*Article*

# On Monotone Embedding in Information Geometry

**Jun Zhang**

Department of Psychology and Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109, USA; E-Mail: junz@umich.edu

**Abstract:** A paper was published (Harsha and Subrahamanian Moosath, 2014) in which the authors claimed to have discovered an extension to Amari's $\alpha$-geometry through a general monotone embedding function. It will be pointed out here that this so-called $(F, G)$-geometry (which includes $F$-geometry as a special case) is identical to Zhang's (2004) extension to the $\alpha$-geometry, where the name of the pair of monotone embedding functions $\rho$ and $\tau$ were used instead of $F$ and $H$ used in Harsha and Subrahamanian Moosath (2014). Their weighting function $G$ for the Riemannian metric appears cosmetically due to a rewrite of the score function in log-representation as opposed to $(\rho, \tau)$-representation in Zhang (2004). It is further shown here that the resulting metric and $\alpha$-connections obtained by Zhang (2004) through arbitrary monotone embeddings is a unique extension of the $\alpha$-geometric structure. As a special case, Naudts' (2004) $\phi$-logarithm embedding (using the so-called $\log_\phi$ function) is recovered with the identification $\rho = \phi$, $\tau = \log_\phi$, with $\phi$-exponential $\exp_\phi$ given by the associated convex function linking the two representations.

In a recent paper that appeared in *Entropy* (Harsha and Subrahamanian Moosath, 2014) [1], the authors proposed an extension to Amari's $\alpha$-geometry, which they call $F$- or $(F, G)$-geometry, where $F$ is a monotone embedding function and $G$ is the weighting function for taking the expectation of random variables in calculating the Riemannian metric ($G = 1$ reduces to $F$-geometry, with the standard Fisher–Rao metric). This paper serves the purpose of pointing out that $(F, G)$-geometry as proposed is the same as what Zhang (2004) [2] has obtained for extending the $\alpha$-geometry and captured in his

subsequent work [4–8]. The metric and affine connections proposed by [1] are identical to [2] apart from the notations: the embedding functions $F$ and $H$ in [1] were denoted as $\rho$ and $\tau$ in [2], and weighting function $G$ in [1] is a trivial rewriting of the convex function $f$ used by [2].

This paper will start in Section 1 with a review of Amari's $\alpha$-geometry and $\alpha$-embedding, a review of Zhang's (2004) [2] extension to $\rho$-embedding with an arbitrary monotone function and a summary of Harsha and Subrahamanian Moosath (2014) [1]. Then, the equivalence of [1] to [2] is shown. In Section 2, after analyzing the group of monotone embedding functions, a stronger statement is made: the construction of [2] is a unique dualistic extension of Amari's $\alpha$-geometry through arbitrary monotone embedding in place of $\alpha$-embedding. As an important special case, we illustrate how the deformed logarithm $\log_\phi$ associated with an arbitrary strictly increasing function $\phi$ as investigated by Naudts (2004) [3] arises naturally from identifying $\phi$ with $\rho$ and with a proper choice of the auxiliary function $f$ as a part of Zhang's theory.

## 1. Equivalence of $(F, G)$-Geometry to Zhang's (2004) [2] $(\rho, \tau)$-Geometry

### 1.1. Amari's $\alpha$-Geometry and $\alpha$-Embedding

The now standard differential geometric characterization of the manifold $\mathcal{M}_\Theta = \{\, p(\cdot \mid \theta), \theta \in \Theta \subseteq \mathbb{R}^n \,\}$ of parametric probability functions $p$ (probability density or probability distributions) is through the Fisher–Rao metric $g_{ij}$ as its Riemannian metric:

$$g_{ij}(\theta) = \mathrm{E}_\mu \left\{ p(\zeta|\theta) \frac{\partial \log p(\zeta|\theta)}{\partial \theta^i} \frac{\partial \log p(\zeta|\theta)}{\partial \theta^j} \right\} \tag{1}$$

and a family of $\alpha$-connections (given by Amari [9,10]) with coefficients $\Gamma^{(\alpha)}$ ($\alpha \in \mathbb{R}$):

$$\Gamma^{(\alpha)}_{ij,k}(\theta) = \mathrm{E}_\mu \left\{ \left( \frac{1-\alpha}{2} \frac{\partial \log p(\zeta|\theta)}{\partial \theta^i} \frac{\partial \log p(\zeta|\theta)}{\partial \theta^j} + \frac{\partial^2 \log p(\zeta|\theta)}{\partial \theta^i \partial \theta^j} \right) \frac{\partial p(\zeta|\theta)}{\partial \theta^k} \right\} . \tag{2}$$

Here, $\mathrm{E}_\mu$ denotes the expectation with respect to a background measure $\mu$ of the random variable denoted by $\zeta$:

$$\mathrm{E}_\mu\{\cdot\} = \int (\cdot) d\mu(\zeta). \tag{3}$$

The $\alpha$-connection is constructed as a convex combination of a pair of conjugate connections $\Gamma, \Gamma^*$

$$\Gamma^{(\alpha)}_{ij,k}(\theta) = \frac{1+\alpha}{2} \Gamma_{ij,k}(\theta) + \frac{1-\alpha}{2} \Gamma^*_{ij,k}(\theta), \tag{4}$$

where $\Gamma \equiv \Gamma^{(1)}$ is frequently called $e$-connection ($\alpha = 1$) and $\Gamma^* \equiv \Gamma^{(-1)}$ called $m$-connection ($\alpha = -1$). A Riemannian manifold $\mathcal{M}_\mu$ with its metric $g$ and the family of $\alpha$-connections $\Gamma^{(\alpha)}$ in the form of (1) and (2) has been called $\alpha$-geometry. Amari's $\alpha$-geometry can be specified in terms of a symmetric $(0, 2)$-tensor $g_{ij}$ (the Fisher–Rao metric) and a totally symmetric $(0, 3)$-tensor $T_{ijk}$ (sometimes called the Amari–Chentsov tensor), which is linked to the $\alpha$-connections via:

$$\Gamma^{(\alpha)}_{ij,k} = \Gamma^{LC}_{ij,k}(\theta) - \frac{\alpha}{2} T_{ijk}(\theta) , \tag{5}$$

where $\Gamma^{LC}_{ij,k}$ is the Levi–Civita connection corresponding to the Riemannian metric $g$.

As an extension of the logarithmic embedding $l(p) = \log p$ of probability density function $p$, an $\alpha$-embedding function [10] is defined through $l^{(\alpha)} : \mathbb{R}^+ \to \mathbb{R}$:

$$
l^{(\alpha)}(t) = \begin{cases} \log t & \alpha = 1 \\ \frac{2}{1-\alpha}\, t^{(1-\alpha)/2} & \alpha \neq 1 \end{cases} . \tag{6}
$$

It is an interesting observation (e.g., p. 46 in [11]) that the $\alpha$-geometry can be recovered under such $\alpha$-representation (scaling) of the probability function, that is the Fisher–Rao metric turns out to be $\alpha$-independent (*i.e.*, embedding independent) and the $\pm 1$-connections precisely the $\alpha$-connections:

$$
g_{ij}(\theta) = \mathrm{E}_\mu \left\{ \frac{\partial l^{(\alpha)}(p(\zeta|\theta))}{\partial \theta^i} \frac{\partial l^{(-\alpha)}(p(\zeta|\theta))}{\partial \theta^j} \right\} , \tag{7}
$$

$$
\Gamma^{(\alpha)}_{ij,k}(\theta) = \mathrm{E}_\mu \left\{ \frac{\partial^2 l^{(\alpha)}(p(\zeta|\theta))}{\partial \theta^i \partial \theta^j} \frac{\partial l^{(-\alpha)}(p(\zeta|\theta))}{\partial \theta^k} \right\} . \tag{8}
$$

A variance of $\alpha$-embedding of a probability function plays an important role in Tsallis statistics; see [12–14]. On the geometric side, [15,16] illuminated that the $\alpha$-scaling of the probability functions leads to a conformal transformation.

### 1.2. Zhang (2004) [2] Extension: $\rho$-Embedding and $(\rho, \tau)$-Geometry

Zhang [2,4,6] obtained generalizations of the $\alpha$-geometry for a pair of monotone embeddings, called $\rho$- and $\tau$-embeddings generalizing $\alpha$-embedding. Given any smooth strictly convex function $f : \mathbb{R} \to \mathbb{R}$, with convex conjugate $f^*$ given by:

$$
f^*(t) = t\, (f')^{-1}(t) - f((f')^{-1}(t)) , \tag{9}
$$

Zhang (2004) defines a pair of conjugate representations [2] (Section 3.2) using two strictly increasing functions $\rho, \tau$ from $\mathbb{R} \to \mathbb{R}$:

(1) we call $\rho$-representation of a probability function $p$ the mapping $p \mapsto \rho(p)$;

(2) we say $\tau$-representation of the probability function $p \mapsto \tau(p)$ is conjugate to $\rho$-representation with respect to a smooth and strictly convex function $f$, or simply $\tau$ is $f$-conjugate to $\rho$, if:

$$
\tau(p) = f'(\rho(p)) = ((f^*)')^{-1}(\rho(p)) , \tag{10}
$$

which can be equivalently written as:

$$
\rho(p) = (f')^{-1}(\tau(p)) = (f^*)'(\tau(p)) . \tag{11}
$$

These equalities in (10) and (11) hold, and they are equivalent, because $f'$ and $(f^*)'$ are both strictly increasing (due to their strict convexity) and that $(f^*)^* = f$, $(f^*)' = (f')^{-1}$. Sometimes, we write $f' = \sigma, (f^*)^{-1} = \sigma^{-1}$ for convenience, so $\sigma(\rho) = \tau, \sigma^{-1}(\tau) = \rho$, for a strictly increasing function $\tau$.

As a first example, we may set $\rho(t) = t, \tau(t) = \log t$. Then, we can derive that $f^*(t) = \exp(t)$ and $f(t) = t \log t - t + 1$. That $\rho(p)$ and $\tau(p)$ are just the $p$ and $\log p$ representation reflects the conventional

dual embeddings that have later been extended to $\phi$- and $\log_\phi$-embedding in ([3]). In Section 2.2, it will be shown that Naudts' $\phi$-logarithm formulation is recovered as a special case of the $(\rho, \tau)$-embedding.

As another example, we may set $\rho(p) = l^{(\beta)}(p)$ to be the $\beta$-representation given by Equation (6); this would have been traditionally called "alpha-embedding", except we use the symbol $\beta$, so that the $\alpha$-parameter will be reserved for indexing $\alpha$-connections. In this case, the conjugate representation is the $(-\beta)$-representation $\tau(p) = l^{(-\beta)}(p)$:

$$\rho(p) = l^{(\beta)}(p) \longleftrightarrow \tau(p) = l^{(-\beta)}(p) \ . \tag{12}$$

In this case, $\rho$ and $\tau$ are conjugate with respect to $f$, where $f$ is given by:

$$f(t) = \frac{2}{1+\beta}\left(\left(\frac{1-\beta}{2}\right)t\right)^{\frac{2}{1-\beta}} \ , \quad f^*(t) = \frac{2}{1-\beta}\left(\left(\frac{1+\beta}{2}\right)t\right)^{\frac{2}{1+\beta}} \ . \tag{13}$$

Based on divergence functions constructed under monotone embedding, Zhang ([2]) showed:

**Proposition 1.** ([2], Proposition 7) *Using an arbitrary monotone embedding function $\rho$ and an arbitrary smooth strictly convex function $f$, a generalization of $\alpha$-geometry is obtained, with metric and $\alpha$-connections taking the form:*

$$g_{ij}(\theta) = E_\mu\left\{ f''(\rho(p(\zeta|\theta))\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^i}\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^j}\right\} \tag{14}$$

$$\Gamma_{ij,k}^{(\alpha)}(\theta) = E_\mu\left\{\frac{1-\alpha}{2}f'''(\rho(p(\zeta|\theta)))A_{ijk} + f''(\rho(p(\zeta|\theta)))B_{ijk}\right\} \ , \tag{15}$$

*where:*

$$A_{ijk}(\zeta,\theta) = \frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^i}\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^j}\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^k} \ , \quad B_{ijk}(\zeta,\theta) = \frac{\partial^2\rho(p(\zeta|\theta))}{\partial\theta^i\partial\theta^j}\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^k} \ . \tag{16}$$

As special cases,

$$\Gamma_{ij,k}(\theta) = E_\mu\left\{ f''(\rho(p))\frac{\partial^2\rho(p)}{\partial\theta^i\partial\theta^j}\frac{\partial\rho(p)}{\partial\theta^k}\right\} \ , \tag{17}$$

$$\Gamma_{ij,k}^*(\theta) = E_\mu\left\{\frac{\partial\rho(p)}{\partial\theta^k}\left(f'''(\rho(p))\frac{\partial\rho(p)}{\partial\theta^i}\frac{\partial\rho(p)}{\partial\theta^j} + f''(\rho(p))\frac{\partial^2\rho(p)}{\partial\theta^i\partial\theta^j}\right)\right\} \ . \tag{18}$$

Furthermore, taking a pair of monotone representations, the metric tensor and affine connections stated in Proposition 1 have dualistic expressions:

**Corollary 1.** ([2], Proposition 8) *Using two arbitrary monotone embedding functions $\rho$ and $\tau$, the metric and $\alpha$-connections of (14)–(16) are:*

$$g_{ij}(\theta) = E_\mu\left\{\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^i}\frac{\partial\tau(p(\zeta|\theta))}{\partial\theta^j}\right\} \ , \tag{19}$$

$$\Gamma_{ij,k}^{(\alpha)}(\theta) = E_\mu\left\{\frac{1-\alpha}{2}\frac{\partial^2\tau(p(\zeta,\theta))}{\partial\theta^i\partial\theta^j}\frac{\partial\rho(p(\zeta|\theta))}{\partial\theta^k} + \frac{1+\alpha}{2}\frac{\partial^2\rho(p(\zeta|\theta))}{\partial\theta^i\partial\theta^j}\frac{\partial\tau(p(\zeta|\theta))}{\partial\theta^k}\right\} \ . \tag{20}$$

As a special case, when $\rho, \tau$ take the familiar alpha-embeddings (12) (using $\beta$ as the parameter), the $\alpha$-connections becomes $(\alpha\beta)$-connections:

$$\Gamma_{ij,k}^{(\alpha)}(\theta) = E_\mu\left\{\left(\frac{1-\alpha\beta}{2}\frac{\partial\log p(\zeta|\theta)}{\partial\theta^i}\frac{\partial\log p(\zeta|\theta)}{\partial\theta^j} + \frac{\partial^2\log p(\zeta|\theta)}{\partial\theta^i\partial\theta^j}\right)\frac{\partial p(\zeta|\theta)}{\partial\theta^k}\right\} \ , \tag{21}$$

with the product $\alpha \cdot \beta$ playing the role of the alpha-parameter indexing the family of connections.

### 1.3. Harsha and Subrahamanian Moosath's (2014) Work [1]

Using a monotone embedding function denoted as $F$ and a weighting function denoted as $G$ ($G = 1$ is a special case to reduce to what they called $F$-geometry), these authors [1] proposed $(F, G)$-metric as (their Equation (33) in [1]):

$$g_{ij}^{F,G} = \mathrm{E}_{\mu} \left\{ pG(p) \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right\} \tag{22}$$

with affine connection given as (their Equation (34)):

$$\Gamma_{ijk}^{F,G} = \mathrm{E}_{\mu} \left\{ p\, G(p) \frac{\partial \log p}{\partial \theta^k} \left( \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} + \left( 1 + \frac{pF''(p)}{F'(p)} \right) \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right) \right\}. \tag{23}$$

Note $\mathrm{E}_p\{(\cdot)\} = \mathrm{E}_{\mu}\{(\cdot)p\}$. (23) is the expression for the $e$-connection ($\alpha = 1$), $\Gamma_{ijk}^{F,G}$. To express the conjugate connection ($m$-connection, $\alpha = -1$), $\Gamma_{ijk}^{H,G}$, a dual embedding function $H$ is introduced, which is shown ([1], Theorem 3.2) to be related to $F$ and $G$ via (their Equation (36)):

$$H'(p) = \frac{G(p)}{pF'(p)}. \tag{24}$$

In such a case, the conjugate connection $\Gamma_{ijk}^{H,G}$ (*sic*, more accurately $(\Gamma^{F,G})_{ijk}^*$) is expressed as (their Equation (37)):

$$\Gamma_{ijk}^{H,G} = \mathrm{E}_{\mu} \left\{ pG(p) \frac{\partial \log p}{\partial \theta^k} \left( \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} + \left( \frac{pG'(p)}{G(p)} - \frac{pF''(p)}{F'(p)} \right) \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right) \right\}. \tag{25}$$

We now show the equivalence of the three expressions (14), (17), (18) from the work [2] with the three corresponding expressions (22), (23), (25) from the work [1].

**Statement 1.** *Equations (14) and (22) give the same Riemannian metric; Equations (17) and (23) give the same affine connection; and Equations (18) and (25) give the same conjugate connection, as long as:*

$$F(p) = \rho(p) , \quad G(p) = (\rho')^2\, p\, f''(\rho(p)). \tag{26}$$

**Proof.** Re-writing (14), and keeping in mind:

$$\frac{\partial \rho(p)}{\partial \theta^i} = \rho'(p) \frac{\partial p}{\partial \theta^i} = p\, \rho'(p) \frac{\partial \log p}{\partial \theta^i}, \tag{27}$$

so:

$$g_{ij}(\theta) = \mathrm{E}_{\mu} \left\{ f''(\rho(p)(p\, \rho'(p))^2 \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right\}. \tag{28}$$

Comparing the above with (22), obviously, $F$ is just $\rho$, and $G$ is linked to $f$ and $\rho$:

$$G(p) = (\rho')^2\, p\, f''(\rho(p)) = p\, \rho'(p)\tau'(p) \tag{29}$$

where we have used (10).

Next, differentiate (27); we obtain:

$$\frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^j} = \frac{\partial p}{\partial \theta^j} \rho'(p) \frac{\partial \log p}{\partial \theta^i} + p\, \rho''(p) \frac{\partial p}{\partial \theta^j} \frac{\partial \log p}{\partial \theta^i} + p\, \rho'(p) \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} \tag{30}$$

$$= p\, \rho'(p) \left( \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} + \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} + \frac{p\rho''(p)}{\rho'(p)} \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right) \tag{31}$$

$$= p\, \rho'(p) \left( \frac{\partial^2 \log p}{\partial \theta^i \partial \theta^j} + \left( 1 + \frac{p\rho''(p)}{\rho'(p)} \right) \frac{\partial \log p}{\partial \theta^i} \frac{\partial \log p}{\partial \theta^j} \right). \tag{32}$$

Identifying $F = \rho$ and making use of (29), we see that (17) is precisely (23).

Finally, differentiate (29),

$$G'(p) = (\rho')^2 f''(\rho(p)) + (\rho')^3 p \, f'''(\rho(p)) + 2\rho'(p)\rho''(p) \, p \, f''(\rho(p)). \tag{33}$$

Therefore,

$$\frac{pG'(p)}{G(p)} - \frac{pF''(p)}{F'(p)} = 1 + \frac{p\rho'(p)f'''(\rho(p))}{f''(\rho(p))} + \frac{p\rho''(p)}{\rho'(p)}. \tag{34}$$

After substituting (34) and (29) into (25) and making use of (31), the expression (18) results. □

**Statement 2.** *The conjugate embedding function $H$ is the same as $\tau$. The conjugate connection (25), when expressed using $H$, has the same form as (23) for $\Gamma_{ij,k}^{G,F}$ using $F$.*

**Proof.** Applying Definition (24) immediately yields $H' = \tau'$. Therefore, (apart from constant) $H(p) = \tau(p)$. Next, we will express (25) explicitly using the conjugate embedding function $H$ (rather than $F$) and the weighting function $G$. That is to say, we will simplify the terms in the middle parenthesis of (25):

$$\frac{pG'(p)}{G(p)} - \frac{pF''(p)}{F'(p)} = p\left(\log\frac{G(p)}{F'(p)}\right)' = p\left(\log(pH'(p))\right)' = p\left(\log p + \log H'(p)\right)' \tag{35}$$

$$= p\left(\frac{1}{p} + \frac{H''(p)}{H'}\right) = 1 + \frac{pH''}{H'(p)} \; . \tag{36}$$

Hence, (25) has the same expression as (23) showing the duality between the embedding function $H$ and the embedding function $F$. □

By Statement 1, starting from $F$ (that is, $\rho$) and $G$ and imposing conjugacy requirement on the pair of affine connections, one is guaranteed to derive $H$ (that is, $\tau$) as the conjugate embedding function.

From Statements 1 and 2, we conclude that, Harsha and Moosath's $F$-embedding [1] replicates the $\rho$-embedding of Zhang (2004) [2]; the conjugate $H$-embedding turns out to be identical to $\tau$-embedding of [2]. Contrary to the authors' claim (Remark 3.7 of [1], p. 2480), $(F, G)$-geometry is identical to Zhang's $(\rho, \tau)$ geometry [2]. In particular, their $F$-geometry is recovered by simply choosing $f$ to satisfy $f''(t) = 1/(\rho^{-1}(t) \, (\rho'(\rho^{-1}(t)))^2)$, for a given $\rho$. The subsequent development in their paper [1], e.g., the definition of the $F$-affine manifold (their Equation (50)), replicates the definition of $\rho$-affine manifold in [2] (Section 3.4).

During the review of their manuscript [1] and in subsequent personal communications, these authors argued that they used a different approach: $(F, G)$-geometry is derived by embedding the manifold into the space of random variables and suitably defining the inner product through using the $F$-expectation (their Equation (15)) and $(F, G)$-expectation (their Equation (32)) as a general weighted expectation of a random variable, while Zhang (2004) [2] derived the geometry through constructing a divergence function. This difference, however, is entirely superficial, because the relationship between divergence functions and geometric structure (metric and affine connection) is well-established by Eguchi's work [17,18] and known to information geometers. Therefore, neither the approach nor the results of Harsha and Moosath's proposed $(F, H, G)$ extension to Amari's $\alpha$-geometry differs from Zhang's proposed $(\rho, \tau, f)$ extension, with the following correspondence in different symbols by the two papers:

$$F \Longleftrightarrow \rho \; , \quad H \Longleftrightarrow \tau \; , \tag{37}$$

$$G(t) \iff t\rho'(t)\tau'(t) = tf''(\rho(t))(\rho'(t))^2 = t(f^*)''(\tau(t))(\tau'(t))^2 \; ; \tag{38}$$

the difference in the representation of score function as $\log$-representation in [1] or under $\rho$ or $\tau$-representation in [2] is cosmetic.

## 2. Uniqueness of $(\rho, \tau)$-Geometry and Representation Duality

### 2.1. Monotone Embedding as a Transformation Group

Monotone representations of any given probability function form a transformation group, with functional composition as group composition operation and the functional inverse as the group inverse operation. This was pointed out by Zhang [6] (Section 2.2.2). We state it as a lemma here.

**Lemma 1.** *Denote $\Omega$ as the set of strictly increasing functions from $\mathbb{R} \to \mathbb{R}$. Then, $(\Omega, \circ)$ forms a group, with $\circ$ denoting functional composition.*

**Proof.** We easily verify that:

(1) closure for $\circ$: for any $\rho_1$, $\rho_2 \in \Omega$, $\rho_2 \circ \rho_1$, defined as $\rho_2(\rho_1(\cdot))$, is strictly increasing, and hence, $\rho_2 \circ \rho_1 \in \Omega$;

(2) existence of unique identity element: the identity function $\iota$, which satisfies $\rho \circ \iota = \iota \circ \rho = \rho$, is strictly increasing, and hence, $\iota \in \Omega$ and is unique;

(3) existence of inverse: for any $\rho \in \Omega$, its functional inverse $\rho^{-1}$, which satisfies $\rho^{-1} \circ \rho = \rho^{-1} \circ \rho = \iota$, is also strictly increasing, and hence, $\rho^{-1} \in \Omega$;

(4) associativity of $\circ$: for any three $\rho_1$, $\rho_2$, $\rho_3 \in \Omega$, then $(\rho_1 \circ \rho_2) \circ \rho_3 = \rho_1 \circ (\rho_2 \circ \rho_3)$.
$\square$

Recall that the derivative of smooth strictly convex functions are strictly increasing functions. From this perspective, $f' = \tau \circ \rho^{-1} = \tau(\rho^{-1}(\cdot))$, $(f^*)' = \rho \circ \tau^{-1} = \rho(\tau^{-1}(\cdot))$, encountered above, are themselves two mutually inverse strictly increasing functions. This is the rationale behind Zhang's ([2]) choice of $f$ (and $f^*$) as the auxiliary function to capture conjugate embedding, rather than using $G$ as in [1]. The following identities are useful; they are obtained by differentiating (10) and (11):

$$f''(\rho(t)) \, \rho'(t) = \tau'(t) \; , \quad (f^*)''(\tau(t)) \, \tau'(t) = \rho'(t) \; ; \tag{39}$$

therefore:

$$f''(\rho(t)) \, (\rho'(t))^2 = (f^*)''(\tau(t)) \, (\tau'(t))^2 \; , \tag{40}$$

and:

$$f''(\rho(t)) \, (f^*)''(\tau(t)) = 1. \tag{41}$$

With respect to (41), taking $\log$ on both sides yields:

$$\log f''(\rho(t)) + \log(f^*)''(\tau(t)) = 0 \; . \tag{42}$$

Move and differentiate:

$$\frac{f'''(\rho(t))\,\rho'(t)}{f''(\rho(t))} = -\frac{(f^*)'''(\tau(t))\,\tau'(t)}{(f^*)''(\tau(t))}\,. \tag{43}$$

Making use of (40) yields:

$$f'''(\rho(t))\,(\rho'(t))^3 = -(f^*)'''(\tau(t))\,(\tau'(t))^3\,. \tag{44}$$

Note the coupling between $f$ and $\rho, \tau$ given by (10), (11), (40) and (44). They allow us to cast (14) and (15) in terms of $f^*$ and $\tau$.

Among the triple $(f, \rho, \tau)$, given any two, the third is specified. In particular, if we arbitrary choose two strictly increasing functions $\rho$ and $\tau$ as embedding functions and require them to be conjugate embeddings, then $f$ is specified by $f'(t) = \tau(\rho^{-1}(t))$. In terms of conjugate function $f^*$, the relation is $(f^*)'(t) = \rho(\tau^{-1}(t))$. The function $f$ (or $f^*$) is important in constructing the general class of divergence function.

### 2.2. Naudts' $\phi$-Logarithm as a Special Case

In his 2004 publication [3], Naudts considered the "deformed" logarithm function as an extension to the exponential family of densities that is log-linear. Given a strictly increasing and strictly positive function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, the $\phi$-logarithm is defined as:

$$\log_\phi(t) = \int_1^t \frac{1}{\phi(s)}\,ds\,, \quad (t > 0). \tag{45}$$

The deformed exponential denoted $\exp_\psi$, is defined by:

$$\exp_\psi(t) = 1 + \int_0^t \psi(s)\,ds. \tag{46}$$

(Naudts (2004) used the notation $\exp_\phi$, so our current rendition has a subtle difference shown as (48) and (49) below.) It can be shown that the deformed functions $\log_\phi$ and $\exp_\psi$ are in fact inverse functions of each other if:

$$\psi(\log_\phi(t)) = \phi(t), \quad \psi(t) = \phi(\exp_\psi(t)). \tag{47}$$

Stated alternatively, the deformed logarithmic function $h(t) = \log_\phi(t)$ can be viewed as the solution to the following integral and its equivalent differential equation:

$$h(t) = \int_1^t \frac{1}{\psi(h(s))}ds \iff \frac{dh}{dt} = \frac{1}{\psi(h(t))}, \tag{48}$$

whereas the deformed exponential function $h(t) = \exp_\psi(t)$ can be viewed as the solution to the following integral and its equivalent differential equation:

$$h(t) = 1 + \int_0^t \phi(h(s))ds \iff \frac{dh}{dt} = \phi(h(t)). \tag{49}$$

We now show that the above formulation can be re-written as $(\rho, \tau)$-embeddings with a particular choice of $f$ (or equivalently, $f^*$) function. Set $\phi(t) = \rho(t)$ and $f^*(t) = \exp_\psi(t)$, so that $(f^*)'(t) = \psi(t)$ from (46). Therefore, we derive:

$$\log_\phi(t) = \psi^{-1}(\phi(t)) = ((f^*)')^{-1}(\rho(t)) = f'(\rho(t)) = \tau(t).$$

That is, when $\phi$ is chosen as $\rho$-representation, the deformed logarithm $\log_\phi$ turns out to be the $\tau$-representation, while the deformed exponential is nothing but $f^*$. The relationship (47) is identical to (10) and (11).

In the $\phi$-logarithm approach, once $\phi$ (that is, $\rho$) is specified, then $\log_\phi$ (that is, $\tau$) is specified, through the integral relation (45). Viewing $\tau(\cdot) = f'(\rho(\cdot))$, the relation (45) essentially specifies a strictly convex function $f$, through its derivative $f'$, which operates on $\rho$.

**Proposition 2.** *Denote $\rho \equiv \phi$. The deformed logarithmic transformation $\phi \to \log_\phi$ given by (45) can be viewed as the function composition $f' : \rho \to f'(\rho)$, where $f$ is given by:*

$$f(\rho(t)) = \rho(t)f'(\rho(t)) - t. \tag{50}$$

*Equivalently, using conjugate function $f^*$ given by (9),*

$$\rho = (f^*)' \circ (f^*)^{-1}, \tag{51}$$

*or*

$$\rho = \frac{1}{((f^*)^{-1})'}. \tag{52}$$

**Proof.** From (45), we write:

$$f'(\rho(t)) = \int_1^t \frac{1}{\rho(s)} \, ds, \tag{53}$$

with unknown $f$. Multiply both sides by $\rho'(t)$ and then integrate from one to $x$; the left-hand side of (53) is:

$$\int_1^x f'(\rho(t)) \, \rho'(t) \, dt = \int_1^x f'(\rho(t)) d(\rho(t)) = f(\rho(x)) - f(\rho(1)).$$

The right-hand side of (53), after the same operation, is:

$$\int_1^x \rho'(t) \, dt \int_1^t \frac{1}{\rho(s)} \, ds = \int_1^x \frac{1}{\rho(s)} \, ds \int_s^x \rho'(t) \, dt = \int_1^x \frac{\rho(x) - \rho(s)}{\rho(s)} \, ds$$

$$= \int_1^x \left( \frac{\rho(x)}{\rho(s)} - 1 \right) ds = \rho(x) \left( \int_1^x \frac{1}{\rho(s)} \, ds \right) - \int_1^x ds = \rho(x) \, f'(\rho(x)) - (x - 1).$$

Clearly, $f'(\rho(1)) = 0$ by (53). We set $f(\rho(1)) = -1$. Comparing expressions from the left- and right-hand side, we obtain (50).

Applying (9), we obtain the equivalent expression:

$$f^*(f'(\rho(t))) = t.$$

That is, $f$ is chosen, such that $f^* \circ f'$ is the inverse function of $\rho$, or:

$$\rho = (f^* \circ f')^{-1} = (f')^{-1} \circ (f^*)^{-1} = (f^*)' \circ (f^*)^{-1}.$$

Hence, (51) holds.

Finally, differentiate the identity:

$$f^*((f^*)^{-1}(t)) = t,$$

we obtain:

$$1 = (f^*)'((f^*)^{-1}(t)) \cdot (f^*)^{-1}(t) = \rho(t) \cdot (f^*)^{-1}(t)$$

upon substituting (51). Hence, (52) holds. $\square$

The expression (51) in Proposition 2 shows that for any $\rho$, if one can find a decomposition: $\rho = g' \circ g^{-1}$ in terms of $g$, then $g$ would be the $\rho$-exponential, $g^{-1}$ the $\rho$-logarithm and $g'$ the linking function. In the case of $\phi \mapsto \log_\phi$ transformation, $g = f^*(t)$.

Naudts' ([3]) deformed logarithm/exponential embedding approach and Zhang's ([2]) $(\rho, \tau)$-embedding approach can be seen as playing complementary roles in information geometry: the former makes it easy to generalize the exponentiation and logarithm as inverse operations obeying desired differential/integral equations, while the latter makes it apparent how conjugate $(\rho, \tau)$-embeddings lead to bidualistic expressions for the underlying geometric structures (metric and conjugate connections).

### 2.3. Uniqueness of $(\rho, \tau)$-Geometry

It is known [19,20] that the Fisher–Rao metric and $\alpha$-connections (equivalently, Amari–Chentsov tensor $T$) are the only invariants of sufficient statistics under the Markov morphism of a random variable. In [22,23], the Fisher–Rao metric has been extended to allow a weighting function. In [2,6], general weighting functions for affine connections were made compatible with the generalized (*i.e.*, weighted) Fisher–Rao metric, since they result from divergence functions that are allowed to have the freedom of monotone embedding. The recent reinvention [1] constructed weighted connections that turned out to be identical to the expressions given by [2]. A natural question is, then, whether Zhang's $(\rho, \tau)$ geometry is the unique construction given the freedom of arbitrary monotone embedding. Below, arguments will be provided, along with a proof, for a positive answer to this question.

First, when a probability function $p(\zeta|\theta)$ (as a function of a random variable indexed by $\zeta$ and a background measure of $\mu$) is embedded into the parametric manifold $\mathcal{M}_\Theta$, there are several traditional choices for tangent vectors: $\partial_i p$, $\partial_i \log p$, $\partial_i \sqrt{p}$, *etc*. Each of these are linked with a weighting function (expectation operator), so that the tangent vectors are zero-mean random variables:

$$0 = \mathrm{E}_\mu\{\partial_i p\} = \mathrm{E}_\mu\{(p)\,\partial_i \log p\} = \mathrm{E}_\mu\{(\sqrt{p})\,\partial_i(\sqrt{p})\} = \cdots \tag{54}$$

where the weighting functions are, respectively, one, $p$, $\sqrt{p}$:

$$0 = \mathrm{E}_\mu\{\partial_i p\} = \mathrm{E}_p\{\partial_i \log p\} = \mathrm{E}_{\sqrt{p}}\{\partial_i(\sqrt{p})\} = \cdots$$

For these various choices, the direction of the tangent vectors are all the same. We can consider the above as special cases of $\rho$-embedding, with $\rho(t) = t, \log t, \sqrt{t}$, respectively. Because $\partial_i(\rho(p)) = \rho'(p)\partial_i p$, so a tangent vector retains its direction with any choice of monotone embedding function.

To investigate the weighting function for general monotone $\rho$-embedding, let us consider the $f$-normalization (foliation) condition, *cf.* [21],

$$\mathrm{E}_\mu\{f(\rho(p))\} = 1, \tag{55}$$

where $f$ is a given convex function. Differentiate the above; we get:

$$0 = \mathrm{E}_\mu\left\{ f'(\rho(p))\frac{\partial \rho'(p)}{\partial \theta^i} \right\} = \mathrm{E}_\mu\{\tau(p)\,\partial_i\rho\}. \tag{56}$$

Therefore, we can see that $\tau(p) = f'(\rho(p))$, what we have called the $f$-conjugate of $\rho$, is precisely the weighting function to make $\partial_i \rho$ a zero-mean random function at any point of $\mathcal{M}_\Theta$ (*i.e.*, for any value of $\theta \in \Theta$).

Next, consider the Fisher–Rao metric (1), which can be written as $E_\mu\{\partial_i p\, \partial_j \log p\} = E_\mu\{\partial_i \log p\, \partial_j p\}$, the pairing of a random function with a random functional under two embeddings $p$ and $\log p$. A natural generalization (see [6]) is to use two (independently chosen) monotone embeddings $\rho, \tau$:

$$g_{ij}(\theta) = E_\mu\{\partial_i \rho\, \partial_j \tau\} = E_\mu\{\partial_j \rho\, \partial_i \tau\} = E_\mu\{\rho'(p)\, \tau'(p)\, \partial_i p\, \partial_j p\}\ . \tag{57}$$

This is precisely (14), with the weighting function for the Riemannian metric as $f''(\rho(p))(\rho'(p))^2 = \tau'(p)\rho'(p)$, when tangent vectors are expressed as $\partial_i p$ (identity representation). When $\rho$-representation or $\tau$-representation is adopted, the weighting function is simply $f''(\rho(p))$ or $(f^*)''(\tau(p))$, respectively.

Third, given $\rho, \tau$ embedding, we can construct two affine connections on the manifold as follows. Differentiate (57),

$$\frac{\partial g_{ij}(\theta)}{\partial \theta^k} = E_\mu \left\{ \frac{\partial^2 \rho(p)}{\partial \theta^k \partial \theta^i} \frac{\partial \tau}{\partial \theta^j} + \frac{\partial^2 \tau(p)}{\partial \theta^k \partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^i} \right\}\ , \tag{58}$$

and compare with the relation that defines conjugate connections:

$$\frac{\partial g_{ij}(\theta)}{\partial \theta^k} = \Gamma_{ki,j}(\theta) + \Gamma^*_{kj,i}(\theta)\ ; \tag{59}$$

we can identify:

$$E_\mu \left\{ \frac{\partial^2 \rho(p)}{\partial \theta^k \partial \theta^i} \frac{\partial \tau(p)}{\partial \theta^j} \right\} \tag{60}$$

with $\Gamma_{ki,j}$ and:

$$E_\mu \left\{ \frac{\partial^2 \tau(p)}{\partial \theta^k \partial \theta^i} \frac{\partial \rho(p)}{\partial \theta^j} \right\} \tag{61}$$

with $\Gamma^*_{kj,i}$, respectively. Their difference is, by definition, the Amari–Chentsov (0,3)-tensor $T$:

$$T_{ijk}(\theta) \equiv E_\mu \left\{ \frac{\partial^2 \tau(p)}{\partial \theta^i \partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^k} - \frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^j} \frac{\partial \tau(p)}{\partial \theta^k} \right\}. \tag{62}$$

**Proposition 3.** *$T$ as given by (62) is a totally symmetric (0,3)-tensor.*

**Proof.** First, we prove that $T(\theta)$ is totally symmetric:

$$T_{ijk} = T_{jik} = T_{ikj} = T_{jki} = T_{kij} = T_{kji}. \tag{63}$$

Since (62) clearly implies $T_{ijk} = T_{jik}$, we only need to establish $T_{ijk} = T_{ikj}$. Applying the chain-rule of differentiation,

$$\frac{\partial}{\partial \theta^i} \left( \frac{\partial \tau(p)}{\partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^k} \right) = \frac{\partial^2 \tau(p)}{\partial \theta^i \partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^k} + \frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^k} \frac{\partial \tau(p)}{\partial \theta^j}\ , \tag{64}$$

$$\frac{\partial}{\partial \theta^i} \left( \frac{\partial \rho(p)}{\partial \theta^j} \frac{\partial \tau(p)}{\partial \theta^k} \right) = \frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^j} \frac{\partial \tau(p)}{\partial \theta^k} + \frac{\partial^2 \tau(p)}{\partial \theta^i \partial \theta^k} \frac{\partial \rho(p)}{\partial \theta^j}\ , \tag{65}$$

and taking into account:

$$\frac{\partial \tau(p)}{\partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^k} = \frac{\partial \tau(p)}{\partial \theta^k} \frac{\partial \rho(p)}{\partial \theta^j} = \tau'(p)\rho'(p) \frac{\partial p}{\partial \theta^j} \frac{\partial p}{\partial \theta^k}\ , \tag{66}$$

(62) becomes:

$$T_{ijk}(\theta) = \mathrm{E}_\mu \left\{ -\left( \frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^k} \frac{\partial \tau(p)}{\partial \theta^j} - \frac{\partial^2 \tau(p)}{\partial \theta^i \partial \theta^k} \frac{\partial \rho(p)}{\partial \theta^j} \right) \right\} = T_{ikj}(\theta) . \tag{67}$$

Next, we prove that $T_{ijk}$ is indeed a (0,3)-tensor. This is done through examining the behavior of $T$ under a coordinate transform $\theta \mapsto \bar{\theta}$, with the (inverse) Jacobian matrix $\frac{\partial \theta^k}{\partial \bar{\theta}^l}$, which affects:

$$\frac{\partial \rho(p)}{\partial \bar{\theta}^i} = \sum_l \frac{\partial \rho(p)}{\partial \theta^l} \frac{\partial \theta^l}{\partial \bar{\theta}^i} , \qquad \frac{\partial \tau(p)}{\partial \bar{\theta}^i} = \sum_l \frac{\partial \tau(p)}{\partial \theta^l} \frac{\partial \theta^l}{\partial \bar{\theta}^i} , \tag{68}$$

and:

$$\frac{\partial^2 \rho(p)}{\partial \bar{\theta}^i \partial \bar{\theta}^j} = \sum_{l,m} \frac{\partial^2 \rho(p)}{\partial \theta^l \partial \theta^m} \frac{\partial \theta^l}{\partial \bar{\theta}^i} \frac{\partial \theta^m}{\partial \bar{\theta}^j} + \sum_l \frac{\partial \rho(p)}{\partial \theta^l} \frac{\partial^2 \theta^l}{\partial \bar{\theta}^i \partial \bar{\theta}^j} , \tag{69}$$

$$\frac{\partial^2 \tau(p)}{\partial \bar{\theta}^i \partial \bar{\theta}^j} = \sum_{l,m} \frac{\partial^2 \tau(p)}{\partial \theta^l \partial \theta^m} \frac{\partial \theta^l}{\partial \bar{\theta}^i} \frac{\partial \theta^m}{\partial \bar{\theta}^j} + \sum_l \frac{\partial \tau(p)}{\partial \theta^l} \frac{\partial^2 \theta^l}{\partial \bar{\theta}^i \partial \bar{\theta}^j} . \tag{70}$$

Therefore:

$$\bar{T}_{ijk}(\bar{\theta}) \equiv \mathrm{E}_\mu \left\{ \frac{\partial^2 \tau(p)}{\partial \bar{\theta}^i \partial \bar{\theta}^j} \frac{\partial \rho(p)}{\partial \bar{\theta}^k} - \frac{\partial^2 \rho(p)}{\partial \bar{\theta}^i \partial \bar{\theta}^j} \frac{\partial \tau(p)}{\partial \bar{\theta}^k} \right\} = \sum_{lmn} \frac{\partial \theta^i}{\partial \bar{\theta}^l} \frac{\partial \theta^j}{\partial \bar{\theta}^m} \frac{\partial \theta^k}{\partial \bar{\theta}^n} T_{lmn}(\theta) . \tag{71}$$

after substituting (69), (70) and (62). $T$ indeed transforms to $\bar{T}$ in a manner that defines a $(0,3)$-tensor. Therefore, the proposition is proven. $\square$

We now cast the Amari–Chentsov tensor $T$ in an alternative form that gives an explicit form of weighting function. Given $\rho, \tau$, because of Lemma 1, there exists another monotone embedding $\sigma$, such that $\sigma(\rho) = \tau$. Differentiating,

$$\frac{\partial \sigma(\rho(p))}{\partial \theta^i} = \sigma'(\rho(p)) \frac{\partial \rho(p)}{\partial \theta^i} . \tag{72}$$

Differentiate again, we obtain:

$$\frac{\partial^2 \sigma(\rho(p))}{\partial \theta^i \partial \theta^j} = \sigma''(\rho(p)) \frac{\partial \rho(p)}{\partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^i} + \sigma'(\rho(p)) \frac{\partial^2 \rho(p)}{\partial \theta^i \partial \theta^j} . \tag{73}$$

Substituting the above into (62), we obtain an expression of $T$ in terms of $\rho$ (which plays the role of embedding function) and $\sigma$ (which plays the role of weighting function):

$$T_{ijk}(\theta) = \mathrm{E}_\mu \left\{ \sigma''(\rho(p)) \frac{\partial \rho(p)}{\partial \theta^i} \frac{\partial \rho(p)}{\partial \theta^j} \frac{\partial \rho(p)}{\partial \theta^k} \right\} . \tag{74}$$

Similarly, we can obtain:

$$T_{ijk}(\theta) = -\mathrm{E}_\mu \left\{ (\sigma^{-1})''(\tau(p)) \frac{\partial \tau(p)}{\partial \theta^i} \frac{\partial \tau(p)}{\partial \theta^j} \frac{\partial \tau(p)}{\partial \theta^k} \right\} . \tag{75}$$

Therefore, under $\tau$-representation, $\sigma^{-1}$ (the inverse function of $\sigma$) serves as the weighting function. Note that $\sigma = f', \sigma^{-1} = (f^*)'$ when $\rho$ and $\tau$ are said to be conjugate. Furthermore, note the negative sign in (75) compared with (74); this precisely reflects "representation duality" with a $\rho \longleftrightarrow \tau$ exchange.

To summarize, because $\alpha$-geometry $\{\mathcal{M}, g, T\}$ is uniquely specified given a Riemannian metric $g$ and the Amari–Chentsov tensor $T$, the above derivations show that they both enjoy the freedom of two monotone/convex functions, with the freedom in specifying $g$ coupled to the freedom in specifying $T$ in the same way that the metric and connections are coupled via Codazzi relation for statistical manifolds. That the weighting functions used to construct linear, symmetric bilinear and totally symmetric trilinear functionals (on random functions) turns out to be $f'(\rho(\cdot))$, $f''(\rho(\cdot))$, $f'''(\rho(\cdot))$, respectively, is noteworthy. See [6] for more discussions.

### 2.4. Representation Duality versus Reference Duality

Going beyond extending $\alpha$-embedding to dual monotonic embeddings, Reference [2] illuminated two different senses of duality in the $\alpha$-geometry. Prior to [2], there have been several different usages of $\alpha$-parameter in Amari's theory of information geometry [10,11]:

(1) parameterizing the divergence functions ($\alpha$-divergences);

(2) parameterizing monotone embedding of probability functions ($\alpha$-embedding);

(3) parameterizing the convex mixture of connections ($\alpha$-connections).

Zhang (2004) [2] showed that (1) and (2) reflect two different types of duality in information geometry, with (1) concerning the reference/comparison status of a pair of points (functions) expressed in the divergence function ("reference duality") and (2) concerning their representation under arbitrary monotone scaling ("representation duality"). Both can lead to (3), the family of $\alpha$-connections. Therefore, care has to be taken in carefully delineating these two kinds of duality; for instance, the $\alpha\beta$-connection we derived in (21) reflects how reference duality and representation duality interacts in the alpha-connections.

The present analysis elaborated representation duality in information geometry by working out the freedom in allowing two (independently chosen) embedding functions $\rho, \tau$ or, equivalently, one embedding function $\rho$ along with a weighting function $f$, while the $(\rho, f)$ pair can be dually chosen to be the $(\tau, f^*)$ pair. Naudts' (2004) [3] $\phi$-logarithm is but a special case of the $(\rho, \tau)$ duality, in which $f'$ plays the role of the "integral-of-the-reciprocal" operation, that is taking the $\log$ of a function. This linkage then leads to $f^*$ and $\tau$ as inverse functions. The phenomena of biduality emerges when exchanging $\rho \longleftrightarrow \tau$ or $(\rho, f) \longleftrightarrow (\tau, f^*)$ leads to invariance of the Riemannian metric, but switches the two connections (the latter half of the statement is equivalent to changing signs of the Amari–Chentsov tensor). Therefore, the present paper, while elaborating the theory developed in [2], re-asserts the distinction between two distinct kinds of duality that was originally confounded in Amari's theory of $\alpha$-geometry, one through the freedom of selecting monotone embedding functions ("representation duality") and the other through the freedom of assigning referential status to points for pair comparison ("reference duality").

Finally, it is noted that the (bi)dualistic structure of the $(\rho, \tau)$-geometry (generalizing $\alpha$-geometry) is preserved in the non-parametric (infinite-dimensional) setting, as well [4,6], with the $\alpha$-connection structure cast in a more general way. Theorem 1 of [4] gives non-parametric expressions of the metric and connections under monotone embedding, mirroring the forms (14) and (15) in the parametric case.

## 3. Conclusion

The Riemannian metric with the pair of conjugate connections derived by Harsha and Moosath [1] are identical to the $(\rho, \tau)$-geometry obtained by Zhang in [2]. The $(\rho, \tau)$-embedding also recovers Naudts' deformed logarithm/exponential formulation. It is further shown in this paper that such $(\rho, \tau)$-geometry obtained is, when $\alpha$-embedding is relaxed to arbitrary monotone embeddings, the unique extension of Amari's $\alpha$-geometry in terms of its representational freedom.

## Acknowledgments

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Harsha, K.V; Subrahamanian Moosath, K.S. $F$-geometry and Amari's $\alpha$-geometry on a statistical manifold. *Entropy* **2014**, *16*, 2472–2487.
2. Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195.
3. Naudts, J. Estimators, escort probabilities, and $\phi$-exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
4. Zhang, J. Referential Duality and Representational Duality on Statistical Manifolds. In Proceedings of the Second International Symposium on Information Geometry and Its Applications, Tokyo, Japan, 12–16 December 2005; pp. 58–67.
5. Zhang, J. Referential duality and representational duality in the scaling of multi-dimensional and infinite-dimensional stimulus space. In *Measurement and Representation of Sensations: Recent Progress in Psychological Theory*; Dzhafarov, E., Colonius, H., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2006.
6. Zhang, J. Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. *Entropy* **2013**, *15*, 5384–5418.
7. Zhang, J. Divergence functions and geometric structures they induce on a manifold. In *Geometric Theory of Information*; Nielsen, F., Ed.; Springer: Cham, Switzerland, 2014; pp. 1–30.
8. Zhang, J. Reference duality and representation duality in information geometry. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt2014)*, Proceedings of 34th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Amboise, France, 21–26 September 2014; Volume 1641; pp. 130–146.
9. Amari, S. Differential geometry of curved exponential families—curvatures and information loss. *Ann. Stat.* **1982**, *10*, 357–385.
10. Amari, S. *Differential Geometric Methods in Statistics*; Lecture Notes in Statistics, Volume 28; Springer: New York, NY, USA, 1985.

11. Amari, S.; Nagaoka, H. *Method of Information Geometry*; Oxford University Press: Oxford, UK, 2000.

12. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **2007**, *370*, 184–193.

13. Naudts, J. Generalised exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149.

14. Ohara, A.; Matsuzoe, H.; Amari, S. A dually flat structure on the space of escort distributions. *J. Phys. Conf. Ser.* **2010**, *201*, 012012.

15. Amari. S.; Ohara, A. Geometry of q-exponential family of probability distributions. *Entropy* **2011**, *13*, 1170–1185.

16. Amari, S.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometry. *Physica A* **2012**, *391*, 4308–4319.

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

18. Eguchi, S. A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Math. J.* **1985**, *15*, 341–391.

19. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; American Mathematics Society: Providence, RI, USA, 1982.

20. Ay, N.; Jost, J.; Le, H.V.; Schwachhöfer, L. Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **2014**, doi: 10.1007/s00440-014-0574-8.

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

22. Burbea, J.; Rao, C.R. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *J. Multivar. Anal.* **1982**, *12*, 575–596.

23. Burbea, J.; Rao, C.R. Differential metrics in probability spaces. *Probab. Math. Stat.* **1984**, *3*, 241–258.