Chapter 1
Divergence Functions and Geometric Structures They Induce on a Manifold

Jun Zhang

Abstract Divergence functions play a central role in information geometry. Given a manifold $\mathfrak{M}$, a divergence function $\mathcal{D}$ is a smooth, nonnegative function on the product manifold $\mathfrak{M} \times \mathfrak{M}$ that achieves its global minimum of zero (with semi-positive definite Hessian) at those points that form its diagonal submanifold $\Delta_{\mathfrak{M}} \subset \mathfrak{M} \times \mathfrak{M}$. In this chapter, we review how such divergence functions induce (i) a statistical structure (i.e., a Riemannian metric with a pair of conjugate affine connections) on $\mathfrak{M}$; (ii) a symplectic structure on $\mathfrak{M} \times \mathfrak{M}$ if they are "proper"; (iii) a Kähler structure on $\mathfrak{M} \times \mathfrak{M}$ if they further satisfy a certain condition. It is then shown that the class of $\mathcal{D}_\Phi$-divergence functions [23], as induced by a strictly convex function Φ on $\mathfrak{M}$, satisfies all these requirements and hence makes $\mathfrak{M} \times \mathfrak{M}$ a Kähler manifold (with Kähler potential given by Φ). This provides a larger context for the α-Hessian structure induced by the $\mathcal{D}_\Phi$-divergence on $\mathfrak{M}$, which is shown to be equiaffine admitting α-parallel volume forms and biorthogonal coordinates generated by Φ and its convex conjugate Φ^*. As the α-Hessian structure is dually flat for $\alpha = \pm 1$, the $\mathcal{D}_\Phi$-divergence provides richer geometric structures (compared to Bregman divergence) to the manifold $\mathfrak{M}$ on which it is defined.

1.1 Introduction

Divergence functions (also called "contrast functions", "york") are non-symmetric measurements of proximity. They play a central role in statistical inference, machine learning, optimization, and many other fields. The most familiar examples include Kullback-Leibler divergence, Bregman divergence [4], α-divergence [1], f-divergence [6], etc. Divergence functions are also a key construct of information

J. Zhang (✉)
Department of Psychology and Department of Mathematics, University of Michigan,
Ann Arbor, MI 48109, USA
e-mail: junz@umich.edu

geometry. Just as L_2-distance is associated with Euclidean geometry, Bregman divergence and Kullback-Leibler divergence are associated with a pair of flat structures (where flatness means free of torsion and free of curvature) that are "dual" to each other; this is called Hessian geometry [18, 19] and it is the dualistic extension of the Euclidean geometry. So just as Riemannian geometry extends Euclidean geometry by allowing non-trivial metric structure, Hessian geometry extends Euclidean geometry by allowing non-trivial affine connections that come in pairs. The pairing of connections are with respect to a Riemannian metric g, which is uniquely specified in the case of Hessian geometry; yet the metric-induced Levi-Civita connection has non-zero curvature in general. The apparent inconvenience is offset by the existence of biorthogonal coordinates in any dually flat (i.e., Hessian) structure and a canonical divergence, along with the tools of convex analysis which is powerful in many practical applications.

In a quite general setting, any divergence function induces a Riemannian metric and a pair of torsion-free connections on the manifold where they are defined [8]. This so-called *statistical structure* is at the core of information geometry. Recently, other geometric structures induced by divergence functions are being investigated, including conformal structure [17], symplectic structure [3, 28], and complex structures [28].

The goal of this chapter is to review the relationship between divergence function and various information geometric structures. In Sect. 1.2, we provide background materials of various geometric structures on a manifold. In Sect. 1.3, we show how these structures can be induced from a divergence function. Starting from a general divergence function which always induces a statistical structure, we define the notion of "properness" for it to be a generating function of a symplectic structure. Imposing a further condition leads to complexification of the product manifold where divergence functions are defined. In Sect. 1.4, we show that a quite broad class of divergence functions, $\mathcal{D}_\Phi$-divergence functions [23] as induced by a strictly convex function, satisfies all these requirements and induces a Kähler structure (Riemannian and complex structures simultaneously) on the tangent bundle. Therefore, just as the full-fledged α-Hessian geometry extends the dually-flat Hessian manifold ($\alpha = \pm 1$), $\mathcal{D}_\Phi$-divergence generalizes Bregman divergence in the "nicest" way possible. Section 1.5 closes with a summary of this approach to information geometric structures through divergence functions.

1.2 Background: Structures on Smooth Manifolds

1.2.1 Differentiable Manifold: Metric and Connection Structures on $T\mathfrak{M}$

A differentiable manifold $\mathfrak{M}$ is a space which locally "looks like" a Euclidean space $\mathbb{R}^n$. By "looks like", we mean that for any base (reference) point $x \in \mathfrak{M}$, there

exists a bijective mapping ("coordinate functions") between the neighborhood of x (i.e., a patch of the manifold) and a subset V of $\mathbb{R}^n$. By locally, we mean that various such mappings must be smoothly related to one another (if they are centered at the same reference point) or consistently glued together (if they are centered at different reference points). Globally, they must cover the entire manifold. Below, we assume that a coordinate system is chosen such that each point is indexed by a vector in V, with the origin as the reference point.

A manifold is specified with certain structures. First, there is an inner-product structure associated with tangent spaces of the manifold. This is given by the metric 2-tensor field g which is, when evaluated at each location x, a symmetric bilinear form $g(\cdot, \cdot)$ of tangent vectors $X, Y \in T_x(\mathfrak{M}) \simeq \mathbb{R}^n$ such that $g(X, X)$ is always positive for all non-zero vector $X \in V$. In local "holonomic" coordinates[1] with bases $\partial_i \equiv \partial/\partial x^i, i = 1, \ldots, n$, (i.e., X, Y are expressed as $X = \sum_i X^i \partial_i, Y = \sum_i Y^i \partial_i$), the components of g are denoted as

$$g_{ij}(x) = g(\partial_i, \partial_j). \tag{1.1}$$

Metric tensor allows us to define distance on a manifold as shortest curve (called "geodesic") connecting two points, to measure angle and hence define orthogonality of vectors—projections of vectors to a lower dimensional submanifold become possible once a metric is given. Metric tensor also provides a linear isomorphism of tangent space with cotangent space at any point on the manifold.

Second, there is a structure implementing the notion of "parallelism" of vector fields and curviness of a manifold. This is given by the affine (linear) connection ∇, mapping two vector fields X and Y to a third one denoted by $\nabla_Y X : (X, Y) \mapsto \nabla_Y X$. Intuitively, it represents the "intrinsic" difference of a tangent vector $X(x)$ at point x and another tangent vector $X(x')$ at a nearby point x', which is connected to x in the direction given by the tangent vector $Y(x)$. Here "intrinsic" means that vector comparison across two neighboring points of the manifold is through a process called "parallel transport," whereby vector components are adjusted as the vector moves across points on the base manifold. Under the local coordinate system with bases $\partial_i \equiv \partial/\partial x^i$, components of ∇ can be written out in its "contravariant" form denoted $\Gamma_{ij}^l(x)$

$$\nabla_{\partial_i} \partial_j = \sum_l \Gamma_{ij}^l \partial_l. \tag{1.2}$$

Under coordinate transform $x \mapsto \tilde{x}$, the new coefficients $\widetilde{\Gamma}$ are related to old ones Γ via

$$\widetilde{\Gamma}_{mn}^l(\tilde{x}) = \sum_k \left(\sum_{i,j} \frac{\partial x^i}{\partial \tilde{x}^m} \frac{\partial x^j}{\partial \tilde{x}^n} \Gamma_{ij}^k(x) + \frac{\partial^2 x^k}{\partial \tilde{x}^m \partial \tilde{x}^n} \right) \frac{\partial \tilde{x}^l}{\partial x^k}; \tag{1.3}$$

---

[1] A holonomic coordinate system means that the coordinates have been properly "scaled" in unit-length with respect to each other such that the directional derivatives commute: their Lie bracket $[\partial_i, \partial_j] = \partial_i \partial_j - \partial_j \partial_i = 0$, i.e., the mixed partial derivatives are exchangeable in their order of application.

A curve whose tangent vectors are parallel along the curve is said to be "auto-parallel".

As a primitive on a manifold, affine connections can be characterized in terms of their (i) torsion and (ii) curvature. The torsion T of a connection Γ, which is a tensor itself, is given by the asymmetric part of the connection $T(\partial_i, \partial_j) = \nabla_{\partial_i}\partial_j - \nabla_{\partial_j}\partial_i = \sum_k T_{ij}^k \partial_k$, where T_{ij}^k is its local representation given as

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

The curviness/flatness of a connection Γ is described by the Riemann curvature tensor R, defined as

$$R(\partial_i, \partial_j)\partial_k = (\nabla_{\partial_i}\nabla_{\partial_j} - \nabla_{\partial_j}\nabla_{\partial_i})\partial_k.$$

Writing $R(\partial_i, \partial_j)\partial_k = \sum_l R_{kij}^l \partial_l$ and substituting (1.2), the components of the Riemann curvature tensor are[2]

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

By definition, R_{kij}^l is anti-symmetric when $i \longleftrightarrow j$.

A connection is said to be flat when $R_{kij}^l(x) \equiv 0$ and $T_{ij}^k \equiv 0$. Note that this is a tensorial condition, so that the flatness of a connection ∇ is a coordinate-independent property even though the local expression of the connection (in terms of Γ) is coordinate-dependent. For any flat connection, there exists a local coordinate system under which $\Gamma_{ij}^k(x) \equiv 0$ in a neighborhood; this is the affine coordinate for the given flat connection.

In the above discussions, metric and connections are treated as separate structures on a manifold. When both are defined on the same manifold, then it is convenient to express affine connection Γ in its "covariant" form

$$\Gamma_{ij,k} = g(\nabla_{\partial_i}\partial_j, \partial_k) = \sum_l g_{lk}\Gamma_{ij}^l. \tag{1.4}$$

Though Γ_{ij}^k is the more primitive quantity that does not involve metric, $\Gamma_{ij,k}$ represents the projection of Γ onto the manifold spanned by the bases ∂_k. The covariant form of Riemann curvature is (c.f. footnote 2)

$$R_{lkij} = \sum_m g_{lm}R_{kij}^m.$$

---

[2] This component-wise notation of Riemann curvature tensor followed standard differential geometry textbook, such as [16]. On the other hand, information geometers, such as [2], adopt the notation that $R(\partial_i, \partial_j)\partial_k = \sum_l R_{ijk}^l \partial_l$, with $R_{ijkl} = \sum_l R_{ijk}^m g_{ml}$.

When the connection is torsion free, R_{lkij} is anti-symmetric when $i \longleftrightarrow j$ or when $k \longleftrightarrow l$, and symmetric when $(i, j) \longleftrightarrow (l, k)$. It is related to the Ricci tensor Ric via $\text{Ric}_{kj} = \sum_{i,l} R_{lkij} g^{il}$.

1.2.2 Coupling Between Metric and Connection: Statistical Structure

A fundamental theorem of Riemannian geometry states that given a metric, there is a unique connection (among the class of torsion-free connections) that "preserves" the metric, i.e., the following condition is satisfied

$$\partial_k g(\partial_i, \partial_j) = g(\widehat{\nabla}_{\partial_k} \partial_i, \partial_j) + g(\partial_i, \widehat{\nabla}_{\partial_k} \partial_j). \tag{1.5}$$

Such a connection, denoted as $\widehat{\nabla}$, is called the Levi-Civita connection. Its component forms, called Christoffel symbols, are specified by the components of the metric tensor as ("Christoffel symbols of the second kind")

$$\widehat{\Gamma}_{ij}^k = \sum_l \frac{g^{kl}}{2} \left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right).$$

and ("Christoffel symbols of the first kind")

$$\widehat{\Gamma}_{ij,k} = \frac{1}{2} \left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} \right).$$

The Levi-Civita connection $\widehat{\Gamma}$ is compatible with the metric g, in the sense that it treats tangent vectors of the shortest curves on a manifold as being parallel (equivalently speaking, it treats geodesics as auto-parallel curves).

It turns out that one can define a kind of "compatibility" relation more general than expressed by (1.5), by introducing the notion of "conjugacy" (denoted by $*$) between two connections. A connection ∇^* is said to be *conjugate* (or dual) to ∇ with respect to g if

$$\partial_k g(\partial_i, \partial_j) = g(\nabla_{\partial_k} \partial_i, \partial_j) + g(\partial_i, \nabla^*_{\partial_k} \partial_j). \tag{1.6}$$

Clearly, $(\nabla^*)^* = \nabla$. Moreover, $\widehat{\nabla}$, which satisfies (1.5), is special in the sense that it is self-conjugate $(\widehat{\nabla})^* = \widehat{\nabla}$.

Because metric tensor g provides a one-to-one mapping between points in the tangent space (i.e., vectors) and points in the cotangent space (i.e., co-vectors), (1.6) can also be seen as characterizing how co-vector fields are to be parallel-transported in order to preserve their dual pairing $\langle \cdot, \cdot \rangle$ with vector fields.

Writing out (1.6):

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ki,j} + \Gamma^*_{kj,i}, \tag{1.7}$$

where analogous to (1.2) and (1.4),

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

so that

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

There is an alternative way of imposing "compatibility" condition between a metric g and a connection ∇, through investigating the behavior of how the metric tensor g behaves under ∇. We introduce a 3-tensor field, called "cubic form", as the covariant derivative of g: $C = \nabla g$, or in component forms

$$C(\partial_i, \partial_j, \partial_k) = (\nabla_{\partial_k} g)(\partial_i, \partial_j) = \partial_k g(\partial_i, \partial_j) - g(\nabla_{\partial_k} \partial_i, \partial_j) - g(\partial_i, \nabla_{\partial_k} \partial_j).$$

Writing out the above:

$$C_{ijk} = \frac{\partial g_{ij}}{\partial x_k} - \Gamma_{ki,j} - \Gamma_{kj,i} (= \Gamma^*_{kj,i} - \Gamma_{kj,i}).$$

From its definition, $C_{ijk} = C_{jik}$, that is, symmetric with respective to its first two indices. It can be further shown that:

$$C_{ijk} - C_{ikj} = \sum_l g_{il} (T^l_{jk} - T^{*l}_{jk})$$

where T, T^* are torsions of ∇ and ∇^*, respectively. Therefore, $C_{ijk} = C_{ikj}$, and hence C is totally symmetric in all (pairwise permutation of) indices, when $T^l_{jk} = T^{*l}_{jk}$. So conceptually, requiring C_{ijk} to be totally symmetric imposes a compatibility condition between g and ∇, making them the so-called "Codazzi pair" (see [20]). The Codazzi pairing generalizes the Levi-Civita coupling whose corresponding cubic form C_{ijk} is easily seen to be identically zero. Lauritzen [10] defined a "statistical manifold" $(\mathfrak{M}, g, \nabla)$ to be a manifold $\mathfrak{M}$ equipped with g and ∇ such that (i) ∇ is torsion free; (ii) $\nabla g \equiv C$ is totally symmetric. Equivalently, a manifold is said to have statistical structure when the conjugate connection ∇^* (with respect to g) of a torsion-free connection ∇ is also torsion-free. In this case, $\nabla^* g = -C$, and that the Levi-Civita connection $\hat{\nabla} = (\nabla + \nabla^*)/2$.

Two torsion-free connections Γ and Γ' are said to be *projectively equivalent* if there exists a function τ such that:

$$\Gamma'^k_{ij} = \Gamma^k_{ij} + \delta^k_i (\partial_j \tau) + \delta^k_j (\partial_i \tau),$$

where δ_i^k is the Kronecker delta. When two connections are projectively equivalent, their corresponding auto-parallel curves have identical shape (i.e., considered as unparameterized curves); these so-called "pre-geodesics" differ only by a change of parameterization τ.

Two torsion-free connections Γ and Γ' are said to be *dual-projectively equivalent* if there exists a function τ such that:

$$\Gamma'_{ij,k} = \Gamma_{ij,k} - g_{ij}(\partial_k \tau).$$

When two connections are dual-projectively equivalent, then their conjugate connections (with respect to g) have identical pre-geodesics (identical shape).

Recall that when the two Riemannian metric g, g' are *conformally equivalent*, i.e., there exists a function τ such that

$$g'_{ij} = e^{2\tau} g_{ij},$$

then their respective Levi-Civita connections $\widehat{\Gamma'}$ and $\widehat{\Gamma}$ are related via

$$\widehat{\Gamma'}_{ij,k} = \widehat{\Gamma}_{ij,k} - (\partial_k \tau)g_{ij} + (\partial_j \tau)g_{ik} + (\partial_i \tau)g_{jk}.$$

(This relation is obtained by directly substituting in the expressions of the corresponding Levi-Civita connections.) This motivates the definition of the more general notion of *conformally-projectively equivalent* of two statistical structures $(\mathfrak{M}, g, \Gamma)$ and $(\mathfrak{M}, g', \Gamma')$, through the existence of two functions ψ, ϕ such that:

$$g'_{ij} = e^{\psi + \phi} g_{ij} \tag{1.8}$$

$$\Gamma'_{ij,k} = \Gamma_{ij,k} - (\partial_k \psi)g_{ij} + (\partial_j \phi)g_{ik} + (\partial_i \phi)g_{jk}. \tag{1.9}$$

When $\phi = const$ (or $\psi = const$), then the corresponding connections are projectively (dual-projectively, resp) equivalent.

1.2.3 Equiaffine Structure and Parallel Volume Form

For a restrictive set of connections, called "equiaffine" connections, the manifold $\mathfrak{M}$ may admit, in a unique way, a volume form $\Omega(x)$ that is "parallel" under the given connection. Here, a volume form is a skew-symmetric multilinear map from n linearly independent vectors to a non-zero scalar at any point $x \in \mathfrak{M}$, and "parallel" is in the sense that $\nabla \Omega = 0$, or $(\partial_i \Omega)(\partial_1, \ldots, \partial_n) = 0$ where

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

Applying (1.2), the equiaffine condition becomes

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

$$= \sum_{l=1}^{n} \sum_{k=1}^{n} \Gamma_{il}^k \, \delta_k^l \, \Omega(\partial_1, \ldots, \partial_n) = \Omega(\partial_1, \ldots, \partial_n) \sum_{l=1}^{n} \Gamma_{il}^l$$

or

$$\sum_{l} \Gamma_{il}^l(x) = \frac{\partial \log \Omega(x)}{\partial x^i}. \tag{1.10}$$

Whether or not a connection is equiaffine is related to the so-called Ricci tensor Ric, defined as the contraction of the Riemann curvature tensor R

$$\text{Ric}_{ij}(x) = \sum_{k} R_{ikj}^k(x). \tag{1.11}$$

For a torsion-free connection $\Gamma_{ij}^k = \Gamma_{ji}^k$, we can verify that

$$\text{Ric}_{ij} - \text{Ric}_{ji} = \frac{\partial}{\partial x^i}\left(\sum_{l} \Gamma_{jl}^l(x)\right) - \frac{\partial}{\partial x^j}\left(\sum_{l} \Gamma_{il}^l(x)\right) \tag{1.12}$$

$$= \sum_{k} R_{kij}^k.$$

One immediately sees that the existence of a function Ω satisfying (1.10) is equivalent to the right side of (1.12) to be identically zero.

Making use of (1.10), it is easy to show that the parallel volume form of a Levi-Civita connection $\widehat{\Gamma}$ is given by

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

Making use of (1.7), the parallel volume forms Ω, Ω^* associated with Γ and Γ^* satisfy (apart from a multiplicative constant which must be positive)

$$\Omega(x)\,\Omega^*(x) = (\widehat{\Omega}(x))^2 = \det[g_{ij}(x)]. \tag{1.13}$$

The equiaffine condition can also be expressed using a quantity related to the cubic form C_{ijk}. We may introduce the *Tchebychev form* (also known as the *first Koszul form*), expressed in the local coordinates,

$$T_i = \sum_{j,k} C_{ijk} g^{jk}. \tag{1.14}$$

A tedious calculation shows that

$$\frac{\partial T_i}{\partial x^j} - \frac{\partial T_j}{\partial x^i} = \frac{\partial}{\partial x^j}\left(\sum_l \Gamma_{li}^l\right) - \frac{\partial}{\partial x^i}\left(\sum_l \Gamma_{lj}^l\right),$$

the righthand side of (1.12). Therefore, an equivalent requirement for equiaffine structure is that Tchebychev 1-form T is "closed":

$$\frac{\partial T_i}{\partial x^j} = \frac{\partial T_j}{\partial x^i}. \tag{1.15}$$

This expresses the integrability condition. When Eq. (1.15) is satisfied, there exits a function ϕ such that $T_i = \partial_i \tau$. Furthermore, it can be shown that

$$\tau = -2\log(\Omega/\widehat{\Omega}).$$

Proposition 1 ([13, 25]) *The necessary and sufficient condition for a torsion-free connection ∇ to be equiaffine is for any of the following to hold:*

1. *There exists a ∇-parallel volume element $\Omega : \nabla\Omega = 0$.*
2. *Ricci tensor of ∇ is symmetric: $Ric_{ij} = Ric_{ji}$.*
3. *Curvature tensor $\sum_k R_{kij}^k = 0$.*
4. *The Tchebychev 1-form T is closed, $dT = 0$.*
5. *There exists a function τ, called Tchebychev potential, such that $T_i = \partial_i \tau$.*

It is known that the Ricci tensor of the Levi-Civita connection is always symmetric—this is why Riemannian volume form $\widehat{\Omega}$ always exists.

1.2.4 α-Structure and α-Hessian Structure

On a statistical manifold, one can define a one-parameter family of affine connections $\Gamma^{(\alpha)}$, called "α-connections" ($\alpha \in \mathbb{R}$):

$$\Gamma_{ij}^{(\alpha)k} = \frac{1+\alpha}{2}\Gamma_{ij}^k + \frac{1-\alpha}{2}\Gamma_{ij}^{*k}. \tag{1.16}$$

Obviously, $\Gamma^{(0)} = \widehat{\Gamma}$ is the Levi-Civita connection. Using cubic form, this amounts to $\nabla^{(\alpha)}g = \alpha C$. The α-parallel volume element is given by:

$$\Omega^{(\alpha)} = e^{-\frac{\alpha}{2}\tau}\widehat{\Omega}$$

where τ is the Tchebychev potential. The Riemannian volume element $\widehat{\Omega}$ is only parallel with respect to the Levi-Civita connection $\widehat{\nabla}$ of g, that is, $\widehat{\nabla}\widehat{\Omega} = 0$, but not other α-connections ($\alpha \neq 0$). Rather, $\nabla^{(\alpha)}\Omega^{(\alpha)} = 0$.

It can be further shown that the curvatures R_{lkij}, R^*_{lkij} for the pair of conjugate connections Γ, Γ^* satisfy

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

So, Γ is flat if and only if Γ^* is flat. In this case, the manifold is said to be "dually flat". When Γ, Γ^* are dually flat, then $\Gamma^{(\alpha)}$ is called "α-transitively flat" [21]. In such case, $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$ is called an α-Hessian structure [26]. They are all compatible with a metric g that is induced from a strictly convex (potential) function, see next subsection.

For an α-Hessian manifold, the Tchebychev form (1.14) is given by

$$T_i = \frac{\partial \log(\det[g_{kl}])}{\partial x^i}$$

and its derivative (known as *the second Koszul form*) is

$$\beta_{ij} = \frac{\partial T_i}{\partial x^j} = \frac{\partial^2 \log(\det[g_{kl}])}{\partial x^i \partial x^j}.$$

1.2.5 Biorthogonal Coordinates

A key feature for α-Hessian manifolds is biorthogonal coordinates, as we shall discuss now. They are the "best" coordinates one can have when the Riemannian metric is non-trivial.

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

$$F_{ij}(x) = \frac{\partial u_i}{\partial x^j}, \quad F^{ij}(u) = \frac{\partial x^i}{\partial u_j}, \quad \sum_l F_{il} F^{lj} = \delta^l_i \tag{1.17}$$

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

$$F_{ij}(x) = g_{ij}(x), \tag{1.18}$$

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 (1.1),

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

In such case, denote

$$g^{ij}(u) = g(\partial^i, \partial^j), \tag{1.19}$$

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

$$\nabla_{\partial^r}\partial^s = \sum_t \Gamma_t^{rs}\partial^t;$$

similarly Γ_t^{*rs} is defined via

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

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

$$\Gamma^{ij,k}(u) = g(\nabla_{\partial^i}\partial^j, \partial^k), \quad \Gamma^{*ij,k}(u) = g(\nabla_{\partial^i}^*\partial^j, \partial^k). \tag{1.20}$$

The affine connections in u-coordinates (expressed in superscript) and in x-coordinates (expressed in subscript) are related via

$$\Gamma_t^{rs}(u) = \sum_k \left(\sum_{i,j} \frac{\partial x^r}{\partial u_i} \frac{\partial x^s}{\partial u_j} \Gamma_{ij}^k(x) + \frac{\partial^2 x^k}{\partial u_r \partial u_s} \right) \frac{\partial u_k}{\partial x^t} \tag{1.21}$$

and

$$\Gamma^{rs,t}(u) = \sum_{i,j,k} \frac{\partial x^r}{\partial u_i} \frac{\partial x^s}{\partial u_j} \frac{\partial x^t}{\partial u_k} \Gamma_{ij,k}(x) + \frac{\partial^2 x^t}{\partial u_r \partial u_s}. \tag{1.22}$$

Similarly relations hold between $\Gamma_t^{*rs}(u)$ and $\Gamma_{ij}^{*k}(x)$, and between $\Gamma^{*rs,t}(u)$ and $\Gamma_{ij,k}^*(x)$.

In analogous to (1.7), we have the following identity

$$\frac{\partial^2 x^t}{\partial u_s \partial u_r} = \frac{\partial g^{rt}(u)}{\partial u_s} = \Gamma^{rs,t}(u) + \Gamma^{*ts,r}(u).$$

Therefore, we have

Proposition 2 *Under biorthogonal coordinates, a pair of conjugate connections* Γ, Γ^* *satisfy*

$$\Gamma^{*ts,r}(u) = -\sum_{i,j,k} g^{ir}(u)g^{js}(u)g^{kt}(u)\Gamma_{ij,k}(x) \tag{1.23}$$

and

$$\Gamma_r^{*\,ts}(u) = -\sum_j g^{js}(u)\Gamma_{jr}^t(x). \tag{1.24}$$

Let us now express parallel volume forms $\Omega(x)$, $\Omega(u)$ under biorthogonal coordinates x or u. Contracting the indices t with r in (1.24), and invoking (1.10), we obtain

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

After integration,

$$\Omega^*(u)\,\Omega(x) = const. \tag{1.25}$$

From (1.13) and (1.25),

$$\Omega(u)\,\Omega^*(x) = const. \tag{1.26}$$

The relations (1.25) and (1.26) 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 (in either x and u coordinates)

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

Clearly,

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

1.2.6 Existence of Biorthogonal Coordinates

From its definition (1.18), we can easily show that

Proposition 3 *A Riemannian manifold with metric g_{ij} admits biorthogonal coordinates if and only if $\frac{\partial g_{ij}}{\partial x^k}$ is totally symmetric*

$$\frac{\partial g_{ij}(x)}{\partial x^k} = \frac{\partial g_{ik}(x)}{\partial x^j}. \tag{1.27}$$

That (1.27) is satisfied for biorthogonal coordinates is evident by virtue of (1.17) and (1.18). Conversely, given (1.27), there must be n functions $u_i(x)$, $i = 1, 2, \ldots, n$

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 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 x- and u-variables satisfy (1.37), and the pair of convex functions, Φ and its conjugate $\widetilde{\Phi}$, are related to g_{ij} and g^{ij} by

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

It follows from the above Lemma that a necessary and sufficient condition for a Riemannian manifold to admit biorthogonal coordinates it that its Levi-Civita connection is given by

$$\widehat{\Gamma}_{ij,k}(x) \equiv \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right) = \frac{1}{2}\frac{\partial g_{ij}}{\partial x^k}.$$

From this, the following can be shown:

Proposition 4 *A Riemannian manifold $\{\mathfrak{M}, g\}$ admits a pair of biorthogonal coordinates x and u if and only if there exists a pair of conjugate connections γ and γ^* such that $\gamma_{ij,k}(x) = 0$, $\gamma^{*rs,t}(u) = 0$.*

In other words, biorthogonal coordinates are affine coordinates for the dually-flat pair of connections. In fact, we can now 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 (1.6). This is to say that we select an affine connection γ such that x is its affine coordinate. From (1.22), when γ^* is expressed in u-coordinates,

$$\gamma^{*rs,t}(u) = \sum_{i,j,k} g^{ir}(u)g^{js}(u)\frac{\partial x^k}{\partial u_t}\frac{\partial g_{ij}(x)}{\partial x^k} + \frac{\partial g^{ts}(u)}{\partial u_r}$$

$$= \sum_{i,j} g^{ir}(u)\left(-\frac{\partial g^{js}(u)}{\partial u_t}g_{ij}(x)\right) + \frac{\partial g^{ts}(u)}{\partial u_r}$$

$$= -\sum_j \delta^r_j \frac{\partial g^{js}(u)}{\partial u_t} + \frac{\partial g^{ts}(u)}{\partial u_r} = 0.$$

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.

1.2.7 Symplectic, Complex, and Kähler Structures

Symplectic structure on a manifold refers to the existence of a closed, non-degenerate 2-tensor, i.e., a skew-symmetric bilinear map $\omega: W \times W \to \mathbb{R}$, with $\omega(X, Y) = -\omega(Y, X)$ for all $X, Y \in W \subseteq \mathbb{R}^{2n}$. For ω to be well-defined, the vector space W is required to be orientable and even-dimensional. In this case, there exists a base $\{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ of W, $\dim(W) = 2n$ such that $\omega(e_i, e_j) = 0, \omega(f_i, f_j) = 0, \omega(e_i, f_j) = \delta_{ij}$ for all indices i, j taking values in $1, \ldots, n$.

Symplectic structure is closely related to inner-product structure (the existence of a positive-definite symmetric bilinear map $G: W \times W \to \mathbb{R}$) and complex structure (linear mapping $J: W \to W$ such that $J^2 = -Id$) on an even-dimensional vector space W. The complex structure J on W is said to be compatible with a symplectic structure ω if $\omega(JX, JY) = \omega(X, Y)$ (symplectomorphism condition) and $\omega(X, JY) > 0$ (taming condition) for any $X, Y \in W$. With ω, J given, $G(X, Y) \equiv \omega(X, JY)$ can be shown to be symmetric and positive-definite, and hence an inner-product on W.

The cotangent bundle $\mathcal{T}^*\mathfrak{M}$ of any manifold $\mathfrak{M}$ admits a canonical symplectic form written as

$$\omega = \sum_{i=1}^{n} dx^i \wedge dp_i,$$

where $(x^1, \ldots, x^n, p_1, \ldots, p_n)$ are coordinates of $\mathcal{T}^*\mathfrak{M}$. That ω is closed can be shown by the existence of the tautological (or Liouville) 1-form

$$\alpha = \sum_{i=1}^{n} p_i dx^i$$

(which can be checked to be coordinate-independent on $\mathcal{T}^*\mathfrak{M}$) and then verifying $\omega = -d\alpha$. Hence, ω is also coordinate-independent. Denote $\partial_i = \partial/\partial x^i$, $\widetilde{\partial}_i = \partial/\partial p_i$ as the base of the tangent bundle $\mathcal{T}\mathfrak{M}$, then

$$\omega(\partial_i, \partial_j) = \omega(\widetilde{\partial}_i, \widetilde{\partial}_j) = 0; \quad \omega(\partial_i, \widetilde{\partial}_j) = -\omega(\widetilde{\partial}_j, \partial_i) = \omega_{ij}. \qquad (1.28)$$

That is, when viewed as 2×2 blocks of $n \times n$ matrix, ω vanishes on diagonal blocks and has non-zero entries ω_{ij} and $-\omega_{ij}$ only on off-diagonal blocks.

The aforementioned linear map J of the tangent space $T_x\mathfrak{M} \simeq W$ at any point $x \in \mathfrak{M}$

$$J : J\partial_i = \widetilde{\partial}_j, J\widetilde{\partial}_j = -\partial_i,$$

gives rise to an "almost complex structure" on $T_x\mathfrak{M}$. For $\mathcal{T}\mathfrak{M}$ to be complex, that is, admitting complex coordinates, an integrable condition needs to be imposed for the J-maps at various base points x of $\mathfrak{M}$, and hence at various tangent spaces $T_x\mathfrak{M}$, to be "compatible" with one another. The condition is that the so-called Nijenhuis tensor N

$$N(X, Y) = [JX, JY] - J[X, JY] - J[JX, Y] - [X, Y]$$

must vanish for arbitrary tangent vector fields X, Y.

The Riemannian metric tensor G on $T\mathfrak{M}$ compatible with ω has the form

$$G_{ij'} \equiv G(\partial_i, \tilde{\partial}_j) = 0;$$
$$G_{i'j} \equiv G(\tilde{\partial}_i, \partial_j) = 0;$$
$$G_{ij} \equiv G(\partial_i, \partial_j) = g_{ij}$$
$$G_{i'j'} \equiv G(\tilde{\partial}_i, \tilde{\partial}_j) = g_{ij}.$$

where $i' = n + i$, $j' = n + j$ and i, j takes values in $1, \ldots, n$. When viewed as 2×2 blocks of $n \times n$ matrix, G vanishes on the off-diagonal blocks and has non-zero entries g_{ij} only on the two diagonal blocks. Such a metric on $T\mathfrak{M}$ is in the form of Sasaki metric, which can also result from an appropriate "lift" of the Riemannian metric on $\mathfrak{M}$ into $T\mathfrak{M}$, via an affine connection on $T\mathfrak{M}$ which induces a splitting of $TT\mathfrak{M}$, the tangent bundle of $T\mathfrak{M}$ as the base manifold. We omit the technical details here, but refer interested readers to Yano and Ishihara [22] and, in the context of statistical manifold, to Matsuzoe and Inoguchi [12].

It is a basic conclusion in symplectic geometry that for any symplectic form, there exists a compatible almost complex structure J. Along with the Riemannian metric, the three structures (G, ω, J) are said to form a *compatible triple* if any two gives rise to the third one. When a manifold has a compatible triple (G, ω, J) in which J is integrable, it is called a Kähler manifold. On a Kähler manifold, using complex coordinates, the metric $\tilde{G}$ associated with the complex line-element

$$ds^2 = \tilde{G}_{ij} dz^i d\bar{z}^j,$$

is given by

$$\tilde{G}_{ij}(z, \bar{z}) = \frac{\partial^2 \Phi}{\partial z^i \partial \bar{z}^j}.$$

Here the real-valued function Φ (of complex variables) is called the "Kähler potential".

It is known that the tangent bundle $T\mathfrak{M}$ of a manifold $\mathfrak{M}$ with a flat connection on it admits a complex structure [7]. As [18] pointed out, Hessian manifold can be seen as the "real" Kähler manifold.

Proposition 5 ([7]) $(\mathfrak{M}, g, \nabla)$ *is a Hessian manifold if and only if* $(T\mathfrak{M}, J, G)$ *is a Kähler manifold, where G is the Sasaki lift of g.*

1.3 Divergence Functions and Induced Structures

1.3.1 Statistical Structure Induced on $T\mathfrak{M}$

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 defined as 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) $\mathcal{D}_i(x, x) = \mathcal{D}_{,j}(x, x) = 0, \forall i, j \in \{1, 2, \ldots, n\}$;
(iii) $-\mathcal{D}_{i,j}(x, x)$ is positive definite.

Here $\mathcal{D}_i(x, y) = \partial_{x^i}\mathcal{D}(x, y)$, $\mathcal{D}_{,i}(x, y) = \partial_{y^i}\mathcal{D}(x, y)$ 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.

On a manifold, divergence functions act as "pseudo-distance" functions that are non-negative but need not be symmetric. That dualistic Riemannian manifold structure (i.e., statistical structure) can be induced from a divergence function was first demonstrated by S. Eguchi.

Proposition 6 ([8, 9]) *A divergence function $\mathcal{D}$ induces a Riemannian metric g and a pair of torsion-free conjugate connections Γ, Γ^* given as*

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

It is easily verifiable that $\Gamma_{ij,k}$, $\Gamma^*_{ij,k}$ as given above are torsion-free[3] 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' [10].

A natural question is whether/how the statistical structures induced from different divergence functions are related. The following is known:

Proposition 7 ([14]) *Let $\mathcal{D}$ be a divergence function and ψ, ϕ be two arbitrary functions. If $\mathcal{D}'(x, y) = e^{\psi(x)+\phi(y)}\mathcal{D}(x, y)$, then $\mathcal{D}'(x, y)$ is also a divergence function, and the induced $(\mathfrak{M}, g', \Gamma')$ and $(\mathfrak{M}, g, \Gamma)$ induced from $\mathcal{D}(x, y)$ are conformally-projectively equivalent. In particular, when $\phi(x) = \text{const}$, then Γ' and Γ are projectively equivalent; when $\psi(y) = \text{const}$, then Γ' and Γ are dual-projectively equivalent.*

---

[3] Conjugate connections which admit torsion has been recently studied by Calin et al. [5] and Matsuzoe [15].

1.3.2 Symplectic Structure Induced on $\mathfrak{M} \times \mathfrak{M}$

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 $\Delta_{\mathfrak{M}} \subset \mathfrak{M} \times \mathfrak{M}$. In this subsection, we investigate the condition under which a divergence function can serve as a "generating function" of a symplectic structure on $\mathfrak{M} \times \mathfrak{M}$. A compatible metric on $\mathfrak{M} \times \mathfrak{M}$ will also be derived.

First, we fix a particular y or a particular x in $\mathfrak{M} \times \mathfrak{M}$—this results in two n-dimensional submanifolds of $\mathfrak{M} \times \mathfrak{M}$ that will be denoted, respectively, $\mathfrak{M}_x \simeq \mathfrak{M}$ (with y point fixed) and $\mathfrak{M}_y \simeq \mathfrak{M}$ (with x point fixed). Let us write out the canonical symplectic form ω_x on the cotangent bundle $T^*\mathfrak{M}_x$ given by

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

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

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

(Recall that the comma separates the variable being in the first slot versus the second slot for differentiation.) It is easy 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 non-degenerate in such a neighborhood of the diagonal $\Delta_{\mathfrak{M}}$.

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

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

(Here $\mathcal{D}_{ij}dx^i \wedge dx^j = 0$ since $\mathcal{D}_{ij}(x, y) = \mathcal{D}_{ji}(x, y)$ always holds.)

Similarly, 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 $\mathfrak{M} \times \mathfrak{M} \to T^*\mathfrak{M}_y$, $(x, y) \mapsto (y, \eta)$ given by

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

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

$$R_{\mathcal{D}}^* \omega_y = -\mathcal{D}_{i,j}(x, y)dx^i \wedge dy^j.$$

Therefore, based on canonical symplectic forms on $\mathcal{T}^*\mathfrak{M}_x$ and $\mathcal{T}^*\mathfrak{M}_y$, we obtained the same symplectic form on $\mathfrak{M} \times \mathfrak{M}$

$$\omega_{\mathcal{D}}(x, y) = -\mathcal{D}_{i,j}(x, y)dx^i \wedge dy^j. \tag{1.29}$$

Proposition 8 *A divergence function $\mathcal{D}$ induces a symplectic form $\omega_{\mathcal{D}}$ (1.29) on $\mathfrak{M} \times \mathfrak{M}$ which is the pullback of the canonical symplectic forms ω_x and ω_y by the maps $L_{\mathcal{D}}$ and $R_{\mathcal{D}}$*

$$L_{\mathcal{D}}^* \omega_y = \mathcal{D}_{i,j}(x, y)dx^i \wedge dy^j = -R_{\mathcal{D}}^* \omega_x \tag{1.30}$$

It was Barndorff-Nielsen and Jupp [3] who first proposed (1.29) as an induced symplectic form on $\mathfrak{M} \times \mathfrak{M}$, apart from a minus sign; the divergence function $\mathcal{D}$ was called a "york". As an example, Bregman divergence $\mathcal{B}_{\Phi}$ (given by (1.33) below) induces the symplectic form $\sum \Phi_{ij}dx^i \wedge dy^j$.

1.3.3 Almost Complex Structure and Hermite Metric on $\mathfrak{M} \times \mathfrak{M}$

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 $J^2 = -Id$. 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

$$\mathcal{D}_{i,j} = \mathcal{D}_{j,i}, \tag{1.31}$$

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 $\Delta_{\mathfrak{M}}$ (Proposition 6). We now require it to be satisfied on $\mathfrak{M} \times \mathfrak{M}$ (at least a neighborhood of $\Delta_{\mathfrak{M}}$).

For divergence functions satisfying (1.31), we can consider inducing 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 easy to verify $G_{\mathcal{D}}$ is invariant under the almost complex structure J. The metric components are given by:

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

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

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

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

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

$$G_{\mathcal{D}} = -\mathcal{D}_{i,j}\left(dx^i dx^j + dy^i dy^j\right).$$

So for $G_{\mathcal{D}}$ to be a Riemannian metric, we require $-\mathcal{D}_{i,j}$ to be positive-definite.

We call a divergence function $\mathcal{D}$ *proper* if and only if $-\mathcal{D}_{i,j}$ is symmetric and positive-definite on $\mathfrak{M} \times \mathfrak{M}$. Just as any divergence function induces a Riemannian structure on the diagonal manifold $\Delta_{\mathfrak{M}}$ of $\mathfrak{M} \times \mathfrak{M}$, any proper divergence function induces a Riemannian structure on $\mathfrak{M} \times \mathfrak{M}$ that is compatible with the symplectic structure $\omega_{\mathcal{D}}$ on it.

1.3.4 Complexification and Kähler Structure on $\mathfrak{M} \times \mathfrak{M}$

We now discuss possible existence of a Kähler structure on the product manifold $\mathfrak{M} \times \mathfrak{M}$. By definition,

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

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

$$\mathcal{D}_{ij} + \mathcal{D}_{,ij} = \kappa \mathcal{D}_{i,j} \tag{1.32}$$

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

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

1.3.5 Canonical Divergence for Hessian Manifold

On dually flat (i.e., Hessian) manifold, there is a canonical divergence as shown below. Recall that the Hessian metric

$$g_{ij}(x) = \frac{\partial^2 \Phi(x)}{\partial x^i \partial x^j}$$

and the dual connections

$$\Gamma_{ij,k}(x) = 0, \quad \Gamma_{ij,k}^*(x) = \frac{\partial^3 \Phi(x)}{\partial x^i \partial x^j \partial x^k}$$

are induced from a convex potential function Φ. In the (biorthogonal) u-coordinates, these geometric quantities can be expressed as

$$g^{ij}(u) = \frac{\partial^2 \widetilde{\Phi}(u)}{\partial u_i \partial u_j}, \quad \Gamma^{*\,ij,k}(u) = 0, \quad \Gamma^{ij,k}(u) = \frac{\partial^3 \widetilde{\Phi}(u)}{\partial u_i \partial u_j \partial u_k},$$

where $\widetilde{\Phi}$ is the convex conjugate of Φ.

Integrating the Hessian structure reveals the so-called Bregman divergence $\mathcal{B}_\Phi(x, y)$ [4] as the generating function:

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle x - y, \partial\Phi(y)\rangle \tag{1.33}$$

where $\partial\Phi = [\partial_1\Phi, \ldots, \partial_n\Phi]$ with $\partial_i \equiv \partial/\partial x^i$ denotes the gradient valued in the co-vector space $\widetilde{\mathbb{R}}^n$, and $\langle \cdot, \cdot \rangle_n$ denotes the canonical pairing of a point/vector $x = [x^1, \ldots, x^n] \in \mathbb{R}^n$ and a point/co-vector $u = [u_1, \ldots, u_n] \in \widetilde{\mathbb{R}}_n$ (dual to $\mathbb{R}^n$):

$$\langle x, u \rangle_n = \sum_{i=1}^n x^i u_i. \tag{1.34}$$

(Where there is no danger of confusion, the subscript n in $\langle \cdot, \cdot \rangle_n$ is often omitted.) A basic fact in convex analysis is that the necessary and sufficient condition for a smooth function Φ to be strictly convex is

$$\mathcal{B}_\Phi(x, y) > 0 \tag{1.35}$$

for $x \neq y$.

Recall that, when Φ is convex, its convex conjugate $\widetilde{\Phi}: \widetilde{V} \subseteq \mathbb{R}_n \to \mathbb{R}$ is defined through the Legendre transform:

$$\widetilde{\Phi}(u) = \langle (\partial \Phi)^{-1}(u), u \rangle - \Phi((\partial \Phi)^{-1}(u)), \tag{1.36}$$

with $\widetilde{\widetilde{\Phi}} = \Phi$ and $(\partial \Phi) = (\partial \widetilde{\Phi})^{-1}$. The function $\widetilde{\Phi}$ is also convex, and through which (1.35) precisely expresses the Fenchel inequality

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

for any $x \in V$, $u \in \widetilde{V}$, with equality holding if and only if

$$u = (\partial \Phi)(x) = (\partial \widetilde{\Phi})^{-1}(x) \longleftrightarrow x = (\partial \widetilde{\Phi})(u) = (\partial \Phi)^{-1}(u), \tag{1.37}$$

or, in component form,

$$u_i = \frac{\partial \Phi}{\partial x^i} \longleftrightarrow x^i = \frac{\partial \widetilde{\Phi}}{\partial u_i}. \tag{1.38}$$

With the aid of conjugate variables, we can introduce the "canonical divergence" $\mathcal{A}_\Phi: V \times \widetilde{V} \to \mathbb{R}_+$ (and $\mathcal{A}_{\widetilde{\Phi}}: \widetilde{V} \times V \to \mathbb{R}_+$)

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

They are related to the Bregman divergence (1.33) via

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

1.4 $\mathcal{D}_\Phi$-Divergence and Its Induced Structures

In this section, we study a particular parametric family of divergence functions, called $\mathcal{D}_\Phi$, induced by a strictly convex function Φ, with α as the parameter. This family was first introduced by Zhang [23], who showed that it included many familiar families (see also [27]). The resulting geometric structures will be studied below.

1.4.1 $\mathcal{D}_\Phi$-Divergence Functions

Recall that, by definition, a strictly convex function $\Phi: V \subseteq \mathbb{R}^n \to \mathbb{R}, x \mapsto \Phi(x)$ satisfies

$$\frac{1-\alpha}{2} \, \Phi(x) + \frac{1+\alpha}{2} \, \Phi(y) - \Phi\left(\frac{1-\alpha}{2} x + \frac{1+\alpha}{2} y\right) > 0 \qquad (1.39)$$

for all $x \neq y$ for any $|\alpha| < 1$ (the inequality sign is reversed when $|\alpha| > 1$). Assume Φ to be sufficiently smooth (differentiable up to fourth order).

Zhang [23] introduced the following family of function on $V \times V$ as indexed by $\alpha \in \mathbb{R}$

$$\mathcal{D}_\Phi^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2} \left(\frac{1-\alpha}{2} \, \Phi(x) + \frac{1+\alpha}{2} \, \Phi(y) - \Phi\left(\frac{1-\alpha}{2} x + \frac{1+\alpha}{2} y\right)\right).$$
$$(1.40)$$

From its construction, $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ is non-negative for $|\alpha| < 1$ due to Eq. (1.39), and for $|\alpha| = 1$ due to Eq. (1.35). For $|\alpha| > 1$, assuming $(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y) \in V$, the non-negativity of $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ can also be proven due to the inequality (1.39) reversing its sign. Furthermore, $\mathcal{D}_\Phi^{(\pm 1)}(x, y)$ is defined by taking $\lim_{\alpha \to \pm 1}$:

$$\mathcal{D}_\Phi^{(1)}(x, y) \;=\; \mathcal{D}_\Phi^{(-1)}(y, x) \;=\; B_\Phi(x, y),$$
$$\mathcal{D}_\Phi^{(-1)}(x, y) \;=\; \mathcal{D}_\Phi^{(1)}(y, x) \;=\; B_\Phi(y, x).$$

Note that $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ satisfies the relation (called "referential duality" in [24])

$$\mathcal{D}_\Phi^{(\alpha)}(x, y) = \mathcal{D}_\Phi^{(-\alpha)}(y, x),$$

that is, exchanging the asymmetric status of the two points (in the directed distance) amounts to $\alpha \leftrightarrow -\alpha$.

1.4.2 Induced α-Hessian Structure on $T\mathfrak{M}$

We start by reviewing a main result from [23] linking the divergence function $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ defined in (1.40) and the α-Hessian structure.

Proposition 9 ([23]) *The manifold* $\{\mathcal{M}, g(x), \Gamma^{(\alpha)}(x), \Gamma^{(-\alpha)}(x)\}$[4] *associated with* $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ *is given by*

$$g_{ij}(x) = \Phi_{ij} \qquad (1.41)$$

---

[4] The functional argument of x (or u-below) indicates that x-coordinate system (or u-coordinate system) is being used. Recall from Sect. 1.2.5 that under x (u, resp) local coordinates, g and Γ, in component forms, are expressed by lower (upper, resp) indices.

and

$$\Gamma_{ij,k}^{(\alpha)}(x) = \frac{1-\alpha}{2}\,\Phi_{ijk}, \quad \Gamma_{ij,k}^{*(\alpha)}(x) = \frac{1+\alpha}{2}\,\Phi_{ijk}. \tag{1.42}$$

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

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

Recall that an 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 (1.16) and $\Gamma^{(\pm 1)}$ are dually flat). From (1.42),

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

with the Levi-Civita connection given as:

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

Straightforward calculation shows that:

Proposition 10 ([26]) *For α-Hessian manifold $\{\mathfrak{M}, g(x), \Gamma^{(\alpha)}(x), \Gamma^{(-\alpha)}(x)\}$,*

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

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

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

(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*

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

It is worth pointing out that while $\mathcal{D}_\Phi$-divergence induces the α-Hessian structure, it is not unique, as the same structure can arise from the following divergence function, which is a mixture of Bregman divergences in conjugate forms:

$$\frac{1-\alpha}{2}\mathcal{B}_\Phi(x,y) + \frac{1+\alpha}{2}\mathcal{B}_\Phi(y,x).$$

1.4.3 The Family of α-Geodesics

The family of auto-parallel curves on α-Hessian manifold have analytic expression. From

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

and substituting (1.42), we obtain

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

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 $a^k, b^k, k = 1, 2\ldots, n$ 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.

1.4.3.1 Related Divergences and Geometries

Note that the metric and conjugated connections in the forms (1.41) and (1.42) are induced from (1.40). Using the convex conjugate $\widetilde{\Phi}\colon \widetilde{V} \to \mathbb{R}$ given by (1.36), we introduce the following family of divergence functions $\widetilde{\mathcal{D}}_{\widetilde{\Phi}}^{(\alpha)}(x,y)$ defined by

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

Explicitly written, this new family of divergence functions is

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

Straightforward calculation shows that $\widetilde{\mathcal{D}}_{\widetilde{\Phi}}^{(\alpha)}(x,y)$ induces the α-Hessian structure $\{\mathcal{M}, g, \Gamma^{(-\alpha)}, \Gamma^{(\alpha)}\}$ where $\Gamma^{(\mp\alpha)}$ are given by (1.42); that is, the pair of α-connections are themselves "conjugate" (in the sense of $\alpha \leftrightarrow -\alpha$) to those induced by $\mathcal{D}_{\Phi}^{(\alpha)}(x,y)$.

If, instead of choosing $x = [x^1, \ldots, x^n]$ as the local coordinates for the manifold $\mathfrak{M}$, we use its biorthogonal counterpart $u = [u_1, \ldots, u_n]$ related to x via (1.38) to index points on $\mathcal{M}$. Under this u-coordinate system, the divergence function $\mathcal{D}_\Phi^{(\alpha)}$ between the same two points on $\mathcal{M}$ becomes

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

Explicitly written,

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

Proposition 11 ([23]) *The α-Hessian manifold $\{\mathfrak{M}, g(u), \Gamma^{(\alpha)}(u), \Gamma^{(-\alpha)}(u)\}$ associated with $\widetilde{\mathcal{D}}_\Phi^{(\alpha)}(u, v)$ is given by*

$$g^{ij}(u) = \widetilde{\Phi}^{ij}(u), \tag{1.43}$$

$$\Gamma^{(\alpha)ij,k}(u) = \frac{1 + \alpha}{2} \widetilde{\Phi}^{ijk}, \quad \Gamma^{*(\alpha)ij,k}(u) = \frac{1 - \alpha}{2} \widetilde{\Phi}_{ijk}, \tag{1.44}$$

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 (1.43) and the same α-connections (1.44) are induced by $\mathcal{D}_{\widetilde{\Phi}}^{(-\alpha)}(u, v) \equiv \mathcal{D}_\Phi^{(\alpha)}(v, u)$—this follows as a simple application of Eguchi relation.

An application of (1.23) gives rise to the following relations:

$$\Gamma^{(\alpha)mn,l}(u) = -\sum_{i,j,k} g^{im}(u) g^{jn}(u) g^{kl}(u) \Gamma_{ij,k}^{(-\alpha)}(x),$$

$$\Gamma^{*(\alpha)mn,l}(u) = -\sum_{i,j,k} g^{im}(u) g^{jn}(u) g^{kl}(u) \Gamma_{ij,k}^{(\alpha)}(x),$$

$$R^{(\alpha)klmn}(u) = \sum_{i,j,\mu,\nu} g^{ik}(u) g^{jl}(u) g^{\mu m}(u) g^{\nu n}(u) R_{ij\mu\nu}^{(\alpha)}(x).$$

The volume form associated with $\Gamma^{(\alpha)}$ is

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

Table 1.1 Various divergence functions under biorthogonal coordinates x or u and their induced geometries

Divergence function	Induced geometry
$\mathcal{D}_{\Phi}^{(\alpha)}(x, y)$	$\left\{ \Phi_{ij}(x), \Gamma(x)^{(\alpha)}, \Gamma(x)^{(-\alpha)} \right\}$
$\mathcal{D}_{\Phi}^{(\alpha)}((\partial\Phi)(x), (\partial\Phi)(y))$	$\left\{ \Phi_{ij}(x), \Gamma(x)^{(-\alpha)}, \Gamma(x)^{(\alpha)} \right\}$
$\mathcal{D}_{\widetilde{\Phi}}^{(\alpha)}(u, v)$	$\left\{ \widetilde{\Phi}^{ij}(u), \Gamma(u)^{(-\alpha)}, \Gamma(u)^{(\alpha)} \right\}$
$\mathcal{D}_{\widetilde{\Phi}}^{(\alpha)}((\partial\widetilde{\Phi})(x), (\partial\widetilde{\Phi})(y))$	$\left\{ \widetilde{\Phi}^{ij}(u), \Gamma(u)^{(\alpha)}, \Gamma(u)^{(-\alpha)} \right\}$

When $\alpha = \pm 1$, $\widetilde{\mathcal{D}}_{\Phi}^{(\alpha)}(u, v)$, as well as $\widetilde{\mathcal{D}}_{\widetilde{\Phi}}^{(\alpha)}(x, y)$ introduced earlier, take the form of Bregman divergence (1.33). In this case, the manifold is dually flat, with Riemann curvature tensor $R_{ij\mu\nu}^{(\pm 1)}(u) = R^{(\pm 1)klmn}(x) = 0$.

We summarize the relations between the convex-based divergence functions and the geometry they generate in Table 1.1. The duality associated with $\alpha \leftrightarrow -\alpha$ is called "referential duality" whereas the duality associated with $x \leftrightarrow u$ is called representational duality [23, 24, 27].

1.4.4 Induced Symplectic and Kähler Structures on $\mathfrak{M} \times \mathfrak{M}$

With respect to the $\mathcal{D}_{\Phi}$-divergence (1.40), observe that

$$\Phi\left(\frac{1-\alpha}{2}x + \frac{1+\alpha}{2}y\right) = \Phi\left((\frac{1-\alpha}{4} + \frac{1+\alpha}{4\sqrt{-1}})z + (\frac{1-\alpha}{4} - \frac{1+\alpha}{4\sqrt{-1}})\bar{z}\right) \equiv \widehat{\Phi}^{(\alpha)}(z, \bar{z}),$$
(1.45)

we have

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

which is symmetric in i, j. Both (1.31) and (1.32) 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

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

Proposition 12 ([28]) *A smooth, strictly convex function* $\Phi: dom(\Phi) \subset \mathfrak{M} \to \mathbb{R}$ *induces a a family of Kähler structure* $(\mathfrak{M}, \omega^{(\alpha)}, G^{(\alpha)})$ *defined on* $dom(\Phi) \times dom(\Phi) \subset \mathfrak{M} \times \mathfrak{M}$ *with*

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

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

which is compatible with the canonical almost complex structure J

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

where X, Y are vector fields on $\mathrm{dom}\Phi \times \mathrm{dom}(\Phi)$;
2. *the Riemannian metric $G^{(\alpha)}$, compatible with J and $\omega^{(\alpha)}$ above, is given by $\Phi_{ij}^{(\alpha)}$*

$$G^{(\alpha)} = \Phi_{ij}^{(\alpha)}(dx^i dx^j + dy^i dy^j);$$

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

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

where $\langle \alpha, a \rangle$ denotes a form α operating on a tensor field a. Therefore, restricting to the diagonal $\Delta_{\mathfrak{M}}$, $g^{(\alpha)}$ reduces to the Riemannian metric induced by the divergence $\mathcal{D}_{\Phi}^{(\alpha)}$ through the Eguchi method.

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

$$\tilde{\Gamma}_{i'jk'} = G^{(\alpha)}\left(\nabla_{\frac{\partial}{\partial x^{i'}}}\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^{k'}}\right) = G^{(\alpha)}\left(\nabla_{\frac{\partial}{\partial y^i}}\frac{\partial}{\partial x^j}, \frac{\partial}{\partial y^k}\right),$$

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

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

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

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

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

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

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

1.5 Summary

In order to construct divergence functions in a principled way, this chapter considered the various geometric structures on the underlying manifold $\mathfrak{M}$ induced from a divergence function. Among the geometric structures considered are: statistical structure (Riemannian metric with a pair of torsion-free dual connections, or by simple construction, a family of α-connections), equiaffine structure (those connections that admit parallel volume forms), and Hessian structure (those connections that are dually flat)—they are progressively more restrictive: while any divergence function will induce a statistical manifold, only canonical divergence (i.e., Bregman divergence) will induce a Hessian manifold. Lying in-between these extremes is the equiaffine α-Hessian geometry induced from, say, the class of $\mathcal{D}_\Phi$-divergence. The α-Hessian structure has the advantage of the existence of biorthogonal coordinates, induced from the convex function Φ and its conjugate; these coordinates are convenient for computation. It should be noted that the above geometric structures, from statistical to Hessian, are all induced on the tangent bundle $\mathcal{T}\mathfrak{M}$ of the manifold $\mathfrak{M}$ on which the divergence function is defined.

On the cotangent bundle $\mathcal{T}^*\mathfrak{M}$ side, a divergence function can be viewed as a generating function for a symplectic structure on $\mathfrak{M} \times \mathfrak{M}$ that can be constructed in a "canonical" way. This imposes a "properness" condition on divergence function, stating that the mixed second derivatives of $\mathcal{D}(x, y)$ with respect to x and y must commute. For such divergence functions, a Riemannian structure on $\mathfrak{M} \times \mathfrak{M}$ can be constructed, which can be seen as an extension of the Riemannian structure on $\Delta_\mathfrak{M} \subset \mathfrak{M} \times \mathfrak{M}$. If a further condition on $\mathcal{D}$ is imposed, then $\mathfrak{M} \times \mathfrak{M}$ may be complexified, so it becomes a Kähler manifold. It was shown that $\mathcal{D}_\Phi$-divergence [23] satisfies this Kählerian condition, in addition to itself being proper—the Kähler potential is simply given by the real-valued convex function Φ. These properties, along with the α-Hessian structure it induces on the tangent bundle, makes $\mathcal{D}_\Phi$ a class of divergence functions that enjoy a special role with "nicest" geometric properties, extending the canonical (Bregman) divergence for dually flat manifolds. This will have implications for machine learning, convex optimization, geometric mechanics, etc.

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, Oxford (2000)
3. Barndorff-Nielsen, O.E., Jupp, P.E.: Yorks and symplectic structures. J. Stat. Plan. Inference **63**, 133–146 (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 Comput. Math. Phys. **7**, 200–217 (1967)
5. Calin, O., Matsuzoe, H., Zhang. J.: Generalizations of conjugate connections. In: Sekigawa, K., Gerdjikov, V., Dimiev, S. (eds.) Trends in Differential Geometry, Complex Analysis and Mathematical Physics: Proceedings of the 9th International Workshop on Complex Structures and Vector Fields, pp. 24–34. World Scientific Publishing, Singapore (2009)
6. Csiszár, I.: On topical properties of f-divergence. Studia Mathematicarum Hungarica **2**, 329–339 (1967)
7. Dombrowski, P.: On the geometry of the tangent bundle. Journal fr der reine und angewandte Mathematik **210**, 73–88 (1962)
8. Eguchi, S.: Second order efficiency of minimum contrast estimators in a curved exponential family. Ann. Stat. **11**, 793–803 (1983)
9. Eguchi, S.: Geometry of minimum contrast. Hiroshima Math. J. **22**, 631–647 (1992)
10. Lauritzen, S.: Statistical manifolds. In: Amari, S., Barndorff-Nielsen, O., Kass, R., Lauritzen, S., Rao, C.R. (eds.) Differential Geometry in Statistical Inference. IMS Lecture Notes, vol. 10, pp. 163–216. Institute of Mathematical Statistics, Hayward (1987)
11. Matsuzoe, H.: On realization of conformally-projectively flat statistical manifolds and the divergences. Hokkaido Math. J. **27**, 409–421 (1998)
12. Matsuzoe, H., Inoguchi, J.: Statistical structures on tangent bundles. Appl. Sci. **5**, 55–65 (2003)
13. Matsuzoe, H., Takeuchi, J., Amari, S.: Equiaffine structures on statistical manifolds and Bayesian statistics. Differ. Geom. Appl. **24**, 567–578 (2006)
14. Matsuzoe, H.: Computational geometry from the viewpoint of affine differential geometry. In: Nielsen, F. (ed.) Emerging Trends in Visual Computing, pp. 103–123. Springer, Berlin, Heidelberg (2009)

15. Matsuzoe, M.: Statistical manifolds and affine differential geometry. Adv. Stud. Pure Math. **57**, 303–321 (2010)
16. Nomizu, K., Sasaki, T.: Affine Differential Geometry—Geometry of Affine Immersions. Cambridge University Press, Cambridge (1994)
17. Ohara, A., Matsuzoe, H., Amari, S.: Conformal geometry of escort probability and its applications. Mod. Phys. Lett. B **26**, 1250063 (2012)
18. Shima, H.: Hessian Geometry. Shokabo, Tokyo (2001) (in Japanese)
19. Shima, H., Yagi, K.: Geometry of Hessian manifolds. Differ. Geom. Appl. **7**, 277–290 (1997)
20. Simon, U.: Affine differential geometry. In: Dillen, F., Verstraelen, L. (eds.) Handbook of Differential Geometry, vol. I, pp. 905–961. Elsevier Science, Amsterdam (2000)
21. Uohashi, K.: On α-conformal equivalence of statistical manifolds. J. Geom. **75**, 179–184 (2002)
22. Yano, K., Ishihara, S.: Tangent and Cotangent Bundles: Differential Geometry, vol. 16. Dekker, New York (1973)
23. Zhang, J.: Divergence function, duality, and convex analysis. Neural Comput. **16**, 159–195 (2004)
24. Zhang, J.: Referential duality and representational duality on statistical manifolds. Proceedings of the 2nd International Symposium on Information Geometry and Its Applications, Tokyo, pp. 58–67 (2006)
25. Zhang, J.: A note on curvature of α-connections on a statistical manifold. Ann. Inst. Stat. Math. **59**, 161–170 (2007)
26. 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, vol. III, Chap. 13, pp. 439–466. Springer, New York (2009)
27. Zhang, J.: Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. Entropy **15**, 5384–5418 (2013)
28. Zhang, J., Li, F.: Symplectic and Kähler structures on statistical manifolds induced from divergence functions. In: Nielson, F., Barbaresco, F. (eds.) Proceedings of the 1st International Conference on Geometric Science of Information (GSI2013), pp. 595–603 (2013)