

Article

Extended Divergence on a Foliation by Deformed Probability Simplexes

Keiko Uohashi 

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

Abstract: This study considers a new decomposition of an extended divergence on a foliation by deformed probability simplexes from the information geometry perspective. In particular, we treat the case where each deformed probability simplex corresponds to a set of q -escort distributions. For the foliation, different q -parameters and the corresponding α -parameters of dualistic structures are defined on each of the various leaves. We propose the divergence decomposition theorem that guides the proximity of q -escort distributions with different q -parameters and compare the new theorem to the previous theorem of the standard divergence on a Hessian manifold with a fixed α -parameter.

Keywords: divergence; exponential family; escort distribution; relative entropy; nonextensive statistics; information geometry; affine geometry



Citation: Uohashi, K. Extended Divergence on a Foliation by Deformed Probability Simplexes. *Entropy* **2022**, *24*, 1736. <https://doi.org/10.3390/e24121736>

Academic Editor: Ali Mohammad-Djafari

Received: 30 October 2022

Accepted: 26 November 2022

Published: 28 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of nonextensive statistics, q -normal distributions and the generalization, q -exponential families, play an important role [1–3]. Since Ohara first pointed out the correspondence between the q -parameter of nonextensive statistics and the α -parameter of information geometry [4,5], the information geometric structure of q -exponential families has been investigated [6–14].

On a set of probability distributions, divergences are usually defined for a fixed α -parameter of the dualistic structure. Using those results, we defined an extended divergence on a foliation by sets of probability distributions, setting different α -parameters on each leaf. In particular, we treated a foliation by deformed probability simplexes [15].

In this paper, we also study deformed probability simplexes corresponding to sets of escort distributions with q -parameters, which satisfy $q = (1 - \alpha)/2$ for α -parameters of information geometry. We clarify the relationship among affine spaces, affine immersions and the extended divergence more than in our previous paper. A comparison with the extended divergence and the duo Bregman divergence used in machine learning is also described [16].

First, we explain the dualistic structures, α -divergences, and the Tsallis relative entropy on the probability simplex, using the concept of affine geometry and information geometry. The relationship between an α -parameter and the Tsallis q -parameter is stated. Next, we describe the dualistic structures and the divergences generated by affine immersions on the deformed probability simplexes corresponding to sets of escort distributions. It also includes topics about Hessian manifolds and their level surfaces. 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. The Tsallis Relative Entropy and the Kullback–Leibler Divergence on the Probability Simplex

In this section, we explain dualistic structures, α -divergences, and the Tsallis relative entropy on the probability simplex [4,5,12].

Let $\mathbf{A}^{n+1}$ be an $(n+1)$ -dimensional real affine space and $\{x^1, \dots, x^{n+1}\}$ be the canonical affine coordinate system on $\mathbf{A}^{n+1}$, i.e., $\tilde{D}dx = 0$, where $\tilde{D}$ is the canonical flat affine connection on $\mathbf{A}^{n+1}$. Let $\mathcal{S}^n$ be a simplex in $\mathbf{A}_+^{n+1}$ defined by

$$\mathcal{S}^n = \{p \mid p \in \mathbf{A}_+^{n+1}, \sum_{i=1}^{n+1} x^i(p) = 1\}. \quad (1)$$

If $x^1(p), \dots, x^{n+1}(p)$ are regarded as probabilities of $n+1$ states, $\mathcal{S}^n$ is called the n -dimensional probability simplex. Let $\{\bar{p}^1, \dots, \bar{p}^n\}$ be an affine coordinate system on $\mathcal{S}^n$ defined by $\bar{p}^i(p) = x^i(p) - x^{n+1}(p)$ for $i = 1, \dots, n$, and

$$\{\partial_1, \dots, \partial_n\}, \quad \text{where } \partial_i|_p = \left(\frac{\partial}{\partial x^i} - \frac{\partial}{\partial x^{n+1}} \right) \Big|_p, \quad p \in \mathcal{S}^n, \quad (2)$$

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

$$g_{ij}(p) \equiv g(\partial_i, \partial_j)|_p = \sum_{k=1}^{n+1} x^k(p) \frac{\partial \log x^k}{\partial x^i} \Big|_p \frac{\partial \log x^k}{\partial x^j} \Big|_p = \frac{1}{x^i(p)} \delta_{ij} + \frac{1}{x^{n+1}(p)}, \quad (3)$$

$$p \in \mathcal{S}^n, \quad i, j = 1, \dots, n,$$

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

$$\nabla_{\partial_i}^{(\alpha)} \partial_j = \sum_{k=1}^n \Gamma_{ij}^{(\alpha)k} \partial_k, \quad (4)$$

$$\Gamma_{ij}^{(\alpha)k}|_p = \frac{1+\alpha}{2} \left(-\frac{1}{x^k(p)} \delta_{ij}^k + x^k(p) g_{ij}(p) \right), \quad i, j, k = 1, \dots, n, \quad (5)$$

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

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

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.

Note that affine connections $\nabla^{(1)}$ and $\nabla^{(-1)}$ in Equations (4)–(6) are the dual connection and the canonical connection, respectively.

It is known that when $n \geq 2$, the curvature of the statistical manifold $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ is a constant value

$$\kappa = \frac{1+\alpha}{2} \frac{1-\alpha}{2} = \frac{1-\alpha^2}{4}.$$

Therefore, the curvature of the dual statistical manifold $(\mathcal{S}^n, \nabla^{(-\alpha)}, g)$ is also $\kappa = (1-\alpha^2)/4$. Iff $\alpha = \pm 1$, the curvature of $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ is zero, and $(\nabla^{(\alpha)}, \nabla^{(-\alpha)}, g)$ is called the dually flat structure.

For $\alpha \neq \pm 1$, an α -divergence $\mathbf{D}^{(\alpha)}$ on $\mathbf{A}_+^{n+1}$ is often defined by

$$\mathbf{D}^{(\alpha)}(p, r) = \frac{4}{1-\alpha^2} \left\{ \frac{1-\alpha}{2} \sum_{i=1}^{n+1} x^i(p) + \frac{1+\alpha}{2} \sum_{i=1}^{n+1} x^i(r) - \sum_{i=1}^{n+1} x^i(p)^{\frac{1-\alpha}{2}} x^i(r)^{\frac{1+\alpha}{2}} \right\}, \quad (7)$$

$$p, r \in \mathbf{A}_+^{n+1}.$$

If $q = (1 - \alpha)/2$, it holds that

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

for the Tsallis relative entropy K_q on $\mathcal{S}^n$ defined by

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

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

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

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

For the Tsallis q -parameter, the curvature of the statistical manifold $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ is $\kappa = q(1 - q)$.

3. Divergences Generated by Affine Immersions as Level Surfaces

In this section, we describe the general theory of affine immersions and divergences related to level surfaces of the Hessian domain.

If the Hessian $\tilde{D}d\varphi = \sum_{i,j} (\partial^2 \varphi) / (\partial x^i \partial x^j) dx^i dx^j$ of a function φ on a domain $\Omega \subseteq \mathbf{A}^{n+1}$ is non-degenerate, the triple $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$ is called a Hessian domain. A statistical manifold is said to be flat if the curvature tensor of its affine connection vanishes. A flat statistical manifold is locally a Hessian domain. Conversely, a Hessian domain is a flat statistical manifold [12,17].

In a previous study, we show the following theorem on the level surfaces of a Hessian function.

Theorem 1 ([18]). *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.*

Here, “1-conformally flat” represents the characterization of surfaces projected by a flat statistical manifold along dual coordinates. We continue to explain the terms used in Theorem 1 and the outline of the proof.

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

$$\begin{aligned} \bar{h}(X, Y) &= e^{\phi} h(X, Y), \\ h(\bar{\nabla}_X Y, Z) &= h(\nabla_X Y, Z) - \frac{1 + \alpha}{2} d\phi(Z) h(X, Y) \\ &\quad + \frac{1 - \alpha}{2} \{ d\phi(X) h(Y, Z) + d\phi(Y) h(X, Z) \}, \quad X, Y, Z \in \mathcal{X}(N). \end{aligned}$$

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 [20,21].

For an $(n + 1)$ -dimensional Hessian domain $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$, an n -dimensional level surface of φ has the dualistic structure as the statistical submanifold structure. On the other hand, the level surface also has the structure induced by the affine immersion. It is essential for Theorem 1 that the statistical submanifold structure coincides with the dualistic structure by the affine immersion on a level surface of φ .

For $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$, let x be the canonical immersion of an n -dimensional level surface M into Ω . Let E be a transversal vector field on M defined by

$$E = -d\varphi(\tilde{E})^{-1}\tilde{E}, \quad (11)$$

where $\tilde{E}$ is the gradient vector field of φ on Ω defined by

$$\tilde{g}(\tilde{X}, \tilde{E}) = d\varphi(\tilde{X}), \quad \tilde{X} \in \mathcal{X}(\Omega). \quad (12)$$

For an affine immersion (x, E) and the canonical flat affine connection $\tilde{D}$ on $\Omega \subseteq \mathbf{A}^{n+1}$, the induced affine connection D^E , the affine fundamental form g^E , the shape operator $\mathcal{S}^E$ and the transversal connection form τ^E on M are defined by

$$D_X Y = D_X^E Y + g^E(X, Y)E, \quad (13)$$

$$D_X E = \mathcal{S}^E(X) + \tau^E(X)E, \quad X, Y \in \mathcal{X}(M). \quad (14)$$

See [21,22]. Then, D^E and g^E coincide with the restricted affine connection of $\tilde{D}$ and the restricted Riemannian metric of $\tilde{g}$, respectively. For the level surface M , the transversal connection form satisfies that $\tau^E \equiv 0$. Therefore, (x, E) is called the equiaffine immersion. It is known that a simply connected statistical manifold can be realized in $\mathbf{A}^{n+1}$ by a non-degenerate equiaffine immersion iff it is 1-conformally flat [19]. Thus, Theorem 1 holds.

Next, we introduce a divergence on a Hessian domain, treating it as a flat statistical manifold.

The canonical divergence ρ of a Hessian domain $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$ is defined by

$$\rho(p, r) = \varphi(p) + \varphi^*(\tilde{r}(r)) + \sum_{i=1}^{n+1} x^i(p)x'_i(r) \quad \text{for } p, r \in \Omega, \quad (15)$$

where $\tilde{r}$ is the gradient mapping from Ω to the dual affine space $\mathbf{A}_{n+1}^*$, i.e.,

$$x'_i = x_i^* \circ \tilde{r} = -\frac{\partial \varphi}{\partial x^i}, \quad (16)$$

and $\{x_1^*, \dots, x_{n+1}^*\}$ is the dual affine coordinate system of $\{x^1, \dots, x^{n+1}\}$. The Legendre transform φ^* of φ is defined by

$$\varphi^* \circ \tilde{r} = -\sum_{i=1}^{n+1} x^i x'_i - \varphi. \quad (17)$$

See [12].

Let ι be the conormal immersion for the affine immersion (x, E) defined by Equation (11), 12. By the definition of a conormal immersion, ι satisfies that

$$\langle \iota(p), Y_p \rangle = 0, \quad Y_p \in T_p M, \quad \langle \iota(p), E_p \rangle = 1 \quad \text{for } p \in M,$$

where $\langle a, b \rangle$ is the pairing of $a \in \mathbf{A}_{n+1}^*$ and $b \in \mathbf{A}^{n+1}$. It is known that the conormal immersion ι coincides with the restriction of the gradient mapping $\tilde{r}$ to the level surface M .

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

$$\rho_{\text{conf}}(p, r) = \langle w(r), v(p) - v(r) \rangle \quad \text{for } p, r \in N.$$

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

The divergence ρ_{conf} is referred to as Kurose geometric divergence in affine geometry and as Fenchel–Young divergence in the machine learning community [23,24]. Since an n -dimensional level surface M of $(\Omega, \tilde{D}, \tilde{g} = \tilde{D}d\varphi)$ is a 1-conformally flat statistical manifold realized by a non-degenerate affine immersion (x, E) , ρ_{conf} on M is as follows:

$$\rho_{\text{conf}}(p, r) = \langle \iota(r), x(p) - x(r) \rangle \quad \text{for } p, r \in M. \quad (18)$$

Let ρ_{sub} be the restriction of the canonical divergence ρ to (M, D, g) as a statistical submanifold of $(\Omega, \tilde{D}, \tilde{g})$. From Equations (15), (17) and (18), the next theorem holds.

Theorem 2 ([20]). For a 1-conformally flat statistical submanifold (M, D, g) of $(\Omega, \tilde{D}, \tilde{g})$, two divergences ρ_{conf} and ρ_{sub} coincide.

4. Deformed Probability Simplexes and Escort Distributions Generated by Affine Immersions

In this section, we explain dualistic structures on deformed probability simplexes, which correspond to sets of escort distributions via affine immersion.

We set $p_i = x^i(p)$, $i = 1, \dots, n+1$ for $p \in \mathcal{S}^n$, where $\mathcal{S}^n$ and $\{x^1, \dots, x^{n+1}\}$ be the probability simplex and the canonical affine coordinate system on $\mathbf{A}^{n+1}$, respectively. For $n+1$ states $p_1, \dots, p_{n+1}$ on $\mathcal{S}^n$ and $0 < q < 1$, if each probability $\mathbf{P}(p_i)$ satisfies

$$\mathbf{P}(p_i) = \frac{(p_i)^q}{\sum_{i=1}^{n+1} (p_i)^q}, \quad i = 1, \dots, n+1, \quad (19)$$

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 $\mathbf{A}_+^{n+1}$ [4,5]. For $0 < q < 1$, let f_q be the affine immersion of $\mathcal{S}^n$ into $\mathbf{A}_+^{n+1}$ defined by

$$x^i(f_q(p)) = \frac{1}{q} (x^i(p))^q, \quad i = 1, \dots, n+1, \quad \text{for } p \in \mathcal{S}^n. \quad (20)$$

Then, the escort distribution $\mathbf{P}$ is also represented as follows:

$$\mathbf{P}(p_i) = \frac{\theta^i}{\sum_{i=1}^{n+1} \theta^i}, \quad \theta^i = \frac{1}{q} (p_i)^q, \quad i = 1, \dots, n+1. \quad (21)$$

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

$$\psi_q = \frac{1}{1-q} \sum_{i=1}^{n+1} (qx^i)^{\frac{1}{q}}, \quad (22)$$

the image $f_q(\mathcal{S}^n)$ is a level surface of ψ_q satisfying $\psi_q = 1/(1-q)$. For $0 < q < 1$, the Hessian matrix of the function ψ_q is positive definite on $\mathbf{A}_+^{n+1}$. Then, ψ_q induces the Hessian structure $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{g}_q \equiv (\partial^2 \psi_q / \partial x^i \partial x^j))$. By definition

$$\tilde{\Gamma}_{ijk} = \sum_{l=1}^{n+1} \tilde{g}_{q\ kl} \tilde{\Gamma}_{ij}^l = \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k}, \quad i, j, k = 1, \dots, n, \quad (23)$$

$$\tilde{D}_{\frac{\partial}{\partial x_i}}^{(\alpha)} \frac{\partial}{\partial x_j} = \frac{1-\alpha}{2} \sum_{k=1}^{n+1} \tilde{\Gamma}_{ij}^k \frac{\partial}{\partial x_k}, \quad \alpha = 1 - 2q, \quad (24)$$

the tetrad $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{D}^{(-1)}, \tilde{g}_q)$ is the dually flat structure. The connection $\tilde{D}^{(0)}$ coincides with the Levi-Civita connection of the Riemannian metric $\tilde{g}_q$.

We denote by D and g_q the restricted $\tilde{D}$ and $\tilde{g}_q$ on $f_q(\mathcal{S}^n)$, and induce the dualistic structure of $(f_q(\mathcal{S}^n), D, g_q)$ as the submanifold structure of $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{g}_q)$. From the discussion in Section 3, $(f_q(\mathcal{S}^n), D, g_q)$ coincides with the dualistic structure induced by the equiaffine immersion (f_q, E_q) , where

$$E_q \equiv -d\psi_q(\tilde{E}_q)^{-1} \tilde{E}_q \quad (25)$$

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

$$\tilde{g}_q(\tilde{X}, \tilde{E}_q) = d\psi_q(\tilde{X}) \quad \text{for } \tilde{X} \in \mathcal{X}(\mathbf{A}_+^{n+1}). \quad (26)$$

The pullback of $(f_q(\mathcal{S}^n), D, g_q)$ to $\mathcal{S}^n$ is (-1) -conformally equivalent to $(\mathcal{S}^n, \nabla^{(\alpha)}, g)$ defined by Equations (3)–(5). In addition, $(f_q(\mathcal{S}^n), D, g_q)$ has a constant curvature $\kappa = q(1-q) = (1-\alpha^2)/4$ [5].

On $(f_q(\mathcal{S}^n), D, g_q)$, the restricted divergence ρ_q from the canonical divergence of $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{g}_q)$ coincides with the geometric divergence by Equation (18) from the affine immersion (f_q, E_q) . For an affine coordinate system $\{x'_1, \dots, x'_{n+1}\}$ on $\mathbf{A}_+^{n+1}$ defined by

$$x'_i = -\frac{\partial \psi_q}{\partial x^i} = -\frac{1}{1-q} (qx^i)^{\frac{1-q}{q}}, \quad (27)$$

the divergence ρ_q of $(f_q(\mathcal{S}^n), D, g_q)$ is described as

$$\rho_q(a, b) = \sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)), \quad a, b \in f_q(\mathcal{S}^n). \quad (28)$$

In addition, the pullback divergence of ρ_q to $\mathcal{S}^n$ coincides with $\mathbf{D}^{(\alpha)}$ and the Tsallis relative entropy K_q [4].

At the end of this section, we mention the divergence of $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{g}_q)$. By Equation (17), the Legendre transform ψ_q^* of ψ_q is

$$\psi_q^*(x'(a)) = -\psi_q(a) + \sum_{i=1}^{n+1} x^i(a)x'_i(a), \quad a \in \mathbf{A}_+^{n+1}. \quad (29)$$

By Equations (15) and (16), the canonical divergence ρ_q of $(\mathbf{A}_+^{n+1}, \tilde{D}, \tilde{g}_q)$ is defined by

$$\rho_q(a, b) = \psi_q(a) - \psi_q(b) + \sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)), \quad a, b \in \mathbf{A}_+^{n+1}, \quad (30)$$

represented by the same symbol ρ_q of $(f_q(\mathcal{S}^n), D, g_q)$.

5. Extended Divergence on a Foliation by Deformed Probability Simplexes

Previous sections described the divergence for each fixed q and each fixed α . This section defines an extended divergence on a foliation by deformed probability simplexes $(f_q(\mathcal{S}^n), D, g_q)$ for all $0 < q < 1$, and shows the divergence decomposition theorem.

The contents of our paper [15] are included but are explained in detail by the setting of affine geometry.

To give the proximity of q -escort distributions with different q -parameters, we define an extended divergence on a foliation by deformed probability simplexes as follows.

Definition 2. Let $S_{fol} = \cup_{0 < q < 1} f_q(\mathcal{S}^n) = \{p \mid p \in \mathbf{A}_+^{n+1}, \sum_{i=1}^{n+1} x^i(p) > 1\}$, which corresponds to a foliation $\mathcal{F} = \{f_q(\mathcal{S}^n) \mid 0 < q < 1\}$. We call a function ρ_{fol} on $S_{fol} \times S_{fol}$ defined by Equation (31) an extended divergence on a foliation by deformed probability simplexes.

$$\rho_{fol}(a, b) \equiv \psi_{q(a)}(a) - \psi_{q(b)}(b) + \sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)) \quad (31)$$

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

The i -th component of the conormal immersion of (f_q, E_q) is $-\partial\psi_q/\partial x^i$. By the right-hand side of Equation (27), the dual coordinate of b , denoted by $x'(b)$, satisfies that

$$-x'(b) \equiv (-x'_1(b), \dots, -x'_{n+1}(b)) \in f_{1-q(b)}(\mathcal{S}^n).$$

Therefore, we consider $f_{1-q}(\mathcal{S}^n)$ as the dual simplex of $f_q(\mathcal{S}^n)$ for $0 < q < 1$. As $q = 1/2$, $f_q(\mathcal{S}^n)$ is self dual [4]. Note that the i -th component of the dual coordinate of b is denoted by $\eta_i(b) = -x'_i(b) = (\partial\psi_q/\partial x^i)|_b$ in [15].

On the extended divergence, the next proposition holds.

Proposition 1. An extended divergence ρ_{fol} on S_{fol} of satisfies that:

(i) If $a, b \in f_{q(a)}(\mathcal{S}^n)$,

$$\rho_{fol}(a, b) = \rho_{q(a)}(a, b) = \mathbf{D}^{(\alpha(a))}(f_{q(a)}^{-1}(a), f_{q(a)}^{-1}(b)),$$

where ρ_q is the divergence of $(f_q(\mathcal{S}^n), D, g_q)$ by Equation (28), $\mathbf{D}^{(\alpha)}$ is an α -divergence defined by Equation (7), and $\alpha(a) = 1 - 2q(a)$.

(ii) In the case of $q(a) \geq q(b)$,

$$\rho_{fol}(a, b) \geq 0 \quad \text{for } (a, b) \in S_{fol} \times S_{fol},$$

and if and only if $a = b$,

$$\rho_{fol}(a, b) = 0.$$

Proof. If $a, b \in f_{q(a)}(\mathcal{S}^n)$, $\psi_{q(a)}(a) = \psi_{q(b)}(a) = \psi_{q(b)}(b)$. By Equations (28) and (31),

$$\rho_{fol}(a, b) = \sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)) = \rho_{q(a)}(a, b).$$

Then, (i) holds. If $1 > q(a) \geq q(b) > 0$, it holds that $\psi_{q(a)}(a) \geq \psi_{q(b)}(b)$ because

$$\psi_{q(a)}(a) = \frac{1}{1 - q(a)}, \quad \psi_{q(b)}(b) = \frac{1}{1 - q(b)} \quad (32)$$

are induced by the definition of $f_q(\mathcal{S}^n)$. In addition, $f_{q(a)}(\mathcal{S}^n)$ and $f_{q(b)}(\mathcal{S}^n)$ are convex surfaces centered on the origin of $\mathbf{A}_+^{n+1}$, and the surfaces $f_{q(a)}(\mathcal{S}^n)$ closer to the origin than $f_{q(b)}(\mathcal{S}^n)$. Then, $\sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)) \geq 0$. Thus, (ii) holds. $\square$

We define the *extended dual divergence* ρ_{fol}^* of ρ_{fol} as follows;

$$\rho_{fol}^*(a, b) \equiv \psi_{q(a)}^*(x'(a)) - \psi_{q(b)}^*(x'(b)) + \sum_{i=1}^{n+1} x^i(b)(x'_i(a) - x'_i(b)) \quad (33)$$

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

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 that*

$$\rho_{fol}^*(b, a) = \rho_{fol}(a, b) \quad \text{for } a \in f_{q(a)}(\mathcal{S}^n), \quad b \in f_{q(b)}(\mathcal{S}^n). \quad (34)$$

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

$$\begin{aligned} \rho_{fol}^*(b, a) &= \psi_{q(b)}^*(x'(b)) - \psi_{q(a)}^*(x'(a)) + \sum_{i=1}^{n+1} x^i(a)(x'_i(b) - x'_i(a)) \\ &= -\psi_{q(b)}(b) - \sum_{i=1}^{n+1} x^i(b)x'_i(b) \\ &\quad + \psi_{q(a)}(a) + \sum_{i=1}^{n+1} x^i(a)x'_i(a) + \sum_{i=1}^{n+1} x^i(a)(x'_i(b) - x'_i(a)) \\ &= \psi_{q(a)}(a) - \psi_{q(b)}(b) + \sum_{i=1}^{n+1} x'_i(b)(x^i(a) - x^i(b)) \\ &= \rho_{fol}(a, b). \end{aligned}$$

□

The extended divergence is related to the duo Bregman (pseudo-)divergence, where the parameters also define the convex functions [16]. To work with the entire parametrized probability distribution families and to explore the application of divergences, we must investigate their relationship.

6. Decomposition of an Extended Divergence

In this section, we explain the orthogonal foliation of $\mathcal{F}$. Next, we give a decomposition of an extended divergence along the orthogonal leaf and the original leaf.

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.

$$\frac{dx'_i}{dt} = x'_i, \quad i = 1, \dots, n+1, \quad (35)$$

where a function x'_i on S_{fol} takes the i -th component of the dual coordinate on $f_q(\mathcal{S}^n)$ as Equation (27) for each $0 < q < 1$. An integral curve of Equation (35) is orthogonal to $f_q(\mathcal{S}^n)$ for each q with respect to the pairing $\langle \cdot, \cdot \rangle$. The set of 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.

$$\frac{dx^i}{dt} = \tilde{E}^i, \quad i = 1, \dots, n+1, \quad \text{on } S_{fol}, \quad (36)$$

$$\tilde{E} = \tilde{E}_q^i = \sum_{j=1}^{n+1} g_q^{ij} \frac{\partial \psi_q}{\partial x^j} \quad \text{on } f_q(\mathcal{S}^n), \quad (37)$$

where (g_q^{ij}) is the inverse matrix of (g_{qij}) . The right-hand side of Equation (37) is calculated using Equations (11) and (12) 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 on the decomposition of the extended divergence.

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

$$x^i(f_q(p)) = \frac{1}{q}(x^i(p))^q, \quad i = 1, \dots, n+1, \quad \text{for } p \in \mathcal{S}^n, \quad (38)$$

$\psi_q \equiv 1/(1-q) \sum_{i=1}^{n+1} (qx^i)^{1/q}$, $E_q \equiv -d\psi(\tilde{E}_q)^{-1}\tilde{E}_q$, $\tilde{E}_q^i \equiv \sum_{j=1}^{n+1} g_q^{ij} \partial\psi_q/\partial x^j$, and g_q is the restriction of $(g_{qij}) = Dd\psi_q$ to $f_q(\mathcal{S}^n)$. Let $a, b \in f_{q(a)}(\mathcal{S}^n)$, $0 < q(a) < 1$, and $c \in \mathcal{S}_{fol} \equiv \cup_{0 < q < 1} f_q(\mathcal{S}^n)$. If there exists an orthogonal leaf $L^\perp \in \mathcal{F}^\perp$ that includes b and c , we have

$$\rho_{fol}(a, c) = \mu \rho_{fol}(a, b) + \rho_{fol}(b, c), \quad x'(c) = \mu x'(b), \quad \mu > 0, \quad (39)$$

where $x'(\cdot)$ 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

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

□

See Figure 1 for a decomposition of extended divergence and graphs of deformed simplexes $f_q(\mathcal{S}^n)$.

A decomposition similar to Equation (39) is also available on a foliation by Hessian level surfaces of one Hessian manifold [20]. Theorem 3 generalizes the previous decomposition.

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

Theorem 4. For a submanifold $(f_q(\mathcal{S}^n), D, g_q)$ of $\mathcal{S}_{fol}$, we denote by $\{x^1, \dots, x^n\}$ an affine coordinate system on $f_q(\mathcal{S}^n)$ such that $Ddx^i = 0$, $i = 1, \dots, n$, and set $g_{qij} = g_q(\partial/\partial x^i, \partial/\partial x^j)$, $(g_q^{ij}) = (g_{qij})^{-1}$. For a fixed point $c \in L^\perp$, the gradient flow on $f_q(\mathcal{S}^n)$ defined by

$$\frac{dx^i}{dt} = - \sum_{j=1}^n g_q^{ij} \frac{\partial}{\partial x^j} \rho_{fol}(a_x, c), \quad i = 1, \dots, n, \quad a_x \in f_q(\mathcal{S}^n) \quad (40)$$

converges to the unique point $b \in L^\perp \cap f_q(\mathcal{S}^n)$, where a_x is a variable point parametrized as $\{x^1(t), \dots, x^n(t)\}$.

Proof. By Theorem 3, for any $a_x \in f_q(\mathcal{S}^n)$, there exists $\mu > 0$ such that

$$\rho_{fol}(a_x, c) = \mu \rho_{fol}(a_x, b) + \rho_{fol}(b, c), \quad x'(c) = \mu x'(b).$$

Equation (40) is described by the dual coordinate system $\{x'_1, \dots, x'_n\}$ on $f_q(\mathcal{S}^n)$ as follows;

$$\frac{dx'_i}{dt} = -\mu \frac{\partial}{\partial x'_j} \rho_{fol}(a_x, b), \quad i = 1, \dots, n. \quad (41)$$

On $f_q(\mathcal{S}^n)$, from Prop. 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 x'_i such that $D^* dx'_i = 0$, $i = 1, \dots, n$, to be derived by x^i [19]. Then, it holds that

$$\frac{dx'_i}{dt} = -\mu(x'_i(a_x) - x'_i(b)), \quad i = 1, \dots, n, \quad (42)$$

and that

$$x'_i = x'_i(b) + (x'_i(a|_{t=0}) - x'_i(b))e^{-\mu t}, \quad i = 1, \dots, n, \quad (43)$$

where $a|_{t=0}$ is an initial point of Equation (40). Then, the gradient flow of Equation (40) converges to $b \in L^\perp \cap f_q(\mathcal{S}^n)$ following a geodesic for the dual coordinate system. $\square$

The gradient flow similar to Equation (40) has been provided on a flat statistical submanifold [25]. The similar one on a Hessian level surface, i.e., a 1-conformally statistical submanifold, has been given in [20]. Theorem 4 generalizes the previous theorems on gradient flows.

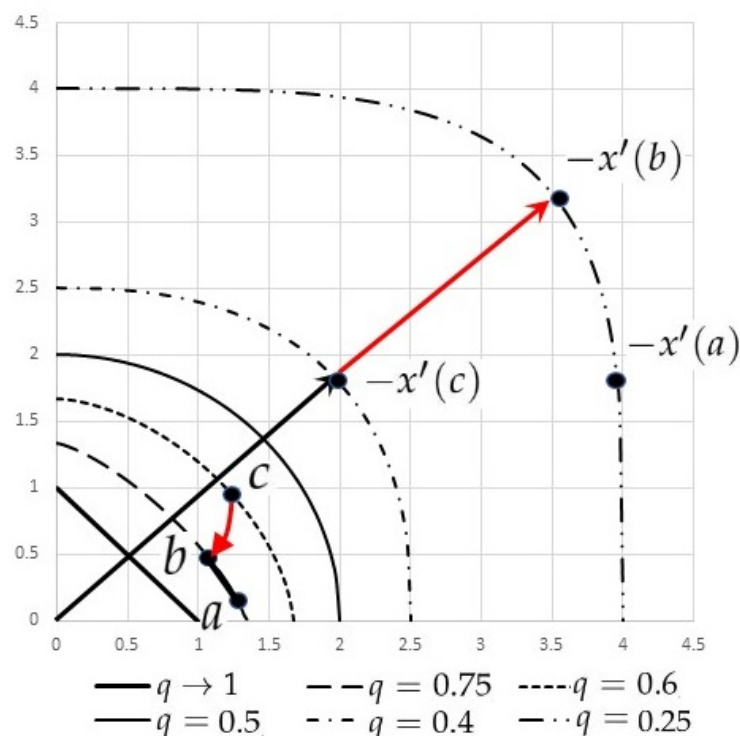


Figure 1. A decomposition of extended divergence $\rho_{fol}(a, c)$, graphs of the standard simplex ($q \rightarrow 1$), and deformed simplexes as $q = 0.75, 0.6, 0.5, 0.4, 0.25$ in $\mathbb{A}_+^2$. For primal coordinates $a, b \in f_{0.75}(\mathcal{S}^1)$, and $c \in f_{0.6}(\mathcal{S}^1)$, dual coordinates satisfy $-x'(a), -x'(b) \in f_{0.25}(\mathcal{S}^1)$, and $-x'(c) \in f_{0.4}(\mathcal{S}^1)$. The primal geodesic between a and b is orthogonal to the dual one between b and c with respect to the metric $g_{0.75}$.

7. Conclusions

This study considers a foliation of deformed probability simplexes corresponding to sets of escort distributions with q -parameters, for the continuous transition of α -parameters on information geometry. Since these are typical q -exponential families, we still need to provide details on the extended divergence and natural definition of the foliation of q -exponential families.

The extended divergence guides the proximity of q -exponential distributions with different q -parameters. Therefore, our theory guarantees the mathematical basis for generalizing methods of machine learning and statistical mechanics to the case of the q -distribution families when different q -parameters are mixed. The decomposition theorem is applied to the problem of the optimal choice of q -parameter. The application methods are open to consideration. It also remains to investigate the relationship with a new λ -duality on nonextensive statistical mechanics with mixed parameters [26,27].

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to the referees for their constructive comments.

Conflicts of Interest: The authors declare 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. Naudts, J. Estimators, escort probabilities, and ϕ -exponential families in statistical physics. *J. Ineq. Pure Appl. Math.* **2004**, *5*, 102.
4. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **2007**, *370*, 184–193. [[CrossRef](#)]
5. 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](#)]
6. 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](#)]
7. 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.
8. Amari, S.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometry. *Phys. A* **2012**, *391*, 4308–4319. [[CrossRef](#)]
9. 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.
10. Matsuzoe, H.; Wada, T. Deformed algebras and generalizations of independence on deformed exponential families. *Entropy* **2015**, *17*, 5729–5751. [[CrossRef](#)]
11. Wada, T.; Matsuzoe, H.; Scarfone, A.M. Dualistic Hessian structures among the thermodynamic potentials in the κ -thermostatistics. *Entropy* **2015**, *17*, 7213–7229. [[CrossRef](#)]
12. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
13. 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](#)]
14. Matsuzoe, H. A sequence of escort distributions and generalizations of expectations on q -exponential family. *Entropy* **2017**, *19*, 7. [[CrossRef](#)]
15. Uohashi, K. A foliation by deformed probability simplexes for transition of α -parameters. In Proceedings of the International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, IHP, Paris, France, 18–22 July 2022.
16. 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](#)]
17. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
18. Uohashi, K.; Ohara, A.; Fujii, T. 1-conformally flat statistical submanifolds. *Osaka J. Math.* **2000**, *37*, 501–507.
19. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J.* **1994**, *46*, 427–433. [[CrossRef](#)]
20. Uohashi, K.; Ohara, A.; Fujii, T. Foliations and divergences of flat statistical manifolds. *Hiroshima Math. J.* **2000**, *30*, 403–414. [[CrossRef](#)]
21. Nomizu, K.; Pinkall, U. On the geometry and affine immersions. *Math. Z.* **1987**, *195*, 165–178. [[CrossRef](#)]

-
22. Nomizu, K.; Sasaki, T. *Affine Differential Geometry: Geometry of Affine Immersions*; Cambridge University Press: Cambridge, UK, 1994.
 23. 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](#)]
 24. Blondel, M.; Martins, A.F.T.; Niculae V. Learning with Fenchel-Young losses. *J. Mach. Learn. Res.* **2020**, *21*, 1–69.
 25. Fujiwara, A.; Amari, S. Gradient systems in view of information geometry. *Phys. D.* **1995**, *80*, 317–327. [[CrossRef](#)]
 26. Zhang, J.; Wong, T.-K.L. λ -Deformation: A canonical framework for statistical manifolds of constant curvature. *Entropy* **2022**, *24*, 193. [[CrossRef](#)]
 27. Wong, T.-K.L.; Zhang, J. Tsallis and Rényi deformations linked via a new λ -duality. *IEEE Trans. Inf. Theory* **2022**, *68*, 5353–5373. [[CrossRef](#)] [[PubMed](#)]