

Article

# Conformal Flattening for Deformed Information Geometries on the Probability Simplex <sup>†</sup>

Atsumi Ohara

Department of Electrical and Electronics, University of Fukui, Bunkyo, Fukui 910-8507, Japan; ohara@fuee.u-fukui.ac.jp

<sup>†</sup> This paper is an extended version of our paper published in SigmaPhi 2014, 2017 and Geometric Science of Information (GSI 2017).

Received: 20 February 2018; Accepted: 8 March 2018; Published: 10 March 2018

**Abstract:** Recent progress of theories and applications regarding statistical models with generalized exponential functions in statistical science is giving an impact on the movement to deform the standard structure of information geometry. For this purpose, various representing functions are playing central roles. In this paper, we consider two important notions in information geometry, i.e., invariance and dual flatness, from a viewpoint of representing functions. We first characterize a pair of representing functions that realizes the invariant geometry by solving a system of ordinary differential equations. Next, by proposing a new transformation technique, i.e., conformal flattening, we construct dually flat geometries from a certain class of non-flat geometries. Finally, we apply the results to demonstrate several properties of gradient flows on the probability simplex.

**Keywords:** representing functions; affine immersion; nonextensive statistical physics; invariance; dually flat structure; Legendre conjugate; gradient flow

---

## 1. Introduction

The theory of information geometry has elucidated abundant geometric properties equipped with a Riemannian metric and mutually dual affine connections. When it is applied to the study of statistical models described by the exponential family, the logarithmic function plays a significant role in giving the standard information geometric structure to the models [1,2].

Inspired by the recent progress of several areas in statistical physics and mathematical statistics [3–10] which have exploited theoretical interests and possible applications for generalized exponential families, one research direction in information geometry is pointing to constructions of deformed geometries based on the standard one, keeping its basic properties. A typical and classical example of such a deformation would be the alpha-geometry [1,2], a statistical definition of which can be regarded as a replacement of the logarithmic function by suitable power functions. Hence, for the purpose of the generalization and flexible applicability, much attention is paid to various uses of such replacements by *representing functions* as important tools [3,4,11,12].

Two major characteristics of the standard structure are dual flatness and invariance [2]. Dual flatness (or Hessian structure [13]) produces fruitful properties such as the existence of canonical coordinate systems, a pair of conjugate potential functions and the canonical divergence (relative entropy). In addition, they are connected with the Legendre duality relation, which is also fundamental in the generalization of statistical physics. On the other hand, the invariance of geometric structure is crucially valuable in developing mathematical statistics. It has been proved [14] that invariance holds for only the structure with a special triple of a Riemannian metric and a pair of mutually dual affine connections, which are respectively called the Fisher information and the alpha-connections (see Section 3 for their definitions). The study of these two characteristics from a viewpoint of representing functions would contribute to our geometrical understanding.

In this paper, we first characterize a pair of representing functions that realizes the invariant information geometric structure. Next, we propose a new transformation to obtain dually flat geometries from a certain class of non-flat information geometries, using concepts from affine differential geometry [15,16]. We call the transformation *conformal flattening*, which is a generalization of the way to realize the corresponding dually flat geometry from the alpha-geometry developed in [17,18]. As applications and easy consequences of the results, we finally show several properties of gradient flows associated with realized dually flat geometries. Focusing on geometric characteristics conserved by the transformation, we discuss the properties such as a relation between geodesics and flows, the first integral of the flows and so on. These properties are new and general. Hence, they refine the arguments of the flows in [18], where only the alpha-geometry is treated.

The paper is organized as follows. In Section 2, we introduce preliminary results, explaining several existing methods to construct the information geometric structure that includes a dually flat structure and the alpha-structure and so on. We also give a short summary of concepts from affine differential geometry, which will be used in this paper. Section 3 provides a characterization of representing functions that realize invariant geometry, i.e., the one equipped with the Fisher information and a pair of the alpha-connections. The characterization is obtained by solving a simple system of ordinary equations. In Section 4, we first obtain a certain class of information geometric structure by regarding representing functions as immersions into an ambient affine space. Then, we demonstrate the conformal flattening to realize the corresponding dually flat structure, and discuss their properties and relations with generalized entropies or escort probabilities [19]. Section 5 exhibits the geometric properties of gradient flows with respect to a conformally realized Riemannian metric. These flows are reduced to the well-known *replicator flow* [20] (Chapter 16) when we consider the standard information geometry. Suitably choosing its pay-off functions, we see that the flow follows a geodesic curve or conserves a divergence from an equilibrium. In the final section, some concluding remarks are made.

Throughout the paper, we use a probability simplex as a statistical model for the sake of simplicity.

## 2. Preliminaries

### 2.1. Information Geometry of $\mathcal{S}^n$ and $\mathbf{R}_+^{n+1}$

Let us represent an element  $p \in \mathbf{R}^{n+1}$  with its components  $p_i$ ,  $i = 1, \dots, n+1$  as  $p = (p_i) \in \mathbf{R}^{n+1}$ . Denote, respectively, the positive orthant by

$$\mathbf{R}_+^{n+1} := \{p = (p_i) \in \mathbf{R}^{n+1} | p_i > 0, i = 1, \dots, n+1\},$$

and the relative interior of the probability simplex by

$$\mathcal{S}^n := \left\{ p \in \mathbf{R}_+^{n+1} \left| \sum_{i=1}^{n+1} p_i = 1 \right. \right\}.$$

Let  $p(X)$  be a probability distribution of a random variable  $X$  taking a value in the finite sample space  $\Omega = \{1, 2, \dots, n, n+1\}$ . We consider a set of distributions  $p(X)$  with positive probabilities, i.e.,  $p(i) = p_i > 0$ ,  $i = 1, \dots, n+1$ , defined by

$$p(X) = \sum_{i=1}^{n+1} p_i \delta_i(X), \quad \delta_i(j) = \delta_i^j \text{ (the Kronecker's delta),}$$

which is identified with  $\mathcal{S}^n$ . A statistical model in  $\mathcal{S}^n$  is represented with parameters  $\zeta = (\zeta^j)$ ,  $j = 1, \dots, d \leq n$  by

$$p_\zeta(X) = \sum_{i=1}^{n+1} p_i(\zeta)\delta_i(X),$$

where each  $p_i$  is smoothly parametrized by  $\zeta$ . For such a statistical model,  $\zeta^j$  can also be regarded as coordinates of the corresponding submanifold in  $\mathcal{S}^n$ . For simplicity, we shall consider the full model, i.e.,  $d = n$  and the parameter set is bijective with  $\mathcal{S}^n$  via  $p_i(\zeta)$ 's.

The information geometric structure [2] on  $\mathcal{S}^n$  denoted by  $(g, \nabla, \nabla^*)$  is composed of the pair of mutually dual torsion-free affine connections  $\nabla$  and  $\nabla^*$  with respect to a Riemannian metric  $g$ . If we write  $\partial_i := \partial/\partial\zeta^i$ ,  $i = 1, \dots, n$ , the mutual duality requires components of  $(g, \nabla, \nabla^*)$  to satisfy

$$\partial_i g_{jk} = \Gamma_{ij,k} + \Gamma_{ik,j}^* \tag{1}$$

Let  $L$  and  $M$  be a pair of strictly monotone (i.e., one-to-one) smooth functions on the interval  $(0, 1)$ . One way of constructing such a structure  $(g, \nabla, \nabla^*)$  is to define the components as follows [2,11]:

$$g_{ij}(p) = \sum_{X \in \Omega} \partial_i L(p_\zeta(X)) \partial_j M(p_\zeta(X)), \quad i, j = 1, \dots, n, \tag{2}$$

$$\Gamma_{ij,k}(p) = \sum_{X \in \Omega} \partial_i \partial_j L(p_\zeta(X)) \partial_k M(p_\zeta(X)), \quad i, j, k = 1, \dots, n, \tag{3}$$

$$\Gamma_{ij,k}^*(p) = \sum_{X \in \Omega} \partial_k L(p_\zeta(X)) \partial_i \partial_j M(p_\zeta(X)), \quad i, j, k = 1, \dots, n. \tag{4}$$

In this paper, we call  $L$  and  $M$  representing functions. It is easy to verify the mutual duality (1). (Positive definiteness of  $g$  needs additional conditions.)

When the curvature tensors of both  $\nabla$  and  $\nabla^*$  vanish,  $(g, \nabla, \nabla^*)$  is called *dually flat* [2]. It is known that  $(g, \nabla, \nabla^*)$  is dually flat if and only if there exist two special coordinate systems denoted by  $\theta^i(p)$  and  $\eta_i(p)$ ,  $i = 1, \dots, n$ , respectively, where  $(\theta^i)$  is  $\nabla$ -affine,  $(\eta_i)$  is  $\nabla^*$ -affine and they are biorthogonal, i.e.,

$$g \left( \frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j} \right) = \delta_i^j.$$

We give examples. For a real number  $\alpha$ , define  $L^{(\alpha)}(u) := 2u^{(1-\alpha)/2}/(1-\alpha)$  and  $L^{(1)} := \ln u$ . If we set  $L(u) = L^{(\alpha)}(u)$  and  $M(u) = L^{(-\alpha)}(u)$ , then they derive the *alpha-structure* [2]  $(g^F, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ , where  $g^F$  is the Fisher information and  $\nabla^{(\pm\alpha)}$  are the alpha-connections (see Section 3). In particular, if we choose  $\alpha = 1$ , it defines the standard dually flat structure  $(g^F, \nabla^{(e)} := \nabla^{(1)}, \nabla^{(m)} := \nabla^{(-1)})$ , where  $\nabla^{(e)}$  and  $\nabla^{(m)}$  are called the e- and m-connection, respectively [2]. Similarly, the  $\phi$ -log geometry [3] can also be introduced in the same way by taking  $L(u) = \log_\phi(u)$  and  $M(u) = u$ .

One traditional way to construct a general information geometric structure  $(g, \nabla, \nabla^*)$ , without using representing functions, is by means of contrast functions (or divergences) [2,21]. In our case, let  $\rho$  be a function on  $\mathcal{S}^n \times \mathcal{S}^n$  satisfying  $\rho(p, r) \geq 0$ ,  $\forall p, r \in \mathcal{S}^n$  with equality if and only if  $p = r$ . For a vector field  $\partial_i$ , let  $(\partial_i)_p$  denote its tangent vector at  $p$ . When we define

$$g_{ij}(p) = -(\partial_i)_p (\partial_j)_r \rho(p, r) \Big|_{p=r}, \quad i, j = 1, \dots, n, \tag{5}$$

$$\Gamma_{ij,k}(p) = -(\partial_i)_p (\partial_j)_p (\partial_k)_r \rho(p, r) \Big|_{p=r}, \quad i, j, k = 1, \dots, n, \tag{6}$$

$$\Gamma_{ij,k}^*(p) = -(\partial_i)_p (\partial_j)_r (\partial_k)_r \rho(p, r) \Big|_{p=r}, \quad i, j, k = 1, \dots, n, \tag{7}$$

we can confirm that (1) holds. If  $g$  is positive definite, we say that  $\rho$  is a *contrast function* or a *divergence* that induces the structure  $(g, \nabla, \nabla^*)$ .

A contrast function  $\rho$  of the form:

$$\rho(p, r) = \psi(\theta(p)) + \varphi(\eta(r)) - \sum_{i=1}^n \theta^i(p)\eta_i(r) \tag{8}$$

always induces the corresponding dually flat structure. Conversely, it is known [2] that if  $(g, \nabla, \nabla^*)$  is dually flat, then there exists the unique contrast function of the form (8) that induces the structure. Hence, it is called the *canonical divergence* of  $(g, \nabla, \nabla^*)$  and we say that the functions  $\psi$  and  $\varphi$  are *potentials*. By setting  $p = r$ , we see that a dually flat structure naturally gives the *Legendre duality* relations at each  $p$ , i.e., the function  $\varphi$ , is the Legendre conjugate of  $\psi$  satisfying

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \varphi}{\partial \eta_i}.$$

Applying the idea of affine hypersurface theory [15] is also one of the other ways to construct the information geometric structure. Let  $D$  be the canonical flat affine connection on  $\mathbf{R}^{n+1}$ . Consider an immersion  $f$  from  $\mathcal{S}^n$  into  $\mathbf{R}^{n+1}$  and a vector field  $\zeta$  on  $\mathcal{S}^n$  that is transversal to the hypersurface  $f(\mathcal{S}^n)$  in  $\mathbf{R}^{n+1}$ . Such a pair  $(f, \zeta)$ , called an *affine immersion*, defines a torsion-free connection  $\nabla$  and the affine fundamental form  $g$  on  $\mathcal{S}^n$  via the Gauss formula as

$$D_X f_*(Y) = f_*(\nabla_X Y) + g(X, Y)\zeta, \quad X, Y \in \mathcal{X}(\mathcal{S}^n), \tag{9}$$

where  $\mathcal{X}(\mathcal{S}^n)$  is the set of tangent vector fields on  $\mathcal{S}^n$  and  $f_*$  denotes the differential of  $f$ . By regarding  $g$  as a (pseudo-) Riemannian metric, one can discuss the realized structure  $(g, \nabla)$  on  $\mathcal{S}^n$ .

We say that  $(f, \zeta)$  is *non-degenerate* and *equiaffine* if  $g$  is non-degenerate and  $D_X \zeta$  is tangent to  $\mathcal{S}^n$  for any  $X \in \mathcal{X}(\mathcal{S}^n)$ , respectively. The latter ensures that the volume element  $\theta$  on  $\mathcal{S}^n$  defined by

$$\theta(X_1, \dots, X_n) = \det(f_*(X_1), \dots, f_*(X_n), \zeta), \quad X_i \in \mathcal{X}(\mathcal{S}^n)$$

is parallel to  $\nabla$  [15] (p.31). It is known [15,16] that there exists a torsion-free dual affine connection  $\nabla^*$  satisfying (1) if and only if  $(f, \zeta)$  is non-degenerate and equiaffine. In this case, the obtained structure  $(g, \nabla, \nabla^*)$  on  $\mathcal{S}^n$  is not dually flat in general. However, there always exists a positive function  $\sigma$  and a dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  on  $\mathcal{S}^n$  that hold the following relations [16]:

$$\tilde{g} = \sigma g, \tag{10}$$

$$g(\tilde{\nabla}_X Y, Z) = g(\nabla_X Y, Z) - d(\ln \sigma)(Z)g(X, Y), \tag{11}$$

$$g(\tilde{\nabla}_X^* Y, Z) = g(\nabla_X^* Y, Z) + d(\ln \sigma)(X)g(Y, Z) + d(\ln \sigma)(Y)g(X, Z). \tag{12}$$

Furthermore, there exists a specific contrast function  $\rho(p, r)$  for  $(g, \nabla, \nabla^*)$  called the *geometric divergence*. Then, a contrast function  $\tilde{\rho}(p, r)$  that induces  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  is given by the *conformal divergence*  $\tilde{\rho}(p, r) = \sigma(r)\rho(p, r)$ . These properties of the structure  $(g, \nabla)$  realized by the non-degenerate and equiaffine immersion are called *1-conformal flatness* [16].

### 3. Characterization of Invariant Geometry by Representing Functions

Suppose that a pair of representing functions  $(L, M)$  defines an information geometric structure  $(g, \nabla, \nabla^*)$  by (2), (3) and (4). In this section, we consider the condition of  $(L, M)$  such that  $(g, \nabla, \nabla^*)$  is invariant. This is equivalent [2,14] to  $g$  which is the Fisher information  $g^F$  defined by

$$g_{ij}^F(p) = \sum_{X \in \Omega} p_X (\partial_i \ln p_X) (\partial_j \ln p_X) \tag{13}$$

and a pair of dual connections satisfies  $\nabla = \nabla^{(\alpha)}$  and  $\nabla^* = \nabla^{(-\alpha)}$  for a certain  $\alpha \in \mathbf{R}$ , where  $\nabla^{(\alpha)}$  is the  $\alpha$ -connection defined by

$$\Gamma_{ij,k}^{(\alpha)} = \sum_{x \in \Omega} p_\zeta \left( \partial_i \partial_j \ln p_\zeta + \frac{1-\alpha}{2} (\partial_i \ln p_\zeta) (\partial_j \ln p_\zeta) \right) (\partial_k \ln p_\zeta). \tag{14}$$

Hence,  $g_{ij}$  expressed in (2) by functions  $L(u)$  and  $M(u)$  coincides with the Fisher information if and only if the following equation holds:

$$\frac{dL}{du} \frac{dM}{du} = 1/u. \tag{15}$$

Similarly, we derive a condition for  $\Gamma_{ij,k}$  expressed in (3) to be the  $\alpha$ -connection. First, note that the following relations hold:

$$\partial_i \ln p_\zeta = (\partial_i p_\zeta) \frac{1}{p_\zeta}, \quad \partial_i \partial_j \ln p_\zeta = (\partial_i \partial_j p_\zeta) \frac{1}{p_\zeta} - (\partial_i p_\zeta) (\partial_j p_\zeta) \frac{1}{p_\zeta^2}. \tag{16}$$

On the other hand, we have

$$\partial_i \partial_j L(p_\zeta) = (\partial_i \partial_j p_\zeta) \frac{dL}{du}(p_\zeta) + (\partial_i p_\zeta) (\partial_j p_\zeta) \frac{d^2L}{du^2}(p_\zeta), \tag{17}$$

$$\partial_k M(p_\zeta) = (\partial_k p_\zeta) \frac{dM}{du}(p_\zeta). \tag{18}$$

Substituting (16), (17) and (18) into (3) and (14), and comparing them, we obtain (15) again and

$$\frac{d^2L}{du^2} \frac{dM}{du} = -\frac{1+\alpha}{2u^2}. \tag{19}$$

Expressing  $L' := dL/du$  and  $L'' := d^2L/du^2$ , we have the following ODE from (15) and (19):

$$\frac{L''}{L'} = -\frac{1+\alpha}{2u}. \tag{20}$$

By integrations, we get

$$\ln L' = -\frac{(1+\alpha)}{2} \ln u + c \tag{21}$$

and

$$L(u) = c_1 u^{(1-\alpha)/2} + c_2, \quad M(u) = c_3 u^{(1+\alpha)/2} + c_4, \tag{22}$$

where  $c$  and  $c_i$ ,  $i = 1, \dots, 4$  are constants with a constraint  $c_1 c_3 = 4/(1-\alpha^2)$ . Thus,  $(L, M)$  is essentially a pair of representing functions that derives the alpha-geometry and there is only freedom of adjusting the constants for the invariance of geometry. If we require solely (15), which implies that only a Riemannian metric  $g$  is the Fisher information  $g^F$ , there still remains much freedom for  $(L, M)$ .

#### 4. Affine Immersion of the Probability Simplex

Now we consider the affine immersion with the following assumptions.

**Assumptions:**

1. The affine immersion  $(f, \zeta)$  is nondegenerate and equiaffine,
2. The immersion  $f$  is given by the component-by-component and common representing function  $L$ , i.e.,

$$f : \mathcal{S}^n \ni p = (p_i) \mapsto x = (x^i) \in \mathbf{R}^{n+1}, \quad x^i = L(p_i), \quad i = 1, \dots, n+1,$$

3. The representing function  $L : (0, 1) \rightarrow \mathbf{R}$  is sign-definite, concave with  $L'' < 0$  and strictly increasing, i.e.,  $L' > 0$ . Hence, the inverse of  $L$  denoted by  $E$  exists, i.e.,  $E \circ L = \text{id}$ .
4. Each component of  $\zeta$  satisfies  $\zeta^i < 0$ ,  $i = 1, \dots, n+1$  on  $\mathcal{S}^n$ .

**Remark 1.** From the assumption 3, it follows that  $L'E' = 1$ ,  $E' > 0$  and  $E'' > 0$ . Regarding sign-definiteness of  $L$ , note that we can adjust  $L(u)$  to  $L(u) + c$  by a suitable constant  $c$  without loss of generality since the resultant geometric structure is unchanged (See Theorem 1) by the adjustment. For a fixed  $L$  satisfying the assumption 3, we can choose  $\xi$  that meets the assumptions 1 and 4. For example, if we take  $\xi^i = -|L(p_i)|$ , then  $(f, \xi)$  is called centro-affine, which is known to be equiaffine [15] (p.37). The assumptions 3 and 4 also assure positive definiteness of  $g$  (the details are described in the proof of Theorem 1). Hence,  $(f, \xi)$  is non-degenerate and we can regard  $g$  as a Riemannian metric on  $S^n$ .

4.1. Conormal Vector and the Geometric Divergence

Define a function  $\Psi$  on  $\mathbf{R}^{n+1}$  by

$$\Psi(x) := \sum_{i=1}^{n+1} E(x^i),$$

then  $f(S^n)$  immersed in  $\mathbf{R}^{n+1}$  is expressed as a level surface of  $\Psi(x) = 1$ . Denote by  $\mathbf{R}_{n+1}$  the dual space of  $\mathbf{R}^{n+1}$  and by  $\langle v, x \rangle$  the pairing of  $x \in \mathbf{R}^{n+1}$  and  $v \in \mathbf{R}_{n+1}$ . The conormal vector [15] (p.57)  $v : S^n \rightarrow \mathbf{R}_{n+1}$  for the affine immersion  $(f, \xi)$  is defined by

$$\langle v(p), f_*(X) \rangle = 0, \quad \forall X \in T_p S^n, \quad \langle v(p), \xi(p) \rangle = 1, \tag{23}$$

for  $p \in S^n$ . Using the assumptions and noting the relations:

$$\frac{\partial \Psi}{\partial x^i} = E'(x^i) = \frac{1}{L'(p_i)} > 0, \quad i = 1, \dots, n + 1,$$

we have

$$v_i(p) := \frac{1}{\Lambda} \frac{\partial \Psi}{\partial x^i} = \frac{1}{\Lambda(p)} E'(x^i) = \frac{1}{\Lambda(p)} \frac{1}{L'(p_i)}, \quad i = 1, \dots, n + 1, \tag{24}$$

where  $\Lambda$  is a normalizing factor defined by

$$\Lambda(p) := \sum_{i=1}^{n+1} \frac{\partial \Psi}{\partial x^i} \xi^i = \sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \xi^i(p). \tag{25}$$

Then, we can confirm (23) using the relation  $\sum_{i=1}^{n+1} X^i = 0$  for  $X = (X^i) \in \mathcal{X}(S^n)$ . Note that  $v : S^n \rightarrow \mathbf{R}_{n+1}$  defined by

$$v_i(p) = \Lambda(p)v_i(p) = \frac{1}{L'(p_i)}, \quad i = 1, \dots, n + 1,$$

also satisfies

$$\langle v(p), f_*(X) \rangle = 0, \quad \forall X \in T_p S^n. \tag{26}$$

Furthermore, it follows, from (24), (25) and the assumption 4, that

$$\Lambda(p) < 0, \quad v_i(p) < 0, \quad i = 1, \dots, n + 1,$$

for all  $p \in S^n$ .

It is known [15] (p.57) that the affine fundamental form  $g$  can be represented by

$$g(X, Y) = -\langle v_*(X), f_*(Y) \rangle, \quad X, Y \in T_p S^n.$$

In our case, it is calculated via (26) as

$$\begin{aligned}
 g(X, Y) &= -\Lambda^{-1} \langle v_*(X), f_*(Y) \rangle - X(\Lambda^{-1}) \langle v, f_*(Y) \rangle \\
 &= -\frac{1}{\Lambda} \sum_{i=1}^{n+1} \left( \frac{1}{L'(p_i)} \right)' L'(p_i) X^i Y^i = \frac{1}{\Lambda} \sum_{i=1}^{n+1} \frac{L''(p_i)}{L'(p_i)} X^i Y^i, \quad X, Y \in T_p \mathcal{S}^n. \tag{27}
 \end{aligned}$$

Hence,  $g$  is positive definite from the assumptions 3 and 4, and we can regard it as a Riemannian metric.

Utilizing these notions from affine differential geometry, we can introduce a geometric divergence [16] as follows:

$$\begin{aligned}
 \rho(p, r) &= \langle v(r), f(p) - f(r) \rangle = \sum_{i=1}^{n+1} v_i(r) (L(p_i) - L(r_i)) \\
 &= \frac{1}{\Lambda(r)} \sum_{i=1}^{n+1} \frac{L(p_i) - L(r_i)}{L'(r_i)}, \quad p, r \in \mathcal{S}^n. \tag{28}
 \end{aligned}$$

It is easily checked that  $\rho$  is actually a contrast function of the 1-conformally flat structure  $(g, \nabla, \nabla^*)$  using (5), (6) and (7).

#### 4.2. Conformal Flattening Transformation

As is described in the preliminary section, by 1-conformally flatness there exists a positive function, i.e., conformal factor  $\sigma$  that relates  $(g, \nabla, \nabla^*)$  with a dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  via the conformal transformation (10), (11) and (12). A contrast function  $\tilde{\rho}$  that induces  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  is given as the conformal divergence:

$$\tilde{\rho}(p, r) = \sigma(r) \rho(p, r), \quad p, r \in \mathcal{S}^n. \tag{29}$$

from the geometric divergence  $\rho$  in (28).

For an arbitrary function  $L$  within our setting given by the four assumptions, we prove that we can construct a dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  by choosing the conformal factor  $\sigma$  carefully. Hereafter, we call this transformation *conformal flattening*.

Define

$$Z(p) := \sum_{i=1}^{n+1} v_i(p) = \frac{1}{\Lambda(p)} \sum_{i=1}^{n+1} \frac{1}{L'(p_i)},$$

then it is negative because each  $v_i(p)$  is negative. The conformal divergence to  $\rho$  with respect to the conformal factor  $\sigma(r) := -1/Z(r)$  is

$$\tilde{\rho}(p, r) = -\frac{1}{Z(r)} \rho(p, r).$$

**Theorem 1.** *If the conformal factor is  $\sigma = -1/Z$ , then the information geometric structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  on  $\mathcal{S}^n$  that is transformed from the 1-conformally flat structure  $(g, \nabla, \nabla^*)$  via (10), (11) and (12) is dully flat.*

Furthermore, the conformal divergence  $\bar{\rho}$  that induces  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  on  $\mathcal{S}^n$  is canonical where Legendre conjugate potential functions and coordinate systems are explicitly given by

$$\theta^i(p) = x^i(p) - x^{n+1}(p) = L(p_i) - L(p_{n+1}), \quad i = 1, \dots, n, \tag{30}$$

$$\eta_i(p) = P_i(p) := \frac{v_i(p)}{Z(p)} = \frac{1/L'(p_i)}{\sum_{k=1}^{n+1} 1/L'(p_k)}, \quad i = 1, \dots, n, \tag{31}$$

$$\psi(p) = -x_{n+1}(p) = -L(p_{n+1}), \tag{32}$$

$$\varphi(p) = \frac{1}{Z(p)} \sum_{i=1}^{n+1} v_i(p)x^i(p) = \sum_{i=1}^{n+1} P_i(p)L(p_i). \tag{33}$$

**Proof.** Using given relations, we first show that the conformal divergence  $\bar{\rho}$  is the canonical divergence for  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$ :

$$\begin{aligned} \bar{\rho}(p, r) &= -\frac{1}{Z(r)} \langle v(r), f(p) - f(r) \rangle = \langle P(r), f(r) - f(p) \rangle \\ &= \sum_{i=1}^{n+1} P_i(r)(x^i(r) - x^i(p)) \\ &= \sum_{i=1}^{n+1} P_i(r)x^i(r) - \sum_{i=1}^n P_i(r)(x^i(p) - x^{n+1}(p)) - \left( \sum_{i=1}^{n+1} P_i(r) \right) x^{n+1}(p) \\ &= \varphi(r) - \sum_{i=1}^n \eta_i(r)\theta^i(p) + \psi(p). \end{aligned} \tag{34}$$

Next, let us confirm that  $\partial\psi/\partial\theta^i = \eta_i$ .

Since  $\theta^i(p) = L(p_i) + \psi(p)$ ,  $i = 1, \dots, n$ , we have

$$p_i = E(\theta^i - \psi), \quad i = 1, \dots, n + 1,$$

by setting  $\theta^{n+1} := 0$ . Hence, we have

$$1 = \sum_{i=1}^{n+1} E(\theta^i - \psi).$$

Differentiating by  $\theta^j$ , we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta^j} \sum_{i=1}^{n+1} E(\theta^i - \psi) = \sum_{i=1}^{n+1} E'(\theta^i - \psi) \left( \delta_j^i - \frac{\partial\psi}{\partial\theta^j} \right) \\ &= E'(x^j) - \left( \sum_{i=1}^{n+1} E'(x^i) \right) \frac{\partial\psi}{\partial\theta^j}. \end{aligned}$$

This implies that

$$\frac{\partial\psi}{\partial\theta^j} = \frac{E'(x^j)}{\sum_{i=1}^{n+1} E'(x^i)} = \eta_j.$$

Together with (34) and this relation,  $\varphi$  is confirmed to be the Legendre conjugate of  $\psi$ .

The dual relation  $\partial\varphi/\partial\eta_i = \theta^i$  follows automatically from the property of the Legendre transform.  $\square$

The following corollary is straightforward because all the quantities in the theorem depend on only  $L$ :



**Corollary 1.** Under the assumptions, the dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  on  $\mathcal{S}^n$ , obtained by following the above conformal flattening, does not depend on the choice of the transversal vector  $\xi$ .

**Remark 2.** Note that the conformal metric is given by  $\tilde{g} = -g/Z$  and is positive definite. Furthermore, the relation (12) means that the dual affine connections  $\nabla^*$  and  $\tilde{\nabla}^*$  are projectively (or -1-conformally) equivalent [15,16]. Hence,  $\nabla^*$  is projectively flat. Furthermore, the above corollary implies that the realized affine connection  $\nabla$  is also projectively equivalent to the flat connection  $\tilde{\nabla}$  if we use the centro-affine immersion, i.e.,  $\tilde{\xi}^i = -L(p_i)$  [15,16]. See Proposition 3 for an application of projective equivalence of affine connections.

**Remark 3.** In our setting, conformal flattening is geometrically regarded as normalization of the conormal vector  $v$ . Hence, the dual coordinates  $\eta_i(p) = P_i(p)$  can be interpreted as a generalization of the escort probability [10,19] (see the following example). Similarly,  $\psi$  and  $-\varphi$  might be seen as the associated Massieu function and entropy, respectively.

**Remark 4.** While the immersion  $f$  is composed of a representing function  $L$  under the assumption 2, the corresponding  $M$  of a single variable does not generally exist for  $(g, \nabla, \nabla^*)$  nor  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$ . From the expressions of the Riemann metrics  $g$  in (27) and  $\tilde{g} = -g/Z$ , we see that the counterparts of the representing functions  $M(p_i)$  would be, respectively,  $-v_i(p)$  and  $P_i(p)$ , but note that they are multi-variable functions of  $p = (p_i)$ .

### 4.3. Examples

If we take  $L$  to be the logarithmic function  $L(t) = \ln(t)$ , then the conformally flattened geometry immediately defines the standard dually flat structure  $(g^F, \nabla^{(1)}, \nabla^{(-1)})$  on the simplex  $\mathcal{S}^n$ . We see that  $-\varphi(p)$  is the entropy, i.e.,  $\varphi(p) = \sum_{i=1}^{n+1} p_i \ln p_i$  and the conformal divergence is the KL divergence (relative entropy), i.e.,  $\tilde{\rho}(p, r) = D^{(KL)}(r||p) = \sum_{i=1}^{n+1} r_i (\ln r_i - \ln p_i)$ .

Next, let the affine immersion  $(f, \xi)$  be defined by the following  $L$  and  $\xi$ :

$$L(t) := \frac{1}{1-q} t^{1-q}, \quad x^i(p) = \frac{1}{1-q} (p_i)^{1-q},$$

and

$$\xi^i(p) = -q(1-q)x^i(p),$$

with  $0 < q$  and  $q \neq 1$ . We see that the immersion is centro-affine scaled by the constant factor  $q(1-q)$ . Then, we see that the immersion realizes the alpha-structure  $(g^F, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  on  $\mathcal{S}^n$  with  $q = (1 + \alpha)/2$ . The geometric divergence is the alpha-divergence, i.e.,

$$\rho(p, r) = \frac{4}{1-\alpha^2} \left( 1 - \sum_{i=1}^{n+1} (p_i)^{(1-\alpha)/2} (r_i)^{(1+\alpha)/2} \right).$$

Following the procedure of conformal flattening described in the above, we have [17]

$$\Psi(x) = \sum_{i=1}^{n+1} ((1-q)x^i)^{1/1-q}, \quad \Lambda(p) = -q, \text{ (constant)}$$

$$v_i(p) = -\frac{1}{q} (p_i)^q, \quad \sigma(p) = -\frac{1}{Z(p)} = \frac{q}{\sum_{k=1}^{n+1} (p_k)^q},$$

and obtain a dually flat structure  $(\tilde{g}^F, \tilde{\nabla}, \tilde{\nabla}^*)$  via the formulas in Theorem 1:

$$\eta_i = P_i = \frac{(p_i)^q}{\sum_{k=1}^{n+1} (p_k)^q}, \quad \theta^i = \frac{1}{1-q} (p_i)^{1-q} - \frac{1}{1-q} (p_{n+1})^{1-q} = \ln_q(p_i) - \psi(p),$$

$$\psi(p) = -\ln_q(p_{n+1}), \quad \varphi(p) = \ln_q\left(\frac{1}{\exp_q(S_q(p))}\right), \quad \tilde{g}^F = -\frac{1}{Z(p)}g^F.$$

Here,  $\ln_q$  and  $S_q(p)$  are the  $q$ -logarithmic function and the Tsallis entropy [10], respectively, defined by

$$\ln_q(t) = \frac{t^{1-q} - 1}{1 - q}, \quad S_q(p) = \frac{\sum_{i=1}^{n+1} (p_i)^q - 1}{1 - q}.$$

Note that the escort probability appears as the dual coordinate  $\eta_i$ .

### 5. An Application to Gradient Flows on $S^n$

Recall the replicator flow on the simplex  $S^n$  for given functions  $f_i(p)$  defined by

$$\dot{p}_i = p_i(f_i(p) - \bar{f}(p)), \quad i = 1, \dots, n + 1, \quad \bar{f}(p) := \sum_{i=1}^{n+1} p_i f_i(p), \tag{35}$$

which is extensively studied in evolutionary game theory. It is known [20] (Chapter 16) that

- (i) the solution to (35) is the gradient flow that maximizes a function  $V(p)$  satisfying

$$f_i = \frac{\partial V}{\partial p_i}, \quad i = 1, \dots, n + 1, \tag{36}$$

with respect to the Shahshahani metric  $g^S$  (See below),

- (ii) the KL divergence is a local Lyapunov function for an equilibrium called the evolutionary stable state (ESS) for the case of  $f_i(p) = \sum_{j=1}^{n+1} a_{ij} p_j$  with  $(a_{ij}) \in \mathbf{R}^{(n+1) \times (n+1)}$ .

The Shahshahani metric  $g^S$  is defined on the positive orthant  $\mathbf{R}_+^{n+1}$  by

$$g_{ij}^S(p) = \frac{\sum_{k=1}^{n+1} p_k}{p_i} \delta_{ij}, \quad i, j = 1, \dots, n + 1.$$

Note that the Shahshahani metric induces the Fisher metric  $g^F$  on  $S^n$ . Further, the KL divergence is the canonical divergence [2] of  $(g^F, \nabla^{(1)}, \nabla^{(-1)})$ . Thus, the replicator dynamics (35) are closely related with the standard dually flat structure  $(g^F, \nabla^{(1)}, \nabla^{(-1)})$ , which associates with exponential and mixture families of probability distributions. In addition, investigation of the flow is also important from a viewpoint of statistical physics governed by the Boltzmann–Gibbs distributions when we choose  $V(p)$  as various physical quantities, e.g., free energy or entropy.

Similarly, when we consider various Legendre relations deformed by  $L$ , it would be of interest to investigate gradient flows on  $S^n$  for a dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  or a 1-conformally flat structure  $(g, \nabla, \nabla^*)$ . Since  $g$  and  $\tilde{g}$  can be naturally extended to  $\mathbf{R}_+^{n+1}$  as a diagonal form (we use the same notation for brevity):

$$g_{ij}(p) = \frac{1}{\Lambda(p)} \frac{L''(p_i)}{L'(p_i)} \delta_{ij}, \quad \tilde{g}_{ij}(p) = -\frac{1}{Z(p)} g_{ij}(p), \quad i, j = 1, \dots, n + 1 \tag{37}$$

from (27), we can define two gradient flows for  $V(p)$  on  $S^n$ . One is the gradient flow for  $g$ , which is

$$\dot{p}_i = g_{ii}^{-1}(f_i - \bar{f}^H), \quad \bar{f}^H(p) := \sum_{k=1}^{n+1} H_k(p) f_k(p), \quad H_i(p) := \frac{g_{ii}^{-1}(p)}{\sum_{k=1}^{n+1} g_{kk}^{-1}(p)}, \tag{38}$$

for  $i = 1, \dots, n + 1$ . It is verified that  $\dot{p}$  is tangent to  $S^n$ , i.e.,  $\dot{p} \in T_p S^n$  and gradient of  $V$ , i.e.,

$$g(X, \dot{p}) = \sum_{i=1}^{n+1} f_i X^i - \bar{f}^H \sum_{i=1}^{n+1} X^i = \sum_{i=1}^{n+1} \frac{\partial V}{\partial p_i} X^i, \quad \forall X = (X^i) \in \mathcal{X}(S^n).$$

In the same way, the other one for  $\tilde{g}$  is defined by

$$\dot{p}_i = \tilde{g}_{ii}^{-1}(f_i - \tilde{f}^H), \quad \tilde{f}^H(p) := \sum_{k=1}^{n+1} H_k(p)f_k(p), \quad i = 1, \dots, n + 1. \tag{39}$$

Note that both the flows reduce to (35) when  $L = \ln$ .

From (37), the following consequence is immediate:

**Proposition 1.** *The trajectories of the gradient flow (38) and (39) starting from the same initial point coincide while velocities of time-evolutions are different by the factor- $Z(p)$ .*

Taking account of the example with respect to the alpha-geometry and the conformally flattened one given in subsection 4.3, the following result shown in [18] can be regarded as a corollary of the above proposition:

**Corollary 2.** *The trajectories of the gradient flow (39) with respect to the conformal metric  $\tilde{g}$  for  $L(t) = t^{1-q}/(1-q)$  coincide with those of the replicator flow (35) while velocities of time-evolutions are different by the factor- $Z(p)$ .*

Next, we particularly consider the case when  $V(p)$  is a potential function or divergences. As for a gradient flow on a manifold equipped with a dually flat structure  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$ , the following result is known:

**Proposition 2.** [22] *Consider the potential function  $\psi(p)$  and the canonical divergence  $\tilde{\rho}(p, r)$  of  $(\tilde{g}, \tilde{\nabla}, \tilde{\nabla}^*)$  for an arbitrary prefixed point  $r$ . The gradient flows for  $V(p) = \pm\psi(p)$  and  $V(p) = \pm\rho(p, r)$  follow  $\tilde{\nabla}^*$ -geodesic curves.*

As is described in Remark 2,  $\nabla^*$  and  $\tilde{\nabla}^*$  are projectively equivalent. One geometrically interesting property of the projective equivalence is that  $\nabla^*$ - and  $\tilde{\nabla}^*$ - geodesic curves coincide up to their parametrizations (i.e., a curve is  $\nabla^*$ -pregeodesic if and only if it is  $\tilde{\nabla}^*$ -pregeodesic) [15] (p.17). Combining this fact with Propositions 1 and 2, we see that the following result holds:

**Proposition 3.** *Let  $r \in S^n$  be an arbitrary prefixed point. The gradient flows (38) for  $V(p) = \pm\rho(p, r) = \pm\tilde{\rho}(p, r)/\sigma(r)$ ,  $V(p) = \pm\tilde{\rho}(p, r)$  and  $V(p) = \pm\psi(p)$  follow  $\tilde{\nabla}^*$ -geodesic curves.*

Finally, we demonstrate here another aspect of the flow (39). Let us particularly consider the following functions  $f_i$ :

$$f_i(p) := \frac{L''(p_i)}{(L'(p_i))^2} \sum_{j=1}^{n+1} a_{ij}P_j(p), \quad a_{ij} = -a_{ji} \in \mathbf{R}, \quad i, j = 1, \dots, n + 1. \tag{40}$$

Note that  $f_i$ s are not integrable, i.e., non-trivial  $V$  satisfying (36) does not exist because of the anti-symmetry of  $a_{ij}$ . Hence, for this case, (39) is no longer a gradient flow. However, we can prove the following result:

**Theorem 2.** *Consider the flow (39) with the functions  $f_i$ s defined in (40) and assume that there exists an equilibrium  $r \in S^n$  for the flow. Then,  $\rho(p, r)$  and  $\tilde{\rho}(p, r)$  are the first integral (conserved quantity) of the flow.*

**Proof.** By substituting (40) into  $\tilde{f}^H(p)$  in (39) and using the expression of  $\tilde{g}_{ii}$  in (37), we have

$$\tilde{f}^H(p) = \frac{1}{\sum_{k=1}^{n+1} L'(p_k)/L''(p_k)} \sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \sum_{j=1}^{n+1} a_{ij}P_i(p).$$

By the relation  $E'(x_i) = 1/L'(p_i)$  and (31), it holds that

$$\sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \sum_{j=1}^{n+1} a_{ij} P_i(p) = \frac{1}{\sum_{k=1}^{n+1} E'(x_k)} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a_{ij} E'(x_i) E'(x_j) = 0.$$

Hence, we see that  $\tilde{f}^H = 0$  and the flow (39) reduces to

$$\dot{p}_i = \tilde{g}_{ii}^{-1}(p) f_i(p), \quad i = 1, \dots, n + 1. \tag{41}$$

Since  $r$  is an equilibrium point, we see from (40) that

$$\sum_{j=1}^{n+1} a_{ij} P_j(r) = 0, \quad i = 1, \dots, n + 1. \tag{42}$$

Then, using (34), (41) and (42), we have

$$\begin{aligned} \frac{d\tilde{\rho}(p, r)}{dt} &= - \sum_{i=1}^{n+1} P_i(r) L'(p_i) \dot{p}_i \\ &= Z(p) \Lambda(p) \sum_{i=1}^{n+1} P_i(r) \frac{(L'(p_i))^2}{L''(p_i)} f_i(p) = Z(p) \Lambda(p) \sum_{i=1}^{n+1} P_i(r) \left( \sum_{j=1}^{n+1} a_{ij} P_j(p) \right) \\ &= Z(p) \Lambda(p) \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} (P_i(p) - P_i(r)) a_{ij} P_j(p) \\ &= Z(p) \Lambda(p) \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} (P_i(p) - P_i(r)) a_{ij} (P_j(p) - P_j(r)) = 0. \end{aligned}$$

Thus,  $\tilde{\rho}(p, r)$  is the first integral of the flow. It follows that  $\rho(p, r)$  is also the first integral of the flow from the definition of conformal divergence (29).  $\square$

**Remark 5.** From proposition 1, the same statement holds for the flow (38). The proposition implies the fact [20] that the KL divergence is the first integral for the replicator flow (35) with the function  $f_i(p)$  in (40) defined by  $L(t) = \ln t$  and  $P_j(p) = p_j$ .

### 6. Conclusions

We have considered two important aspects of information geometric structure, i.e., invariance and dual flatness, from a viewpoint of representing functions. As for the invariance of geometry, we have proved that a pair of representing functions that derives the alpha-structure is essentially unique. On the other hand, we have shown the explicit formula of conformal flattening that transforms 1-conformally flat structures on the simplex  $\mathcal{S}^n$  realized by affine immersions to the corresponding dually flat structures. Finally, we have discussed several geometric properties of gradient flows associated to two structures.

Presently, our analysis is restricted to the probability simplex, i.e., the space of discrete probability distributions. For the continuous case, the similar or related results are obtained in [23,24] without using affine immersions. Extensions of the results obtained in this paper to continuous probability space and the exploitation of relations to the literature are left for future work.

The conformal flattening can also be applied to the computationally efficient construction of a Voronoi diagram with respect to the geometric divergences [18]. Exploring the possibilities of other applications would be of interest.

**Acknowledgments:** Part of the results is adapted and reprinted with permission from Springer Customer Service Centre GmbH (licence No: 4294160782766): Springer Nature, *Geometric Science of Information* LNCS 10589 Nielsen, F., Barbaresco, F., Eds., (Article Name:) On affine immersions of the probability simplex and their

conformal flattening, (Author:) A. Ohara, (Copyright:) Springer International Publishing AG 2017 [25]. The author is partially supported by JSPS Grant-in-Aid (C) 15K04997.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Amari, S.I. *Differential-Geometrical Methods in Statistics*; Lecture Notes in Statistics Series 28; Springer: New York, NY, USA, 1985.
2. Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; Translations of Mathematical Monographs Series 191; AMS & Oxford University Press: Oxford, UK, 2000.
3. Naudts, J. Continuity of a class of entropies and relative entropies. *Rev. Math. Phys.* **2004**, *16*, 809.
4. Eguchi, S. Information geometry and statistical pattern recognition. *Sugaku Expos.* **2006**, *19*, 197.
5. Grünwald, P.D.; Dawid, A.P. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Ann. Statist.* **2004**, *32*, 1367.
6. Fujisawa, H.; Eguchi, S. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* **2008**, *99*, 2053.
7. Naudts, J. The  $q$ -exponential family in statistical Physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405.
8. Naudts, J. *Generalized thermostatics*; Springer: Berlin, Germany, 2010.
9. Ollila, E.; Tyler, D.; Koivunen, V.; Poor, V. Complex elliptically symmetric distributions : Survey, new results and applications. *IEEE Trans. Sig. Proc.* **2012**, *60*, 5597.
10. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: Berlin, Germany, 2009.
11. Zhang, J. Divergence Function, Duality, and Convex Analysis. *Neural Comput.* **2004**, *16*, 159.
12. Wada, T.; Matsuzoe, H. Conjugate representations and characterizing escort expectations in information geometry. *Entropy* **2017**, *19*, 309.
13. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
14. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; AMS: Providence, RI, USA, 1982.
15. Nomizu, K.; Sasaki, T. *Affine Differential Geometry*; Cambridge University Press: Cambridge, UK, 1993.
16. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J.* **1994**, *46*, 427.
17. Ohara, A.; Matsuzoe, H.; Amari, S.I. A dually flat structure on the space of escort distributions. *J. Phys. Conf. Ser.* **2010**, *201*, 012012.
18. Ohara, A.; Matsuzoe, H.; Amari, S.I. Conformal geometry of escort probability and its applications. *Mod. Phys. Lett. B* **2012**, *26*, 1250063.
19. Tsallis, C.; Mendes, M.S.; Plastino, A.R. The role of constraints within generalized nonextensive statistics. *Physica A* **1998**, *261*, 534.
20. Hofbauer, J.; Sigmund, K. *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection*; Cambridge University Press: Cambridge, UK, 1988.
21. Eguchi, S. Geometry of minimum contrast. *Hiroshima Math. J.* **1992**, *22*, 631.
22. Fujiwara, A.; Amari, S.I. Gradient systems in view of information geometry. *Physica D* **1995**, *80*, 317.
23. Amari, S.I.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. *Physica A* **2012**, *391*, 4308.
24. Matsuzoe, H. Hessian structures on deformed exponential families and their conformal structures. *Diff. Geo. Appl.* **2014**, *35*, 323.
25. Ohara, A. On affine immersions of the probability simplex and their conformal flattening. In *Geometric Science of Information*; Nielsen, F., Barbaresco, F., Eds.; Springer: Berlin, Germany, 2017.

