Information Geometry (2019) 2:77–98
https://doi.org/10.1007/s41884-019-00018-x

RESEARCH PAPER

From Hessian to Weitzenböck: manifolds with torsion-carrying connections

Jun Zhang[1] · **Gabriel Khan**[1]

Received: 22 December 2018 / Revised: 4 July 2019 / Published online: 18 July 2019

Abstract

We investigate affine connections that have zero curvature but not necessarily zero torsion. Slightly generalizing from what is known as Weitzenböck connections, such non-flat connections (we call "pseudo-Weitzenböck connections") can be constructed from any frame (a set of linearly independent vector fields), not just the orthonormal frames, with their torsions vanishing if and only if the frame is a coordinate frame. In such situations, the notion of biorthogonal frames generalizes the notion of biorthogonal coordinates of a Hessian manifold, with respect to any given Riemannian metric g (not necessarily Hessian). Our main theorem shows that the pair of pseudo-Weitzenböck connections, each adapted to one of the pair of g-biorthogonal frames, are g-conjugate to each other. As a result, the pseudo-Weitzenböck connection pair generalize dually flat connections characteristic of Hessian manifolds, by being both curvature-free yet admitting (generally unequal) torsions. These results allow us to construct a pseudo-Weitzenböck connection for the manifold of parametric statistical models and treat it as a "statistical manifold admitting torsion".

Keywords Codazzi coupling · Conjugate connection · Biorthogonal · Curvature · Torsion · Dually flat · Partially flat

1 Introduction

The most well-understood case of information geometry are dually flat manifolds. These are Riemannian manifolds which admit a conjugate pair of affine connections that are both curvature- and torsion-free, or simply "flat." As such, there exists a pair of affine coordinates for which the corresponding affine connection has vanishing

✉ Jun Zhang
junz@umich.edu

Gabriel Khan
gabekhan@umich.edu

1 University of Michigan, Ann Arbor, USA

coefficients. A dually flat manifold is also called a *Hessian manifold*, because, when evaluated on the coordinates associated to either of the flat connections, the Riemannian metric takes the form of the Hessian (second-derivative) of a strictly convex potential. Hessian manifolds enjoy especially nice properties, including the existence of a pair of convex potentials conjugate to each other, of the canonical form of divergence function expressed in these potentials and affine coordinates, of the Pythagorean law associated to such divergence function, etc. Finally, the pair of affine coordinates are biorthogonal with respect to the Riemannian (in fact Hessian) metric.

For any flat connection ∇ on a manifold $\mathbb{M}$, we can associate a family of Riemannian metrics g^Φ taking the form of the Hessian operator $Hess_\nabla$ of some smooth convex function Φ

$$g^\Phi = Hess_\nabla(\Phi)$$

which, in the affine coordinates $x^i (i = 1, \ldots, n = \dim \mathbb{M})$ of the flat connection ∇, is simply

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

where ∂_{x^j} is the shorthand for $\partial/\partial x^j$.

It is worth noting that not all metrics can be written in this form. Reference [5] provides more details on the curvature and cohomological obstructions to the existence of such a metric.

Any pair (g, ∇) of a Riemannian metric and a connection leads to a well-defined conjugate connection ∇^*. When ∇ is flat, ∇^* is necessarily curvature-free. However, ∇^* may carry torsion even though ∇ is torsion-free. Torsion of ∇^* vanishes under the condition that the flat ∇ is Codazzi coupled to g, which gives rise to the dually flat situation. These well-known facts were recapitulated in Sect. 2.1.

In the interest of understanding the duality in Hessian manifold in more detail, we seek to understand a pair of curvature-free connections that are conjugate connections with respect to some Riemannian metric g, while dropping the assumption that such connections are torsion-free. Specifically, we want to investigate the following two cases that deviate from a Hessian manifold, despite $\mathbb{M}$ being equipped with a curvature-free connection ∇ (with its g-conjugate connection ∇^* necessarily also curvature-free):

(i) ∇ carries torsion but is Codazzi coupled to g;
(ii) ∇ is torsion-free (and hence flat), but may not be Codazzi coupled to g.

In both cases, one can define uniquely the g-conjugate connection, although the Riemannian metric g itself may no longer be a Hessian metric.

In Sect. 2.2, we will first examine in detail the Hessian, or second-order derivative, $Hess_\nabla(\Phi)$ of a function Φ on the manifold as a bilinear form, and investigate the role of curvature and torsion in the definition of Hessian metric g^Φ. Our Proposition 3, along with Lemma 2, shows that the bilinear form $Hess_\nabla(\Phi)$ is

(i) symmetric if and only if ∇ is torsion-free;

(ii) Codazzi coupled to a torsion-free ∇ if and only if its curvature R^{∇} satisfies $d\Phi(R^{\nabla}) = 0$.

Statement (i) shows that torsion-freeness of ∇ is necessary for $Hess_{\nabla}(\Phi)$ to be considered as a metric. Statement (ii) shows that furthermore, curvature-freeness of ∇ will lead to Codazzi coupling of ∇ with $Hess_{\nabla}(\Phi)$. Therefore for any flat connection ∇, a potential Φ induces a dually flat structure $(Hess_{\nabla}(\Phi), \nabla, \nabla^*)$, where ∇^* is dual of ∇ with respect to $Hess_{\nabla}(\Phi)$. Assuming Φ to be convex, then $Hess_{\nabla}(\Phi) = Hess_{\nabla^*}(\Phi^*)$ can be taken as the Riemannian metric—here the function Φ^* is the Legendre conjugate of Φ (Sect. 2.3).

Having understood the importance of Codazzi coupling for Hessian manifolds, we move on (Sect. 3) to study a pair of g-conjugate connections that are curvature-free but generally admit torsion. Connections that are curvature-free yet carry torsion have been previously studied and are known as Weitzenböck connections. These connections were originally introduced by Weitzenböck as an alternative for using curvature to study Riemannian geometry. They are connections which preserve a frame, and hence are curvature-free (but not necessarily torsion-free). The now-classic construction uses an orthogonal frame (with respect to a given Riemannian metric) to construct a connection that keeps that frame parallel. As we will see, this connection will be curvature-free, but may admit torsion. To do this construction globally, the manifold must be parallelizable, so there is a topological obstruction to the global existence of such a connection. Traditionally, these connections have been studied to understand teleparallelism as an alternative to the Einstein gravity theory [8,11]. There the Weitzenböck connections are metric-connections that in general carry torsion.

In information geometry, one often considers non-metric connections. As such, we will drop the assumption (made in the construction of Weitzenböck connection) that the choice of frame be orthogonal with respect to a given metric, and discuss how this affects the theory. To distinguish this case from the classical case in which the connection is metric, we refer to these connections as "pseudo-Weitzenböck." Given a pseudo-Weitzenböck connection and the associated frame, we can define (i) the g-conjugate connection and (ii) g-biorthogonal frame. It will be shown (Theorem 10 in Sect. 3.3) that the pseudo-Weitzenböck connection adapted to a g-biorthogonal frame is precisely the g-conjugate connection of the pseudo-Weitzenböck connection adapted to the original frame. In this way we successfully construct dually curvature-free manifolds which admit biorthogonal *frames*, a relaxation of the dually flat manifolds which admit biorthogonal *coordinates*.

In Sect. 4, we study g-conjugate connections for which one is flat and the other is pseudo-Weitzenböck. This has been incidentally referred to as "partially flat" [13], in contrast with the "dually flat" (Hessian) situation. The pair of connections, though conjugate with respect to g, are nevertheless *not* Codazzi coupled to it—one connection is torsion-free and the other can carry torsion. In fact, this is an example of statistical manifold admitting torsion or SMAT [12], with a pair of curvature-free connections. This happens when the Riemannian metric is not Hessian. In Sect. 4.1, we construct pseudo-Weitzenböck connection for the manifold of parametric probability models as the conjugate cconnection with respect to the Fisher–Rao metric. More specifically, the parameter x in the parametric statistical model $p(\cdot|x)$ is treated as an affine coordinate for a (locally) flat connection. A frame biorthogonal (with respect to the Fisher–Rao metric) to the x coordinate frame is then constructed. This pair of affine connections, while both curvature-free, are not in general Codazzi coupled to the Fisher–Rao metric because of their difference in torsion. This characterization of the manifold of parametric statistical models is very different from the now-classic formulation [3,6], where the Fisher–Rao metric is accompanied by the family of α-connections that are all torsion-free and that are all Codazzi coupled to the Fisher–Rao metric. To put it in another way, while the classic approach treats the parametric statistical models as a statistical manifold (with torsion-free connections generally containing curvatures), our approach treats these models as a statistical manifold admitting torsion (with vanishing curvatures for the pair of conjugate connections). In the previous case, the manifold becomes Hessian when ∇-connection and ∇^*-connection are dually curvature-free (because they are already torsion-free), whereas in the latter case, the manifold becomes Hessian when the pseudo-Weitzenböck connection becomes torsion-free. As an example, we use univariate normal distributions to illustrate our calculations in Sect. 4.2.

In Sect. 5, we discuss the implication of this shift from torsion-free connections to curvature-free connections in characterizing parametric statistical models, as well as other issues comparing geometry of statistics and geometry of physics.

Throughout the paper, Einstein summation notation over repeated indices is in effect.

2 Affine manifold and Hessian metric

2.1 *g*-Conjugate connection

We first recall some basic definitions and facts in information geometry.

Definition 1 (*g-conjugate connection*) Given any connection ∇ and an arbitrary Riemannian metric g, the g-conjugate connection ∇^* is defined as the (unique) connection that jointly preserves g:

$$Z(g(X, Y)) = g(\nabla_Z X, Y) + g(X, \nabla^*_Z Y), \tag{1}$$

where X, Y, Z are all vector fields on $\mathbb{M}$.

It was shown that the above definition is independent of whether it is with respect to the first or second slot of g, and that the conjugation operation $*$ is involutive: $(\nabla^*)^* = \nabla$. For more details, see the discussions in [16].

Curvature R^∇ and torsion T^∇ of an affine connection ∇ are defined by, respectively,

$$R^\nabla(X, Y)Z = \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]}Z$$

and

$$T^\nabla(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

The curvature and torsion of conjugate connections ∇ and ∇^* may be related as follows.

Lemma 2 *With respect to a pair of conjugate connections* ∇, ∇^* *on any Riemannian manifold* $(\mathbb{M}, g)$*,*

(i) *their curvature tensors* R^{∇}, R^{∇^*} *satisfy*

$$g(R^{\nabla}(Z, W)X, Y) + g(R^{\nabla^*}(Z, W)Y, X) \equiv 0; \tag{2}$$

(ii) *their torsion tensors* T^{∇}, T^{∇^*} *satisfy*

$$g(T^{\nabla^*}(Z, X) - T^{\nabla}(Z, X), Y) \equiv (\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y). \tag{3}$$

Note that torsion relation (3) holds regardless of curvature of ∇, ∇^*, while curvature relation (2) holds regardless of torsions of ∇, ∇^*. Due to their importance, we recapitulate their proofs below.

Proof

$$\begin{aligned}
0 &\equiv Z(Wg(X, Y)) - W(Zg(X, Y)) - [Z, W]g(X, Y) \\
&= Z(g(\nabla_W X, Y)) + Z(g(X, \nabla^*_W Y)) \\
&\quad - W(g(\nabla_Z X, Y)) - W(g(X, \nabla^*_Z Y)) \\
&\quad - g(\nabla_{[Z,W]} X, Y) - g(X, \nabla^*_{[Z,W]} Y) \\
&= g(\nabla_Z \nabla_W X, Y) + g(\nabla_W X, \nabla^*_Z Y) \\
&\quad + g(\nabla_Z X, \nabla^*_W Y) + g(X, \nabla^*_Z \nabla^*_W Y) \\
&\quad - g(\nabla_W \nabla_Z X, Y) - g(\nabla_Z X, \nabla^*_W Y) \\
&\quad - g(\nabla_W X, \nabla^*_Z Y) - g(X, \nabla^*_W \nabla^*_Z Y) \\
&\quad - g(\nabla_{[Z,W]} X, Y) - g(X, \nabla^*_{[Z,W]} Y) \\
&= g(\nabla_Z \nabla_W X, Y) + g(X, \nabla^*_Z \nabla^*_W Y) \\
&\quad - g(\nabla_W \nabla_Z X, Y) - g(X, \nabla^*_W \nabla^*_Z Y) \\
&\quad - g(\nabla_{[Z,W]} X, Y) - g(X, \nabla^*_{[Z,W]} Y) \\
&= g(R^{\nabla}(Z, W)X, Y) + g(R^{\nabla^*}(Z, W)Y, X).
\end{aligned}$$

Here, the first line is due to the definition of Lie bracket $[\cdot, \cdot]$, and last line due to the definition of curvature tensor R. Also,

$$\begin{aligned}
(\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y) &= Zg(X, Y) - g(\nabla_Z X, Y) - g(X, \nabla_Z Y) \\
&\quad - (Xg(Z, Y) - g(\nabla_X Z, Y) - g(Z, \nabla_X Y)) \\
&= g(\nabla^*_Z X, Y) - g(\nabla_Z X, Y) \\
&\quad - (g(\nabla^*_X Z, Y) - g(\nabla_X Z, Y)) \\
&= g(\nabla^*_Z X - \nabla^*_X Z, Y) - g(\nabla_Z X - \nabla_X Z, Y) \\
&= g(T^{\nabla^*}(Z, X) - T^{\nabla}(Z, X), Y).
\end{aligned}$$

Here, the second equality used the definition of conjugate connection, and the last equality is due to cancelation of $[Z, X]$ term in $T^{\nabla^*}(Z, X) - T^{\nabla}(Z, X)$. □

A consequence of Eq. (2) is that if ∇ is curvature-free, then so is ∇^*. Likewise, the consequence of Eq. (3) is that ∇ and ∇^* carry the same amount of torsion if and only if

$$(\nabla_Z g)(X, Y) = (\nabla_X g)(Z, Y),$$

which is known as the "Codazzi coupling" of (g, ∇). Both (2) and (3) are well-known facts in information geometry. It is easily verified that (g, ∇) is Codazzi coupled if and only if (g, ∇^*) is Codazzi coupled. A connection is called *flat* when it is both curvature-free and torsion-free. A manifold is called *dually flat* when it carries two flat connections ∇ and ∇^* that form a conjugate pair with respect to the Hessian metric constructed from either ∇ or ∇^* (see below).

2.2 Hessian metric

Let us now recall the Hessian operator (second derivative) on a function Φ defined on a manifold, defined as the covariant derivative of the 1-form $d\Phi$,

$$Hess_\nabla(\Phi)(X, Y) = (\nabla_X d\Phi)(Y)$$

or explicitly,

$$Hess_\nabla(\Phi)(X, Y) \equiv XY\Phi - d\Phi(\nabla_X Y).$$

Note that $Hess$ is defined only after ∇ is specified. It is a bilinear form of X, Y—linearity in X is easy to shown, while linearity in Y needs verification, by invoking Leibniz rule of derivative. $Hess_\nabla(\Phi)(X, Y)$ is sometimes also written as $(\nabla d\Phi)(X, Y)$, to emphasize the bilinearity of the operator $Hess_\nabla(\Phi) \equiv \nabla d\Phi$. Note that in general

$$Hess_\nabla(\Phi)(X, Y) \neq Hess_\nabla(\Phi)(Y, X).$$

In fact, their difference

$$\begin{aligned} Hess_\nabla(\Phi)(X, Y) - Hess_\nabla(\Phi)(Y, X) &= [X, Y]\Phi - d\Phi(\nabla_X Y - \nabla_Y X) \\ &= d\Phi([X, Y] - \nabla_X Y + \nabla_Y X) \\ &= d\Phi(T^\nabla(Y, X)). \end{aligned}$$

As such, $Hess_\nabla(\Phi)(X, Y)$ is symmetric in X, Y for all functions Φ if and only if ∇ is torsion-free.

We now impose this torsion-free condition on ∇ and introduce the notion of a "Hessian metric", namely a Riemannian metric g such that

$$g(X, Y) = Hess_\nabla(\Phi)(X, Y) \equiv XY\Phi - d\Phi(\nabla_X Y).$$

Positive-definiteness of g imposes further requirements on the function Φ, such as convexity (see the end of Sect. 2.3 for a discussion of the notion of "convexity" of a function on a manifold). In local coordinates ($X = \partial_{x^i}$, $Y = \partial_{x^j}$) it takes the form

$$Hess_\nabla(\Phi)(\partial_{x^i}, \partial_{x^j}) = \partial_{x^i}\partial_{x^j}\Phi - \Gamma^k_{ij}\partial_{x^k}\Phi,$$

where $\nabla_{\partial_{x^i}}\partial_{x^j} \equiv \Gamma^k_{ij}\partial_{x^k}$. Torsion-freeness of ∇ is reflected as $\Gamma^k_{ij} = \Gamma^k_{ji}$.

Given a Riemannian metric g, we can ask whether there are torsion-free connections ∇ so that

$$g(X, Y) = Hess_\nabla(\Phi)(X, Y).$$

There are no geometric obstructions to the existence of such connections (and in fact, such connections exist in abundance for any metric). However, we can ask under what circumstances we can force g to be both $Hess_\nabla$ as well as $Hess_{\nabla^*}$, where ∇^* is conjugate to ∇ with respect to $g = Hess_\nabla$. To answer this question, let us examine the role of curvature of ∇ for the existence of a Hessian metric.

Proposition 3 *For a torsion-free ∇ and an arbitrary smooth function Φ, then $(\nabla, Hess_\nabla(\Phi))$ is Codazzi coupled, that is,*

$$(\nabla_Z Hess_\nabla(\Phi))(X, Y) = (\nabla_X Hess_\nabla(\Phi))(Z, Y),$$

if and only if $d\Phi(R^\nabla) = 0$.

Proof Denote

$$\begin{aligned} C(X, Y, Z) &\equiv (\nabla_Z Hess_\nabla(\Phi))(X, Y) \\ &\equiv Z(Hess_\nabla(\Phi)(X, Y)) - Hess_\nabla(\Phi)(\nabla_Z X, Y) - Hess_\nabla(\Phi)(X, \nabla_Z Y). \end{aligned}$$

Using the definition of $Hess_\nabla(\Phi)$, we write out:

$$\begin{aligned} Z(Hess_\nabla(\Phi)(X, Y)) &= Z(XY\Phi) - Z(d\Phi(\nabla_X Y)) \\ &= ZX(Y\Phi) - (d\Phi(\nabla_Z(\nabla_X Y)) + (\nabla_Z d\Phi)(\nabla_X Y)) \\ &= ZX(Y\Phi) - d\Phi(\nabla_Z(\nabla_X Y)) - Hess_\nabla(\Phi)(Z, \nabla_X Y); \\ Hess_\nabla(\Phi)(\nabla_Z X, Y) &= (\nabla_Z X)(Y\Phi) - d\Phi(\nabla_{\nabla_Z X} Y). \end{aligned}$$

Therefore,

$$\begin{aligned} C(X, Y, Z) = (ZX - \nabla_Z X)(Y\Phi) &- d\Phi\left(\nabla_Z(\nabla_X Y) - \nabla_{\nabla_Z X} Y\right) \\ &- Hess_\nabla(\Phi)(Z, \nabla_X Y) - Hess_\nabla(\Phi)(X, \nabla_Z Y). \end{aligned}$$

Exchange X and Z to obtain a similar expression, and then calculate $\Delta \equiv C(X, Y, Z) - C(Z, Y, X)$, we obtain

$$\begin{aligned}\Delta &= T^{\nabla}(X,Z)(Y\Phi) + d\Phi(R^{\nabla}(X,Z)Y + \nabla_{T^{\nabla}(Z,X)}Y)\\ &\quad - Hess_{\nabla}(\Phi)(Z,\nabla_X Y) - Hess_{\nabla}(\Phi)(X,\nabla_Z Y)\\ &\quad + Hess_{\nabla}(\Phi)(X,\nabla_Z Y) + Hess_{\nabla}(\Phi)(Z,\nabla_X Y)\\ &= d\Phi(R^{\nabla}(X,Z)Y) + T^{\nabla}(X,Z)(d\Phi(Y)) - d\Phi(\nabla_{T^{\nabla}(X,Z)}Y).\end{aligned}$$

Note that the above relationship holds for an arbitrary connection ∇ that may have non-vanishing curvature and torsion. When $T^{\nabla} = 0$ (and hence $Hess_{\nabla}(\Phi)$ is symmetric), then we obtain that $\Delta = 0$, or $C(X,Y,Z) = C(Z,Y,X)$, if and only if $d\Phi(R^{\nabla}) = 0$. Since $C(X,Y,Z) = C(Z,Y,X)$ is, by definition, Codazzi coupling of ∇ with $Hess_{\nabla}(\Phi)$, the Proposition is proved. ◻

An immediate corollary of Proposition 3 is

Corollary 4 *Any flat connection ∇ on a manifold $\mathbb{M}$ is Codazzi coupled to $Hess_{\nabla}(\Phi)$ for any smooth Φ on $\mathbb{M}$.*

These formulae show that care must be taken in order to meaningfully generalize Hessian metrics based on non-flat connections. If one only requires that the metric be Hessian with respect to a single torsion-free connection, then for any metric one can find such connections in spades. However, if one also requires that such a metric be Codazzi coupled to the conjugate connections (which is necessary for $g = Hess_{\nabla^*}\Phi^*$, see below), then the curvature of ∇ must vanish in the direction of $d\Phi$.

2.3 Manifold with flat connections

Starting from a flat connection and associated affine coordinates, there are infinitely many associated local Hessian metrics, each of which is determined by the choice of the smooth convex potential function expressed in those coordinates.

Now suppose a manifold $\mathbb{M}$ is equipped with two flat connections ∇ and ∇^*, with associated affine coordinates x and u respectively (i.e., they respectively render ∇ and ∇^* to have vanishing coefficients). Then we may construct two local Hessian metrics $g^{\Phi} = Hess_{\nabla}(\Phi)$ and $g^{\Phi^*} = Hess_{\nabla^*}(\Phi^*)$ using two smooth functions Φ, Φ^* that are convex under x and u, respectively. Here,

$$g^{\Phi}(\partial_{x^i},\partial_{x^j}) = \frac{\partial^2\Phi}{\partial x^i\partial x^j}, \quad g^{\Phi^*}(\partial_{u^i},\partial_{u^j}) = \frac{\partial^2\Phi^*}{\partial u^i\partial u^j}.$$

Proposition 5 *The following statements are equivalent:*

(i) *the two convex functions Φ and Φ^* form a Legendre pair:*

$$u^i = \frac{\partial\Phi(x)}{\partial x^i} \quad and \quad x^i = \frac{\partial\Phi^*(u)}{\partial u^i};$$

(ii) *the two connections ∇ and ∇^* are conjugate with respect to both $Hess_{\nabla}(\Phi)$ and $Hess_{\nabla^*}(\Phi^*)$, in which case the two Hessian metrics g^{Φ} and g^{Φ^*} are the same (and denoted by g):*

$$g = Hess_{\nabla}(\Phi) = Hess_{\nabla^*}(\Phi^*);$$

(iii) *x and u form biorthogonal coordinates with respect to either g^{Φ} or g^{Φ^*}:*

$$g^{\Phi}(\partial_{x^i}, \partial_{u^j}) = g^{\Phi^*}(\partial_{x^i}, \partial_{u^j}) = \delta_{ij} \ ;$$

(iv) *the connection $\frac{1}{2}(\nabla + \nabla^*)$ is the Levi-Civita connection for both g^{Φ} and g^{Φ^*}:*

$$\frac{1}{2}(\nabla + \nabla^*)g^{\Phi} = \frac{1}{2}(\nabla + \nabla^*)g^{\Phi^*} = 0.$$

So while any flat connection can be associated with infinitely many Hessian metrics, each indexed by a Φ, two flat connections ∇, ∇^* (and hence two smooth convex functions Φ, Φ^*) are "related" to each other if they are required to be conjugate with respect to both Hessian metrics they induce; in such case their associated Hessian metric is one and the same, and the two convex functions are conjugate to each other. It is also the metric with respect to which the averaged connection $\frac{1}{2}(\nabla + \nabla^*)$ is the Levi-Civita connection. In this case we speak of ∇, ∇^* as (a pair of) "dually flat" connections.

A remark about convexity of functions on a manifold $\mathbb{M}$ is in order. A function on $\mathbb{M}$ is convex with respect to a given connection, or more generally with respect to some family of curves. Though a Hessian metric can be defined without using any specific coordinate systems, and transforms covariantly as a (0,2)-tensor under coordinate change, convexity of a function depends on the choice of coordinates (in terms of the accompanying affine connection). As a result, $Hess_{\nabla}(\Phi^*)$, which may be distinct from $Hess_{\nabla^*}(\Phi)$, is not necessarily positive definite, so may not define a metric. Suppose $\Phi(x)$ is convex in x, $\Phi^*(u)$ is convex in u, but $\tilde{\Phi}(x) = \Phi^*(u(x)) = x^k \frac{\partial \Phi}{\partial x^k} - \Phi(x)$ may not be convex in x. To see this,

$$\begin{aligned} Hess_{\nabla}(\tilde{\Phi})(\partial_{x^i}, \partial_{x^j}) &= \frac{\partial^2}{\partial x^i \partial x^j}\left(x^k \frac{\partial \Phi}{\partial x^k} - \Phi(x)\right) \\ &= \frac{\partial}{\partial x^i}\left(x^k \frac{\partial^2 \Phi}{\partial x^j \partial x^k}\right) \\ &= \frac{\partial^2 \Phi}{\partial x^i \partial x^j} + x^k \frac{\partial^3 \Phi}{\partial x^k \partial x^i \partial x^j}. \end{aligned}$$

Clearly, positive-definiteness is not guaranteed.

3 Pseudo-Weitzenböck connections

3.1 Connection adapted to a frame

Let us start by defining a local frame on a smooth manifold $\mathbb{M}$, with $n = \dim(\mathbb{M})$. A frame $\mathfrak{B} = \{\mathrm{b}_1, \ldots, \mathrm{b}_n\}$ is a collection of n local linearly independent vector fields $\{\mathtt{b}_i\}_{i=1}^n$ on $\mathbb{M}$. Assume that $\mathbb{M}$ is prescribed with a local coordinate system $x = \{x^i\}_{i=1}^n$, then the local expression of a frame $\mathfrak{B} = \{\mathtt{b}_i\}_{i=1}^n$ using this coordinate system is $\mathtt{b}_i = B_i^j \partial_{x^j}$, where B_i^j is an $n \times n$ matrix, assumed to be of full rank, and hence invertible. Below, B^{-1} denotes the matrix inverse of B:

$$(B^{-1})_l^i \, B_j^l = \delta_j^i = B_l^i \, (B^{-1})_j^l.$$

A frame is said to be global when it is defined on the entire manifold $\mathbb{M}$. As mentioned before, this implies that $\mathbb{M}$ is parallelizable, so is subject to topological obstructions.

Suppose B-matrix is the Jacobian matrix of coordinate transform: $x \longrightarrow y$

$$(B^{-1})_j^l = \frac{\partial y^l}{\partial x^j} \longleftrightarrow B_l^j = \frac{\partial x^j}{\partial y^l}. \tag{4}$$

In this case, then

$$\mathtt{b}_i = \frac{\partial x^l}{\partial y^i} \frac{\partial}{\partial x^l} = \frac{\partial}{\partial y^i} := \partial_{y^i}.$$

So we say that the frame $\{\mathtt{b}_i\}_{i=1}^n$ forms a coordinate frame.

The necessary and sufficient condition for (4) is

$$\partial_{x^i} (B^{-1})_j^l = \partial_{x^j} (B^{-1})_i^l. \tag{5}$$

Necessity is obvious. As for sufficiency, note that when the above is satisfied, then for each l there exists a function $y^l = y^l(x)$ such that

$$(B^{-1})_j^l = \frac{\partial y^l}{\partial x^j}.$$

Definition 6 (*Adapted connection*) An affine connection ∇ is said to be *adapted* to a frame $\mathfrak{B} = \{\mathtt{b}_1, \ldots, \mathtt{b}_n\}$ if $\nabla_{\mathtt{b}_i} \mathtt{b}_j = 0$ holds $\forall i, j$.

We can now construct the adapted connection $\nabla^{\mathfrak{B}}$ explicitly.

Proposition 7 *Given a frame $\mathfrak{B}$, construct $\nabla^{\mathfrak{B}} = B\, \partial(B^{-1})$, or in component forms:*

$$\Gamma_{ik}^j = B_l^j (\partial_{x^i} (B^{-1})_k^l) = -(B^{-1})_k^l (\partial_{x^i} B_l^j). \tag{6}$$

Then

$$\nabla_{\mathtt{b}_i}^{\mathfrak{B}} \mathtt{b}_j \equiv 0.$$

Proof

$$\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j = B_i^l \, \nabla_{\partial_{x^l}} \left(B_j^k \partial_{x^k} \right)$$
$$= B_i^l \left(\left(\partial_{x^l} B_j^k \right) \partial_{x^k} + B_j^k \, \Gamma^s_{lk} \partial_{x^s} \right).$$

Substituting Eq. (6) for the expression of Γ, we find that the terms in the parenthesis vanish, so

$$\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j = 0.$$

□

We call $\nabla^{\mathfrak{B}}$ as given by Eq. (6) the "pseudo-Weitzenböck" connection adapted to the frame $\mathfrak{B}$. Since we did not assume that the frame $\mathfrak{B}$ to be orthonormal, we cannot assume any nice form for B^{-1}. It is worth noting that the construction of the pseudo-Weitzenböck connections is identical to that of a Weitzenböck connection. For clarity, we reserve the term Weitzenböck connection for the case where $\mathfrak{B} = \mathfrak{B}^\star$, (see Definition 9 and Sect. 5) where the frame is orthonormal with respect to a given metric g. Historically, this assumption is often made, and we wish to drop the assumption of metricity to apply this concept to information geometry.

Note that $\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j = 0$ means that $\mathfrak{B}$ is parallel along the connection. Pseudo-Weitzenböck connections are not unique. In fact, any choice of frame yields a pseudo-Weitzenböck connection; given two frames $\mathfrak{B}$ and $\hat{\mathfrak{B}}$, $\nabla^{\mathfrak{B}} = \nabla^{\hat{\mathfrak{B}}}$ iff there is an element of $C \in GL(n)$ so that $B \equiv C\hat{B}$. Note that this is fairly restrictive, as C must be constant on $\mathbb{M}$, whereas $\mathfrak{B}$ and $\hat{\mathfrak{B}}$ vary throughout $\mathbb{M}$.

It is worth noting that any curvature-free connection can be locally written as a pseudo-Weitzenböck connection with respect to some frame. To see this, pick a point $p \in \mathbb{M}$ and a frame $\mathfrak{B}_p$ at tangent space $T_p\mathbb{M}$. If we define $\mathfrak{B}$ in a neighborhood of p by parallel transport of $\mathfrak{B}_p$, then this frame is parallel under the connection, and gives a local trivializing frame of the connection. This construction is restricted to a locally trivializing neighborhood of the tangent bundle, so will not extend globally unless the tangent bundle is trivial. Therefore, in order to construct a pseudo-Weitzenböck connection globally, one must consider parallelizable Riemannian manifolds.

3.2 Curvature and torsion of adapted connection

The coefficient expression (Christoffel symbol) Γ of an affine connection ∇ is coordinate-dependent, and does not transform tensorially. Torsion and curvature of ∇, however, are tensors. In this subsection, we study curvature and torsion of a connection $\nabla^{\mathfrak{B}}$ adapted to a frame $\mathfrak{B}$, and the implications of $\mathfrak{B}$ being a coordinate frame.

3.2.1 Torsion

An adapted connection may have non-vanishing torsion. In fact, we have the following result.

Proposition 8 *The torsion of an adapted connection* $\nabla^{\mathfrak{B}}$ *vanishes if and only if* $\mathfrak{B}$ *is a coordinate frame, that is* $[\mathtt{b}_i, \mathtt{b}_j] = 0$.

Proof We know that ∇ is torsion-free iff $\Gamma^j_{ik} = \Gamma^j_{ki}$. From (6), this condition translates into

$$(B^{-1})^l_k(\partial_{x^i} B^j_l) = (B^{-1})^l_i(\partial_{x^k} B^j_l),$$

or

$$B^k_i(\partial_{x^k} B^j_l) = B^k_l(\partial_{x^k} B^j_i).$$

That is

$$\mathtt{b}_i\mathtt{b}_l = \mathtt{b}_l\mathtt{b}_i$$

or $[\mathtt{b}_i, \mathtt{b}_l] = 0$.

The torsion-free condition also translates to

$$-B^j_l(\partial_{x^i}(B^{-1})^l_k) = -B^j_l(\partial_{x^k}(B^{-1})^l_i)$$

which, after canceling out $-B^j_l$ on both sides, reduces to Eq. (5). □

We also provide an explicit calculation to show that $\nabla^{\mathfrak{B}}$ is torsion-free iff $[\mathtt{b}_i, \mathtt{b}_j] = 0$. This involves writing out explicitly the torsion tensor of an adapted connection,

$$T(\mathtt{b}_i, \mathtt{b}_j) = \nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j - \nabla^{\mathfrak{B}}_{\mathtt{b}_j}\mathtt{b}_i - [\mathtt{b}_i, \mathtt{b}_j].$$

The first two terms automatically vanish, leaving the following

$$T(\mathtt{b}_i, \mathtt{b}_j) = -[\mathtt{b}_i, \mathtt{b}_j].$$

When non-vanishing, torsion tensor can be expressed $T(\partial_{x^i}, \partial_{x^j}) = T^k_{ij}\partial_{x^k}$ for some coefficients T^k_{ij}. Therefore, we have

$$T(\mathtt{b}_i, \mathtt{b}_j) = B^l_i\, B^k_j\, (B^{-1})^m_s\, T^s_{lk}\, \mathtt{b}_m.$$

3.2.2 Curvature

Since $\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_j \equiv 0$, any adapted connection always has zero-curvature, as we now show.

We invoke the curvature formula:

$$\begin{aligned} R(\mathtt{b}_i, \mathtt{b}_j)\mathtt{b}_k &= \nabla^{\mathfrak{B}}_{\mathtt{b}_i}(\nabla^{\mathfrak{B}}_{\mathtt{b}_j}\mathtt{b}_k) - \nabla^{\mathfrak{B}}_{\mathtt{b}_j}(\nabla^{\mathfrak{B}}_{\mathtt{b}_i}\mathtt{b}_k) - \nabla^{\mathfrak{B}}_{[\mathtt{b}_i, \mathtt{b}_j]}\mathtt{b}_k \\ &= \nabla^{\mathfrak{B}}_{\mathtt{b}_i}0 - \nabla^{\mathfrak{B}}_{\mathtt{b}_j}0 + \nabla^{\mathfrak{B}}_{T(\mathtt{b}_i, \mathtt{b}_j)}\mathtt{b}_k \end{aligned}$$

$$\begin{aligned} &= B_i^l\, B_j^n\, (B^{-1})_s^m\, T_{ln}^s \nabla_{\mathrm{b}_m}^{\mathfrak{B}} \mathrm{b}_k \\ &= 0. \end{aligned}$$

We remark that trying to do this calculation with Christoffel symbols becomes tedious; the reason is that it is not obvious when to extract the torsion terms to get cancellation. As a result, it is much more productive to do the calculation using the fact that $\mathfrak{B}$ is parallel along the connection.

3.3 Biorthogonal frames and conjugate connections

On a Riemannian manifold $(\mathbb{M}, g)$, one can define "orthonormal frame" as a frame $\mathfrak{B}$ satisfying $g(\mathrm{b}_i, \mathrm{b}_j) = \delta_{ij}$. As a generalization, given any frame $\mathfrak{B}$, there is the notion of "biorthogonal frame" $\mathfrak{B}^\star$ of $\mathfrak{B}$ with respect to an arbitrarily given Riemannian metric g.

Definition 9 (*g-biorthogonal frame*) Given any frame $\mathfrak{B} = \{\mathrm{b}_i\}_{i=1}^n$, the g-biorthogonal frame is defined as the (unique) frame $\mathfrak{B}^\star = \{\mathrm{b}_i^\star\}_{i=1}^n$ that is biorthogonal with respect to the given g:

$$g(\mathrm{b}_i, \mathrm{b}_j^\star) \equiv \delta_{ij}.$$

Writing $\mathfrak{B}^*$ in the same local coordinate system as B, the vectors $b_i^\star$ are

$$\mathrm{b}_i^\star = (B^\star)_i^k \partial_{x^k}.$$

Note that the biorthogonality condition is simply

$$\delta_{ij} = g\left(B_i^l \partial_{x^l}, (B^\star)_j^k \partial_{x^k}\right) = B_i^l\, (B^\star)_j^k\, g(\partial_{x^l}, \partial_{x^k}) = B_i^l\, (B^\star)_j^k\, g_{lk}.$$

Therefore, given g, the uniquely-defined $B^\star$-matrix biorthogonal to the B-matrix is

$$(B^\star)_j^k = (B^{-1})_l^i\, g^{lk}\, \delta_{ij},$$

where g^{lk} is the matrix inverse of $g_{ij} = g(\partial_{x^i}, \partial_{x^j})$. Here the Kronecker delta δ_{ij} remains in the expression because according to our notation B^{-1} (and B) must maintain one upper index and one lower index.

Given $\mathfrak{B}$ and an arbitrary g, we can also define a pseudo-Weitzenböck connection on $\mathfrak{B}^\star$ as well. So we will get a pair of adapted connections $\nabla^{\mathfrak{B}}$ and $\nabla^{\mathfrak{B}^\star}$.

Recall that given any connection ∇ and an arbitrary g, we can also define the g-conjugate connection ∇^* so as to jointly preserve g according to Eq. (1). The question now arises as whether these two notions are compatible, i.e., $\left(\nabla^{\mathfrak{B}}\right)^* = \nabla^{(\mathfrak{B}^\star)}$. In other words, we would like to know whether the adapted connection $\nabla^{\mathfrak{B}^\star}$ of the g-biorthogonal frame $\mathfrak{B}^\star$ is g-conjugate to the connection $\nabla^{\mathfrak{B}}$ adapted to $\mathfrak{B}$. The next Theorem gives a positive answer.

Theorem 10 *With respect to any Riemannian metric g, the g-conjugation of a connection adapted to a frame $\mathfrak{B}$ equals the connection adapted to the g-biorthogonal frame $\mathfrak{B}^\star$:*

$$\left(\nabla^{\mathfrak{B}}\right)^* = \nabla^{(\mathfrak{B}^\star)}.$$

Proof From

$$\delta_{ij} = g_{lk}\, B_i^l\, (B^\star)_j^k,$$

differentiating both sides with respect to x^s, we find that

$$\begin{aligned} 0 &= \partial_{x^s}\left(g_{lk}\, B_i^l\, (B^\star)_j^k\right) \\ &= (\partial_{x^s} g_{lk})\, B_i^l\, (B^\star)_j^k + g_{lk}\, (\partial_{x^s} B_i^l)\, (B^\star)_j^k + g_{lk}\, B_i^l (\partial_{x^s}\, (B^\star)_j^k). \end{aligned}$$

Using the definition of the connection coefficients,

$$0 = (\partial_{x^s} g_{lk})\, B_i^l\, (B^\star)_j^k - g_{lk}\, (B_i^r \Gamma_{sr}^l)\, (B^\star)_j^k - g_{lk}\, B_i^l\, ((B^\star)_j^r \Gamma^{\star k}_{sr}).$$

Multiplying through by the $(B^{-1})_m^i((B^\star)^{-1})_n^j$ and relabeling the indices, we obtain

$$\partial_{x^k} g_{ij} = g_{lj}\Gamma_{ki}^l + g_{li}(\Gamma^\star)_{kj}^l.$$

This means that the connection $\nabla^{\mathfrak{B}}$ and $\nabla^{\mathfrak{B}^\star}$ are g-conjugate to each other. □

In the current setting, when ∇ takes the form of an adapted connection to some frame $\mathfrak{B}$, we have $\nabla^{\mathfrak{B}}_{\mathrm{b}_i}\mathrm{b}_j = 0$, so

$$Hess_{\nabla^{\mathfrak{B}}}(\Phi)(\mathrm{b}_i, \mathrm{b}_j) \;=\; \mathrm{b}_i\mathrm{b}_j\Phi \;\neq\; \mathrm{b}_j\mathrm{b}_i\Phi \;=\; Hess_{\nabla^{\mathfrak{B}}}(\Phi)(\mathrm{b}_j, \mathrm{b}_i).$$

Here $\mathrm{b}_i\Phi \equiv d\Phi(\mathrm{b}_i)$ is the directional or first derivative of a function, which itself is another funciton on the manifold.

3.4 Exchanging curvature for torsion

Given a flat connection ∇^* on a Riemannian manifold $(\mathbb{M}, g)$, our previous discussion shows that the g-conjugate connection ∇ is curvature-free, but will have torsion in general. If we wish to deal with torsion-free connections without making the assumption of Codazzi coupling, then we can symmetrize ∇.

To do so, we recall the fact that for any affine connection ∇, the new connection $\widetilde{\nabla}$ given by

$$\widetilde{\nabla} = \nabla - \frac{1}{2}\, T^{\nabla},$$

or explicitly

$$\widetilde{\nabla}_X Y = \nabla_X Y - \frac{1}{2} T^{\nabla}(X, Y),$$

is necessarily torsion-free, but will admit curvature. Because $\widetilde{\nabla}$ is torsion-free whereas ∇ is curvature-free, we have "exchanged" the torsion of ∇ for the curvature of $\widetilde{\nabla}$, and vice-versa. It is worth emphasizing that ∇ and $\widetilde{\nabla}$ have the same geodesics. To see this, note that in a coordinate chart $\{x^i\}$, the geodesic equations for any affine connection (whether torsion-carrying or not), namely,

$$\frac{d^2 x^i}{ds^2} + \Gamma^i_{jk} \frac{dx^j}{ds} \frac{dx^k}{ds} = 0$$

are symmetric in j and k. As such, any antisymmetry in Γ^i_{jk} will be eliminated after summation. In concrete terms, this means that if two connections differ only by torsion, then their associated geodesic equations are the same, and hence they have the same solution and identical geodesic curves. Since ∇ and $\widetilde{\nabla}$ differ only by torsion, this is exactly the situation at work here. This understanding of "exchanging torsion for curvature" may be helpful to understanding "partially flat" geometry [13]. For instance, we can apply results using curvature to understand the geometry associated with $\widetilde{\nabla}$. Doing so, we can obtain results about ∇, since the two connections have the same geodesics.

When ∇ and ∇^* are a pair of g-conjugate, curvature-free connections, their averaged connection

$$\nabla^{(0)} = \frac{1}{2}(\nabla + \nabla^*),$$

even if torsion-carrying, is necessarily parallel to g, $\nabla^{(0)} g = 0$. Its torsion $T^{(0)}$ is

$$T^{(0)} = \frac{1}{2}(T^{\nabla} + T^{\nabla^*}).$$

We consider the symmetrization of $\nabla^{(0)}$ as

$$\begin{aligned}
\widetilde{\nabla^{(0)}} &= \nabla^{(0)} - \frac{1}{2} T^{(0)} \\
&= \frac{1}{2}(\nabla + \nabla^*) - \frac{1}{4}(T^{\nabla} + T^{\nabla^*}) \\
&= \frac{1}{2}\left(\nabla - \frac{1}{2} T^{\nabla}\right) + \frac{1}{2}\nabla^* \\
&= \frac{1}{2}(\widetilde{\nabla} + \nabla^*),
\end{aligned}$$

where the last-but-one step used the fact $T^{\nabla^*} = 0$ by our assumption. Note that though $\widetilde{\nabla^{(0)}}$ is torsion-free now (as is $\widetilde{\nabla}$), it is *not* the Levi-Civita connection ∇^{LC} on $\mathbb{M}$ (the unique connection that it both torsion-free and parallel to metric g), because $\widetilde{\nabla^{(0)}} g \neq 0$ in general. To obtain the Levi-Civita connection ∇^{LC} from $\nabla^{(0)}$, one needs to substract the *contorsion tensor* K from $\nabla^{(0)}$:

$$\nabla^{LC} = \nabla^{(0)} - K,$$

where

$$K_{ij}^{l} = \frac{1}{2}\left(T_{ij}^{(0)l} - g_{mj}g^{kl}T_{ik}^{(0)m} + g_{mi}g^{kl}T_{kj}^{(0)m}\right). \tag{7}$$

Contorsion tensor K and torsion tensor $T^{(0)}$ are related by

$$T_{ij}^{(0)l} = K_{ij}^{l} - K_{ji}^{l}$$

and $K_{ij}^{l} \neq -K_{ji}^{l}$ in general. Introducing the covariant version $C_{ijk} \equiv g_{lk}K_{ij}^{l}$. Then $C_{ijk} = -C_{ikj}$, though $C_{ijk} \neq -C_{jik}$ in general. The condition $C_{ijk} = -C_{jik}$ is satisfied if and only if $K_{ij}^{l} = \frac{1}{2}T_{ij}^{(0)l}$; equivalently the covariant contorsion tensor C (and hence the covariant torsion) is totally skew-symmetric.

4 Parametric statistical models: a new characterization

In this section, we provide a new presentation of Information Geometry of parametric statistical models.

4.1 Constructing the pseudo-Weitzenböck connection

Recall the standard setting where a parametric family of density functions, $p(\cdot|x)$, called a parametric statistical model, is the association $x \mapsto p(\cdot|x)$ of a point $x = [x^1, \ldots, x^n]$, so here x is in a connected open subset of $\mathbb{R}^n$ and serves as a local coordinate chart of $\mathbb{M}$. The Fisher–Rao metric and the α-connections are given by

$$g_{ij}(x) = \int_{\Omega} d\omega \left\{ p(\omega|x) \frac{\partial \log p(\omega|x)}{\partial x^i} \frac{\partial \log p(\omega|x)}{\partial x^j} \right\};$$

$$\Gamma_{ij,k}^{(\alpha)}(x) = \int_{\Omega} d\omega \frac{\partial p(\omega|x)}{\partial x^k} \left(\frac{1-\alpha}{2} \frac{\partial \log p(\omega|x)}{\partial x^i} \frac{\partial \log p(\omega|x)}{\partial x^j} + \frac{\partial^2 \log p(\omega|x)}{\partial x^i \partial x^j} \right).$$

The α-connection and $(-\alpha)$-connection are conjugate to each other with respect to the Fisher–Rao metric g. Note that all α-connections are torsion-free; yet generally they have non-zero curvatures, with curvature of $(\pm\alpha)$-connections equal but opposite sign of each other, see Lemma 2. When the curvatures of (± 1)-connections vanish, g takes the form of a Hessian metric. It is important to keep in mind that each member of the α-connection is Codazzi coupled to the Fisher–Rao metric g.

Now let us take a different perspective about the manifold $\mathbb{M}$ of parametric statistical models $p(\cdot|x)$. We take x to be a local coordinate system, and associate a flat connection ∇^* to it. We continue to take Fisher–Rao metric g as the Riemannian metric on $\mathbb{M}$. Denote by ∇ the g-conjugate of our flat connection ∇^*. Because ∇ is necessarily curvature-free, our results in Sect. 2 tell us that ∇ is flat if and only if g is Codazzi coupled to ∇^*. In general ∇ will not be torsion-free; ∇ is a pseudo-Wietzenbock connection. This has been referred to as a "partially flat" manifold [13], which admits a pair of curvature-free g-conjugate connections, one of which is torsion-free as well. Partially flat structure is a slightest relaxation to the dually flat Hessian structure, by allowing one of the connections, say ∇, to admit torsion. The metric g may still be Hessian, but the flat connection (∇^*) is no longer Codazzi coupled to g. In the literature, a manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ for which ∇^* is torsion-free is called a "statistical manifold admitting torsion" or SMAT [12], and ∇ and g are coupled by the following relation:

$$(\nabla_Z g)(X, Y) - (\nabla_X g)(Z, Y) = g(T^{\nabla}(X, Z), Y).$$

Below, we actually describe parametric statistical model as SMAT by constructing *the* biorthogonal frame $\mathfrak{B}$ based on g being the Fisher–Rao metric.

Given a parametric statistical model $p(\cdot|x)$, let us treat its parameter x as the affine coordinates for the flat connection ∇^*, i.e., the Christoffel symbol Γ^{*i}_{jk} vanishes. Writing out the equation of conjugate connections ∇, ∇^* under this coordinate chart

$$\frac{\partial g_{ij}}{\partial x^k} = g_{lj}\Gamma^l_{ki} + g_{il}\Gamma^{*l}_{kj} = g_{lj}\Gamma^l_{ki}\,.$$

Therefore, the pseudo-Weitzenböck connection ∇ of the parametric statistical model is

$$\Gamma^j_{ki} = g^{jl}\frac{\partial g_{il}}{\partial x^k}\,, \tag{8}$$

with g^{ij} denoting the elements of the matrix inverse of g, the Fisher–Rao metric. It can be readily verified that such connection is always curvature-free, but carries torsion

$$T^j_{ik} = g^{jl}\left(\frac{\partial g_{kl}}{\partial x^i} - \frac{\partial g_{il}}{\partial x^k}\right).$$

In general, $T \neq 0$, unless $\frac{\partial g_{il}}{\partial x^k}$ is totally symmetric, i.e., g is Hessian.

The contorsion tensor K evaluated from (7) turns out to be, for this case,

$$K^l_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^j}\right).$$

Note that in the above calculation, $\Gamma^{(0)} = \frac{1}{2}\Gamma$, so $T^{(0)} = \frac{1}{2}T$, because $T^* = 0$. The quantity

$$\Gamma^{(0)l}_{ij} - K^l_{ij} = \frac{1}{2}g^{lk}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right)$$

is precisely the Levi-Civita connection of g.

Even though Γ^j_{ki} given by (8) may carry torsion, we still compute its geodesic equation

$$\frac{d^2x^j}{ds^2} + g^{jl}\frac{\partial g_{il}}{\partial x^k}\frac{dx^i}{ds}\frac{dx^k}{ds} = 0$$

which turns out to be

$$\frac{d}{ds}\left(g_{ij}\frac{dx^j}{ds}\right) = 0.$$

The solutions are given by

$$g_{ij}\frac{dx^j}{ds} = const, \qquad i = 1, 2, \ldots, n.$$

Of course, as discussed earlier, the torsion of Γ^j_{ki} is *not* captured by the geodesic curves themselves; it describes the "screw" component of the motion with axis of rotation precisely the tangent direction of the curve.

The associated frame, for which (8) is the adapted connection, is

$$\mathfrak{B} = \{\mathbf{b}_i\}_{i=1}^n = \{\delta_{il}\, g^{lj}\,\partial_{x^j},\ i = 1, 2, \ldots, n.\},$$

with

$$B_i^j = \delta_{il}\, g^{lj} \longleftrightarrow (B^{-1})^k_j = \delta^{kl} g_{lj}.$$

This frame is nothing but "natural gradients" as referred to in the machine learning community after Amari [4]. Here the nominal Kronecker delta δ_{il} and δ^{lj} serve to conform to our summation convention of using one lower index and one upper index.

For comparison with standard model Eq. (8), the resulting family of α-connections from our model, denoted as $\widetilde{\Gamma}^{(\alpha)}_{ki,j} = \frac{1+\alpha}{2} g_{lj}\Gamma^l_{ki}$, takes the form:

$$\widetilde{\Gamma}^{(\alpha)}_{ki,j}(x) = \frac{1+\alpha}{2}\left(\int_\Omega d\omega \left\{\frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^i}\frac{\partial p(\omega|x)}{\partial x^j} + \frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^j}\frac{\partial p(\omega|x)}{\partial x^i}\right\} + \int_\Omega d\omega\, p(\omega|x)\,\frac{\partial \log p(\omega|x)}{\partial x^i}\frac{\partial \log p(\omega|x)}{\partial x^j}\frac{\partial \log p(\omega|x)}{\partial x^k}\right).$$

Torsion of $\widetilde{\Gamma}$ is given by

$$\widetilde{T}^j_{ik} = \frac{1+\alpha}{2}\, g^{jl}\int_\Omega d\omega \left\{\frac{\partial^2 \log p(\omega|x)}{\partial x^i \partial x^l}\frac{\partial p(\omega|x)}{\partial x^k} - \frac{\partial^2 \log p(\omega|x)}{\partial x^k \partial x^l}\frac{\partial p(\omega|x)}{\partial x^i}\right\}.$$

4.2 Univariate normal distribution: an example

As an example of this approach for statistical manifolds, we consider the univariate normal family on the real line ($-\infty < \omega < \infty$)

$$\mathcal{N}(\omega|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\omega-\mu)^2}{2\sigma^2}\right)$$

with parameters $m = \sqrt{2}\,\mu$ and σ. These are a natural choice of parameters, as μ and σ are the mean and the standard deviation, respectively and the factor of $\sqrt{2}$ is for convenience. Reparametrizing, it is possible to consider this statistical manifold as an exponential family using the natural parameters x, in which case it inherits a dually flat geometry. The natural coordinates (x^1, x^2) and expectation coordinates (u_1, u_2) are well-known:

$$x^1 = \frac{\mu}{\sigma^2}, \quad x^2 = -\frac{1}{2\sigma^2};$$
$$u_1 = \mu, \quad u_2 = \mu^2 + \sigma^2.$$

When treating those coordinates as affine coordinates for the dually flat connections, the Fisher–Rao metric g becomes the Hessian metric with potential Φ (expressed in x-coordinates) given by

$$\Phi(x) = -\frac{x^1 \cdot x^1}{4x^2} - \frac{1}{2}\log(-x^2).$$

The conjugate potential Φ^* (expressed in u-coordinates) is

$$\Phi^*(u) = \frac{1}{2}\log(u_2 - u_1 \cdot u_1).$$

As the mean (μ) and standard deviation (σ) parameters of the univariate normal model are intrinsically meaningful in statistics, it is desirable, in the geometric framework, to treat these parameters as affine coordinates for some flat connection. In these coordinates, the Fisher–Rao metric is

$$g = \frac{2}{\sigma^2}(dm^2 + d\sigma^2).$$

As such, if we consider the coordinate frame

$$\left\{\frac{\partial}{\partial m}, \frac{\partial}{\partial \sigma}\right\},$$

its biorthogonal frame is

$$\left\{\frac{\sigma^2}{2}\frac{\partial}{\partial m}, \frac{\sigma^2}{2}\frac{\partial}{\partial \sigma}\right\}.$$

By computing the Lie bracket of the biorthogonal frame, we find that

$$\left[\frac{\sigma^2}{2}\frac{\partial}{\partial m}, \frac{\sigma^2}{2}\frac{\partial}{\partial \sigma}\right] = -\frac{\sigma^3}{2}\frac{\partial}{\partial m}.$$

This is the torsion of the pseudo-Weitzenböck connection adapted to the g-biorthogonal frame. This biorthogonal frame is not a coordinate frame; the torsion of the g-conjugate connection is non-zero.

Note that the Fisher–Rao metric, when expressed in the (m, σ)-coordinates, is *not* Hessian, as expected. These pseudo-Weitzenböck connection derived above, while carrying torsion, has geodesics which are reparametrizations of straight lines in the upper half-plane. This fact does not hold in general, but turns out in the present case because the Fisher–Rao metric, though not Hessian, is so simple for our choice of parametrization.

To summarize, we have contructed a presentation of the univariate normal family (as a manifold of upper half plane), not as a manifold of dually flat (Hessian manifold) in the conventionally-adopted natural and expectation coordinates, but as a statistical manifold admitting torsion (SMAT) in the original (m, σ)-coordinates.

5 Discussion

In this paper, we provided a weakening of the Hessian manifold with dually flat connections by studying the connections that are curvature-free but carry torsion.

Weitzenböck connections are well-studied in "teleparallel" theory of gravitation. Amari [1,2] also studied the role of torsion in plasticity theory. In order to generalize the notion of Weitzenböck connection to information geometry, we drop one of the common assumptions employed in defining Weitzenböck connections—that the given frame is orthonormal with respect to a (semi-)Riemannian metric. This assumption has forced the connection to be metric, which is not desirable when studying manifolds of statistical models. To construct an affine connection with torsion, a good first case is when the curvature vanishes, in which case the given connection is said to be "pseudo-Weitzenböck."

For each flat connection ∇, the group $GL(n)$ acts transitively and freely on the set of parallel frames, so there is a choice of parallel frames for each curvature-free connection. A pseudo-Weitzenböck connection on a Riemannian manifold is metric iff there is an element of $GL(n)$ that transforms $\mathfrak{B}$ into an orthogonal frame. This is more general than the assumption that the frame $\mathfrak{B}$ is orthonormal everywhere, so it is possible that $\nabla^{\mathfrak{B}}$ is metric even if B is not orthonormal. Put another way, when $\mathfrak{B} \equiv \mathfrak{B}^{\star}$, the connection $\nabla^{\mathfrak{B}}$ is metric. However, it is possible for the connection $\nabla^{\mathfrak{B}}$ to be metric even when $\mathfrak{B} \neq \mathfrak{B}^{\star}$.

Suppose that $\mathfrak{B}$ is a coordinate frame. If $\nabla^{\mathfrak{B}^{\star}}$ is torsion-free, then g must be Hessian. It may be worthwhile to see whether this allows us to determine conditions for a given metric to be Hessian.The state-of-the-art results on this problem are given by Armstrong and Amari [5], who were able to compute the two curvature obstructions for a 4-metric to admit a Hessian metric. One of the obstructions is that the Pontryagin forms must vanish, but the second obstruction is a mysterious cubic expression in the curvature. This is not all of the obstructions, but are the only two curvature obstructions in 4 dimensions.

If $\mathfrak{B}^1$ is dual to $\mathfrak{B}$ with respect to g^1 and $\mathfrak{B}^2$ is dual to $\mathfrak{B}$ with respect to another metric g^2, then we do not expect for $\mathfrak{B}^1$ to necessarily be g-biorthogonal to $\mathfrak{B}^2$ for some third metric g. This is unlikely to be true because metrics do not form a group. However, we would like to understand the possible biorthogonal frames in better detail. We could however ask if the image of the $\mathfrak{B}$ under possible biorthogonalities is transitive in some sense. That is to say, whether by dualizing with respect to several different metrics in succession, we can find an arbitrary frame pointwise.

Since pseudo-Weitzenböck connections are highly non-unique, it is worthwhile to ask whether one can construct canonical examples of pseudo-Weitzenböck connections. A naive approach would be to try to minimize the torsion. However, this produces coordinate frames, which are still non-unique. One approach may be to try to minimize the torsion for both the connection and its g-conjugate. This minimizer will still not be unique in general, but this seems more promising. To see that the minimizer may not be unique, not if a Riemannian manifold is Hessian (in which case the connections can be chosen to be dually flat), it may be possible to express a Hessian metric in terms of many different coordinates and choices of convex function. For instance, a hyperbolic plane can be expressed as the statistical manifold of the negative trinomial, the univariate normal family, or the inverse Gaussian distribution. Each of these yields different dually flat structures on hyperbolic space. We expect that in higher dimensional spaces, where there are topological and curvature obstructions to a metric being Hessian, the program to find canonical curvature-free conjugate connections may be more successful, at least for generic metrics.

Classical information geometry is centered around the notion of statistical manifolds, or manifolds admitting a statistical structure. There are two equivalent definitions,

(i) Lauritzen's [15] viewpoint: $(\mathbb{M}, g, \nabla, \nabla^*)$ where the pair of g-conjugated connections ∇ and ∇^* are both torsion-free;
(ii) Kurose's [14] viewpoint: $(\mathbb{M}, g, \nabla)$ where ∇ is torsion-free and Codazzi coupled to g.

With application to parametric statistical models, the Riemannian metric is the Fisher–Rao metric and the pair of conjugate connections are the (± 1)-connections. These are "canonical" objects once the parametric statistical model $p(\cdot|x)$ is specified, and there have been theories that show that they are unique second- and third-order invariants for parametric statistical models (see [7,9]). In particular, such geometry is generated by divergence (contrast) functions. Here we provide another "canonical" construction of the manifold of parametric statistical model as a "partially flat" geometry under which both conjugate connections are curvature-free. In other words, our construction of this manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ is such that ∇^* is flat and ∇ is curvature-free but usually carries torsion, while g is still the Fisher–Rao metric. This is a particular kind of statistical manifold admitting torsion (SMAT) that can be generated by "pre-contrast functions" [12]. Compared to statistical manifold $(\mathbb{M}, g, \nabla, \nabla^*)$ à la Lauritzen, our alternative approach selects a pair of connections both of which are, instead of torsion-free, curvature-free. Compared to statistical manifold $(\mathbb{M}, g, \nabla)$ à la Kurose, our alternative approach selects a connection that is, instead of Codazzi coupled, SMAT-coupled to g. More details can be found in a follow-up paper [17]. The switch of emphasis from curvature to torsion may lead to interesting reformulation of information geometry. This is in parallel to works in theoretical physics which reformulate the theory of general relativity in terms of a Weitzenböck connection (see also the work of Fernandez and Bloch [10] in application to mechanics).

Acknowledgements The project is supported by DARPA/ARO Grant W911NF-16-1-0383 ("Information Geometry: Geometrization of Science of Information", PI: Zhang).

References

1. Amari, S.: On some primary structures of non-Riemannian plasticity theory. RAAG Mem. Unifying Study Basic Probl. Eng. Phys. Sci. Means Geom. Div. D **3**, 99–108 (1962)
2. Amari, S.: A geometrical theory of moving dislocations and anelasticity. RAAG Mem. Unifying Study Basic Probl. Eng. Phys. Sci. Means Geom. Div. D **4**, 142–152 (1968)
3. Amari, S.: Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics. Springer, New York (1985)
4. Amari, S.: Natural gradient works efficiently in learning. Neural Comput. **10**(2), 251–276 (1998)
5. Amari, S., Armstrong, J.: Curvature of Hessian manifolds. Differ. Geom. Appl. **33**, 1–12 (2014)
6. Amari, S., Nagaoka, H.: Methods of Information Geometry, vol. 191. American Mathematical Society, Providence (2000)
7. Ay, N., Jost, J., Vân Lê, H., Schwachhöfer, L.: Information geometry and sufficient statistics. Probab. Theory Relat. Fields **162**(1–2), 327–364 (2015)
8. De Andrade, V.C., Guillen, L.C.T., Pereira, J.G.: Teleparallel gravity: an overview. In: The Ninth Marcel Grossmann Meeting: On Recent Developments in Theoretical and Experimental General Relativity, Gravitation and Relativistic Field Theories (In 3 Volumes), pp. 1022–1023. World Scientific, Singapore (2002)
9. Dowty, J.G.: Chentsov's theorem for exponential families. Inf. Geom. **1**(1), 117–135 (2018)
10. Fernandez, O.E., Bloch, A.M.: The Weitzenböck connection and time reparameterization in nonholonomic mechanics. J. Math. Phys. **52**(1), 012901 (2011)
11. Garecki, J.: Teleparallel equivalent of general relativity: a critical review **(arXiv preprint)**. arXiv:1010.2654 (2010)
12. Henmi, M., Matsuzoe, H.: Geometry of pre-contrast functions and non-conservative estimating functions. In: AIP Conference Proceedings, vol. 1340, pp. 32–41. AIP (2011)
13. Henmi, M., Matsuzoe, H.: Statistical manifolds admitting torsion and partially flat spaces. In: Geometric Structures of Information, pp. 37–50. Springer, Cham (2018)
14. Kurose, T.: Dual connections and affine geometry. Math. Z. **203**(1), 115–121 (1990)
15. Lauritzen, S.L.: Statistical manifolds. In: Differential Geometry in Statistical Inference. IMS Lecture Notes Monograph Series, vol. 10, pp. 163–216. Institute of Mathematical Statistics (1987)
16. Tao, J., Zhang, J.: Transformations and coupling relations for affine connections. Differ. Geom. Appl. **49**, 111–130 (2016)
17. Zhang, J., Khan, G.: New geometry of parametric statistical models. In: Geometric Sciences of Information. Springer (2019, in press)