Symplectic and Kähler Structures on Statistical Manifolds Induced from Divergence Functions

Jun Zhang[1,⋆] and Fubo Li[2]

[1] University of Michigan, Ann Arbor, Michigan, USA
[2] Sichuan University, Chengdu, China
junz@umich.edu

Abstract. Divergence functions play a central role in information geometry. Given a manifold $\mathcal{M}$, a divergence function $\mathcal{D}$ is a smooth, non-negative function on the product manifold $\mathcal{M} \times \mathcal{M}$ that achieves its global minimum of zero (with semi-positive definite Hessian) at those points that form its diagonal submanifold $\mathcal{M}_x$. It is well-known (Eguchi, 1982) that the statistical structure on $\mathcal{M}$ (a Riemmanian metric with a pair of conjugate affine connections) can be constructed from the second and third derivatives of $\mathcal{D}$ evaluated at $\mathcal{M}_x$. Here, we investigate Riemannian and symplectic structures on $\mathcal{M} \times \mathcal{M}$ as induced from $\mathcal{D}$. We derive a necessary condition about $\mathcal{D}$ for $\mathcal{M} \times \mathcal{M}$ to admit a complex representation and thus become a Kähler manifold. In particular, Kähler potential is shown to be globally defined for the class of Φ-divergence induced by a strictly convex function Φ (Zhang, 2004). In such case, we recover α-Hessian structure on the diagonal manifold $\mathcal{M}_x$, which is equiaffine and displays the so-called called "reference-representation biduality."

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. Divergence functions induce the statistical structure of a manifold — a statistical structure consists of a Riemannian metric along with a pair of torsion-free affine connections that are conjugate to each other with respect to the metric. When the conjugate connections are Ricci-symmetric, i.e., when the connections are equiaffine, then the manifold admits, in addition, a pair of parallel volumn forms. Those geometric structures on tangent bundles were reviewed in Zhang and Matsuzoe (2009). In this paper, we investigate how to use divergence functions to construct geometric structures of the cotangent bundle, specifically, the symplectic structure of a statistical manifold. For the class of Φ-divergence functions (Zhang, 2004), the statistical manifold admits, in addition, a compatible complex structure, and hence becomes a Kähler manifold.

⋆ Corresponding author.

F. Nielsen and F. Barbaresco (Eds.): GSI 2013, LNCS 8085, pp. 595–603, 2013.

1 Statistical Manifolds Induced by Divergence Functions

1.1 Φ-Divergence Functions

Definition. A divergence function $\mathcal{D} : \mathfrak{M} \times \mathfrak{M} \to \mathbb{R}_{\geq 0}$ on a manifold $\mathfrak{M}$ under a local chart $V \subseteq \mathbb{R}^n$ 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) $\mathcal{D}_i(x, x) = \mathcal{D}_{,j}(x, x) = 0, \forall i, j \in \{1, 2, \cdots, n\}$;
(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 point x and of point y, respectively, $\mathcal{D}_{i,j}(x, y) = \partial_{x^i} \partial_{y^j} \mathcal{D}(x, y)$ the second-order mixed derivative, etc.

Zhang (2004) proposed the construction of a general family of divergece functions based on a convex function Φ. This class of divergence functions, called Φ-divergence here, included many familiar examples, such as Bregman divergence (Bregman, 1967), Kullback-Leibler divergence, f-divergence (Csiszar, 1967), α-divergence (Amari, 1985), U-divergence (Eguchi, 2008), Jenson difference (Rao, 1987), etc.

Let $\Phi : V \subseteq \mathbb{R}^n \to \mathbb{R}, x \mapsto \Phi(x)$ be a strictly convex function. For any two points $x \in \mathfrak{M}$, $y \in \mathfrak{M}$ and any real number $\alpha \in (-1, 1)$, strict convexity of Φ guarantees

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

The inequality sign is reversed when $|\alpha| > 1$) (with equality holding only when $x = y$). Assuming Φ to be sufficiently smooth, a family of functions on $V \times V$, as indexed by $\alpha \in \mathbb{R}$, can be constructed as Φ-divergence functions, denoted $\mathcal{D}_\Phi^{(\alpha)}$ (Zhang, 2004):

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

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 $\lim_{\alpha \to \pm 1}$:

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

where $\mathcal{B}_\Phi$ is the Bregman divergence

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle x - y, \partial\Phi(y) \rangle \tag{2}$$

where $\partial\Phi = [\partial_1\Phi, \cdots, \partial_n\Phi]$ and $\langle \cdot, \cdot \rangle_n$ denotes the canonical pairing of $x = [x^1, \cdots, x^n] \in V$ and $u = [u_1, \cdots, u_n] \in \widetilde{V}$ (dual to V): $\langle x, u \rangle_n = \sum_{i=1}^n x^i u_i$.

1.2 α-Hessian Structure Induced from Φ-Divergence

A statistical manifold $(\mathfrak{M}, g, \Gamma, \Gamma^*)$ is equipped with a one-parameter family of affine connections, the so-called (Amari, 1985) "α-connections $\Gamma^{(\alpha)}$ ($\alpha \in \mathbb{R}$), where $\Gamma^{(\alpha)} = \frac{1+\alpha}{2}\Gamma + \frac{1-\alpha}{2}\Gamma^*$ and $\Gamma^{(0)} = \widehat{\Gamma}$. It is well-known that a statistical structure $\{\mathfrak{M}, g, \Gamma, \Gamma^*\}$ can be induced from a divergence function.

Lemma 1 (Eguchi, 1983). *A divergence function induces a Riemannian metric g and a pair of torsion-free conjugate connections Γ, Γ^* given as*

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

It is easily verifiable that $\Gamma_{ij,k}, \Gamma^*_{ij,k}$ as given above are torsion free[1] and satisfy the conjugacy condition with respect to the induced metric g_{ij}. Hence $\{\mathfrak{M}, g, \Gamma, \Gamma^*\}$ as induced is a "statistical manifold' (Lauritzen, 1987).

Applying the Eguchi construction to $\mathcal{D}_\Phi$, we have

Proposition 1 (Zhang, 2004). *The manifold $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$ associated with $\mathcal{D}^{(\alpha)}_\Phi(x,y)$ is given by*

$$g_{ij}(x) = \Phi_{ij} \tag{3}$$

and

$$\Gamma^{(\alpha)}_{ij,k}(x) = \frac{1-\alpha}{2}\,\Phi_{ijk}, \quad \Gamma^{*(\alpha)}_{ij,k}(x) = \frac{1+\alpha}{2}\,\Phi_{ijk}. \tag{4}$$

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

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

The manifold $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$ is called an "α-Hessian manifold". It is equipped with an α-independent metric and a family of α-transitively flat[2] connections $\Gamma^{(\alpha)}$, i.e., $\Gamma^{(\alpha)}$, with $\Gamma^{*(\alpha)}_{ij,k} = \Gamma^{(-\alpha)}_{ij,k}$ and Levi-Civita connection $\widehat{\Gamma}_{ij,k}(x) = \frac{1}{2}\Phi_{ijk}$. The concept of α-Hessian manifold is a proper generalization of the dually flat "Hessian manifold" (Shima, 2001; Shima and Yagi, 1997) induced from a strict convex function. Straightforward calculation shows that:

Proposition 2 (Zhang, 2004; Zhang, 2007; Zhang and Matsuzoe, 2009). *For α-Hessian manifold $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$,*

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

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

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

[1] Conjugate connections which admit torsion has been recently studied by Calin, Matsuzoe, and Zhang (2009) and Matsuzoe (2010).

[2] Uohashi (2002) called $\Gamma^{(\alpha)}$ α-transitively flatness when Γ, Γ^* are dually flat.

(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

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

α-Hessian manifold manifests the so-called "reference-representation biduality" (Zhang 2004; 2006; Zhang and Matsuzoe, 2009). Its auto-parallel curves have analytic expression. From

$$\frac{d^2 x^i}{ds^2} + \sum_{j,k} \Gamma^{(\alpha)i}_{jk} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0$$

and substituting (4), we obtain

$$\sum_i \Phi_{ki} \frac{d^2 x^i}{ds^2} + \frac{1-\alpha}{2} \sum_{i,l} \Phi_{kij} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0 \longleftrightarrow \frac{d^2}{ds^2} \Phi_k \left(\frac{1-\alpha}{2} x \right) = 0.$$

So the auto-parallel curves of an α-Hessian manifold all have the form

$$\Phi_k \left(\frac{1-\alpha}{2} x \right) = a^k s^{(\alpha)} + b^k$$

where the scalar s is the arc length and $a^k, b^k, k = 1, c \ldots, n$ are constant vectors (determined by a point and the direction along which the auto-parallel curve flows through). For $\alpha = -1$, the auto-parallel curves are given by $u^k = \Phi_k(x) = a^k s + b^k$ are affine coordinates as previously noted.

2 Symplectic Structue Induced by Divergence Functions

A divergence function $\mathcal{D}$ 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 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 $\mathcal{D}$ (called "york" there), defined as (apart from a minus sign that we add here)

$$\omega_{\mathcal{D}}(x, y) = -\mathcal{D}_{i,j}(x, y) dx^i \wedge dy^j \tag{5}$$

(the comma separates the variable being in the first slot versus the second slot for differentiation). For example, Bregman divergence $\mathcal{B}_\Phi$ (given by (2)) 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 $\mathcal{D}$, we 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 of the diagonal $\Delta_{\mathfrak{M}} \subset \mathfrak{M} \times \mathfrak{M}$, the map $L_{\mathcal{D}}$ is a diffeomorphism since the Jacobian matrix of the map

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

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

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

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

(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

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

Using $R_{\mathcal{D}}$ to pullback ω_y to $\mathfrak{M} \times \mathfrak{M}$ yields an analogous formula.

Proposition 3. *The symplectic form $\omega_{\mathcal{D}}$ (as in Eqn. 5) 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 $T^*\mathfrak{M}_y$:*

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

2.1 Almost Complex, Hermite Metric, and Kähler Structures

An almost complex structure J on $\mathfrak{M} \times \mathfrak{M}$ is defined by a vector bundle isomorphism (from $\mathcal{T}\mathfrak{M} \times \mathfrak{M}$ 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

$$\mathcal{D}_{i,j} = \mathcal{D}_{j,i} \ , \tag{6}$$

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 $\Delta_{\mathfrak{M}}$ (Lemma 1). 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:

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

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

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

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

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

In short, 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 induces a Riemannian structure on $\mathfrak{M} \times \mathfrak{M}$ that is compatible with $\omega_{\mathcal{D}}$.

We now discuss the Kähler structure on the product manifold $\mathfrak{M} \times \mathfrak{M}$. 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 $z = x + \sqrt{-1}y$,

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

so

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

If $\mathcal{D}$ satisfies

$$\mathcal{D}_{ij} + \mathcal{D}_{,ij} = \kappa \mathcal{D}_{i,j} \tag{7}$$

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

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

2.2 Kähler Structure Induced from Φ-Divergence

With respect to the Φ-divergence (1), observe that

$$\Phi\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right) = \Phi\Big((\frac{1-\alpha}{4} + \frac{1+\alpha}{4\sqrt{-1}})z + (\frac{1-\alpha}{4} - \frac{1+\alpha}{4\sqrt{-1}})\bar{z}\Big) \equiv \widehat{\Phi}^{(\alpha)}(z,\bar{z}), \tag{8}$$

we have

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

which is symmetric in i, j. Both (6) and (7) 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 1. *A smooth, 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, is given by*

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

3. the Kähler structure

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

with the Kähler potential given by

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

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

For the diagonal manifold $\Delta_{\mathfrak{M}} = \{(x,x) : x \in \mathfrak{M}\}$, a basis of its 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

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

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.

References

1. Amari, S.: Differential Geometric Methods in Statistics. Lecture Notes in Statistics, vol. 28. Springer, New York (1985) (reprinted in 1990)
2. Amari, S., Nagaoka, H.: Method of Information Geometry. AMS Monograph. Oxford University Press (2000)
3. Barndorff-Nielsen, O.E., Jupp, P.E.: Yorks and Symplectic Structures. Journal of Statistical Planning and Inferrence 63, 63133–63146 (1997)
4. Bregman, L.M.: 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 (1967)
5. Calin, O., Matsuzoe, H.: Zhang. J. Generalizations of conjugate connections. In: Trends in Differential Geometry, Complex Analysis and Mathematical Physics: Proceedings of the 9th International Workshop on Complex Structures and Vector Fields, pp. 24–34 (2009)
6. Csiszár, I.: On topical properties of f-divergence. Studia Mathematicarum Hungarica 2, 329–339 (1967)
7. Eguchi, S.: Second order efficiency of minimum contrast estimators in a curved exponential family. Annals of Statistics 11, 793–803 (1983)

8. Eguchi, S.: Information divergence geometry and the application to statistical machine learning. In: Emmert-Streib, F., Dehmer, M. (eds.) Information Theory and Statistical Learning, pp. 309–332. Springer (2008)
9. Lauritzen, S.: Statistical manifolds. In: Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, S., Rao, C.R. (eds.) Differential Geometry in Statistical Inference, Hayward, CA. IMS Lecture Notes, vol. 10, pp. 163–216 (1987)
10. Matsuzoe, M.: Statistical manifolds and affine differential geometry. Advanced Studies in Pure Mathematics 57, 303–321 (2010)
11. Rao, C.R.: Differential metrics in probability spaces. In: Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, S., Rao, C.R. (eds.) Differential Geometry in Statistical Inference, Hayward, CA. IMS Lecture Notes, vol. 10, pp. 217–240 (1987)
12. Shima, H.: Hessian Geometry, Shokabo, Tokyo (2001) (in Japanese)
13. Shima, H., Yagi, K.: Geometry of Hessian manifolds. Differential Geometry and its Applications 7, 277–290 (1997)
14. Uohashi, K.: On α-conformal equivalence of statistical manifolds. Journal of Geometry 75, 179–184 (2002)
15. Zhang, J.: Divergence function, duality, and convex analysis. Neural Computation 16, 159–195 (2004)
16. Zhang, J.: Referential duality and representational duality on statistical manifolds. In: Proceedings of the Second International Symposium on Information Geometry and its Applications, Tokyo, pp. 58–67 (2006)
17. Zhang, J.: A note on curvature of α-connections on a statistical manifold. Annals of Institute of Statistical Mathematics 59, 161–170 (2007)
18. Zhang, J., Matsuzoe, H.: Dualistic differential geometry associated with a convex function. In: Gao, D.Y., 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, ch. 13, vol. III, pp. 439–466. Springer (2009)