Advances in Mechanics and Mathematics

Volume III

Dedicated to Gilbert Strang on the Occasion of His 70th Birthday

Edited by

David Y. Gao & Hanif D. Sherali
Virginia Polytechnic Institute & State University
Blacksburg, VA 24061, USA
E-mails: gao@vt.edu, hanifs@vt.edu

Chapter 13
Dualistic Riemannian Manifold Structure Induced from Convex Functions

Jun Zhang and Hiroshi Matsuzoe

Key words: Legendre–Fenchel duality, biorthogonal coordinates, Riemannian metric, conjugate connections, equiaffine geometry, parallel volume form, affine immersion, Hessian geometry

13.1 Introduction

Convex analysis has wide applications in science and engineering, such as mechanics, optimization and control, theoretical statistics, mathematical economics and game theory, and so on. It offers an analytic framework to treat systems and phenomena that depart from linearity, based on an elegant mathematical characterization of the notion of "duality" (Rockafellar, 1970, 1974, Ekeland and Temam, 1976). Recent work of David Gao (2000) further provided a comprehensive and unified treatment of duality principles in convex and nonconvex systems, greatly enriching the theoretical foundation and scope of applications.

Central to convex analysis is the Legendre–Fenchel transform, and duality between two sets of variables defined on a pair of vector spaces that are dual with respect to each other. When the convex functions involved are smooth, these variables are in one-to-one correspondence; they can actually be viewed as two coordinate systems on a certain Riemannian manifold. This is the viewpoint from the so-called information geometry (Amari, 1985, Amari and Nagaoka, 2000), and it is investigated at great length in this chapter.

Jun Zhang · Department of Psychology, University of Michigan, Ann Arbor, MI 48109, U.S.A., e-mail: junz@umich.edu

Hiroshi Matsuzoe · Department of Computer Science and Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
e-mail: matsuzoe@nitech.ac.jp

The link between convex functions and Riemannian geometry is shown to be severalfold. First, the pair of convex functions conjugate to one another are the potential functions that induce the Riemannian metric. Second, the two sets of variables are special coordinate systems of the manifold in that they are "biorthogonal;" that is, the Jacobian of coordinate transformation between them is precisely the Riemannian metric. It turns out that biorthogonal coordinates are global coordinates for a pair of dually flat connections on the Riemannian manifold. Third, the Fenchel inequality provides a natural way to construct directed ("pseudo-") distance over the convex point set; this is the Bregman divergence (a.k.a. canonical divergence), which gives rise to the dually flat connections. Finally, the geometric structure (Riemannian metric, conjugate/dual connections) can be induced from graph immersions of a convex function into a higher-dimensional affine space.

Our goal in this chapter is to review such a geometric view of convex functions and the associated conjugacy/duality, as well as provide some new results. We review the background of convex analysis and Riemannian geometry (and affine hypersurface theory) in Section 13.2, with attention to the well-established relation between biorthogonal coordinates and dually flat (also called "Hessian") manifolds. In Section 13.3, we develop the full-fledged α-Hessian geometry, which extends the dually flat Hessian manifold ($\alpha = \pm 1$), and give an example from theoretical statistics when such geometry arises; this parallels the generalization of the convex-induced divergence function with arbitrary α (Zhang, 2004) from Bregman divergence ($\alpha = \pm 1$). To close, we give a summary and discuss some open problems in Section 13.4.

13.2 Convex Functions and Riemannian Geometry

13.2.1 Convex Functions and the Associated Divergence Functions

A strictly convex (or simply "convex") function $\Phi \colon V \subseteq \mathbb{R}^n \to \mathbb{R}, x \mapsto \Phi(x)$ is defined by

$$\frac{1-\alpha}{2}\,\Phi(x) + \frac{1+\alpha}{2}\,\Phi(y) - \Phi\left(\frac{1-\alpha}{2}\,x + \frac{1+\alpha}{2}y\right) > 0 \tag{13.1}$$

for all $x \neq y$ for any $|\alpha| < 1$ (the inequality sign is reversed when $|\alpha| > 1$). In this chapter, V (and $\widetilde{V}$ below) identifies a subset of $\mathbb{R}^n$ both as a point set and as a vector space. We assume Φ to be sufficiently smooth (differentiable up to fourth order). Define

$$\mathcal{B}_\Phi(x, y) = \Phi(x) - \Phi(y) - \langle x - y, \partial\Phi(y)\rangle, \tag{13.2}$$

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 $\widetilde{V} \subseteq \mathbb{R}^n$, and $\langle\cdot,\cdot\rangle_n$ denotes the canonical pairing of a point/vector $x = [x^1, \ldots, x^n] \in V$ and a point/co-vector $u = [u_1, \ldots, u_n] \in \widetilde{V}$ (dual to V):

$$\langle x, u\rangle_n = \sum_{i=1}^{n} x^i u_i. \tag{13.3}$$

(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 convex is

$$\mathcal{B}_\Phi(x, y) > 0 \tag{13.4}$$

for $x \neq y$. We remark that $\mathcal{B}_\Phi$ is sometimes called "Bregman divergence" (Bregman, 1967), widely used in convex optimization literature (Della Pietra et al., 2002, Bauschke, 2003, Bauschke and Combettes, 2003, Bauschke et al., 2003).

Zhang (2004) introduced the following family of functions on $V \times V$ as indexed by $\alpha \in \mathbb{R}$,

$$\mathcal{D}_\Phi^{(\alpha)}(x, y) = \frac{4}{1-\alpha^2}\left(\frac{1-\alpha}{2}\,\Phi(x) + \frac{1+\alpha}{2}\,\Phi(y) - \Phi\left(\frac{1-\alpha}{2}\,x + \frac{1+\alpha}{2}y\right)\right). \tag{13.5}$$

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

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

Note that $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ satisfies the relation (called "referential duality" in Zhang, 2006a)

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

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

From its construction, $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ is nonnegative for $|\alpha| < 1$ due to equation (13.1), and for $|\alpha| = 1$ due to equation (13.4). For $|\alpha| > 1$, assuming $(((1-\alpha)/2)\,x + ((1+\alpha)/2)\,y) \in V$, the nonnegativity of $\mathcal{D}_\Phi^{(\alpha)}(x, y)$ can also be proven due to the inequality (13.1) reversing its sign. Therefore, we have

Lemma 13.1. *For a smooth function $\Phi\colon V \subseteq \mathbb{R}^n \to \mathbb{R}$, the following conditions are equivalent (for $x, y \in V$).*

(i) *Φ is strictly convex.*
(ii) *$\mathcal{D}_\Phi^{(1)}(x, y) \geq 0$.*
(iii) *$\mathcal{D}_\Phi^{(-1)}(x, y) \geq 0$.*

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

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

$$\widetilde{\Phi}(u) = \langle (\partial\Phi)^{-1}(u), u\rangle - \Phi((\partial\Phi)^{-1}(u)), \tag{13.6}$$

with $\widetilde{\widetilde{\Phi}} = \Phi$ and $(\partial\Phi) = (\partial\widetilde{\Phi})^{-1}$. The function $\widetilde{\Phi}$ is also convex, and through which (13.4) 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

$$u = (\partial\Phi)(x) = (\partial\widetilde{\Phi})^{-1}(x) \longleftrightarrow x = (\partial\widetilde{\Phi})(u) = (\partial\Phi)^{-1}(u), \tag{13.7}$$

or, in component form,

$$u_i = \frac{\partial\Phi}{\partial x^i} \longleftrightarrow x^i = \frac{\partial\widetilde{\Phi}}{\partial u_i}\,. \tag{13.8}$$

With the aid of conjugate variables, we can introduce the "canonical divergence" $\mathcal{A}_{\Phi}\colon V \times \widetilde{V} \to \mathbb{R}_+$ (and $\mathcal{A}_{\widetilde{\Phi}}\colon \widetilde{V} \times V \to \mathbb{R}_+$) where $\mathbb{R}_+ = \mathbb{R}^+ \cup \{0\}$

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

They are related to the Bregman divergence (13.2) via

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

Bregman (or canonical) divergence[1] provides a measure of directed distance between two points; that is, it is nonnegative for all values of $x, y \in V$, and vanishes only when $x = y$. More formally, a divergence function $\mathcal{D}\colon V \times V \to \mathbb{R}_+$ is a smooth function (differentiable up to third order) that satisfies

(i) $\mathcal{D}(x,y) \geq 0\ \forall x, y \in V$ with equality holding if and only if $x = y$,
(ii) $\partial_{x^i}\mathcal{D}(x,y)|_{x=y} = \partial_{y^i}\mathcal{D}(x,y)\big|_{x=y} = 0$,
(iii) $\partial_{x^i}\partial_{y^j}\mathcal{D}(x,y)\big|_{x=y}$ is negative definite.

Here ∂_{x^i} denotes partial derivative with respect to the ith component of the x-variable only.[2]

[1] The divergence function, also called the "contrast function," is a terminology arising out of the theoretical statistics literature. It has nothing to do with the divergence operation in vector calculus.

[2] The reader should not confuse the shorthand notations ∂_i with ∂_{x^i} (or ∂_{y^i}): the former operates on a function defined on V such as $\Phi\colon x \mapsto \Phi(x) \in \mathbb{R}$, whereas the latter operates on a function defined on $V \times V$ such as $\mathcal{D}\colon (x,y) \mapsto \mathcal{D}(x,y) \in \mathbb{R}_+$.

13.2.2 Differentiable Manifold: Metric and Connection Structures

A differentiable manifold $\mathfrak{M}$ is a space that locally "looks like" a Euclidean space $\mathbb{R}^n$. By "looks like," we mean that for any base (reference) point $p \in \mathfrak{M}$, there exists a bijective mapping ("coordinate functions") between the neighborhood of p (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) and globally cover the entire manifold. Below, we assume that a coordinate system is chosen such that each point is indexed by $x \in V$, 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 tensor field g which is, when evaluated at each location x (omitted in our notation), a symmetric bilinear form $g(\cdot,\cdot)$ of tangent vectors $X, Y \in T_p(\mathfrak{M}) \simeq \mathbb{R}^n$ such that $g(X,X)$ is always positive for all nonzero vectors X. In local coordinates 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

$$g_{ij}(x) = g(\partial_i, \partial_j). \tag{13.9}$$

The metric tensor allows us to define distance on a manifold as the shortest curve (called "geodesic") connecting two points. It also allows the measurement of angles and hence defines orthogonality of a vector to a submanifold. Projections of vectors to a lower-dimensional submanifold become possible once a metric is given.

Second, there is a structure associated with the notion of "parallelism" of vector fields on a manifold. This is given by the affine (linear) connection (or simply "connection") ∇, mapping two vector fields X and Y to a third one denoted by $\nabla_Y X\colon (X,Y) \mapsto \nabla_Y X$. Intuitively, it represents the "intrinsic" difference of the vector field X from its value at point x and its value at a nearby point connected to x (in the direction given by Y). Here "intrinsic" means that vector comparison at two neighboring locations of the manifold is through a process called "parallel transport," whereby a vector's components are adjusted as it 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 "contravariant" form denoted Γ_{ij}^l (which is a collection of n^3 functions of x),

$$\nabla_{\partial_i} \partial_j = \sum_l \Gamma_{ij}^l \, \partial_l. \tag{13.10}$$

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

$$\widetilde{\Gamma}^l_{mn}(\tilde{x}) = \sum_k \left(\sum_{i,j} \frac{\partial x^i}{\partial \tilde{x}^m} \frac{\partial x^j}{\partial \tilde{x}^n} \Gamma^k_{ij}(x) + \frac{\partial^2 x^k}{\partial \tilde{x}^m \partial \tilde{x}^n} \right) \frac{\partial \tilde{x}^l}{\partial x^k}. \tag{13.11}$$

A curve whose tangent vectors are intrinsically parallel along it is called an "auto-parallel curve."

As a primitive on a manifold, affine connections can be characterized in terms of their torsion and curvature. The torsion T of a connection Γ, which is a tensor itself, is given by the asymmetric part of the connection $T(\partial_i, \partial_j) = \nabla_{\partial_i}\partial_j - \nabla_{\partial_j}\partial_i = \sum_k T^k_{ij}\partial_k$, where T^k_{ij} is its local representation[3] given as

$$T^k_{ij}(x) = \Gamma^k_{ij}(x) - \Gamma^k_{ji}(x).$$

The curviness/flatness of a connection Γ is described by the 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 $R(\partial_i, \partial_j)\partial_k = \sum_l R^l_{kij}\partial_l$ and substituting (13.10), the components of the curvature tensor are[4]

$$R^l_{kij}(x) = \frac{\partial \Gamma^l_{jk}(x)}{\partial x^i} - \frac{\partial \Gamma^l_{ik}(x)}{\partial x^j} + \sum_m \Gamma^l_{im}(x)\Gamma^m_{jk}(x) - \sum_m \Gamma^l_{jm}(x)\Gamma^m_{ik}(x).$$

By definition, R^l_{kij} is antisymmetric when $i \leftrightarrow j$. A connection is said to be flat when $R^l_{kij}(x) \equiv 0$. 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 highly coordinate-dependent. For any flat connection, there exists a local coordinate system under which $\Gamma^k_{ij}(x) \equiv 0$ in a neighborhood; this is the affine coordinate for a flat connection.

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

$$\Gamma_{ij,k} = g(\nabla_{\partial_i}\partial_j, \partial_k) = \sum_l g_{lk}\Gamma^l_{ij}. \tag{13.12}$$

Although Γ^k_{ij} is the more primitive quantity that does not involve the metric, $\Gamma_{ij,k}$ represents the projection of an intrinsically differentiated vector field onto the manifold spanned by the bases ∂_k. The covariant form of the curva-

[3] Here and below, we restrict to holonomic coordinate systems in $\mathbb{R}^n$ only, where all coordinate bases commute $[\partial_i, \partial_j] = 0$ for $i \neq j$.

[4] This componentwise notation of curvature tensor here follows standard differential geometry textbooks, such as Nomizu and Sasaki (1994). On the other hand, information geometers, such as Amari and Nagaoka (2000), adopt the notation $R(\partial_i, \partial_j)\partial_k = \sum_l R^l_{ijk}\partial_l$, with $R_{ijkl} = \sum_l R^m_{ijk} g_{ml}$.

ture tensor is (cf. footnote 4)

$$R_{lkij} = \sum_m g_{lm} \, R^m_{kij}.$$

When the connection is torsion free, R_{lkij} is antisymmetric when $i \leftrightarrow j$ or when $k \leftrightarrow l$, and symmetric when $(i,j) \leftrightarrow (l,k)$. It is related to the Ricci tensor Ric (to be defined in (13.27) below) via $\mathrm{Ric}_{kj} = \sum_{i,l} R_{lkij} g^{il}$.

13.2.3 Dualistic Structure on a Manifold: Compatibility Between Metric and Connection

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; that is, the following condition is satisfied:

$$\partial_k g(\partial_i, \partial_j) = g(\widehat{\nabla}_{\partial_k}\partial_i, \partial_j) + g(\partial_i, \widehat{\nabla}_{\partial_k}\partial_j). \tag{13.13}$$

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

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

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 is compatible with the metric, in the sense that it treats tangent vectors of the shortest curves on a manifold as being parallel (or equivalently, auto-parallel curves are also geodesics).

It turns out that one can define a kind of "compatibility" relation more general than expressed by (13.13), by introducing the notion of "conjugacy" (denoted by $*$) between two connections. A connection ∇^* is said to be "conjugate" to ∇ with respect to g if

$$\partial_k g(\partial_i, \partial_j) = g(\nabla_{\partial_k}\partial_i, \partial_j) + g(\partial_i, \nabla^*_{\partial_k}\partial j). \tag{13.14}$$

Clearly, $(\nabla^*)^* = \nabla$. Moreover, $\widehat{\nabla}$, which satisfies (13.13), 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), (13.14) 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 (13.14) explicitly,

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma_{ki,j} + \Gamma^*_{kj,i}, \tag{13.15}$$

where analogous to (13.10) and (13.12),

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

so that

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

In the following, a manifold $\mathfrak{M}$ with a metric g and a pair of conjugate connections Γ, Γ^* with respect to g is called a "Riemannian manifold with dualistic structure," and denoted by $\{\mathfrak{M}, g, \Gamma, \Gamma^*\}$. Obviously, Γ and Γ^* satisfy the relation (in either covariant or contravariant forms)

$$\widehat{\Gamma} = \frac{1}{2}(\Gamma + \Gamma^*).$$

More generally, in information geometry, a one-parameter family of affine connections $\Gamma^{(\alpha)}$, called "α-connections" ($\alpha \in \mathbb{R}$), is introduced (Amari, 1985, Amari and Nagaoka, 2000)

$$\Gamma^{(\alpha)} = \frac{1+\alpha}{2}\Gamma + \frac{1-\alpha}{2}\Gamma^*. \tag{13.16}$$

Obviously, $\Gamma^{(0)} = \widehat{\Gamma}$.

It can be 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" (Uohashi, 2002). In such case, $\{\mathfrak{M}, g, \Gamma^{(\alpha)}, \Gamma^{(-\alpha)}\}$ is called an "α-Hessian manifold," or a manifold with α-Hessian structure.

13.2.4 Biorthogonal Coordinate Transformation

Consider coordinate transform $x \mapsto u$,

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

where the Jacobian matrix J is given by

$$J_{ij}(x) = \frac{\partial u_i}{\partial x^j}, \qquad J^{ij}(u) = \frac{\partial x^i}{\partial u_j}, \qquad \sum_l J_{il}J^{lj} = \delta_i^j, \tag{13.17}$$

where δ_i^j is the 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

$$J_{ij}(x) = g_{ij}(x), \tag{13.18}$$

then the x-coordinate system and the u-coordinate system are said to be "biorthogonal" to each other because, from the definition of a metric tensor (13.9),

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

In such a case, denote

$$g^{ij}(u) = g(\partial^i, \partial^j), \tag{13.19}$$

which equals $J^{ij}(u)$, the Jacobian of the inverse coordinate transform $u \mapsto x$. Also introduce the (contravariant version) of the affine connection Γ under the u-coordinate system 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 is denoted by superscripted Γ and Γ^*:

$$\Gamma^{ij,k}(u) = g(\nabla_{\partial^i}\partial^j, \partial^k), \qquad \Gamma^{*ij,k}(u) = g(\nabla^*_{\partial^i}\partial^j, \partial^k). \tag{13.20}$$

As in (13.11), the affine connections in u-coordinates (expressed in superscript) and in x-coordinates (expressed in subscript) are related via

$$\Gamma_t^{rs}(u) = \sum_k \left(\sum_{i,j} \frac{\partial x^r}{\partial u_i}\frac{\partial x^s}{\partial u_j}\Gamma_{ij}^k(x) + \frac{\partial^2 x^k}{\partial u_r \partial u_s} \right) \frac{\partial u_k}{\partial x^t} \tag{13.21}$$

and

$$\Gamma^{rs,t}(u) = \sum_{i,j,k} \frac{\partial x^r}{\partial u_i}\frac{\partial x^s}{\partial u_j}\frac{\partial x^t}{\partial u_k}\Gamma_{ij,k}(x) + \frac{\partial^2 x^t}{\partial u_r \partial u_s}. \tag{13.22}$$

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

Analogous to (13.15), we have the following identity,

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

which leads to

Proposition 13.1. *Under biorthogonal coordinates, the component forms of the metric tensor satisfy*

$$\sum_k g^{ik}(u) g_{kj}(x) = \delta^i_j$$

while the pair of conjugate connections Γ, Γ^* satisfies

$$\Gamma^{*ts,r}(u) = -\sum_{i,j,k} g^{ir}(u) g^{js}(u) g^{kt}(u) \Gamma_{ij,k}(x) \tag{13.23}$$

and

$$\Gamma_r^{*\,ts}(u) = -\sum_j g^{js}(u) \Gamma^t_{jr}(x). \tag{13.24}$$

Next, we discuss the conditions under which biorthogonal coordinates exist on an arbitrary Riemannian manifold. From its definition (13.18), we can easily show that

Lemma 13.2. *A Riemannian manifold $\mathfrak{M}$ with metric g_{ij} admits biorthogonal coordinates if and only if $\partial g_{ij}/\partial x^k$ is totally symmetric,*[5]

$$\frac{\partial g_{ij}(x)}{\partial x^k} = \frac{\partial g_{ik}(x)}{\partial x^j}. \tag{13.25}$$

That (13.25) is satisfied for biorthogonal coordinates is evident by virtue of (13.17) and (13.18). Conversely, given (13.25), there must be n functions $u_i(x)$, $i = 1, 2, \ldots, n$ such that

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

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

The above identity, in turn, implies that there exists 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 (13.7), and the pair of convex functions, Φ and its conjugate $\widetilde{\Phi}$, is 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 Lemma 13.2 that a necessary and sufficient condition for a Riemannian manifold to admit biorthogonal coordinates is 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 13.2. *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 dually flat conjugate connections.*

In fact, we can now define a pair of torsion-free connections by

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

and show that they are conjugate with respect to g; that is, they satisfy (13.14). This is to say that we select an affine connection γ such that x is its affine coordinate. From (13.22), when γ^* is expressed in u-coordinates,

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

This implies that u is an affine coordinate system with respect to γ^*. Therefore, biorthogonal coordinates are affine coordinates for a pair of dually flat connections. Such a manifold $\{\mathfrak{M}, g, \gamma, \gamma^*\}$ is called a “Hessian manifold” (Shima, 2007, Shima and Yagi, 1997). It is a special case of the α-Hessian manifold (introduced in Section 13.3.2).

13.2.5 Equiaffine Structure and Parallel Volume Form on a Manifold

For a restrictive class of connections, called "equiaffine" connections, the manifold $\mathfrak{M}$ may admit uniquely a parallel volume form $\omega(x)$. Here, a volume form is a skew-symmetric multilinear map from n linearly independent vectors to a nonzero scalar, and "parallel" is in the sense that $(\partial_i\omega)(\partial_1, \ldots, \partial_n) = 0$ where

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

Applying (13.10), 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

$$\sum_{l} \Gamma_{il}^{l}(x) = \frac{\partial \log \omega(x)}{\partial x^i}. \tag{13.26}$$

Whether a connection is equiaffine is related to the so-called Ricci tensor Ric, defined as the contraction of the curvature tensor R,

$$\mathrm{Ric}_{ij}(x) = \sum_{k} R_{ikj}^{k}(x). \tag{13.27}$$

For a torsion-free connection $\Gamma_{ij}^{k} = \Gamma_{ji}^{k}$, applying the definition of the curvature tensor R to the above yields

$$\mathrm{Ric}_{ij} - \mathrm{Ric}_{ji} = \frac{\partial}{\partial x^i}\left(\sum_{l}\Gamma_{jl}^{l}(x)\right) - \frac{\partial}{\partial x^j}\left(\sum_{l}\Gamma_{il}^{l}(x)\right) = \sum_{k} R_{kij}^{k}. \tag{13.28}$$

One immediately sees that the existence of a function ω satisfying (13.26) is equivalent to the right side of (13.28) being identically zero. In other words, the necessary and sufficient condition for a torsion-free connection to be equiaffine is that its Ricci tensor is symmetric, $\mathrm{Ric}_{ij} = \mathrm{Ric}_{ji}$, or equivalently, $\sum_k R^k_{kij} = 0$.

Making use of (13.26), it is easy to show that the parallel volume form of a Levi-Civita connection $\widehat{\Gamma}$ is given by

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

The parallel volume forms ω, ω^* associated with Γ and Γ^* satisfy (apart from a positive, multiplicative constant)

$$\omega(x)\,\omega^*(x) = (\widehat{\omega}(x))^2 = \det[g_{ij}(x)], \tag{13.29}$$

$$\omega(u)\,\omega^*(u) = (\widehat{\omega}(u))^2 = \det[g^{ij}(u)]. \tag{13.30}$$

Let us now consider the parallel volume forms under biorthogonal coordinates. Contracting the indices t with r in (13.24), and invoking (13.26), we obtain

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

After integration,

$$\omega^*(u)\,\omega(x) = const. \tag{13.31}$$

From (13.29)–(13.31),

$$\omega(u)\,\omega^*(x) = const. \tag{13.32}$$

The relations (13.31) and (13.32) indicate that the volume forms of the pair of conjugate connections, when expressed in biorthogonal coordinates respectively, are inversely proportional to each other. Note that $\omega(x) = \omega(\partial_1, \ldots, \partial_n)$ and $\omega^*(x) = \omega^*(\partial_1, \ldots, \partial_n)$, as skew-symmetric multilinear maps, transform to $\omega(u) = w(\partial^1, \ldots, \partial^n)$ and $\omega^*(u) = \omega^*(\partial^1, \ldots, \partial^n)$ via

$$\omega(x) = \det[J_{ij}(x)]\omega(u) \longleftrightarrow \omega^*(x) = \det[J^{ij}(u)]\omega^*(u),$$

where $\det[J_{ij}(x)] = \det[g_{ij}(x)] = (\det[J^{ij}(u)])^{-1} = (\det[g^{ij}(u)])^{-1}$.

When the pair of equiaffine connections Γ, Γ^* are further assumed to be dually flat, then the entire family of α-connections $\Gamma^{(\alpha)}$ given by (13.16) are equiaffine (Takeuchi and Amari, 2005, Matsuzoe et al., 2006, Zhang, 2007). The $\Gamma^{(\alpha)}$-parallel volume element $\omega^{(\alpha)}$ can be shown to be given by

$$\omega^{(\alpha)} = \omega^{(1+\alpha)/2}(\omega^*)^{(1-\alpha)/2}.$$

Clearly,

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

13.2.6 *Affine Hypersurface Immersion (of Co-Dimension One)*

We next discuss dualistic geometry from convex functions as related to hypersurfaces in affine space, which is the subject of study in affine differential geometry (Simon et al., 1991, Nomizu and Sasaki, 1994).

Let $\mathbb{A}^{n+1}$ be the standard affine space of dimension $n+1$, and $\mathfrak{M}$ an n-dimensional manifold immersed into $\mathbb{A}^{n+1}$ as a hypersurface with affine coordinates $f = [f^1, \ldots, f^{n+1}]$; that is, $f\colon \mathfrak{M} \to \mathbb{A}^{n+1}$. Assume that the local coordinate system on $\mathfrak{M}$ is $x = [x^1, \ldots, x^n]$. Let $\xi = [\xi^1, \ldots, \xi^{n+1}]$ be a vector field defined on $\mathfrak{M}$ that is "transversal," that is, nowhere tangential to $\mathfrak{M}$. Denote the vector space associated with $\mathbb{A}^{n+1}$ as V, with $\dim(V) = n+1$, and the canonical pairing of V with its dual vector space $\widetilde{V}$ (with $\dim(\widetilde{V}) = n+1$) as $\langle\ ,\ \rangle_{n+1}$; see (13.3). The duplet $\{f, \xi\}$ is called an "affine immersion." In local coordinates, they can be explicitly written as functions of x: $\{f(x), \xi(x)\}$, where f is valued in $\mathbb{A}$ and ξ is valued in V.

Because the tangent space $T_p(\mathfrak{M})$ is spanned by

$$\left\{ \left[\frac{\partial f^a}{\partial x^1}, \ldots, \frac{\partial f^a}{\partial x^n} \right],\ a = 1, \ldots, n+1 \right\},$$

we may decompose the second derivatives of f as

$$\frac{\partial^2 f^a}{\partial x^i \partial x^j} = \sum_{k=1}^{n} \Gamma_{ij}^k \frac{\partial f^a}{\partial x^k} + h_{ij}\xi^a \qquad (i, j = 1, \ldots, n), \tag{13.33}$$

where $h_{ij} = h_{ji}$ (called "induced bilinear form" or "affine fundamental form"); if f is convex, then h_{ij} is positive definite. The set of coefficients Γ_{ij}^k is called the "induced connection" on $\mathfrak{M}$, because it is induced by a flat connection on $\mathbb{A}^{n+1}$. Under coordinate transform, these coefficients can be shown to transform according to (13.21). Similarly, decompose the derivative of ξ^a as

$$\frac{\partial \xi^a}{\partial x^i} = -\sum_{k=1}^{n} S_i^k \frac{\partial f^a}{\partial x^k} + \tau_i \xi^a, \tag{13.34}$$

where S_i^k is known as the "affine shape operator," and τ_i is a 1-form on $\mathfrak{M}$ called the "transversal connection form;" when $\tau = 0$ everywhere on $\mathfrak{M}$, the affine immersion $\{f, \xi\}$ is called "equiaffine."

We define a volume form ω on $\mathfrak{M}$ arising out of the immersion of $\{f, \xi\}$,

$$\omega(\partial_1, \ldots, \partial_n) = \mathrm{Det}(\partial_1 f, \ldots, \partial_n f, \xi),$$

where Det is the determinant form on $\mathbb{A}^{n+1}$, and $\partial_i f$ is the vector field $\partial_i f = [\partial_i f^1, \ldots, \partial_i f^{n+1}]$. The covariant derivative of ω is given as follows (see Nomizu and Sasaki, 1994):

$$(\nabla_{\partial_i}\omega)(\partial_1,\ldots,\partial_n) = \tau_i\omega(\partial_1,\ldots,\partial_n).$$

This implies that the induced volume form ω is parallel with respect to the induced connection ∇ if and only if $\{f,\xi\}$ is equiaffine: $\tau = 0$.

In order to consider the geometry induced from convex functions and biorthogonal coordinates, we consider a special kind of affine immersion called "graph immersion:"

$$f = [x^1,\ldots,x^n,\Phi(x)], \qquad \xi = [0,\ldots,0,1], \tag{13.35}$$

where Φ is some nondegenerate (in particular, convex) function. Applying (13.33), we obtain the induced connection $\Gamma_{ij}^k(x) = 0$ and the affine fundamental form $h_{ij}(x)$ as the Hessian of Φ,

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

Thus the geometry of a graph affine immersion coincides with the Hessian geometry induced from a convex function. Because the transversal vector field ξ is parallel along f, from (13.34), obviously $\mathfrak{M}$ has an equiaffine structure.

We can define the "dual" of graph immersion, $\{\tilde{f},\tilde{\xi}\}$, mapping $\mathfrak{M}$ to $\mathbb{A}^{n+1}$ as another graph. Here $\tilde{f} = [u_1,\ldots,u_n,\widetilde{\Phi}(u)]$, with $\widetilde{\Phi}$ and u given by (13.6) and (13.7), respectively. The transversal vector field $\tilde{\xi} = (0,\ldots,0,1)$ is valued in $\widetilde{V}$, the dual vector space. The affine fundamental form $\widetilde{h}$ is

$$\widetilde{h}^{ij}(u) = \frac{\partial^2\widetilde{\Phi}(u)}{\partial u_i\partial u_j}.$$

Because of the identity

$$\frac{\partial^2\widetilde{\Phi}(u)}{\partial u_i\partial u_j} = \sum_{k,l}\frac{\partial x^k}{\partial u_i}\frac{\partial x^l}{\partial u_j}\frac{\partial^2\Phi(x)}{\partial x^k\partial x^l},$$

such affine fundamental form transforms as a 0–2 tensor

$$\widetilde{h}^{ij}(u) = \sum_{k,l}\frac{\partial x^k}{\partial u_i}\frac{\partial x^l}{\partial u_j}h_{kl}(x)$$

(even though second derivatives in general do not transform in a tensorlike fashion). This means that for dual graph immersions $\{f,\xi\}$ and $\{\widetilde{f},\widetilde{\xi}\}$, the induced affine fundamental form is one and the same $\widetilde{h} = h$. The induced objects $\{\mathfrak{M},h,\Gamma,\Gamma^*\}$ form a Hessian structure (i.e., induced connections are dually flat).

More generally, for an arbitrary affine immersion, we can introduce the notion of "co-normal mapping," defined as $\zeta\colon \mathfrak{M} \to \widetilde{V}$ as

$$\langle \xi(x), \zeta(x) \rangle_{n+1} = 1, \tag{13.36}$$

$$\langle \partial_i f(x), \zeta(x) \rangle_{n+1} = 0 \quad (i = 1, \ldots, n); \tag{13.37}$$

that is,

$$\sum_{a=1}^{n+1} \xi^a(x)\, \zeta_a(x) = 1, \qquad \sum_{a=1}^{n+1} \frac{\partial f^a}{\partial x^i}(x)\, \zeta_a(x) = 0 \qquad (i = 1, \ldots, n).$$

Intuitively, the co-normal map is a uniquely defined "normal" vector of the tangent hyperplane at $f(x)$. (This property comes from (13.37).) The co-normal map is not a unit vector; the "length" of the map is normalized by (13.36). Note that the word "length" and "normal" are in quotation marks because no metric has ever been introduced on V or $\widetilde{V}$; normalization is through the pairing operation $\langle \cdot, \cdot \rangle$.

When $\{f, \xi\}$ is equiaffine, then the co-normal map ζ can be viewed as an immersion from $\mathfrak{M}$ to $\mathbb{A}^{n+1}$ (Nomizu and Sasaki, 1994, p. 57). Specifically, $\zeta(\mathfrak{M})$ is taken to be (the negative of) the positional vector field (with respect to a center point) in addition to being the transversal vector field. In this case $\{\tilde{f}, \zeta\} = \{-\zeta, \zeta\}$ is an affine immersion, called the "co-normal immersion" of $\{f, \xi\}$. We also call $\{-\zeta, \zeta\}$ a "centroaffine immersion" because the immersion has a center, with the position vector $-\zeta$ (the first element in the duplet) transversal to its image $\mathfrak{M}$. We denote by $\widetilde{\Gamma}$, $\widetilde{h}$, $\widetilde{\tau}$, $\widetilde{S}$, … the induced objects of $\{-\zeta, \zeta\}$. Then we have the following formulae (see Simon et al., 1991);

$$\widetilde{\Gamma}_{kj,i} = -\Gamma_{ki,j} + \partial_k h_{ij}, \tag{13.38}$$

$$\widetilde{h}_{ij} = \sum_{k=1}^{n} S_i^k h_{kj}, \tag{13.39}$$

$$\widetilde{\tau}_i = 0,$$

$$\widetilde{S}_j^i = \delta_j^i.$$

Equation (13.38) implies that ∇ and $\widetilde{\nabla}$ are mutually conjugate with respect to h. Note that Γ and $\widetilde{\Gamma}$ are, respectively, the induced connections when $\mathfrak{M}$ is immersed into $\mathbb{A}^{n+1}$ in two distinct ways, $\{f, \xi\}$ and $\{-\zeta, \zeta\}$.

Suppose that $\{f, \xi\}$ is a graph affine immersion with respect to some convex function, and $\{-\zeta, \zeta\}$ is the co-normal immersion of $\{f, \xi\}$. From (13.34), the affine shape operator S of $\{f, \xi\}$ vanishes. This implies that $\widetilde{h} = 0$ from (13.39). Thus, although the co-normal map of an equiaffine immersion is a centroaffine hypersurface in $\mathbb{A}^{n+1}$, the co-normal map of graph immersion has its image lie on an affine hyperplane in $\mathbb{A}^{n+1}$.

For an affine immersion $\{f, \xi\}$ and the co-normal immersion $\{-\zeta, \zeta\}$, we define the "geometric divergence" $\mathcal{G}$ on any two points on $\mathfrak{M}$ by

$$\mathcal{G}(x,y) = \langle f(x) - f(y), \zeta(y) \rangle_{n+1} = \sum_{a=1}^{n+1} (f^a(x) - f^a(y))\zeta_a(y).$$

For a graph immersion given by (13.35), we can explicitly solve for ζ from (13.36) and (13.37):

$$\zeta = [-\partial_1 \Phi, \ldots, -\partial_n \Phi, 1].$$

Therefore, the expression for geometric divergence becomes

$$\mathcal{G}(x,y) = -\langle x - y, (\partial \Phi)(y) \rangle_n + \Phi(x) - \Phi(y) \equiv \mathcal{B}_\Phi(x,y);$$

geometric divergence is nothing but Bregman divergence (13.2), see Kurose (1994) and Matsuzoe (1998).

13.2.7 Centroaffine Immersion of Co-Dimension Two

Now we consider affine immersion of $\mathfrak{M}$ (with $\dim(\mathfrak{M}) = n$) into a co-dimension two affine space $\mathbb{A}^{n+2}$ (rather than the co-dimension one affine space $\mathbb{A}^{n+1}$ as discussed in the last section). In this case, in addition to specifying the immersion, denoted by $f : \mathfrak{M} \to \mathbb{A}^{n+2}$, we need to specify two noncollinear vector fields, both "transversal" on $\mathfrak{M}$. The vector space is denoted as V with $\dim(V) = n+2$; the dual vector space is denoted as $\widetilde{V}$ with $\dim(\widetilde{V}) = n+2$. To simply the situation, we consider centroaffine immersion such that one of the transversal vector fields is the (negative of the) positional vector $-f$ and the other is, as before, denoted ξ, that is, the affine immersion is denoted as $\{f, -f, \xi\}$; the elements are valued in $\mathbb{A}^{n+2}$, V, V, respectively. The second derivatives of f and ξ are decomposed as follows (for $i,j = 1, \ldots, n$; $a = 1, \ldots, n+2$);

$$\frac{\partial^2 f^a}{\partial x^i \partial x^j} = \sum_{k=1}^{n} \Gamma_{ij}^k \frac{\partial f^a}{\partial x^k} + h_{ij}\xi^a - t_{ij} f^a,$$

$$\frac{\partial \xi^a}{\partial x^i} = -\sum_{k=1}^{n} S_i^k \frac{\partial f^a}{\partial x^k} + \tau_i \xi^a - \kappa_i f^a.$$

As in affine immersion of co-dimension one, we call Γ_{ij}^k the "induced connection," h_{ij} the "affine fundamental form," τ_i the "transversal connection form," and S_i^k the "affine shape operator." Below, we assume that h is positive definite (i.e., f is strictly convex) and $\tau = 0$ (the centroaffine immersion is equiaffine).

We denote the "dual map" of $\{f, -f, \xi\}$ as another centroaffine map taking the form of $\{\tilde{f}, -\tilde{f}, \zeta\}$, where the elements are valued in $\mathbb{A}^{n+2}$, $\widetilde{V}$, $\widetilde{V}$, respectively; $\tilde{f}$ and ζ are specified by

$$\langle \tilde{f}(x), \xi(x) \rangle_{n+2} = 1, \ \ \langle \zeta(x), \xi(x) \rangle_{n+2} = 0,$$
$$\langle \tilde{f}(x), f(x) \rangle_{n+2} = 0, \ \ \langle \zeta(x), f(x) \rangle_{n+2} = 1,$$
$$\langle \tilde{f}(x), \partial_i f(x) \rangle_{n+2} = 0, \ \ \langle \zeta(x), \partial_i f(x) \rangle_{n+2} = 0 \quad (i = 1, \ldots, n),$$

or explicitly

$$\sum_{a=1}^{n+2} \tilde{f}_a(x) \xi^a(x) = 1, \qquad \sum_{a=1}^{n+2} \zeta_a(x) \xi^a(x) = 0,$$
$$\sum_{a=1}^{n+2} \tilde{f}_a(x) f^a(x) = 0, \qquad \sum_{a=1}^{n+2} \zeta_a(x) f^a(x) = 1, \tag{13.40}$$
$$\sum_{a=1}^{n+2} \tilde{f}_a(x) \frac{\partial f^a}{\partial x^i}(x) = 0, \qquad \sum_{a=1}^{n+2} \zeta_a(x) \frac{\partial f^a}{\partial x^i}(x) = 0 \qquad (i = 1, \ldots, n). \tag{13.41}$$

Denote the induced objects as $\widetilde{\Gamma}$, $\widetilde{h}$, $\widetilde{\tau}$, $\ldots$; we have the following formulae (see Nomizu and Sasaki, 1994, Matsuzoe, 1998);

$$\begin{aligned} \partial_k h_{ij} &= \Gamma_{ki,j} + \widetilde{\Gamma}_{kj,i}, \\ \widetilde{h}_{ij} &= h_{ij}, \\ \widetilde{\tau}_i &= 0. \end{aligned} \tag{13.42}$$

We remark that (13.42) is different from (13.39) of the co-dimension one case. If a centroaffine immersion $\{f, -f, \xi\}$ induces $\{g, \Gamma\}$ on $\mathfrak{M}$, then the dual map $\{\tilde{f}, -\tilde{f}, \zeta\}$ induces $\{g, \widetilde{\Gamma}\}$ on $\mathfrak{M}$. This implies that the theory of centroaffine immersions of co-dimension two is more useful than that of affine immersions of co-dimension one when we discuss the duality of statistical manifold.

Consider the special case of graph immersion (of co-dimension two) $\{f, -f, \xi\}$; that is,

$$f = [x^1, \ldots, x^n, \Phi(x), 1], \qquad \xi = [0, \ldots, 0, 1, 0], \tag{13.43}$$

where $\Phi(x)$ is some convex function. If $\{f, -f, \xi\}$ has other representations, they are centroaffinely congruent (linearly congruent) to (13.43); hence it suffices to consider (13.43).

From straightforward calculations, the dual map $\{\tilde{f}, -\tilde{f}, \zeta\}$ of $\{f, -f, \xi\}$ takes the form

$$\tilde{f} = [-u_1, \ldots, -u_n, 1, \widetilde{\Phi}(u)], \qquad \zeta = [0, \ldots, 0, 0, 1]. \tag{13.44}$$

The left side equation in (13.40) then gives

$$-\sum_{i=1}^{n} x^i u_i + \Phi(x) + \widetilde{\Phi}(u) = 0,$$

and the left side equation in (13.41) is

$$-u_i + \frac{\partial \Phi}{\partial x^i}(x) = 0.$$

Thus, $\widetilde{\Phi}$ is the convex conjugate of Φ as in (13.6), and $u = [u_1, \ldots, u_n]$ is the conjugate variable as in (13.7). For graph immersion, it is easy to check that $\Gamma_{ki,j} = 0$, $S_i^k = 0$, $t_{ij} = 0$, $\tau_i = 0$, $\kappa_i = 0$ for all indices and

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

The same is true for induced objects in dual immersion.

Just as in the case of equiaffine immersion $\{f, \xi\}$ of co-dimension one and the associated co-normal map $\{-\zeta, \zeta\}$, we can construct the geometric divergence $\mathcal{G}$ on $\mathfrak{M}$ for centroaffine immersion $\{f, -f, \xi\}$ of co-dimension two and the associated dual map $\{\tilde{f}, -\tilde{f}, \zeta\}$:

$$\begin{aligned}\mathcal{G}(x,y) &= \langle \tilde{f}(y), f(x) - f(y) \rangle_{n+2} \\ &= \langle \tilde{f}(y), f(x) \rangle_{n+2} \\ &= \sum_{a=1}^{n+2} \tilde{f}_a(y) f^a(x).\end{aligned}$$

For graph immersion, we substitute f and $\tilde{f}$ in (13.43) and (13.44) to yield

$$\begin{aligned}\mathcal{G}(x,y) &= -\langle x, (\partial\Phi)(y) \rangle_n + \Phi(x) + \widetilde{\Phi}((\partial\Phi)(y)) \\ &\equiv \mathcal{B}_{\Phi}(x,y).\end{aligned}$$

In both the equiaffine immersion of co-dimension one (discussed in Section 13.2.6) and centroaffine immersion of co-dimension two (discussed here), the notion of geometric divergence is a generalization of the Bregman (canonical) divergence on a dually flat space.

Proposition 13.3. (Kurose, 1994, Matsuzoe, 1998) *Let Φ be a strictly convex function on $\mathbb{R}^n$. Then geometric divergence $\mathcal{G}(x,y) \colon V \times V \to \mathbb{R}$ induced by the affine immersion of Φ as a graph in $\mathbb{A}^{n+1}$ or by the centroaffine immersion of Φ as a graph in $\mathbb{A}^{n+2}$ equals the Bregman divergence $\mathcal{B}_{\Phi}(x,y)$.*

13.3 The α-Hessian Structure Associated with Convex-Induced Divergence

The discussion at the end of the last section anticipate a close relation between convex functions and the Riemannian structure on a differentiable manifold whose coordinates are the variables of the convex functions. On such a manifold, divergence functions take the role of pseudo-distance functions that are nonnegative but need not be symmetric. That dualistic Riemannian manifold structures can be induced from a divergence function was first demonstrated by S. Eguchi.

Lemma 13.3. (Eguchi, 1983, 1992) *A divergence function induces a Riemannian metric g and a pair of conjugate connections Γ, Γ^* given as*

$$g_{ij}(x) = -\partial_{x^i}\partial_{y^j}\mathcal{D}(x,y)\Big|_{y=x}; \tag{13.45}$$

$$\Gamma_{ij,k}(x) = -\partial_{x^i}\partial_{x^j}\partial_{y^k}\mathcal{D}(x,y)\Big|_{y=x}; \tag{13.46}$$

$$\Gamma^*_{ij,k}(x) = -\partial_{y^i}\partial_{y^j}\partial_{x^k}\mathcal{D}(x,y)\Big|_{y=x}. \tag{13.47}$$

It is easily verifiable that $g_{ij}, \Gamma_{ij,k}, \Gamma^*_{ij,k}$ as given above satisfy (13.15). Furthermore, under arbitrary coordinate transform, these quantities behave properly as desired. Equations (13.45)–(13.47) link a divergence function $\mathcal{D}$ to the dualistic Riemannian structure $\{\mathfrak{M}, g, \Gamma, \Gamma^*\}$.

Applying Lemma 13.3 to Bregman divergence $\mathcal{B}_\Phi(x,y)$ given by (13.2) yields

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

and

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

Calculating their curvature tensors shows the pair of connections are dually flat. It is commonly referred to, in affine geometry literature, as the "Hessian manifold" (see Section 13.2.4), although in the study by Shima (2007), the potential function Φ need not be convex but only semidefinite. In u-coordinates, these geometric quantities can be expressed as

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

where $\widetilde{\Phi}$ is the convex conjugate of Φ. Below, this link from convex functions to Riemannian manifold is explored in greater detail.

13.3.1 The α-Hessian Geometry

We start by reviewing a main result from Zhang (2004) linking the divergence function $\mathcal{D}^{(\alpha)}_\Phi(x,y)$ defined in (13.5) and the α-Hessian structure.

Proposition 13.4. (Zhang, 2004) *The manifold* $\{\mathfrak{M}, g_x, \Gamma_x^{(\alpha)}, \Gamma_x^{(-\alpha)}\}$[6] *associated with* $\mathcal{D}_{\Phi}^{(\alpha)}(x,y)$ *is given by*

$$g_{ij}(x) = \Phi_{ij} \tag{13.48}$$

and

$$\Gamma_{ij,k}^{(\alpha)}(x) = \frac{1-\alpha}{2}\,\Phi_{ijk}, \qquad \Gamma_{ij,k}^{*(\alpha)}(x) = \frac{1+\alpha}{2}\,\Phi_{ijk}. \tag{13.49}$$

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

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

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 (13.16) and $\Gamma^{(\pm 1)}$ are dually flat). From (13.49),

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

with the Levi-Civita connection given as:

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

Straightforward calculation shows that:

Corollary 13.1. *For* α*-Hessian manifold* $\{\mathfrak{M}, g_x, \Gamma_x^{(\alpha)}, \Gamma_x^{(-\alpha)}\}$,

(i) *The 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*

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

The reader is reminded that the metric and conjugated connections in the forms (13.48) and (13.49) are induced from (13.5). Using the convex conjugate $\widetilde{\Phi}\colon \widetilde{V} \to \mathbb{R}$ given by (13.6), we introduce the following family of divergence functions $\widetilde{\mathcal{D}}_{\widetilde{\Phi}}^{(\alpha)}(x,y)\colon V \times V \to \mathbb{R}_+$ defined by

---

[6] The subscript in x (or u below) indicates that the x-coordinate system (or u-coordinate system, resp.) is being used. Recall from Section 13.2.4 that under x (u, resp.) local coordinates g and Γ, in component forms, are expressed by lower (upper, resp.) indices.

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

Explicitly written, this new family of divergence functions is

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

Straightforward calculation shows that $\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x,y)$ induces the α-Hessian structure $\{\mathfrak{M}, g_x, \Gamma^{(-\alpha)}_x, \Gamma^{(\alpha)}_x\}$ where $\Gamma^{(\mp\alpha)}$ are given by (13.49); that is, the pair of α-connections are themselves "conjugate" (in the sense of $\alpha \leftrightarrow -\alpha$) to those induced by $\mathcal{D}^{(\alpha)}_{\Phi}(x,y)$.

13.3.2 Biorthogonal Coordinates on α-Hessian Manifold

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

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

Explicitly written,

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

Recalling our notation (13.19) and (13.20), we have

Corollary 13.2. *The α-Hessian manifold $\{\mathfrak{M}, g_u, \Gamma^{(\alpha)}_u, \Gamma^{(-\alpha)}_u\}$ associated with $\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u,v)$ is given by*

$$g^{ij}(u) = \widetilde{\Phi}^{ij}(u), \tag{13.50}$$

$$\Gamma^{(\alpha)ij,k}(u) = \frac{1+\alpha}{2}\,\widetilde{\Phi}^{ijk}, \qquad \Gamma^{*(\alpha)ij,k}(u) = \frac{1-\alpha}{2}\,\widetilde{\Phi}_{ijk}. \tag{13.51}$$

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

Table 13.1 Divergence functions and induced geometry

Divergence Function	Defined as	Induced Geometry
$\mathcal{D}^{(\alpha)}_{\Phi}(x,y)$	$V \times V \to \mathbb{R}_+$	$\left\{\Phi_{ij}, \Gamma^{(\alpha)}_x, \Gamma^{(-\alpha)}_x\right\}$
$\mathcal{D}^{(\alpha)}_{\widetilde{\Phi}}((\partial\Phi)(x),(\partial\Phi)(y))$	$V \times V \to \mathbb{R}_+$	$\left\{\Phi_{ij}, \Gamma^{(-\alpha)}_x, \Gamma^{(\alpha)}_x\right\}$
$\mathcal{D}^{(\alpha)}_{\widetilde{\Phi}}(u,v)$	$\widetilde{V} \times \widetilde{V} \to \mathbb{R}_+$	$\left\{\widetilde{\Phi}^{ij}, \Gamma^{(-\alpha)}_u, \Gamma^{(\alpha)}_u\right\}$
$\mathcal{D}^{(\alpha)}_{\Phi}((\partial\widetilde{\Phi})(u),(\partial\widetilde{\Phi})(v))$	$\widetilde{V} \times \widetilde{V} \to \mathbb{R}_+$	$\left\{\widetilde{\Phi}^{ij}, \Gamma^{(\alpha)}_u, \Gamma^{(-\alpha)}_u\right\}$

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

We remark that the same metric (13.50) and the same α-connections (13.51) are induced by $\mathcal{D}^{(-\alpha)}_{\widetilde{\Phi}}(u,v) \equiv \mathcal{D}^{(\alpha)}_{\widetilde{\Phi}}(v,u)$; this follows as a simple application of Lemma 13.3.

An application of (13.23) gives rise to the following relations.

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

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

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

When $\alpha = \pm 1$, $\widetilde{\mathcal{D}}^{(\alpha)}_{\Phi}(u,v)$, as well as $\widetilde{\mathcal{D}}^{(\alpha)}_{\widetilde{\Phi}}(x,y)$ introduced earlier, take the form of Bregman divergence (13.2). In this case, the manifold is dually flat, with curvature tensor $R^{(\pm 1)}_{ij\mu\nu}(x) = R^{(\pm 1)klmn}(u) = 0$.

We summarize the relations between the convex-induced divergence functions and the geometry they generate in Table 13.1.

13.3.3 Applications of α-Hessian Geometry

Finally, we give an application of the α-Hessian geometry in mathematical statistics. A statistical model is a set of (what we call) ζ-functions $\zeta \mapsto p(\zeta)$, where a ζ-function is an element of some function space $\mathbb{B} = \{p(\cdot)\colon \mathcal{X} \to$

$\mathbb{R}, p(\zeta) > 0\}$ over a σ-finite set $\mathcal{X}$ with dominant measure μ. A parametric model $\mathfrak{M}_\theta$ is defined as

$$\mathfrak{M}_\theta = \{p(\cdot|\theta) \in \mathbb{B}, \theta \in V \subseteq \mathbb{R}^n\}.$$

That is, $\mathfrak{M}_\theta$ forms a smooth manifold with θ as coordinates.

One can define divergence functionals to measure the directed distance between two ζ-functions p and q. The most familiar is the Kullback–Leibler divergence. With the aid of a smooth and strictly convex function $f\colon \mathbb{R} \to \mathbb{R}$ and a strictly increasing function $\rho\colon \mathbb{R} \to \mathbb{R}$, one can show that the following is a general form of convex-induced divergence functional.

$$\frac{4}{1-\alpha^2} \int_{\mathcal{X}} \frac{1-\alpha}{2} f(\rho(p)) + \frac{1+\alpha}{2} f(\rho(q)) - f\left(\frac{1-\alpha}{2}\rho(p) + \frac{1+\alpha}{2}\rho(q)\right) d\mu, \tag{13.52}$$

since it is nonnegative and equals zero if and only if $p(\zeta) = q(\zeta)$ almost surely. A parametric model $p(\cdot|\theta) \in \mathfrak{M}_\theta$ is said to be "ρ-affine" if there exists a set of linearly independent functions $\lambda_i(\zeta) \in \mathbb{B}$ such that

$$\rho(p(\zeta|\theta)) = \sum_i \theta^i \lambda_i(\zeta).$$

The parameter $\theta = [\theta^1, \ldots, \theta^n]$ is called the "natural parameter" of a ρ-affine parametric model, and the functions $\lambda_1(\zeta), \ldots, \lambda_n(\zeta)$ are the affine basis functions. Examples of ρ-affine manifold include the so-called "alpha-affine manifolds" (Amari, 1985, Amari and Nagaoka, 2000), where $\rho(\cdot)$ takes on the following form (indexed by $\beta \in [-1, 1]$),

$$l^{(\beta)}(t) = \begin{cases} \log t & \beta = 1, \\ \dfrac{2}{1-\beta}\, t^{(1-\beta)/2} & \beta \in [-1, 1). \end{cases}$$

When a parametric model is ρ-affine, the function

$$\Phi(\theta) = \int_{\mathcal{X}} f(\rho(p(\zeta|\theta)))\, d\mu = \int_{\mathcal{X}} f\left(\sum_i \theta^i \lambda_i(\zeta)\right) d\mu$$

can be shown to be strictly convex. Therefore, the divergence functional in (13.52) takes the form of the divergence function $\mathcal{D}_\Phi^{(\alpha)}(\theta_p, \theta_q)$ on $V \times V$ given by

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

This is exactly (13.5)! An immediate consequence is that a ρ-affine manifold is the α-Hessian manifold, with metric and affine connections given by Proposition 13.4.

For any ζ-function $\zeta \mapsto p(\zeta)$, we now define

$$\eta_i = \int_{\mathcal{X}} f'(\rho(p(\zeta)))\, \lambda_i(\zeta)\, d\mu$$

such that $\eta = [\eta_1, \ldots, \eta_n] \in \widetilde{V} \subseteq \mathbb{R}^n$. We call η the "expectation parameter" of $p(\zeta)$ with respect to the set of (affine basis) functions $\lambda_1(\zeta), \ldots, \lambda_n(\zeta)$. It can be easily verified that for the ρ-affine parametric models,

$$\eta_i = \frac{\partial \Phi(\theta)}{\partial \theta^i}.$$

Define

$$\Phi^*(\theta) = \int_{\mathcal{X}} \tilde{f}(f'(\rho(p(\zeta|\theta))))\, d\mu,$$

where $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$ is the Fenchel conjugate of f; then $\widetilde{\Phi}(\eta) \equiv \Phi^*((\partial\Phi)^{-1}(\eta))$ is the Fenchel conjugate of $\Phi(\theta)$. The pair of convex functions Φ, Φ^* induces η, θ via:

$$\frac{\partial \Phi(\theta)}{\partial \theta^i} = \eta_i \longleftrightarrow \frac{\partial \widetilde{\Phi}(\eta)}{\partial \eta_i} = \theta^i\,.$$

In theoretical statistics, we can call $\Phi(\theta)$ the generalized cumulant generating function (or partition function), and $\widetilde{\Phi}(\eta)$ the generalized entropy function. Natural parameter θ and expectation parameter η, which form biorthogonal coordinates, play important roles in statistical inference.

13.4 Summary and Open Problems

For two smooth, strictly convex functions $\Phi, \widetilde{\Phi}$ that are mutually conjugate, the variables $u = \partial\Phi(x)$ and $x = \partial\widetilde{\Phi}(u)$ are in one-to-one correspondence. It has been shown in this chapter that such a pair of variables can be viewed as biorthogonal coordinate systems on a Riemannian manifold whose metric is the second derivative of Φ when the x-coordinate system is used (or of $\widetilde{\Phi}$ when the u-coordinate system is used). Furthermore, a family of affine connections (indexed by α) can be defined with nonzero curvatures except for $\alpha = \pm 1$, the dually flat case (the so-called "Hessian manifold"). Each of these α-connections is equiaffine and admits a parallel volume form, and the entire family is induced from the divergence function $\mathcal{D}_{\Phi}^{(\alpha)}$ (or $\widetilde{\mathcal{D}}_{\tilde{\Phi}}^{(\alpha)}$) associated with any convex function Φ (or $\widetilde{\Phi}$).

Our analysis revealed that the conjugate ($\pm\alpha$)-connections reflect two kinds of duality embodied by the convex-induced divergence function. The first is referential duality related to the choice of the reference and the comparison status for the two points (x versus y) for computing the value of the divergence $\mathcal{D}_{\Phi}^{(\alpha)}(x,y) = \mathcal{D}_{\Phi}^{(-\alpha)}(y,x)$. The second is representational duality related to the construction of two families of divergence functions, $\mathcal{D}_{\Phi}^{(\alpha)}(x,y)$ versus $\mathcal{D}_{\tilde{\Phi}}^{(\alpha)}((\partial\Phi)(x),(\partial\Phi)(y))$, using conjugate convex functions (see Table 13.1 in Section 13.3). The geometric quantities expressed in x-coordinates and expressed in u-coordinates are related to each other via Proposition 13.1. When $\alpha = \pm 1$, the two members of divergence functions coincide (and become Bregman divergence), so that the two kinds of duality reveal themselves as biduality:

$$\mathcal{D}_{\Phi}^{(-1)}(x,y) = \mathcal{D}_{\tilde{\Phi}}^{(-1)}(\partial\Phi(y),\partial\Phi(x)) = \mathcal{D}_{\tilde{\Phi}}^{(1)}(\partial\Phi(x),\partial\Phi(y)) = \mathcal{D}_{\Phi}^{(1)}(y,x) \ ,$$

which is compactly written in the form of canonical divergence as

$$\mathcal{A}_{\Phi}(x,v) = \mathcal{A}_{\tilde{\Phi}}(v,x) \ .$$

The relation between convex-induced divergence functions and α-connections is intriguing; that α as a convex mixture parameter coincides with α as indexing the family of connections is remarkable! We know that, in general, there may be many families of divergence functions that could yield the same α-connections. An explicit construction is as follows. Take the families of divergence functions ($\gamma \in \mathbb{R}$, $\beta \in [-1,1]$)

$$\frac{1+\beta}{2}\mathcal{D}_{\Phi}^{(\gamma)}(x,y) + \frac{1-\beta}{2}\mathcal{D}_{\Phi}^{(-\gamma)}(x,y) \, ,$$

which induce an α-Hessian structure whose metric and conjugate connections are given in the forms (13.48) and (13.49), with α taking the value of $\beta\gamma$. The nonuniqueness of divergence functions giving rise to the family of α-connections invites the question of how to characterize the convex-induced divergence functions from the perspective of α-Hessian geometry. There is reason to believe that such axiomatization is possible because (i) the form of divergence function for the dually flat manifold ($\alpha = \pm 1$) is unique, namely, the Bregman divergence $\mathcal{B}_{\Phi}$; (ii) Lemma 13.1 gives that $\mathcal{D}^{(\alpha)} \geq 0$ if and only if $\mathcal{B}_{\Phi} \geq 0$ for any smooth function Φ. This hints at a deeper connection yet to be understood between convexity of a function and the α-Hessian geometry.

Another topic that needs further investigation is with respect to affine hypersurface realization of the α-Hessian manifold. We know that in affine immersion, geometric divergence is a generalization of the canonical divergence of dually flat (i.e., Hessian) manifolds. How to model the nonflat manifold with a general α value remains an open question. In particular, is there a generalization of geometric divergence that mirrors the way a convex-induced divergence function $\mathcal{D}_{\Phi}^{(\alpha)}$ generalizes Bregman divergence $\mathcal{B}_{\Phi}$ (or equivalently, the canonical divergence $\mathcal{A}_{\Phi}$)?

Finally, how do we extend the above analysis to an infinite-dimensional setting? The use of convex analysis (in particular, Young function and Orlicz space) to model the infinite-dimensional probability manifold yields fruitful insights for understanding difficult topological issues (Pistone and Sempi, 1995). It would thus be a worthwhile effort to extend the notion of biorthogonal coordinates to the infinite-dimensional manifold to study nonparametric information geometry. To this end, it would also be useful to extend the affine hypersurface theory to the infinite-dimensional setting and provide the formulation for co-dimension one affine immersion and co-dimension two centroaffine immersion. Here, affine hypersurfaces are submanifolds (resulting from normalization and positivity constraints on probability density functions; see, e.g., Zhang and Hasto, 2006) of an ambient manifold of unrestricted Banach space functions. Preliminary analyses (Zhang, 2006b) show that such an ambient manifold is flat for all α-connections, $\alpha \in \mathbb{R}$. So it provides a natural setting (i.e., affine space) in which probability densities can be embedded as an affine hypersurface. The value of such a viewpoint for statistical inference remains a topic for future exploration.

References

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

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

Bauschke, H.H. (2003). Duality for Bregman projections onto translated cones and affine subspaces. *J. Approx. Theory* **121**, 1–12.

Bauschke, H.H. and Combettes, P.L. (2003). Iterating Bregman retractions. *SIAM J. Optim.* **13**, 1159–1173.

Bauschke, H.H., Borwein, J.M., and Combettes, P.L. (2003). Bregman monotone optimization algorithms. *SIAM J. Control Optim.* **42**, 596–636.

Bregman, L.M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Physics* **7**, 200–217.

Della Pietra, S., Della Pietra, V., and Lafferty, J. (2002). Duality and auxiliary functions for Bregman distances. *Technical Report CMU-CS-01-109*, School of Computer Science, Carnegie Mellon University.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statistics* **11**, 793–803.

Eguchi, S. (1992). Geometry of minimum contrast. *Hiroshima Math. J.* **22**, 631–647.

Ekeland, I. and Temam, R. (1976). *Convex Analysis and Variational Problems.* SIAM, Amsterdam.

Gao, D.Y. (2000). *Duality Principles in Nonconvex Systems: Theory, Methods and Applications.* Kluwer Academic, Dordrecht, xviii+454 pp.

Kurose, T. (1994). On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J.* **46**, 427–433.

Matsuzoe, H. (1998). On realization of conformally-projectively flat statistical manifolds and the divergences. *Hokkaido Math. J.* **27**, 409–421.

Matsuzoe, H., Takeuchi, J., and Amari. S (2006). Equiaffine structures on statistical manifolds and Bayesian statistics. *Differential Geom. Appl.* **24**, 567–578.

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

Pistone, G. and Sempi, C. (1995). An infinite dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statistics* **33**, 1543–1561.

Rockafellar, R.T. (1970). *Convex Analysis*. Princeton University Press.

Rockafellar, R.T. (1974). *Conjugate Duality and Optimization*. SIAM, Philadelphia.

Shima, H. (2007). *The Geometry of Hessian Structures*. World Scientific, Singapore.

Shima, H. and Yagi, K. (1997). Geometry of Hessian manifolds. *Differential Geom. Appl.* **7**, 277–290.

Simon, U., Schwenk-Schellschmidt, A., and Viesel, H. (1991). *Introduction to the Affine Differential Geometry of Hypersurfaces*. Lecture Notes, Science University of Tokyo.

Takeuchi, J. and Amari, S. (2005). α-Parallel prior and its properties. *IEEE Trans. Inf. Theory* **51**, 1011–1023.

Uohashi, K. (2002). On α-conformal equivalence of statistical manifolds. *J. Geom.* **75**, 179–184.

Zhang, J. (2004). Divergence function, duality, and convex analysis. *Neural Comput.* **16**, 159–195.

Zhang, J. (2006a). Referential duality and representational duality in the scaling of multidimensional and infinite-dimensional stimulus space. In: Dzhafarov, E. and Colonius, H. (Eds.) *Measurement and Representation of Sensations: Recent Progress in Psychological Theory*. Lawrence Erlbaum, Mahwah, NJ.

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

Zhang, J. (2007). A note on curvature of α-connections on a statistical manifold. *Ann. Inst. Statist. Math.* **59**, 161–170.

Zhang, J. and Hasto, P. (2006). Statistical manifold as an affine space: A functional equation approach. *J. Math. Psychol.* **50**, 60–65.