedge d\mathcal{D}_i(x,y) = dx^i \wedge (\mathcal{D}_{ij}(x,y)dx^j + \mathcal{D}_{i,j}dy^j) = \mathcal{D}_{i,j}(x,y)dx^i \wedge dy^j. \tag{3.2}$$

(Here $\mathcal{D}_{ij}dx^i \wedge dx^j = 0$ since $\mathcal{D}_{ij}(x,y) = \mathcal{D}_{ji}(x,y)$ always holds.) On the other hand, we consider the canonical symplectic form $\omega_y = dy^i \wedge d\eta^i$ on $\mathfrak{M}_y$ and define a map $R_{\mathcal{D}}$ from $\mathfrak{M} \times \mathfrak{M} \to \mathcal{T}^*\mathfrak{M}_y, (x,y) \mapsto (y,\eta)$ given by

$$R_{\mathcal{D}}: \quad (x,y) = (y, \mathcal{D}_{,i}(x,y)dy^i).$$

We then use $R_{\mathcal{D}}$ to pullback ω_y to $\mathfrak{M} \times \mathfrak{M}$, and get an analogous formula. To summarize:

Theorem 3.1. *$\omega_{\mathcal{D}}$ defined on $\mathfrak{M} \times \mathfrak{M}$ is the pullback by the maps $L_{\mathcal{D}}$ and $R_{\mathcal{D}}$ of the canonical symplectic form ω_x defined on $T^*\mathfrak{M}_x$ and ω_y defined on $T^*\mathfrak{M}_y$:*

$$\begin{aligned} L_{\mathcal{D}}^* \omega_x &= \mathcal{D}_{i,j}(x,y)dx^i \wedge dy^j = \omega_{\mathcal{D}}, \\ R_{\mathcal{D}}^* \omega_y &= -\mathcal{D}_{i,j}(x,y)dx^i \wedge dy^j = -\omega_{\mathcal{D}}. \end{aligned}$$

3.1 Almost complex structure, Hermite metric, and Kähler structure

An almost complex structure J on $\mathfrak{M} \times \mathfrak{M}$ is defined by a vector bundle isomorphism (from $\mathfrak{M} \times \mathfrak{M}$ to itself), with the property that $J^2 = -I$. Requiring J to be compatible with $\omega_{\mathcal{D}}$, that is,

$$\omega_{\mathcal{D}}(JX, JY) = \omega_{\mathcal{D}}(X,Y), \ \forall X, Y \in \mathcal{T}_{x,y}\mathfrak{M} \times \mathfrak{M},$$

we may obtain a constraint on the divergence function $\mathcal{D}$. From

$$\omega_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}\left(J\frac{\partial}{\partial x^i}, J\frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial y^i}, -\frac{\partial}{\partial x^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial y^i}\right).$$

we require, and subsequently call a divergence function "proper" if and only if

$$\mathcal{D}_{i,j} = \mathcal{D}_{j,i} \ , \tag{3.3}$$

or explicitly

$$\frac{\partial^2 \mathcal{D}}{\partial x^i \partial y^j} = \frac{\partial^2 \mathcal{D}}{\partial x^j \partial y^i}.$$

Note that this condition is always satisfied on $\Delta_{\mathfrak{M}}$, by the definition of a divergence function $\mathcal{D}$, which has allowed us to define a Riemannian structure on $\Delta_{\mathfrak{M}}$ (Section 2.3). We now require it to be satisfied on $\mathfrak{M} \times \mathfrak{M}$ (at least a neighborhood of $\Delta_{\mathfrak{M}}$.

For proper divergence functions, we can induce a metric $g_{\mathcal{D}}$ on $\mathfrak{M} \times \mathfrak{M}$ — the induced Riemannian (Hermit) metric $g_{\mathcal{D}}$ is defined by

$$g_{\mathcal{D}}(X,Y) = \omega_{\mathcal{D}}(X, JY).$$

It is easily to verify $g_{\mathcal{D}}$ is invariant under the almost complex structure J. The metric components are given by:

$$g_{ij} = g_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, J\frac{\partial}{\partial x^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = -\mathcal{D}_{i,j}\ ,$$

$$g_{,ij} = g_{\mathcal{D}}\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial y^i}, J\frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial y^i}, -\frac{\partial}{\partial x^j}\right) = -\mathcal{D}_{j,i}\ ,$$

$$g_{i,j} = g_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}\left(\frac{\partial}{\partial x^i}, J\frac{\partial}{\partial y^j}\right) = \omega_{\mathcal{D}}(\frac{\partial}{\partial x^i}, -\frac{\partial}{\partial x^j}) = 0\ .$$

So the desired metric on $\mathfrak{M} \times \mathfrak{M}$ is

$$g_{\mathcal{D}} = -\mathcal{D}_{i,j}\Big(dx^i dx^j + dy^i dy^j\Big).$$

So, while a divergence function induces a Riemannian structure on the diagonal manifold $\Delta_{\mathfrak{M}}$ of $\mathfrak{M} \times \mathfrak{M}$, a proper divergence function indeces a Riemannian structure on $\mathfrak{M} \times \mathfrak{M}$.

We now discuss the Kähler structure on the product space $\mathfrak{M} \times \mathfrak{M}$. By definition,

$$\begin{aligned} ds^2 &= g_{\mathcal{D}} - \sqrt{-1}\omega_{\mathcal{D}} = -\mathcal{D}_{i,j}\Big(dx^i \otimes dx^j + dy^i \otimes dy^j\Big) + \sqrt{-1}\mathcal{D}_{i,j}\Big(dx^i \otimes dy^j - dy^i \otimes dx^j\Big) \\ &= -\mathcal{D}_{i,j}\Big(dx^i + \sqrt{-1}dy^i\Big) \otimes \Big(dx^j - \sqrt{-1}dy^j\Big) = -\mathcal{D}_{i,j} dz^i \otimes d\bar{z}^j. \end{aligned}$$

Now introduce complex coordinates $z = x + \sqrt{-1}y$

$$\mathcal{D}(x,y) = \mathcal{D}\left(\frac{z+\bar{z}}{2}, \frac{z-\bar{z}}{2\sqrt{-1}}\right) \equiv \widehat{D}(z, barz)$$

so

$$\frac{\partial^2 \mathcal{D}}{\partial z^i \partial \bar{z}^j} = \frac{1}{4}(\mathcal{D}_{ij} + \mathcal{D}_{,ij}) = \frac{1}{2}\frac{\partial^2 \tilde{\mathcal{D}}}{\partial z^i \partial \bar{z}^j}$$

If $\mathcal{D}$ satisfy

$$\mathcal{D}_{ij} + \mathcal{D}_{,ij} = \kappa \mathcal{D}_{i,j} \tag{3.4}$$

where κ is a constant, then $\mathfrak{M} \times \mathfrak{M}$ admits a Kähler potential (and hence is a Kähler manifold)

$$ds^2 = \frac{\kappa}{2}\frac{\partial^2 \tilde{\mathcal{D}}}{\partial z^i \partial \bar{z}^j} dz^i \otimes d\bar{z}^j.$$

3.2 Kähler structure induced from $\mathcal{D}_\Phi$-divergence

Observe that

$$\Phi\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right) = \Phi\Big((\frac{1-\alpha}{4} + \frac{1+\alpha}{4\sqrt{-1}})z + (\frac{1-\alpha}{4} - \frac{1+\alpha}{4\sqrt{-1}})\bar{z}\Big) \equiv \widehat{\Phi}^{(\alpha)}(z,\bar{z}), \tag{3.5}$$

we have

$$\frac{\partial^2 \widehat{\Phi}^{(\alpha)}}{\partial z^i \partial \bar{z}^j} = \frac{1+\alpha^2}{8}\Phi_{ij}\Big((\frac{1-\alpha}{4} + \frac{1+\alpha}{4\sqrt{-1}})z + (\frac{1-\alpha}{4} - \frac{1+\alpha}{4\sqrt{-1}})\bar{z}\Big)$$

which is symmetric in i, j. Both (3.3) and (3.4) are satisfied. The symplectic form, under the complex coordinates, is given by

$$\omega^{(\alpha)} = \Phi_{ij}\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}\right) dx^i \wedge dy^j = \frac{4\sqrt{-1}}{1+\alpha^2}\frac{\partial^2 \widehat{\Phi}^{(\alpha)}}{\partial z^i \partial \bar{z}^j} dz^i \wedge d\bar{z}^j$$

and the line-element is given by

$$ds^{2(\alpha)} = \frac{8}{1+\alpha^2}\frac{\partial^2 \widehat{\Phi}^{(\alpha)}}{\partial z^i \partial \bar{z}^j} dz^i \otimes d\bar{z}^j.$$

Theorem 3.2. *A smooth and strictly convex function $\Phi : U \subset \mathfrak{M} \to \mathbb{R}$ induces a a family of Kähler structure $(\mathfrak{M}, \omega^{(\alpha)}, g^{(\alpha)})$ defined on $U \times U \subset \mathfrak{M} \times \mathfrak{M}$ with*

1. *the symplectic form $\omega^{(\alpha)}$ is given by*

$$\omega^{(\alpha)} = \Phi^{(\alpha)}_{ij} dx^i \wedge dy^j$$

which is compatiable with the canonical almost complex structure

$$\omega^{(\alpha)}(JX, JY) = \omega^{(\alpha)}(X, Y),$$

where X, Y are vector fields on $U \times U$;

2. *the Riemannian metric $g^{(\alpha)}$, compatible with J and $\omega^{(\alpha)}$ above, given by*

$$g^{(\alpha)} = \Phi^{(\alpha)}_{ij}(dx^i dx^j + dy^i dy^j);$$

3. *the Kähler structure*

$$ds^{2(\alpha)} = \Phi^{(\alpha)}_{ij} dz^i \otimes d\bar{z}^j = \frac{8}{1+\alpha^2}\frac{\partial^2 \widehat{\Phi}^{(\alpha)}}{\partial z^i \partial \bar{z}^j}.$$

with the Kähler potential given by

$$\frac{2}{1+\alpha^2}\widehat{\Phi}^{(\alpha)}(z, \bar{z}).$$

Here, $\Phi^{(\alpha)}_{ij} = \Phi_{ij}\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right)$.

For the diagonal $\Delta_U = \{(x, x) \in \mathfrak{M} \times \mathfrak{M}\}$, a basis of the tangent space $\mathcal{T}_{(x,x)}\Delta_{\mathfrak{M}}$ can be selected as

$$e_i = \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x^i} + \frac{\partial}{\partial y^i}\right).$$

The Riemannian metric on the diagonal, induced from $g^{(\alpha)}$ is

$$\begin{aligned} & g^{(\alpha)}(e_i, e_j)|_{x=y} = \langle g^{(\alpha)}, e_i \otimes e_j \rangle \\ = \; & \langle \Phi^{(\alpha)}_{kl}(dx^k \otimes dx^l + dy^k \otimes dy^l), \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x^i} + \frac{\partial}{\partial y^i}\right) \otimes \frac{1}{\sqrt{2}}\left(\frac{\partial}{\partial x^j} + \frac{\partial}{\partial y^j}\right)\rangle = \Phi^{(\alpha)}_{ij}(x, x) = \Phi_{ij}(x). \end{aligned}$$

Therefore, restricting to the diagonal $\Delta_{\mathfrak{M}}$, $g^{(\alpha)}$ reduces to the Riemannian metric induced by the divergence $D_\Phi^{(\alpha)}$ through the Eguchi method.

We next calculate the Levi-Civita connection $\tilde{\Gamma}$ asociated with $g^{(\alpha)}$. Denote $x^{i'} = y^i$, and that

$$\tilde{\Gamma}_{i'jk'} = \langle \nabla_{\frac{\partial}{\partial x^{i'}}} \frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^{k'}} \rangle = \langle \nabla_{\frac{\partial}{\partial y^i}} \frac{\partial}{\partial x^j}, \frac{\partial}{\partial y^k} \rangle,$$

and so on. The Levi-Civita connection on $\mathfrak{M} \times \mathfrak{M}$ is

$$\tilde{\Gamma}_{ijk} = \frac{1}{2}\Big(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\Big) = \frac{1-\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

$$\tilde{\Gamma}_{ijk'} = \frac{1}{2}\Big(\frac{\partial g_{ik'}}{\partial x^j} + \frac{\partial g_{jk'}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^{k'}}\Big) = -\frac{1+\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

$$\tilde{\Gamma}_{i'jk'} = \tilde{\Gamma}_{ij'k'} = \frac{1}{2}\Big(\frac{\partial g_{ik'}}{\partial x^{j'}} + \frac{\partial g_{j'k'}}{\partial x^i} - \frac{\partial g_{ij'}}{\partial x^{k'}}\Big) = \frac{1-\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

$$\tilde{\Gamma}_{i'jk} = \tilde{\Gamma}_{ij'k} = \frac{1}{2}\Big(\frac{\partial g_{ik}}{\partial x^{j'}} + \frac{\partial g_{j'k}}{\partial x^i} - \frac{\partial g_{ij'}}{\partial x^k}\Big) = \frac{1+\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

$$\tilde{\Gamma}_{i'j'k} = \frac{1}{2}\Big(\frac{\partial g_{i'k}}{\partial x^{j'}} + \frac{\partial g_{j'k}}{\partial x^{i'}} - \frac{\partial g_{i'j'}}{\partial x^k}\Big) = -\frac{1-\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

$$\tilde{\Gamma}_{i'j'k'} = \frac{1}{2}\Big(\frac{\partial g_{i'k'}}{\partial x^{j'}} + \frac{\partial g_{j'k'}}{\partial x^{i'}} - \frac{\partial g_{i'j'}}{\partial x^{k'}}\Big) = \frac{1+\alpha}{4}\Phi_{ijk}^{(\alpha)}.$$

This implies that $\mathfrak{M} \times \mathfrak{M}$ is an α-Hessian manifold.

References

[1] Amari, S. (1982). Differential geometry of curved exponential families – curvatures and information loss. *Annals of Statistics*, 10: 357-385.

[2] Amari, S. (1985). *Differential Geometric Methods in Statistics (Lecture Notes in Statistics, 28)*, Springer-Verlag, New York. Reprinted in 1990.

[3] Amari, S. and Nagaoka, H. (2000). *Method of Information Geometry*, AMS Monograph, Oxford University Press.

[4] O.E.Barndorff-Nielsen and P.E.Jupp, Yorks and Symplectic Structures, Journal of Statistical Planning and Inferrence, 63(1997)133-146.

[5] Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Physics*, **7**, 200-217.

[6] Csiszár, I. (1967). On topical properties of f-divergence. *Studia Mathematicarum Hungarica*, **2**, 329339.

[7] Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Annals of Statistics*, **11**, 793-803.

[8] Eguchi, S. (1992). Geometry of minimum contrast. *Hiroshima Mathematical Journal*, **22**, 631-647.

[9] Lauritzen, S. (1987). Statistical manifolds. In Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, S., and Rao, C.R. (Eds.) *Differential Geometry in Statistical Inference*, IMS Lecture Notes, Vol. 10, Hayward, CA (pp. 163-216).

[10] Nomizu, K. and Sasaki, T. (1994). *Affine Differential Geometry – Geometry of Affine Immersions*, Cambridge University Press.

[11] Rao, C.R. (1987). Differential metrics in probability spaces. In Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, S., and Rao, C.R. (Eds.) *Differential Geometry in Statistical Inference*, IMS Lecture Notes, Vol. 10, Hayward, CA (pp. 217-240).

[12] Shima, H. (2001). *Hessian Geometry*, Shokabo, Tokyo (in Japanese).

[13] Shima, H. and Yagi, K. (1997). Geometry of Hessian manifolds. *Differential Geometry and its Applications*, **7**, 277-290.

[14] Uohashi, K. (2002). On α-conformal equivalence of statistical manifolds. *Journal of Geometry*, **75**: 179-184.

[15] Zhang, J. (2004). Divergence function, duality, and convex analysis. *Neural Computation*, **16**, 159-195.

[16] Zhang, J. (2006). Referential duality and representational duality on statistical manifolds. *Proceedings of the Second International Symposium on Information Geometry and Its Applications*, Tokyo (pp 58-67).

[17] Zhang, J. (2007). A note on curvature of α-connections on a statistical manifold. *Annals of Institute of Statistical Mathematics*, **59**, 161-170.

[18] Zhang, J. and Matsuzuo, H. (2009). Dualistic differential geometry associated with a convex function. In Gao D.Y. and Sherali, H.D. (Eds) Advances in Applied Mathematics and Global Optimization (Dedicated to Gilbert Strang on the Occasion of His 70th Birthday), Advances in Mechanics and Mathematics, Vol. III, Chapter 13, Springer (pp 439-466).