physical sciences forum

MDPI

Proceeding Paper

A Foliation by Deformed Probability Simplexes for Transition of α-Parameters †

Keiko Uohashi

Faculty of Engineering, Tohoku Gakuin University, Tagajo 985-8537, Miyagi, Japan; uohashi@mail.tohoku-gakuin.ac.jp

† Presented at the 41st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, 18–22 July 2022.

Abstract: This study considers dualistic structures of the probability simplex from the information geometry perspective. We investigate a foliation by deformed probability simplexes for the transition of α-parameters, not for a fixed α-parameter. We also describe the properties of extended divergences on the foliation when different α-parameters are defined on each of the various leaves.

Keywords: exponential family; escort distribution; Tsallis entropy; dualistic structure; divergence; affine geometry

Citation: Uohashi, K. A Foliation by Deformed Probability Simplexes for Transition of α-Parameters. *Phys. Sci. Forum* **2022**, *5*, 53. https://doi.org/10.3390/psf2022005053

Academic Editors: Frédéric Barbaresco, Ali Mohammad-Djafari, Frank Nielsen and Martino Trassinelli

Published: 28 March 2023

1. Introduction

For instance, since the Cauchy distribution and the Student's t-distribution are q-Gaussians, a set of q-normal distributions is considered a typical q-exponential family and has been related to nonextensive statistical mechanics [1,2]. Sets of q-normal distributions and q-exponential families have been investigated from the information geometry perspective, sometimes for nonextensive statistical mechanics and sometimes independently of it [3–10]. Deformed q-exponential families are defined using the deformed logarithm and reciprocal-deformed exponential functions. For instance, deformed q-exponential families have been used for studying escort distributions [11]. Their Hessian and conformal structures have also been investigated [12–14].

The current study considers a foliation by deformed probability simplexes representing sets of escort distributions, which are typical q-exponential families for the continuous transition of α-parameters on the information geometry. Previous studies on escort distributions are for a fixed α-parameter or among several α-parameters. However, foliations and divergence decomposition in dually flat spaces using mixed parameterizations are crucial. Therefore, we investigate extended divergences on the foliation, setting different α-parameters on each leaf.

First, we explain the dualistic structures, α-divergences, and the Tsallis relative entropy on the probability simplex. Next, we describe the divergences generated by affine immersions as level surfaces on the deformed probability simplexes corresponding to sets of escort distributions. We then define an extended divergence on a foliation by deformed probability simplexes. Finally, we propose a new decomposition of an extended divergence on the foliation.

2. Dualistic Structures and Divergences on the Probability Simplex

Let $\mathcal{S}^n$ be the n-dimensional probability simplex, i.e.,

$$\mathcal{S}^n = \{p := (p_i) | p_i > 0, \sum_{i=1}^{n+1} p_i = 1\}, \tag{1}$$

where $p_1, \ldots, p_{n+1}$ are the probabilities of $n+1$ states. Let $\{\bar{p}_1, \ldots, \bar{p}_n\}$ be an affine coordinate system on $\mathcal{S}^n$, where $\bar{p}_i \equiv p_i - p_{n+1}$ for $i = 1, \ldots, n$, and

$$\{\partial_i \equiv \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}\}_{i=1}^n \tag{2}$$

be a frame of a tangent vector field on $\mathcal{S}^n$. The Fisher metric $g = (g_{ij})$ on $\mathcal{S}^n$ is defined by

$$g_{ij} \equiv g(\partial_i, \partial_j) = \sum_{k=1}^{n+1} p_k \frac{\partial \log p_k}{\partial p_i} \frac{\partial \log p_k}{\partial p_j} = \frac{1}{p_i}\delta_{ij} + \frac{1}{p_{n+1}}, \quad i, j = 1, \ldots, n, \tag{3}$$

where δ_{ij} is Kronecker's delta. We define an α-connection $\nabla^{(\alpha)}$ on $\mathcal{S}^n$ by

$$\nabla^{(\alpha)}_{\partial_i} \partial_j = \sum_{k=1}^{n} \Gamma^{(\alpha)k}_{ij} \partial_k, \tag{4}$$

$$\Gamma^{(\alpha)k}_{ij} = \frac{1+\alpha}{2}(-\frac{1}{p_k}\delta^k_{ij} + p_k g_{ij}), \quad i, j, k = 1, \ldots, n, \tag{5}$$

where $\delta^k_{ij} = 1$ if $i = j = k$, and $\delta^k_{ij} = 0$ if others. Then, the Levi-Civita connection ∇ of g coincides with $\nabla^{(0)}$. For $\alpha \in \mathbf{R}$, we have

$$Xg(Y, Z) = g(\nabla^{(\alpha)}_X Y, Z) + g(Y, \nabla^{(-\alpha)}_X Z) \qquad \text{for } X, Y, Z \in \mathcal{X}(\mathcal{S}^n), \tag{6}$$

where $\mathcal{X}(\mathcal{S}^n)$ is the set of all smooth tangent vector fields on $\mathcal{S}^n$. Then, $\nabla^{(-\alpha)}$ is called the dual connection of $\nabla^{(\alpha)}$. For each α, $\nabla^{(\alpha)}$ is torsion-free and $\nabla^{(\alpha)} g$ is symmetric. Therefore, the triple $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ is a statistical manifold, and $(\mathcal{S}^n, \nabla^{(-\alpha)}, g)$ the dual statistical manifold of it [9,12,13].

For $\alpha \neq \pm 1$, an α-divergence $\mathbf{D}^{(\alpha)}$ is defined by

$$\mathbf{D}^{(\alpha)}(p, r) = \frac{4}{1-\alpha^2}\{\frac{1-\alpha}{2}\sum_{i=1}^{n+1} p_i + \frac{1+\alpha}{2}\sum_{i=1}^{n+1} r_i - \sum_{i=1}^{n+1} (p_i)^{\frac{1-\alpha}{2}} (r_i)^{\frac{1+\alpha}{2}}\}, \quad p, r \in \mathbf{R}^{n+1}_+. \tag{7}$$

For $q = (1-\alpha)/2$, it is known that

$$\mathbf{D}^{(\alpha)}(p, r) = \frac{1}{q} K_q(p, r), \quad p, r \in \mathcal{S}^n, \tag{8}$$

for the Tsallis relative entropy, K_q defined by

$$K_q(p, r) \equiv -\sum_{i=1}^{n+1} p_i \ln_q(\frac{r_i}{p_i}) = \frac{1}{1-q}\{1 - \sum_{i=1}^{n+1} (p_i)^q (r_i)^{1-q}\}, \quad p, r \in \mathcal{S}^n, \tag{9}$$

where $\ln_q$ is the q-logarithmic function defined by

$$\ln_q x \equiv \frac{x^{1-q} - 1}{1-q}, \quad q \neq 1, \ x > 0 \tag{10}$$

[1,2]. The Tsallis relative entropy K_q converges to the Kullback–Leibler divergence as $q \to 1$, because $\lim_{q\to 1} \ln_q x = \log x$. In information geometric view, the α-divergence $\mathbf{D}^{(\alpha)}$ converges to the Kullback–Leibler divergence as $\alpha \to -1$.

3. Deformed Probability Simplexes and Escort Distributions

For $n+1$ states $p_1, \ldots, p_{n+1}$ on $\mathcal{S}^n$ and $0 < q < 1$, if each probability $\mathbf{P}(p_i)$ satisfies

$$\mathbf{P}(p_i) = \frac{(p_i)^q}{\sum_{i=1}^{n+1}(p_i)^q}, \quad i = 1, \ldots, n+1, \tag{11}$$

the probability distribution $\mathbf{P}$ is called the escort distribution [1,2], where $(p_i)^q$ is p_i powered by q.

It realizes the dualistic structure of a set of escort distributions via the affine immersion into $\mathbf{R}_+^{n+1}$ [12,13]. Let f_q be the affine immersion of $\mathcal{S}^n$ into $\mathbf{R}_+^{n+1}$ defined by

$$f_q : p = (p_i) \mapsto \theta = (\theta^i), \quad \theta^i = \frac{1}{q}(p_i)^q, \quad i = 1, \ldots, n+1, \tag{12}$$

where $\{\theta^1, \ldots, \theta^{n+1}\}$ is the canonical coordinate system on $\mathbf{R}^{n+1}$. Then, the escort distribution $\mathbf{P}$ is represented as follows:

$$\mathbf{P}(p_i) = \frac{\theta^i}{\sum_{i=1}^{n+1}\theta^i}, \quad i = 1, \ldots, n+1. \tag{13}$$

For a function ψ_q on $\mathbf{R}_+^{n+1}$ defined by

$$\psi_q(\theta) = \frac{1}{1-q}\sum_{i=1}^{n+1}(q\theta^i)^{\frac{1}{q}}, \tag{14}$$

the image $f_q(\mathcal{S}^n)$ is a level surface of ψ_q satisfying $\psi_q(\theta) = 1/(1-q)$. For $0 < q < 1$, the Hessian matrix of the function ψ_q is positive definite on $\mathbf{R}_+^{n+1}$. Then, ψ_q induces the Hessian structure $(\mathbf{R}_+^{n+1}, D, h \equiv (\partial^2\psi_q/\partial\theta^i\partial\theta^j))$, where D is the canonical flat affine connection [9,15]. By definition

$$\tilde{\Gamma}_{ijk} = \sum_{l=1}^{n+1} h_{kl}\tilde{\Gamma}_{ij}^l = \frac{\partial^3\psi_q}{\partial\theta^i\partial\theta^j\partial\theta^k}, \quad i, j, k = 1, \ldots, n, \tag{15}$$

$$D^{(\alpha)}_{\frac{\partial}{\partial\theta_i}}\frac{\partial}{\partial\theta_j} = \frac{1-\alpha}{2}\sum_{k=1}^{n+1}\tilde{\Gamma}_{ij}^k\frac{\partial}{\partial\theta_k}, \quad \alpha = 1-2q, \tag{16}$$

the tetrad $(\mathbf{R}_+^{n+1}, D, D^{(-1)}, h)$ is the dual flat structure.

The submanifold structure of $f_q(\mathcal{S}^n)$ induced by $(\mathbf{R}_+^{n+1}, D, D^{(-1)}, h)$ coincides with the dualistic structure induced by the equiaffine immersion (f_q, E), where

$$E \equiv -d\psi_q(\tilde{E})^{-1}\tilde{E} \tag{17}$$

for the gradient vector field $\tilde{E}$ of ψ_q on $\mathbf{R}_+^{n+1}$ defined by

$$h(\tilde{X}, \tilde{E}) = d\psi_q(\tilde{X}) \quad \text{for } \tilde{X} \in \mathcal{X}(\mathbf{R}_+^{n+1}) \tag{18}$$

(cf. Theorem 2) [13,16,17]. In Equation (19), the induced affine connection D^E is the restricted D. The affine fundamental form h^E is the restricted h. The operator S^E is called the shape operator. If the transversal connection form satisfies $\tau^E \equiv 0$, (f_q, E), then it is called the equiaffine immersion [18].

$$D_X Y = D_X^E Y + h^E(X,Y)E, \quad D_X E = S^E(X) + \tau^E(X)E \qquad \text{for } X, Y \in \mathcal{X}(f_q(\mathcal{S}^n)). \tag{19}$$

For the restricted D and h on $f_q(\mathcal{S}^n)$, we use the same notations. The pullback of $(f_q(\mathcal{S}^n), D, h)$ to $\mathcal{S}^n$ is (-1)-conformally equivalent to $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ defined by Equations (3)–(5). In addition, $(f_q(\mathcal{S}^n), D, h)$ has a constant curvature $\kappa = q(1-q) = (1-\alpha^2)/4$ [13].

4. Divergences Generated by Affine Immersions as Level Surfaces

Let $\tilde{D}$ be the canonical flat affine connection on an $(n+1)$-dimensional real affine space $\mathbf{A}^{n+1}$. The following theorem is known on the level surfaces of a Hessian function.

Theorem 1 ([16]). *Let M be a simply connected n-dimensional level surface of φ on an $(n+1)$-dimensional Hessian domain $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$ with a Riemannian metric $\tilde{g}$ and suppose that $n \geq 2$. If we consider $(\Omega, \tilde{D}, \tilde{g})$ a flat statistical manifold, (M, D, g) is a 1-conformally flat statistical submanifold of $(\Omega, \tilde{D}, \tilde{g})$, where D and g denote the connection and the Riemannian metric on M induced by $\tilde{D}$ and $\tilde{g}$, respectively.*

For $\alpha \in \mathbf{R}$, statistical manifolds (N, ∇, h) and $(N, \bar{\nabla}, \bar{h})$ are α-conformally equivalent if there exists a function ϕ on N such that

$$\bar{h}(X,Y) = e^{\phi} h(X,Y), \quad X, Y, Z \in \mathcal{X}(N),$$

$$h(\bar{\nabla}_X Y, Z) = h(\nabla_X Y, Z) - \frac{1+\alpha}{2} d\phi(Z) h(X,Y) + \frac{1-\alpha}{2}\{d\phi(X)h(Y,Z) + d\phi(Y)h(X,Z)\}.$$

If $(N, \bar{\nabla}, \bar{h})$ is 1-conformally equivalent to a flat statistical manifold (N, ∇, h), $(N, \bar{\nabla}, \bar{h})$ is called a 1-conformally flat statistical manifold. A statistical manifold (N, ∇, h) is 1-conformally flat iff the dual statistical manifold (N, ∇', h) is (-1)-conformally flat [19].

In terms of affine geometry, (N, ∇', h) and $(N, \bar{\nabla}', h)$ are (-1)-conformally equivalent if and only if ∇' and $\bar{\nabla}'$ are projectively equivalent [17,20].

The conformal immersion w for an affine immersion (v, ξ) satisfies that $\langle w(b), \xi_b \rangle = 1$, where b is a point on the surface, and $\langle a, b \rangle$ a pairing of $a \in \mathbf{A}^*_{n+1}$ and $b \in \mathbf{A}^{n+1}$. The next definition is given in relation to affine immersions and divergences.

Definition 1 ([19]). *Let (N, ∇, h) be a 1-conformally flat statistical manifold realized by a non-degenerate affine immersion (v, ξ) into $\mathbf{A}^{n+1}$, and w the conormal immersion for v. Then, the divergence ρ_{conf} of (N, ∇, h) is defined by*

$$\rho_{conf}(a, b) = \langle w(b), v(a) - v(b) \rangle \quad \textit{for} \;\; a, b \in N. \tag{20}$$

The ρ_{conf} definition is independent of the choice of a realization of (N, ∇, h).

This divergence ρ_{conf} is referred to as Kurose geometric divergence in affine geometry and as Fenchel–Young divergence in the machine learning community [21]. The canonical divergence ρ of a flat statistical manifold $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$ is defined by

$$\rho(a, b) = \varphi(a) + \varphi^*(\tilde{w}(b)) + \sum_{i=1}^{n+1} \tilde{v}^i(a) \tilde{v}'_i(b) \quad \text{for} \;\; a, b \in \Omega, \tag{21}$$

where $(\tilde{v}^i)$, $(\tilde{v}'_i = -\partial\varphi/\partial\tilde{v}_i)$, and φ^* are the primal coordinate, the dual coordinate, and the Legendre transform of φ, respectively [9]. The gradient mapping $\tilde{w}$ is defined by $\tilde{w} = -\partial\varphi/\partial\tilde{v} \in \mathbf{A}^*_{n+1}$. For a 1-conformally flat statistical submanifold (M, D, g) of a Hessian domain $(\Omega, \tilde{D}, \tilde{g})$, we denote by ρ_{sub} the restriction of the divergence ρ. Then, the next theorem holds.

Theorem 2 ([17]). *For a 1-conformally flat statistical submanifold (M, D, g) of $(\Omega, \tilde{D}, \tilde{g})$, two divergences ρ_{conf} and ρ_{sub} coincide.*

On the level surface $(f_q(\mathcal{S}^n), D, h)$ in Section 3, the restricted divergence from the canonical divergence of $(\mathbf{R}^{n+1}_+, D, h)$ coincides with the geometric divergence by Equation (20) for the affine immersion (f_q, E). In addition, the pullback divergence to $\mathcal{S}^n$ coincides with $\mathbf{D}^{(\alpha)}$ and the Tsallis relative entropy K_q [12].

5. Extended Divergence on a Foliation by Deformed Probability Simplexes

Previous sections described the dualistic structure, the affine immersion, and the divergence for each fixed q. This section defines a divergence on a foliation by deformed probability simplexes $(f_q(\mathcal{S}^n), D, h)$ for all $0 < q < 1$, and shows the divergence decomposition property. We apply the discussion on $\mathbf{A}^*_{n+1}$ and $\mathbf{A}^{n+1}$ in Section 4 to the one on $\mathbf{R}^*_{n+1\,+}$ and $\mathbf{R}^{n+1}_+$.

Let ρ_q be the divergence on $f_q(\mathcal{S}^n)$ defined by the affine immersion (f_q, E_q) by Equations (17) and (18).

Let $S_{fol} = \cup_{0<q<1} f_q(\mathcal{S}^n) \subset \mathbf{R}^{n+1}_+$, which corresponds to a foliation $\mathcal{F} = \{f_q(\mathcal{S}^n)|0 < q < 1\}$. We define a function ρ_{fol} on $S_{fol} \times S_{fol}$ as follows:

$$\rho_{fol}(a,b) \equiv \psi_{q(a)}(a) - \psi_{q(b)}(b) - \sum_{i=1}^{n+1} \eta_i(b)(\theta^i(a) - \theta^i(b)) \quad \text{for } a \in f_{q(a)}(\mathcal{S}^n),\ b \in f_{q(b)}(\mathcal{S}^n), \tag{22}$$

$$0 < q(a) < 1, \quad 0 < q(b) < 1,$$

where

$$\eta_i(b) \equiv \frac{\partial \psi_{q(b)}(b)}{\partial \theta^i} = \frac{1}{1-q(b)} (q(b)\theta^i(b))^{\frac{1-q(b)}{q(b)}}, \quad i = 1, \dots, n+1. \tag{23}$$

We identify the dual space $\mathbf{R}^*_{n+1\,+}$ with $\mathbf{R}^{n+1}_+$. The i-th component of the conformal immersion of (f_q, E_q) is $-\partial\psi_q/\partial\theta^i$. Then, the dual coordinate of b, denoted by $\eta(b)$, satisfies that $\eta(b) \equiv (\eta_1(b), \dots, \eta_{n+1}(b)) \in f_{1-q(b)}(\mathcal{S}^n)$ [12,17]. The next proposition holds.

Proposition 1. *The function ρ_{fol} satisfies that:*
(i) If $a, b \in f_{q(a)}(\mathcal{S}^n)$, $\rho_{fol}(a,b) = \rho_{q(a)}(a,b) = \mathbf{D}^{(\alpha(a))}(f^{-1}_{q(a)}(a), f^{-1}_{q(a)}(b))$, where $\alpha(a) = 1 - 2q(a)$.
(ii) In the case of $q(a) \geq q(b)$, $\rho_{fol}(a,b) \geq 0$ for $(a,b) \in S_{fol} \times S_{fol}$, and $\rho_{fol}(a,b) = 0$ if and only if $a = b$.

Proof. The Legendre transform of $\psi_{q(b)}$ is defined by $\psi^*_{q(b)}(\eta(b)) = -\psi_{q(b)}(b) + \sum_{i=1}^{n+1} \theta^i(b)\eta_i(b)$. By Equation (21), (i) holds. The definition of $\mathcal{S}^n$ induces that

$$\psi_{q(a)}(a) = \frac{1}{1-q(a)}, \quad \psi_{q(b)}(b) = \frac{1}{1-q(b)}. \tag{24}$$

If $1 > q(a) > q(b) > 0$, it holds that $\psi_{q(a)}(a) > \psi_{q(b)}(b)$. In addition, $f_{q(a)}(\mathcal{S}^n)$ and $f_{q(b)}(\mathcal{S}^n)$ are convex centro-affine hypersurfaces, and $f_{q(a)}(\mathcal{S}^n)$ is more on the origin side than $f_{q(b)}(\mathcal{S}^n)$. Then, $-\sum_{i=1}^{n+1} \eta_i(b)(\theta^i(a) - \theta^i(b)) \geq 0$. Thus, (ii) holds. □

Definition 2. *We refer to ρ_{fol} defined by Equations (22) and (23) as an extended divergence on the foliation S_{fol}.*

We define the *extended dual divergence* ρ^*_{fol} of ρ_{fol} as follows:

$$\rho^*_{fol}(\eta(a), \eta(b)) \equiv \psi^*_{q(a)}(\eta(a)) - \psi^*_{q(b)}(\eta(b)) - \sum_{i=1}^{n+1} \theta^i(b)(\eta_i(a) - \eta_i(b)) \tag{25}$$

$$\text{for } a \in f_{q(a)}(\mathcal{S}^n),\ b \in f_{q(b)}(\mathcal{S}^n), \quad 0 < q(a) < 1, \quad 0 < q(b) < 1,$$

where ψ^*_q is the Legendre transform of ψ_q for $0 < q < 1$. Then, the following holds.

Proposition 2. *The functions ρ_{fol} and ρ^*_{fol} satisfy the following:*

$$\rho^*_{fol}(\eta(b), \eta(a)) = \rho_{fol}(a,b) \quad \text{for } a \in f_{q(a)}(\mathcal{S}^n),\ b \in f_{q(b)}(\mathcal{S}^n). \tag{26}$$

Proof. By the definition of the Legendre transform, we have

$$\begin{aligned}
\rho^*_{fol}(\eta(b),\eta(a)) &= \psi^*_{q(b)}(\eta(b)) - \psi^*_{q(a)}(\eta(a)) - \sum_{i=1}^{n+1}\theta^i(a)(\eta_i(b)-\eta_i(a)) \\
&= -\psi_{q(b)}(b) + \sum_{i=1}^{n+1}\theta^i(b)\eta_i(b) + \psi_{q(a)}(a) - \sum_{i=1}^{n+1}\theta^i(a)\eta_i(a) - \sum_{i=1}^{n+1}\theta^i(a)(\eta_i(b)-\eta_i(a)) \\
&= \psi_{q(a)}(a) - \psi_{q(b)}(b) - \sum_{i=1}^{n+1}\eta_i(b)(\theta^i(a)-\theta^i(b)) \;=\; \rho_{fol}(a,b).
\end{aligned}$$

□

The extended divergence using Equations (22) and (23) is related to the duo Bregman (pseudo-)divergence where the parameters also define the convex functions [22]. Their relationship will be studied in future works.

6. Decomposition of an Extended Divergence

At the beginning of this section, to make a decomposition theorem of an extended divergence, we give flows which are orthogonal to each leaf of $\mathcal{F}$.

For the foliation $\mathcal{F} = \{f_q(\mathcal{S}^n) | 0 < q < 1\}$, we consider the flow on S_{fol} defined using the following equation:

$$\frac{d\eta_i}{dt} = \eta_i, \quad i = 1, \dots, n+1, \tag{27}$$

where a function η_i on S_{fol} takes the i-th component of the dual coordinate on $f_q(\mathcal{S}^n)$ as Equation (23) for each $0 < q < 1$. An integral curve of Equation (27) is orthogonal to $f_q(\mathcal{S}^n)$ for each q with respect to the pairing $\langle \;,\; \rangle$. The set of the integral curves becomes the orthogonal foliation of $\mathcal{F}$. We denote it by $\mathcal{F}^\perp$.

Translating into the primal coordinate system, we have the next equation on S_{fol}:

$$\frac{d\theta^i}{dt} = \tilde{E}^i, \qquad \tilde{E}^i \equiv \tilde{E}^i_q = \sum_{j=1}^{n+1} h_q^{ij}\frac{\partial\psi_q}{\partial\theta^j} \;\text{ if }\; (\theta^i) \in f_q(\mathcal{S}^n), \quad i = 1, \dots, n+1. \tag{28}$$

The right-hand side of Equation (28) is calculated using Equations (17) and (18) for ψ_q. A leaf of $\mathcal{F}^\perp$ is an integral curve of the vector field $\tilde{E}$ that takes the value $\tilde{E}_q$ on $f_q(\mathcal{S}^n)$ for each q.

The following theorem is about the decomposition of the extended divergence.

Theorem 3. *Let $\mathcal{S}^n$ be the probability simplex, and $(f_q(S^n), D, h_q = Dd\psi_q)$ the 1-conformally flat statistical manifold generated by the affine immersion (f_q, E_q), where f_q is defined as*

$$f_q : p = (p_i) \mapsto \theta = (\theta^i), \quad \theta^i = \frac{1}{q}(p_i)^q, \quad i = 1, \dots, n+1, \tag{29}$$

$\psi_q(\theta) \equiv 1/(1-q)\sum_{i=1}^{n+1}(q\theta^i)^{1/q}$, $E_q \equiv -d\psi(\tilde{E}_q)^{-1}\tilde{E}_q$, and $\tilde{E}^i_q \equiv \sum_{j=1}^{n+1} h_q^{ij}\,\partial\psi_q/\partial\theta^j$. Let $a, b \in f_{q(a)}(\mathcal{S}^n)$, $0 < q(a) < 1$, and $c \in S_{fol} \equiv \cup_{0<q<1} f_q(\mathcal{S}^n)$. If there exists an orthogonal leaf $L^\perp \in \mathcal{F}^\perp$, which includes b and c, we have

$$\rho_{fol}(a,c) = \mu\,\rho_{fol}(a,b) + \rho_{fol}(b,c), \quad \eta(c) = \mu\,\eta(b), \;\; \mu > 0, \tag{30}$$

where $\eta(\cdot)$ is the dual coordinate of $f_q(\mathcal{S}^n)$ for each q.

Proof. From $a, b \in f_{q(a)}(\mathcal{S}^n)$, it holds that $\psi_{q(a)}(a) = \psi_{q(b)}(b)$, where $q(b) = q(a)$. By the definition in Equations (22) and (23), we have

$$\begin{aligned}
\rho_{fol}(a,c) &= \psi_{q(a)}(a) - \psi_{q(c)}(c) - \sum_{i=1}^{n+1} \eta_i(c)(\theta^i(a) - \theta^i(c)) \\
&= \psi_{q(b)}(b) - \psi_{q(c)}(c) - \sum_{i=1}^{n+1} \{\eta_i(c)(\theta^i(a) - \theta^i(b)) + \eta_i(c)(\theta^i(b) - \theta^i(c))\} \\
&= -\mu \sum_{i=1}^{n+1} \eta_i(b)(\theta^i(a) - \theta^i(b)) + \{\psi_{q(b)}(b) - \psi_{q(c)}(c) - \sum_{i=1}^{n+1} \eta_i(c)(\theta^i(b) - \theta^i(c))\} \\
&= \mu\, \rho_{fol}(a,b) + \rho_{fol}(b,c).
\end{aligned}$$

□

A decomposition similar to Equation (30) on a foliation of Hessian level surfaces is also available [17]. Theorem 3 generalizes the previous decomposition.

Finally, we describe the gradient flow on a leaf $f_q(S^n)$ using the extended divergence.

Theorem 4. *For a submanifold $(f_q(S^n), D, h_q)$ of S_{fol}, we denote by $(\theta^1, \ldots, \theta^n)$ an affine coordinate system on $f_q(S^n)$ such that $D d\theta^i = 0$, $i = 1, \ldots, n$, and set $h_{q\,ij} = h_q(\partial/\partial\theta^i, \partial/\partial\theta^j)$, $(h_q^{ij}) = (h_{q\,ij})^{-1}$. The gradient flow on $f_q(S^n)$ defined by*

$$\frac{d\theta^i}{dt} = -\sum_{j=1}^{n} h_q^{ij} \frac{\partial}{\partial\theta^j} \rho_{fol}(a_\theta, c), \quad a_\theta \in f_q(S^n), \quad c \in L^\perp, \quad i = 1, \ldots, n \tag{31}$$

converges to the unique point $b \in L^\perp \cap f_q(S^n)$, where a_θ is a variable point parametrized by (θ^i).

Proof. By Theorem 3, there exists $\mu > 0$ such that $\eta(c) = \mu\ \eta(b)$ and $\rho_{fol}(a_\theta, c) = \mu\ \rho_{fol}(a_\theta, b) + \rho_{fol}(b, c)$ for any $a_\theta \in f_q(S^n)$. Equation (31) is described by the dual coordinate system $(\eta_1, \ldots, \eta_n)$ on $f_q(S^n)$ as follows:

$$\frac{d\eta_i}{dt} = -\mu \frac{\partial}{\partial\theta^j} \rho_{fol}(a_\theta, b), \quad i = 1, \ldots, n. \tag{32}$$

On $f_q(S^n)$, from Proposition 1 (i), ρ_{fol} coincides with the geometric divergence ρ_q, generated by the affine immersion (f_q, E_q). The geometric divergence generates the dual coordinate η_i such that $D^* d\eta_i = 0$, $i = 1, \ldots, n$, to be derived by θ^i [19]. Then, it holds that

$$\frac{d\eta_i}{dt} = -\mu(\eta_i(a_\theta) - \eta_i(b)), \quad i = 1, \ldots, n. \tag{33}$$

and that

$$\eta_i = \eta_i(b) + (\eta_i(a|_{t=0}) - \eta_i(b))e^{-\mu}, \quad i = 1, \ldots, n, \tag{34}$$

where $a|_{t=0}$ is an initial point of Equation (31). Then, the gradient flow of Equation (31) converges to $b \in L^\perp \cap f_q(S^n)$ following a geodesic for the dual coordinate system. □

The gradient flow similar to Equation (31) has been provided on a flat statistical submanifold [23]. The similar one on a Hessian level surface, i.e., a 1-conformally statistical submanifold, has been given in [17].

7. Conclusions

This study considers a foliation of probability simplexes, which are typical q-exponential families, for the continuous transition of α-parameters on information geometry. We still need to provide details on the extended divergence and natural definition of the foliation of q-exponentials.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: I am grateful to the referees for their constructive comments.

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

References

1. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: New York, NY, USA, 2009.
2. Naudts, J. *Generalised Thermostatistics*; Springer: London, UK, 2011
3. Ohara, A.; Wada, T. Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations. *J. Phys. A Math. Theor.* **2010**, *43*, 035002. [CrossRef]
4. Matsuzoe, H.; Ohara, A. Geometry for q-exponential families. In *Recent Progress in Differential Geometry and Its Related Fields*; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific Publishing: Hackensack, NJ, USA, 2011; pp. 55–71.
5. Amari, S.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometry. *Physica A* **2012**, *391*, 4308–4319. [CrossRef]
6. Matsuzoe, H.; Henmi, M. Hessian structures and divergence functions on deformed exponential families. In *Geometric Theory of Information, Signals and Communication Technology*; Nielsen, F., Ed.; Springer: Basel, Switzerland, 2014; pp. 57–80.
7. Matsuzoe, H.; Wada, T. Deformed algebras and generalizations of independence on deformed exponential families. *Entropy* **2015**, *17*, 5729–5751. [CrossRef]
8. Wada, T.; Matsuzoe, H.; Scarfone, A.M. Dualistic Hessian structures among the thermodynamic potentials in the κ-thermostatistics. *Entropy* **2015**, *17*, 7213–7229. [CrossRef]
9. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
10. Scarfone, A.M.; Matsuzoe, H.; Wada, T. Information geometry of κ-exponential families: Dually-flat, Hessian and Legendre structures. *Entropy* **2018**, *20*, 436. [CrossRef] [PubMed]
11. Naudts, J. Estimators, escort probabilities, and ϕ-exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
12. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **2007**, *370*, 184–193. [CrossRef]
13. Ohara, A. Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation. *Eur. Phys. J. B* **2009**, *70*, 15–28. [CrossRef]
14. Matsuzoe, H. A sequence of escort distributions and generalizations of expectations on q-exponential family. *Entropy* **2017**, *19*, 7. [CrossRef]
15. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
16. Uohashi, K.; Ohara, A.; Fujii, T. 1-conformally flat statistical submanifolds. *Osaka J. Math.* **2000**, *37*, 501–507.
17. Uohashi, K.; Ohara, A.; Fujii, T. Foliations and divergences of flat statistical manifolds. *Hiroshima Math. J.* **2000**, *30*, 403–414. [CrossRef]
18. Nomizu, K.; Sasaki, T. *Affine Differential Geometry: Geometry of Affine Immersions*; Cambridge University Press: Cambridge, UK, 1994.
19. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J.* **1994**, *46*, 427–433. [CrossRef]
20. Nomizu, K.; Pinkal, U. On the geometry and affine immersions. *Math. Z.* **1987**, *195*, 165–178. [CrossRef]
21. Azoury, K.S.; Warmuth, M.K. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Mach. Learn.* **2001**, *43*, 211–246. [CrossRef]
22. Nielsen, F. Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences. *Entropy* **2022**, *24*, 421. [CrossRef] [PubMed]
23. Fujiwara, A.; Amari, S. Gradient systems in view of information geometry. *Physica D* **1995**, *80*, 317–327. [CrossRef]