Conformal flattening on the probability simplex and its applications to Voronoi partitions and centroids

Atsumi Ohara

Department of Electrical and Electronics Engineering, University of Fukui
Bunkyo 3-9-1, Fukui 910-8507, Japan
ohara@fuee.u-fukui.ac.jp

Abstract. A certain class of information geometric structure can be conformally transformed to dually flat one. This paper studies the transformation on the probability simplex from a viewpoint of *affine differential geometry* and provides its applications. By restricting affine immersions with certain conditions, the probability simplex is realized to be 1-conformally flat statistical manifolds immersed in $\mathbf{R}^{n+1}$. Using this fact, we introduce a concept of *conformal flattening* for such manifolds in order to obtain the corresponding dually flat statistical (Hessian) ones with conformal divergences, and show explicit forms of potential functions, dual coordinates. Finally, we demonstrate applications of the flattening to nonextensive statistical physics, Voronoi partitions and weighted centroids on the probability simplex with respect to *geometric divergences*, which is not necessarily of Bregman type.

Keywords: Conformal flattening, Affine differential geometry, Escort probability, Geometric divergence, Conformal divergence

1 Introduction

In the theory of information geometry for statistical models, the logarithmic function is crucially significant to give a standard information geometric structure for exponential family [1, 2]. By changing the logarithmic function to another one we can deform the standard structure to a new one while keeping its basic property as a statistical manifold, which consists of a pair of mutually dual affine connections (∇, ∇^*) with respect to Riemannian metric g. There exist several ways [6–9] to introduce functions to deform a statistical manifold structure and these functions are sometimes called *embedding* or *representing functions*.

Affine immersion [4] can be regarded as one of possible ways. Further, Kurose [5] has proved that *1-conformally flat* statistical manifolds (See Appendix) realized by a certain class of affine immersions can be conformally transformed to dually flat ones, which are the most fruitful information geometric structures.

In this paper we call the transformation *conformal flattening* and give its explicit formula in order to elucidate the relations between representing functions and realized information geometric structures. We also discuss its applicability to computational geometric topics. These are interpreted as generalizations of the results in [11, 12], where the arguments are limited to conformal flattening of the alpha-geometry [1, 2] (See also section 2.4).

The paper is organized as follows: In section 2 we first discuss the affine immersion of the probability simplex and its geometric structure realized by the associated *geometric divergence*. Next, the conformally flattening transformation is given and the obtained dually flat structure with the associated *conformal divergence* is investigated. Section 3 describes applications of the conformal flattening. We consider a Voronoi partition and a weighted centroid with respect to the geometric divergence on the probability simplex. While geometric divergences are not of Bregman type in general, geometric properties such as conformality and projectivity are well utilized in these topics. We also see that *escort probabilities*, which are interpreted as the dual affine coordinates for the flattened geometry, play important roles. Section 4 includes concluding remarks. Finally, a short review on statistical manifolds and affine differential geometry is given in Appendix.

2 Affine immersion of the probability simplex

Let $\mathcal{S}^n$ be the relative interior of the probability simplex, i.e.,

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

where $\mathbf{R}_+$ denotes the set of positive numbers.

Consider an affine immersion [4] (f, ξ) of the simplex $\mathcal{S}^n$ (see also Appendix). Let D be the canonical flat affine connection on $\mathbf{R}^{n+1}$. Further, let f be an immersion from $\mathcal{S}^n$ into $\mathbf{R}^{n+1}$ and ξ be a transversal vector field on $\mathcal{S}^n$ (cf. figure 1). For a given *affine immersion* (f, ξ) of $\mathcal{S}^n$, the induced torsion-free connection ∇ and the affine fundamental form h are defined from the Gauss formula by

$$D_X f_*(Y) = f_*(\nabla_X Y) + h(X, Y)\xi, \quad X, Y \in \mathcal{X}(\mathcal{S}^n), \tag{1}$$

where f_* is the differential of f and $\mathcal{X}(\mathcal{S}^n)$ is the set of vector fields on $\mathcal{S}^n$.

It is well known [5, 4] that the realized geometric structure $(\mathcal{S}^n, \nabla, h)$ is a statistical manifold if and only if (f, ξ) is nondegenerate and equiaffine, i.e., h is nondegenerate and $D_X\xi$ is tangent to $\mathcal{S}^n$ for any $X \in \mathcal{X}(\mathcal{S}^n)$. Furthermore, the statistical manifold $(\mathcal{S}^n, \nabla, h)$ is 1-conformally flat [5] (but not necessarily dually flat nor of constant curvature).

Now we consider the affine immersion with the following assumptions.

Assumptions:

1. The affine immersion (f, ξ) is nondegenerate and equiaffine,

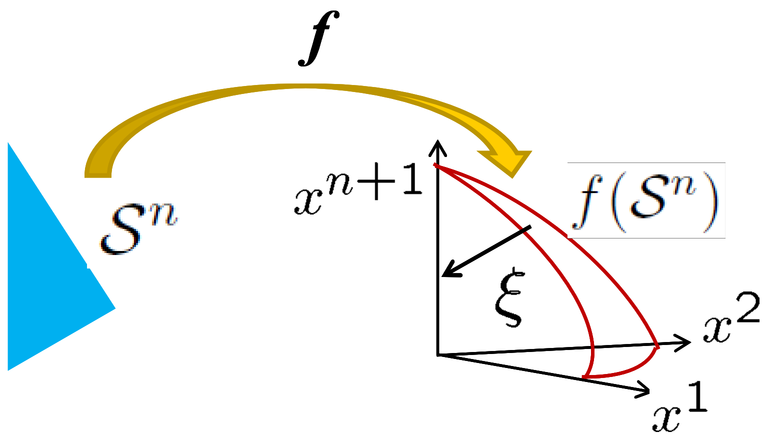

Fig. 1. An affine immersion (f, ξ) from $\mathcal{S}^n$ to $\mathbf{R}^{n+1}$

2. Let $\{x^i\}$ be an affine coordinate system for D on $\mathbf{R}^{n+1}$. The immersion f is given by the component-by-component and a common representing function L, i.e.,

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

3. The representing function $L : (0,\ 1) \to \mathbf{R}$ is sign-definite (or non-zero), concave with $L'' < 0$ and strictly increasing, i.e., $L' > 0$, Hence, the inverse of L denoted by E exists, i.e., $E \circ L = \mathrm{id}$.
4. Each component of ξ satisfies $\xi^i < 0,\ i = 1, \cdots, n+1$ 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 Proposition 1) by the adjustment. For a fixed L satisfying the assumption 3, we can choose ξ that meets the assumptions 1 and 4. For example, if we take $\xi^i = -|L(p_i)|$ then (f, ξ) is called *centro-affine*, which is known to be equiaffine [4]. The assumptions 3 and 4 also assure positive definiteness of h (The details are described in the proof of Proposition 1). Hence, (f, ξ) is non-degenerate and we can regard h as a Riemannian metric on $\mathcal{S}^n$.

2.1 Conormal vector and the geometric divergence

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

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

then $f(\mathcal{S}^n)$ 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 $\langle \nu, x \rangle$ the pairing of $x \in \mathbf{R}^{n+1}$ and $\nu \in \mathbf{R}_{n+1}$. The conormal vector [4] $\nu : \mathcal{S}^n \to \mathbf{R}_{n+1}$ for the affine immersion (f, ξ) is defined by

$$\langle \nu(p), f_*(X) \rangle = 0, \ \forall X \in T_p\mathcal{S}^n, \qquad \langle \nu(p), \xi(p) \rangle = 1 \tag{2}$$

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

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

we have

$$\nu_i(p) := \frac{1}{\Lambda} \frac{\partial \Psi}{\partial x^i} = \frac{1}{\Lambda(p)} E'(x^i) = \frac{1}{\Lambda(p)} \frac{1}{L'(p_i)}, \quad i = 1, \cdots, n+1, \tag{3}$$

where Λ is a normalizing factor defined by

$$\Lambda(p) := \sum_{i=1}^{n+1} \frac{\partial \Psi}{\partial x^i} \xi^i = \sum_{i=1}^{n+1} \frac{1}{L'(p_i)} \xi^i(p). \tag{4}$$

Then we can confirm (2) using the relation $\sum_{i=1}^{n+1} X^i = 0$ for $X = (X^i) \in \mathcal{X}(\mathcal{S}^n)$. Note that $v : \mathcal{S}^n \to \mathbf{R}_{n+1}$ defined by

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

also satisfies

$$\langle v(p), f_*(X) \rangle = 0, \ \forall X \in T_p\mathcal{S}^n. \tag{5}$$

Further, it follows, from (3), (4) and the assumption 4, that

$$\Lambda(p) < 0, \quad \nu_i(p) < 0, \quad i = 1, \cdots, n+1,$$

for all $p \in \mathcal{S}^n$.

It is known [4] that the affine fundamental form h can be represented by

$$h(X, Y) = -\langle \nu_*(X), f_*(Y) \rangle, \quad X, Y \in T_p\mathcal{S}^n.$$

In our case, it is calculated via (5) as

$$\begin{aligned} h(X, Y) &= -\Lambda^{-1} \langle v_*(X), f_*(Y) \rangle - X(\Lambda^{-1}) \langle v, f_*(Y) \rangle \\ &= -\frac{1}{\Lambda} \sum_{i=1}^{n+1} \left(\frac{1}{L'(p_i)} \right)' L'(p_i) X^i Y^i = \frac{1}{\Lambda} \sum_{i=1}^{n+1} \frac{L''(p_i)}{L'(p_i)} X^i Y^i. \end{aligned}$$

Since h is positive definite from the assumptions 3 and 4, we can regard it as a Riemannian metric.

Utilizing these notions from affine differential geometry, we can introduce the function ρ on $\mathcal{S}^n \times \mathcal{S}^n$, which is called a *geometric divergence* [5], as follows:

$$\begin{aligned}\rho(p,r) &= \langle \nu(r), f(p) - f(r)\rangle = \sum_{i=1}^{n+1} \nu_i(r)(L(p_i) - L(r_i)) \\ &= \frac{1}{\Lambda(r)} \sum_{i=1}^{n+1} \frac{L(p_i) - L(r_i)}{L'(r_i)}, \qquad p, r \in \mathcal{S}^n. \end{aligned} \tag{6}$$

We can easily see that ρ is a contrast function [10, 2] of the geometric structure $(\mathcal{S}^n, \nabla, h)$ because it holds that

$$\rho[X|] = 0, \quad h(X,Y) = -\rho[X|Y], \tag{7}$$

$$h(\nabla_X Y, Z) = -\rho[XY|Z], \quad h(Y, \nabla^*_X Z) = -\rho[Y|XZ], \tag{8}$$

where $\rho[X_1 \cdots X_k | Y_1 \cdots Y_l]$ stands for

$$\rho[X_1 \cdots X_k | Y_1 \cdots Y_l](p) := (X_1)_p \cdots (X_k)_p (Y_1)_r \cdots (Y_l)_r \rho(p,r)|_{p=r}$$

for $p, r \in \mathcal{S}^n$ and $X_i, Y_j \in \mathcal{X}(\mathcal{S}^n)$.

2.2 Conformal divergence and 1-conformal transformation

Let σ be a positive function on $\mathcal{S}^n$. Associated with the geometric divergence ρ, the *conformal divergence* [5] of ρ with respect to a conformal factor $\sigma(r)$ is defined by

$$\tilde{\rho}(p,r) = \sigma(r)\rho(p,r), \qquad p, r \in \mathcal{S}^n. \tag{9}$$

The divergence $\tilde{\rho}$ can be proved to be a contrast function for $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$, which is 1-conformally transformed geometric structure from $(\mathcal{S}^n, \nabla, h)$, where $\tilde{h}$ and $\tilde{\nabla}$ are given by

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

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

When there exists such a positive function σ that relates $(\mathcal{S}^n, \nabla, h)$ with $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ as in (10) and (11), they are called 1-conformally equivalent and $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ is also a statistical manifold [5].

2.3 Main result

Generally, the induced structure $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ from the conformal divergence $\tilde{\rho}$ is not also dually flat, which is the most abundant structure in information geometry. However, by choosing the conformal factor σ carefully, we can demonstrate that $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ is dually flat. Hereafter, we call such a transformation as *conformal flattening*.

Define

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

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

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

Proposition 1. *If the conformal factor is given by $\sigma = -1/Z$, then the statistical manifold $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ that is 1-conformally transformed from $(\mathcal{S}^n, \nabla, h)$ via (10) and (11) is dully flat. Further, $\tilde{\rho}$ is the canonical divergence where mutually dual pair of affine coordinates (θ^i, η_i) and a pair of potential functions (ψ, φ) are explicitly given by*

$$\theta^i(p) = x^i(p) - x^{n+1}(p) = L(p_i) - L(p_{n+1}), \quad i = 1, \cdots, n \tag{12}$$

$$\eta_i(p) = \frac{\nu_i(p)}{Z(p)} =: P_i(p), \quad i = 1, \cdots, n, \tag{13}$$

$$\psi(p) = -x_{n+1}(p) = -L(p_{n+1}), \tag{14}$$

$$\varphi(p) = \frac{1}{Z(p)} \sum_{i=1}^{n+1} \nu_i(p) x^i(p) = \sum_{i=1}^{n+1} P_i(p) L(p_i). \tag{15}$$

Proof. Using given relations, we first show that the conformal divergence $\tilde{\rho}$ is the canonical divergence [2] for $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$:

$$\begin{aligned}
\tilde{\rho}(p, r) &= -\frac{1}{Z(r)} \langle \nu(r), f(p) - f(r) \rangle = \langle P(r), f(r) - f(p) \rangle \\
&= \sum_{i=1}^{n+1} P_i(r)(x^i(r) - x^i(p)) \\
&= \sum_{i=1}^{n+1} P_i(r) x^i(r) - \sum_{i=1}^{n} P_i(r)(x^i(p) - x^{n+1}(p)) - \left(\sum_{i=1}^{n+1} P_i(r) \right) x^{n+1}(p) \\
&= \varphi(r) - \sum_{i=1}^{n} \eta_i(r) \theta^i(p) + \psi(p).
\end{aligned} \tag{16}$$

Next, let us confirm that $\partial\psi/\partial\theta^i = \eta_i$. Since $\theta^i(p) = L(p_i) + \psi(p),\ i = 1, \cdots, n$, we have

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

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

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

Differentiating by θ^j, we have

$$0 = \frac{\partial}{\partial\theta^j}\sum_{i=1}^{n+1} E(\theta^i - \psi) = \sum_{i=1}^{n+1} E'(\theta^i - \psi)\left(\delta^i_j - \frac{\partial\psi}{\partial\theta^j}\right)$$
$$= E'(x^j) - \left(\sum_{i=1}^{n+1} E'(x^i)\right)\frac{\partial\psi}{\partial\theta^j}.$$

This implies that

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

Together with (16) and this relation, φ is confirmed to be the Legendre transform of ψ.

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

Remark 2. Since the conformal metric is $\tilde{h} = -h/Z$, it is also positive definite. The dual affine connections ∇^* and $\tilde{\nabla}^*$ are known to be projectively equivalent [5]. Hence, ∇^* is projectively (or -1-conformally) flat. Further, the following corollary implies that the realized affine connection ∇ is also projectively equivalent to the flat connection $\tilde{\nabla}$ if we use the centro-affine immersion, i.e., $\xi^i = -L(p_i)$ [4,5] (See also Appendix). Note that the expressions of the dual coordinates $\eta_i(p) = P_i(p)$ can be interpreted as a generalization of the *escort probability* [13] because it is a normalization of deformed probabilities $1/L'(p_i)$ (see the following subsection).

Corollary 1. *The choice of ξ does not affect the obtained dually flat structure $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$.*

Proof. We have the following alternative expressions of $\eta_i = P_i$ with respect to L and E:

$$P_i(p) = \frac{1/L'(p_i)}{\displaystyle\sum_{k=1}^{n+1} 1/L'(p_k)} = \frac{E'(x_i)}{\displaystyle\sum_{i=1}^{n+1} E'(x_i)} > 0, \quad i = 1, \cdots, n.$$

Hence, all the expressions in proposition 1 does not depend on ξ, and the statement follows. Q.E.D.

2.4 Examples

Ex.1) If we take L to be the logarithmic function $L(t) = \ln(t)$, the conformally flattened geometry immediately defines the standard dually flat structure $(g^{\mathrm{F}}, \nabla^{(1)}, \nabla^{(-1)})$ on the simplex $\mathcal{S}^n$, where g^F denotes the Fisher metric. 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., $\tilde{\rho}(p, r) = D^{(\mathrm{KL})}(r||p) = \sum_{i=1}^{n+1} r_i(\ln r_i - \ln p_i)$.

Ex.2) Next let the affine immersion (f, ξ) be defined by the following L and ξ:

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

and

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

with $0 < q$ and $q \neq 1$, then it realizes the alpha-geometry [2] $(\mathcal{S}^n, \nabla^{(\alpha)}, g^F)$ with $q = (1+\alpha)/2$. Since the immersion (f, ξ) is centro-affine and the length of ξ is suitably scaled, $(\mathcal{S}^n, \nabla^{(\alpha)}, g^F)$ is of constant curvature $\kappa = (1-\alpha^2)/4$. The associated geometric divergence is the alpha-divergence, i.e.,

$$\rho(p, r) = D^{(\alpha)}(p, r) = \frac{4}{1-\alpha^2}\left(1 - \sum_{i=1}^{n+1}(p_i)^{(1-\alpha)/2}(r_i)^{(1+\alpha)/2}\right). \tag{17}$$

Following the procedure of conformally flattening described in the above, we have [11]

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

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

and obtain dually flat structure $(\tilde{h}, \tilde{\nabla}, \tilde{\nabla}^*)$ via the formulas in proposition 1:

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

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

Here, $\ln_q$ and $S_q(p)$ are the *q-logarithmic function* and the *Tsallis entropy* [13], respectively defined by

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

3 Construction of Voronoi partitions and centroids with respect to geometric divergences

In the previous section we have seen that various geometric divergences ρ can be constructed on the statistical manifold $\mathcal{S}^n$ by changing the representing function L and the transversal vector field ξ.

We demonstrate interesting applications of the conformal flattening to topics related with computational geometry, which are Voronoi partitions and centroids for the geometric divergence on a 1-conformally flat statistical manifold. We find that escort probabilities P_i (dual coordinates η_i) play important roles.

In this section, subscripts by Greek letters such as p_λ are used to denote the λ-th point in $\mathcal{S}^n$ among given ones while subscripts by Roman letters such as p_i denote the i-th coordinate of a point $p = (p_i) \in \mathcal{S}^n$.

3.1 Voronoi partitions

Let ρ be a geometric divergence defined in (6) on a 1-conformal statistical manifold $(\mathcal{S}^n, \nabla, h)$. For given m points $p_\lambda,\ \lambda = 1, \cdots, m$ on $\mathcal{S}^n$ we define *Voronoi regions* on $\mathcal{S}^n$ with respect to the geometric divergence ρ as follows:

$$\mathrm{Vor}^{(\rho)}(p_\lambda) := \bigcap_{\mu \neq \lambda} \{r \in \mathcal{S}^n | \rho(p_\lambda, r) < \rho(p_\mu, r)\}, \quad \lambda = 1, \cdots, m.$$

An *Voronoi partition (diagram)* on $\mathcal{S}^n$ is a collection of the Voronoi regions and their boundaries. For example, if we take $L(t) = t^{1-q}/(1-q)$ as in section 2.4, the corresponding Voronoi partition is the one with respect to the alpha-divergence $D^{(\alpha)}$ in (17) on $(\mathcal{S}^n, \nabla^{(\alpha)}, g^F)$ (Cf. the figures in [12]). Note that $D^{(\alpha)}$ approaches the Kullback-Leibler (KL) divergence if $\alpha \to -1$, and $D^{(0)}$ is called the Hellinger distance. Further, the partition is also equivalent to that with respect to *Rényi divergence* [14] defined by

$$D_\alpha(p, r) := \frac{1}{\alpha - 1} \ln \sum_{i=1}^{n+1} (p_i)^\alpha (r_i)^{1-\alpha}$$

because of their one-to-one functional relationship.

The acclaimed algorithm using projection of a convex polyhedron [15, 16] has been known to commonly work well to construct Voronoi partitions for the KL divergence [17–19] as well as the Euclidean distance. Furthermore, the algorithm is generally applicable if a divergence function δ is of *Bregman type* [20], which is represented by the remainder of the first order Taylor expansion of a convex potential function in a suitable coordinate system. Geometrically speaking, this implies that

i) the divergence δ is a *canonical divergence* [2] associated with a dually flat structure, i.e, it is of Bregman type:

$$\begin{aligned} \delta(p, r) &= \psi(\theta(r)) + \varphi(\eta(r)) - \sum_{i=1}^{n} \theta^i(p)\eta_i(r) \\ &= \varphi(\eta(r)) - \left\{ \varphi(\eta(p)) + \sum_{i=1}^{n} \theta^i(p) \left(\eta_i(r) - \eta_i(p)\right) \right\}, \qquad (18) \\ \theta^i &= \frac{\partial \varphi(\eta)}{\partial \eta_i}, \quad i = 1, \cdots, n, \end{aligned}$$

ii) its affine coordinate system $\eta = (\eta_i)$ is chosen to realize the corresponding Voronoi partitions. In this coordinate system with one extra complementary coordinate the polyhedron is expressed as the upper envelope of m hyperplanes tangent to the potential function $\varphi(\eta)$ at $\eta(p_\lambda),\ \lambda = 1, \cdots, m$.

Unfortunately a problem for the case of our Voronoi partition is that the geometric divergences ρ on $\mathcal{S}^n$ is *not* of Bregman type generally, i.e., they *cannot* be represented as a remainder of any convex potentials as in (18).

The following theorem, however, claims that the problem is resolved via Proposition 1. In other words, we can still apply the projection algorithm by conformally flattening a statistical manifold $(\mathcal{S}, \nabla, h)$ to a dually flat structure $(\mathcal{S}, \tilde{\nabla}, \tilde{h})$ and by invoking the conformal divergence $\tilde{\rho}$, which is always of Bregman type, and escort probabilities $\eta_i(p) = P_i(p)$ as a coordinate system.

The similar result is proved in [12] for the case of the α-divergence $D^{(\alpha)}$. However, the proof there was based on the fact that $(\mathcal{S}^n, \nabla^{(\alpha)}, g^F)$ is a statistical manifold *of constant curvature* in order to use the *modified Pythagorean relation* (See Appendix). In the following theorem, the assumption is relaxed to a 1-conformally flat statistical manifold $(\mathcal{S}, \nabla, h)$ and we prove with the usual Pythagorean relation on dually flat space.

Here, we denote the space of escort distributions by $\mathcal{E}^n$ and represent the point on $\mathcal{E}^n$ by $P = (P_1, \cdots, P_n)$ because $P_{n+1} = 1 - \sum_{i=1}^n P_i$ and $\mathcal{E}^n$ is also the probability simplex.

Theorem 1. *i) The bisector of two points p_λ and p_μ defined by $\{r|\rho(p_\lambda, r) = \rho(p_\mu, r)\}$ is a simultaneously ∇^*- and $\tilde{\nabla}^*$-autoparallel hypersurface on $\mathcal{S}^n$.*
ii) Let $\mathcal{H}_\lambda, \lambda = 1, \cdots, m$ be the hyperplane in $\mathcal{E}^n \times \mathbf{R}$ which is respectively tangent at $(P(p_\lambda), \varphi(p_\lambda))$ to the hypersurface $\{(P, y) = (P(p), \varphi(p))|p \in \mathcal{S}^n\}$. The Voronoi partition with respect to ρ can be constructed on $\mathcal{E}^n$ by projecting the upper envelope of all $\mathcal{H}_\lambda$'s along the y-axis.

Proof. i) We construct a bisector for points p_λ and p_μ. Consider the $\tilde{\nabla}$-geodesic $\tilde{\gamma}$ connecting p_λ and p_μ, and let $\bar{p}$ be the midpoint on $\tilde{\gamma}$ satisfying $\tilde{\rho}(p_\lambda, \bar{p}) = \tilde{\rho}(p_\mu, \bar{p})$. Note that the point $\bar{p}$ satisfies $\rho(p_\lambda, \bar{p}) = \rho(p_\mu, \bar{p})$ by the conformal relation (9). Denote by $\mathcal{B}$ the $\tilde{\nabla}^*$-autoparallel hypersurface that is orthogonal to $\tilde{\gamma}$ at $\bar{p}$ with respect to the conformal metric $\tilde{h}$. Note that $\mathcal{B}$ is simultaneously ∇^*-autoparallel because of the projective equivalence of ∇^* and $\tilde{\nabla}^*$ as is mentioned in Remark 2.

Using these setup and the fact that $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ is dually flat, we have the following relation from the Pythagorean theorem [2]

$$\tilde{\rho}(p_\lambda, r) = \tilde{\rho}(p_\lambda, \bar{p}) + \tilde{\rho}(\bar{p}, r) = \tilde{\rho}(p_\mu, \bar{p}) + \tilde{\rho}(\bar{p}, r) = \tilde{\rho}(p_\mu, r),$$

for all $r \in \mathcal{B}$. Using the conformal relation (9) again, we have $\rho(p_\lambda, r) = \rho(p_\mu, r)$ for all $r \in \mathcal{B}$. Hence, $\mathcal{B}$ is a bisector of p_λ and p_μ.

ii) Recall the conformal relation (9) between ρ and $\tilde{\rho}$, then we see that $\mathrm{Vor}^{(\rho)}(p_\lambda) = \mathrm{Vor}^{(\tilde{\rho})}(p_\lambda)$ holds on $\mathcal{S}^n$, where

$$\mathrm{Vor}^{(\tilde{\rho})}(p_\lambda) := \bigcap_{\mu \neq \lambda} \{r \in \mathcal{S}^n | \tilde{\rho}(p_\lambda, r) < \tilde{\rho}(p_\mu, r)\}.$$

Proposition 1 and the Legendre relations (16) imply that $\tilde{\rho}(p_\lambda, r)$ is represented with the escort probabilities, i.e., the dual coordinates $(P_i) = (\eta_i)$ by

$$\tilde{\rho}(p_\lambda, r) = \varphi(P(r)) - \left(\varphi(P(p_\lambda)) + \sum_{i=1}^{n} \frac{\partial \varphi}{\partial P_i}(p_\lambda)\{P_i(r) - P_i(p_\lambda)\} \right),$$

By definition the hyperplane $\mathcal{H}_\lambda$ is expressed by

$$\mathcal{H}_\lambda = \left\{ (P(r), y(r)) \,\middle|\, y(r) = \varphi(P(p_\lambda)) + \sum_{i=1}^{n} \frac{\partial \psi^*}{\partial P_i}(p_\lambda)\{P_i(r) - P_i(p_\lambda)\},\ r \in \mathcal{S}^n \right\}.$$

Hence, we have $\tilde{\rho}(p_\lambda, r) = \varphi(P(r)) - y(r)$. Thus, we see, for example, that the bisector on $\mathcal{E}^n$ for p_λ and p_μ is represented as a projection of $\mathcal{H}_\lambda \cap \mathcal{H}_\mu$. Thus, the statement follows. Q.E.D.

As a special case of the above theorem for $\rho = D^{(\alpha)}$, examples of Voronoi partitions with respect to $D^{(\alpha)}$ on usual probability simplex $\mathcal{S}^n$ and escort probability simplex $\mathcal{E}^n$ are given with their figures in [12].

Remark 3. Voronoi partitions for broader class of divergences that are not necessarily associated with any convex potentials are theoretically studied [21] from more general affine differential geometric points of views.

On the other hand, if the domain is extended from $\mathcal{S}^n$ to the positive orthant $\mathbf{R}_+^{n+1}$, then the α-divergence there can be expressed as a Bregman divergence [1, 2, 22]. Hence, the α-geometry on $\mathbf{R}_+^{n+1}$ is dually flat. Using this property, α-Voronoi partitions on $\mathbf{R}_+^{n+1}$ is discussed in [23].

However, while both of the above mentioned methods require constructions of the convex polyhedrons in the space of dimension $d = n + 2$, the new one proposed in this paper does in the space of dimension $d = n + 1$. Since it is known [24] that the optimal computational time of polyhedrons depends on the dimension d by $O(m \log m + m^{\lfloor d/2 \rfloor})$, the new one is slightly better when n is even and m is large.

3.2 Weighted Centroids

Let $p_\lambda,\ \lambda = 1, \cdots, m$ be given m points on $\mathcal{S}^n$ and $w_\lambda > 0,\ \lambda = 1, \cdots, m$ be their weights. Define the *weighted ρ-centroid* $c^{(\rho)} \in \mathcal{S}^n$ by the minimizer of the following problem:

$$\min_{p \in \mathcal{S}^n} \sum_{\lambda=1}^{m} w_\lambda \rho(p, p_\lambda).$$

Theorem 2. *The weighted ρ-centroid $c^{(\rho)}$ for given m points $p_1, \cdots, p_m$ on $\mathcal{S}^n$ is expressed by*

$$P_i(c^{(\rho)}) = \frac{1}{\sum_{\lambda=1}^{m} w_\lambda Z(p_\lambda)} \sum_{\lambda=1}^{m} w_\lambda Z(p_\lambda) P_i(p_\lambda), \quad i = 1, \cdots, n+1,$$

with weights w_λ, escort probabilities $P(p_\lambda)$ and the conformal factors $\sigma(p_\lambda) = -1/Z(p_\lambda) > 0$ for $p_\lambda,\ \lambda = 1, \cdots, m$.

Proof. Denote $\theta^i(p)$ by θ^i simply. Using (9), we have

$$\begin{aligned}
\sum_{\lambda=1}^m w_\lambda \rho(p, p_\lambda) &= -\sum_{\lambda=1}^m w_\lambda Z(p_\lambda) \tilde{\rho}(p, p_\lambda) \\
&= -\sum_{\lambda=1}^m w_\lambda Z(p_\lambda) \left\{ \psi(\theta) + \psi^*(\eta(p_\lambda)) - \sum_{i=1}^n \theta^i \eta_i(p_\lambda) \right\}.
\end{aligned}$$

Then the optimality condition is

$$\frac{\partial}{\partial \theta^i} \sum_{\lambda=1}^m w_\lambda \rho(p, p_\lambda) = -\sum_{\lambda=1}^m w_\lambda Z(p_\lambda) \{\eta_i - \eta_i(p_\lambda)\} = 0, \quad i = 1, \cdots, n,$$

where $\eta_i = \eta_i(p)$. Thus, the statements for $i = 1, \cdots, n$ hold from $\eta_i = P_i$ in Proposition 1. For $i = n + 1$, we have as follows:

$$\begin{aligned}
P_{n+1}(c^{(\rho)}) &= 1 - \sum_{i=1}^n P_i(c^{(\rho)}) \\
&= \frac{1}{\sum_{\lambda=1}^m w_\lambda Z(p_\lambda)} \sum_{\lambda=1}^m w_\lambda Z(p_\lambda) \left\{ 1 - \sum_{i=1}^n P_i(p_\lambda) \right\} \\
&= \frac{1}{\sum_{\lambda=1}^m w_\lambda Z(p_\lambda)} \sum_{\lambda=1}^m w_\lambda Z(p_\lambda) P_{n+1}(p_\lambda).
\end{aligned}$$

Q.E.D.

4 Concluding remarks

We have realized 1-conformally flat structures $(\mathcal{S}^n, \nabla, h)$ by changing affine immersions (f, ξ) or representing functions L, considered their conformal flattening and explicitly derived the corresponding dually flat structure, i.e., mutually dual potentials and affine coordinate systems.

Applications of the conformal flattening to topics in computational geometry are also demonstrated. As a result the geometric divergence, which is not generally of Bregman type, can be easily treated via the traditional computation algorithm. Recently, conformal divergences for Bregman-type divergences are proposed from different viewpoints and their properties are exploited [25, 26].

Extensions of the conformal flattening to other non-flat statistical manifolds or families of continuous probability distributions are left in the future work. While relations with the gradient flows (replicator flows, in a special case) on $(\mathcal{S}^n, \nabla, h)$ or $(\mathcal{S}^n, \tilde{\nabla}, \tilde{h})$ can be found in [27], searching for the other applications of the technique would be also of interest.

Acknowledgements

A. O. is partially supported by JSPS Grant-in-Aid (C) 15K04997.

References

1. Amari, S.-I. *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, vol. 28, Springer, New York, (1985).
2. S-I. Amari and H. Nagaoka, Methods of Information Geometry, Trans. Math. Monogr. vol 191 (American Mathematical Society and Oxford University Press) (2000).
3. Shima, H. *The geometry of Hessian structures*, World Scientific, Singapore (2007).
4. K. Nomizu and T. Sasaki, Affine Differential Geometry, Cambridge University Press, Cambridge, (1993).
5. T. Kurose, On the divergences of 1-conformally flat statistical manifolds, Tohoku Math. J. 46, 427–433 (1994).
6. J. Zhang Divergence Function, Duality, and Convex Analysis, Neural Computation 16, 159–195 (2004).
7. S. Eguchi, Information geometry and statistical pattern recognition, Sugaku Expositions,19, 197–216 (2006), (originally 2004 Sugaku 56 380–99 in Japanese).
8. J. Naudts, Continuity of a class of entropies and relative entropies, Reviews in Mathematical Physics, 16, 6, 809–822 (2004).
9. K. V. Harsha and K. S. Subrahamanian Moosath, F-geometry and Amari's α-geometry on a statistical manifold, Entropy, 16(5), 2472–2487 (2014).
10. S. Eguchi, Geometry of minimum contrast, Hiroshima Math. J. 631-647 (1992).
11. A. Ohara, H. Matsuzoe and S.-I. Amari, A dually flat structure on the space of escort distributions J. Phys.: Conf. Ser. 201, 012012 (2010).
12. A. Ohara, H. Matsuzoe and S. Amari, Conformal geometry of escort probability and its applications, Modern Physics Letters B, 26, 10 1250063 (14 pages) (2012).
13. C. Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer, New York (2009).
14. A. Rényi, On measures of entropy and information *Proc. 4th Berkeley Symp. Math. Statist. Probability* **1** (Univ. Calif. Press: Berkeley, 547-561 (1961).
15. H. Edelsbrunner and R. Seidel, Voronoi diagrams and arrangements, *Discrete Comput. Geom.* **1**, 25-44 (1986).
16. H. Edelsbrunner, *Algorithms in Combinatorial Geometry* Springer-Verlag, (1987).
17. K. Onishi and N. Takayama, Construction of Voronoi diagram on the Upper Half-Plane *IEICE Trans. on Fundamentals*, **E79-A**, 533-539 (1996).
18. K. Onishi and H. Imai, Voronoi diagram in statistical parametric space by Kullback-Leibler divergence, *Proc. 13th ACM-SIAM Symposium on Computational Geometry*, 463-465 (1997).
19. H. Imai and M. Inaba, Divergence-Based Geometric Clustering and Its Underlying Discrete Proximity Structures, *IEICE Trans. Inf. & Syst.*, **E83-D**, 27-35 (2000).
20. J.-D. Boissonnat, F. Nielsen and N. Nock, Bregman Voronoi diagram, *Discrete Comput. Geom.*, **44**, 281-307 (2010).
21. H. Matsuzoe, Computational Geometry from the Viewpoint of Affine Differential Geometry, *F.Nielsen eds. Emerging Trends in Visual Computing*, Berlin/Heidelberg: Springer, 103-123 (2009).

22. A. Ohara, Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization, *Physics Letters*, A 370 (2007) 184-193.
23. F. Nielsen and R. Nock, The dual Voronoi diagrams with respect to representational Bregman divergences, International Symposium on Voronoi Diagrams (ISVD), DTU Lyngby, Denmark (IEEE Press, New York, 2009).
24. B. Chazelle, An optimal convex hull algorithm in any fixed dimension, *Discrete and Computational Geometry* **10**, 377-409 (1993).
25. F. Nielsen and R. Nock, Total Jensen divergences: definition, properties and clustering, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2015).
26. R. Nock, F. Nielsen, S.-I. Amari, On Conformal Divergences and their Population Minimizers *IEEE Trans. IT*, 62, 1, 527-538 (2016).
27. A. Ohara, Conformal flattening for deformed information geometries on the probability simplex, Entropy, 20, 186 (2018).

A A short review of statistical manifolds and affine differential geometry

We shortly summarize the basic notions and results in information geometry [1, 2], Hessian domain [3] and affine differential geometry [4, 5], which are used in this paper. See for the details and proofs in the literature.

A.1 Statistical manifolds

For a torsion-free affine connection ∇ and a pseudo-Riemannian metric g on a manifold $\mathcal{M}$, the triple $(\mathcal{M}, \nabla, g)$ is called a *statistical (Codazzi) manifold* if it admits another torsion-free affine connection ∇^* satisfying

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \tag{19}$$

for arbitrary X, Y and Z in $\mathcal{X}(\mathcal{M})$, where $\mathcal{X}(\mathcal{M})$ is the set of all tangent vector fields on $\mathcal{M}$. We say that ∇ and ∇^* *duals* of each other with respect to g, and (g, ∇, ∇^*) is called *dualistic structure* on $\mathcal{M}$.

A statistical manifold $(\mathcal{M}, \nabla, g)$ is said to be of *constant curvature* $\kappa \in \mathbf{R}$ if the curvature tensor R of ∇ satisfies

$$R(X, Y)Z = \kappa\{g(Y, Z)X - g(X, Z)Y\}. \tag{20}$$

When the constant κ is zero, the statistical manifold is called *flat*, or *dually flat*, because the dual curvature tensor R^* of ∇^* also vanishes automatically [2, 3].

For $\alpha \in \mathbf{R}$, statistical manifolds $(\mathcal{M}, \nabla, g)$ and $(\mathcal{M}, \tilde{\nabla}, \tilde{g})$ are said to be *α-conformally equivalent* [5] if there exists a positive function σ on $\mathcal{M}$ satisfying

$$\begin{aligned} \tilde{g}(X, Y) &= \sigma g(X, Y) \\ g(\tilde{\nabla}_X Y, Z) &= g(\nabla_X Y, Z) - \frac{1+\alpha}{2}(d\ln\sigma)(Z)g(X, Y) \\ &\quad + \frac{1-\alpha}{2}\{(d\ln\sigma)(X)g(Y, Z) + (d\ln\sigma)(Y)g(X, Z)\}. \end{aligned}$$

Statistical manifolds $(\mathcal{M}, \nabla, g)$ and $(\mathcal{M}, \tilde{\nabla}, \tilde{g})$ are α-conformally equivalent if and only if $(\mathcal{M}, \nabla^*, g)$ and $(\mathcal{M}, \tilde{\nabla}^*, \tilde{g})$ are $-\alpha$-conformally equivalent. In particular, -1-conformal equivalence means *projective equivalence* of ∇ and $\tilde{\nabla}$, which implies that a ∇-pregeodesic curve is simultaneously $\tilde{\nabla}$-pregeodesic [4]. A statistical manifold $(\mathcal{M}, \nabla, g)$ is called *α-conformally flat* if it is locally α-conformally equivalent to a flat statistical manifold. It is known that a statistical manifold is of constant curvature if and only if it is ± 1-conformally flat, when $\dim \mathcal{M} \geq 3$ [5].

A.2 Affine differential geometry

Let $\mathcal{M}$ be an n-dimensional manifold and consider an *affine immersion* [4] (f, ξ), which is the pair of an immersion f from $\mathcal{M}$ into $\mathbf{R}^{n+1}$ and a transversal vector field ξ along $f(\mathcal{M})$. By a given affine immersion (f, ξ) of $\mathcal{M}$ and the usual flat affine connection D of $\mathbf{R}^{n+1}$, the Gauss and Weingarten formulas are respectively obtained as follows:

$$\begin{aligned} D_X f_*(Y) &= f_*(\nabla_X Y) + h(X,Y)\xi, \\ D_X \xi &= -f_*(SX) + \tau(X)\xi. \end{aligned}$$

Here, ∇, h, S and τ are called, respectively, *induced connection, affine fundamental form, affine shape operator* and *transversal connection form*. In this case, we say that the affine immersion realizes $(\mathcal{M}, \nabla, h)$ in $\mathbf{R}^{n+1}$. If h is non-degenerate (resp. $\tau = 0$ on $\mathcal{M}$), the affine immersion (f, ξ) is called *non-degenerate* (resp. *equiaffine*). It is known that non-degenerate and equiaffine (f, ξ) realizes a statistical manifold $(\mathcal{M}, \nabla, h)$ by regarding h as a pseudo-Riemannian metric g.

Such a statistical manifold is characterized as follows:

Proposition 2. *[5] A simply connected statistical manifold $(\mathcal{M}, \nabla, g)$ can be realized by a non-degenerate and equiaffine immersion if and only if it is 1-conformally flat.*

Let a point o in $\mathbf{R}^{n+1}$ be chosen as origin and consider an immersion f from $\mathcal{M}$ to $\mathbf{R}^{n+1}\backslash\{o\}$ so that $\xi = -\overrightarrow{of(p)}$ is transversal to $f(\mathcal{M})$ for $p \in \mathcal{M}$. Such an affine immersion (f, ξ) is called *centro-affine*, where the Weingarten formula is $D_X \xi = -f_*(X)$, or $S = I$ and $\tau = 0$. This implies that a centro-affine immersion, if it is non-degenerate, realizes a statistical manifold of constant curvature because of the Gauss equation:

$$R(X,Y)Z = h(X,Z)SX - h(X,Z)SY.$$

Further, the realized affine connection ∇ is projectively flat [4].

Denote the dual space of $\mathbf{R}^{n+1}$ by $\mathbf{R}_{n+1}$ and the pairing of $x \in \mathbf{R}^{n+1}$ and $y \in \mathbf{R}_{n+1}$ by $\langle y, x \rangle$. Define a map $\nu : \mathcal{M} \to \mathbf{R}_{n+1}\backslash\{o\}$ as follows:

$$\langle \nu_p, \xi_p \rangle = 1, \quad \langle \nu_p, f_*(X) \rangle = 0 \quad (\forall X \in T_p\mathcal{M}).$$

Such ν_p is uniquely defined and is called the *conormal vector*.

The pair $(\nu, -\nu)$ can be regarded as a centro-affine immersion from $\mathcal{M}$ into the dual space $\mathbf{R}_{n+1}$ equipped with the usual flat connection D^*. The formulas are

$$D^*_X(\nu_* Y) = \nu(\nabla^*_X Y) + h^*(X,Y)(-\nu),$$
$$D^*_X(-\nu) = -\nu_*(X),$$

where $h^*(X,Y) = h(SX,Y)$, and ∇^* is dual of ∇ with respect to h. Hence, when (f,ξ) realizes a statistical manifold $(\mathcal{M}.\nabla,h)$ with $S = I$, then $(\nu,-\nu)$ realizes its dual statistical manifold $(\mathcal{M},\nabla^*,h)$ [4]. Both manifolds are of constant curvature.

For a statistical manifold $(\mathcal{M},\nabla,h)$ realized by a non-degenerate and equiaffine immersion (f,ξ), we can define a *contrast function* ρ that induces the structure $(\mathcal{M},\nabla,h)$

$$\rho(p,q) = \langle \nu(q), f(p) - f(q)\rangle, \quad (p,q \in \mathcal{M}).$$

The function ρ is called the *geometric divergence* of $(\mathcal{M},\nabla,h)$ [5]. For a statistical manifold $(\mathcal{M},\tilde{\nabla},\tilde{h})$ that is 1-conformally equivalent to $(\mathcal{M},\nabla,h)$, one of its contrast function is given by $\tilde{\rho}(p,q) = \sigma(q)\rho(p,q)$ for a certain positive function σ. The contrast function $\tilde{\rho}$ is called the *conformal divergence* [5].

A statistical manifold $(\mathcal{M},\nabla,g)$ of constant curvature κ is studied from a viewpoint of affine differential geometry [5]. It is known that $(\mathcal{M},\nabla,g)$ realized in $\mathbf{R}^{n+1}$ has the following geometric properties:

P1 For three points p,q and r in $\mathcal{M}$ let the ∇-geodesic connecting p and q and the ∇^*-geodesic connecting q and r are orthogonal at q. Then the following modified Pythagorean relation holds:

$$\rho(p,r) = \rho(p,q) + \rho(q,r) - \kappa\rho(p,q)\rho(q,r),$$

P2 An arbitrary ∇-geodesic on $\mathcal{M}$ is the intersection of a two-dimensional subspace in $\mathbf{R}^{n+1}$ and $\mathcal{M}$,

P3 The volume element θ on $\mathcal{M}$ induced from $\mathbf{R}^{n+1}$ satisfies $\nabla\theta = 0$,

and so on. A typical example of the statistical manifold of non-zero constant curvature is the alpha-geometry $(\mathcal{S}^n, \nabla^{(\alpha)}, g^F)$, where $\kappa = (1-\alpha^2)/4$. In this case, the modified Pythagorean relation induces the widely-known nonextensivity relation of Tsallis entropy [22, Remark 2].