Conjugate Representations and Characterizing Escort Expectations in Information Geometry

Tatsuaki Wada [1],* and Hiroshi Matsuzoe [2]

[1]   Department of Electrical and Electronic Engineering, Ibaraki University, Nakanarusawa-cho, Hitachi, Ibaraki 316-8511, Japan

[2]   Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan; matsuzoe@nitech.ac.jp

*   Correspondence: tatsukai.wada.to@vc.ibaraki.ac.jp; Tel.: +81-294-38-5110

Abstract: Based on the maximum entropy (MaxEnt) principle for a generalized entropy functional and the conjugate representations introduced by Zhang, we have reformulated the method of information geometry. For a set of conjugate representations, the associated escort expectation is naturally introduced and characterized by the generalized score function which has zero-escort expectation. Furthermore, we show that the escort expectation induces a conformal divergence.

Keywords: conjugate representations; information geometry; escort expectations; deformed exponential functions; conformal divergences

1. Introduction

Information geometry (IG) [1,2] is a differential geometrical method based on a Riemannian metric on a statistical manifold, which is constructed from a given parameterized probability distribution function (pdf) $p_\theta(x)$. It provides a useful tool to study, for example, the dually flat structures of a statistical manifold. Recently, much effort has been made to study some deformed exponential families of pdfs, in which the standard exponential function $\exp(x)$ and its inverse function $\ln(x)$ are replaced with a deformed exponential and its inverse function, which is called a deformed logarithmic function. Among different deformed exponential functions, relatively well known ones include Tsallis' q-deformed exponential [3] and Kaniadakis' κ-deformed exponential functions [4]. Naudts [5] introduced the so-called φ-logarithmic function in terms of a positive increasing function $\varphi(x)$, and studied the generalized thermostatistics. It is shown that a q-deformed relative entropy is proportional to Amari's α-divergence and is related with the α-geometry on the statistical manifold with a constant curvature [6]. In order to construct a suitable statistical manifold in IG, usually the α-representation (rep.), or α-immersion, of a pdf is used. It is well known that the α-rep. works fine with an exponential family but does not necessarily work fine for a non-exponential pdf, e.g., a κ-deformed exponential family [4]. A generalization of the α-rep. is *conjugate representations*, or (ρ, τ)-reps., by Zhang [7]. He also introduced the (ρ, τ)-divergences from the point of view of "representation duality". By finding out a suitable conjugate rep. for κ-deformed exponential pdf, the IG of the κ-generalized thermostatistics [8] was studied. We further studied the IG structures among the thermodynamic potentials in the κ-thermostatistics [9], in which the escort pdfs and escort expectations play an important role. Zhang [10] further showed that his conjugate reps. also include Naudts' φ-logarithm [5] as a special case. In this way, Zhang's conjugate reps. are very useful as a generalization of α-reps in IG. Amari [11] showed that (ρ, τ)-divergences generate a dually flat structure in the manifold of positive measures and in that of positive-definite matrices. In this contribution, we reformulate the IG structures based on Zhang's conjugate reps. and the maximum entropy (MaxEnt) principle for a generalized entropy functional. Our approach is different from the previous

works [10,11] in that we relate the ρ-rep. of a pdf $p(x)$ with the Lagrange multipliers in the MaxEnt problem. This enables us to introduce a generalized score function and characterize the escort expectations.

The rest of the paper is organized as follows. The next section provides us with the preliminaries for the basics of IG for the exponential families of pdf. In Section 3, after a brief review of conjugate reps. introduced by Zhang [7], we reformulate the IG structures based on the conjugate reps. and discuss the maximum entropy (MaxEnt) principle for a generalized entropy. For a set of conjugate reps., the associated generalized score function is introduced. The escort rep. and escort expectation are then naturally induced. Section 4 relates the conformal divergence to the difference of the entropies in terms of the escort expectations. The final section is devoted to our concluding remarks. Throughout the paper, we use the abbreviations ∂_i for $\partial/\partial\theta^i$, and ∂^i for $\partial/\partial\eta^i$.

2. Preliminaries

Information geometry [1,2] provides us a useful tool for studying a family

$$\mathcal{S} = \left\{ p_\theta(x) \ \middle| \ p_\theta(x) > 0, \int dx \, p_\theta(x) = 1 \right\}, \tag{1}$$

of a probability distribution function (pdf) $p_\theta(x)$ characterized by a set of real parameters $\boldsymbol{\theta} = (\theta^1, \theta^2 \dots, \theta^M)$. $\mathcal{S}$ is called a (M-dimensional) *statistical model* and the pdf $p_\theta(x)$ of $\mathcal{S}$ can be regarded as a point in a differential manifold $\mathcal{M}$ with local coordinates $\{\theta^i\}$. $\mathcal{M}$ is called a *statistical manifold* and a Riemannian metric on $\mathcal{M}$ is provided by the Fisher information matrix g_{ij}^{F} [2],

$$g_{ij}^{\mathrm{F}}(\boldsymbol{\theta}) = \mathbb{E}_{\mathrm{p}_\theta} \left[\partial_i \ell_\theta(x) \, \partial_j \ell_\theta(x) \right], \quad i,j = 1, 2, \dots, M, \tag{2}$$

where $\ell_\theta(x) \equiv \ln p_\theta(x)$. In this contribution, we assume that g^{F} is positive definite, and $\mathbb{E}_{\mathrm{p}_\theta}[\cdot]$ stands for the linear expectation with respect to the pdf $p_\theta(x)$.

A manifold $\mathcal{M}$ is said to be *e*-flat (*exponential-flat*) if a set of coordinate systems $\{\theta^i\}$ satisfies

$$\mathbb{E}_{\mathrm{p}_\theta} \left[\partial_i \partial_j \ell_\theta(x) \, \partial_k \ell_\theta(x) \right] = 0, \quad \forall \, i, j, k, \tag{3}$$

identically. Any set of coordinates $\{\theta^i\}$ satisfying (3) is called *e*-affine coordinates. A well-known example of *e*-flat manifolds is the exponential family

$$\mathcal{S}_{\exp} = \left\{ p_\theta(x) \ \middle| \ p_\theta(x) = \exp \left[\sum_{m=1}^{M} \theta^m F_m(x) - \Psi(\boldsymbol{\theta}) \right], \int dx \, p_\theta(x) = 1 \right\}, \tag{4}$$

where each $F_m(x)$ is a given function of a random value x and $\Psi(\boldsymbol{\theta})$ is the normalization factor of a pdf $p_\theta(x)$. From the normalization of the pdf $p_\theta(x)$, we see that

$$\mathbb{E}_{\mathrm{p}_\theta} \left[\partial_i \ell_\theta(x) \right] = \partial_i \mathbb{E}_{\mathrm{p}_\theta} [1] = 0, \tag{5}$$

and, for the exponential family, we have

$$\partial_i \partial_j \ell_\theta(x) = -\partial_i \partial_j \Psi(\boldsymbol{\theta}), \tag{6}$$

which does not depend on x. Hence, the condition (3) is satisfied and one confirms that the exponential family is *e*-flat. In addition, for the exponential family, we have

$$\partial_i \ell_\theta(x) = f_i(x) - \partial_i \Psi(\boldsymbol{\theta}). \tag{7}$$

Taking the expectation of both sides and using Equation (5), we see that the *m*-affine coordinates (*η*-coordinates) of the exponential family are given by

$$\eta_i = \mathbb{E}_{\mathrm{p}_\theta}\left[f_i(x)\right] = \partial_i \Psi(\boldsymbol{\theta}). \tag{8}$$

Accounting for Equations (7) and (8), from definition (2), we obtain

$$g_{ij}^{\mathrm{F}}(\boldsymbol{\theta}) = \mathbb{E}_{\mathrm{p}_\theta}\left[\left(f_i - \mathbb{E}_{\mathrm{p}_\theta}\left[f_i\right]\right)\left(f_j - \mathbb{E}_{\mathrm{p}_\theta}\left[f_j\right]\right)\right], \quad i,j = 1,2,\ldots,M, \tag{9}$$

which is the covariance matrix for the statistical model $\mathcal{S}_{\mathrm{exp}}$.

A manifold $\mathcal{M}$ is said to be *m*-flat (*mixture-flat*) if a set of coordinate systems $\{\eta_i\}$ satisfies

$$\mathbb{E}_{\mathrm{p}_\theta}\left[\frac{1}{p_\eta(x)}\partial^i\partial^j p_\eta(x)\,\partial^k \ln p_\eta(x)\right] = 0, \quad \forall\, i,j,k, \tag{10}$$

identically. In this case, the set of coordinates $\{\eta_i\}$ is called *m*-affine coordinates.

In a dually flat structure, the $\boldsymbol{\theta}$- and $\boldsymbol{\eta}$-coordinates are related by the Legendre transformation

$$\Psi(\boldsymbol{\theta}) + \Psi^*(\boldsymbol{\eta}) - \boldsymbol{\theta}\cdot\boldsymbol{\eta} = 0, \tag{11}$$

$$\theta^i = \partial^i \Psi^*(\boldsymbol{\eta}), \tag{12}$$

$$\eta_i = \partial_i \Psi(\boldsymbol{\theta}), \tag{13}$$

where $\Psi(\boldsymbol{\theta})$ and $\Psi^*(\boldsymbol{\eta})$ are Legendre–Fenchel dual to each other and are called θ- and η-potential functions, respectively. The canonical divergence function [2] for a set of two pdf $p_{\theta_p}(x)$ and $r_{\theta_r}(x)$ can be defined by

$$D(p|r) \equiv \Psi(\boldsymbol{\theta}_p) - \Psi(\boldsymbol{\theta}_r) - \nabla\Psi(\boldsymbol{\theta}_r)\cdot(\boldsymbol{\theta}_p - \boldsymbol{\theta}_r), \tag{14}$$

which is a Bregman divergence with the convex function $\Psi(\boldsymbol{\theta})$.

For a dually flat manifold $\mathcal{S}$, Pythagorean relation is generalized in terms of divergence. Let p, r, s be three probability distributions in $\mathcal{S}$. When the *e*-geodesic connecting p and r is orthogonal at r to the *m*-geodesic connecting r and s, the following generalized Pythagorean relation [2] holds.

$$D(p|r) + D(r|s) = D(p|s). \tag{15}$$

As is well known, maximizing the Boltzmann–Gibbs–Shannon (BGS) entropy

$$S^{\mathrm{BGS}} \equiv -\int dx\, p(x)\ln p(x) = \mathbb{E}_{\mathrm{p}_\theta}\left[-\ln p\right], \tag{16}$$

under the *M*-constraints

$$\mathbb{E}_{\mathrm{p}_\theta}\left[F_m(x)\right] = U_m, \quad m = 1,2,\cdots,M, \tag{17}$$

for a given set of U_m and the normalization $\int dx\, p(x) = 1$, leads to the optimized pdf belonging to the exponential family $\mathcal{S}_{\mathrm{exp}}$. The Lagrange multipliers are the control parameters $\{\theta^m\}$ for the above *M*-constraints. From the normalization of an exponential pdf, we readily obtain the θ-potential function $\Psi(\boldsymbol{\theta})$ as

$$\Psi(\boldsymbol{\theta}) = \ln\left(\int dx\, \exp\left[\sum_{m=1}^{M}\theta^m F_m(x)\right]\right). \tag{18}$$

We note that, in addition to Equation (2), the Fisher metric g^{F} can be written equivalently in other different expressions

$$g^{\mathrm{F}}_{ij} = \int dx\, \partial_i p_\theta(x)\, \partial_j \ell_\theta(x) \tag{19}$$

$$= -\int dx\, p_\theta(x)\, \partial_i\, \partial_j \ell_\theta(x) \tag{20}$$

$$= \int dx\, \frac{1}{p_\theta(x)}\, \partial_i p_\theta(x) \partial_j p_\theta(x). \tag{21}$$

In particular, combining Equation (6) with (20), we readily confirm the important relation

$$g^{\mathrm{F}}_{ij} = \partial_i \partial_j \Psi(\boldsymbol{\theta}), \tag{22}$$

that is, the Fisher metric coincides with the Hessian matrix of the θ-potential function $\Psi(\boldsymbol{\theta})$. It is known that an exponential family naturally has the dualistic Hessian structures and their canonical divergences coincide with the Kullback–Leibler divergences. Furthermore, using Equation (8), the Fisher matrix can also be rewritten as

$$g^{\mathrm{F}}_{ij} = \partial_i \eta_j = \partial_i \mathbb{E}_{\mathrm{p}_\theta}\left[f_j \right], \tag{23}$$

which holds for the exponential family $\mathcal{S}_{\exp}$.

In general, the dual affine connections are induced from the metric. By applying ∂_i to Equation (19) for g^{F}, we see that the following relation holds

$$\partial_i g^{\mathrm{F}}_{jk} = \Gamma^{(e)}_{ij,k} + \Gamma^{(m)}_{ij,k}, \tag{24}$$

where the Christoffel symbol of the first kind for the e-affine connection and that for the m-affine connection are defined by

$$\Gamma^{(e)}_{ij,k} \equiv \int dx\, \partial_k p_\theta(x) \partial_i \partial_j \ell_\theta(x) = \mathbb{E}_{\mathrm{p}_\theta}\left[\partial_k \ell_\theta\, \partial_i \partial_j \ell_\theta \right], \tag{25}$$

$$\Gamma^{(m)}_{ij,k} \equiv \int dx\, \partial_i \partial_j p_\theta(x) \partial_k \ell_\theta(x) = \mathbb{E}_{\mathrm{p}_\theta}\left[\frac{1}{p_\theta(x)} \partial_i \partial_j p_\theta(x)\, \partial_k \ell_\theta \right], \tag{26}$$

respectively. In addition, we can introduce a cubic form

$$C_{ijk} \equiv \Gamma^{(m)}_{ij,k} - \Gamma^{(e)}_{ij,k}, \tag{27}$$

which characterizes the difference between the affine connection $\nabla^{(e)}$ (or $\nabla^{(m)}$) and Levi–Civita connection $\nabla^{(0)}$ through the relations

$$\Gamma^{(e)}_{ij,k} = \Gamma^{(0)}_{ij,k} - \frac{1}{2} C_{ijk}, \tag{28}$$

$$\Gamma^{(m)}_{ij,k} = \Gamma^{(0)}_{ij,k} + \frac{1}{2} C_{ijk}. \tag{29}$$

3. Conjugate Representations

Here, we briefly review Zhang's conjugate representations [7]. For a parameterized probability density function (pdf) $p_\theta(x)$ with a set of real parameters $\{\theta^m\}, m = 1, 2, \ldots, M$, information geometry is founded by Prof. Amari [1], based on his α-representations (reps.) defined by

$$\ell^\alpha\big(p_\theta(x)\big) = \frac{2}{1-\alpha}\, p_\theta^{\frac{1-\alpha}{2}}(x), \quad \ell^{-\alpha}\big(p_\theta(x)\big) = \frac{2}{1+\alpha}\, p_\theta^{\frac{1+\alpha}{2}}(x), \tag{30}$$

for a real parameter $\alpha \neq 1$, and on the α-divergence

$$D^{\alpha}(p|r) = \frac{4}{1-\alpha^2} \int dx \left\{ \left(\frac{1-\alpha}{2} \right) p(x) + \left(\frac{1+\alpha}{2} \right) r(x) - p^{\frac{1-\alpha}{2}}(x) r^{\frac{1+\alpha}{2}}(x) \right\}. \tag{31}$$

As a generalization of the α-reps., Zhang [7] introduced the *conjugate representations* as follows.

Definition 1. *A ρ-representation of a real positive number ξ is a mapping $\xi \mapsto \rho(\xi)$, where $\rho(\xi)$ is a strictly monotone function. For a smooth and strictly convex function $f(\rho)$, a τ-representation: $\xi \mapsto \tau(\xi)$ is said to be* **conjugate** *to the ρ-representation with respect to $f(\rho)$ if the following relations are satisfied,*

$$\tau(\xi) = \left. \frac{df(\rho)}{d\rho} \right|_{\rho=\rho(\xi)}, \quad \rho(\xi) = \left. \frac{df^{\star}(\tau)}{d\tau} \right|_{\tau=\tau(\xi)}, \tag{32}$$

where the convex functions $f(\rho)$ and $f^{\star}(\tau)$ are Legendre dual to each other:

$$f(\rho) = \rho\tau(\rho) - f^{\star}(\tau(\rho)), \quad f^{\star}(\tau) = \rho(\tau)\tau - f(\rho(\tau)). \tag{33}$$

By utilizing the conjugate reps., the associated Bregman divergence can be defined as

$$D_{f,\rho}(p|r) = \int dx \Big[f(\rho(p(x))) - f(\rho(r(x))) - \tau(r(x))(\rho(p(x)) - \rho(r(x))) \Big], \tag{34}$$

$$D_{f^{\star},\tau}(p|r) = \int dx \Big[f^{\star}(\tau(p(x))) - f^{\star}(\tau(r(x))) - \rho(r(x))(\tau(p(x)) - \tau(r(x))) \Big]. \tag{35}$$

The α-rep. is, of course, an example of the conjugate reps., and they are related as follows.

$$\rho_{\alpha}(p) = \ell^{\alpha}(p), \qquad\qquad \tau_{\alpha}(p) = \ell^{-\alpha}(p), \tag{36}$$

$$f_{\alpha}(\rho_{\alpha}) = \frac{2}{1+\alpha} \left(\frac{1-\alpha}{2} \rho_{\alpha} \right)^{\frac{2}{1-\alpha}}, \quad f_{\alpha}^{\star}(\tau_{\alpha}) = \frac{2}{1-\alpha} \left(\frac{1+\alpha}{2} \tau_{\alpha} \right)^{\frac{2}{1+\alpha}}. \tag{37}$$

The α-divergence (31) is expressed as $D_{f_{\alpha},\rho_{\alpha}}(p|r)$.

Remark 1. *We assume that ρ- and τ-functions satisfy the suitable regularity conditions throughout this paper. It is important to describe the domains and the tangents of the relevant ρ- and τ-functions. However, this is a very difficult matter in general. For example, consider a statistical manifold which is a set of q-Gaussian distributions, and using the α-rep. (36). We see that it is an α-affine manifold with $\alpha = 2q - 1$ [1,7]. In this case, if the domain Ω (the total sample space) is $\mathbf{R}$, then α must satisfy $1 < \alpha < 5$. If $\Omega = \mathbf{R}^2$, then α must satisfy $1 < \alpha < 3$. The lower bound comes from the regularity conditions of the statistical manifold (Amari and Nagaoka [1], Chapter 2), and the upper bounds come from the integrability conditions of probability densities. In this way, the regularity conditions for a set of ρ- and τ-functions are not determined from these functions themselves only, but depend on the total sample space and the given statistical model. Some arguments have been given in our previous paper [12].*

3.1. MaxEnt

For a set $(\rho, \tau, f(\rho), f^{\star}(\tau))$ of conjugate reps., let us introduce a generalized entropy functional S defined by

$$S := -\int dx f^{\star}(\tau(p(x))), \tag{38}$$

and consider the following MaxEnt problem.

$$\frac{\delta}{\delta p(x)} \left[S + \sum_{m=1}^{M} \theta^m \int dx\, \tau(p(x))\, F_m(x) - \gamma \int dx\, \tau(p(x)) \right] = 0, \tag{39}$$

where $F_m(x)$ is a given function of x, and $\theta^m, m = 1, 2, \ldots, M$ and γ are the Lagrange multipliers. Using the relation $df^\star(\tau)/d\tau = \rho$ defined in Equation (32), this MaxEnt problem leads to

$$\rho(p(x))\, \tau'(p(x)) = \sum_{m=1}^{M} \theta^m F_m(x)\, \tau'(p(x)) - \gamma(\boldsymbol{\theta})\, \tau'(p(x)), \tag{40}$$

where $\tau'(p)$ stands for $d\tau(p)/dp$. We assume $\tau'(p(x)) \neq 0$ because if $\tau'(p) = 0$ then the τ-rep. is a constant mapping, which fails to work as a rep., or immersion, of a pdf $p(x)$. We thus obtain

$$\rho(p(x)) = \sum_{m=1}^{M} \theta^m F_m(x) - \gamma(\boldsymbol{\theta}). \tag{41}$$

Remark 2. *Note that unless $\tau(p) = p$, the constraints of this generalized MaxEnt problem are neither the standard expectations $\int dx p(x) F_m(x)$ nor the normalization $\int dx p(x) = 1$ of the pdf $p(x)$. However, the solution of this MaxEnt problem is expressed in terms of the inverse function of $\rho(p)$ as*

$$p(x) = \rho^{-1} \left(\sum_{m=1}^{M} \theta^m F_m(x) - \gamma(\boldsymbol{\theta}) \right). \tag{42}$$

Definition 2. *For any given ρ-rep., the generalized score function $s_\rho(x)$ is defined by*

$$s_\rho(x) := \partial_i \rho(p(x)). \tag{43}$$

Remark 3. *In the above MaxEnt setting, substituting Equation (42) into the generalized score function, we obtain that*

$$s_\rho(x) = F_i(x) - \partial_i \gamma. \tag{44}$$

Theorem 1. *For any set of conjugate reps., and the associated generalized score function $s(x)$,*

$$\int dx\, \tau(p(x))\, s_\rho(x) = \partial_i \int dx\, f(\rho(p(x))) \tag{45}$$

holds.

Proof. From the definition (32) of the conjugate reps., we see $\tau = df(\rho)/d\rho$. It follows that

$$\tau(p(x))\, \partial_i \rho(p(x)) = \frac{df(\rho)}{d\rho}\, \partial_i \rho(p(x)) = \partial_i f(\rho(p(x))), \tag{46}$$

and integrating both sides by x, we obtain the result. $\square$

Definition 3 (Escort rep.)**.** *For a given ρ-rep. which satisfies $d\rho(p)/dp \neq 0$, we can introduce a new $\tilde{\tau}$-rep., which is called the **escort rep.** of a pdf $p(x)$ and is defined by*

$$\tilde{\tau}(p) := \frac{c}{\frac{d\rho(p)}{dp}} = c\, \frac{d\rho^{-1}(\xi)}{d\xi}\bigg|_{\xi = \rho(p)}, \tag{47}$$

where c is an appropriate constant and $\rho^{-1}(\xi)$ is the inverse function of $\rho(\xi)$.

Remark 4. *For the α-reps., we have*

$$\rho_\alpha^{-1}(\xi) = \left(\frac{1-\alpha}{2} \xi \right)^{\frac{2}{1-\alpha}}. \tag{48}$$

We thus see that

$$\tilde{\tau}(p) = \frac{2}{1+\alpha} \frac{d\rho_\alpha^{-1}(\xi)}{d\xi} \bigg|_{\xi=\rho_\alpha(p)} = \frac{2}{1+\alpha} p^{\frac{1+\alpha}{2}} = \tau_\alpha(p), \tag{49}$$

which states that $\tau_\alpha(p)$ is a self-escort rep. with the constant $c = 2/(1+\alpha)$.

One of the merits of introducing the escort rep. $\tilde{\tau}(p)$ is the next theorem.

Theorem 2. *A $\tilde{\tau}$-rep. satisfies*

$$\int dx \, \tilde{\tau}(p(x)) s_\rho(x) = \int dx \, \tilde{\tau}(p(x)) \partial_i \rho(p(x)) = 0, \tag{50}$$

Proof. Substituting the relation

$$\partial_i \rho(p) = \frac{d\rho(p)}{dp} \partial_i p, \tag{51}$$

into Equation (50) leads to

$$\partial_i \int dx \, p(x) = \partial_i 1 = 0, \tag{52}$$

because the pdf $p(x)$ is normalized. □

For this $\tilde{\tau}$-rep., we can introduce the associated convex functions $\tilde{f}(\rho)$ and $\tilde{f}^\star(\tilde{\tau})$ that satisfy

$$\frac{d\tilde{f}(\rho)}{d\rho} = \tilde{\tau}, \quad \frac{d\tilde{f}^\star(\tilde{\tau})}{d\tilde{\tau}} = \rho, \tag{53}$$

respectively.

Note that combining Theorem 1 with Theorem 2 leads to

$$\int dx \, \tilde{\tau}(p(x)) s_\rho(x) = \partial_i \int dx \, \tilde{f}(\rho(p(x))) = 0. \tag{54}$$

We then obtain the following corollary

Corollary 1. *For the conjugate reps. ρ and $\tilde{\tau}$, the associated $\tilde{f}(\rho)$ function satisfies that*

$$\int dx \tilde{f}(\rho(p(x))) = c \int dx \, p(x) = c, \tag{55}$$

which is the constant c defined in Equation (47) for any normalized pdf $p(x)$.

Remark 5. *For the α-rep., we see that*

$$f_\alpha(\rho_\alpha(p)) = \frac{2}{1+\alpha} p. \tag{56}$$

Definition 4 (Escort pdf and escort exprectation)**.** *Define the escort pdf $P^{\text{esc}}(x)$ with regards to a pdf $p(x)$ by utilizing the escort rep. $\tilde{\tau}(p)$ as follows.*

$$P^{\text{esc}}(x) := \frac{\tilde{\tau}(p(x))}{\int dx \ \tilde{\tau}(p(x))}, \tag{57}$$

and define the escort expectation with regards to $p(x)$ of a given function $F_i(x)$ as

$$\mathbb{E}_{\text{P}_{\text{esc}}}[F_i] := \int dx \ P^{\text{esc}}(x) F_i(x). \tag{58}$$

Theorem 3. *In the MaxEnt setting of Equation* (39), *the score function $s_\rho(x)$ has zero-escort expectation, i.e., $\mathbb{E}_{\text{P}_{\text{esc}}}[s_\rho(x)] = 0$, and it follows that*

$$\partial_i \gamma(\boldsymbol{\theta}) = \mathbb{E}_{\text{P}_{\text{esc}}}[F_i]. \tag{59}$$

Proof. From Equation (41) we have

$$\mathbb{E}_{\text{P}_{\text{esc}}}[s_\rho] = \mathbb{E}_{\text{P}_{\text{esc}}}[\partial_i \rho(p(x))] = \mathbb{E}_{\text{P}_{\text{esc}}}[F_i] - \partial_i \gamma, \tag{60}$$

where we used $\mathbb{E}_{\text{P}_{\text{esc}}}[1] = 1$. Since $\mathbb{E}_{\text{P}_{\text{esc}}}[s_\rho] = 0$, we obtain Equation (59). $\square$

Remark 6. *In our formalism, the escort expectation is characterized by the generalized score function $s_\rho(x)$ which is unbiased, i.e., $s_\rho(x)$ has zero-escort expectation.*

We see that the Lagrange multiplier $\gamma(\boldsymbol{\theta})$ is the θ-potential $\check{\Psi}(\boldsymbol{\theta})$ associated with the escort expectation. The dual affine coordinate $\boldsymbol{\eta}^{\text{esc}}$ is

$$\eta_i^{\text{esc}} = \partial_i \gamma(\boldsymbol{\theta}) = \mathbb{E}_{\text{P}_{\text{esc}}}[F_i], \tag{61}$$

and the associated Riemannian metric and cubic form are

$$g_{ij}^{\text{esc}}(\boldsymbol{\theta}) = \partial_i \partial_j \gamma(\boldsymbol{\theta}), \tag{62}$$

$$C_{ijk}^{\text{esc}}(\boldsymbol{\theta}) = \partial_i \partial_j \partial_k \gamma(\boldsymbol{\theta}), \tag{63}$$

respectively. Since g^{esc} is a Hessian metric, the statistical manifold described by the $\boldsymbol{\theta}$- and $\boldsymbol{\eta}^{\text{esc}}$-coordinates is dually flat.

4. Conformal Divergence

Let us consider the Bregman divergence (35) of the escort reps., i.e.,

$$D_{\tilde{f}^\star, \tilde{\tau}}(p|r) = \int dx \left[\tilde{f}^\star(\tilde{\tau}(p(x))) - \tilde{f}^\star(\tilde{\tau}(r(x))) - \rho(r(x))(\tilde{\tau}(p(x)) - \tilde{\tau}(r(x))) \right]. \tag{64}$$

The next theorem is a main result of this contribution.

Theorem 4. *The relative escort expectation of ρ-reps. is the conformal (or scaled) divergence of $D_{\tilde{f}^\star, \tilde{\tau}}(p|r)$ with the scaling factor $1/\int dx \tilde{\tau}(p(x))$, i.e.,*

$$\mathbb{E}_{\text{P}_{\text{esc}}}[\rho(p(x))] - \mathbb{E}_{\text{P}_{\text{esc}}}[\rho(r(x))] = \frac{1}{\int dx \tilde{\tau}(p(x))} D_{\tilde{f}^\star, \tilde{\tau}}(p|r). \tag{65}$$

Proof. From Corollary 1, we see that

$$\int dx \tilde{f}^{\star}(\tilde{\tau}(p(x))) = \int dx \tilde{\tau}(p(x)) \rho(p(x)) - c,\tag{66}$$

where c is an appropriate constant for any normalized pdf $p(x)$, and it follows that

$$\int dx \left[\tilde{f}^{\star}(\tilde{\tau}(p(x))) - \tilde{f}^{\star}(\tilde{\tau}(r(x))) \right] = \int dx \left[\tilde{\tau}(p(x)) \rho(p(x)) - \tilde{\tau}(r(x)) \rho(r(x)) \right].\tag{67}$$

Substituting this relation into Equation (64) leads to

$$D_{\tilde{f}^{\star}, \tilde{\tau}}(p|r) = \int dx \tilde{\tau}(p(x)) \left\{ \rho(p(x)) - \rho(r(x)) \right\}.\tag{68}$$

Dividing both sides by $\int dx \tilde{\tau}(p(x))$ and using the escort expectation, we obtain the result. □

Remark 7. *As an example of Theorem 4, let us consider the α-rep. case. Since $\tilde{\tau}_\alpha = \tau_\alpha$ as shown in Remark 4, it follows that $\tilde{f}_\alpha^{\star}(\tilde{\tau}) = f_\alpha^{\star}(\tau)$. The corresponding escort pdf becomes*

$$P_\alpha^{\text{esc}}(x) = \frac{p^{\frac{1+\alpha}{2}}}{\int dx \, p^{\frac{1+\alpha}{2}}},\tag{69}$$

and Equation (65) becomes

$$\frac{2}{1-\alpha} \left(\mathbb{E}_{\text{P}_{\text{esc}}} \left[p^{\frac{1-\alpha}{2}} \right] - \mathbb{E}_{\text{P}_{\text{esc}}} \left[r^{\frac{1-\alpha}{2}} \right] \right) = \frac{(1+\alpha)/2}{\int dx \, p^{\frac{1+\alpha}{2}}(x)} \, D_{\tilde{f}_\alpha^{\star}, \tau_\alpha}(p|r).\tag{70}$$

When we set $q = (1+\alpha)/2$ this relation becomes

$$\mathbb{E}_{\text{P}_{\text{esc}}} \left[\ln_q p \right] - \mathbb{E}_{\text{P}_{\text{esc}}} \left[\ln_q r \right] = \frac{q}{\int dx \, p^q(x)} \, D_{\tilde{f}_\alpha^{\star}, \tau_\alpha}(p|r),\tag{71}$$

which was first shown by Matsuzoe and Ohara [13].

5. Concluding Remarks

We have discussed and reformulated the method of information geometry in terms of the conjugate reps. introduced by Zhang [7]. For an appropriate set of conjugate reps., the MaxEnt principle for a generalized entropy relates the associated Lagrange multipliers to the corresponding ρ rep. (41) of the optimal pdf (42). For a generalized score function (2), the escort rep. and escort expectation are then naturally induced. The conformal divergence is related to the difference of the entropies in terms of the escort expectations, as shown in Theorem 4.

In previous work [9], we studied, for the κ-deformed exponential family, the dualistic Hessian geometries among the thermodynamic potentials in the κ-deformed thermostatistics, and found that there exist two different kinds of dual affine-coordinates: one η is associated with the standard expectation; and the other η^{esc} is associated with the escort expectation. There, the double escort distributions, i.e., the escort of the escort distributions, play an important role. For the q-deformed exponential family, one of the authors (H.M.) further studied a sequence (or hierarchy) of escort distributions [14]. We think that these results are not specific to the q- or κ-deformed exponential pdf. We believe that these results [9,14] can be systematically studied by applying the reformulated method developed in this work. Further studies are needed and will be carried out in future work.

Author Contributions: Tatsuaki Wada designed the main subject of this research and mainly wrote the manuscript. Hiroshi Matsuzoe commented on the manuscript at all stages. All authors equally promoted the research and discussed the results. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Amari, S.-I.; Nagaoka, H. *Method of Information Geometry*; Oxford University Press: Oxford, UK, 2000.
2. Amari, S.-I. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
3. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: New York, NY, USA, 2009.
4. Kaniadakis, G. Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions. *Entropy* **2013**, *15*, 3983–4010.
5. Naudts, J. *Generalized Thermostatistics*; Springer: Berlin, Germany, 2011.
6. Amari, S.-I.; Ohara, A. A Geometry of q-Exponential Family of Probability Distributions. *Entropy* **2011**, *13*, 1170–1185.
7. Zhang, J. Divergence function, duality and convex analysis. *Neural Comput.* **2004**, *16*, 159–195.
8. Wada, T.; Scarfone, A.M. Information geometry on the κ-thermostatistics. *Entropy* **2015**, *17*, 1204–1217.
9. Wada, T.; Matsuzoe, H.; Scarfone, A.M. Dualistic Hessian structures among the thermodynamic potentials in the κ-thermostatistics. *Entropy* **2015**, *17*, 7213–7229.
10. Zhang, J. On monotone embedding in information geometry. *Entropy* **2015**, *17*, 4485–4499.
11. Amari, S.-I. Information Geometry of Positive Measures and Positive-Definite Matrices: Decomposable Dually Flat Structure. *Entropy* **2014**, *16*, 2131–2145.
12. Matsuzoe, H.; Wada, T. Deformed Algebras and Generalizations of Independence on Deformed Exponential Families. *Entropy* **2015**, *17*, 5729–5751.
13. Matsuzoe, H.; Ohara, A. Geometry for q-exponential families. In *Recent Progress in Differential Geometry and Its Related Fields, Proceedings of the 2nd International Colloquium on Differential Geometry and Its Related Fields, Veliko Tarnovo, Bulgaria, 6–10 September 2010*; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific: Hackensack, NJ, USA, 2011; pp. 55–71.
14. Matsuzoe, H. A sequence of escort distributions and generalizations of expectations on q-exponential family. *Entropy* **2017**, *19*, 7.