Canonical Divergence for Flat *α*-Connections: Classical and Quantum

Domenico Felice [1,*] **and Nihat Ay** [1,2,3]

[1] Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
[2] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA
[3] Faculty of Mathematics and Computer Science, University of Leipzig, PF 100920, 04009 Leipzig, Germany
* Correspondence: felice@mis.mpg.de

check for updates

Abstract: A recent canonical divergence, which is introduced on a smooth manifold M endowed with a general dualistic structure (g, ∇, ∇^*), is considered for flat *α*-connections. In the classical setting, we compute such a canonical divergence on the manifold of positive measures and prove that it coincides with the classical *α*-divergence. In the quantum framework, the recent canonical divergence is evaluated for the quantum *α*-connections on the manifold of all positive definite Hermitian operators. In this case as well, we obtain that the recent canonical divergence is the quantum *α*-divergence.

Keywords: information geometry; quantum relative entropy; alpha connections; canonical divergence; Kullback-Leibler divergence

1. Introduction

Methods of Information Geometry (IG) [1] are ubiquitous in physical sciences and encompass both classical and quantum systems [2]. The natural object of study in IG is a quadruple (M, g, ∇, ∇^*) given by a smooth manifold M, a Riemannian metric g, and a pair of affine connections on M, which are dual with respect to g,

$$X \, g\,(Y, Z) \,=\, g\,(\nabla_X Y, Z) + g\,(Y, \nabla^*_X Z) \,, \tag{1}$$

for all sections $X, Y, Z \in \mathcal{T}(M)$. The quadruple (M, g, ∇, ∇^*) is called a statistical manifold whenever the dual connections are both torsion-free [3]. Actually, the notion of the statistical manifold, introduced by Lauritzen [4], is usually referred to the triple (M, g, T), where $T(X, Y, Z) = g\,(\nabla^*_X Y - \nabla_X Y, Z)$ is a three-symmetric tensor. However, when ∇ and ∇^* are both torsion-free connections, the structures (M, g, ∇, ∇^*) and (M, g, T) are equivalent [3]. When M is a manifold of probability distributions, $g \equiv g^F$ is the Fisher metric, and $\nabla \equiv \nabla^{(e)}$ and $\nabla^* \equiv \nabla^{(m)}$ are the exponential and mixture connections [5], IG has been successfully applied to many fields, such as statistical inference, control systems theory, and neural networks (see [6] and the references therein for a comprehensive literature on applications of IG).

The geometric structure of a statistical manifold is encoded by a smooth function $\mathcal{D} : M \times M \to \mathbb{R}$ such that:

$$\mathcal{D}(p, q) \geq 0, \quad \text{and} \quad \mathcal{D}(p, q) = 0 \quad \text{iff} \quad p = q, \tag{2}$$

for all $p, q \in M$ [7]. The dualistic structure (g, ∇, ∇^*) of M is then recovered in the following way:

$$g_{ij}(p) = -\left. \partial_i \partial'_j \mathcal{D}(\xi_p, \xi_q)\right|_{p=q} = \left. \partial'_i \partial'_j \mathcal{D}(\xi_p, \xi_q)\right|_{p=q} \tag{3}$$

$$\Gamma_{ijk}(p) = -\left. \partial_i \partial_j \partial'_k \mathcal{D}(\xi_p, \xi_q)\right|_{p=q}, \qquad \Gamma^*_{ijk}(p) = -\left. \partial'_i \partial'_j \partial_k \mathcal{D}(\xi_p, \xi_q)\right|_{p=q}. \tag{4}$$

Here,

$$\partial_i = \frac{\partial}{\partial \xi_p^i} \quad \text{and} \quad \partial_i' = \frac{\partial}{\partial \xi_q^i}$$

and $\{\boldsymbol{\xi}_p := (\xi_p^1, \ldots, \xi_p^n)\}$ and $\{\boldsymbol{\xi}_q := (\xi_q^1, \ldots, \xi_q^n)\}$ are local coordinate systems of p and q, respectively. Moreover, $\Gamma_{ijk} = \mathrm{g}\left(\nabla_{\partial_i}\partial_j, \partial_k\right)$, $\Gamma_{ijk}^* = \mathrm{g}\left(\nabla_{\partial_i}^*\partial_j, \partial_k\right)$ are the symbols of the dual connections ∇ and ∇^*, respectively. The function $\mathcal{D}$ is called a divergence or contrast function of the statistical manifold $(\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ [8].

The function $\mathcal{D}$ is called a flat divergence if the dualistic structure $(\mathrm{g}, \nabla, \nabla^*)$ introduced on a smooth manifold M by Equations (3) and (4) is flat, namely the curvature tensors $\mathcal{R}(\nabla)$ and $\mathcal{R}^*(\nabla^*)$ are zero. In this particular case, an attempt to connect IG with physics was established in [9]. There, the connection between IG and integrable dynamical systems was bridged by a divergence, which is canonical in some sense. More precisely, for $q \in \mathrm{M}$, the gradient flows $\dot{\boldsymbol{\xi}} = -\mathrm{grad}_{\boldsymbol{\xi}}\, \mathcal{D}(\boldsymbol{\xi}_q, \cdot)$ and $\dot{\boldsymbol{\xi}} = -\mathrm{grad}_{\boldsymbol{\xi}}\, \mathcal{D}(\cdot, \boldsymbol{\xi}_q)$ converge to the point q along the ∇^*-geodesic and the ∇-geodesic, respectively. In this context, when M is the manifold of Gaussian distributions with mean μ and variance σ^2, the Kullback-Leibler (KL) divergence induces a dualistic structure given by the Fisher metric and the (e), (m) connections. In this case, the dynamics of the above-mentioned gradient flows, given now in terms of the KL divergence, turns out to be the Uhlenbeck-Ornstein process $\mu = \mu_0 + v(t + \tau)$, $\sigma^2 = 2D(t + \tau)$ characterized by the drift coefficient v and diffusion coefficient D [9]. Further connections between IG and dynamical systems can be found in [10], where certain gradient flows on Gaussian and multinomial distributions are characterized as completely integrable Hamiltonian systems.

The Kullback-Leibler divergence has been effectively employed to quantify the complexity of a system described by a probability distribution p in terms of its deviation from an exponential family of probability distributions [11]. The quantum version of the KL divergence, namely the quantum relative entropy, induces on the manifold of quantum states a dually flat structure given by a quantum version of the Fisher metric tensor and two flat connections, also called the mixture and the exponential connections, which are dual in the sense of Equation (1) [12]. Moreover, the quantum relative entropy has been used to quantify the many-party correlations of a composite quantum state ρ as the deviation of it from a Gibbs family of quantum states [13]. The utility of the quantum relative entropy as a measure of the complexity for quantum states was shown in [14], where algorithms for its evaluation were studied. Furthermore, in that context, the many-party correlations were related to the entanglement of quantum systems as defined in [15].

A generalization of the flat structures induced by the Kullback-Leibler divergence on the finite classical systems and by the quantum relative entropy on the finite quantum states is provided by the classical α-divergence and the quantum α-divergence, respectively. Both generate a one-parameter family of connections, the α-connections, which are dual with respect to the Fisher metric in the classical case [1], whereas in the quantum case, they are dual with respect to the Wigner-Yanase-Dyson metric [16].

From a physical viewpoint, it is worth remarking that the Boltzmann-Gibbs distribution in statistical physics is an exponential family such that an invariant flat structure is given to the underlying manifold in terms of the (e)-connection [6]. Tsallis generalized the concept of the Boltzmann-Gibbs distribution by introducing a generalized entropy, called the q-entropy, for studying various phenomena not included in the conventional Boltzmann-Gibbs framework [17,18]. Actually, the α-geometry, which is induced by the α-divergence, covers the geometry of q-entropy physics [19]. Therefore, the α-divergence can be understood as a generalization of the classical KL divergence also from a physical standpoint. In the quantum case, the geometry of q-entropy physics has been successfully employed for carrying out a criterion that detects the critical frontier, which has separable states on one side and quantum entangled ones on the other [20].

The purpose of the present article is to consider a recent canonical divergence, introduced by Ay and Amari in [21], as a tool for unifying classical and quantum information geometry. In particular, we

aim to prove that this canonical divergence is the classical α-divergence when computed on the space of positive measures, as well as the quantum α-divergence if evaluated on the manifold of positive definite Hermitian operators.

2. Canonical Divergence and the Inverse Problem in Information Geometry

The inverse problem within information geometry concerns the search for a divergence function $\mathcal{D}$, which recovers a given dual structure (g, ∇, ∇^*) of a smooth manifold M according to Equations (3) and (4). The solution to this problem was provided by Matumoto, who showed that such a divergence always exists for any statistical manifold [22]. Nonetheless, this solution is not unique, and infinitely many divergences can be defined on M, which gives the same dualistic structure. For this reason, seeking a divergence that can be considered as the most natural is of utmost importance. To this end, Amari and Nagaoka defined a Bregman-type divergence on dually flat manifolds [1]. This one has relevant properties such as the generalized Pythagorean theorem and the geodesic projection theorem, and it is commonly assessed as the natural solution of the inverse problem in information geometry for dually flat manifolds. This is exactly why the Amari and Nagaoka divergence is referred to as the canonical divergence of dually flat statistical manifolds. However, the need for a general canonical divergence, which applies to any dualistic structure, is a very crucial issue, as pointed out in [23]. According to the theory developed in [3], a divergence function $\mathcal{D}$ of a statistical manifold (M, g, ∇, ∇^*) is called canonical if:

- $\mathcal{D}$ generates the dualistic structure (g, ∇, ∇^*) based on Equations (3) and (4);
- $\mathcal{D}$ is one half of the squared Riemannian distance, i.e., $\mathcal{D}(p, q) = \dfrac{1}{2}\, d(p, q)^2$, when the statistical manifold is self-dual, namely when $\nabla = \nabla^*$ coincides with the Levi–Civita connection of g;
- $\mathcal{D}$ is the canonical divergence of the Bregman type when (M, g, ∇, ∇^*) is dually flat.

Ay and Amari recently defined a divergence for a general dualistic structure in terms of geodesic integration of the inverse exponential map [21]. It turns out that such a divergence satisfies all the above-mentioned requirements. Therefore, it can be viewed as a canonical divergence for a general dualistic structure. For $p, q \in$ M, consider the ∇-geodesic $\widetilde{\gamma}(t)$ $(0 \leq t \leq 1)$ connecting q with p, the recent canonical divergence introduced in [21] is then defined by:

$$\mathcal{D}(p, q) := \int_0^1 \left\langle X_t(p), \dot{\widetilde{\gamma}}(t) \right\rangle_{\widetilde{\gamma}(t)} dt, \quad X_t(p) := \exp_{\widetilde{\gamma}(t)}^{-1}(p). \tag{5}$$

The ∇-exponential map, $\exp : TM \to M$, is defined by $\exp(X) = \gamma_X(1)$ whenever the ∇-geodesic $\gamma_X(t)$, satisfying $\dot{\gamma}_X(0) = X$, exists on an interval of t containing $[0, 1]$. Therefore, if $\gamma : [0, 1] \to$ M is the ∇-geodesic such that $\gamma(0) = p$ and $\gamma(1) = q$, the inverse at p of the exponential map is given by $\exp_p^{-1}(q) := \dot{\gamma}(0)$. According to this definition, for every $t \in [0, 1]$, we can consider the ∇-geodesic $\gamma_t(s)$ such that $\gamma_t(0) = p$ and $\gamma_t(1) = \gamma(t)$ and then define the ∇-velocity vector at p by $X_p(\gamma(t)) := \exp_p^{-1}(\gamma(t)) = \dot{\gamma}_t(0)$. In this way, the vector field $X_t(p)$ of Equation (5) turns out to be given by $X_t(p) = P_{\gamma(t)} X_p(\gamma(t)) = t\, \dot{\gamma}(t)$. Here, $P : T_pM \to T_{\gamma(t)}M$ is the ∇-parallel transport from p to $\gamma(t)$. In light of all this, the divergence $\mathcal{D}(p, q)$ assumes the following useful expression:

$$\mathcal{D}(p, q) = \int_0^1 t\, \|\dot{\gamma}(t)\|^2 \, dt. \tag{6}$$

Analogously, the dual function of $\mathcal{D}(p, q)$ is defined as the ∇^*-geodesic integration of the inverse of the ∇^*-exponential map [21]. Therefore, we have for the dual divergence $\mathcal{D}^*$ a similar expression as Equation (6) for the canonical divergence $\mathcal{D}$:

$$\mathcal{D}^*(p, q) = \int_0^1 t\, \|\dot{\gamma}^*(t)\|^2 \, dt, \tag{7}$$

where $\gamma^*(t)$ $(0 \leq t \leq 1)$ is the ∇^*-geodesic connecting p with q.

The canonical divergence given by Equation (6) has been recently proposed as a tool for unifying classical and quantum information geometry [24]. In particular, it has been considered on the simplex of probability measures:

$$\mathcal{S} := \left\{ p = \sum_i p_i \, \delta_i \in \mathbb{R}^n \mid p_i > 0 \text{ for all } i, \text{ and } \sum_i p_i = 1 \right\}, \tag{8}$$

as well as on the manifold of quantum finite states,

$$\mathrm{S} := \left\{ \rho : \mathcal{H} \to \mathbb{C} \mid \rho = \rho^\dagger > 0, \, \mathrm{Tr}\rho = 1 \right\}, \tag{9}$$

where $\mathcal{H}$ denotes a finite-dimensional Hilbert space and ρ is any Hermitian operator on $\mathcal{H}$. The natural dualistic structure on the simplex $\mathcal{S}$ is given, in classical information geometry, in terms of the Fisher metric g^{F} and two flat connections, the mixture $\nabla^{(m)}$ and the exponential $\nabla^{(e)}$ ones, which are dual with respect to g^{F} in the sense of Equation (1) [5]. Very remarkably, the Fisher metric is the only monotone Riemannian metric (up to a positive factor) on the class of finite probability simplices [25]. The quantum version of a monotone Riemannian metric on the manifold of quantum finite states S is given in terms of the notion of stochastic mapping [26]. More precisely, let $\mathcal{A}$ be the set of all Hermitian operators on the Hilbert space $\mathcal{H}$; a linear mapping $T : \mathcal{A} \to \mathcal{A}$ is said to be stochastic if $T(\mathrm{S}) \subset \mathrm{S}$ and T is completely positive [27]. Furthermore, a family $\{\langle\langle \, , \, \rangle\rangle_\rho \mid \rho \in \mathrm{S}\}$ of inner products on $\mathcal{A}$ is said to be monotone if $\langle\langle A, A \rangle\rangle_\rho \geq \langle\langle T(A), T(A) \rangle\rangle_{T(\rho)}$ for any arbitrary stochastic map T and for every $\rho \in \mathrm{S}$ [26]. Due to Petz, there are infinitely many monotone inner products on $\mathcal{H}$. Therefore, the quantum analogue of the Fisher metric is not unique [28]. However, when the flat connections are required to be torsion-free, a natural dualistic structure on the manifold of quantum states S is the one induced by the Bogoliubov-Kubo-Mori (BKM) inner product [12]. Furthermore, it turns out that the only monotone metrics that make the mixture connection $\nabla^{(m)}$ and the exponential connection $\nabla^{(e)}$ dual are the scalar multiples of the BKM metric [29]. When the canonical divergence (6) is computed on the simplex $\mathcal{S}$, it is shown to be the Kullback-Leibler divergence,

$$\mathcal{D}(p, q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right), \quad p, q \in \mathcal{S}, \tag{10}$$

which proves that $\mathcal{D}(p, q)$ recovers the natural dualistic structure on $\mathcal{S}$ given by the Fisher metric g^{F} and the dually flat connections $\nabla^{(m)}$ and $\nabla^{(e)}$ [24]. Analogously, when $\mathcal{D}$ is considered on the manifold of quantum states S, it has been proven that:

$$\mathcal{D}(\rho, \sigma) = \mathrm{Tr}\,\rho \left(\log \rho - \log \sigma\right), \quad \rho, \sigma \in \mathrm{S}, \tag{11}$$

where Tr denotes the trace operator on the finite-dimensional Hilbert space of density matrices [24]. The function on the right-hand side of Equation (11) is called the quantum relative entropy, and it recovers the dual structure of S given by the metric induced by the BKM inner product and the flat connections $\nabla^{(m)}$ and $\nabla^{(e)}$ [1].

In this article, we aim to investigate the canonical divergence (6) on the manifold of positive measures, as well as on the space of positive definite Hermitian operators for the more general α-connections. In the classical information geometry, the one-parameter family of the α-connections is defined on the manifold of positive measures by the linear combination of the mixture and exponential connections [6],

$$\nabla^\alpha = \frac{1-\alpha}{2} \nabla^{(m)} + \frac{1+\alpha}{2} \nabla^{(e)}. \tag{12}$$

It turns out that ∇^α and $\nabla^{-\alpha}$ are dual with respect to the Fisher metric g^F in the sense of Equation (1). In quantum information geometry, α-connections appeared also in terms of the Amari α-embeddings [30]. While in the classical information geometry, the two definitions coincide, this is no longer true in quantum information geometry [31]. In the present paper, we consider the definition of the quantum α-connections by means of the Amari α-embedding, which excludes that they can be obtained by the convex mixture of $\nabla^{(m)}$ and $\nabla^{(e)}$ connections. The natural inner product that makes the quantum α-connections dual in the sense of Equation (1) is the Wigner-Yanase-Dyson (WYD) metric, which appeared for the first time in the context of quantum information geometry in the work of Hasegawa [32]. Actually, it turns out that this is the only monotone metric (up to a scalar multiple) that makes the quantum ∇^α and $\nabla^{-\alpha}$ dual [16].

3. Classical Flat Alpha-Divergence

We represent measures on the set $\{1, \ldots, n\}$ as elements of $\mathbb{R}^n$. In this representation, the Dirac measures δ_i, $i = 1, \ldots, n$ form the canonical basis of $\mathbb{R}^n$. The n-dimensional cone of positive measures on the set $\{1, \ldots, n\}$ is then defined by:

$$\mathcal{M}_+ := \mathbb{R}^n_+ = \left\{ p = \sum_{i=1}^n p_i \, \delta_i \in \mathbb{R}^n \mid p_i > 0 \text{ for all } i \right\}. \tag{13}$$

The very natural Riemannian metric on $\mathcal{M}_+$ is the Fisher metric [6], which is defined by:

$$g_p^F(X, Y) = \sum_{i=1}^n \frac{1}{p_i} X_i Y_i, \tag{14}$$

for all $p \in \mathcal{M}_+$ and $X, Y \in T_p \mathcal{M}_+$. Here, $T_p \mathcal{M}_+$ denotes the tangent space to $\mathcal{M}_+$ at p. Given the mixture and the exponential connections, we can define the α-connection on $\mathcal{M}_+$ by using Equation (12). Recalling that $T_p \mathcal{M}_+ \equiv \mathbb{R}^n$, we can write any $X \in T_p \mathcal{M}_+$ with respect to the canonical basis of $\mathbb{R}^n$. Hence, the (m)-connection on $\mathcal{M}_+$ reads as follows:

$$\left. \nabla_X^{(m)} Y \right|_p = \sum_{i=1}^n \frac{\partial Y_i}{\partial X}(p) \, \delta_i, \tag{15}$$

for all vector fields $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ on $\mathcal{M}_+$. Here, $\partial/\partial X$ denotes the derivative in the direction $X \in T_p \mathcal{M}_+$. On the contrary, the (e)-connection is given by:

$$\left. \nabla_X^{(e)} Y \right|_p = \sum_{i=1}^n \left(\frac{\partial Y_i}{\partial X}(p) - \frac{1}{p_i} X_i Y_i \right) \delta_i. \tag{16}$$

Therefore, by applying Equation (12), we can describe the α-connection for the manifold $\mathcal{M}_+$ as follows,

$$
\begin{aligned}
\left. \nabla_X^\alpha Y \right|_p &= \frac{1-\alpha}{2} \left. \nabla_X^{(m)} Y \right|_p + \frac{1+\alpha}{2} \left. \nabla_X^{(e)} Y \right|_p \\
&= \left. \nabla_X^{(m)} Y \right|_p + \frac{1+\alpha}{2} \left(\left. \nabla_X^{(e)} Y \right|_p - \left. \nabla_X^{(m)} Y \right|_p \right) \\
&= \sum_{i=1}^n \left(\frac{\partial Y_i}{\partial X}(p) - \frac{1+\alpha}{2} \frac{X_i Y_i}{p_i} \right) \delta_i,
\end{aligned}
\tag{17}
$$

for all vector fields X, Y on $\mathcal{M}_+$. It turns out that the dualistic structure $(g^F, \nabla^\alpha, \nabla^{-\alpha})$ is dually flat, i.e., the Riemannian curvature tensors of ∇^α and $\nabla^{-\alpha}$ are zero [6]. Furthermore, very naturally, such an α-flat dualistic structure is induced on $\mathcal{M}_+$ by the α-divergence $D^{(\alpha)}$, which is commonly assessed

as the canonical solution of the inverse problem for recovering the dually flat α-structure $(g^F, \nabla^\alpha, \nabla^{-\alpha})$ on $\mathcal{M}_+$ [6]. Given $p, q \in \mathcal{M}_+$, the α-divergence between p and q is given by:

$$D^{(\alpha)}(p,q) = \sum_{i=1}^{n} \left(\frac{2}{1-\alpha} q_i + \frac{2}{1+\alpha} p_i - \frac{4}{1-\alpha^2} q_i^{\frac{1+\alpha}{2}} p_i^{\frac{1-\alpha}{2}} \right). \tag{18}$$

The flat α-divergence (18) has been deeply studied, and its features have been widely discussed in the literature (see [3,6] for more details). In particular, we point out that $D^{(\alpha)}$ is a continuous function of the parameter α, and the limit $\alpha \to -1$ gives the well-known Kullback-Leibler divergence on the manifold of positive measures,

$$\lim_{\alpha \to -1} D^{(\alpha)}(p,q) = \sum_{i=1}^{n} \left(q_i - p_i - p_i \log \frac{q_i}{p_i} \right), \quad p,q \in \mathcal{M}_+, \tag{19}$$

whereas the limit $\alpha \to +1$ gives:

$$\lim_{\alpha \to +1} D^{(\alpha)}(p,q) = \sum_{i=1}^{n} \left(p_i - q_i - q_i \log \frac{p_i}{q_i} \right), \quad p,q \in \mathcal{M}_+. $$

Let us observe that, if we restrict (19) to the simplex of probability distributions, namely when $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, then we obtain the function (10). At this point, it is worth mentioning the close connection between the α-divergence (18) and the Tsallis relative entropy, or q-divergence, on the manifold of positive measures. The q-divergence on $\mathcal{M}_+$ is defined by (see [33] for more details):

$$D_q(p,q) = \frac{1}{1-q} \sum_{i=1}^{n} \left(q\, p_i + (1-q)\, q_i - p_i^q q_i^{1-q} \right), \quad p,q \in \mathcal{M}_+. \tag{20}$$

By setting $\alpha = 1 - 2q$, we can easily verify that the flat α-divergence (18) and the Tsallis q-divergence (20) coincide up to a scaling factor. Further, the limit $q \to 1$, that is $\alpha \to -1$, recovers the Kullback-Leibler divergence,

$$\lim_{q \to 1} D_q(p,q) = \lim_{\alpha \to -1} \frac{1-\alpha}{2} D^{(\alpha)}(p,q) = \sum_{i=1}^{n} \left(q_i - p_i - p_i \log \frac{q_i}{p_i} \right), \quad p,q \in \mathcal{M}_+.$$

In this section, we aim to compute the canonical divergence (6) on $\mathcal{M}_+$ for the α-connections given by (17). In order to achieve this result, we need to consider the geodesic with respect to the ∇^α-connection. Let $p, q \in \mathcal{M}_+$, a curve $\gamma : [0,1] \to \mathcal{M}_+$ from p to q is an α-geodesic iff $\nabla^\alpha_{\dot{\gamma}} \dot{\gamma} = 0$ and $\gamma(0) = p$, $\gamma(1) = q$. From Equation (17), we then obtain the following geodesic equations:

$$\ddot{\gamma}_i - \frac{1+\alpha}{2} \frac{\dot{\gamma}_i^2}{\gamma_i} = 0, \, i = 1, \ldots, n, \quad \gamma(0) = p, \, \gamma(1) = q, \tag{21}$$

where $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$. Hence, the solution of (21) is the α-geodesic from p to q, and it is given by:

$$\gamma^{(\alpha)}(t) = \left((1-t)\, p^{\frac{1-\alpha}{2}} + t\, q^{\frac{1-\alpha}{2}} \right)^{\frac{2}{1-\alpha}}, \quad t \in [0,1]. \tag{22}$$

At this point, we can apply Equation (6) to compute the canonical divergence $\mathcal{D}(p,q)$ on the manifold $\mathcal{M}_+$ of positive measures. From Equation (14), we obtain:

$$
\begin{aligned}
\|\dot{\gamma}^{(\alpha)}(t)\|^2_{\gamma^{(\alpha)}(t)} &= g^{\mathrm{F}}_{\gamma^{(\alpha)}(t)}\left(\dot{\gamma}^{(\alpha)}(t), \dot{\gamma}^{(\alpha)}(t)\right) = \sum_{i=1}^{n} \frac{\left(\dot{\gamma}^{(\alpha)}_i\right)^2}{\gamma^{(\alpha)}_i} \\
&= \sum_{i=1}^{n} \frac{4}{(1-\alpha)^2} \frac{\left((1-t)\,p^{\frac{1-\alpha}{2}} + t\,q^{\frac{1-\alpha}{2}}\right)^{2\frac{1+\alpha}{1-\alpha}}}{\left((1-t)\,p^{\frac{1-\alpha}{2}} + t\,q^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}} \left(q^{\frac{1-\alpha}{2}} - p^{\frac{1-\alpha}{2}}\right)^2 \\
&= \sum_{i=1}^{n} \frac{4}{(1-\alpha)^2} \left((1-t)\,p^{\frac{1-\alpha}{2}} + t\,q^{\frac{1-\alpha}{2}}\right)^{\frac{2\alpha}{1-\alpha}} \left(q^{\frac{1-\alpha}{2}} - p^{\frac{1-\alpha}{2}}\right)^2.
\end{aligned}
$$

Now, we can compute the integral in (6) by performing an integration by parts:

$$
\begin{aligned}
\mathcal{D}(p,q) &= \sum_{i=1}^{n} \int_0^1 t\,\frac{4}{(1-\alpha)^2} \left[\left((1-t)\,p_i^{\frac{1-\alpha}{2}} + t\,q_i^{\frac{1-\alpha}{2}}\right)^{\frac{2\alpha}{1-\alpha}} \left(q_i^{\frac{1-\alpha}{2}} - p_i^{\frac{1-\alpha}{2}}\right)\right] \left(q_i^{\frac{1-\alpha}{2}} - p_i^{\frac{1-\alpha}{2}}\right)\,\mathrm{d}t \\
&= \sum_{i=1}^{n} \left\{ \left[t\,\frac{4}{(1-\alpha)^2}\frac{1-\alpha}{1+\alpha} \left((1-t)\,p_i^{\frac{1-\alpha}{2}} + t\,q_i^{\frac{1-\alpha}{2}}\right)^{\frac{1+\alpha}{1-\alpha}}\right]_0^1 \left(q_i^{\frac{1-\alpha}{2}} - p_i^{\frac{1-\alpha}{2}}\right) \right. \\
&\qquad\left. - \int_0^1 \frac{4}{(1-\alpha)^2}\frac{1-\alpha}{1+\alpha} \left((1-t)\,p_i^{\frac{1-\alpha}{2}} + t\,q_i^{\frac{1-\alpha}{2}}\right)^{\frac{1+\alpha}{1-\alpha}} \left(q_i^{\frac{1-\alpha}{2}} - p_i^{\frac{1-\alpha}{2}}\right)\,\mathrm{d}t \right\} \\
&= \sum_{i=1}^{n} \frac{4}{1-\alpha^2} \left\{ q_i^{\frac{1+\alpha}{2}} \left(q_i^{\frac{1-\alpha}{2}} - p_i^{\frac{1-\alpha}{2}}\right) - \int_0^1 \left((1-t)\,p_i^{\frac{1-\alpha}{2}} + t\,q_i^{\frac{1-\alpha}{2}}\right)^{\frac{1+\alpha}{1-\alpha}} \left(q_i^{\frac{1-\alpha}{2}} - p^{\frac{1-\alpha}{2}}\right)\,\mathrm{d}t \right\} \\
&= \sum_{i=1}^{n} \frac{4}{1-\alpha^2} \left(q_i - p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}}\right) - \frac{4}{1-\alpha^2}\frac{1-\alpha}{2} \left[\left((1-t)\,p_i^{\frac{1-\alpha}{2}} + t\,q_i^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}}\right]_0^1 \\
&= \sum_{i=1}^{n} \left\{ \frac{4}{1-\alpha^2}\,q_i - \frac{4}{1-\alpha^2}\,p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}} - \frac{2}{1+\alpha}\,q_i + \frac{2}{1+\alpha}\,p_i \right\}.
\end{aligned}
$$

This proves that:

$$
\mathcal{D}(p,q) = \sum_{i=1}^{n} \left(\frac{2}{1-\alpha}\,q_i + \frac{2}{1+\alpha}\,p_i - \frac{4}{1-\alpha^2}\,q_i^{\frac{1+\alpha}{2}} p_i^{\frac{1-\alpha}{2}} \right), \tag{23}
$$

that is the canonical divergence (6) coincides with the α-divergence (18) on the manifold $\mathcal{M}_+$ of positive measures.

4. Quantum Flat Alpha-Divergence

The goal of this section is to compute the canonical divergence of Equation (6) on the manifold of positive definite Hermitian operators endowed with the quantum α-connections. In the classical case, the definition of the α-connection ∇^α can be equivalently given by Equation (12) or by means of the well-known Amari α-embedding [6]. In the quantum setting, by exploiting the linear structure of the manifold, we can introduce the quantum mixture connection $\nabla^{(m)}$ on the space of positive definite Hermitian operators. On the other hand, the quantum exponential connection $\nabla^{(e)}$ is introduced by relying on the linear structure of logarithms of the manifold. Both connections, $\nabla^{(m)}$ and $\nabla^{(e)}$, are flat. Furthermore, they turn out to be dual in the sense of Equation (1) with respect to the metric induced by the BKM inner product [12]. Actually, this metric is (up to a scalar multiple) the only one that makes $\nabla^{(m)}$ and $\nabla^{(e)}$ dual on the manifold of positive definite Hermitian operators. A generalization of these two natural connections is provided by the quantum α-connections, which, on the manifold under consideration, are flat ones, as well. In our approach, these α-connections are introduced in

terms of the Amari α-embeddings, and then, they cannot be obtained by the convex combination of the mixture and the exponential connections as in the classical setting [16]. It turns out that for $\alpha \in [-3, 3]$, the connections ∇^α and $\nabla^{-\alpha}$ are torsion-free and dual with respect to the WYD metric [34]. However, the target spaces of our approach are L^r, with $r = \dfrac{2}{1-\alpha}$, and then, we restrict our discussion to the range $\alpha \in (-1, 1)$.

Recall that in order to compute the canonical divergence (6), we need to get the α-geodesic γ_α between any two positive Hermitian operators ρ_1 and ρ_2. To do this, we shall describe the α-connections in terms of the ∇^α-parallel transport, which is introduced very naturally on the manifold of positive Hermitian operators. Given $\mathcal{B}(\mathcal{H})$, the algebra of linear operators on an N-dimensional complex Hilbert space $\mathcal{H}$, the subspace of Hermitian operators is an N^2-dimensional real vector space defined by:

$$\mathcal{A} := \{\rho \in \mathcal{B}(\mathcal{H}) \,|\, \rho = \rho^\dagger\}, \tag{24}$$

where $\rho^\dagger = \bar{\rho}^t$ and t, here, denotes the transpose matrix. Therefore, the manifold of all positive definite Hermitian operators, or more simply, quantum operators, is given by:

$$\mathrm{M}_+ := \{\rho \in \mathcal{A} \,|\, \rho > 0\}. \tag{25}$$

The α-embedding:

$$l_\alpha : \mathrm{M}_+ \to \mathcal{A}, \quad l_\alpha(\rho) := \frac{2}{1-\alpha} \rho^{\frac{1-\alpha}{2}}, \tag{26}$$

maps the manifold of quantum operators into the vector space $\mathcal{A}$ for all $\alpha \in (-1, 1)$ [1]. Furthermore, for $\rho \in \mathrm{M}_+$, we have that $\mathrm{T}_\rho \mathrm{M}_+ = \mathrm{T}_\rho \mathcal{A} = \mathcal{A}$. Hence, the α-embedding supplies a useful representation of the tangent bundle of M_+. In fact, by considering the subspace of $\mathcal{A}$ given by:

$$\mathcal{A}^{(\alpha)} := \left\{A \in \mathcal{A} \,\Big|\, \mathrm{Tr}\left(\rho^{\frac{1-\alpha}{2}} A\right) = 0\right\}, \qquad \rho \in \mathrm{M}_+, \tag{27}$$

we can then define the isomorphism:

$$(l_\alpha)_{*(\rho)} : \mathrm{T}_\rho \mathrm{M}_+ \to \mathcal{A}^{(\alpha)}, \quad (l_\alpha)_{*(\rho)}(X) := (l_\alpha \circ \gamma)'(0), \tag{28}$$

where $\gamma : (-\varepsilon, \varepsilon) \to \mathrm{M}_+$ is a curve such that $\gamma'(0) = X$. This isomorphism provides the α-representation of the tangent space $\mathrm{T}_\rho \mathrm{M}_+$ [16]. In particular, if $\{\xi^1, \ldots, \xi^n\}$ is a coordinate system for M_+, then the α-representation of the basis $\left\{\dfrac{\partial}{\partial \xi^1}, \ldots, \dfrac{\partial}{\partial \xi^n}\right\}$ of $\mathrm{T}_\rho \mathrm{M}_+$ is $\left\{\dfrac{\partial l_\alpha}{\partial \xi^1}, \ldots, \dfrac{\partial l_\alpha}{\partial \xi^n}\right\}$, where $n = N^2$. Finally, for any vector field $X \in \mathcal{T}(\mathrm{M}_+)$, we have that its α-representation is defined by:

$$(X)^{(\alpha)}(\rho) := (l_\alpha)_{*(\rho)} X_\rho. \tag{29}$$

From Equation (26), we may observe that the $l_\alpha(\rho) \in \mathrm{M}_+$ for all $\rho \in \mathrm{M}_+$. Now, since $\mathrm{T}_\rho \mathrm{M}_+ = \mathcal{A}$, we can simply define the α-parallel transport on M_+ by:

$$\Pi_{\rho_1,\rho_2}^{(\alpha)} : \mathrm{T}_{\rho_1}\mathrm{M}_+ \to \mathrm{T}_{\rho_2}\mathrm{M}_+, \quad \Pi_{\rho_1,\rho_2}^{(\alpha)}(X) := (l_\alpha)_{*(\rho_2)}^{-1}\left((l_\alpha)_{*(\rho_1)}(X)\right), \quad \forall\, \rho_1, \rho_2 \in \mathrm{M}_+. \tag{30}$$

Therefore, we find that the α-representation of the covariant derivative ∇^α associated with $\Pi^{(\alpha)}$ is:

$$\left(\nabla_{\partial_i}^\alpha \partial_j\right)^{(\alpha)} = \frac{\partial^2 l_\alpha(\rho)}{\partial \xi^i \partial \xi^j}, \tag{31}$$

where $\{\xi^1, \ldots, \xi^n\}$ is an arbitrary coordinate system for M_+ and $n = N^2$ is the topological dimension of the tangent space $T_\rho M_+ = \mathcal{A}$. In order to show that M_+ is ∇^α-flat, consider a basis $\{A_1, \ldots, A_n\}$ of $\mathcal{A}$. Then, there exist real numbers $\{\theta^1, \ldots, \theta^n\}$ such that we can write $\rho^{\frac{1-\alpha}{2}} \in \mathcal{A}$ as follows:

$$\frac{2}{1-\alpha} \rho^{\frac{1-\alpha}{2}} = \theta^1 A_1 + \ldots + \theta^n A_n,$$

for every $\rho \in M_+$. We can see from Equation (31) that:

$$\left(\nabla^\alpha_{\partial_i} \frac{\partial}{\partial \theta^j} \right)^{(\alpha)} = \frac{\partial^2 l_\alpha(\rho)}{\partial \theta^i \partial \theta^j} = \frac{\partial A_j}{\partial \theta^i} = 0.$$

This proves that $\{\theta^1, \ldots, \theta^n\}$ is a ∇^α-affine coordinate system for M_+, and then, M_+ is ∇^α-flat [16]. As a consequence, for any couple of points $\rho_1, \rho_2 \in M_+$, we can write the α-geodesic from ρ_1 to ρ_2 in these ∇^α-affine coordinates as follows:

$$\gamma_\alpha(t) = \frac{2}{1-\alpha} \left(t \rho_2^{\frac{1-\alpha}{2}} + (1-t) \rho_1^{\frac{1-\alpha}{2}} \right). \tag{32}$$

In order to perform the computation of the integral (6), we need to calculate the norm $\|\dot{\gamma}_\alpha(t)\|^2_{\gamma_\alpha(t)}$. To do this, we have to specify a suitable metric on the tangent space TM_+. As discussed above, we select the metric induced by the WYD inner product because this metric is the only one (up to a scalar multiple) that makes the quantum flat α-connections dual in the sense of Equation (1) [29]. For any X, Y vector fields on M_+, this metric turns out to be defined by:

$$g^\alpha_\rho(X, Y) = \mathrm{Tr}\left(X^{(\alpha)} Y^{(-\alpha)} \right), \quad X, Y \in T_\rho M_+, \tag{33}$$

where $X^{(\alpha)}$ denotes the α-representation of the tangent vector X, and it is given by Equation (29) [35]. It is worth noticing that the limit $\lim_{\alpha \to \pm 1} g^\alpha$ gives the metric induced by the BKM inner product on the manifold of density operators [12].

In order to write the WYD metric with respect to the ∇^α-affine coordinate system $\{\theta^1, \ldots, \theta^n\}$, we may observe that:

$$\frac{\partial l_\alpha(\rho)}{\partial \theta^i} = A_i, \tag{34}$$

where A_i is a vector of the basis $\{A_1, \ldots, A_n\} \subset \mathcal{A}$. This implies that, with this coordinate system, the (α)-representation $X^{(\alpha)}$ of a vector field $X \in \mathcal{T}(M_+)$ is the vector field X itself. In addition, we also have that:

$$l_{-\alpha} = \frac{1+\alpha}{2} \rho^{\frac{1+\alpha}{2}} = \left(\frac{2}{1+\alpha} \right) \left(\frac{1-\alpha}{2} \right)^{\frac{1+\alpha}{1-\alpha}} l_\alpha^{\frac{1+\alpha}{1-\alpha}}. \tag{35}$$

Therefore, in the α-affine coordinate system $\{\theta^1, \ldots, \theta^n\}$, the components of the metric tensor (33) are given by:

$$\begin{aligned}
g^\alpha_{ij}(\theta) &= \mathrm{Tr}\left(\frac{\partial l_\alpha}{\partial \theta^i} \frac{\partial l_{-\alpha}}{\partial \theta^j} \right) \\
&= \frac{2}{1-\alpha} \left(\frac{1-\alpha}{2} \right)^{\frac{1+\alpha}{1-\alpha}} \mathrm{Tr}\left(\frac{\partial l_\alpha}{\partial \theta^i} l_\alpha^{\frac{2\alpha}{1-\alpha}} \frac{\partial l_\alpha}{\partial \theta^j} \right).
\end{aligned}$$

The quantum dually flat structure $(g^\alpha, \nabla^\alpha, \nabla^{-\alpha})$ for the manifold of positive definite Hermitian operators (or quantum operators) so far described can be also obtained through Equations (3) and (4) when the following divergence is considered,

$$D^{(\alpha)}(\rho_1, \rho_2) := \frac{4}{1-\alpha^2} \mathrm{Tr}\left(\frac{1-\alpha}{2}\rho_1 + \frac{1+\alpha}{2}\rho_2 - \rho_1^{\frac{1-\alpha}{2}}\rho_2^{\frac{1+\alpha}{2}} \right), \quad \rho_1, \rho_2 \in M_+ \tag{36}$$

This function is called the quantum α-divergence, and it has been introduced on M_+ as the generalization of the α-divergence (18) of the positive measures [36]. Carrying on this line of reasoning, we can introduce a q-divergence on the manifold of positive Hermitian operators, as well. In analogy with the classical case, we can set $\alpha = 1 - 2q$ in the argument of the trace Tr in Equation (36), and then, we can write the following expression,

$$D_q(\rho_1, \rho_2) := \frac{1}{1-q} \mathrm{Tr}\left[q\rho_1 + (1-q)\rho_2 - \rho_1^q \rho_2^{1-q} \right], \quad \rho_1, \rho_2 \in M_+, \tag{37}$$

which corresponds to the quantum α-divergence up to the same scalar factor as in the classical case. It is worth noting that this function is different from the extensions of the Tsallis relative entropy to the positive operators in the literature. For example, in [37], the quantum q-divergence was introduced in the following way,

$$\widetilde{D}_q(\rho_1, \rho_2) := \frac{\mathrm{Tr}[\rho_1] - \mathrm{Tr}\left[\rho_1^q \rho_2^{1-q}\right]}{1-q}, \quad \rho_1, \rho_2 \in M_+, \tag{38}$$

for all $q \in [0,1)$. However, both functions, D_q and $\widetilde{D}_q$, reduce to the Tsallis relative entropy when restricted to the manifold S of density operators,

$$D_q(\rho_1, \rho_2) := \frac{1 - \mathrm{Tr}\left[\rho_1^q \rho_2^{1-q}\right]}{1-q}, \quad \rho_1, \rho_2 \in S.$$

Analogously, we can restrict the quantum α-divergence (36) to the set of density operators, and since $\mathrm{Tr}\,\rho_1 = \mathrm{Tr}\,\rho_2 = 1$, we can easily verify that (36) becomes:

$$D^{(\alpha)}(\rho_1, \rho_2) = \frac{4}{1-\alpha^2}\left(1 - \mathrm{Tr}\left[\rho_1^{\frac{1-\alpha}{2}}\rho_2^{\frac{1+\alpha}{2}} \right] \right), \quad \rho_1, \rho_2 \in S. \tag{39}$$

Again, we can see that, on the manifold of density operators, the quantum α-divergence coincides (up to a scalar factor) with the Tsallis relative entropy when we set $\alpha = 1 - 2q$. The quantum α-divergence (39) was introduced and studied by Hasegawa in [32], and it turns out to be a continuous function of the parameter α. In particular, we have that:

$$\lim_{\alpha \to -1} D^{(\alpha)}(\rho_1, \rho_2) = \mathrm{Tr}\left[\rho_1 \left(\log \rho_1 - \log \rho_2 \right) \right],$$

that is the quantum relative entropy $\mathcal{D}(\rho_1, \rho_2)$ as previously given in Equation (11). Furthermore, the limit $\alpha \to +1$ gives $\lim_{\alpha \to +1} D^{(\alpha)}(\rho_1, \rho_2) = \mathcal{D}(\rho_2, \rho_1)$.

At this point, we have in our hands all the ingredients necessary to compute the canonical divergence defined in (6) between ρ_1 and ρ_2 on the manifold M_+ of positive definite Hermitian operators,

$$\mathcal{D}(\rho_1, \rho_2) := \int_0^1 t\, \|\dot{\gamma}_\alpha(t)\|_{\gamma_\alpha(t)}^2 \, \mathrm{d}t, \tag{40}$$

where $\gamma_\alpha(t)$ is the ∇^α-geodesic from ρ_1 to ρ_2. In the ∇^α-affine coordinates $\{\theta^1,\ldots,\theta^N\}$, this reads as in Equation (32). Here, the norm $\|\cdot\|^2$ is induced by the WYD metric given in Equation (33). Hence, in this case, the canonical divergence (40) can be written as:

$$\mathcal{D}(\rho_1,\rho_2) = \int_0^1 t\,\mathrm{Tr}\left((\dot\gamma_\alpha(t))^{(\alpha)}\,(\dot\gamma_\alpha(t))^{(-\alpha)}\right)\,\mathrm{d}t\,, \tag{41}$$

where $(\dot\gamma_\alpha(t))^{(\alpha)}$ and $(\dot\gamma_\alpha(t))^{(-\alpha)}$ denote the (α) and the $(-\alpha)$ representations of $\dot\gamma(t)$. In order to compute the α representations of $\dot\gamma(t)$, we consider Equation (29) together with Equations (34) and (35). This yields:

$$\begin{aligned}
(\dot\gamma_\alpha(t))^{(\alpha)} &= \dot\gamma_\alpha(t) \\
(\dot\gamma_\alpha(t))^{(-\alpha)} &= \frac{2}{1-\alpha}\left(\frac{1-\alpha}{2}\right)^{\frac{1+\alpha}{1-\alpha}}\frac{1+\alpha}{1-\alpha}\,(\gamma_\alpha(t))^{\frac{2\alpha}{1-\alpha}}\,\dot\gamma_\alpha(t) \\
&= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}(\gamma_\alpha(t))^{\frac{2\alpha}{1-\alpha}}\,\dot\gamma_\alpha(t)\,.
\end{aligned}$$

We can plug these expressions in Equation (41). Hence, by performing an integration by parts, we get:

$$\begin{aligned}
\mathcal{D}(\rho_1,\rho_2) &= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}\int_0^1 t\,\mathrm{Tr}\left(\dot\gamma_\alpha(t)\,(\gamma_\alpha(t))^{\frac{2\alpha}{1-\alpha}}\,\dot\gamma_\alpha(t)\right)\,\mathrm{d}t \\
&= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}\mathrm{Tr}\left(\int_0^1 t\,\dot\gamma_\alpha(t)\,(\gamma_\alpha(t))^{\frac{2\alpha}{1-\alpha}}\,\dot\gamma_\alpha(t)\,\mathrm{d}t\right) \\
&= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}\mathrm{Tr}\left(\left[t\dot\gamma_\alpha(t)\frac{1-\alpha}{\alpha+1}(\gamma_\alpha(t))^{\frac{1+\alpha}{1-\alpha}}\right]_0^1 - \frac{1-\alpha}{\alpha+1}\int_0^1 \dot\gamma_\alpha(t)\,(\gamma_\alpha(t))^{\frac{1+\alpha}{1-\alpha}}\,\mathrm{d}t\right) \\
&= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}\mathrm{Tr}\left(\left[t\dot\gamma_\alpha(t)\frac{1-\alpha}{1+\alpha}(\gamma_\alpha(t))^{\frac{1+\alpha}{1-\alpha}}\right]_0^1 - \frac{1-\alpha}{1+\alpha}\frac{1-\alpha}{2}\left[(\gamma_\alpha(t))^{\frac{2}{1-\alpha}}\right]_0^1\right)\,.
\end{aligned}$$

At this point, we can use Equation (32) to obtain:

$$\begin{aligned}
\mathcal{D}(\rho_1,\rho_2) &= \left(\frac{1-\alpha}{2}\right)^{\frac{2\alpha}{1-\alpha}}\mathrm{Tr}\left(\frac{2}{1-\alpha}\left(\rho_2^{\frac{1-\alpha}{2}} - \rho_1^{\frac{1-\alpha}{2}}\right)\frac{1-\alpha}{1+\alpha}\left(\frac{2}{1-\alpha}\right)^{\frac{1+\alpha}{1-\alpha}}\rho_2^{\frac{1-\alpha}{2}\frac{1+\alpha}{1-\alpha}}\right. \\
&\qquad\left. -\frac{1-\alpha}{1+\alpha}\frac{1-\alpha}{2}\left(\frac{2}{1-\alpha}\right)^{\frac{2}{1-\alpha}}\left(\rho_2^{\frac{1-\alpha}{2}\frac{2}{1-\alpha}} - \rho_1^{\frac{1-\alpha}{2}\frac{2}{1-\alpha}}\right)\right) \\
&= \frac{4}{1-\alpha^2}\mathrm{Tr}\left(\left(\rho_2^{\frac{1-\alpha}{2}} - \rho_1^{\frac{1-\alpha}{2}}\right)\rho_2^{\frac{1+\alpha}{2}}\right) - \frac{1-\alpha}{1+\alpha}\frac{2}{1-\alpha}\mathrm{Tr}\left(\rho_2 - \rho_1\right) \\
&= \frac{4}{1-\alpha^2}\mathrm{Tr}\rho_2 - \frac{2}{1+\alpha}\mathrm{Tr}\rho_2 + \frac{2}{1+\alpha}\mathrm{Tr}\rho_1 - \frac{4}{1-\alpha^2}\mathrm{Tr}\left(\rho_1^{\frac{1-\alpha}{2}}\rho_2^{\frac{1+\alpha}{2}}\right) \\
&= \frac{2}{1+\alpha}\mathrm{Tr}\rho_1 + \frac{2}{1-\alpha}\mathrm{Tr}\rho_2 - \frac{4}{1-\alpha^2}\mathrm{Tr}\left(\rho_1^{\frac{1-\alpha}{2}}\rho_2^{\frac{1+\alpha}{2}}\right)\,.
\end{aligned}$$

Finally, we can conclude that:

$$\mathcal{D}(\rho_1,\rho_2) = \frac{4}{1-\alpha^2}\mathrm{Tr}\left(\frac{1-\alpha}{2}\rho_1 + \frac{1+\alpha}{2}\rho_2 - \rho_1^{\frac{1-\alpha}{2}}\rho_2^{\frac{1+\alpha}{2}}\right)\,,\quad \rho_1,\rho_2 \in \mathrm{M}_+\,, \tag{42}$$

which corresponds to the α-canonical divergence on the manifold of positive definite matrices.

5. Conclusions

The present article is a follow-up of the recently published paper [24]. In the latter one, the authors showed that the canonical divergence defined in Equation (6) provides a powerful tool for unifying classical and quantum information geometry. In particular, such a divergence was proven to coincide with the Kullback-Leibler divergence on the simplex of probability distributions and with the quantum relative entropy on the manifold of quantum states. The effectiveness of the Kullback-Leibler divergence was ascribed in the context of complex systems for quantifying how much a probability measure deviates from the non-interacting states that are modeled by exponential families of probabilities [11]. On the other hand, the quantum relative entropy turned out to be relevant for providing a measure of the many-party correlations of a quantum state from a Gibbs family [13], which in turn was related to the entanglement of quantum systems as defined in [15].

The α-divergence (18) is a generalization of the relative entropy (10). Moreover, the flat α-geometry induced by the α-divergence on the manifold of positive measures constitutes a generalization of the dually flat structure given by the Fisher metric and the mixture and exponential connections. Very remarkably, the α-geometry covers the geometry of q-entropy physics [19]. This bridges a very nice connection between the α-geometry and the generalized statistical mechanics established by Tsallis [17,18]. On the quantum side, the α-divergence (36) was introduced as the generalization of the α-divergence (18) of the positive measures [36]. The quantum α-geometry of the manifold of positive definite matrices is flat, as well, and turns out to be a generalization of the quantum geometry given by the quantum Fisher metric and the quantum mixture and exponential connections induced on the manifold of quantum states by the BKM inner product [16]. In the quantum setting as well, the connection between the α-geometry and the q-entropy physics by Tsallis provides a physical interpretation of the α-divergence. Indeed, a conditional form of the q-entropy was employed for the exact calculation, on some systems, of the separable-entangled separatrix [17].

In this article, we computed the canonical divergence (6) for flat α-connections on the manifold of positive measures, as well as on the manifold of positive definite Hermitian operators. Therefore, we proved that the divergence introduced by Ay and Amari in [21] reduces to the classical α-divergence (18) and to the quantum α-divergence (36). Actually, the equivalence between the canonical divergence (6) and the classical α-divergence was primarily shown in [21]. There, the derivation of $D^{(\alpha)}$ was grounded on the idea of a squared distance function associated to the α-connections through the vector field $X_t(p)$ given in Equation (5). However, the α-divergence $D^{(\alpha)}$ does not share all the properties of a squared distance function, unless $\alpha = 0$. For this reason, in the present paper, we obtained the equivalence between $\mathcal{D}$ and $D^{(\alpha)}$ in a different way, by considering the α-divergence as a function of a more general structure.

The present paper is conceived within a project that aims to characterize a general canonical divergence for a given dualistic structure (g, ∇, ∇^*) of a smooth manifold M. This project started with the work by Ay and Amari [21] and later developed in [3], where the concept of the "canonical divergence" was clearly depicted. A considerable effort towards the definition of a general canonical divergence was put forward in [38], where a divergence function was introduced through an extensive investigation of the geodesic geometry of the dualistic structure (g, ∇, ∇^*). Actually, this latter divergence turns out to coincide with the divergence (6) when the dualistic structure (g, ∇, ∇^*) is flat. Further work around this topic was presented in [39], where the very recent divergence was compared with other divergence functions present in the literature. It is commonly accepted that the α-divergence on the simplex is obtained by restricting (18) to the set of normalized positive measures (see [1,6] for more details). Then, the α-divergence of probability distributions in the simplex reads as follows,

$$D^{(\alpha)}(p,q) = \sum_{i=1}^{n} \left(1 - \frac{4}{1-\alpha^2} p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}} \right) . \tag{43}$$

One can easily verify that for $\alpha = 0$, this divergence is closely related to the Hellinger distance, which is the distance in the ambient space $\mathcal{M}_+$ [3]. However, when $\alpha = 0$, the α-connection is the Levi–Civita connection of the Fisher metric. Therefore, in this case, the canonical divergence has to be one half the square of the Fisher distance, as discussed in Section 2, which is different from $D^{(0)}$ [3]. It would be interesting to evaluate the very recent divergence introduced in [38] on the simplex of probability distributions, which is not α-flat, and to compare it with the α-divergence (43).

On the quantum side, an intriguing work within the above-mentioned project is to consider that very recent divergence function, introduced in [38], on the manifold of pure quantum states, where a dually flat structure does not exist [23]. This will constitute the object of study of a forthcoming investigation.

Author Contributions: The authors have equally contributed to the manuscript. They all have read and approved its final version.

References

1. Amari, S.I.; Nagaoka, H. *Methods of Information Geometry*; Oxford University Press: Oxford, UK, 2000.
2. Felice, D.; Cafaro, C.; Mancini, S. Information Geometric Methods for Complexity. *Chaos* **2018**, *28*, 032101. [CrossRef] [PubMed]
3. Ay, N.; Jost, J.; Van Le, H.; Schwachhöfer, L. *Information Geometry*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017.
4. Lauritzen, S.L. *Chapter 4: Statistical Manifolds*; Differential Geometry in Statistical Inference; Institute of Mathematical Statistics: Hayward, CA, USA, 1987; pp. 163–216.
5. Amari, S.-I. Differential Geometry of Curved Exponential Families-Curvatures and Information Loss. *Ann. Statist.* **1982**, *10*, 357–387. [CrossRef]
6. Amari, S.-I. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
7. Eguchi, S. A differential geometric approach to statistical inference on the basis of contrast functions. *Hiroshima Math. J.* **1985**, *15*, 341–391. [CrossRef]
8. Eguchi, S. Geometry of Minimum Contrast. *Hiroshima Math. J.* **1992**, *22*, 631–647. [CrossRef]
9. Fujiwara, A.; Amari, S.-I. Gradient systems in view of information geometry. *Physica D* **1995**, *80*, 317–327. [CrossRef]
10. Nakamura, Y. Completely integrable gradient systems on the manifolds of Gaussian and multinomial distributions. *Jpn. J. Ind. Appl. Math.* **1993**, *10*, 179. [CrossRef]
11. Ay, N. Information geometry on complexity and stochastic interaction. *Entropy* **2015**, *17*, 2432–2458. [CrossRef]
12. Nagaoka, H. Differential Geometrical Aspects of Quantum State Estimation and Relative Entropy. In *Quantum Communications and Measurement*; Belavkin, V.P., Hirota, O., Hudson, R.L. Eds.; Springer: Boston, MA, USA, 1995.
13. Weis, S.; Knauf, A.; Ay, N.; Zhao, M.J. Maximizing the divergence from a hierarchical model of quantum states. *Open Syst. Inf. Dyn.* **2015**, *22*, 1550006. [CrossRef]
14. Niekamp, S.; Galla, T.; Kleinmann, M.; Gühne, O. Computing complexity measures for quantum states based on exponential families. *J. Phys. A Math. Theor.* **2013**, *46*, 125301. [CrossRef]
15. Vedral, V.; Plenio, M.B.; Rippin, M.A.; Knight, P.L. Quantifying entanglement. *Phys. Rev. Lett.* **1997**, *78*, 2275–2279. [CrossRef]
16. Grasselli, M.H. Duality, monotonicity and the Wigner-Yanase-Dyson metrics. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **2004**, *7*, 215–232. [CrossRef]
17. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
18. Tsallis, C. Beyond Boltzmann–Gibbs–Shannon in Physics and Elsewhere. *Entropy* **2019**, *21*, 696. [CrossRef]

19. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **2007**, *370*, 184–193. [CrossRef]

20. Abe, S.; Rajagopal, A.K. Nonadditive conditional entropy and its significance for local realism. *Physica A* **2001**, *249*, 157. [CrossRef]

21. Ay, N.; Amari, S.-I. A Novel Approach to Canonical Divergences within Information Geometry. *Entropy* **2015**, *7*, 8111–8129. [CrossRef]

22. Matumoto, T. Any statistical manifold has a contrast function—on the C^3-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J.* **1993**, *23*, 327–337. [CrossRef]

23. Ay, N.; Tuschmann, W. Duality versus dual flatness in quantum information geometry. *J. Math. Phys.* **2003**, *44*, 1512–1518. [CrossRef]

24. Felice, D.; Mancini, S.; Ay, N. Canonical divergence for measuring classical and quantum complexity. *Entropy* **2019**, *21*, 435. [CrossRef]

25. Čencov, N.N. *Statistical Decision Rules and Optimal Inference*; American Mathematical Society: Providence, RI, USA, 1982.

26. Petz, D.; Sudàr, C. Geometries of Quantum States. J. Math. Phys. **1996**, *37*, 2662. [CrossRef]

27. Holevo, A.S. *Probabilistic and Statistical Aspects of Quantum Theory*; North-Holland: Amsterdam, The Netherlands, 1982.

28. Petz, D. Monotone metrics on matrix spaces. *Linear Algebra Appl.* **1996**, *244*, 81–96. [CrossRef]

29. Grasselli, M.R.; Streater, R.F. On the uniqueness of the Chentsov metric in quantum information geometry. *Inn. Dimens. Anal. Quantum Probab. Relat. Top.* **2001**, *4*, 173–182. [CrossRef]

30. Hasegawa, H. Exponential and mixture families in quantum statistics: dual structure and unbiased parameter estimation. *Rep. Math. Phys.* **1997**, *39*, 49–68. [CrossRef]

31. Jenčovà, A. Geometry of quantum states: dual connections and divergence functions. *Rep. Math. Phys.* **2001**, *47*, 121–138. [CrossRef]

32. Hasegawa, H. α-divergence of the noncommutative information geometry. *Rep. Math. Phys.* **1993**, *33*, 87–93. [CrossRef]

33. Martins, A.F.T.; Smith, N.A.; Xing, E.P.; Aguiar, P.M.Q.; Figueiredo, M.A.T. Nonextensive Information Theoretic Kernels on Measures. *J. Mach. Learn. Res.* **2009**, *10*, 935–975.

34. Jenčovà, A. Flat connections and Wigner-Yanase-Dyson metrics. *Rep. Math. Phys.* **2003**, *52*, 331–351. [CrossRef]

35. Jenčovà, A. Quantum information geometry and standard purification. *J. Math. Phys.* **2002**, *43*, 2187. [CrossRef]

36. Amari, S.-I.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci. Tech. Sci.* **2010**, *58*, 183–195. [CrossRef]

37. Furuichi, S.; Yanagi, K.; Kuriyama, K. Fundamental properties of Tsallis relative entropy. *J. Math. Phys.* **2004**, *45*, 4868–4877. [CrossRef]

38. Felice, D.; Ay, N. Towards a canonical divergence within Information Geometry. *arXiv* **2018**, arXiv:1806.11363.

39. Felice, D.; Ay, N. Divergence functions in Information Geometry. *arXiv* **2019**, arXiv:1903.02379.