

Review

# $\lambda$ -Deformation: A Canonical Framework for Statistical Manifolds of Constant Curvature

Jun Zhang<sup>1,2,\*</sup> and Ting-Kam Leonard Wong<sup>3</sup> 

<sup>1</sup> Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1109, USA

<sup>2</sup> Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1109, USA

<sup>3</sup> Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 1A1, Canada; tkl.wong@utoronto.ca

\* Correspondence: junz@umich.edu

**Abstract:** This paper systematically presents the  $\lambda$ -deformation as the canonical framework of deformation to the dually flat (Hessian) geometry, which has been well established in information geometry. We show that, based on deforming the Legendre duality, all objects in the Hessian case have their correspondence in the  $\lambda$ -deformed case:  $\lambda$ -convexity,  $\lambda$ -conjugation,  $\lambda$ -biorthogonality,  $\lambda$ -logarithmic divergence,  $\lambda$ -exponential and  $\lambda$ -mixture families, etc. In particular,  $\lambda$ -deformation unifies Tsallis and Rényi deformations by relating them to two manifestations of an identical  $\lambda$ -exponential family, under subtractive or divisive probability normalization, respectively. Unlike the different Hessian geometries of the exponential and mixture families, the  $\lambda$ -exponential family, in turn, coincides with the  $\lambda$ -mixture family after a change of random variables. The resulting statistical manifolds, while still carrying a dualistic structure, replace the Hessian metric and a pair of dually flat conjugate affine connections with a conformal Hessian metric and a pair of projectively flat connections carrying constant (nonzero) curvature. Thus,  $\lambda$ -deformation is a canonical framework in generalizing the well-known dually flat Hessian structure of information geometry.

**Keywords:** Legendre duality;  $\lambda$ -duality;  $\lambda$ -exponential family;  $\lambda$ -mixture family; conformal Hessian; constant curvature space



**Citation:** Zhang, J.; Wong, T.-K.L.

$\lambda$ -Deformation: A Canonical Framework for Statistical Manifolds of Constant Curvature. *Entropy* **2022**, *24*, 193. <https://doi.org/10.3390/e24020193>

Academic Editor: Miguel Rubi

Received: 15 December 2021

Accepted: 18 January 2022

Published: 27 January 2022

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



**Copyright:** © 2022 by the authors. 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

Information geometry is a differential-geometric framework for studying finite-dimensional statistical models that coherently integrates the following notions:

- (i) A differentiable manifold  $\mathcal{M}$  consisting of probability density functions or finite measures on a common sample space;
- (ii) A divergence function  $D[p||p']$  that defines an asymmetric proximity between points  $p, p'$  in  $\mathcal{M}$ ;
- (iii) A Riemannian metric  $g$  plus a pair of torsion-free dual (conjugate) affine connections  $\nabla, \nabla^*$  on  $\mathcal{M}$ .

For completeness, we recall that a pair of affine connections  $\nabla, \nabla^*$  on  $\mathcal{M}$  are said to be dual (or conjugate) with respect to a Riemannian metric  $g$  if for any vector fields  $X, Y$ , and  $Z$  on  $\mathcal{M}$ , one has:

$$Zg(X, Y) = g(\nabla_Z X, Y) + g(X, \nabla_Z^* Y). \quad (1)$$

Here,  $(\mathcal{M}, g, \nabla, \nabla^*)$  is called a *dualistic structure*. When  $D$  is the Kullback–Leibler divergence (or more generally,  $f$ -divergence), the induced Riemannian metric  $g$  is the Fisher–Rao metric, and the induced cubic form  $C = \nabla^* - \nabla$  is the Amari–Chentsov tensor [1]. It can be shown that the Fisher–Rao metric and the Amari–Chentsov tensor are unique invariants, of respectively second and third orders, under sufficient statistics on the manifold  $\mathcal{M}$  [2].

Geometrically, the standard model (denoted the  $\mathcal{S}$ -model in this paper) uses a pair of affine connections that are torsion-free, though in general, they are not curvature-free. An alternative, “partially flat” model (denoted the  $\mathcal{P}$ -model in this paper) was recently investigated in [3], leading to the notion of “statistical mirror symmetry” [4]. Under the  $\mathcal{P}$ -model, the affine connections  $\nabla$  and  $\nabla^*$  are allowed to carry torsion, but are both curvature-free. See [4] for the geometric properties of the  $\mathcal{P}$ -model leading to a symplectic-to-complex correspondence characteristic of mirror Calabi–Yau manifolds studied in string theory and mathematical physics.

Within the usual  $\mathcal{S}$ -model, a special case is the *dually flat geometry* where the Riemannian metric can be expressed under special coordinate systems as a Hessian metric. Two prominent examples are the exponential family and the mixture family, where the Hessian metric coincides with the Fisher–Rao metric. The Hessian geometry is said to be dually flat because the Riemann curvature tensors of both the primal and the dual connections vanish; the corresponding primal and dual affine coordinate systems are linked via Legendre transformations by a pair of convex potentials. For an exponential family, these coordinates are precisely the natural (canonical) and mixture (expectation) coordinate systems, respectively. Note that the Hessian metric itself is not flat as its Levi-Civita connection contains curvature in general.

Between the well-understood dually flat Hessian geometry and the full-blown  $\mathcal{S}$ -model, there is a wide range of geometries capturing various probability models. Of special interest are generalizations of the exponential family, namely *deformed exponential families*. The  $\phi$ -exponential family was introduced in the context of statistical physics [5]; it was later shown [6] to be equivalent to the  $U$ -model [7] motivated by applications in machine learning—[6] revealed that both the  $\phi$ - and  $U$ -models can be generated from the  $(\rho, \tau)$ -model [8] through the mechanism of “gauge selection”. The  $(\rho, \tau)$ -metric generalizes the Fisher–Rao metric and may lead to a conformal Hessian metric for a  $\phi$ -exponential family. However, the connections are typically not curvature-free unless a special type of gauge is selected; this underlies the geometric characterization of the  $q$ -exponential model of Tsallis by [9–11].

In recent years, the second author [12], motivated by previous works with Pal on mathematical finance and optimal transport [13–16], studied a class of deformed exponential families generating *constant curvatures* through the use of a new divergence function called *logarithmic divergence*. By constant (information geometric) curvature, we mean that both the primal and dual Riemann curvature tensors have (the same) constant sectional curvature with respect to  $g$ . In [17], the present authors developed a unified framework, based on the notions of  $\lambda$ -duality and the  $\lambda$ -exponential family, which appears to provide a *canonical* extension of the dually flat geometry to the constant curvature case. Previously, statistical manifolds with constant curvature were studied using the abstract tools of affine differential geometry; see, e.g., [1,18] (also see [19]). Our framework provides a concrete approach and an explicit construction that elucidates how the properties of the exponential family and the dually flat geometry may be extended to the constant curvature case. In this paper, a careful exposition of the  $\lambda$ -deformation framework is provided from the perspective of  $\lambda$ -duality, namely the  $\lambda$ -deformation of Legendre duality.

The rest of the paper is organized as follows. In Section 2, we review the standard  $\mathcal{S}$ -model of information geometry with a focus on the dually flat geometry, based on convex duality and Bregman divergence, of the exponential and mixture families. The section closes with a preview of  $\lambda$ -deformation by introducing a suite of four deformation functions, as two pairs of mutually inverse functions:  $\log_\lambda$  versus  $\exp_\lambda$  and  $\kappa_\lambda$  versus  $\gamma_\lambda$ , with the first pair deforming  $\log$  and  $\exp$  and the second pair deforming the identity function. In Section 3, we describe the  $\lambda$ -duality, which deforms the standard convex duality. In particular, we compare  $\lambda$ -duality and standard Legendre duality and show their relations to each other upon a change of parameterization. In Section 4, we define the  $\lambda$ -gradient and then the  $\lambda$ -logarithmic divergence and study the constant curvature information geometry the latter induces. In Section 5, we relate  $\lambda$ -divergence to Rényi entropy by introducing the

$\lambda$ -exponential and  $\lambda$ -mixture family. The two expressions of the  $\lambda$ -exponential family under divisive and subtractive normalization correspond to, respectively, Rényi deformation and Tsallis deformation. Section 6 concludes with a comparison of  $\lambda$ -deformation with the standard dually flat (Hessian) framework.

## 2. The Standard Model of Information Geometry

### 2.1. The Standard Model

We begin by recalling the standard framework (referred to as the  $\mathcal{S}$ -model) of parametric information geometry [1,20]. Let  $\mathcal{M}$  be a finite-dimensional differentiable manifold with dimension  $d$  and  $\theta = (\theta^1, \dots, \theta^d)$  be a local coordinate system. The most important case is where  $\mathcal{M}$  is a manifold of parametric probability density functions. However, the idea of deforming Legendre duality to  $\lambda$ -duality and hence dually flat (Hessian) manifolds to manifolds of constant curvature discussed in Sections 3 and 4 is entirely general and does not rely on  $\mathcal{M}$  being a manifold of probability density functions.

Let  $(\mathcal{X}, \mu)$  be a measure space, where  $\mu$  is called the reference (or dominating) measure. Let  $\Theta \subset \mathbb{R}^d$  be an open domain. A parametric family of density functions is a mapping  $\theta \in \Theta \mapsto p(\cdot|\theta)$ , where each  $p(\cdot|\theta)$  is a probability density function with respect to  $\mu$ , i.e.,  $\int_{\mathcal{X}} p(\zeta|\theta) d\mu(\zeta) = 1$ . We assume that the family is sufficient regular such that all analytical operations (such as differentiation under the integral sign) can be performed as needed.

While a dualistic structure  $(\mathcal{M}, g, \nabla, \nabla^*)$  can be defined abstractly, in practice, it is often constructed by a *divergence*, namely a smooth, non-negative function  $D[\cdot|\cdot]$  on  $\mathcal{M} \times \mathcal{M}$  such that  $D[p||p'] = 0$  only if  $p = p'$  and the (0,2)-tensor  $g$  it induces on  $\mathcal{M}$  (see (2) below) is positive definite. Intuitively,  $D[p||p']$  defines a notion of “asymmetric distance” between points  $p$  and  $p'$  of  $\mathcal{M}$ . When  $\mathcal{M}$  is a manifold of density functions, a prominent example is the *Kullback–Leibler (KL) divergence* (relative entropy) given by:

$$H[p||p'] = \int p \log \frac{p}{p'} d\mu.$$

When dealing with parametric probability families,  $p$  and  $p'$  are replaced by  $p(\cdot|\theta)$  and  $p(\cdot|\theta')$ , then we denote  $D[p||p']$  as  $D(\theta, \theta')$  with an abuse of notation, that is:

$$D[p(\cdot|\theta)||p(\cdot|\theta')] \equiv D(\theta, \theta'),$$

and similarly for  $H$  as well—the notation of  $[p||p']$  in the divergence for probability density functions emphasizes the non-symmetry in  $p, p'$ ; see [1].

Eguchi [21] showed that any divergence function (called a “contrast function” there) induces a dualistic structure  $(\mathcal{M}, g, \nabla, \nabla^*)$ . In local coordinates, given  $D(\theta, \theta')$ , the components  $g_{ij}$  of the metric  $g$  are given by:

$$g_{ij}(\theta) = - \left. \frac{\partial^2}{\partial \theta^i \partial \theta^j} D(\theta, \theta') \right|_{\theta=\theta'} = \left. \frac{\partial^2}{\partial \theta^i \partial \theta^j} D(\theta, \theta') \right|_{\theta=\theta'}, \tag{2}$$

and the Christoffel symbols of the conjugate connections  $\nabla$  and  $\nabla^*$  are given respectively by:

$$\Gamma_{ij,k}(\theta) = - \left. \frac{\partial^3}{\partial \theta^i \partial \theta^j \partial \theta^k} D(\theta, \theta') \right|_{\theta=\theta'}, \quad \Gamma_{ij,k}^*(\theta) = - \left. \frac{\partial^3}{\partial \theta^i \partial \theta^j \partial \theta^k} D(\theta, \theta') \right|_{\theta=\theta'}. \tag{3}$$

Conversely, given any dualistic structure  $(\mathcal{M}, g, \nabla, \nabla^*)$ , there exists a divergence  $D$  that induces it, but this  $D$  is not unique in general [22]. Thus, the standard model  $\mathcal{S}$  is completely encoded by the choice of a divergence function.

### 2.2. Dually Flat Geometry

The most important example of a dualistic structure is the *dually flat geometry*, which is induced by a Bregman divergence [23]. Let  $\mathcal{M}$  be prescribed with an affine coordinate chart

$\theta \in \Theta$  on an open convex set  $\Theta \subset \mathbb{R}^d$ . Let  $\phi : \Theta \rightarrow \mathbb{R}$  be a differentiable convex function; specifically, we assumed that  $\phi$  is  $C^2$  and its Hessian  $D^2\phi$  is strictly positive definite. The Bregman divergence of  $\phi$  is defined by:

$$B_\phi(\theta, \theta') = \phi(\theta) - \phi(\theta') - D\phi(\theta') \cdot (\theta - \theta'), \quad \theta, \theta' \in \Theta,$$

where  $D\phi(\theta) = (\frac{\partial}{\partial\theta^1}\phi(\theta), \dots, \frac{\partial}{\partial\theta^d}\phi(\theta))^\top$  is the Euclidean gradient and  $a \cdot b$  denotes the standard dot product. We call  $\theta \in \Theta$  the *primal coordinates*, and  $\eta = D\phi(\theta)$  the *dual coordinates*, where the inverse of  $D\phi$  is given by  $\theta = D\phi^*(\eta)$ . Here, the Legendre conjugate  $\phi^*$  (or convex conjugate) of  $\phi$  is defined by:

$$\phi^*(\eta) = \sup_{\theta} (\theta \cdot \eta - \phi(\theta)). \tag{4}$$

Then, the components  $g_{ij}$  of the Riemannian metric  $g$ , under the respective local coordinate system, are given by:

$$g_{ij}(\theta) = \frac{\partial^2}{\partial\theta^i\partial\theta^j}\phi(\theta), \quad g_{ij}(\eta) = \frac{\partial^2}{\partial\eta^i\partial\eta^j}\phi^*(\eta). \tag{5}$$

In particular,  $g$  is a *Hessian metric* with potential  $\phi$  (resp.  $\phi^*$ ) under  $\theta$  (resp.  $\eta$ ). Furthermore, the Christoffel symbols of  $\nabla$  and  $\nabla^*$  are given respectively by:

$$\Gamma_{ij,k}(\theta) = 0, \quad \Gamma_{ij,k}^*(\eta) = 0. \tag{6}$$

From (6), we see that the Riemann curvature tensors of both  $\nabla$  and  $\nabla^*$  vanish. Thus, we call this a *dually flat* geometry. Furthermore, a  $\nabla$ -geodesic (resp.  $\nabla^*$ -geodesic) is a constant velocity straight line under the  $\theta$  (resp.  $\eta$ ) coordinate system.

Moreover, the  $\theta$  and  $\eta$  coordinates are *biorthogonal* in the sense that:

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta^j}\right) = \delta_{ij}, \tag{7}$$

and the Bregman divergence takes the forms of:

$$B_\phi(\theta, \theta'(\eta')) = A_\phi(\theta, \eta') = A_{\phi^*}(\eta', \theta) = B_{\phi^*}(\eta', \eta(\theta))$$

with  $\eta = D\phi(\theta)$  and  $\theta' = D\phi^*(\eta')$ , where  $A$  is called the *canonical divergence*:

$$A_\phi(\theta, \eta') = \phi(\theta) + \phi^*(\eta') - \theta \cdot \eta' = A_{\phi^*}(\eta', \theta). \tag{8}$$

Following [24,25], we call the equality between two expressions of  $B$  and the equality between two expressions of  $A$  in (8) *reference–representation biduality*. In [26], the identity (8) was used to motivate a family of *Fenchel–Young losses* in the context of regularized prediction in machine learning. Last but not least, the Bregman divergence satisfies the *generalized Pythagorean theorem*: given points  $P, Q$ , and  $R$ , we have the equality:

$$B_\phi(\theta_Q, \theta_P) + B_\phi(\theta_R, \theta_Q) = B_\phi(\theta_R, \theta_P)$$

if and only if the  $\nabla$ -geodesic between  $Q$  and  $R$  and the  $\nabla^*$ -geodesic between  $Q$  and  $P$  meet  $g$ -orthogonally at  $Q$ . As we will see in Section 4, all the properties above have natural generalizations under our  $\lambda$ -framework. We stress that the dually flat geometry depends crucially on classical convex (or Legendre) duality, as seen from (4) and (8).

### 2.3. Exponential and Mixture Families

The dually flat Hessian geometry arises naturally in the *exponential* and *mixture families* of probability densities. Given a reference measure  $\mu$  on a state space  $\mathcal{X}$ , an *exponential family* is a parameterized probability density  $p^{(e)}(\cdot|\theta)$  of the form:

$$p^{(e)}(\zeta|\theta) = e^{\theta \cdot F(\zeta) - \phi(\theta)}, \tag{9}$$

where  $\theta = (\theta^1, \dots, \theta^d) \in \Theta \subseteq \mathbb{R}^d$  and  $F(\zeta) = (F_1(\zeta), \dots, F_d(\zeta))$  is a vector of *sufficient statistics*. In (9), the *cumulant generating function*  $\phi$ , defined by:

$$\phi(\theta) = \log \int e^{\theta \cdot F} d\mu,$$

enforces the normalization  $\int p^{(e)} d\mu = 1$ . The exponential family generalizes the Boltzmann–Gibbs distribution in statistical physics, where  $Z(\theta) = e^{\phi(\theta)}$  is called the partition function.

The information geometry of the exponential family begins with the observation that  $\phi$  is convex. Then,  $\phi$  defines a Bregman divergence  $B_\phi$  giving rise to a dually flat structure. It can be shown that the Bregman divergence is a KL-divergence:

$$B_\phi(\theta, \theta') = H[p^{(e)}(\cdot|\theta') || p^{(e)}(\cdot|\theta)].$$

The induced Riemannian metric  $\mathbf{g}$ , the Fisher–Rao metric given by (in matrix components  $g_{ij}$ ), becomes a Hessian metric  $D^2\phi$ :

$$g_{ij}(\theta) = \int \left( \frac{\partial}{\partial \theta^i} \log p^{(e)}(\zeta|\theta) \right) \left( \frac{\partial}{\partial \theta^j} \log p^{(e)}(\zeta|\theta) \right) p^{(e)}(\zeta|\theta) d\mu = \frac{\partial^2}{\partial \theta^i \partial \theta^j} \phi(\theta).$$

Equivalently,  $\mathbf{g}(\theta)$  is the covariance matrix of the sufficient statistics  $F$ :

$$g_{ij}(\theta) = \int p^{(e)}(\zeta|\theta) \left( F_i(\zeta) - \int p^{(e)}(\zeta|\theta) F_i(\zeta) \right) \left( F_j(\zeta) - \int p^{(e)}(\zeta|\theta) F_j(\zeta) \right) d\mu.$$

Furthermore, the dual coordinate  $\eta = D\phi(\theta)$  is the *expectation coordinates* given by:

$$\eta = \int p^{(e)}(\zeta|\theta) F(\zeta) d\mu,$$

and the dual potential function  $\phi^*$  is, as a function of  $\eta$ , the negative Shannon entropy:

$$\phi^*(\eta) = -H[p^{(e)}(\cdot|\theta)] = \int p^{(e)}(\zeta|\theta) \log p^{(e)}(\zeta|\theta) d\mu.$$

A theoretical justification for the exponential family is that it maximizes the Shannon entropy under the constraints of its expected value of the vector of random functions  $F(\cdot)$ .

The *mixture family* is another probability family that is very useful in both theory and applications. Let  $P_0(\zeta), P_1(\zeta), \dots, P_d(\zeta)$  be a set of affinely independent probability densities with respect to the same dominating measure  $\mu$ . Given mixture parameters  $\eta_i > 0$  for  $i = 0, \dots, d$  with  $\sum_{i=0}^d \eta_i = 1$ , the mixture family  $p^{(m)}(\cdot|\eta)$  is defined by:

$$p^{(m)}(\zeta|\eta) = \sum_{i=0}^d \eta_i P_i(\zeta) = P_0(\zeta) + \sum_{i=1}^d \eta_i (P_i(\zeta) - P_0(\zeta)),$$

where  $(\eta_1, \dots, \eta_d)$  may be taken as the independent parameters. It can be shown that the negative Shannon entropy:

$$\psi(\eta) = -H[p^{(m)}(\cdot|\eta)] = \int p^{(m)}(\zeta|\eta) \log p^{(m)}(\zeta|\eta) d\mu$$

of a mixture family is convex in  $\eta$ . Using  $\psi$  as the potential function, we have:

$$B_{\psi}(\eta, \eta') = H[p^{(m)}(\cdot|\eta) || p^{(m)}(\cdot|\eta')],$$

which is again a KL-divergence and induces a dually flat geometry. In summary, the exponential and mixture families are both dually flat when the geometry is induced by the KL-divergence.

For completeness, we note that the convex conjugate of  $\psi(\eta)$  is:

$$\psi^*(\theta) = - \int P_0(\zeta) \log p^{(m)}(\zeta|\eta) d\mu,$$

with conjugate parameters  $\theta = D\psi$  given by:

$$\theta^i = \int (P_i(\zeta) - P_0(\zeta)) \log p^{(m)}(\zeta|\eta) d\mu.$$

#### 2.4. Deforming exp and log

The exponential function used in the exponential family:

$$p^{(e)}(\zeta|\theta) = \exp\{\theta \cdot F(\zeta) - \phi(\theta)\} = \frac{e^{\theta \cdot F(\zeta)}}{Z(\theta)}$$

allows the cumulant generating function  $\phi(\theta)$  (also called the potential function) and the partition function  $Z(\theta)$  to be linked by the simple relation  $\phi = \log Z$ . The equivalence of using  $\phi$  as subtractive normalization and  $Z$  as divisive normalization of the same exponential family  $\int p^{(e)}(\zeta|\theta) d\mu = 1$  is due to the elementary, but crucial property  $\exp(x + y) = \exp(x) \exp(y)$  of the exponential function. Using a functional form other than exp (exponential function) or log (logarithm function) is referred to as *deformation* in information geometric (statistical and information theoretic) contexts, and the resulting probability families are called “deformed” families. Typically, this is performed by regarding log, or equivalently exp, as special cases of some parametric class of functions that include them as special members.

More generally, the exponential/logarithmic function can be considered within a non-parametric function space that includes exp or log as a special member. Several approaches can be found in the literature, including the  $\phi$ -deformed exponential approach by Naudts [5,27,28], the conjugate  $(\rho, \tau)$ -embedding approach by the first author [8,25,29], and the  $U$ -model by Eguchi [7,30]. The  $\phi$ -model and  $U$ -model are both one-function models, while the  $(\rho, \tau)$ -model uses two free functions. It eventually became clear in the 2018 paper [6] by Naudts and the first author that (i) the  $\phi$ - and  $U$ -model turned out to be equivalent; (ii) they are special cases of the  $(\rho, \tau)$ -model upon a particular fixing of the “gauge freedom”; (iii) the corresponding  $(\rho, \tau)$ -geometry of the manifold of the  $\phi$ -exponential family can have different appearances (gauges freedom), such as a Hessian geometry (under one type of gauge selection) and a conformal Hessian geometry (under another type of gauge selection). The work [6] unified the intermediary results in [10,11,31] and provided a general deformation framework that preserves the rigid interlocking of: (i) the functional form of entropy, cross-entropy, and relative entropy (divergence); (ii) the functional form of the deformed probability family with the corresponding normalization and potential and the duality between the natural and expectation parameterizations; (iii) the expressions of the Riemannian metric (Fisher–Rao metric in general and Hessian metric in particular) and of the conjugate connections. Some of these concepts have their correspondence in nonparametric probability families as well [32–34].

Although the  $(\rho, \tau)$ -model may admit a conformal Hessian metric (more rigorously stated: the  $\phi$ -exponential family with the  $(\rho, \tau)$ -metric under a certain gauge will lead to conformal Hessian geometry), the dual connections are not projectively flat (as the geometry studied by [12]). As a result, while the connections are not flat (torsion-free, but

not curvature-free), they are not in general of the constant-curvature-type either. Therefore they are “too general” and do not generate the space of constant curvatures.

2.5. Highlights of  $\lambda$ -Deformation

Here enters  $\lambda$ -deformation as a middle ground [17]. The  $\lambda$ -deformation theory absorbs the  $q$ -deformation model of Tsallis and the  $F^{(\pm\alpha)}$  model of Wong [12] in deforming the exponential family and unifies the subtractive and divisive normalization—this is an occasion where subtractive and divisive normalizations are still linked by a simple reparameterization of the probability family.

Let us introduce some notations. Consider the following deformed logarithm and exponential functions (note the slight difference to the  $\log_q$  notation used by Tsallis, in the way how the subscript indicates the deformation parameter):

$$\log_\lambda(t) = \frac{1}{\lambda} (t^\lambda - 1), \quad \exp_\lambda(t) = (1 + \lambda t)^{1/\lambda}.$$

More precisely, we define  $\exp_\lambda : \mathbb{R} \rightarrow [0, \infty]$  (where  $\lambda \in \mathbb{R}, \lambda \neq 0$ ) by:

$$\exp_\lambda(t) = [1 + \lambda t]_+^{1/\lambda},$$

where  $[a]_+ = \max\{a, 0\}$ . In our analysis, we assumed implicitly that  $1 + \lambda t > 0$ , which is shown to hold for  $\lambda$ -duality, so the subscript  $+$  can be omitted. Furthermore,  $\frac{d}{dt} \exp_\lambda(t) = [\exp_\lambda(t)]^{1-\lambda}$ , so  $\exp_\lambda(\cdot)$  is convex if and only if  $\lambda < 1$ . For this reason, we restricted  $\lambda$  to this range as in [9,28]. Below, we also took  $\log t = -\infty$  whenever  $t \leq 0$ . Note that our notation differs slightly from Tsallis’ indexing of the deformed logarithm and exponential functions; see Section 5.

Next, we construct another pair of inverse functions  $\kappa_\lambda, \gamma_\lambda$  by:

$$\kappa_\lambda = \log \circ \exp_\lambda, \quad \gamma_\lambda = \log_\lambda \circ \exp,$$

where  $\circ$  denotes function composition. Explicitly, they are:

$$\kappa_\lambda(t) = \frac{1}{\lambda} \log(1 + \lambda t), \quad \gamma_\lambda(t) = \frac{1}{\lambda} (e^{\lambda t} - 1). \tag{10}$$

This suite of four functions, namely  $\exp_\lambda, \log_\lambda$  as an inverse pair and  $\kappa_\lambda, \gamma_\lambda$  as another inverse pair, is called  $\lambda$ -deformation and used in the discussions of  $\lambda$ -convexity,  $\lambda$ -conjugation, and  $\lambda$ -duality. Regular exponential and logarithmic functions are recovered when  $\lambda \rightarrow 0$ , whence both  $\kappa_\lambda$  and  $\gamma_\lambda$  reduce to the identity function.

Using these four functions, Wong and Zhang [17] developed the  $\lambda$ -deformation framework to solve the problem of relating the exponential family under subtractive normalization:

$$p^{(\lambda)}(\zeta|\theta) = \exp_\lambda(\theta \cdot F(\zeta) - \phi_\lambda(\theta))$$

to that under divisive normalization:

$$p^{(\lambda)}(\zeta|\vartheta) = \exp_\lambda(\vartheta \cdot F(\zeta)) e^{-\varphi_\lambda(\vartheta)}.$$

There, the same  $\lambda$ -deformed exponential family can be expressed by two parameterizations  $\theta$  and  $\vartheta$  linked through:

$$\theta = \vartheta e^{-\lambda \varphi_\lambda(\vartheta)} \iff \vartheta = \frac{\theta}{1 - \lambda \phi_\lambda(\theta)},$$

while the normalization functions  $\phi_\lambda$  and  $\varphi_\lambda$  (with different domains) are linked through:

$$\phi_\lambda(\theta) = \gamma_{-\lambda}(\varphi_\lambda(\vartheta)) \iff \varphi_\lambda(\vartheta) = \kappa_{-\lambda}(\phi_\lambda(\theta)).$$

The  $\lambda$ -deformation framework led to a unified way of looking at the Tsallis entropy (related to the subtractive denormalization) and Rényi entropy (related to the divisive normalization), as well as generating new insights into the distinction between the exponential and mixture families through the lens of deformation theory. To understand this deformation better, we describe the underlying mathematical framework of  $\lambda$ -deformation.

### 3. Deforming the Legendre Duality: $\lambda$ -Duality

In this section, we describe the  $\lambda$ -duality and its link to the standard Legendre duality. We start by defining the notions of  $\lambda$ -conjugate and  $\lambda$ -convexity/ $\lambda$ -concavity, then draw a parallel to the regular Legendre duality. We proceed to establish a formal correspondence between the  $\lambda$ -duality and classical convex duality, including the associated notions of the  $\lambda$ -gradient,  $\lambda$ -logarithmic divergence, etc. Some of the derivations are illustrative, yet heuristic—a rigorous analysis in the spirit of Rockafellar [35] is yet to be performed in future research.

#### 3.1. Legendre Duality and Bregman Divergence Reviewed

Recall from (4) that the *convex conjugate* of a function  $f$  on  $\mathbb{R}^d$  is defined by:

$$f^*(u) = \sup_x (x \cdot u - f(x)), \quad u \in \mathbb{R}^d. \quad (11)$$

It can be proven that:

- (i)  $f^*$  is convex;
- (ii)  $((f^*)^*)^* = f^*$ ;
- (iii)  $(f^*)^* = f$  if  $f$  is convex and lower semicontinuous.

When  $f$  is further differentiable, then the Legendre transformation:

$$u = Df(x),$$

which can be motivated by the first-order condition in (11), defines a “dual variable”  $u$ , satisfying the Fenchel identity:

$$f(x) + f^*(u) = x \cdot u.$$

We have  $x = Df^*(u)$ , provided the second derivative or  $D^2f$  is positive definite. The function  $f$  also defines a Bregman divergence  $B_f$  given by:

$$B_f(x, x') = f(x) - f(x') - Df(x') \cdot (x - x') \geq 0. \quad (12)$$

The Bregman divergence satisfies the *reference–representation biduality* [24,25] in the sense that:

$$B_f(x, x') = B_{f^*}(u', u)$$

where  $u = Df(x)$ ,  $u' = Df(x')$ . Note that when  $f$  is convex and differentiable, the non-negativity of the Bregman divergence encodes the fact that for any  $x, x'$ :

$$f(x) - f(x') \geq Df(x') \cdot (x - x').$$

#### 3.2. $\lambda$ -Deformation of Legendre Duality

The main idea behind the  $\lambda$ -deformation of the Legendre duality (“ $\lambda$ -duality”) is to replace the term  $x \cdot u$  in (11) by a monotone transformation of  $x \cdot u$ . Given a parameter  $\lambda \in \mathbb{R} \setminus \{0\}$ , later revealed to be the *curvature parameter* of the information geometric characterization, we replace the term  $x \cdot u$  by:

$$\kappa_\lambda(x \cdot u) = \frac{1}{\lambda} \log(1 + \lambda x \cdot u), \quad (13)$$

where  $\kappa_\lambda(t)$  and its inverse  $\gamma_\lambda(t)$  are given by (10). With this in mind we give the following definition.

**Definition 1** ( $\lambda$ -conjugation). Let  $\Omega, \Omega' \subset \mathbb{R}^d$ . Given a function  $f : \Omega \rightarrow \mathbb{R}$ , we define its  $\lambda$ -conjugate  $f^{(\lambda)}$  by:

$$f^{(\lambda)}(u) = \sup_{x \in \Omega} (\kappa_\lambda(x \cdot u) - f(x)), \quad u \in \Omega'. \tag{14}$$

Generalized convex dualities have been heavily used in optimal transport theory [36,37] to characterize the optimal transport plans; in this context, it is called the  $c$ -duality where  $c$  is the cost function of the transport problem. A major novelty of our framework is that the functional form of  $\kappa_\lambda$  (and of  $\gamma_\lambda$ ) leads to explicit formulas, which are not available in the general case. We remark that this is closely related to the fact that the associated information geometry has constant curvature  $\lambda$ .

It turns out that the  $\lambda$ -conjugation defined by (14) corresponds to an appropriately generalized notion of convexity or concavity, through the aid of the function  $\gamma_\lambda$  given by (10). Henceforth, we let  $\lambda \in \mathbb{R} \setminus \{0\}$  be a fixed constant.

**Definition 2** ( $\lambda$ -exponential convexity and concavity). Let  $\Omega \subset \mathbb{R}^d$  be an open convex set. A function  $f : \Omega \rightarrow \mathbb{R}$  is said to be  $\lambda$ -exponentially convex (" $\lambda$ -convex"), or  $\lambda$ -exponentially concave (" $\lambda$ -concave"), if:

$$G_{\lambda,f}(x) = (\gamma_\lambda \circ f)(x) = \frac{1}{\lambda} (e^{\lambda f(x)} - 1)$$

is convex, or concave, on  $\Omega$ . When  $f$  is  $C^2$ , we have equivalently that  $f$  is  $\lambda$ -convex, or  $\lambda$ -concave, if the Hessian of  $G_{\lambda,f} \equiv \gamma_\lambda \circ f$  is positive definite, or negative definite.

Note that the additive term  $-1/\lambda$  in the above definition of  $G_{\lambda,f}(x) = \frac{1}{\lambda}(e^{\lambda f(x)} - 1)$  is not necessary; it is included so that  $\lim_{\lambda \rightarrow 0} G_{\lambda,f}(x) = f(x)$ , meaning that in the limiting case of zero-convexity is just ordinary convexity.

It is easily shown that, for  $\lambda > 0$  a fixed positive number,

- (i)  $f$  is  $\lambda$ -convex if and only if  $-f$  is  $(-\lambda)$ -concave;
- (ii)  $f$  is  $\lambda$ -concave if and only if  $-f$  is  $(-\lambda)$ -convex.

**Proposition 1.** Given any  $f : \Omega \rightarrow \mathbb{R}$ , we define variable  $\tilde{x}$ , which has range  $\tilde{\Omega} \subset \mathbb{R}^d$ , and function  $g : \tilde{\Omega} \rightarrow \mathbb{R}$  by:

$$\tilde{x} = x e^{-\lambda f(x)} = x (1 - \lambda G_{-\lambda,f}(x)), \tag{15}$$

$$g(\tilde{x}) = \frac{1}{-\lambda} (e^{-\lambda f(x)} - 1) = \gamma_{-\lambda}(f(x)) = G_{-\lambda,f}(x). \tag{16}$$

Then, the convex (Legendre) conjugate  $g^*$  of the function  $g$ :

$$g^*(u) = \sup_{\tilde{x} \in \tilde{\Omega}} (\tilde{x} \cdot u - g(\tilde{x}))$$

is related to the  $\lambda$ -conjugate  $f^{(\lambda)}$  of the function  $f$  via:

$$g^*(u) = \frac{1}{\lambda} (e^{\lambda f^{(\lambda)}(u)} - 1) = \gamma_\lambda(f^{(\lambda)}(u)) = G_{\lambda,f^{(\lambda)}}(u). \tag{17}$$

**Proof.** We first prove the following identities:

$$\begin{aligned} (1 + \lambda x \cdot u)e^{-\lambda f(x)} &= e^{-\lambda f(x)} + \lambda e^{-\lambda f(x)} x \cdot u \\ &= (1 - \lambda g(\tilde{x})) + \lambda \tilde{x} \cdot u \\ &= 1 + \lambda (\tilde{x} \cdot u - g(\tilde{x})) \end{aligned}$$

where, going from the first to the second line, we used (15) and the fact:

$$1 - \lambda g(\tilde{x}) = e^{-\lambda f(x)},$$

which is a re-write of the definition of  $g$  given by (16).

With the above identity, we can proceed to prove this proposition. For  $u \in \Omega'$ , we have:

$$\begin{aligned} f^{(\lambda)}(u) &= \sup_{x \in \Omega} \left( \frac{1}{\lambda} \log(1 + \lambda(x \cdot u)) - f(x) \right) \\ &= \sup_{\tilde{x} \in \tilde{\Omega}} \frac{1}{\lambda} \log(1 + \lambda(\tilde{x} \cdot u - g(\tilde{x}))) \\ &= \frac{1}{\lambda} \log \left( 1 + \lambda \sup_{\tilde{x} \in \tilde{\Omega}} (\tilde{x} \cdot u - g(\tilde{x})) \right) \\ &= \frac{1}{\lambda} \log(1 + \lambda g^*(u)) \\ &= \kappa_\lambda(g^*(u)). \end{aligned}$$

Recasting the above relation yields (17).  $\square$

Recall that from convex analysis,  $g^*$  is always a convex function regardless of whether  $g$  is convex (by the property of Legendre conjugation). The expression of  $g^*(u) = \gamma_\lambda(f^{(\lambda)}(u))$  in (17) therefore implies that  $f^{(\lambda)}$  is  $\lambda$ -convex, by the definition of  $\lambda$ -convexity.

**Corollary 1.** For any  $f : \Omega \rightarrow \mathbb{R}$ , its  $\lambda$ -conjugate  $f^{(\lambda)}(u)$  as defined by (14) is a  $\lambda$ -convex function of  $u$  on  $\Omega'$  (note  $\Omega'$  may not necessarily be convex).

**Proof.** We can also give a direct proof (essentially reversing the steps of the proof of Proposition 1).

$$\begin{aligned} g^*(u) &= \sup_{\tilde{x} \in \tilde{\Omega}} (\tilde{x} \cdot u - g(\tilde{x})) \\ &= \sup_{x \in \Omega} \left( e^{-\lambda f(x)}(x \cdot u) - \frac{1}{-\lambda} (e^{-\lambda f(x)} - 1) \right) \\ &= \sup_{x \in \Omega} \frac{1}{\lambda} \left( (1 + \lambda x \cdot u)e^{-\lambda f(x)} - 1 \right) \\ &= \frac{1}{\lambda} \left( \sup_{x \in \Omega} e^{\log(1 + \lambda x \cdot u) - \lambda f(x)} - 1 \right) \\ &= \frac{1}{\lambda} \left( e^{\sup_{x \in \Omega} (\log(1 + \lambda x \cdot u) - \lambda f(x))} - 1 \right) \\ &= \frac{1}{\lambda} \left( e^{\lambda f^{(\lambda)}(u)} - 1 \right) \\ &= \gamma_\lambda(f^{(\lambda)}(u)). \end{aligned}$$

$\square$

Corollary 1 is the extension of the claim that for any  $f$ , the standard Legendre conjugate  $f^*$  as given by (11) is always a convex function. Because of this, we can prove, in analogy to the standard Legendre conjugation  $*$ , the following relations:

- (i)  $((f^{(\lambda)})^{(\lambda)})^{(\lambda)} = f^{(\lambda)}$  for any  $f$ .
- (ii)  $(f^{(\lambda)})^{(\lambda)} = f$  if  $f$  is  $\lambda$ -convex.

### 3.3. Relations Between the $\lambda$ -Duality and Legendre Duality

We proceed to establish a formal relationship between the  $\lambda$ -duality and the ordinary Legendre duality, by relating the  $\lambda$ -conjugation of a  $\lambda$ -convex function  $f$ , denoted by  $f^{(\lambda)}$ , to the standard Legendre conjugation of a function (denoted by  $*$ ).

We continue the analysis performed in Proposition 1. Taking  $\lambda$ -conjugation for a second time,

$$\begin{aligned} (f^{(\lambda)})^{(\lambda)}(x) &= \sup_{u \in \Omega'} \frac{1}{\lambda} (\log(1 + \lambda(x \cdot u))) - f^{(\lambda)}(u) \\ &= \sup_{\tilde{u} \in \Omega'} \frac{1}{\lambda} (\log(1 + \lambda(x \cdot \tilde{u} - \tilde{g}(\tilde{u})))) \\ &= \frac{1}{\lambda} \log \left( 1 + \lambda \sup_{\tilde{u} \in \Omega'} (x \cdot \tilde{u} - \tilde{g}(\tilde{u})) \right) \\ &= \frac{1}{\lambda} \log(1 + \lambda \tilde{g}^*(x)) \\ &= \kappa_\lambda(\tilde{g}^*(x)). \end{aligned}$$

Here, the variable  $\tilde{u}$  is defined by:

$$\tilde{u} = u e^{-\lambda f^{(\lambda)}(u)},$$

and the function  $\tilde{g}$  by:

$$\tilde{g}(\tilde{u}) = \gamma_{-\lambda}(f^{(\lambda)}(u)) = G_{-\lambda, f^{(\lambda)}}(u). \tag{18}$$

In the event when  $f$  is  $\lambda$ -convex, then  $(f^{(\lambda)})^\lambda = f$ . Therefore:

$$\tilde{g}^*(x) = \gamma_\lambda(f(x)) = G_{\lambda, f}(x).$$

Therefore,  $\tilde{g}(\tilde{u}) = (G_{\lambda, f})^*(\tilde{u})$ . That is, the function  $\tilde{g}$  is just the (regular) Legendre conjugation  $*$  of the function  $G_{\lambda, f}(x)$ . In  $\tilde{u}$  parameterization, the  $\tilde{g}$  function has the expression of (18) with  $\tilde{u}$  and  $u$  related by (20). This parallels the fact that  $g(\tilde{x}) = (G_{\lambda, f^{(\lambda)}})^*(\tilde{x})$ , and in  $\tilde{x}$  parameterization, the  $g$  function has the expression of (16) with  $\tilde{x}$  and  $x$  related by (19).

Summarizing the above, we have:

**Theorem 1** (Connecting  $\lambda$ -duality to Legendre duality). *Let  $f$  be a  $\lambda$ -convex function and  $f^{(\lambda)}$  be its  $\lambda$ -conjugate. Denote two functions  $g$  and  $\tilde{g}$ :*

$$\begin{aligned} g(\tilde{x}) &= G_{-\lambda, f}(x) = \gamma_{-\lambda}(f(x)), \\ \tilde{g}(\tilde{u}) &= G_{-\lambda, f^{(\lambda)}}(u) = \gamma_{-\lambda}(f^{(\lambda)}(u)), \end{aligned}$$

where the two variables  $\tilde{x}$  and  $\tilde{u}$  are given by:

$$\tilde{x} = x e^{-\lambda f(x)} \iff x = \frac{\tilde{x}}{1 - \lambda g(\tilde{x})}, \tag{19}$$

$$\tilde{u} = u e^{-\lambda f^{(\lambda)}(u)} \iff u = \frac{\tilde{u}}{1 - \lambda \tilde{g}(\tilde{u})}. \tag{20}$$

Then, the following statements are equivalent:

(i) The  $(x, u)$  variables satisfy the  $\lambda$ -duality of a pair of  $\lambda$ -convex functions  $(f, f^{(\lambda)})$ :

$$\kappa_\lambda(x \cdot u) = f(x) + f^{(\lambda)}(u); \tag{21}$$

(ii) The  $(\tilde{x}, u)$  variables satisfy the Legendre duality of a pair of convex functions  $(g, g^*)$ :

$$\tilde{x} \cdot u = g(\tilde{x}) + g^*(u) \tag{22}$$

with:

$$g^*(u) = G_{\lambda, f^{(\lambda)}}(u) = \gamma_\lambda(f^{(\lambda)}(u));$$

(iii) The  $(x, \tilde{u})$  variables satisfy the Legendre duality of a pair of convex functions  $(\tilde{g}, \tilde{g}^*)$ :

$$x \cdot \tilde{u} = \tilde{g}^*(x) + \tilde{g}(\tilde{u}) \tag{23}$$

with:

$$\tilde{g}^*(x) = G_{\lambda, f}(x) = \gamma_\lambda(f(x)).$$

**Proof.** To prove the equivalence of (21) and (22), we re-write the latter as:

$$e^{-\lambda f(x)} x \cdot u = \gamma_{-\lambda}(f(x)) + \gamma_\lambda(f^{(\lambda)}(u)) = \frac{1}{\lambda} \left( e^{\lambda f^{(\lambda)}(u)} - e^{-\lambda f(x)} \right),$$

where we inserted the following relations:

$$g(\tilde{x}) = \gamma_{-\lambda}(f(x)), \quad g^*(u) = \gamma_\lambda(f^{(\lambda)}(u))$$

and replaced  $\tilde{x}$  by  $x$  using (19). Multiplying  $e^{\lambda f(x)}$  on both sides, we obtain:

$$x \cdot u = \frac{1}{\lambda} \left( e^{\lambda(f^{(\lambda)}(u)+f(x))} - 1 \right) = \gamma_\lambda(f^{(\lambda)}(u) + f(x)). \tag{24}$$

Noting  $(\gamma_\lambda)^{-1} = \kappa_\lambda$  verifies (21).

To prove the equivalence of (21) and (23), we rely on an analogous identity:

$$(1 + \lambda x \cdot u) e^{-\lambda f^{(\lambda)}(u)} = 1 + \lambda (x \cdot \tilde{u} - \tilde{g}(\tilde{u})),$$

where:

$$\tilde{u} = u e^{-\lambda f^{(\lambda)}(u)}, \quad \tilde{g}(\tilde{u}) = \gamma_{-\lambda}(f^{(\lambda)}(u)).$$

We have, after multiplying  $e^{\lambda f^{(\lambda)}(u)}$  on both sides of (24),

$$x \cdot u e^{-\lambda f^{(\lambda)}(u)} = \frac{1}{\lambda} \left( e^{\lambda f(x)} - e^{-\lambda f^{(\lambda)}(u)} \right) = \gamma_\lambda(f(x)) + \gamma_{-\lambda}(f^{(\lambda)}(u)) = \tilde{g}^*(x) + \tilde{g}(\tilde{u}),$$

where the last step used:

$$\tilde{g}^*(x) = \gamma_\lambda((f^{(\lambda)})^{(\lambda)}(x)), \quad \tilde{g}(\tilde{u}) = \gamma_{-\lambda}(f^{(\lambda)}(u)).$$

Noting  $(f^{(\lambda)})^{(\lambda)} = f$  due to  $f$  assumed to be  $\lambda$ -convex, then (23) follows.  $\square$

We see that the functions  $\gamma_\lambda$  and  $\gamma_{-\lambda}$  serve as link functions from the  $(f, f^{(\lambda)})$ -pair of the  $\lambda$ -deformed Legendre conjugation to the  $(g, g^*)$ -pair and the  $(\tilde{g}, \tilde{g}^*)$ -pair of the regular Legendre conjugation.

#### 4. $\lambda$ -Logarithmic Divergence and Its Dualistic Geometry

In this section, we study the  $\lambda$ -deformation of the Bregman (canonical) divergence function and the resulting dualistic geometry (Riemannian metric and dual connections), which correspond to the  $\lambda$ -duality. This involves first establishing the  $\lambda$ -deformation to the gradient operation (so-called  $\lambda$ -gradient), which then leads to the so-called  $\lambda$ -logarithmic divergence function as deformation to the Bregman divergence. Finally, we show that

the resulting Riemannian metric is a conformal Hessian metric, while the resulting dual connections are projectively flat (with constant curvature). The conformal and projective factor is parameterized by  $\lambda$ , which gives the curvature of the constant curvature space.

4.1.  $\lambda$ -Gradient

**Definition 3** ( $\lambda$ -gradient). For  $x \in \Omega$ , define the  $\lambda$ -gradient  $D^{(\lambda)}f$  by:

$$D^{(\lambda)}f(x) = \frac{1}{1 - \lambda Df(x) \cdot x} Df(x). \tag{25}$$

The work of [17] (Theorem 2.2) showed the above formula for deforming the gradient of a function motivated by the  $\lambda$ -duality setting. For mathematical convenience, it is proven under some regularity conditions; a full generalization along the lines of [35] is a natural direction for further research.

**Theorem 2** ( $\lambda$ -gradient for  $\lambda$ -duality). Let  $\lambda \neq 0$ , and let  $f$  be a  $\lambda$ -exponentially convex function that is  $C^2$  on some open convex set  $\Omega \subset \mathbb{R}^d$ , such that (a)  $D^2G_{\lambda,f}$  is strictly positive definite and (b)  $1 - \lambda Df(x) \cdot x > 0$  on  $\Omega$ . Then we have

- (i)  $D^{(\lambda)}f$  is a  $C^1$ -diffeomorphism from  $\Omega$  to its range  $\Omega'$ .
- (ii) Denote  $u = D^{(\lambda)}f(x)$ . We have  $1 + \lambda x \cdot u > 0$ , and the following identity holds:

$$f(x) + f^{(\lambda)}(u) = \frac{1}{\lambda} \log(1 + \lambda x \cdot u) \equiv \kappa_\lambda(x \cdot u).$$

- (iii) Furthermore,  $x = D^{(\lambda)}f^{(\lambda)}(u)$ .

Note that the  $\lambda$ -gradient  $D^{(\lambda)}f$  differs from the regular gradient  $Df$  by a scalar multiplication. The duality between  $x$  and  $u$  under the  $\lambda$ -duality is mediated by a dual variable  $u = D^{(\lambda)}f(x)$ , which plays an important role in what follows.

Let:

- (a)  $u = D^{(\lambda)}f(x)$  denote the  $\lambda$ -conjugate variable corresponding to  $x$  with respect to  $f(x)$ ;
- (b)  $\hat{u} = Dg(\tilde{x})$  be the Legendre conjugate variable corresponding to  $\tilde{x}$  with respect to  $g(\tilde{x})$ ;
- (c)  $x = D^{(\lambda)}f^{(\lambda)}(u)$  denote the  $\lambda$ -conjugate variable corresponding to  $u$  with respect to  $f^{(\lambda)}(u)$ ;
- (d)  $\hat{x} = D\tilde{g}(\tilde{u})$  be the Legendre conjugate variable corresponding to  $\tilde{u}$  with respect to  $\tilde{g}(\tilde{u})$ .

Is there a simple relationship between them? The following proposition says  $u(x) = \hat{u}(\tilde{x})$ , where  $\tilde{x}$  and  $x$  are linked by (19), and  $x(u) = \hat{x}(\tilde{u})$ , where  $\tilde{u}$  and  $u$  are linked by (20).

**Proposition 2.** We have:

$$u = D_x^{(\lambda)}f(x) = D_{\tilde{x}}g(\tilde{x}), \quad x = D_u^{(\lambda)}f^{(\lambda)}(u) = D_{\tilde{u}}\tilde{g}(\tilde{u})$$

Here, we add the subscript to D to emphasize the argument with respect to which the derivative is taken.

**Proof.** We use matrix notations where the gradient is regarded as a column vector. Applying the multivariate chain rule to (17), we have:

$$(D_{\tilde{x}}g(\tilde{x}))^\top = e^{-\lambda f(x)} (D_x f(x))^\top \frac{\partial x}{\partial \tilde{x}}(\tilde{x}),$$

where  $\frac{\partial x}{\partial \tilde{x}}(\tilde{x})$  is the Jacobian of the transformation  $\tilde{x} \mapsto x$  and  $(\cdot)^\top$  denotes the transpose. For two vectors  $x$  and  $y$ , their outer product is denoted by  $xy^\top$ , which is a rank-one square matrix with the  $(i, j)$ -entry  $x^i y^j$ .

From (15), we have:

$$\frac{\partial \tilde{x}}{\partial x}(x) = e^{-\lambda f(x)} \left( I - \lambda x (D_x f(x))^\top \right).$$

Since  $1 - \lambda D_x f(x) \cdot x > 0$  by assumption, we can invert the Jacobian by the Sherman–Morrison formula (see [12], Proposition 4) to obtain:

$$\frac{\partial x}{\partial \tilde{x}}(\tilde{x}) = e^{\lambda f(x)} \left( I + \frac{\lambda x (D_x f(x))^\top}{1 - \lambda D_x f(x) \cdot x} \right).$$

Plugging this into the above, we have:

$$(D_{\tilde{x}} g(\tilde{x}))^\top = \frac{(D_x f(x))^\top}{1 - \lambda D_x f(x) \cdot x}.$$

Using (25) to relate  $D_x^{(\lambda)} f(x)$  to  $D_x f(x)$ , the first relation involving  $D_{\tilde{x}} g(\tilde{x})$  is proven. The proof of the second relation in this proposition is analogous.  $\square$

Just as ordinary convexity leads to the notion of Bregman divergence (12), the notion of  $\lambda$ -exponential convexity leads to a generalization that we call the  $\lambda$ -logarithmic divergence. Henceforth, we let  $f : \Omega \rightarrow \mathbb{R}$  be a  $\lambda$ -exponentially convex function on an open convex domain  $\Omega \subset \mathbb{R}^d$ , and we assumed that the regularity conditions in Theorem 2 hold.

#### 4.2. $\lambda$ -Logarithmic Divergence

By the definition of the  $\lambda$ -convexity, we have that  $G_{\lambda, f}(x) = \gamma_\lambda(f(x))$  is convex on  $\Omega$ . By the ordinary convexity of  $G_{\lambda, f}$ , we have:

$$G_{\lambda, f}(x) - G_{\lambda, f}(x') \geq DG_{\lambda, f}(x') \cdot (x - x'), \quad x, x' \in \Omega.$$

In terms of  $f$ , we have, after some manipulations,

$$\gamma_\lambda(f(x) - f(x')) \geq Df(x') \cdot (x - x').$$

Since  $\gamma_\lambda$  is increasing, we have:

$$f(x) - f(x') \geq (\gamma_\lambda)^{-1}(Df(x') \cdot (x - x')) = \kappa_\lambda(Df(x') \cdot (x - x')).$$

This motivates the following definition.

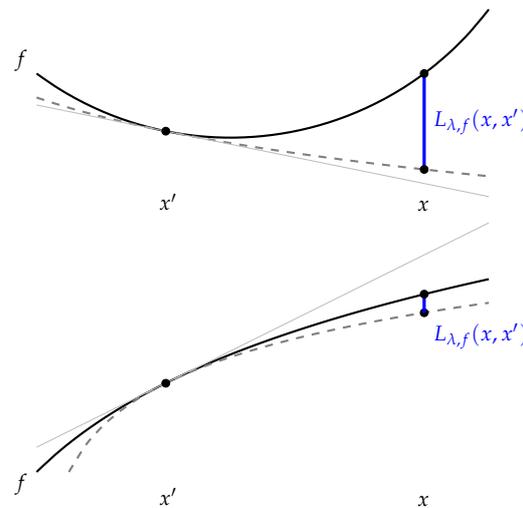
**Definition 4** ( $\lambda$ -logarithmic divergence). We define the  $\lambda$ -logarithmic divergence of  $f$  by:

$$\begin{aligned} L_{\lambda, f}(x, x') &= f(x) - f(x') - \kappa_\lambda(Df(x') \cdot (x - x')) \\ &= f(x) - f(x') - \frac{1}{\lambda} \log(1 + \lambda Df(x') \cdot (x - x')), \quad x, x' \in \Omega. \end{aligned} \tag{26}$$

See Figure 1 for a graphical illustration. We note that the logarithmic correction in (26) corresponds to a logarithmic first-order approximation, based at  $x'$ , which is possible due to the  $\lambda$ -exponential convexity of  $f$ . We also note that when  $\lambda > 0$ , it is possible that  $L_{\lambda, f}(x, x') = \infty$ . Nevertheless,  $L_{\lambda, f}(x, x')$  is finite when  $x$  and  $x'$  are sufficiently close. Formally, letting  $\lambda \rightarrow 0$  in (26) recovers the Bregman divergence.

#### 4.3. $\lambda$ -Logarithmic Divergence in Different Forms

We now prove a lemma about the relationship of the variables  $x, u$  and gradients or  $\lambda$ -gradients of  $f$  or  $f^{(\lambda)}$ . We assumed, for convenience, that  $1 + \lambda x \cdot u > 0$  for all  $x \in \Omega, u \in \Omega'$ .



**Figure 1.** Illustration of the  $\lambda$ -logarithmic divergence. Top:  $\lambda = -1$  and  $f(x) = \sqrt{x(10 - x)}$ . Bottom:  $\lambda = 1$  and  $f(x) = 2 \log x$ . In both cases,  $x' = 4$  and  $x = 8$ , and we plot the function on the interval  $(2, 9)$ . Note that the first-order logarithmic approximation (dashed grey curve) supports the graph of  $f$  from below.

**Lemma 1.** Given  $u = D_x^{(\lambda)} f(x)$  or equivalently  $x = D_u^{(\lambda)} f^{(\lambda)}(u)$ , for arbitrary  $x', u'$  (such that the expressions are well defined), we have the following identities:

$$\kappa_\lambda(u \cdot x') - \kappa_\lambda(u \cdot x) = \kappa_\lambda(u \cdot (x' - x) (\Pi_\lambda)^{-1}), \tag{27}$$

$$\kappa_\lambda(u' \cdot x) - \kappa_\lambda(u \cdot x) = \kappa_\lambda((u' - u) \cdot x (\Pi_\lambda)^{-1}). \tag{28}$$

where  $\Pi_\lambda$  is a multiplicative factor (function of  $x$  or  $u$ ) given by:

$$\Pi_\lambda \equiv 1 + \lambda x \cdot u = \frac{1}{1 - \lambda Df(x) \cdot x} = \frac{1}{1 - \lambda Df^{(\lambda)}(u) \cdot u}.$$

**Proof.** Since  $u = D_x^{(\lambda)} f(x)$ , substituting (25), we have:

$$u \cdot x = \frac{Df(x) \cdot x}{1 - \lambda Df(x) \cdot x}$$

and:

$$1 + \lambda u \cdot x = \frac{1}{1 - \lambda Df(x) \cdot x}$$

so:

$$1 + \lambda u \cdot x' = \frac{1 + \lambda Df(x) \cdot (x' - x)}{1 - \lambda Df(x) \cdot x} = (1 + \lambda u \cdot x)(1 + \lambda Df(x) \cdot (x' - x)).$$

Taking the logarithm and rearranging, we obtain (27).

On the other hand, because:

$$x = D_u^{(\lambda)} f^{(\lambda)}(u) = \frac{Df^{(\lambda)}(u)}{1 - \lambda Df^{(\lambda)}(u) \cdot u},$$

we also have:

$$1 + \lambda x \cdot u = \frac{1}{1 - \lambda Df^{(\lambda)}(u) \cdot u}.$$

The proof of (28) is similar.  $\square$

In this above lemma,  $x'$  and  $u'$  are arbitrary; it is interesting that a modified form of “linearity” holds even though  $\kappa_\lambda$  is itself nonlinear. As a consequence, we have an alternative expression for  $L_{\lambda, f}(x, x')$ .

**Proposition 3.**  $L_{\lambda,f}(x, x')$  defined by (26) can also be written as:

$$L_{\lambda,f}(x, x') = f(x) - f(x') - \kappa_{\lambda}(x \cdot u') + \kappa_{\lambda}(x' \cdot u').$$

where  $u' = D^{(\lambda)}f(x')$ .

Of course, we may express the  $\lambda$ -logarithmic divergence using the conjugate variables  $u, u'$  as well. Indeed, we have the analogous reference–representation biduality (see [24,25]) that is characteristic of Bregman divergence and canonical divergence for dually flat spaces, that is (8). See [38] for the reference–representation biduality of a general  $c$ -divergence (which includes both the Bregman and logarithmic divergences) based on optimal transport.

**Theorem 3.** The  $\lambda$ -logarithmic divergence satisfies the reference–representation biduality, namely:

$$L_{\lambda,f^{(\lambda)}}(u', u) = L_{\lambda,f}(x, x'),$$

where  $u = D^{(\lambda)}f(x)$  and  $u' = D^{(\lambda)}f(x')$ . Moreover, define the  $\lambda$ -deformed canonical divergence  $A_{\lambda,f}$  by:

$$A_{\lambda,f}(x, u') = f(x) + f^{(\lambda)}(u') - \frac{1}{\lambda} \log(1 + \lambda x \cdot u') = A_{\lambda,f^{(\lambda)}}(u', x).$$

We have:

$$L_{\lambda,f}(x, x') = A_{\lambda,f}(x, u') = A_{\lambda,f^{(\lambda)}}(u', x) = L_{\lambda,f^{(\lambda)}}(u', u).$$

Proposition 3 also allows us to derive our next theorem (Theorem 4) linking  $\lambda$ -logarithmic divergence and Bregman divergence (also see [19] for a discussion of conformal divergence in the affine immersion setting).

**Theorem 4.** The canonical forms of the  $\lambda$ -logarithmic divergence  $A_{\lambda,f}$  and  $A_{\lambda,f^{(\lambda)}}$  are related to the canonical forms of the Bregman divergence  $A_{g^*}$  and  $A_{\tilde{g}}$  via a conformal transformation and the non-linear link function  $\kappa_{-\lambda}$ :

$$\begin{aligned} A_{\lambda,f^{(\lambda)}}(u', x) &= \kappa_{-\lambda}\left(e^{-\lambda f^{(\lambda)}(u')} A_{g^*}(u', \tilde{x})\right), \\ &= A_{\lambda,f}(x, u') = \kappa_{-\lambda}\left(e^{-\lambda f(x)} A_{\tilde{g}^*}(x, \tilde{u}')\right). \end{aligned}$$

**Proof.**

$$\begin{aligned} A_{\lambda,f^{(\lambda)}}(u', x) &= f^{(\lambda)}(u') + f(x) - \frac{1}{\lambda} \log(1 + \lambda u' \cdot x) \\ &= f^{(\lambda)}(u') - \frac{1}{\lambda} \log\left((1 + \lambda u' \cdot x)e^{-\lambda f(x)}\right) \\ &= \frac{1}{\lambda} \log(1 + \lambda g^*(u')) - \frac{1}{\lambda} \log\left(e^{-\lambda f(x)} + \lambda u' \cdot (xe^{-\lambda f(x)})\right) \\ &= \frac{1}{\lambda} \log(1 + \lambda g^*(u')) - \frac{1}{\lambda} \log(1 - \lambda g(\tilde{x}) + \lambda u' \cdot \tilde{x}) \\ &= -\frac{1}{\lambda} \log\left(\frac{1 + \lambda(u' \cdot \tilde{x} - g(\tilde{x}))}{1 + \lambda g^*(u')}\right) \\ &= -\frac{1}{\lambda} \log\left(1 + \lambda \frac{u' \cdot \tilde{x} - g(\tilde{x}) - g^*(u')}{1 + \lambda g^*(u')}\right) \\ &= -\frac{1}{\lambda} \log\left(1 - \lambda \frac{A_{g^*}(u', \tilde{x})}{1 + \lambda g^*(u')}\right) \\ &= -\frac{1}{\lambda} \log\left(1 - \lambda \left(e^{-\lambda f^{(\lambda)}(u')} A_{g^*}(u', \tilde{x})\right)\right) \\ &= \kappa_{-\lambda}\left(e^{-\lambda f^{(\lambda)}(u')} A_{g^*}(u', \tilde{x})\right). \end{aligned}$$

The proof of the second line of Theorem 4 is similar. We have  $A_{\lambda, f^{(\lambda)}}(u', x) = A_{\lambda, f}(x, u')$  from Theorem 3.  $\square$

4.4. Dualistic Geometry of  $\lambda$ -Logarithmic Divergence

Regard  $x \in \Omega$  as the primal (global) coordinate system of a manifold  $\mathcal{M}$ . As described in Section 2.1, we may use the  $\lambda$ -logarithmic divergence  $L_{\lambda, f}$  of  $f$  to construct a dualistic structure  $(\mathcal{M}, \mathbf{g}, \nabla, \nabla^*)$ . In this subsection, we provide explicit expressions of the corresponding coefficients and state some key geometric consequences.

We begin with the Riemannian metric.

**Theorem 5.** *The Riemannian metric  $\mathbf{g}$  induced from  $L_{\lambda, f}(x, x')$  is given in primal coordinate  $x$  by:*

$$\mathbf{g}(x) = D^2f(x) + \lambda(Df(x))(Df(x))^\top = e^{-\lambda f(x)} D^2G_{\lambda, f}(x). \tag{29}$$

**Proof.** According to (2), we perform direct differentiation of (26):

$$g_{ij}(x) = \left. \frac{\partial^2}{\partial x^i \partial x^j} L_{\lambda, f}(x, x') \right|_{x=x'}$$

and obtain the expression of (29).  $\square$

By symmetry, under the dual coordinate system  $u = D^{(\lambda)}f(x)$ , we have:

$$\mathbf{g}(u) = D^2f^{(\lambda)}(u) + \lambda(Df^{(\lambda)}(u))(Df^{(\lambda)}(u))^\top.$$

From the first equality in (29), we see that  $\mathbf{g}$  is a *rank-one correction* of the Hessian matrix  $D^2f(x)$ . From the second equality, we see that  $\mathbf{g}$  is in fact a *conformal Hessian metric*, i.e., it has the form  $\mathbf{g}(x) = e^{-\lambda f(x)} \mathbf{g}_0(x)$ , where  $\mathbf{g}_0(x) = D^2G_{\lambda, f}(x)$  is the Hessian metric induced by the convex function  $G_{\lambda, f}(x) = \frac{1}{\lambda}(e^{\lambda f(x)} - 1)$ . This conclusion is entirely anticipated from Theorem 4.

To compute the Christoffel symbols of the primal and dual connections, we need an expression of the inverse of the Riemannian metric  $\mathbf{g}(x)$  as a matrix. This is provided by the following proposition.

**Proposition 4.** *The metric  $\mathbf{g}$  can be expressed as:*

$$\mathbf{g}(x) = \frac{1}{\Pi_\lambda(x)} \left( I_d - \frac{\lambda}{\Pi_\lambda(x)} u x^\top \right) \frac{\partial u}{\partial x}(x), \tag{30}$$

where  $\frac{\partial u}{\partial x}$  is the Jacobian matrix of the coordinate transformation  $x \mapsto u$  and  $I_d$  is the  $d \times d$  identity matrix with Kronecker  $\delta_{ij}$  as its entries. Here:

$$\Pi_\lambda(x) = 1 + \lambda x \cdot u = \frac{1}{1 - \lambda Df(x) \cdot x}$$

and  $\Pi_\lambda(x) > 0$  for  $x \in \Omega$  and  $u = D^{(\lambda)}f(x)$ , due to Part (ii) of Theorem 2.

Moreover, the inverse of  $\mathbf{g}(x)$  can be expressed as:

$$(\mathbf{g}(x))^{-1} = \Pi_\lambda(x) \frac{\partial x}{\partial u}(u) (I_d + \lambda u x^\top). \tag{31}$$

**Proof.** Using the  $\lambda$ -logarithmic divergence represented as the generalized canonical divergence  $A_{\lambda, f}$  (26), we apply (2) to obtain:

$$\begin{aligned}
 g_{ij}(x) &= -\frac{\partial^2}{\partial x^i \partial x'^j} L_{\lambda, f}(x, x') \Big|_{x=x'} \\
 &= -\frac{\partial^2}{\partial x^i \partial x'^j} \left\{ f(x) + f^{(\lambda)}(u') - \frac{1}{\lambda} \log(1 + \lambda x \cdot u') \right\} \Big|_{x=x'} \\
 &= \frac{1}{\Pi_\lambda(x)} \left\{ \frac{\partial u^i}{\partial x^j} - \frac{\lambda}{\Pi_\lambda(x)} u^i \sum_{k=1}^d x^k \frac{\partial u^k}{\partial x^j} \right\}.
 \end{aligned}$$

Expressing the above expression using matrix notations gives (30). Formula (31) follows by inverting (30) using the Sherman–Morrison formula.  $\square$

Under the dualistic structure induced by a  $\lambda$ -logarithmic divergence, the primal and dual coordinate vector fields are no longer biorthogonal in the sense of (7). Nevertheless, we have the following generalization. Again, we write  $\Pi_\lambda(x) = 1 + \lambda x \cdot u$ .

**Corollary 2.** *The inner product of the coordinate vector fields  $\frac{\partial}{\partial x^i}$  and  $\frac{\partial}{\partial u^j}$  is given by a  $\lambda$ -deformed “biorthogonality” relation:*

$$\mathfrak{g} \left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial u^j} \right) = \frac{1}{\Pi_\lambda(x)} \delta_{ij} - \frac{\lambda}{\Pi_\lambda(x)^2} x^j u^i.$$

**Proof.** Write  $\frac{\partial}{\partial u^j} = \sum_{m=1}^d \frac{\partial x^m}{\partial u^j} \frac{\partial}{\partial x^m}$ . Then:

$$\mathfrak{g} \left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial u^j} \right) = \sum_{m=1}^d \frac{\partial x^m}{\partial u^j} \mathfrak{g} \left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^m} \right).$$

Simplifying the expression using (30) gives the result. For details, see ([12], Proposition 8).  $\square$

**Theorem 6.** *The Christoffel symbols of the primal connection  $\nabla$  are given by:*

$$\Gamma_{ij,k}(x) = -\frac{\lambda}{\Pi_\lambda(x)^2} \left( u^j \frac{\partial u^i}{\partial x^k} + u^i \frac{\partial u^j}{\partial x^k} \right) + \frac{2\lambda^2}{\Pi_\lambda(x)^3} \sum_{\ell=1}^d u^i u^j x^\ell \frac{\partial u^\ell}{\partial x^k},$$

where  $\Pi_\lambda(x) = 1 + \lambda x \cdot u$  as in Proposition 4. Furthermore, let  $\Gamma_{ij}^k = \sum_{\ell=1}^d \Gamma_{ij,\ell} g^{\ell k}$  be the Christoffel symbol of the second kind, then:

$$\Gamma_{ij}^k(x) = \frac{-\lambda}{\Pi_\lambda(x)} \left( u^i \delta_j^k + u^j \delta_i^k \right) = -\lambda \left( \frac{\partial f}{\partial x^i}(x) \delta_j^k + \frac{\partial f}{\partial x^j}(x) \delta_i^k \right), \tag{32}$$

where  $\delta$  is the Kronecker delta.

Similarly, under the dual coordinate system  $u$ , the Christoffel symbol (of the second kind) of the dual connection  $\nabla^*$  is given by:

$$\Gamma_{ij}^{*k}(u) = -\lambda \left( \frac{\partial f^{(\lambda)}}{\partial u^i}(u) \delta_j^k + \frac{\partial f^{(\lambda)}}{\partial u^j}(u) \delta_i^k \right). \tag{33}$$

**Proof.** This is a straightforward computation using (3) and Proposition 4. The details, which are a minor modification of the proof of ([12], Proposition 5), are omitted.  $\square$

Although the curvatures of  $\nabla$  and  $\nabla^*$  are nonzero, it can be shown that  $\nabla$  and  $\nabla^*$  are both projectively flat, i.e., each of them is projectively equivalent to a flat connection. Specifically, any  $\nabla$ -geodesic (resp.  $\nabla^*$ -geodesic) is a time-reparameterized straight line under the  $x$  (resp.  $u$ ) coordinate system.

**Theorem 7.** *The sectional curvatures of  $\nabla$  and  $\nabla^*$  with respect to  $g$  are both equal to  $\lambda$ .*

**Proof.** See ([12], Theorem 15).  $\square$

Using the dual projective flatness and Corollary 2, Reference ([12], Theorem 16) showed that the  $\lambda$ -logarithmic divergence satisfies a generalized Pythagorean theorem, which generalizes the property of Bregman divergence outlined in Section 2.2.

**Theorem 8** (Generalized Pythagorean theorem). *Let  $P, Q, R \in \mathcal{M}$ . Then:*

$$L_{\lambda,f}(x_Q, x_P) + L_{\lambda,f}(x_R, x_Q) = L_{\lambda,f}(x_R, x_P)$$

*if and only if the  $\nabla$ -geodesic between  $Q$  and  $R$  and the  $\nabla^*$ -geodesic between  $Q$  and  $P$  meet  $g$ -orthogonally at  $Q$ .*

To summarize, the dually flat geometry becomes a dually projectively flat geometry with constant sectional curvature  $\lambda$ , and the Hessian metric becomes a conformal Hessian metric. Nevertheless, the primal and dual geodesics are still straight lines (up to time reparametrizations), and the generalized Pythagorean theorem holds.

We say that the above  $\lambda$ -deformation framework is “canonical” because the statistical manifold  $(\mathcal{M}, g, \nabla, \nabla^*)$ , with a conformal Hessian metric  $g_{ij}$  given by (29) and a pair of dual projectively flat affine connections  $\Gamma_{ij}^k, \Gamma_{ij}^{*k}$  given by (32) and (33), is the only statistical structure with constant curvature ([12], Theorem 15). Moreover, given such a statistical manifold, one can construct locally a  $\lambda$ -logarithmic divergence, which induces the given geometry.

### 5. Linking $\lambda$ -Deformation to Rényi Entropy and Divergence

#### 5.1. Relation Between Tsallis’ and Rényi’s Deformation Expressions

Recall that Tsallis [39], in the context of statistical physics, introduced the generalized entropy:

$$H_{\lambda}^{\text{Tsallis}}[p] = \int p \log_{\lambda} \left( \frac{1}{p} \right) d\mu = \frac{1}{\lambda} \left( \int (p(\zeta))^{\lambda} - 1 \right) d\mu;$$

note that we use  $\lambda$  here in place of  $q = 1 - \lambda$  as in [40].

Tsallis entropy is related to Rényi entropy [41], defined as:

$$H_{\lambda}^{\text{Rényi}}[p] := \frac{1}{\lambda} \log \left( \int p^{1-\lambda}(\zeta) d\mu \right),$$

through a monotonic transformation:

$$H_{\lambda}^{\text{Tsallis}}[p] = \frac{1}{\lambda} \left( e^{\lambda H_{\lambda}^{\text{Rényi}}[p]} - 1 \right).$$

In our current notation,

$$H_{\lambda}^{\text{Tsallis}}[p] = \gamma_{\lambda} \left( H_{\lambda}^{\text{Rényi}}[p] \right) \iff H_{\lambda}^{\text{Rényi}}[p] = \kappa_{\lambda} \left( H_{\lambda}^{\text{Tsallis}}[p] \right).$$

Rényi divergence (with Rényi index  $1 - \lambda$ ) is defined by:

$$H_{\lambda}^{\text{Rényi}}[p||p'] = \frac{-1}{\lambda} \log \int (p(\zeta))^{1-\lambda} (p'(\zeta))^{\lambda} d\mu.$$

Rényi divergence is additive: given two product measures  $p_1 \otimes p_2$  and  $p'_1 \otimes p'_2$ , we have:

$$H_{\lambda}^{\text{Rényi}}[p_1 \otimes p_2 || p'_1 \otimes p'_2] = H_{\lambda}^{\text{Rényi}}[p_1 || p'_1] + H_{\lambda}^{\text{Rényi}}[p_2 || p'_2].$$

Because Tsallis entropy is not additive, this has been used as an argument for favoring Rényi entropy as a physical concept over Tsallis entropy; see [28] (Section 9.3) and [42].

### 5.2. $\lambda$ -Exponential Family

Under the  $\lambda$ -deformation, there is an intrinsic link between the subtractive and divisive normalizations of the  $\lambda$ -deformed exponential family. Starting with the observation:

$$e^{\kappa_\lambda(t)} = (1 + \lambda t)^{1/\lambda} = \exp_\lambda(t),$$

we investigate the identity:

$$(1 + \lambda \vartheta \cdot F(\zeta))^{1/\lambda} e^{-\varphi_\lambda(\vartheta)} = (1 + \lambda(\vartheta \cdot F(\zeta) - \phi_\lambda(\vartheta)))^{1/\lambda}.$$

Taking the  $\lambda$ -th power and equating both sides, we obtain the conditions for the above identity to hold:

$$\begin{aligned} \theta &= \vartheta e^{-\lambda \varphi_\lambda(\vartheta)} \iff \vartheta = \frac{\theta}{1 - \lambda \phi_\lambda(\theta)}, \\ \phi_\lambda(\theta) &= \frac{1}{-\lambda}(e^{-\lambda \varphi_\lambda(\theta)} - 1) \iff \varphi_\lambda(\vartheta) = -\frac{1}{\lambda} \log(1 - \lambda \phi_\lambda(\theta)). \end{aligned}$$

This fact led us to define a  $\lambda$ -exponential family that can be normalized both subtractively and divisively: the former denoted by  $p(\zeta|\theta)$  and the latter denoted by  $p(\zeta|\vartheta)$ .

**Proposition 5** (Reparameterization equivalence). *Let  $\lambda \neq 0$ . With respect to a given reference measure  $\mu$  and a fixed vector of random functions  $F(\zeta) = (F_1(\zeta), \dots, F_d(\zeta))$ , the  $\lambda$ -exponential family is given by  $p^{(\lambda)}(\zeta|\theta)$  under subtractive normalization and by  $p^{(\lambda)}(\zeta|\vartheta)$  under divisive normalization; they are reparameterizations of each other:*

$$p^{(\lambda)}(\zeta|\theta) = \exp_\lambda(\theta \cdot F(\zeta) - \phi_\lambda(\theta)) = \exp_\lambda(\vartheta \cdot F(\zeta)) e^{-\varphi_\lambda(\vartheta)} = p^{(\lambda)}(\zeta|\vartheta). \tag{34}$$

Here, the function  $\phi_\lambda(\theta)$  is called *subtractive*  $\lambda$ -potential and used for *subtractive* normalization, while  $\varphi_\lambda(\vartheta)$  is called *divisive*  $\lambda$ -potential and used for *divisive* normalization. Note that  $\phi_\lambda$  and  $\varphi_\lambda$  may not have same domains. They satisfy:

$$\phi_\lambda(\theta) = \gamma_{-\lambda}(\varphi_\lambda(\vartheta)) \iff \varphi_\lambda(\vartheta) = \kappa_{-\lambda}(\phi_\lambda(\theta)),$$

where:

$$\kappa_{-\lambda}(t) = -\frac{1}{\lambda} \log(1 - \lambda t), \quad \gamma_{-\lambda}(t) = \frac{1}{\lambda} (1 - e^{-\lambda t}).$$

Note again that we use  $\vartheta$  for the divisive normalization setting and distinguish it from  $\theta$  for the subtractive normalization setting. For later convenience, we also note:

$$e^{-\lambda \varphi_\lambda(\vartheta)} = 1 - \lambda \phi_\lambda(\theta).$$

#### 5.2.1. Under Subtractive Normalization

The deformed exponential family takes the form:

$$p(\zeta|\theta) = \exp_\lambda(\theta \cdot F(\zeta) - \phi_\lambda(\theta)) \tag{35}$$

where  $\theta \cdot F(\zeta) = \sum_{i=1}^d \theta^i F_i(\zeta)$ , and the subtractive  $\lambda$ -potential  $\phi_\lambda(\theta)$  is specified by the normalization:

$$1 = \int p(\zeta|\theta) d\mu = \int \exp_\lambda(\theta \cdot F(\zeta) - \phi_\lambda(\theta)) d\mu.$$

This leads to:

$$\frac{\partial \phi_\lambda}{\partial \theta^i} = \int \tilde{p}(\zeta|\theta) F_i(\zeta) d\mu$$

with the escort transformation given by:

$$\tilde{p}(\zeta|\theta) = \frac{(p(\zeta|\theta))^{1-\lambda}}{\int (p(\zeta|\theta))^{1-\lambda} d\mu}. \tag{36}$$

Clearly, when  $\lambda \rightarrow 0$ , we recover the regular exponential family (9). It was Tsallis who introduced the  $q$ -exponential family, where  $q = 1 - \lambda$ .

### 5.2.2. Under Divisive Normalization

To deform the exponential family through divisive normalization, we use a smooth monotone function  $\kappa_\lambda(\cdot)$  and define a parametric probability family, which takes the form:

$$\log p(\zeta|\vartheta) = \kappa_\lambda(\vartheta \cdot F(\zeta)) - \varphi_\lambda(\vartheta).$$

Note that we use the symbol  $\vartheta$  to distinguish it from the parameter  $\theta$  in the subtractive case. Here:

$$\varphi_\lambda(\vartheta) = \log \int e^{\kappa_\lambda(\vartheta \cdot F(\zeta))} d\mu$$

is the divisive normalization function, and it was assumed that:

$$\int e^{\kappa_\lambda(\vartheta \cdot F(\zeta))} d\mu < \infty$$

in the domain of  $\vartheta$  (the natural parameter set). It is possible that the support of the density depends on the parameter  $\vartheta$ , as in the case of the  $q$ -exponential family; see [17]. To avoid technicalities, we assumed that the support of  $p(\zeta|\vartheta)$  is independent of  $\vartheta$ .

Writing out  $\kappa_\lambda(\cdot)$ , the resulting family is:

$$p(\zeta|\vartheta) = (1 + \lambda \vartheta \cdot F(\zeta))^{1/\lambda} e^{-\varphi_\lambda(\vartheta)}, \tag{37}$$

where the divisive  $\lambda$ -potential  $\varphi_\lambda(\vartheta)$  is given by:

$$\varphi_\lambda(\vartheta) = \log \int (1 + \lambda \vartheta \cdot F(\zeta))^{1/\lambda} d\mu \tag{38}$$

is finite on the parameter set. This family unifies the  $\mathcal{F}^{(\pm\lambda)}$ -families introduced in [12].

### 5.3. $\lambda$ -Mixture Family

We next define a mixture-type family dual to the  $\lambda$ -exponential family, in an analogous way that an exponential family is dual to the mixture family. The form of the family is justified by its compatibility with the  $\lambda$ -duality.

**Definition 5** ( $\lambda$ -mixture family). *Let  $\lambda \neq 0, 1$  be given. The  $\lambda$ -mixture family with respect to a fixed set of densities  $P_0(\zeta), P_1(\zeta), \dots, P_d(\zeta)$  is defined by:*

$$p^{(\lambda)}(\zeta|\eta) = \frac{1}{Z_\lambda(\eta)} \left( \sum_{i=0}^d \eta_i \tilde{P}_i(\zeta) \right)^{1/(1-\lambda)}, \tag{39}$$

where  $\eta = (\eta_1, \dots, \eta_d)$  is the mixture parameter satisfying  $0 \leq \eta_i \leq 1$  and  $\eta_0 = 1 - \sum_{i=1}^d \eta_i > 0$ . Here,  $\tilde{P}_i, i = 0, 1, \dots, d$  denotes the escort transformation, as given by (36), of the given  $P_i$ 's:

$$\tilde{P}_i(\zeta) = \frac{(P_i(\zeta))^{1-\lambda}}{\int (P_i(\zeta))^{1-\lambda} d\mu},$$

where the denominator is assumed to exist, and  $Z_\lambda(\eta)$  represents the integral:

$$Z_\lambda(\eta) = \int \left( \sum_{i=0}^d \eta_i \tilde{P}_i(\zeta) \right)^{1/(1-\lambda)} d\mu,$$

which is assumed to converge for all  $\eta$  and to be differentiable under the integral sign.

Denote:

$$C_i = \int (P_i(\zeta))^{1-\lambda} d\mu$$

and:

$$\tilde{\eta}_i = \frac{1}{h_\lambda(\eta)} \frac{\eta_i}{C_i}$$

where:

$$h_\lambda(\eta) = \sum_{i=0}^d \frac{\eta_i}{C_i}.$$

Then,  $0 \leq \tilde{\eta}_i \leq 1$  and  $\sum_{i=0}^d \tilde{\eta}_i = 1$ . We can express  $p^{(\lambda)}$  now in  $\tilde{\eta}$ :

$$\begin{aligned} p^{(\lambda)} &= \frac{1}{Z_\lambda(\eta)} \left( \sum_{i=0}^d \eta_i \tilde{P}_i(\zeta) \right)^{1/(1-\lambda)} = \frac{h_\lambda(\eta)}{Z_\lambda(\eta)} \left( \sum_{i=0}^d \tilde{\eta}_i (P_i(\zeta))^{1-\lambda} \right)^{1/(1-\lambda)} \\ &= e^{\log h_\lambda(\eta) - \log Z_\lambda(\eta)} \left( \left( 1 - \sum_{i=1}^d \tilde{\eta}_i \right) (P_0(\zeta))^{1-\lambda} + \sum_{i=1}^d \tilde{\eta}_i (P_i(\zeta))^{1-\lambda} \right)^{1/(1-\lambda)} \\ &= e^{\log h_\lambda(\eta) - \log Z_\lambda(\eta)} \left( 1 + \sum_{i=1}^d \tilde{\eta}_i \frac{(P_i(\zeta))^{1-\lambda} - (P_0(\zeta))^{1-\lambda}}{(P_0(\zeta))^{1-\lambda}} \right)^{1/(1-\lambda)} P_0(\zeta). \end{aligned}$$

Setting:

$$F_i(\zeta) = \frac{1}{1-\lambda} \left( \left( \frac{P_i(\zeta)}{P_0(\zeta)} \right)^{1-\lambda} - 1 \right),$$

with  $d\nu = P_0(\zeta)d\mu$ , the density of the  $\lambda$ -mixture family  $p^{(\lambda)}$  with respect to the new measure  $\nu$  now has the form:

$$p^{(\lambda)}(\zeta|\tilde{\eta}) = (1 + (1-\lambda)\tilde{\eta} \cdot F(\zeta))^{1/(1-\lambda)} e^{-\psi_{1-\lambda}(\tilde{\eta})},$$

where  $\psi_{1-\lambda}(\tilde{\eta}) = \log Z_\lambda(\eta) - \log h_\lambda(\eta)$ . Thus, we showed the following:

**Proposition 6** (Relation between  $\lambda$ -exponential and  $\lambda$ -mixture families). *Suppose  $\lambda \neq 0, 1$ . A  $\lambda$ -mixture family with pure densities:*

$$P(\zeta) = \{P_0(\zeta), P_1(\zeta), \dots, P_d(\zeta)\}$$

*becomes a  $\lambda$ -exponential family with the vector of random functions:*

$$F(\zeta) = \{F_1(\zeta), \dots, F_d(\zeta)\}$$

*after a transformation of the dominating measure  $d\mu \rightarrow d\nu = P_0(\zeta)d\mu$  and the random variables  $P(\zeta) \rightarrow F(\zeta)$ :*

$$F_i(\zeta) = \frac{1}{1-\lambda} \left( \left( \frac{P_i(\zeta)}{P_0(\zeta)} \right)^{1-\lambda} - 1 \right) = \log_{1-\lambda} \left( \frac{P_i(\zeta)}{P_0(\zeta)} \right),$$

*and a reparameterization  $\eta \rightarrow \tilde{\eta}$ :*

$$\tilde{\eta}_i = \frac{1}{h_\lambda C_i} \eta_i \iff \eta_i = \tilde{\eta}_i (h_\lambda C_i)$$

*with:*

$$h_\lambda = \sum_{i=0}^d \frac{\eta_i}{C_i} = \sum_{i=0}^d \tilde{\eta}_i C_i.$$

### 5.4. Potential Functions as Rényi Entropies

We now show that our  $\lambda$ -duality framework is naturally compatible with the  $\lambda$ -exponential and  $\lambda$ -mixture families, with Rényi entropy and Rényi divergence replacing Shannon entropy and Kullback–Leibler divergence. In what follows, we assume  $\lambda < 1$ .

**Proposition 7** (For the  $\lambda$ -exponential family). *With respect to the  $\lambda$ -exponential family defined by (37) with divisive potential function  $\varphi_\lambda$  given by (38), we have:*

- (i)  $\varphi_\lambda(\vartheta)$  is  $\lambda$ -convex. Moreover,  $1 - \lambda D\varphi_\lambda(\vartheta) \cdot \vartheta > 0$ .
- (ii) The  $\lambda$ -conjugate variable  $\eta = D^{(\lambda)}\varphi_\lambda(\vartheta) = D\varphi_\lambda(\vartheta)$  is the the escort expectation:

$$\eta = \frac{\int (p(\zeta|\vartheta))^{1-\lambda} F(\zeta) d\mu}{\int (p(\zeta|\vartheta))^{1-\lambda} d\mu} = \int \tilde{p}(\zeta|\vartheta) F(\zeta) d\mu.$$

- (iii) The  $\lambda$ -conjugate function  $\psi_\lambda(\eta)$  with respect to  $\varphi_\lambda(\vartheta)$  is given by:

$$\psi_\lambda(\eta) = -H_\lambda^{\text{Rényi}}[p(\cdot|\vartheta)].$$

- (iv) The  $\lambda$ -logarithmic divergence is the Rényi divergence:

$$L_{\lambda,\varphi_\lambda}(\vartheta, \vartheta') = H_\lambda^{\text{Rényi}}[p(\cdot|\vartheta')||p(\cdot|\vartheta)].$$

**Proposition 8** (For the  $\lambda$ -mixture family). *With respect to the  $\lambda$ -mixture family given by (39) with its potential function  $\psi_\lambda(\eta)$  given by:*

$$\psi_\lambda(\eta) = \frac{1-\lambda}{\lambda} \log \int \left( \sum_{i=0}^d \eta_i \tilde{P}_i \right)^{1/(1-\lambda)} d\mu = \frac{1-\lambda}{\lambda} \log Z_\lambda(\eta),$$

we have:

- (i) The potential function  $\psi_\lambda(\eta)$  is a  $\lambda$ -convex function of  $\eta$ .
- (ii) The potential function  $\psi_\lambda(\eta)$  is given by:

$$\psi_\lambda(\eta) = -H_\lambda^{\text{Rényi}}[p(\cdot|\eta)].$$

- (iii) The  $\lambda$ -logarithmic divergence is the Rényi divergence:

$$L_{\lambda,\psi_\lambda}(\eta, \eta') = H_\lambda^{\text{Rényi}}[p(\cdot|\eta')||p(\cdot|\eta)].$$

The proofs of the above Proposition 7 (about the  $\lambda$ -exponential family) and Proposition 8 (about the  $\lambda$ -mixture family) can be found in [17].

## 6. Summary and Conclusions

Our paper summarizes a canonical approach to deforming exponential and mixture families and the associated dually flat Hessian geometry. The  $\lambda$ -exponential family we introduced has two parameterizations (35) and (37):

$$p^{(\lambda)}(\zeta|\cdot) = \exp_\lambda(\vartheta \cdot F(\zeta))e^{-\varphi_\lambda(\vartheta)} = \exp_\lambda(\vartheta \cdot F(\zeta) - \varphi_\lambda(\vartheta)).$$

The two expressions reflect subtractive and divisive normalizations—a typical example of the former is the  $q$ -exponential family with associated Tsallis entropy, whereas an example of the latter is the  $\mathcal{F}^{(\pm\alpha)}$ -family and the associated Rényi entropy. These two versions of deformation to an exponential family are two faces of the same coin; furthermore, the  $\lambda$ -exponential family is also linked to the  $\lambda$ -mixture family, when  $\lambda \neq 0, 1$ , via a reparameterization of the random functions  $F(\zeta)$  above.

The coincidence of these two parameterizations of the deformed family is associated with the  $\lambda$ -duality, which is the main focus of our exposition. The  $\lambda$ -duality is a “deformation” (see Table 1) of the usual Legendre duality reviewed in Section 3.1. In a nutshell, instead of convex functions, we worked with  $\lambda$ -convex functions  $f$  such that  $\frac{1}{\lambda}(e^{\lambda f} - 1)$  is convex, for a fixed  $\lambda \neq 0$ . Furthermore, instead of the convex conjugate, we used the  $\lambda$ -conjugate given by:

$$f^{(\lambda)}(u) = \sup_x \left( \frac{1}{\lambda} \log(1 + \lambda x \cdot u) - f(x) \right).$$

The expression of the  $\lambda$ -duality:

$$\kappa_\lambda(x \cdot u) = f(x) + f^{(\lambda)}(u),$$

turns out to be a re-write of the Legendre duality between  $\tilde{x}$  and  $u$ :

$$\tilde{x} \cdot u = g(\tilde{x}) + g^*(u), \quad \text{with } \tilde{x} = x e^{-\lambda f(x)};$$

and a re-write of the Legendre duality between  $x$  and  $\tilde{u}$ :

$$x \cdot \tilde{u} = \tilde{g}^*(x) + \tilde{g}(\tilde{u}), \quad \text{with } \tilde{u} = u e^{-\lambda f^{(\lambda)}(u)}.$$

Therefore,  $\lambda$ -duality is in essence the Legendre duality with a  $\lambda$ -dependent rescaling of the variables:

$$\tilde{x} = x e^{-\lambda f(x)} \iff x = \frac{\tilde{x}}{1 - \lambda g(\tilde{x})}$$

and:

$$\tilde{u} = u e^{-\lambda f^{(\lambda)}(u)} \iff u = \frac{\tilde{u}}{1 - \lambda \tilde{g}(\tilde{u})}.$$

The two pairs of convex functions  $g, g^*$  and  $\tilde{g}, \tilde{g}^*$  are linked with the pair of  $\lambda$ -convex functions  $f, f^{(\lambda)}$  via:

$$\begin{aligned} g(\tilde{x}) &= G_{-\lambda, f}(x) = \gamma_{-\lambda}(f(x)) = (\gamma_{-\lambda} \circ f)(x); \\ g^*(u) &= G_{\lambda, f^{(\lambda)}}(u) = \gamma_\lambda(f^{(\lambda)}(u)) = (\gamma_\lambda \circ f^{(\lambda)})(u); \\ \tilde{g}(\tilde{u}) &= G_{-\lambda, f^{(\lambda)}}(u) = \gamma_{-\lambda}(f^{(\lambda)}(u)) = (\gamma_{-\lambda} \circ f^{(\lambda)})(u); \\ \tilde{g}^*(x) &= G_{\lambda, f}(x) = \gamma_\lambda(f(x)) = (\gamma_\lambda \circ f)(x). \end{aligned}$$

The  $\lambda$ -duality leads to nontrivial mathematical questions, e.g., a differential calculus in the spirit of Rockafellar and analogous to functions of the Legendre type. Some of the derivations in the current paper were heuristic, and a complete and rigorous development is left for future research.

Coming back to the probability families, we first verified that the subtractive potential  $\phi_\lambda(\theta)$  is convex in  $\theta$  and the divisive potential  $\varphi_\lambda(\vartheta)$  is  $\lambda$ -convex in  $\vartheta$ . Subtractive normalization using  $\phi_\lambda(\theta)$  is associated with the regular Legendre duality, whereas divisive normalization using  $\varphi_\lambda(\vartheta)$  is associated with the  $\lambda$ -duality. This gives an interpretation of the distinctiveness of Rényi entropy (used in the latter) from Tsallis entropy (used in the former) based on their intimate connection to the  $\lambda$ -duality (for  $\lambda \neq 0$ ) or to the Legendre duality. As  $\lambda$  is the parameter that controls the curvature in the Riemannian geometry of these probability families (see [12]), our framework provides a simple parametric deformation from the dually flat geometry (of the exponential model) to the dually projectively flat geometry (of the  $\lambda$ -exponential model). We expect that this framework will generate new insights in the applications of the  $q$ -exponential family and related concepts in statistical physics and information science.

**Table 1.** Generalization of objects from the Hessian (dually flat) geometry to the  $\lambda$ -deformed (dually projectively flat) geometry.

Objects	Conventional ( $\lambda = 0$ )	$\lambda$ -Deformed
transformation	Legendre	$\lambda$ -Legendre
conjugation	$\sup_x(x \cdot u - f(x))$	$\sup_x(\kappa_\lambda(x \cdot u) - f(x))$
potentials	convex	$\lambda$ -convex
associated divergence	Bregman	$\lambda$ -logarithmic
Riemannian metric	Hessian	conformal Hessian
affine connections	dually flat	dually projectively flat
curvature of connections	0	constant $\lambda \neq 0$
biorthogonal coordinates	$(x, u)$	$(\tilde{x}, u)$ or $(x, \tilde{u})$
probability family	exponential	$\lambda$ -exponential
probability family	mixture	$\lambda$ -mixture
associated divergence	Kullback–Leibler	Rényi
associated entropy	Shannon	Rényi/Tsallis

**Author Contributions:** Conceptualization, J.Z.; Formal analysis, J.Z. and T.-K.L.W.; Investigation, T.-K.L.W.; Writing—original draft, J.Z.; Writing—review & editing, J.Z. and T.-K.L.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** J.Z. is supported by United States Air Force Office of Scientific Research, grant number AFOSR-FA9550-19-1-0213. T.-K.L.W. is supported by NSERC Discovery Grant RGPIN-2019-04419 and a Connaught New Researcher Award.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

- Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000; Volume 191.
- Ay, N.; Jost, J.; Van Lê, H.; Schwachhöfer, L. Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **2015**, *162*, 327–364. [[CrossRef](#)]
- Zhang, J.; Khan, G. From Hessian to Weitzenböck: Manifolds with torsion-carrying connections. *Inf. Geom.* **2019**, *2*, 77–98. [[CrossRef](#)]
- Zhang, J.; Khan, G. Statistical mirror symmetry. *Differ. Geom. Its Appl.* **2020**, *73*, 101678. [[CrossRef](#)]
- Naudts, J. Estimators, escort probabilities, and  $\phi$ -exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
- Naudts, J.; Zhang, J. Rho–tau embedding and gauge freedom in information geometry. *Inf. Geom.* **2018**, *1*, 79–115. [[CrossRef](#)]
- Murata, N.; Takenouchi, T.; Kanamori, T.; Eguchi, S. Information geometry of U-Boost and Bregman divergence. *Neural Comput.* **2004**, *16*, 1437–1481. [[CrossRef](#)]
- Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195. [[CrossRef](#)]
- Amari, S.I.; Ohara, A. Geometry of  $q$ -exponential family of probability distributions. *Entropy* **2011**, *13*, 1170–1185. [[CrossRef](#)]
- Amari, S.I.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 4308–4319. [[CrossRef](#)]
- Ohara, A.; Matsuzoe, H.; Amari, S.I. Conformal geometry of escort probability and its applications. *Mod. Phys. Lett. B* **2012**, *26*, 1250063. [[CrossRef](#)]
- Wong, T.K.L. Logarithmic divergences from optimal transport and Rényi geometry. *Inf. Geom.* **2018**, *1*, 39–78. [[CrossRef](#)]
- Pal, S.; Wong, T.K.L. Multiplicative Schrödinger problem and the Dirichlet transport. *Probab. Theory Relat. Fields* **2020**, *178*, 613–654. [[CrossRef](#)]
- Pal, S.; Wong, T.K.L. The geometry of relative arbitrage. *Math. Financ. Econ.* **2016**, *10*, 263–293. [[CrossRef](#)]

15. Pal, S.; Wong, T.K.L. Exponentially concave functions and a new information geometry. *Ann. Probab.* **2018**, *46*, 1070–1113. [[CrossRef](#)]
16. Wong, T.K.L. Information Geometry in Portfolio Theory. In *Geometric Structures of Information*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 105–136.
17. Wong, T.K.L.; Zhang, J. Tsallis and Rényi deformations linked via a new  $\lambda$ -duality. *arXiv* **2021**, arXiv:2107.11925.
18. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tohoku Math. J. Second Ser.* **1994**, *46*, 427–433. [[CrossRef](#)]
19. Wong, T.K.L.; Yang, J. Logarithmic divergence: Geometry and interpretation of curvature. In *International Conference on Geometric Science of Information*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 413–422.
20. Amari, S.I. *Information Geometry and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2016.
21. Eguchi, S. Geometry of minimum contrast. *Hiroshima Math. J.* **1992**, *22*, 631–647. [[CrossRef](#)]
22. Matumoto, T. Any statistical manifold has a contrast function—On the C3-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J.* **1993**, *23*, 327–332. [[CrossRef](#)]
23. Nagaoka, H.; Amari, S.I. *Differential Geometry of Smooth Families of Probability Distributions*; Technical Report METR 82-7; University of Tokyo: Tokyo, Japan, 1982.
24. Zhang, J. Referential duality and representational duality on statistical manifolds. In Proceedings of the Second International Symposium on Information Geometry and Its Applications, Tokyo, Japan, 12–16 December 2005; Volume 1216, pp. 58–67.
25. Zhang, J. Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. *Entropy* **2013**, *15*, 5384–5418. [[CrossRef](#)]
26. Blondel, M.; Martins, A.F.; Niculae, V. Learning with Fenchel-Young losses. *J. Mach. Learn. Res.* **2020**, *21*, 1–69.
27. Naudts, J. Generalized exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149. [[CrossRef](#)]
28. Naudts, J. *Generalized Thermostatistics*; Springer: Berlin/Heidelberg, Germany, 2011.
29. Zhang, J. On monotone embedding in information geometry. *Entropy* **2015**, *17*, 4485–4499. [[CrossRef](#)]
30. Eguchi, S. Information geometry and statistical pattern recognition. *Sugaku Expos.* **2006**, *19*, 197–216.
31. Matsuzoe, H. Hessian structures on deformed exponential families and their conformal structures. *Differ. Geom. Its Appl.* **2014**, *35*, 323–333. [[CrossRef](#)]
32. Newton, N.J. An infinite-dimensional statistical manifold modelled on Hilbert space. *J. Funct. Anal.* **2012**, *263*, 1661–1681. [[CrossRef](#)]
33. Montrucchio, L.; Pistone, G. Deformed exponential bundle: The linear growth case. In *International Conference on Geometric Science of Information*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 239–246.
34. De Andrade, L.H.; Vieira, F.L.; Cavalcante, C.C. On Normalization Functions and  $\phi$ -Families of Probability Distributions. In *Progress in Information Geometry: Theory and Applications*; Springer Nature Switzerland AG: Cham, Switzerland, 2021.
35. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
36. Villani, C. *Topics in Optimal Transportation*; American Mathematical Society: Providence, RI, USA, 2003.
37. Villani, C. *Optimal Transport: Old and New*; Springer: Berlin/Heidelberg, Germany, 2008.
38. Wong, T.K.L.; Yang, J. Pseudo-Riemannian geometry encodes information geometry in optimal transport. *Inf. Geom.* **2021**, 1–29. [[CrossRef](#)]
39. Tsallis, C. Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
40. Tsallis, C. What are the numbers that experiments provide. *Quim. Nova* **1994**, *17*, 468–471.
41. Rényi, A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*; The Regents of the University of California: Oakland, CA, USA, 1961.
42. Van Erven, T.; Harremoës, P. Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [[CrossRef](#)]