Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Towards a Canonical Divergence within Information Geometry

by

Domenico Felice and Nihat Ay

Preprint no.: 43 2018

Towards a Canonical Divergence within Information Geometry

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

[1] *Max Planck Institute for Mathematics in the Sciences*
Inselstrasse 22–04103 Leipzig, Germany
[2] *Max Planck Institute for Mathematics in the Sciences*
Inselstrasse 22–04103 Leipzig, Germany
Santa Fe Institute, Santa Fe, NM 87501, USA

In Riemannian Geometry geodesics are integral curves of the gradient of Riemannian distance. We extend this classical result to the framework of Information Geometry. In particular, we prove that the rays of level-sets defined by a pseudo-distance are generated by the sum of two tangent vectors. By relying on these vectors, we propose a novel definition of divergence and its dual function. We prove that the new divergence defines a dual structure $(\mathrm{g}, \nabla, \nabla^*)$ of a statistical manifold M. Additionally, we show that this divergence reduces to the canonical divergence proposed by Ay and Amari in the case of: (a) self-duality, (b) dual flatness, (c) statistical geometric analogue of the concept of symmetric spaces in Riemannian Geometry. The case (c) leads to a further comparison of the novel divergence with the one introduced by Henmi and Kobayashi.

PACS numbers: Classical differential geometry (02.40.Hw), Riemannian geometries (02.40.Ky), Inverse problems (02.30.Zz).

I. INTRODUCTION

The Inverse Problem within Information Geometry [4] concerns the search for a divergence function $\mathcal{D}$ which generates a given statistical structure $(\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. A significant attempt to solve this problem has been put forth in [5]. Here, the authors defined a divergence function relying on the inverse exponential map.

A *statistical manifold* is a C^∞ manifold M endowed with a *dual structure* $(\mathrm{g}, \nabla, \nabla^*)$ such that

$$X\ \mathrm{g}\,(Y, Z) = \mathrm{g}\,(Y, \nabla^*_X Z) + \mathrm{g}\,(\nabla_X Y, Z)\,,\ \forall\ X, Y, Z \in \mathcal{T}(\mathrm{M}), \tag{1}$$

where $\mathcal{T}(\mathrm{M})$ denotes the space of vector fields on M, namely C^∞ sections $X : \mathrm{M} \to \mathrm{TM}$, $X_p \in \mathrm{T}_p\mathrm{M}$. The dual connections ∇ and ∇^* are both torsion free. The notion of statistical manifold, introduced

*Electronic address: domenico.felice@unicam.it
†Electronic address: nay@mis.mpg.de

by Lauritzen [15], is usually referred to the triple $(\mathrm{M}, \mathrm{g}, T)$, where $T(X,Y,Z) = \mathrm{g}\left(\nabla^*_X Y - \nabla_X Y, Z\right)$ is a 3-symmetric tensor. However, when ∇ and ∇^* are both torsion free connections, then the structures $(\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ and $(\mathrm{M}, \mathrm{g}, T)$ are equivalent [6].

A distance-like function $\mathcal{D} : \mathrm{M} \times \mathrm{M} \to \mathbb{R}$ satisfies the following conditions

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

The function $\mathcal{D}$ is called a *divergence* or *contrast function* on M [4] if the matrix

$$\mathrm{g}_{ij}(p) = -\, \partial_i \partial'_j \mathcal{D}(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q} = \partial'_i \partial'_j \mathcal{D}(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q} \tag{3}$$

is strictly positive definite everywhere on M. Here,

$$\partial_i = \frac{\partial}{\partial \xi^i_p} \quad \text{and} \quad \partial'_i = \frac{\partial}{\partial \xi^i_q}$$

and $\{\boldsymbol{\xi}_p := (\xi^1_p, \ldots, \xi^n_p)\}$ and $\{\boldsymbol{\xi}_q := (\xi^1_q, \ldots, \xi^n_q)\}$ are local coordinate systems of p and q, respectively. Moreover, $\{\boldsymbol{\partial}_p = (\partial_1, \ldots, \partial_n)\}$ and $\{\boldsymbol{\partial}'_q = (\partial'_1, \ldots, \partial'_n)\}$ are the correspondent local frames on $\mathrm{T}_p\mathrm{M}$ and $\mathrm{T}_q\mathrm{M}$, respectively.

Conversely, given a dual structure $(\mathrm{g}, \nabla, \nabla^*)$ on M, the distance-like function (2) is compatible with $(\mathrm{g}, \nabla, \nabla^*)$ if g is obtained by (3) and furthermore the following holds [3]:

$$\Gamma_{ijk}(p) = -\, \partial_i \partial_j \partial'_k \mathcal{D}(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q}, \qquad \Gamma^*_{ijk}(p) = -\, \partial'_i \partial'_j \partial_k \mathcal{D}(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q}\ , \tag{4}$$

where $\Gamma_{ijk} = \mathrm{g}\left(\nabla_{\partial_i}\partial_j, \partial_k\right)$, $\Gamma^*_{ijk} = \mathrm{g}\left(\nabla^*_{\partial_i}\partial_j, \partial_k\right)$ are the symbols of the dual connections ∇ and ∇^*, respectively. In this article, we address our investigation to the latter issue, namely we will try to figure out a canonical divergence that recovers the dual structure of a given statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$.

Matumoto [19] showed that a divergence exists for any such statistical manifold. However, it is not unique and there are infinitely many divergences that give the same dual structure. When a manifold is dually flat, a canonical divergence was introduced by Amari and Nagaoka [4], which is a Bregman divergence. Extensions of the canonical divergence within conformal geometry have been analysed by Kurose [14] and Matsuzoe [18]. The canonical divergence has relevant properties concerning the generalized Pythagorean theorem and the geodesic projection theorem [1]. For this reason the issue of finding a general canonical divergence for a given statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is of uppermost importance. In [5] the definition of a canonical divergence for a general $\mathcal{S}$ is given by using the geodesic integration of the inverse exponential map. This one is interpreted as a *difference* vector that translates q to p for all q, p suitably close in M.

To be more precise, the inverse exponential map provides a generalization to M of the notion of difference vector of the linear vector space. In detail, let $p, q \in \mathbb{R}^n$, the difference between p and q is given by the vector $p - q$ pointing to p (see side (A) of Fig. I). Then, the difference between p and q in M is supplied by the exponential map of the connection ∇ [Appendix A]. In particular, assuming that $p \in \mathrm{U}_q$ and $\mathrm{U}_q \subset \mathrm{M}$ is a ∇-geodesic neighborhood of q, the difference vector from q to p is defined as (see (B) of Fig. I)

$$\mathrm{X}_q(p) := \mathrm{X}(q,p) := \exp_q^{-1}(p) = \dot{\gamma}_{q,p}(0) \ , \tag{5}$$

where $\gamma_{q,p}$ is the ∇-geodesic from q to p laying in U_q. Clearly, by fixing $p \in \mathrm{M}$ and letting q vary in M, we obtain a vector field $\mathrm{X}(\cdot, p)$ whenever a ∇-geodesic from q to p exists. From here on, we equally use both the notations, $\mathrm{X}(q,p)$ and $\mathrm{X}_q(p)$, for representing the difference vector from q to p.

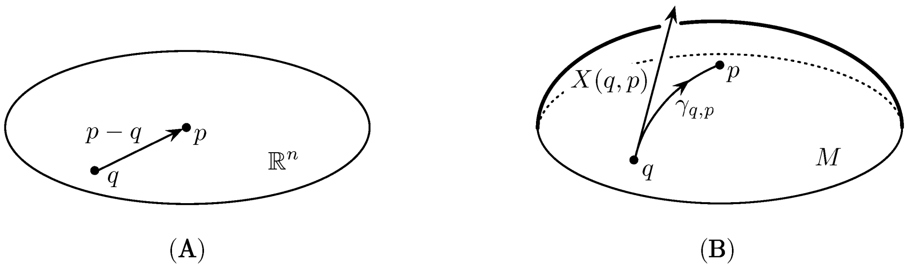

FIG. 1: On the left, (A) illustrates the difference vector $p - q$ in the linear vector space $\mathbb{R}^n$; whereas, in (B) we can see the difference vector $\mathrm{X}(q,p) = \dot{\gamma}_{q,p}(0)$ in M as the inverse of the exponential map at q (This Figure comes from [5]).

Therefore, the divergence proposed by Ay and Amari in [5] is defined as the path integral

$$D_\gamma(p,q) := \int_0^1 \langle \mathrm{X}_t(p), \dot{\gamma}(t) \rangle_{\gamma(t)} \ dt \ , \tag{6}$$

where γ is the ∇-geodesic from q to p and $\langle \cdot, \cdot \rangle_{\gamma(t)}$ denotes the inner product with respect to g evaluated at $\gamma(t)$. In Eq. (6), $\mathrm{X}_t(p)$ is the vector field along $\gamma(t)$ given by Eq. (5) as follows,

$$\mathrm{X}_t(p) = \mathrm{X}(\gamma(t), p) = \exp_{\gamma(t)}^{-1}(p) \ . \tag{7}$$

After elementary computation Eq. (6) reduces to [5],

$$D_\gamma(p,q) = \int_0^1 \ t \|\dot{\gamma}_{p,q}(t)\|^2 \ dt \ , \tag{8}$$

where $\gamma_{p,q}(t)$ is the ∇-geodesic from p to q. If we consider definition (6) for general path γ then $D_\gamma(p,q)$ will be depending on γ. On the contrary, if the vector field $\mathrm{X}_t(p)$ is integrable, then $D_\gamma(p,q) =: D(p,q)$ turns out to be independent of the path from q to p.

The dual divergence of D_γ has been defined in terms of the inverse exponential map with respect to the ∇^*-connection. It turned out to be closely related to the divergence of the article [11]. Here the authors applied the Hook's law to a "∇^* spring" and defined the divergence as the physical work that is necessary to move a unit mass from q to p along the ∇-geodesic γ connecting them against the force field described by the inverse exponential map of ∇^*.

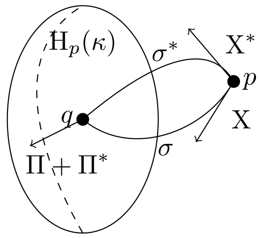

FIG. 2: Π is the parallel transport with respect to ∇ of $\mathrm{X}(\equiv \dot\sigma(0))$ along σ^*, while Π^* is the parallel transport with respect to ∇^* of $\mathrm{X}^*(\equiv \dot\sigma^*(0))$ along σ. The sum $\Pi + \Pi^*$ is orthogonal to the level-hypersurface $\mathrm{H}_p(\kappa)$ of constant pseudo-squared-norm $r_p(q)$.

In this manuscript we propose a novel definition of divergence by combining both, the approach carried out in [5] and the one in [11]. We significantly investigate the intrinsic structure of the dual geometry of a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. Slightly moving from the definition (5) of the difference vector $\mathrm{X}(q,p)$, we define a vector at q in the following way. Consider the unique ∇-geodesic σ from p to q and

$$\mathrm{X}_p(q) := \exp_p^{-1}(q) = \dot\sigma(0) \in \mathrm{T}_p\mathrm{M} \ .$$

We then parallel translate [Appendix A] it with respect to the ∇-connection along the ∇^*-geodesic σ^* from p to q (see Fig. I), and obtain

$$\Pi_q(p) := \mathrm{P}_{\sigma^*}\mathrm{X}_p(q) \in \mathrm{T}_q\mathrm{M} \ . \tag{9}$$

(Note that $\Pi_q(p)$ corresponds to minus a difference vector). At this point, fixed $p \in \mathrm{M}$ and let q be varied in M, we can have a vector field $\Pi_q(p) \in \mathcal{T}(\mathrm{M})$ whenever ∇ and ∇^* geodesics from p to q exist. Analogously, we define the dual vector of $\Pi_q(p)$ as the parallel transport with respect to ∇^* of $\dot\sigma^*(0)$,

$$\Pi_q^*(p) := \mathrm{P}_\sigma^*\mathrm{X}_p^*(q), \tag{10}$$

where

$$\mathrm{X}_p^*(q) := \overset{*}{\exp}{}_p^{-1}(q) = \dot\sigma^*(0) \ . \tag{11}$$

A relevant result of this article shows that the sum $\Pi_q + \Pi_q^*$ generates the rays of level-hypersurface defined in terms of the pseudo-squared-norm

$$r_p(q) := \langle \exp_p^{-1}(q), \overset{*}{\exp}_p^{-1}(q)\rangle_p \ . \tag{12}$$

In particular, we have the following result

Theorem I.1. *Given a statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ *and* $p \in \mathrm{M}$*, let us assume that there exists a neighborhood* $\mathrm{U}_p \subset \mathrm{M}$ *of* p *such that all* $q \in \mathrm{U}_p$ *can be connected with* p *by* ∇ *and* ∇^* *geodesics. Then we have*

$$\mathrm{grad}_q r_p = \Pi_q(p) + \Pi_q^*(p), \tag{13}$$

where $r_p(q)$ *is defined by Eq.* (12) *.*

The proof of Theorem I.1 is deferred to Section II within several steps.

Clearly, the function $r : \mathrm{M} \times \mathrm{M} \to \mathbb{R}$ defined by $(p, q) \mapsto r_p(q)$ is symmetric in its arguments, namely $r_p(q) = r_q(p)$. This suggests that $r_p(q)$ is not a good candidate to move it forward as the general canonical divergence. To support this claim, the well-known canonical divergence $D[p : q]$ described in [4] in the case of dually flat manifold is non-symmetric in p and q. In this case, a sort of symmetry is recovered in terms of the dual divergence $D^*[p : q]$ as follows,

$$D[q : p] = D^*[p : q] \ . \tag{14}$$

Hence, the sum of these divergences $D[p : q] + D^*[p : q]$ turns out to be symmetric in its arguments. For this reason, we propose to define a novel divergence by relying on Π_q and its dual function by relying on Π_q^* in the way that their sum gives the pseudo-squared-norm r_p. Before doing this, we introduce two functions (*Phi-functions* in this article) and prove that they generate the dual geometry of $\mathcal{S}$. The relevance of these *Phi-functions* is due to the fact that is provides the local decomposition of Π_q and Π_q^* in terms of gradient vectors. Moreover, we are able to decompose the pseuo-squared-norm $r_p(q)$ in terms of *Phi-function* and divergence function. Afterwards, we prove that the novel canonical divergence is closely connected with the *Phi-function.* This result allows us to establish a symmetry property related to the (14), however in a more general context where just the torsion free-ness of connections ∇ and ∇^* is required.

Further investigation is devoted to the connection between the novel divergence and the divergence proposed by Ay and Amari in [5]. We show that our divergence corresponds to the one of Ay and Amari in self-dual manifolds, in dually flat manifolds and in statistical manifolds analogue to the symmetric spaces in Riemannian Geometry. By this correspondence, the novel divergence inherits all the nice properties owned by the divergence of Ay and Amari. In particular, in the case of dually flat manifolds it is the same as the canonical divergence defined in terms of the Bregman divergence [4] of M.

Finally, we carry out the comparison between our approach and the one presented in [11] and prove a close connection between our divergence and the one introduced by Henmi and Kobayashi.

The layout of this article is as follows. In Section II we develop our approach by extending the celebrated Lemma of Gauss to the more general context of Infomration Geometry. Then we prove Theorem I.1. In Section III we define the novel divergence function and prove its consistency with respect to the dual structure. Section IV is devoted to the comparison between the novel divergence and the divergence of Ay and Amari. In addition, we discuss the approach presented in [11]. In Section V we draw some conclusions by outlining the results obtained in this work and discussing possible extensions. Useful tools of statistical differential geometry appear in [Appendix A].

II. GRADIENT VECTOR FIELDS IN STATISTICAL MANIFOLDS

Given a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ we can recover the Levi-Civita connection by averaging the dual connections ∇ and ∇^* [1],

$$\overline{\nabla}_{\mathrm{LC}} = \frac{1}{2}\left(\nabla + \nabla^*\right) \ . \tag{15}$$

In Riemannian Geometry [16], the celebrated Gauss Lemma tells us that the vector field which defines a geodesic line is the gradient of the function

$$d_p(q) = \frac{1}{2}\langle \overline{\exp}_p^{-1}(q), \overline{\exp}_p^{-1}(q)\rangle_p, \tag{16}$$

for every $p, q \in \mathrm{M}$ suitably close each other. Here $\overline{\exp}_p^{-1}(q)$ denotes the inverse of the exponential map with respect to the Levi-Civita connection $\overline{\nabla}_{\mathrm{LC}}$. To be more precise, consider $B_\varepsilon(p) \subset \mathrm{M}$ a geodesic ball centred at p, where ε is a positive number which gives the diffeomorphism of $\overline{\exp}_p : B_\varepsilon(p) \to \overline{\exp}_p\left(B_\varepsilon(p)\right)$ over its image. Then, for every $q \in B_\varepsilon(p) \setminus \{p\}$ the gradient of the function $d_p(q)$ is given by [21]

$$\mathrm{grad}_q \ d_p = \dot{\bar{\sigma}}(1), \tag{17}$$

where $\bar{\sigma}$ is the $\overline{\nabla}_{\mathrm{LC}}$-geodesic from p to q. This means that

$$\mathrm{grad}_q \ d_p = \overline{\mathrm{P}}_{\bar{\sigma}}\overline{\mathrm{X}}_p(q),$$

where $\overline{\mathrm{X}}_p(q) = \overline{\exp}_p^{-1}(q)$ and $\overline{\mathrm{P}}$ denotes the parallel transport with respect to the Levi-Civita connection $\overline{\nabla}_{\mathrm{LC}}$. From a geometrical point of view, the geodesic rays from p are all orthogonal to the *geodesic* sphere

$$\overline{\mathrm{S}}_\kappa = \{q \in \mathrm{M} \ : \ \langle \overline{\exp}_p^{-1}(q), \overline{\exp}_p^{-1}(q)\rangle_p = \kappa\},$$

for $\kappa \leq \varepsilon$. The proof of this classical result in Riemannian Geometry relies on the function

$$t \mapsto \langle \frac{\mathrm{d}}{\mathrm{d}t}\overline{\exp}_p\,(tX_p)\,, \frac{\mathrm{d}}{\mathrm{d}t}\overline{\exp}_p\,(tX_p)\rangle_{\overline{\exp}_p(tX_p)}, \tag{18}$$

which is constant with respect to t for all $X_p \in B_\varepsilon(p)$. As a consequence, every vector field $X_q \in B_\varepsilon(p)$ can be decomposed in the following way,

$$X_q = \lambda(q)\mathrm{grad}_q d_p + W_q, \tag{19}$$

where $\lambda(q)$ is a coefficient depending on q and W_q is tangent to the hypersurface $\overline{\mathrm{S}}_\kappa$.

Remark II.1. *In [2] the authors proposed the function*

$$D[p:q] := \langle \exp_p^{-1}(q), \exp_p^{-1}(q)\rangle_p \tag{20}$$

as the Standard Divergence of the statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$*. Here,* $\exp_p$ *denotes the exponential map with respect to the* ∇*-connection. In contrast to the Levi-Civita connection, the function of Eq.* (18) *now computed by the exponential map of* ∇ *is not constant with respect to* t*. However, this definition turned out to be unsatisfactory because it is unable, at least in general, to recover the dual structure of* $\mathcal{S}$*.*

Given a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$, in the rest of this article we use the following working hypothesis

$$\text{(I)} \qquad \forall\, p \in \mathrm{M}\ \exists\ \mathrm{U}_p \subset \mathrm{M} \text{ such that } \forall q \in \mathrm{U}_p\ \exists\,!\ \nabla - \text{geodesic and } \nabla^* - \text{geodesic connecting } p \text{ and } q\ . \tag{21}$$

In order to prove Theorem I.1 we introduce a pseudo-energy functional as follows. Given $p \in \mathrm{M}$ and $q \in \mathrm{U}_p$, let us consider $\gamma(t) : [0,1] \to \mathrm{M}$ an arbitrary path connecting them. Let us define the functional over the set of paths connecting p and q as,

$$\mathrm{L}(\gamma) := \int_0^1 \langle \dot\gamma(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\gamma(t)}\ dt, \tag{22}$$

where $\langle\cdot,\cdot\rangle_{\gamma(t)}$ denotes the inner product with respect to g evaluated at $\gamma(t)$ and P_t is the parallel transport along γ with respect to the ∇-connection. When γ is a ∇^*-geodesic we have that L assumes a very useful form.

Proposition II.1. *Let* $\sigma^* : \mathrm{I} \to \mathrm{M}$ *be a* ∇^**-geodesic connecting* p *and* q*. Then*

$$\mathrm{L}(\sigma^*) = \langle \mathrm{X}^*(p,q), \mathrm{X}(p,q)\rangle_p, \tag{23}$$

where $\mathrm{X}^*(p,q) = \overset{*}{\exp}_p^{-1}(q)$ *and* $\overset{*}{\exp}_p$ *is the exponential map of* ∇^* *connection at* p*.*

Proof. Consider the map

$$t \mapsto \langle \dot{\sigma}^*(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\sigma^*(t)},$$

where P_t denotes the parallel transport along $\sigma^*(t)$. Then, by taking the derivative with respect to t it trivially follows from relation (1) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \dot{\sigma}^*(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\sigma^*(t)} = \langle \nabla_t \mathrm{P}_t\mathrm{X}(p,q), \dot{\sigma}^*(t)\rangle_{\sigma^*(t)} + \langle \nabla_t^* \dot{\sigma}^*(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\sigma^*(t)} \ ,$$

where ∇_t and ∇_t^* are covariant derivatives [Appendix A] with respect to ∇ and ∇^* connections, respectively. By recalling that $\mathrm{P}_t\mathrm{X}(p,q)$ is the parallel transport with respect to the ∇-connection along σ^* we trivially have that $\nabla_t\mathrm{P}_t\mathrm{X}(p,q) \equiv 0$. Analogously, we have that $\nabla_t^*\dot{\sigma}^*(t) \equiv 0$ because $\dot{\sigma}^*$ is parallel along σ^* with respect to the ∇^*-connection. Therefore, we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \dot{\sigma}^*(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\sigma^*(t)} = 0,$$

and finally, we arrive at

$$\langle \dot{\sigma}^*(t), \mathrm{P}_t\mathrm{X}(p,q)\rangle_{\sigma^*(t)} = \langle \dot{\sigma}^*(0), \mathrm{X}(p,q)\rangle_p \ .$$

Hence, we can conclude by noticing that $\dot{\sigma}^*(0) = \overset{*}{\exp}_p^{-1}(q) = \mathrm{X}^*(p,q)$. □

Remark II.2. *The functional* L *can be also computed over a* ∇*-geodesic* σ *from* p *to* q*. In this case, it assumes the following expression*

$$\mathrm{L}(\sigma) = \int_0^1 \|\dot{\sigma}(t)\|^2_{\sigma(t)}\ dt, \tag{24}$$

where the integrand is now not constant with respect to t*.*

Before proving Theorem I.1, we now investigate the intrinsic geometry of geodesics due to the duality (1) of the affine connections ∇ and ∇^*. To this aim, let us consider the hypersurface $\mathrm{H}_p(\kappa) \subset \mathrm{M}$ defined as follows

$$\mathrm{H}_p(\kappa) := \{q \in \mathrm{M} \mid \langle \mathrm{X}_p(q), \mathrm{X}_p^*(q)\rangle_p = \kappa\} \ , \tag{25}$$

where

$$\mathrm{X}_p(q) = \exp_p^{-1}(q), \qquad \mathrm{X}^*_p(q) = \overset{*}{\exp}_p{}^{-1}(q)\ .$$

From assumption (I) of Eq. (21) it immediately follows that, if $\mathrm{H}_p(\kappa) \subset \mathrm{U}_p$ then $\mathrm{H}_p(\kappa)$ is really an hypersurface within M. We will soon prove that the combination of vectors $\Pi_q(p)$ and $\Pi^*_q(p)$ defines the rays of the hypersurface $\mathrm{H}_p(\kappa)$. In particular, we will show that $\Pi_q(p) + \Pi^*_q(p)$ is orthogonal to $\mathrm{H}_p(\kappa)$ at each $q \in \mathrm{H}_p(\kappa)$.

Owing to the duality structure of the statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$, we can introduce in $\mathrm{T}_p\mathrm{M}$ two notions of geodesic pseudo-spheres.

Definition II.1. *Let $p \in \mathrm{M}$ and $\mathrm{U}_p \subset \mathrm{M}$ be a neighborhood of p as in* (21)*. Consider the set $\mathcal{E}_p$ of all tangent vectors X_p for which there exist ∇ and ∇^* geodesics σ and σ^*, respectively, such that*

$$\sigma(0) = p,\ \dot\sigma(0) = X_p \qquad \text{and} \qquad \sigma^*(0) = p,\ \dot\sigma^*(0) = X_p\ .$$

Then we define

$$\mathrm{S}_p(\kappa) := \Big\{ X_p \in \mathrm{T}_p\mathrm{M} \mid \langle \exp_p^{-1}(\overset{*}{\exp}_p\,(X_p)), X_p\rangle_p = \kappa \Big\} \tag{26}$$

and

$$\mathrm{S}^*_p(\kappa) := \Big\{ X_p \in \mathrm{T}_p\mathrm{M} \mid \langle \overset{*}{\exp}_p{}^{-1}\left(\exp_p(X_p)\right), X_p\rangle_p = \kappa \Big\}. \tag{27}$$

Remark II.3. *Both the sets, $\mathrm{S}_p(\kappa)$ and $\mathrm{S}^*_p(\kappa)$ are hypersurfaces of $\mathrm{T}_p\mathrm{M}$ because of the assumption* (21)*. In addition, we can trivially see that the image of $\mathrm{S}_p(\kappa)$ through the exponential map of ∇^*-connection is given by*

$$\overset{*}{\exp}_p\,(\mathrm{S}_p(\kappa)) = \mathrm{H}_p(\kappa),$$

and the action of the exponential map of ∇-connection on $\mathrm{S}^*_p(\kappa)$ gives

$$\exp_p\left(\mathrm{S}^*_p(\kappa)\right) = \mathrm{H}_p(\kappa),$$

where $\mathrm{H}_p(\kappa)$ is defined in (25)*.*

Spheres of Def. II.1 are not the same but almost the same object. Indeed, consider the map

$$\begin{aligned} &I_p : \mathrm{S}_p(\kappa) \to \mathrm{S}^*_p(\kappa) \\ &I_p(x) := \exp_p^{-1}\left(\overset{*}{\exp}_p\,(x)\right),\ \forall\ x := X_p \in \mathrm{S}_p(\kappa)\ . \end{aligned} \tag{28}$$

Then we have,

Proposition II.2. *The map* $I_p : \mathrm{S}_p(\kappa) \to \mathrm{S}^*_p(\kappa)$ *defined by Eq.* (28) *is an isomorphism of vector spaces. In addition, the following diagram*

$$\begin{array}{ccc} \mathrm{S}_p(\kappa) & \xrightarrow{\overset{*}{\exp}_p} & \mathrm{H}_p(\kappa) \\ {\scriptstyle I_p}\big\downarrow & \nearrow_{\exp_p} & \\ \mathrm{S}^*_p(\kappa) & & \end{array}$$

is commutative.

Proof. Consider $x \in \mathrm{S}_p(\kappa)$. Firstly, we have that $I_p(x) \in \mathrm{S}^*_p(\kappa)$. Indeed,

$$\begin{aligned} \langle \overset{*}{\exp}{}_p^{-1} \left(\exp_p(I_p(x))\right), I_p(x)\rangle_p &= \langle x, \exp_p^{-1}\left(\overset{*}{\exp}_p (x)\right)\rangle_p \\ &= \kappa \ . \end{aligned}$$

Consider now the map $\tilde{I}_p : \mathrm{S}^*_p(\kappa) \to \mathrm{S}_p(\kappa)$ defined by

$$\tilde{I}_p(x) = \overset{*}{\exp}{}_p^{-1} \left(\exp_p(x)\right) \ .$$

Then, we can trivially see that

$$\begin{aligned} I_p \circ \tilde{I}_p(x) = x \ , &\quad \forall x \in \mathrm{S}_p(\kappa) \\ \tilde{I}_p \circ I_p(x) = x \ , &\quad \forall x \in \mathrm{S}^*_p(\kappa) \ . \end{aligned}$$

Therefore, we can conclude that $\tilde{I}_p = I_p^{-1}$. In order to prove that the diagram is commutative, let us consider $q = \overset{*}{\exp}_p (x)$ by some $x \in \mathrm{S}_p(\kappa)$. From Remark II.3 we know that $q \in \mathrm{H}_p(\kappa)$. In addition, by the definition (28) we also have that $q = \exp_p(I_p(x))$ □

We now proceed to prove the First Variational Formula of the functional L. In order to pursue this goal, let us firstly introduce the notion of path variation. Given an arbitrary path $\gamma : [0, 1] \to \mathrm{M}$ from p to q, we call $\Sigma : (-\varepsilon, \varepsilon) \times [0, 1] \to \mathrm{M}$ a variation of γ if $s \mapsto \Sigma^{(t)}(s)$ and $t \mapsto \Sigma_s(t)$ are smooth curves, $\Sigma_s(0) = p$ for all $s \in (-\varepsilon, \varepsilon)$ and $\Sigma(0, t) \equiv \gamma(t)$. In addition, Σ is a ∇-*geodesic variation* if $\Sigma^{(t)}(s)$ and $\Sigma_s(t)$ are ∇-geodesic. A vector field along Σ is a smooth map $\Xi : (-\varepsilon, \varepsilon) \times \mathrm{I} \to \mathrm{TM}$ such that $\Xi(s, t) \in \mathrm{T}_{(s,t)}\mathrm{M}$ for each (s, t). Two very special vector fields are defined as follows

$$\partial_t \Sigma(s, t) := \frac{\mathrm{d}}{\mathrm{d}t}\Sigma_s(t) = \dot{\Sigma}_s(t) \ , \quad \partial_s \Sigma(s, t) := \frac{\mathrm{d}}{\mathrm{d}s}\Sigma^{(t)}(s) = \dot{\Sigma}^{(t)}(s) \ . \tag{29}$$

Finally, $V(t) = \partial_s \Sigma^*(0, t) \in \mathcal{T}(\gamma)$ is called the *variation* vector field of Σ.

Given a variation $\Sigma(s,t)$ of an arbitrary path $\gamma(t)$ from p to q we can consider for every $s \in (-\varepsilon,\varepsilon)$ the vector,

$$\mathrm{X}_p(s) \equiv \mathrm{X}(p,\Sigma_s(1)) := \exp_p^{-1}(\Sigma_s(1)) \ , \tag{30}$$

which is the velocity vector at p of the ∇-geodesic connecting p and $\Sigma_s(1)$. Therefore, the First Variational Formula of L is proved by the following Proposition.

Proposition II.3. *Let* $\gamma : [0,1] \to \mathrm{M}$ *be a smooth curve and* $\Sigma^* : (-\varepsilon,\varepsilon)\times[0,1] \to \mathrm{M}$ *be a variation of* γ*. Let* $V \in \mathcal{T}(\gamma)$ *be the variation vector field of* Σ^**. Finally, let us define the functional* $\mathrm{L}(s) := \mathrm{L}(\Sigma^*_s)$*. Then we have*

$$\frac{\mathrm{dL}}{\mathrm{d}s}(0) = \langle V(t), \mathrm{P}_t\mathrm{X}_p(q)\rangle_{\gamma(t)}\big|_0^1 + \int_0^1 \langle\dot\gamma(t), \nabla_s\mathrm{P}_t\mathrm{X}_p(q)\rangle_{\gamma(t)}dt, \tag{31}$$

where P_t *denotes the parallel transport with respect to* ∇ *along the curve* $\gamma(t)$*.*

Proof. Let us first see the definition of L evaluated at Σ^*_s:

$$\mathrm{L}(s) \equiv \mathrm{L}(\Sigma^*_s) = \int_0^1 \langle\frac{\mathrm{d}\Sigma^*_s}{\mathrm{d}t}(t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\ dt\ , \tag{32}$$

where $\mathrm{P}_{s,t}$ is the parallel transport along the curve $t \mapsto \Sigma^*_s(t)$ with respect to the ∇-connection.

Therefore, by taking the derivative and exploiting relation (1) we obtain

$$\frac{\mathrm{dL}}{\mathrm{d}s}(s) \ = \ \int_0^1 \left(\langle\nabla^*_s\frac{\mathrm{d}\Sigma^*_s}{\mathrm{d}t}(t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)} + \langle\frac{\mathrm{d}\Sigma^*_s}{\mathrm{d}t}(t), \nabla_s\mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\right)\ dt, \tag{33}$$

where $\nabla_s = \nabla_{\dot\Sigma^{*(t)}(s)}$ and $\nabla^*_s = \nabla^*_{\dot\Sigma^{*(t)}(s)}$ are the covariant derivatives along $\Sigma^{*(t)}(s)$ with respect to ∇ and ∇^*, respectively. Since the connection ∇^* is torsion-free we have that $\nabla^*_s\left(\partial_t\Sigma^*(s,t)\right) = \nabla^*_t\left(\partial_s\Sigma^*(s,t)\right)$. Therefore, by means of the following computations

$$\begin{aligned}
\frac{\mathrm{dL}}{\mathrm{d}s}(s) \ &= \ \int_0^1 \ \left(\langle\nabla^*_t\partial_s\Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)} + \langle\partial_t\Sigma^*(s,t), \nabla_s\mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\right)\ dt \\
&= \int_0^1 \ \left(\frac{\mathrm{d}}{\mathrm{d}t}\langle\partial_s\Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)} - \langle\partial_s\Sigma^*(s,t), \nabla_t\mathrm{P}_{s,t}\mathrm{X}_p(s))\rangle_{\Sigma^*_s(t)}\right) dt \\
&\quad + \int_0^1 \ \left(\langle\partial_t\Sigma^*(s,t), \nabla_s\mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\right)\ dt \\
&= \int_0^1 \ \left(\frac{\mathrm{d}}{\mathrm{d}t}\langle\partial_s\Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)} + \langle\partial_t\Sigma^*(s,t), \nabla_s\mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\right)\ dt\ ,
\end{aligned}$$

where we used $\nabla_t\mathrm{P}_{s,t}\mathrm{X}_p(s) = 0$ and definitions (29), we arrive at

$$\frac{\mathrm{dL}}{\mathrm{d}s}(s) = \langle\partial_s\Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\big|_0^1 + \int_0^1 \langle\dot\Sigma^*_s(t), \nabla_s\mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\ dt\ . \tag{34}$$

Finally, recalling that $\Sigma^*(0,t) \equiv \gamma(t)$ and $\partial_s \Sigma^*(0,t) = V(t)$ we obtain that

$$\frac{d\mathrm{L}}{ds}(0) = \langle V(t), \mathrm{P}_t \mathrm{X}_p(0)\rangle_{\gamma(t)}\Big|_0^1 + \int_0^1 \langle \dot{\gamma}(t), \nabla_s \mathrm{P}_t \mathrm{X}_p(0)\rangle_{\gamma(t)} dt. \tag{35}$$

In the end, we get the statement (31) by noticing from (30) that

$$\mathrm{X}_p(s)\big|_{s=0} = \exp_p^{-1}(\Sigma(0,1)) = \exp_p^{-1}(\gamma(1)) = \mathrm{X}_p(q) \qquad \square$$

Remark II.4. *The expression $\nabla_s \mathrm{P}_{s,t} \mathrm{X}_p(s)$ in Eq.* (34) *can be actually written as $\mathrm{P}_{s,t} \nabla_s \mathrm{X}_p(s)$. To prove it, let us choose a frame $\{e_i(s,t)\}$ that is parallel translating along $\Sigma^*(s,t)$. Then, we have that*

$$\mathrm{P}_{s,t} \mathrm{X}_p(s) = p^i(s,t) e_i(s,t)$$

where $p^i(s,t)$ is the principal part of $\mathrm{P}_{s,t} \mathrm{X}_p(s)$ [23] and the summation over i is intended. Then,

$$\mathrm{P}_{s,t} \nabla_s \mathrm{X}_p(s) = (\partial_s p^i)(s,t) e_i(s,t)$$

since $\{e_i(s,t)\}$ is parallel along $\Sigma^*(s,t)$. Finally, from

$$\nabla_s \mathrm{P}_{s,t} \mathrm{X}_p(s) = (\partial_s p^i)(s,t) e_i(s,t)$$

we arrive at

$$\nabla_s \mathrm{P}_{s,t} \mathrm{X}_p(s) = \mathrm{P}_{s,t} \nabla_s \mathrm{X}_p(s) = \mathrm{P}_{s,t} \partial_s \mathrm{X}_p(s) \ .$$

We are now in position to prove the following relevant Theorem.

Theorem II.1. *Let $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ be a statistical manifold. Consider $p \in \mathrm{M}$ and $\mathrm{U}_p \subset \mathrm{M}$ under the assumption* (21). *Then, for all $q \in \mathrm{H}_p(\kappa)$, the sum $\Pi_q(p) + \Pi^*_q(p)$ of the parallel transports of the vectors $\mathrm{X}_p(q)$ and $\mathrm{X}^*_p(q)$ along ∇^* and ∇ geodesics, respectively, starting from p are all orthogonal to $\mathrm{H}_p(\kappa)$ at q.*

Proof. Let us consider a curve within the pseudo-sphere $\mathrm{S}_p(\kappa)$, namely $\tau : (-\varepsilon, \varepsilon) \to \mathrm{S}_p(\kappa)$, such that $\tau(0) = \mathrm{X}^*_p(q) = \overset{*}{\exp}_p^{-1}(q)$. The map $\Sigma^*(s,t) := \overset{*}{\exp}_p(t\ \tau(s))$ is a ∇^*-geodesic variation of the ∇^*-geodesic $\sigma^*(t) := \overset{*}{\exp}_p(t\ \tau(0))$. At the same time, by means of the map I_p defined by Eq. (28), we may also consider a ∇-geodesic variation $\Sigma(s,t) = \exp_p(t\ I_p(\tau(s)))$ of the ∇-geodesic $\sigma(t) = \exp_p(t\ I_p(\tau(0)))$. Indeed, by definition, we have that $I_p(\tau(0)) = \mathrm{X}_p(q)$. In addition we also have that

$$\Sigma^*(s,1) = \Sigma(s,1) \quad \forall s \in (-\varepsilon, \varepsilon) \ . \tag{36}$$

Let us now evaluate the functional L at $\Sigma_s^*(t)$. From Eq. (22) we have that

$$\mathrm{L}(s) = \int_0^1 \langle \dot{\Sigma}^*{}_s(t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma_s^*(t)}\ \mathrm{d}t\ ,$$

where again $\mathrm{P}_{s,t}$ is the parallel transport along the curve $t \mapsto \Sigma_s^*(t)$ and $\mathrm{X}_p(s)$ is given by Eq. (30). Since $\Sigma_s^*(t)$ is ∇^*-geodesic then $\dot{\Sigma}_s^*(t)$ is parallel with respect to ∇^*. For this reason and from Eq. (1) we immediately get,

$$\mathrm{L}(s) = \langle \dot{\Sigma}_s^*(0), \mathrm{X}_p(s)\rangle_p\ .$$

Now, we observe that $\dot{\Sigma}_s^*(0) = \tau(s)$. Moreover, by Eq. (30) we have that $\mathrm{X}_p(s) = \exp_p(\Sigma_s^*(1))$. Since now $\Sigma_s^*(t)$ is ∇^*-geodesic we have that $\Sigma_s^*(1) = \overset{*}{\exp}_p\ (\tau(s))$. Hence, we have that $\mathrm{X}_(s) = \exp_p^{-1}\left(\overset{*}{\exp}_p\ (\tau(s)\right)$. From this, we obtain the following result on L,

$$\begin{aligned}\mathrm{L}(s) &= \langle \dot{\Sigma}_s^*(0), \mathrm{X}_p(s)\rangle_p \\ &= \langle \tau(s), \exp_p^{-1}\left(\overset{*}{\exp}_p\ (\tau(s))\right)\rangle_p = \kappa\ ,\end{aligned} \tag{37}$$

where the last equality follows by the assumption that $\tau(s) \in \mathrm{S}_p(\kappa)$. Finally, we trivially get

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{L}(s)\bigg|_{s=0} = 0\ . \tag{38}$$

Consider now the First Variational Formula (34) of $\mathrm{L}(s)$, that for the sake of readability we rewrite above,

$$\frac{\mathrm{dL}}{\mathrm{d}s}(s) = \langle \partial_s \Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma_s^*(t)}\big|_0^1 + \int_0^1 \langle \dot{\Sigma}_s^*(t), \nabla_s \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma_s^*(t)}\ dt\ .$$

Let us now investigate the second term of the right hand side of (34). First of all, let us rely on a local frame of $\mathrm{T}_p\mathrm{M}$ that we parallel transport with respect to ∇ along both the families of curves, $\Sigma(s,t)$ and $\Sigma^*(s,t)$. From Remark II.4 we know that $\nabla_s \mathrm{P}_{s,t}\mathrm{X}_p(s) = \mathrm{P}_{s,t}\nabla_s \mathrm{X}_p(s)$. Then, since $\Sigma_s^*(t)$ is ∇^*-geodesic we also know that $\dot{\Sigma}_s^*(t)$ is the parallel transport with respect to the ∇^*-connection of the velocity of $\Sigma_s^*(t)$ at $t = 0$. For these reasons we have that the integrand function can be written as $\langle \dot{\Sigma}_s^*(0), \partial_s \mathrm{X}_p(s)\rangle_p$. We now evaluate this inner product along the ∇-geodesic $\Sigma_s(t)$. So, we have

$$\begin{aligned}\langle \dot{\Sigma}_s^*(0), \partial_s \mathrm{X}_p(s)\rangle_p &= \langle \mathrm{P}_{s,t}^* \dot{\Sigma}_s^*(0), \mathrm{P}_{s,t}\partial_s \mathrm{X}_p(s)\rangle_{\Sigma_s(t)} \\ &= \langle \mathrm{P}_{s,t}^* \dot{\Sigma}_s^*(0), \nabla_s \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma_s(t)},\end{aligned}$$

where we used the invariance of the inner product under $\mathrm{P}_{s,t}^*$ and $\mathrm{P}_{s,t}$ and again the Remark II.4.

We may now observe that $\Sigma_s(t)$ is a ∇-geodesic and $\mathrm{X}_p(s) = \dot{\Sigma}_s(0)$. Then $\mathrm{P}_{s,t}\mathrm{X}_p(s)$ is nothing but $\dot{\Sigma}_s(t)$. Now, by means of the torsion-freeness of ∇ we have that $\nabla_s\partial_t\Sigma(s,t) = \nabla_t\partial_s\Sigma(s,t)$.

So, according to Eq. (1) we arrive at

$$\begin{aligned}\langle \mathrm{P}^*_{s,t}\dot{\Sigma}^*_s(0), \nabla_s\dot{\Sigma}_s(t)\rangle_{\Sigma_s(t)} &= \langle \mathrm{P}^*_{s,t}\dot{\Sigma}^*_s(0), \nabla_t\partial_s\Sigma(s,t)\rangle_{\Sigma_s(t)} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\langle \mathrm{P}^*_{s,t}\dot{\Sigma}^*_s(0), \partial_s\Sigma(s,t)\rangle_{\Sigma_s(t)} - \langle \nabla^*_t\mathrm{P}^*_{s,t}\dot{\Sigma}^*_s(0), \partial_s\Sigma(s,t)\rangle_{\Sigma_s(t)}\ . \end{aligned} \quad (39)$$

Let now observe that $\dot{\Sigma}^*_s(0)$ is the velocity vector at p of the ∇^*-geodesic connecting p and $\Sigma^*_s(1) = \Sigma_s(1)$. Then, in accordance to Eq. (30) we use the following notation,

$$\dot{\Sigma}^*_s(0) = \mathrm{X}^*_p(s) \equiv \mathrm{X}^*(p, \Sigma_s(1)) := \overset{*}{\exp}{}_p^{-1}(\Sigma_s(1))\ . \quad (40)$$

Going back to Eq. (34), we substitute (39) into the integral and, since $\nabla^*_t\mathrm{P}^*_{s,t}\dot{\Sigma}^*_s(0) \equiv 0$, we obtain

$$\frac{\mathrm{dL}}{\mathrm{d}s}(s) = \langle \partial_s\Sigma^*(s,t), \mathrm{P}_{s,t}\mathrm{X}_p(s)\rangle_{\Sigma^*_s(t)}\Big|_0^1 + \langle \partial_s\Sigma(s,t), \mathrm{P}^*_{s,t}\mathrm{X}^*_p(s)\rangle_{\Sigma_s(t)}\Big|_0^1\ .$$

By evaluating this expression at $s=0$, we get

$$\frac{\mathrm{dL}}{\mathrm{d}s}(0) = \langle \partial_s\Sigma^*(0,1), \Pi_q(p)\rangle_q + \langle \partial_s\Sigma(0,1), \Pi^*_q(p)\rangle_q\ . \quad (41)$$

Indeed, firstly we have

$$\mathrm{X}_q(s)\big|_{s=0} = \exp_p^{-1}(q) = \mathrm{X}_p(q), \quad \mathrm{X}^*_q(s)\big|_{s=0} = \overset{*}{\exp}{}_p^{-1}(q) = \mathrm{X}^*_p(q)\ .$$

In addition, $\Sigma^*(0,t) = \sigma^*(t)$, $\Sigma(0,t) = \sigma(t)$ and then

$$\Pi_q(p) = \mathrm{P}_{\sigma^*}\mathrm{X}_p(q), \quad \Pi^*_q(p) = \mathrm{P}^*_\sigma\mathrm{X}^*_p(q)\ .$$

Finally, since $\Sigma^{*(0)}(s) = p = \Sigma^{(0)}(s)$ for every $s \in (-\varepsilon, \varepsilon)$ we also have that

$$\partial_s\Sigma^*(0,0) = O_p = \partial_s\Sigma(0,0)\ ,$$

where O_p is the null element of $\mathrm{T}_p\mathrm{M}$.

In order to conclude the proof of the Theorem, let us observe that

$$\begin{aligned}\partial_s\Sigma(0,1) &= \left(\mathrm{d}\exp_p\right)_{I_p(\mathrm{X}^*_p(q))}\left[\partial_s I_p(\tau(s))\big|_{s=0}\right] \\ &= \left(\mathrm{d}\exp_p\right)_{I_p(\mathrm{X}^*_p(q))}\left[(\mathrm{d}I_p)_{\mathrm{X}^*_p(q)}(\tau'(0))\right] \\ &= \left(\mathrm{d}\exp_p\right)_{I_p(\mathrm{X}^*_p(q))}\left[\left(\mathrm{d}\exp_p\right)^{-1}_{I_p(\mathrm{X}^*_p(q))}\left(\mathrm{d}\,\overset{*}{\exp}{}_p\right)_{\mathrm{X}^*_p(q)}(\tau'(0))\right] \\ &= \left(\mathrm{d}\exp_p\right)_{I_p(\mathrm{X}^*_p(q))}\left[\left(\mathrm{d}\exp_p\right)^{-1}_{I_p(\mathrm{X}^*_p(q))}(\partial_s\Sigma^*(0,1))\right] \\ &= \partial_s\Sigma^*(0,1)\ .\end{aligned}$$

In the end, from Eq. (38) and Eq. (41) we obtain

$$0 = \langle \partial_s \Sigma^*(0,1), \Pi_q(p) + \Pi^*_q(p) \rangle_q \tag{42}$$

for an arbitrary tangent vector $\partial_s \Sigma^*(0,1)$ of $\overset{*}{\exp}_p\,(\mathrm{S}_p(\kappa)) = \mathrm{H}_p(\kappa)$ $\square$

Proof of the Theorem I.1. Consider $p \in \mathrm{M}$ and $\mathrm{U}_p \subset \mathrm{M}$ under the assumption (21). The psuedo-squared-norm $r_p(q)$ is defined for all $q \in \mathrm{U}_p$ by

$$r_p(q) = \langle \exp_p^{-1}(q), \overset{*}{\exp}_p{}^{-1}(q) \rangle_p \; .$$

In order to prove that

$$\mathrm{grad}_q r_p = \Pi_q(p) + \Pi^*_q(p),$$

let us step back into the proof of Theorem II.1. A variation of the end point q can be given in terms of the ∇^*-geodesic variation $\Sigma^*(s,t) = \overset{*}{\exp}_p\,(t\;\tau(s))$ of the ∇^*-geodesic $\sigma^*(t) = \overset{*}{\exp}_p\,(t\;\tau(0))$ as well as in terms of the ∇-geodesic variation $\Sigma(s,t) = \exp_p(t\; I_p(\tau(s)))$ of the ∇-geodesic $\sigma(t) = \exp_p(t\; I_p(\tau(0)))$. Again, $I_p : \mathcal{E}_p \to \mathcal{E}_p$ is the map defined by

$$I_p(x) = \exp_p^{-1}\left(\overset{*}{\exp}_p\,(x)\right), \quad \forall\; x := X_p \in \mathcal{E}_p,$$

where $\mathcal{E}_p$ is the set of tangent vectors at p to M given in the Def. (II.1). The curve $\tau(s)$ is now just a curve within $\mathcal{E}_p$. A variation of the end point q is then given by $\Sigma^*(s,1) = \overset{*}{\exp}_p\,(\tau(s))$ for $s \in (-\varepsilon, \varepsilon)$.

In addition, we know from Eq. (23) that the pseudo-squared-norm $r_p(q)$ is achieved by the computation of L over the ∇^*-geodesic σ^*. In this way, we have that the differential $(\mathrm{d}r_p)_q$ at q of $r_p(q)$ can be evaluated as the derivative with respect to s at $s = 0$ of the functional L,

$$\frac{\mathrm{dL}}{\mathrm{d}s}(s)\bigg|_{s=0} \equiv (\mathrm{d}r_p)_q \; , \tag{43}$$

where $\mathrm{L}(s) = \mathrm{L}(\Sigma^*_s)$. On the other hand, the differential of r_p is uniquely expressed in terms of the gradient as follows,

$$(\mathrm{d}r_p)_q(w) = \langle \mathrm{grad}_q r_p, w \rangle_q, \quad \forall\; w := W_p \in \mathrm{T}_p\mathrm{M} \; .$$

Hence, we obtain

$$\begin{aligned} \langle \mathrm{grad}_q r_p, \partial_s \Sigma^*(0,1) \rangle_q &= \frac{\mathrm{dL}}{\mathrm{d}s}(s)\bigg|_{s=0} \\ &= \langle \Pi_q(p) + \Pi^*_q(p), \partial_s \Sigma^*(0,1) \rangle_q, \end{aligned}$$

which gives us that $\mathrm{grad}_q r_p = \Pi_q(p) + \Pi^*_q(p)$ because of the arbitrariness of $\partial_s \Sigma^*(0,1)$. □

Owing to the dual structure of a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$, we can prove Theorem II.1 by interchanging the role of ∇ and ∇^*, as well. In this case we need to consider a function L^* different from the one L of Eq. (22). The new one is defined over the set of path connecting p and q in terms of the parallel transport with respect to ∇^*-connection, and it reads as follows,

$$\mathrm{L}^*(\gamma) := \int_0^1 \langle \dot{\gamma}(t), \mathrm{P}^*_t \mathrm{X}^*_p(q) \rangle_{\gamma(t)} \, dt, \tag{44}$$

where P^*_t is the parallel transport along $\gamma(t)$ with respect to the ∇^*-connection and $\mathrm{X}^*_p(q) = \overset{*}{\exp}{}_p^{-1}(q)$. In this case the pseudo-squared-norm is achieved by computing L^* over the ∇-geodesic $\sigma(t)$ connecting p and q.

Proposition II.4. *Let $\sigma : \mathrm{I} \to \mathrm{M}$ be a ∇-geodesic connecting p and q. Then*

$$\mathrm{L}^*(\sigma) = \langle \exp_p^{-1}(q), \overset{*}{\exp}{}_p^{-1}(q) \rangle_q \ . \tag{45}$$

Proof. The proof relies on the Eq. (1) as well as in the case of Prop. II.1.

By evaluating L^* over the ∇^*-geodesic σ^* we obtain

$$\mathrm{L}^*(\sigma^*) = \int_0^1 \|\dot{\sigma}^*(t)\|^2_{\sigma^*(t)} \mathrm{d}t$$

where $\|\dot{\sigma}^*(t)\|_{\sigma^*(t)}$ is not constant with respect to t as well as in the case of L computed over the ∇-geodesic $\sigma(t)$.

Also for L^* we can rely on a First Variational Formula. It is stated in the following Proposition and the proof is avoided as it can be easily given by going back through the proof of Prop. II.3.

Proposition II.5. *Let $\gamma : [0,1] \to \mathrm{M}$ be a smooth curve and $\Sigma : (-\varepsilon, \varepsilon) \times [0,1] \to \mathrm{M}$ be a variation of γ. Let $V \in \mathcal{T}(\gamma)$ be the variation vector field of Σ. Finally, define the functional $\mathrm{L}^*(s) := \mathrm{L}^*(\Sigma_s)$. Then we have*

$$\frac{\mathrm{dL}^*}{\mathrm{d}s}(0) = \langle V(t), \mathrm{P}^*_t \mathrm{X}^*_p(q) \rangle_{\gamma(t)} \big|_0^1 + \int_0^1 \langle \dot{\gamma}(t), \nabla^*_s \mathrm{P}^*_t \mathrm{X}^*_p(q) \rangle_{\gamma(t)} dt, \tag{46}$$

where P^*_t denotes the parallel transport with respect to ∇^* along the curve $\gamma(t)$.

Finally, Theorem (II.1) can be proved by resorting to Eq. (46) and following the same methods carried out on ∇-connection and functional L.

In the rest of this section we consider a self-dual manifold in order to prove that Theorem I.1 is consistent with the Riemannian case. A statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is called self-dual when $\nabla = \nabla^*$. Therefore, from Eq. (15) we recover the Riemannian structure of M. In addition, we have that the ∇ and ∇^* geodesics coincide and then we obtain

$$\Pi_q(p) = \mathrm{P}_{\sigma^*}\dot{\sigma}(0) = \dot{\sigma}(1) = \mathrm{P}^*_{\sigma}\dot{\sigma}^*(0) = \Pi^*_q(p)\ .$$

By applying Theorem I.1 to the case of self-duality we then get

$$\mathrm{grad}_q r_p = \Pi_q(p) + \Pi^*_q(p) = 2\dot{\sigma}(1),$$

where σ is now the $\overline{\nabla}_{\mathrm{LC}}$-geodesic from p to q. We observe that, in this particular case, the pseudo-squared-norm becomes

$$r_p(q) = \langle \overline{\exp}_p^{-1}(q), \overline{\exp}_p^{-1}(q)\rangle_q = 2d_p(q)$$

where $d_p(q)$ is given by Eq. (16). Finally we can get

$$\dot{\sigma}(1) = \mathrm{grad}_q d_p = \frac{1}{2}\mathrm{grad}_q r_p = \dot{\sigma}(1)\ .$$

Hence, for a self-dual manifold the result of Theorem I.1 is in accordance with the one of the Riemannian case obtained through the celebrated Gauss Lemma.

III. GENERAL CANONICAL DIVERGENCE

Theorem I.1 identifies the appropriate vector fields for defining the differential of the pseudo-squared-norm $r_p(q)$ given by Eq. (12). Thanks to the metric structure of the statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$, we can express the differential $(\mathrm{d}r_p)_q$ in terms of the gradient $\mathrm{grad}_q r_p$,

$$(\mathrm{d}r_p)_q(X_q) = \langle \mathrm{grad}_q r_p, X_q\rangle_q, \quad \forall\ X_q \in \mathrm{T}_q\mathrm{M}\ . \tag{47}$$

Let $\gamma : [0, 1] \to \mathrm{M}$ be a path from p to q, we assume that $\gamma(t)$ lies in a neighborhood U_p of p as in (21), for all $t \in [0, 1]$. Then consider ∇ and ∇^* geodesics σ_t and σ^*_t, respectively, connecting p and $\gamma(t)$. Letting t vary in $[0, 1]$ we obtain two vector fields along γ,

$$\Pi_t(p) = \mathrm{P}_{\sigma^*_t}\mathrm{X}_p(t), \qquad \mathrm{X}_p(t) = \exp_p^{-1}(\gamma(t)) \tag{48}$$

$$\Pi^*_t(p) = \mathrm{P}^*_{\sigma_t}\mathrm{X}^*_p(t), \qquad \mathrm{X}^*_p(t) = \overset{*}{\exp}{}_p^{-1}(\gamma(t))\ . \tag{49}$$

From Theorem I.1 we can recover the pseudo-squared-norm $r_p(q)$ by composing the inner product of the curve velocity $\dot\gamma(t)$ with the vector field $\Pi_t(p) + \Pi_t^*(p)$,

$$\begin{aligned}\int_0^1 \langle \Pi_t(p) + \Pi_t^*(p), \dot\gamma(t)\rangle_{\gamma(t)}\ dt &= \int_0^1 \langle \mathrm{grad}_{\gamma(t)}\ r_p, \dot\gamma(t)\rangle_{\gamma(t)}\ dt \\ &= \int_0^1 \mathrm{d}_{\gamma(t)}\ r_p\left(\dot\gamma(t)\right)\ dt \\ &= \int_0^1 \frac{\mathrm{d}\ r_p\circ\gamma}{\mathrm{d}t}(t)\ dt \\ &= r_p(\gamma(1)) - r_p(\gamma(0)) \\ &= r_p(q), \end{aligned} \tag{50}$$

where, obviously, we have $r_p(p) \equiv 0$. Therefore, from computation (50) we obtain that the sum

$$\int_0^1 \langle \Pi_t(p), \dot\gamma(t)\rangle_{\gamma(t)}\ dt + \int_0^1 \langle \Pi_t^*(p), \dot\gamma(t)\rangle_{\gamma(t)}\ dt = r_p(q) \tag{51}$$

is independent of the particular path from p to q.

We now introduce two functions by integrating the two vector fields Π_t and Π_t^* along ∇^*-geodesic and ∇-geodesic, respectively. Soon after, we show that these functions are potential function for the geometry (1) according to relations (3) and (4). In addition, they play a key role for decomposing Π_t and Π_t^* in terms of gradient vector fields.

Definition III.1. *Let $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ be a statistical manifold. Consider $p \in \mathrm{M}$ and U_p as in* (21)*. For every $q \in \mathrm{U}_p$ we define the function $\varphi : \mathrm{M} \to \mathbb{R}$ by the path integration of the vector field $\Pi_t(p)$ along the ∇^*-geodesic $\sigma^*(t)$ from p to q,*

$$\varphi(p, q) := \varphi_p(q) := \int_0^1 \langle \Pi_t(p), \dot\sigma^*(t)\rangle_{\sigma^*(t)}\ \mathrm{d}t\ . \tag{52}$$

Analogously, we define the dual of φ_p by the path integration of the vector field $\Pi_t^*(p)$ along the ∇-geodesic $\sigma(t)$ from p to q,

$$\varphi^*(p, q) := \varphi_p^*(q) := \int_0^1 \langle \Pi_t^*(p), \dot\sigma(t)\rangle_{\sigma(t)}\ \mathrm{d}t\ . \tag{53}$$

*We refer to φ and φ^* as **Phi-functions**.*

A. Consistency Theorem

In order to show that $\varphi(p, q)$ allows to recover the dual geometry of $\mathcal{S}$, we have to prove its consistency with the dual structure $(\mathrm{g}, \nabla, \nabla^*)$ of M. This means that in a neighborhood of the diagonal set of $\mathrm{M} \times \mathrm{M}$ we need to verify that relations (3) and (4) are satisfied. We start by formulating a nice and useful representation of $\varphi(p, q)$.

Lemma III.1. *The Phi-function φ of Definition III.1 is given by*

$$\varphi(p,q) = \int_0^1 \langle I_p(t\mathrm{X}_p^*(q)), \mathrm{X}_p^*(q)\rangle_p \ \mathrm{d}t, \tag{54}$$

where $I_p(t\mathrm{X}_p^*(q)) = \exp_p^{-1}\left(\overset{*}{\exp}_p\ (t\mathrm{X}_p^*(q))\right)$.

Proof. Consider the ∇^*-geodesic $\sigma_t^*(s)$ that connects p to $\sigma^*(t)$, namely $\sigma_t^*(0) = p$ and $\sigma_t^*(1) = \sigma^*(t)$. Then, $\sigma_t^*(s) = \sigma^*(st)$. Therefore, a classical result in Riemannian geometry tells us that [16],

$$\dot{\sigma}_t^*(1) = t\ \dot{\sigma}^*(t)\ .$$

By substituting this expression into Eq. (52) we immediately obtain that

$$\varphi(p,q) = \int_0^1 \frac{1}{t}\langle \Pi_t(p), \dot{\sigma}_t^*(1)\rangle_{\sigma^*(t)} \ \mathrm{d}t\ . \tag{55}$$

Let us now recall that

$$\Pi_t(p) = \mathrm{P}_{\sigma_t^*}\mathrm{X}_p(\sigma^*(t)), \qquad \dot{\sigma}_t^*(1) = \mathrm{P}^*_{\sigma_t^*}\dot{\sigma}_t^*(0)\ .$$

Then, because of the invariance of the inner product under the combined action of P and P^* we get

$$\varphi(p,q) = \int_0^1 \frac{1}{t}\langle \mathrm{X}_p(t), \dot{\sigma}_t^*(0)\rangle_p \ \mathrm{d}t\ .$$

Finally, we may observe that

$$\mathrm{X}_p(t) = \exp_p^{-1}(\sigma^*(t)) = \exp_p^{-1}\left(\overset{*}{\exp}_p\ (t\ \mathrm{X}_p^*(q))\right), \qquad \dot{\sigma}_t^*(0) = t\ \mathrm{X}_p^*(q)$$

because σ^* is a ∇^*-geodesic and σ_t^* is a re-parametrization of σ^*. Hence, we obtain

$$\begin{aligned}\varphi(p,q) &= \int_0^1 \frac{1}{t}\langle \exp_p^{-1}\left(\overset{*}{\exp}_p\ (t\ \mathrm{X}_p^*(q))\right), t\ \mathrm{X}_p^*(q)\rangle_p \ \mathrm{d}t \\ &= \int_0^1 \langle I_p(t\ \mathrm{X}_p^*(t)), \mathrm{X}_p^*(q)\rangle_p \ \mathrm{d}t,\end{aligned}$$

by recalling definition (28) of the map I_p. □

Let us now assume that p and q are close to each other, that is

$$z^i = \xi_q^i - \xi_p^i \tag{56}$$

is small. Here $\{\boldsymbol{\xi}_p\}$ and $\{\boldsymbol{\xi}_q\}$ are local coordinates at p and q, respectively. Then, Taylor expansion of $\varphi(p,q)$ up to $O\left(\|\boldsymbol{z}\|^3\right)$ leads to the following result.

Proposition III.1. *Consider* $\|\boldsymbol{z}\| = \|\boldsymbol{\xi}_q - \boldsymbol{\xi}_p\|$ *small enough. Then, the function* $\varphi(p,q)$ *is expanded up to* $O\left(\|z\|^3\right)$ *as follows*

$$\varphi(p,q) = \frac{1}{2}\ \mathrm{g}_{ij}(p)\ z^i z^j + \frac{1}{6}\Lambda_{ijk}(p)\ z^i z^j z^k + O\left(\|z\|^3\right), \tag{57}$$

where

$$\Lambda_{ijk}(p) = 2\Gamma^*_{ijk}(p) + \Gamma_{ijk}(p)\ . \tag{58}$$

Proof. Let us consider the representation (54) of the function $\varphi(p,q)$. Then, recall that $I_p(t\ \mathrm{X}^*_p(t))$ is nothing but the velocity vector at p of the ∇-geodesic σ_t from p to $\sigma^*(t)$. On the other hand, $\mathrm{X}^*_p(q)$ is the velocity vector at p of the ∇^*-geodesic σ^* from p to q. Therefore, we need to Taylor expand up to $O\left(\|\boldsymbol{z}\|^4\right)$ with respect to the local coordinate $\{\boldsymbol{\xi}\}$ the following expression

$$\int_0^1 \mathrm{g}_{ij}(p)\ \dot{\sigma}_t(0)^i\ \dot{\sigma}^*(0)^j\ \mathrm{d}t\ . \tag{59}$$

The local coordinates $\boldsymbol{\xi}(t)$ of the ∇-geodesic $\sigma^*(t)$ in Taylor series are given by

$$\xi^j(t) = \xi^j_p + tz^j + \frac{t}{2}(1-t)\Gamma^{*j}_{\mu\nu}(p)z^\mu z^\nu + O\left(\|z\|^3\right), \tag{60}$$

where the summation over μ and ν is intended. Then we obtain,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma^*(0)^j = z^j + \frac{1}{2}\Gamma^{*j}_{\mu\nu}(p)z^\mu z^\nu + O\left(\|z\|^3\right). \tag{61}$$

In addition we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_t(0)^i = \sigma^*(t)^i - \xi^i_p + \frac{1}{2}\Gamma^i_{\mu\nu}(p)(\sigma^*(t)^\mu - \xi^\mu_p)(\sigma^*(t)^\nu - \xi^\nu_p) + O\left(\|\sigma^*(t) - \xi_p\|^3\right)\ .$$

Now, as

$$\sigma^*(t)^j - \xi^j_p = tz^j + \frac{t}{2}\Gamma^{*j}_{\mu\nu}(p)z^\mu z^\nu - \frac{t^2}{2}\Gamma^{*j}_{\mu\nu}(p)z^\mu z^\nu + O\left(\|z\|^3\right)$$

we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_t(0)^i = tz^i + \frac{t}{2}\Gamma^{*i}_{\mu\nu}z^\mu z^\nu + \frac{t^2}{2}(\Gamma^i_{\mu\nu}(p) - \Gamma^{*i}_{\mu\nu}(p))z^\mu z^\nu + O\left(\|z\|^3\right)\ . \tag{62}$$

At this point, we can substitute Eqs. (61), (62) into Eq. (59)

$$\begin{aligned}
\int_0^1 \mathrm{g}_{ij}(p)\ \dot{\sigma}_t(0)^i\ \dot{\sigma}^*(0)^j\ \mathrm{d}t\ =&\ \int_0^1 g_{ij}(p)\left[tz^i + \frac{t}{2}\Gamma^{*i}_{\mu\nu}(p)z^\mu z^\nu + \frac{t^2}{2}(\Gamma^i_{\mu\nu}(p) - \Gamma^{*i}_{\mu\nu}(p))z^\mu z^\nu\right]\\
&\times\left[z^j + \frac{1}{2}\Gamma^{*j}_{\mu\nu}(p)z^\mu z^\nu\right]\ \mathrm{d}t\\
=&\ \frac{1}{2}\mathrm{g}_{ij}(p)z^i z^j + \frac{1}{2}\mathrm{g}_{ij}(p)\Gamma^{*i}_{\mu\nu}(p)z^\mu z^\nu z^j\\
&+\frac{1}{6}\mathrm{g}_{ij}(p)z^i\left(\Gamma^j_{\mu\nu}(p) - \Gamma^{*j}_{\mu\nu}(p)\right)z^\mu z^\nu\ .
\end{aligned}$$

Finally, by symmetrizing the indices because of the multiplication $z^i z^j z^k$, we obtain

$$\varphi(p,q) = \frac{1}{2}\mathrm{g}_{ij}(p) z^i z^j + \frac{1}{6}\Lambda_{ijk}(p) z^i z^j z^k,$$

where $\Lambda_{ijk}(p) = 2{\Gamma^*}_{ijk}(p) + \Gamma_{ijk}(p)$ is obtained by recalling that $g_{il}\Gamma^l_{jk} = \Gamma_{ijk}$. □

Theorem III.1. *Consider a statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. *Let* $\varphi : \mathrm{M} \times \mathrm{M} \to \mathbb{R}$ *be a two point function defined by Eq.* (52). *Then we have*

$$\partial'_j \partial'_i \varphi(p,q)\big|_{p=q} = \mathrm{g}_{ij}(p) \tag{63}$$

$$\partial_k \partial'_j \partial'_i \varphi(p,q)\big|_{p=q} = -\overset{*}{\Gamma}_{ijk}(p), \tag{64}$$

where $\partial_i = \dfrac{\partial}{\partial \xi^i_p}$, $\partial'_i = \dfrac{\partial}{\partial \xi^i_q}$ *and* $\{\boldsymbol{\xi}_p\}$, $\{\boldsymbol{\xi}_q\}$ *are local coordinates at p and q, respectively.*

Proof. Consider the Taylor series (57) of $\varphi(p,q)$. By differentiating it with respect to $\boldsymbol{\xi}_q$ we obtain,

$$\partial'_i \varphi(p,q) = \mathrm{g}_{ij}(p) z^j + \frac{1}{2}\Lambda_{ijk}(p) z^j z^k \tag{65}$$

$$\partial'_j \partial'_i \varphi(p,q) = \mathrm{g}_{ij}(p) + \Lambda_{ijk}(p) z^k \ . \tag{66}$$

By evaluating $\partial'_j \partial'_i \varphi(p,q)$ at $\boldsymbol{\xi}_q = \boldsymbol{\xi}_p$, i.e. $\boldsymbol{z} = 0$, we obtain

$$\partial_j \partial_i \varphi(p,q)\big|_{p=q} = \mathrm{g}_{ij}(p) \ .$$

In addition, we differentiate Eq. (66) with respect to $\boldsymbol{\xi}_p$ and evaluate it at $\boldsymbol{z} = 0$. This computation leads to

$$\begin{aligned} \partial_k \partial'_j \partial'_i \varphi(p,q)\big|_{p=q} &= \partial_k \mathrm{g}_{ij}(p) - \Lambda_{ijk}(p) + \partial_k \Lambda_{ijk} z^k\Big|_{z=0} \\ &= \partial_k \mathrm{g}_{ij}(p) - \Lambda_{ijk}(p) \\ &= \Gamma_{ijk} + \overset{*}{\Gamma}_{ijk} - 2\overset{*}{\Gamma}_{ijk} - \Gamma_{ijk} = -\overset{*}{\Gamma}_{ijk}, \end{aligned} \tag{67}$$

where we used Eq. (58) and the relation $\partial_k \mathrm{g}_{ij} = \Gamma_{ijk} + \Gamma^*_{ijk}$. □

Remark III.1. *Consider the dual Phi-function* $\varphi^*(p,q)$. *By interchanging the role of* ∇ *connection with* ∇^* *connection, we obtain from Prop. III.1 the following Taylor expansion for* φ^* *up to* $O(\|\boldsymbol{z}\|^3)$,

$$\varphi^*(p,q) = \frac{1}{2}\ \mathrm{g}_{ij}(p)\ z^i z^j + \frac{1}{6}\Lambda^*_{ijk}(p)\ z^i z^j z^k + O\left(\|\boldsymbol{z}\|^3\right), \tag{68}$$

where

$$\Lambda^*_{ijk}(p) = 2\Gamma_{ijk}(p) + \overset{*}{\Gamma}_{ijk}(p) \ . \tag{69}$$

Then, by repeating the same argument as in the proof of Theorem III.1 we get

$$\mathrm{g}_{ij}(p) = \partial'_i \partial'_j \varphi^*(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q}, \quad \Gamma_{ijk}(p) = -\,\partial'_i \partial'_j \partial_k \varphi^*(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)\big|_{p=q} \ . \tag{70}$$

B. Local Decomposition of Π and Π^*

In this section we describe the local decomposition of vector fields $\Pi_q(p)$ and $\Pi^*_q(p)$ in terms of gradient vector fields. To this aim, let us consider the representation (54) of $\varphi(p,q) \equiv \varphi_p(q)$. Then we have that,

$$\varphi_p(q) \equiv \left(\varphi_p \circ \overset{*}{\exp}_p\right)(\mathrm{X}^*_p(q)) = \int_0^1 \frac{1}{t} \langle \exp_p^{-1}\left(\overset{*}{\exp}_p\,(t\mathrm{X}^*_p(q))\right), t\mathrm{X}^*_p(q)\rangle_p \,\mathrm{d}t \ , \tag{71}$$

where we have written $q = \overset{*}{\exp}_p\,(\mathrm{X}^*_p(q))$.

Very recently, normal coordinates for manifolds with an affine geometry of general form are constructed [12] under the assumption that all geometric objects are *real* analytic. A very remarkable feature of a normal coordinate system in a neighborhood of any $p \in \mathrm{M}$ is that

$$\mathrm{g}_{ij} = \delta_i^j, \qquad \Gamma_{ijk} \equiv 0,$$

where δ_i^j denotes the delta-Dirac function. Given a dual structure $(\mathrm{g}, \nabla, \nabla^*)$, we can indifferently choose to rely on normal coordinates with respect to ∇ or ∇^* since for our purpose we need just having that $\mathrm{g}_{ij} = \delta_i^j$.

Theorem III.2. *Given a statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$*, let us consider* $p \in \mathrm{M}$ *and* $\mathrm{U}_p \subset \mathrm{M}$ *as in the assumption* (21)*. Then*

$$\Pi_q(p) = \mathrm{grad}_q \varphi_p + V_q, \qquad V_q \in \mathrm{T}_q\mathrm{M} \quad \textit{and} \quad \langle V_q, \dot{\sigma}^*(1)\rangle_q = 0 \ . \tag{72}$$

In addition, decomposition (72) *is unique in* U_p.

Proof. Let us consider a system of normal coordinate $\{\xi^i\}$ in U_p. Then, the local expression of the Riemannian gradient of the divergence function with respect to $\{\xi^i\}$ is given by

$$\mathrm{grad}_q \varphi_p = \frac{\partial \varphi_p}{\partial \xi^i} \partial_i,$$

where $\partial_i = \dfrac{\partial}{\partial \xi^i}$ as usual and the summation over i is intended. By identifying $\mathrm{T}_{\mathrm{X}^*_p(q)}(\mathrm{T}_p\mathrm{M})$ to $\mathrm{T}_p\mathrm{M}$ in the canonical way, we can write $\mathrm{X}^*_p(q) = \xi^i \partial_i$. Hence, by setting

$$\Xi(t\ \mathrm{X}^*_p(q)) = \langle \exp_p^{-1}\left(\overset{*}{\exp}_p\,(t\ \mathrm{X}^*_p(q))\right), t\ \mathrm{X}^*_p(q)\rangle_p$$

we have that

$$\begin{aligned}
\langle \mathrm{X}_p^*(q), \mathrm{grad}_{\mathrm{X}_p^*(q)}\left(\varphi_p \circ \overset{*}{\exp}_p\right)\rangle_p &= \xi^i \frac{\partial}{\xi^i} \int_0^1 \frac{1}{t}\, \Xi(t\ \mathrm{X}_p^*(q))\ \mathrm{d}t \\
&= \xi^i \int_0^1 \frac{\partial}{\partial \xi^i} \Xi(t\ \mathrm{X}_p^*(q))\ \mathrm{d}t \\
&= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} \Xi(t\ \mathrm{X}_p^*(q))\ \mathrm{d}t \\
&= \Xi(\mathrm{X}_p^*(q)) - \Xi(0) = \Xi(\mathrm{X}_p^*(q)),
\end{aligned}$$

because $\Xi(0) = \langle \exp_p(p), 0\rangle_p = 0$. Collecting the last results, we have obtained that

$$\mathrm{grad}_{\mathrm{X}_p^*(q)}\left(\varphi_p \circ \overset{*}{\exp}_p\right) = \mathrm{X}_p(q) + V_p,$$

where $\langle V_p, \mathrm{X}_p^*(q)\rangle_p = 0$ and we used the equivalence

$$\Xi(\mathrm{X}_p^*(q)) \equiv \langle \mathrm{X}_p(q), \mathrm{X}_p^*(q)\rangle_p \ .$$

At this point, we can observe that $(\mathrm{d}\varphi_p)_q = (\mathrm{d}\varphi_p)_q \circ (\mathrm{d}\ \overset{*}{\exp}_p)_{\mathrm{X}_p^*(q)}$. Then, we can perform the following computation,

$$\begin{aligned}
\langle \Pi_q(p), \dot{\sigma}^*(1)\rangle_q = \langle \mathrm{X}_p(q), \mathrm{X}_p^*(q)\rangle_p &= \langle \mathrm{grad}_{\mathrm{X}_p^*(q)}\left(\varphi_p \circ \overset{*}{\exp}_p\right), \mathrm{X}_p^*(q)\rangle_p = \mathrm{d}(\varphi_p \circ \overset{*}{\exp}_p)_{\mathrm{X}_p^*(q)}(\mathrm{X}_p^*(q)) \\
&= (\mathrm{d}\varphi_p)_q\left((\mathrm{d}\ \overset{*}{\exp}_p)_{\mathrm{X}_p^*(q)}(\mathrm{X}_p^*(q))\right) = \langle \mathrm{grad}_q \varphi_p, \dot{\sigma}^*(1)\rangle_q \ ,
\end{aligned}$$

where we employed the well-known equivalence $(\mathrm{d}\ \overset{*}{\exp}_p)_{\mathrm{X}_p^*(q)}(\mathrm{X}_p^*(q)) = \dot{\sigma}^*(1)$ and σ^* is the ∇^*-geodesic from p to q. From the last chain of equalities, we immediately obtain that

$$\Pi_q(p) = \mathrm{grad}_q \varphi_p + V_q$$

being $V_q \in \mathrm{T}_q\mathrm{M}$ uniquely defined by $\Pi_q(p) - \mathrm{grad}_q \varphi_p$ and $\langle V_q, \dot{\sigma}^*(1)\rangle_q = 0$.

In order to prove that decomposition (72) is unique, suppose that there exists another decomposition of $\Pi_q(p)$ satisfying conditions of Theorem III.2, i.e. $\Pi_q(p) = \mathrm{grad}_q \tilde{\varphi}_p(q) + \tilde{V}_q$ with $\langle \tilde{V}_q, \dot{\sigma}^*(1)\rangle_q = 0$. In addition, let us assume that $\tilde{\varphi}_p(p) = 0$. We have then,

$$0 = \mathrm{grad}_q\left(\varphi_p - \tilde{\varphi}_p\right) + V_q - \tilde{V}_q \quad \text{and} \quad \langle \dot{\sigma}^*(1), V_q - \tilde{V}_q\rangle_q = 0.$$

It is evident that

$$\langle \dot{\sigma}^*(1), \mathrm{grad}_q\left(\varphi_p - \tilde{\varphi}_p\right)\rangle_q = 0 \ . \tag{73}$$

By relying again on normal coordinates with respect to ∇, we have that Eq. (73) is a homogeneous first-order differential equation. Then, we can conclude that $\mathrm{grad}_q \varphi = \mathrm{grad}_q \tilde{\varphi}$, because of the assumption $\tilde{\varphi}_p(p) = 0$ and $V_q = \tilde{V}_q$. □

Remark III.2. *Methods in the proof of Theorem III.2 are inspired by [22], where the author presented a decomposition for vector fields on a Riemannian manifold of non-positive curvature with application to non-linear mechanics and irreversible thermodynamics.*

Remark III.3. *Likewise the vector* $\Pi_q(p)$*, we can decompose also the vector* $\Pi^*_q(p)$ *in terms of a gradient vector. This can be trivially carried out by relying on the dual Phi-function* φ^*_p *and repeating same methods of the proof of Theorem III.2. Then, we obtain*

$$\Pi^*_q(p) = \mathrm{grad}_q \varphi^*_p + V^*_q , \qquad V^*_q \in \mathrm{T}_q\mathrm{M} \quad \textit{and} \quad \langle V^*_q, \dot{\sigma}(1)\rangle_q = 0 \; . \tag{74}$$

C. Canonical Divergence

Theorem I.1 suggests the way to single out the appropriate vector field for defining the novel divergence and its dual function. In addition, Theorem III.2 strengthens this choice and we are driven to the following definition.

Definition III.2. *Let us consider the statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ *and* $p \in \mathrm{M}$*. Assume also that there exists* $\mathrm{U}_p \subset \mathrm{M}$ *as in* (21)*. Then, for every* $q \in \mathrm{U}_p$ *we define the function* $\mathcal{D} : \mathrm{M} \to \mathbb{R}$ *by the path integral of* $\Pi_t(p)$ *along the* ∇*-geodesic* $\sigma(t)$ *from* p *to* q*,*

$$\mathcal{D}(p, q) := \mathcal{D}_p(q) := \int_0^1 \langle \Pi_t(p), \dot{\sigma}(t)\rangle_{\sigma(t)} \; \mathrm{d}t \; . \tag{75}$$

The dual function $\mathcal{D}^*(p, q)$ *of* $\mathcal{D}(p, q)$ *is instead defined by the path integral of* $\Pi^*_t(p)$ *along the* ∇^**-geodesic* $\sigma^*(t)$ *from* p *to* q*,*

$$\mathcal{D}^*(p, q) := \mathcal{D}^*_p(q) := \int_0^1 \langle \Pi^*_t(p), \dot{\sigma}^*(t)\rangle_{\sigma^*(t)} \; \mathrm{d}t \; . \tag{76}$$

We refer to $\mathcal{D}(p, q)$ *as the* ***canonical divergence*** *on* M *from* p *to* q*. Analogously, we refer to* $\mathcal{D}^*(p, q)$ *as the* ***dual canonical divergence*** *on* M *from* p *to* q*.*

We have defined our canonical divergence $\mathcal{D}_p(q)$ based on the metric g and the affine connection ∇. In addition, we supplied the ∇^*-connection by the definition of $\Pi_t(p)$. It is then natural to require that this divergence is consistent in the sense that Eq. (3) and Eq. (4) are satisfied. Before addressing this issue, we can observe that from Theorem I.1 and Definitions III.1, III.2 we may obtain the following relations among the pseudo-squared-norm $r_p(q)$, the Phi-functions $\varphi_p(q)$ and $\varphi^*_p(q)$, and the canonical divergences $\mathcal{D}_p(q)$ and $\mathcal{D}^*_p(q)$,

$$r_p(q) = \mathcal{D}_p(q) + \varphi^*_p(q), \qquad r_p(q) = \mathcal{D}^*_p(q) + \varphi_p(q) \; . \tag{77}$$

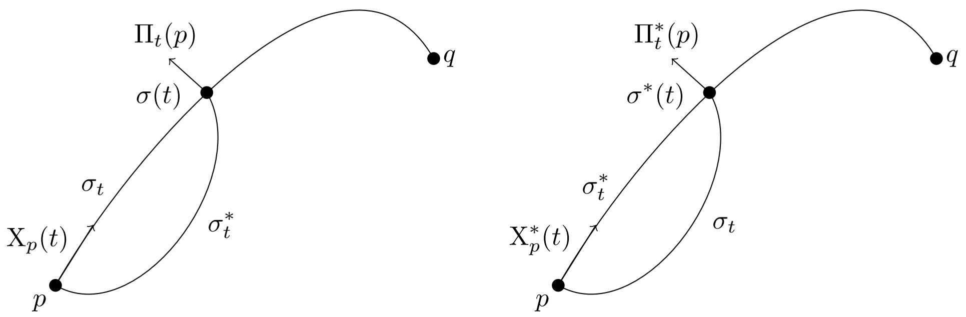

FIG. 3: On the left side of the figure we can see the ∇-geodesic σ connecting p and q and the vector field $\Pi_t(p)$ at $\sigma(t)$. This is obtained by parallel translating with respect to ∇ the vector $\mathrm{X}_p(t) = \exp_p(\sigma(t))$ along the ∇^*-geodesic σ_t^*. On the right side of the figure we can see the ∇^*-geodesic connecting p and q and the vector field $\Pi_t^*(p)$ at $\sigma^*(t)$. This is obtained by parallel translating with respect to ∇^* the vector $\mathrm{X}_p^*(t) = \overset{*}{\exp}_p^{-1}(\sigma^*(t))$ along the ∇-geodesic σ_t.

In fact, since the pseudo-squared-norm $r_p(q)$ is independent of the particular path from p to q, from Eq. (51) we can compute $r_p(q)$ whether along the ∇-geodesic $\sigma(t)$ or along the ∇^*-geodesic $\sigma^*(t)$. Then, by recalling Definitions III.1, III.2 we get relations (77). In addition, by means of decomposition (72) we also have that

$$\begin{aligned}\mathcal{D}_p(q) &= \int_0^1 \langle \Pi_t(p), \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t \\ &= \int_0^1 \langle \mathrm{grad}_t \varphi_p, \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t + \int_0^1 \langle V_t(p), \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t \\ &= \varphi_p(q) + \int_0^1 \langle V_t(p), \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t\ ,\end{aligned} \tag{78}$$

where $V_t = \Pi_t - \mathrm{grad}_t \varphi_p$ and we assumed that $\varphi_p(p) = 0$. Moreover, from decomposition (74) we get

$$\mathcal{D}_p^*(q) = \varphi_p^*(q) + \int_0^1 \langle V_t^*(p), \dot{\sigma}^*(t)\rangle_{\sigma^*(t)}\ \mathrm{d}t\ , \tag{79}$$

where $V_t^* = \Pi_t^* - \mathrm{grad}_t \varphi_p^*$ and we assumed that $\varphi_p^*(p) = 0$. By combining (77) and (78), (79) we trivially obtain that

$$\int_0^1 \langle V_t(p), \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t = \int_0^1 \langle V_t^*(p), \dot{\sigma}^*(t)\rangle_{\sigma^*(t)}\ \mathrm{d}t\ . \tag{80}$$

Let us now step back to the issue of $\mathcal{D}$-consistency with the geometry of the statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. Since the geometry is determined by the derivatives of $\mathcal{D}(\boldsymbol{\xi}_p, \boldsymbol{\xi}_q)$ at $p = q$, we consider the case where p and q are close to each other, that is $z^i = \xi_q^i - \xi_p^i$ is small for all i. We then evaluate the canonical divergence by Taylor expansion up to $O\left(\|\boldsymbol{z}\|^3\right)$.

Proposition III.2. *When $\|z\| = \|\boldsymbol{\xi}_q - \boldsymbol{\xi}_p\|$ is small, the canonical divergence $\mathcal{D}$ is expanded as*

$$\mathcal{D}(p,q) = \frac{1}{2}\mathrm{g}_{ij}(p) z^i z^j + \frac{1}{6}\Lambda_{ijk}(p) z^i z^j z^k + O(\|\boldsymbol{z}\|^3) \ , \tag{81}$$

where $\Lambda_{ijk}(p) = 2 \stackrel{*}{\Gamma}_{ijjk} + \Gamma_{ijk}$.

Proof. By looking at Eq. (75) we need to Taylor expand with respect to the local coordinate $\{\boldsymbol{\xi}\}$ the following factors

$$\mathrm{g}_{ij}(\sigma(t)), \quad \frac{\mathrm{d}}{\mathrm{d}t}\sigma^i(t), \quad \mathrm{P}^j_{\sigma_t^*}\left(\mathrm{X}_p(t)\right),$$

where $\mathrm{X}_p(t) = \frac{\mathrm{d}}{\mathrm{d}t}\sigma_t(0)$ and $\mathrm{P}^j_{\sigma_t^*}$ is the jth component of the parallel transport with respect to ∇-connection. Here, $\sigma_t(s)$ is the ∇-geodesic from p to $\sigma(t)$. The Taylor expansion of the metric tensor is given by

$$\mathrm{g}_{ij}(\sigma(t)) = \mathrm{g}_{ij}(p) + t\partial_k \mathrm{g}_{ij}(p) z^k + O(\|\boldsymbol{z}\|^2), \tag{82}$$

where $\partial_k = \frac{\partial}{\partial \xi_p^k}$. Consider now the local coordinates $\boldsymbol{\xi}(t)$ of the geodesic $\sigma(t)$. By Taylor expanding it, we obtain

$$\xi^i(t) = \xi_p^i + tz^i + \frac{t}{2}(1-t)\Gamma^i_{\mu\nu}(p) z^\mu z^\nu + O\left(\|\boldsymbol{z}\|^3\right), \tag{83}$$

where the summation over μ and ν is intended. Then we obtain,

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma^i(t) = z^i + \frac{1}{2}(1-2t)\Gamma^i_{\mu\nu}(p) z^\mu z^\nu + O\left(\|\boldsymbol{z}\|^3\right). \tag{84}$$

Consider now the ∇-geodesic $\sigma_t(s)$. From Eq. (84) we obtain the following expression for $\mathrm{X}_p^j(t)$,

$$\mathrm{X}_p^j(t) = \frac{\mathrm{d}}{\mathrm{d}t}\sigma_t(0) = \xi^j(t) - \xi_p^j + \frac{1}{2}\Gamma^j_{\mu\nu}(\sigma(t))(\xi^\mu(t) - \xi_p^\mu)(\xi^\nu(t) - \xi_p^\nu).$$

In addition, we have that

$$\xi^j(t) - \xi_p^j = tz^j + \frac{t}{2}(1-t)\Gamma^j_{\mu\nu}(p) z^\mu z^\nu.$$

Then, we arrive at

$$\mathrm{X}_p^j(t) = tz^j + \frac{t}{2}\Gamma^j_{\mu\nu}(p) z^\mu z^\nu. \tag{85}$$

In the end, recalling that ∇^*-geodesic σ_t^* connects p and $\sigma(t)$, we use the following Taylor expansion of the parallel transport with respect to ∇ along σ_t^* [7],

$$\mathrm{P}^j_{\sigma_t^*}(\mathrm{X}_p(t)) = \mathrm{X}^j_p(t) - \Gamma^j_{\mu\nu}(p)(\mathrm{X}^\mu_p(t))(\sigma^\nu(t) - \xi^\nu_p)$$

and from Eq. (85) we obtain

$$\mathrm{P}^j_{\sigma_t^*}(\mathrm{X}_p(t)) = tz^j + \frac{t}{2}(1-2t)\Gamma^j_{\mu\nu}(p)z^\mu z^\nu + O\left(\|\boldsymbol{z}\|^3\right). \tag{86}$$

We are now ready to provide the Taylor series of the path integral Eq. (75). By collecting Eqs. (82), (84) and (86) we obtain the following expression for $\mathcal{D}(p,q)$,

$$\begin{aligned}\mathcal{D}(p,q) = \int_0^1 dt &\left[\mathrm{g}_{ij}(p) + t\partial_k \mathrm{g}_{ij}(p)z^k + O(\|\boldsymbol{z}\|^2)\right] \\ \times &\left[z^i + \frac{1}{2}(1-2t)\Gamma^i_{\mu\nu}(p)z^\mu z^\nu + O\left(\|\boldsymbol{z}\|^3\right)\right] \\ \times &\left[tz^j + \frac{t}{2}(1-2t)\Gamma^j_{\mu\nu}(p)z^\mu z^\nu + O\left(\|\boldsymbol{z}\|^3\right)\right] \,\mathrm{d}t.\end{aligned} \tag{87}$$

Finally, by computing this integral up to $O\left(\|\boldsymbol{z}\|^4\right)$ and recalling the relation $\Gamma_{ijk} = g_{li}\Gamma^l_{jk}$ we arrive at

$$\begin{aligned}\mathcal{D}(p,q) &= \frac{1}{2}\mathrm{g}_{ij}(p)z^iz^j + \frac{1}{3}\partial_k\mathrm{g}_{ij}(p)z^iz^jz^k - \frac{1}{6}\Gamma_{ijk}(p)z^iz^jz^k \\ &= \frac{1}{2}\mathrm{g}_{ij}(p)z^iz^j + \frac{1}{6}\left(2\partial_k\mathrm{g}_{ij}(p)z^iz^jz^k - \Gamma_{ijk}(p)z^iz^jz^k\right).\end{aligned}$$

Now, from the relation $\partial_k \mathrm{g}_{ij} = \Gamma_{ijk} + \Gamma^*_{ijk}$, we obtain

$$\mathcal{D}(p,q) = \frac{1}{2}\mathrm{g}_{ij}(q)z^iz^j + \frac{1}{6}\left(2\Gamma^*_{ijk} + \Gamma_{ijk}\right). \tag{88}$$

Eq. (88) can be reduced to Eq. (81) by using Eq. (58). □

Remark III.4. *By comparing Prop. III.2 with Prop. III.1 we can see that the Phi-function* φ_p *and the canonical divergence* $\mathcal{D}_p$ *coincide in a neighborhood of* p *up to* $O(\|\boldsymbol{z}\|^3)$*. This immediately leads the canonical divergence* $\mathcal{D}(p,q)$ *to satisfying the consistency with the geometry of* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ *by means of Theorem III.1. The same argument can be used for* $\mathcal{D}^*(p,q)$*. By interchanging the role of* ∇ *and* ∇^* *in* (81) *and by Remark III.1, it can be trivially proved that* $\mathcal{D}^*(p,q)$ *generates the dual structure of* M *in the same way as stated by Eq.* (70)*.*

Eguchi introduced in [9] the concept of the *contrast function* in order to construct statistical structures on a given manifold M. A contrast function $\rho(p,q)$ is defined everywhere on $\mathrm{M}\times\mathrm{M}$. For a function $\rho(p,q)$ to be a contrast function, it is required that

$$\rho(p,q) \geq 0, \qquad \rho(p,q) = 0 \iff p = q\ ,$$

and

$$\partial_i \partial_j\ \rho(p,q)\big|_{q=p} = -\partial_i \partial'_j\ \rho(p,q)\Big|_{q=p} = \mathrm{g}_{ij}(p)$$

is strictly positive definite on M. If $\rho(p,q)$ is a contrast function,

$$\Gamma^{\rho}_{ijk}(p) := -\partial_i \partial_j \partial'_k \rho(p,q)\Big|_{q=p}\,, \qquad \Gamma^{\rho^*}_{ijk}(p) := -\partial'_i \partial'_j \partial_k \rho(p,q)\Big|_{q=p}$$

define torsion free affine dual connections with respect to the Riemannian metric g. Now, the purpose of the present article is to recover a given dual structure $(\mathrm{g}, \nabla, \nabla^*)$ on a manifold M by means of the divergence function since our investigation has been addressed from the very beginning to the inverse problem. In order to pursue this aim, it is enough to consider a contrast function to be defined in a neighborhood of the diagonal set of $\mathrm{M} \times \mathrm{M}$. Then we have the following result.

Theorem III.3. *Consider a statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. *Then the canonical divergence* $\mathcal{D}(p,q)$ *and the Phi-function* $\varphi(p,q)$ *are both non-negative in a neighborhood of the diagonal set* Δ *of* $\mathrm{M}\times\mathrm{M}$ *and vanish only on* Δ. *Furthermore, they both induce the dual structure* $(\mathrm{g}, \nabla, \nabla^*)$ *of* M.

Proof. By means of Eq. (81) we obtain that, if p and q are sufficiently close to each other, then

$$\mathcal{D}(p,q) \geq 0, \qquad \mathcal{D}(p,q) = 0 \iff p = q\ .$$

By Remark III.4 we also know that $\mathcal{D}(p,q)$ generates the dual structure of M. These properties ensure that the canonical divergence $\mathcal{D}(p,q)$ is a contrast function on M.

Analogously, by Pro. III.1 and, in particularly, from Eq. (57) we have that

$$\varphi(p,q) \geq 0, \qquad \varphi(p,q) = 0 \iff p = q\ ,$$

when p and q are sufficiently close each to other, as well. Furthermore, by Theorem III.1 also φ induces the dual structure $(\mathrm{g}, \nabla, \nabla^*)$. Then, $\varphi(p,q)$ is a contrast function on M, too. □

Although Theorem III.3 holds when p and q are sufficiently close to each other, a contrast function can be defined everywhere on $\mathrm{M} \times \mathrm{M}$ by means of an appropriate positive function. However, we will prove that our canonical divergence $\mathcal{D}(p,q)$ turns out to be positive under sufficient conditions whenever p and q can be connected by unique ∇-geodesic and unique ∇^*-geodesic (see Pro. IV.1).

From Theorem III.3 we know that both the functions, the canonical divergence $\mathcal{D}$ and the Phi-function φ, generate the same dual structure on M. This leads to a very close link between them. In order to prove this connection, we need the following

Lemma III.2. *Let* $\widetilde{\mathrm{M}} \subset \mathrm{M}$ *be a smooth submanifold of* M *of codimension* ≥ 1*. Let* $p \in \mathrm{M}\backslash\widetilde{\mathrm{M}}$*. Then the function*

$$\varphi_p : \widetilde{\mathrm{M}} \to \mathbb{R}^+, \quad q \mapsto \varphi_p(q)$$

takes its (local) minimum whenever $q \in \mathrm{U}_p$*, where* U_p *is as in Eq.* (21)*.*

Proof. The proof trivially follows by the decomposition of $\Pi_q(p)$ given by Theorem III.2 and by the definition (52) of $\varphi_p(q)$. □

Proposition III.3. *Consider the canonical divergence* $\mathcal{D}(p,q)$ *and the Phi-function* $\varphi(p,q)$ *of a given statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$*. Then, there exists a function* $f : [0, K] \to \mathbb{R}^+$ $(K > 0)$ *with* $f(0) = 0$ *and* $f'(0) > 0$*, such that*

$$\mathcal{D}_p(q) = f\left(\varphi_p(q)\right) \ . \tag{89}$$

Proof. By means of definition (9), we can consider the functional

$$\Omega(\gamma) := \int_0^1 \langle \Pi_t(p), \dot{\gamma}(t)\rangle_{\gamma(t)}\ \mathrm{d}t, \tag{90}$$

where γ is an arbitrary path from p to q. By Lemma III.2 it takes a local minimum when $\gamma \equiv \sigma^*$, where σ^* is the ∇^*-geodesic from p to q. Consider now the level hypersuface of $\mathcal{D}$ given by $\mathrm{H}_{\mathcal{D}} = \{q \in \mathrm{M} \mid \mathcal{D}_p(q) = \kappa\}$. For a point $q \in \mathrm{H}_{\mathcal{D}}$ we define

$$L_q := \left\{p \in \mathrm{M} \mid \mathcal{D}_p(q) = \min_{q' \in \mathrm{H}_{\mathcal{D}}} \mathcal{D}_p(q')\right\} \ .$$

According to the theory of *minimum contrast geometry* by Eguchi [9], we know that $\{L_q\}_{q\in\mathrm{H}_{\mathcal{D}}}$ is (locally) a foliation of M with 1-dimensional leaves such that

(i) each leaf L_q is orthogonal to $\mathrm{H}_{\mathcal{D}}$ at q,

(ii) the second fundamental form with respect to ∇ of L_q is zero at q.

From Theorem III.3 we have that $\mathcal{D}(p,q) \geq 0$ for p and q suitably close to each other. Therefore, by the uniqueness of the decomposition (72) it is clear that level hypersurfaces of $\mathcal{D}$ and φ coincide as families and in particular they have identical gradient flows. This implies that there exists a monotonic function f with $f(0) = 0$ such that $\mathcal{D}_p(q) = f\left(\varphi_p(q)\right)$. □

Proposition III.4. *For a given statistical manifold* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$*, let us consider the Phi-functions* $\varphi_q(p)$ *and* $\varphi^*_p(q)$*. Both are contrast functions for* M*. In addition, there exists a function* $h : [0, K] \to \mathbb{R}^+$ $(K > 0)$ *with* $h(0) = 0$ *and* $h'(0) > 0$*, such that*

$$\varphi_q(p) = h\left(\varphi^*_p(q)\right) \ . \tag{91}$$

Proof. The *Phi-function* $\varphi(q,p) = \varphi_q(p)$ is obtained from (52) by interchanging the role of p and q. Therefore, from Eq. (57) we obtain that the Taylor expansion of $\varphi_q(p)$ reads as follows,

$$\varphi_q(p) = \frac{1}{2}\ \mathrm{g}_{ij}(q)\zeta^i\zeta^j + \frac{1}{6}\ \Lambda_{ijk}(q)\zeta^i\zeta^j\zeta^k + O\left(\|\zeta\|^3\right), \quad \Lambda_{ijk}(q) = 2{\Gamma^*}_{ijk}(q) + \Gamma_{ijk}(q), \quad \zeta = \boldsymbol{\xi}_p - \boldsymbol{\xi}_q \ .$$

By repeating the argument in the proof of Theorem III.1, we obtain that

$$\partial_i\partial_j\partial'_k\varphi_q(p)\big|_{p=q} = -\Gamma^*_{ijk}(p) \ ,$$

where $\partial_i = \dfrac{\partial}{\partial \xi^i_p}$ and $\partial'_i = \dfrac{\partial}{\partial \xi^i_q}$. Now, by comparing this result with the Remark III.1 we conclude that $\varphi_q(p)$ generates the dual structure of M in the same way of the dual *Phi-function* $\varphi^*_p(q)$.

Given the level hypersurface $\mathrm{H}_\varphi = \{q \in \mathrm{M} \mid \varphi_q(p) = \kappa\}$ we can define for a point $q \in \mathrm{H}_\varphi$

$$L^*_q := \left\{p \in \mathrm{M} \mid \varphi_q(p) = \min_{q' \in \mathrm{H}_\varphi} \varphi_{q'}(p)\right\} \ .$$

Again, we know that $\{L^*_q\}_{q\in\mathrm{H}_\varphi}$ is (locally) a foliation of M with 1-dimensional leaves such that

(i) each leaf L^*_q is orthogonal to H_φ at q,

(ii) the second fundamental form with respect to ∇^* of L^*_q is zero at q.

Therefore, the local minimum of φ_q is obtained by integrating Π^*_t along a ∇-geodesic. This implies that level hypersurfaces of φ^*_p and φ_q coincide as families and in particular they have identical gradient flows. For this reason, there exists a monotonic function h with $h(0) = 0$ such that $\varphi_q(p) = h\left(\varphi^*_p(q)\right)$. □

In general, the canonical divergence $\mathcal{D}(p,q)$ is not symmetric, i.e. $\mathcal{D}(p,q) \neq \mathcal{D}(q,p)$. However, Proposition III.3 and Proposition III.4 suggest the following symmetry property.

Theorem III.4. *Given* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ *a statistical manifold, let us consider the canonical divergences* $\mathcal{D}(q,p)$ *and* $\mathcal{D}^*(p,q)$ *that, both, generate the dual structure* $(\mathrm{g}, \nabla, \nabla^*)$ *of* M*. Then, there exists a function* $\Upsilon : [0, K] \to \mathbb{R}^+$ *satisfying the conditions* $\Upsilon(0) = 0, \quad \Upsilon'(0) > 0$ *such that*

$$\mathcal{D}(q,p) = \Upsilon(\mathcal{D}^*(p,q)) \ . \tag{92}$$

Proof. From the symmetry of the pseudo-squared-norm $r_p(q)$ in its arguments we have that

$$\mathcal{D}_q(p) + \varphi^*_q(p) = \mathcal{D}^*_p(q) + \varphi_p(q) \ .$$

This implies that $\mathcal{D}_q(p) = \mathcal{D}^*_p(q) + \varphi_p(q) - \varphi^*_q(p)$. Now by Eq. (89) and Eq. (91) we can write

$$\mathcal{D}_q(p) = (f^* + Id - h^{-1})\varphi_p(q) \ ,$$

where f^* is the dual of f in the sense that it provides the same relation as Eq. (89) but between φ^* and $\mathcal{D}^*$. Moreover, Id is the identity map and h^{-1} is the inverse of the map h in Eq. (91). Finally, we get statement (92) by defining $\Upsilon := f^* + Id - h^{-1}$. □

IV. COMPARISON WITH THE DIVERGENCE OF AY AND AMARI

In this section we compare the canonical divergence defined by Eq. (75) with respect to the one proposed by Ay and Amari in [5]. Recall that the latter has been defined by path integration of the vector field $\mathrm{X}_t(q) = \exp^{-1}_{\sigma(t)}(q)$ along the ∇-geodesic $\sigma(t)$. In particular, when the ∇-geodesic σ goes from p to q, the divergence of Ay and Amari assumes the nice form (8), i.e.

$$D(p,q) = \int_0^1 t\|\dot{\sigma}(t)\|^2 \ \mathrm{d}t \ .$$

In order to carry out this comparison, let us consider for each $t \in [0,1]$ the loop Σ_t based at p and passing by $\sigma(t)$. This is defined as follows,

$$\Sigma_t(s) = \begin{cases} \sigma^*_t(2s), & s \in [0,1/2] \\ \\ \sigma_t(2-2s), & s \in [1/2,1] \end{cases} , \tag{93}$$

where ∇ and ∇^* geodesics σ_t and σ^*_t go from p to $\sigma(t)$. By means of Lemma A.1 in [Appendix A] we know that, if Σ_t lies in a sufficiently small neighborhood of p, then

$$\mathrm{P}_{\Sigma_t}\mathrm{X}_p(t) = \mathrm{X}_p(t) + \mathcal{R}_{\Sigma_t}\left(\mathrm{X}^*_p(t), \mathrm{X}_p(t)\right), \tag{94}$$

where

$$\mathcal{R}_{\Sigma_t}\left(\mathrm{X}^*_p(t), \mathrm{X}_p(t)\right) := \int_{B_t} \frac{\mathrm{P}\left[\mathcal{R}\left(\mathrm{X}^*(t), \mathrm{X}(t)\right)\mathrm{X}(t)\right]}{\|\mathrm{X}^*_p(t) \wedge \mathrm{X}_p(t)\|} \ \mathrm{d}A \tag{95}$$

with $\mathrm{X}^*(t)$ and $\mathrm{X}(t)$ being the parallel transport with respect to ∇ of $\mathrm{X}^*_p(t)$ and $\mathrm{X}_p(t)$, respectively, from p to each point of B_t along the unique ∇-geodesic joining them. Here, $\mathcal{R}$ is the curvature tensor [Appendix A] of ∇, B_t denotes the disk defined by the curve Σ_t and $\mathrm{X}_p(t) = \overset{*}{\exp}_p^{-1}(\sigma(t))$, $\mathrm{X}_p(t) = \exp_p^{-1}(\sigma(t))$ are linearly independent on B_t. In addition, P within the integral denotes the parallel translation from each point in B_t to p along the unique ∇-geodesic segment joining them. Now, we can write P_{Σ_t} as the parallel transport with respect to ∇ along σ_t^* and along σ_t, but in the reversed direction. In particular we have that

$$\mathrm{P}_{\Sigma_t}\mathrm{X}_p(t) = \left(\mathrm{P}_{\sigma_t}^{-1} \circ \mathrm{P}_{\sigma_t^*}\right)\mathrm{X}_p(t) \ .$$

Then, from Eq. (94) and by the definition of vector $\Pi_t(p)$, we obtain

$$\begin{aligned} \Pi_t(p) &= \mathrm{P}_{\sigma_t^*}\mathrm{X}_p(t) \\ &= \mathrm{P}_{\sigma_t}\mathrm{X}_p(t) + \ \mathrm{P}_{\sigma_t}\left[\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\right] \ . \end{aligned} \tag{96}$$

Let us recall that the canonical divergence $\mathcal{D}_p$ is defined by means of the path integration of the vector field $\Pi_t(p)$ along the ∇-geodesic $\sigma(t)$ from p to q. Since $\mathrm{X}_p(t) = \exp_p^{-1}(\sigma(t))$ is the velocity vector at p of the ∇-geodesic σ_t, we have that

$$\mathrm{P}_{\sigma_t}\mathrm{X}_p(t) = \dot{\sigma}_t(1) = t\ \dot{\sigma}(t) \ . \tag{97}$$

Therefore, by Eq. (75) we obtain

$$\mathcal{D}(p,q) = \int_0^1 t\ \|\dot{\sigma}(t)\|^2\ \mathrm{d}t + \int_0^1 \ \langle \mathrm{P}_{\sigma_t}\left[\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\right], \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t \ . \tag{98}$$

The decomposition of $\mathcal{D}(p,q)$ given by Eq. (98) allows us to provide sufficient conditions for the positivity of $\mathcal{D}(p,q)$ whenever exist unique ∇- geodesic and unique ∇^*-geodesic both connecting p and q.

Proposition IV.1. *Consider $p \in \mathrm{M}$ and a neighborhood $\mathrm{U}_p \subset \mathrm{M}$ of p as the one in (21). Let us assume the following conditions on the Riemannian curvature tensor R [Appendix A],*

$$\begin{aligned} &(i) \qquad \nabla\ \mathrm{R} \equiv 0 \\ &(ii) \qquad \mathrm{R}\left(X,Y,Y,Y\right) \geq 0 \quad \forall\ X,\ Y \in \mathcal{T}(\mathrm{M}) \ . \end{aligned} \tag{99}$$

Then, we have that

$$\mathcal{D}(p,q) \geq 0 \quad \forall\ q \in \mathrm{U}_p, \qquad \mathcal{D}(p,q) = 0 \iff p = q \ . \tag{100}$$

Proof. In order to prove this statement, let us decompose the canonical divergence $\mathcal{D}(p,q)$ according to Eq. (98). By $\nabla\mathrm{R} \equiv 0$ we know that the curvature tensor is invariant under all parallel translations with respect to ∇-connection [10]. Therefore, by Eq. (95) and by recalling the definition of P below Eq. (95) we obtain

$$\begin{aligned}
\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right) &= \int_{B_t} \frac{\mathrm{P}\left[\mathcal{R}\left(\mathrm{X}^*(t), \mathrm{X}(t)\right)\mathrm{X}(t)\right]}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \\
&= \int_{B_t} \frac{\mathcal{R}\left(\mathrm{PX}^*(t), \mathrm{PX}(t)\right)\mathrm{PX}(t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \\
&= \int_{B_t} \frac{\mathcal{R}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\mathrm{X}_p(t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \\
&= \varepsilon_t\, \mathcal{R}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\mathrm{X}_p(t)\ , \qquad (101)
\end{aligned}$$

where

$$\varepsilon_t := \frac{\mathrm{Area}(B_t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\ .$$

Moreover, from Eq. (97) we have that

$$\begin{aligned}
\int_0^1 \langle \mathrm{P}_{\sigma_t}\left[\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\right], \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t &= \int_0^1 \frac{\varepsilon_t}{t}\ \mathrm{R}\left(\mathrm{P}_{\sigma_t}\mathrm{X}_p^*(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t)\right)\ \mathrm{d}t \\
&\geq 0
\end{aligned}$$

because of Condition (ii) in Eq. (99). Finally, from Eq. (98) we arrive at $\mathcal{D}(p,q) \geq 0$ for all $q \in \mathrm{U}_p$ with $\mathcal{D}(p,q) = 0$ iff $p = q$. □

By replacing ∇ and R in Eq. (99) with ∇^* and R^*, respectively, we obviously obtain that $\mathcal{D}^*(p,q) \geq 0$ for all $q \in \mathrm{U}_p$ with $\mathcal{D}^*(p,q) = 0$ iff $p = q$, as well.

A. Divergence in self dual manifolds

A statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is self-dual when $\nabla = \nabla^*$. In this case $\mathcal{S}$ reduces to a Riemannian manifold endowed with the Levi-Civita connection. Indeed, it is well-known that [1],

$$\overline{\nabla}_{\mathrm{LC}} = \frac{1}{2}\left(\nabla + \nabla^*\right).$$

Since $\nabla = \nabla^*$ then ∇ and ∇^* geodesics coincide. For this reason we have

$$\mathrm{X}_p(t) = \mathrm{X}_p^*(t), \quad \forall\ t \in [0,1]\ .$$

Indeed, in our approach $\mathrm{X}_p(t)$ and $\mathrm{X}_p^*(t)$ are tangent vectors at p of ∇ and ∇^* geodesics, respectively. Therefore, by recalling the definition (95) of $\mathcal{R}_{\Sigma_t}$ we obtain that the second term of the right hand side of Eq. (98) reduces to zero. This trivially follows by the skew-symmetry of the curvature tensor that implies $\mathcal{R}(\mathrm{X}(t), \mathrm{X}(t)) \equiv 0$ for all $t \in [0,1]$. Therefore, we have that.

$$\int_0^1 \langle \mathrm{P}_{\sigma_t}\left[\mathcal{R}_{\Sigma_t}\left(\mathrm{X}(t), \mathrm{X}(t)\right)\right], \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t\ \equiv 0$$

where now $\mathrm{X}(t)$ is the parallel transport of the tangent vector $\mathrm{X}_p(t)$ at p of $\overline{\nabla}_{\mathrm{LC}}$-geodesic $\sigma(t)$ and P denotes the parallel transport with respect to $\overline{\nabla}_{\mathrm{LC}}$-connection. This leads to

$$\mathcal{D}(p,q) = \int_0^1 t\ \|\dot{\sigma}(t)\|^2\ \mathrm{d}t \tag{102}$$

and proves that our divergence coincides with the one of Ay and Amari in the case of self-dual manifold.

In addition, we know from classical Riemannian Geometry that the term $\|\dot{\sigma}(t)\|^2$ is constant with respect to the parameter t. Then, we can conclude that the novel canonical divergence corresponds to the energy of the $\overline{\nabla}_{\mathrm{LC}}$-geodesic $\sigma(t)$ from p to q, that is

$$\mathcal{D}(p,q) = \frac{1}{2} d^2(p,q)\ ,$$

where $d(p,q)$ is the Riemannian distance between p and q.

B. Divergence in dually flat manifolds

The statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is called dually flat when the curvature tensors of ∇ and ∇^* are zero, i.e. $\mathcal{R} = \mathcal{R}^* \equiv 0$. Then, since $\mathcal{R} \equiv 0$ implies that $\mathcal{R}_{\Sigma_t} \equiv 0$ we have that in this particular case Eq. (98) reduces to

$$\mathcal{D}(p,q) = \int_0^1 t\ \|\dot{\sigma}(t)\|^2\ \mathrm{d}t\ .$$

This proves that also in case of dually flat manifold our divergence coincides with the one of Ay and Amari.

In a dually flat manifold we can rely on an affine coordinate system $\theta = (\theta^1, \ldots, \theta^n)$ and a potential function $\phi(\theta)$. In addition, the dual affine coordinates $\eta = (\eta_1, \ldots, \eta_n)$ are given by

$$\eta_i = \frac{\partial \phi(\theta)}{\partial \theta^i}, \qquad i = 1, \ldots, n\ .$$

The dual potential is then defined as

$$\phi^*(\eta) = \phi(\theta) - \theta \cdot \eta$$

where $\theta \cdot \eta = \theta^i\ \eta_i$ and θ is a function of η. Since by definition of affine coordinate system we have that $\Gamma_{ijk}(\theta) \equiv 0$, then the geodesic connecting p and q assumes the form

$$\theta(t) = \theta_p + t(\theta_q - \theta_p)\ .$$

Hence, the velocity is constant

$$\dot{\theta} = z = \theta_q - \theta_p \ .$$

The novel canonical divergence from p to q is defined by

$$\mathcal{D}(p,q) = \int_0^1 t\ \mathrm{g}_{ij}(\theta(t))\ z^i z^j\ \mathrm{d}t\ .$$

Since $\mathrm{g}_{ij} = \partial_i \partial_j \phi$, we have

$$\begin{aligned}\mathcal{D}(p,q) &= \int_0^1 t\ \partial_i\partial_j\phi(\theta_p + tz) z^i z^j\ \mathrm{d}t \\ &= \int_0^1 t\ \ddot{\phi}(\theta(t))\ \mathrm{d}t \\ &= -\int_0^1 \dot{\phi}(\theta(t))\ \mathrm{d}t + \Big[t\dot{\phi}(\theta(t))\Big]_0^1 \\ &= \phi(\theta_p) + \phi^*(\eta_q) - \theta_p \cdot \eta_q\ .\end{aligned}$$

This shows that our divergence is the same as the canonical divergence defined in terms of the Bregman divergence of M.

By considering the dual divergence $\mathcal{D}^*(p,q)$, Eq. (98) assumes in the dually flat case the following form,

$$\mathcal{D}^*(p,q) = \int_0^1 t\ \|\dot{\sigma}^*(t)\|^2\ \mathrm{d}t\ ,$$

that corresponds to the dual version of the divergence of Ay and Amari [5]. By using the dual affine coordinates $\{\eta\}$ and the dual potential function $\phi^*(\eta)$ we obtain the expression of $\mathcal{D}^*(p,q)$ as Bregman divergence of M which proves that $\mathcal{D}_q(p) = \mathcal{D}^*_p(q)$.

Let us now consider the case of α-connections within the class of conjugate symmetric manifolds [Appendix A]. Recall that a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is conjugate symmetric when the curvature tensors of ∇ and ∇^* are the same, i.e. $\mathcal{R} \equiv \mathcal{R}^*$. A very remarkable relation between the curvature tensor of $\overset{\alpha}{\nabla}$ (see [Appendix A]) and $\mathcal{R}$ is provided in [25] and for conjugate symmetric manifolds reads as follows,

$$\overset{\alpha}{\mathcal{R}}(X,Y)Z = \frac{1}{2}\ \mathcal{R}(X,Y)Z + \frac{1-\alpha^2}{2}\Big(\widetilde{T}(Y, \widetilde{T}(X,Z)) - \widetilde{T}(X, \widetilde{T}(Y,Z)\Big) \tag{103}$$

for all $X, Y, Z \in \mathcal{T}(\mathrm{M})$. Here, $\widetilde{T}(X,Y)$ is the "difference tensor" and it is defined by [Appendix A]

$$\widetilde{T}(X,Y) = \nabla_X Y - \nabla^*_X Y\ . \tag{104}$$

Therefore, if $\mathcal{R} = \mathcal{R}^* \equiv 0$, then the α curvature tensor is given in terms of the difference tensor $\widetilde{T}$. Now, by resorting to the affine local coordinate $\{\theta\}$ of ∇-connection, we have that $\Gamma_{ijk}(\theta) \equiv 0$. Hence, from Eq. (103) we have that the local expression of α curvature is given by

$$\overset{\alpha}{\mathcal{R}}{}_{lij}^{\,k} = \frac{1-\alpha^2}{2}\left[{}^*\Gamma_{im}^{k}\,{}^*\Gamma_{jl}^{m} - {}^*\Gamma_{il}^{m}\,{}^*\Gamma_{jm}^{k}\right] = \frac{1-\alpha^2}{2}\,\overset{*}{\mathcal{R}}{}_{jil}^{\,k}\ ,$$

which is zero because $\overset{*}{\mathcal{R}} \equiv 0$. As a result, from Eq. (95) and Eq. (98) we obtain the following expression for the α-divergence,

$$\mathcal{D}^{\alpha}(p,q) = \int_0^1 t\|\dot{\sigma}^{\alpha}(t)\|^2\ \mathrm{d}t, \tag{105}$$

where $\sigma^{\alpha}(t)$ is the $\overset{\alpha}{\nabla}$-geodesic from p to q. This proves that also in the case of α-connection our divergence coincides with the one of Ay and Amari.

C. Divergence of Henmi and Kobayashi

In this section we address our investigation to the divergence proposed by Henmi and Kobayashi in [11]. Given a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$, the authors have considered the function

$$W^*(p\|q) := -\int_0^1 \left[g_{ij}(\bar{\sigma}^*(t))\frac{\mathrm{d}}{\mathrm{d}t}\bar{\sigma}^{*i}(t)\frac{\mathrm{d}}{\mathrm{d}s}\bar{\sigma}_t^j(0)\right]\ \mathrm{d}t\ , \tag{106}$$

where $\bar{\sigma}^*(t)$ is the ∇^*-geodesic from q to p and $\bar{\sigma}_t(s)$ is the ∇-geodesic from $\bar{\sigma}^*(t)$ to q. The vector field

$$\mathrm{F}(\bar{\sigma}^*(t)) := \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \bar{\sigma}_t(s) \tag{107}$$

has been interpreted in [11] as the force field (*the stress*) that is obtained by applying *Hook's law* to the ∇-geodesic $\bar{\sigma}_t(s)$ at each $\bar{\sigma}^*(t) = \bar{\sigma}_t(0)$. The ∇-geodesic $\bar{\sigma}_t(s)$ is here understood as a *spring* which is stretched by F from the *equilibrium* state q. By means of this interpretation, the function $W^*(p\|q)$ is the *work* which is necessary to move a point of unit mass from q to p along the ∇^*-geodesic $\bar{\sigma}^*(t)$ against the force field $\mathrm{F}(\sigma^*)(t)$.

In [11], the authors showed that if

$$\begin{aligned} &(i) \qquad \mathrm{R}(X,Y,Y,Y) = 0 \quad \forall\ X,Y \in \mathcal{T}(\mathrm{M}) \\ &(ii) \qquad \nabla\ \mathrm{R} = 0, \end{aligned} \tag{108}$$

are satisfied, then each ∇-geodesic emanating from q is perpendicular to the level-hyepersurface of the functional $W_*(\cdot\|q)$, where

$$W_*(p\|q) := -\int_0^1 \left[g_{ij}(\sigma(t))\frac{\mathrm{d}}{\mathrm{d}t}\sigma^i(t)\frac{\mathrm{d}}{\mathrm{d}s}\bar{\sigma}_t^j(0)\right]\ dt \tag{109}$$

and the ∇-geodesic $\bar{\sigma}(t)$ substitutes the path $\bar{\sigma}^*(t)$ connecting q and p. In (108) R denotes the Riemannian curvature tensor (see [Appendix A]). In this way they proved that their divergence function is independent of the particular path from q to p. In addition, they proved that the integral curves of grad $W_*(\cdot\|q)$ coincides with the ∇-geodesics starting from q. For the rest of this subsection we assume that conditions (108) hold.

Let us now consider the effects that conditions (108) have on our divergence. Firstly, consider Eq. (95). Since now $\nabla\mathcal{R} \equiv 0$, we have that the curvature tensor is invariant under all parallel translations with respect to the ∇-connection. From this analysis it follows that

$$\begin{aligned}\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right) &= \int_{B_t} \frac{\mathrm{P}\left[\mathcal{R}\left(\mathrm{X}^*(t), \mathrm{X}(t)\right)\mathrm{X}(t)\right]}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \\ &= \int_{B_t} \frac{\mathcal{R}\left(\mathrm{PX}^*(t), \mathrm{PX}(t)\right)\mathrm{PX}(t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \\ &= \int_{B_t} \frac{\mathcal{R}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\mathrm{X}_p(t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\, \mathrm{d}A \ ,\end{aligned}$$

where the last equality is obtained by recalling the definition of P below Eq. (95). More explicitly, we can write

$$\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right) = \varepsilon_t\ \mathcal{R}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\mathrm{X}_p(t), \quad \varepsilon_t = \frac{\mathrm{Area}(B_t)}{\|\mathrm{X}_p^*(t) \wedge \mathrm{X}_p(t)\|}\ . \tag{110}$$

Therefore, the second term of the right hand side in Eq. (98) becomes

$$\begin{aligned}\int_0^1 \langle \mathrm{P}_{\sigma_t}\left[\mathcal{R}_{\Sigma_t}\left(\mathrm{X}_p^*(t), \mathrm{X}_p(t)\right)\right], \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t &= \int_0^1 \varepsilon_t\ \langle \mathcal{R}\left(\mathrm{P}_{\sigma_t}\mathrm{X}_p^*(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t)\right)\mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \dot{\sigma}(t)\rangle_{\sigma(t)}\ \mathrm{d}t \\ &= \int_0^1 \frac{\varepsilon_t}{t}\ \langle \mathcal{R}\left(\mathrm{P}_{\sigma_t}\mathrm{X}_p^*(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t)\right)\mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t)\rangle_{\sigma(t)}\ \mathrm{d}t \\ &= \int_0^1 \frac{\varepsilon_t}{t}\ \mathrm{R}\left(\mathrm{P}_{\sigma_t}\mathrm{X}_p^*(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t), \mathrm{P}_{\sigma_t}\mathrm{X}_p(t)\right)\ \mathrm{d}t \\ &= 0\ ,\end{aligned}$$

where the first equality follows by $\nabla\mathcal{R} \equiv 0$, the second one follows by recalling the implication of the relation $\sigma_t(s) = \sigma(st)$ on the derivative and the last equality follows from Condition (i) in Eq. (108). As result, we obtain that under Conditions (108) our divergence is given by

$$\mathcal{D}(p, q) = \int_0^1 t\ \|\dot{\sigma}(t)\|^2\ \mathrm{d}t,$$

which proves that also in this case our divergence coincides with the one of Ay and Amari.

In order to single out a connection between the divergence $\mathcal{D}(p, q)$ proposed in the present article and the one introduced in [11], consider the definition of $\mathcal{D}(q, p)$. By interchanging p and q

in Eq. (75), we obtain

$$\mathcal{D}(q,p) \equiv \mathcal{D}_q(p) = \int_0^1 \langle \Pi_t(q), \dot{\bar{\sigma}}(t)\rangle_{\bar{\sigma}(t)} \,\mathrm{d}t \ , \tag{111}$$

where the ∇-geodesic $\bar{\sigma}(t)$ goes from q to p and

$$\Pi_t(q) = \mathrm{P}_{\sigma_t^*}\mathrm{X}_q(t),$$

with σ_t^* being the ∇^*-geodesic from q to $\sigma(t)$ and $\mathrm{X}_q(t) = \exp_q^{-1}(\sigma(t))$. We now investigate the relation between $\Pi_t(q)$ and the vector field $\mathrm{F}(\sigma(t)) = \dot{\bar{\sigma}}_t(0)$. Firstly, we may observe that

$$\mathrm{X}_q(t) = -\mathrm{P}_{\sigma_t}^{-1}\dot{\bar{\sigma}}_t(0),$$

where σ_t is the ∇-geodesic from q to $\bar{\sigma}(t)$. Then, by repeating the same methods as the ones that have led to Eq. (96) together the nice representation (110) of $\mathcal{R}_{\Sigma_t}$ under Conditions (108), we obtain that

$$\begin{aligned}\Pi_t(q) &= -\mathrm{P}_{\sigma_t^*} \circ \mathrm{P}_{\sigma_t}^{-1}\dot{\bar{\sigma}}_t(0) \\ &= -\dot{\bar{\sigma}}_t(0) + \varepsilon_t\ \mathcal{R}\left(\mathrm{X}_q^*(t), \dot{\bar{\sigma}}_t(0)\right)\dot{\bar{\sigma}}_t(0) \ .\end{aligned} \tag{112}$$

Since $\bar{\sigma}_t(s)$ is a re-parametrization of the ∇-geodesic $\bar{\sigma}(t)$, namely $\bar{\sigma}_t(s) = \bar{\sigma}((1-t)s+t)$, then $\dot{\bar{\sigma}}_t(0) = (1-t)\dot{\bar{\sigma}}(t)$. Therefore, by integrating $\Pi_t(q)$ along $\bar{\sigma}(t)$ we get $\mathcal{D}(q,p)$ and from Eq. (112) we are able to establish the connection between our divergence and the one of Henmi and Kobayashi,

$$\begin{aligned}\mathcal{D}(q,p) &= \int_0^1 \langle \Pi_t(q), \dot{\bar{\sigma}}(t)\rangle_{\bar{\sigma}(t)} \,\mathrm{d}t \\ &= -\int_0^1 \langle \dot{\bar{\sigma}}_t(0), \dot{\bar{\sigma}}(t)\rangle_{\bar{\sigma}(t)} \,\mathrm{d}t + \int_0^1 \frac{1}{1-t}\,\mathrm{R}\left(\mathrm{X}_q^*(t), \dot{\bar{\sigma}}_t(0), \dot{\bar{\sigma}}_t(0), \dot{\bar{\sigma}}_t(0)\right)\mathrm{d}t \\ &= W^*(p\|q),\end{aligned} \tag{113}$$

where we used $\nabla\mathcal{R} \equiv 0$ and the last equality follows from Condition (i) of (108) and by recalling definition (106). The function $W^*(p\|q)$ is actually the dual divergence of $W(p\|q)$ and in [11] the following symmetric property has been proved,

$$W^*(p\|q) = f^*\left(W(q\|p)\right), \tag{114}$$

where f^* is a function such that $f^*(0) = 0$ and $(f^*)'(0) = 1$. Finally, from Eq. (114) we can carry out the following connection between $\mathcal{D}(q,p)$ and $W(q\|p)$,

$$\mathcal{D}(q,p) = f^*\left(W(q\|p)\right) \ . \tag{115}$$

V. CONCLUDING REMARKS

The main result obtained in this work is Theorem II.1, which gives a geometric interpretation of Theorem I.1. In particular, we showed that the rays of hypersurfaces H_p at q are generated by the sum of the tangent vectors $\Pi_q + \Pi_q^*$. Hypersurfaces H_p are level-sets of the pseudo-squared-norm $r_p(q)$. The symmetry of $r_p(q)$ suggested that it is not a good candidate to put forward for defining the novel canonical divergence since, commonly, divergences are not necessarily symmetric. However, in classical Information Theory [13] we recover such a symmetry by adding to the Kullback-Leibler divergence its dual function that is obtained from the Kullback-Leibler divergence by just exchanging the role of p and q.

Inspired by this property of the Kullback-Leibler divergence for positive and probability measures, we made the choice to define the novel divergence by integrating Π along the ∇-geodesic connecting p to q. The dual divergence is then defined by the path integration of Π^* along the ∇^*-geodesic connecting p to q. Nevertheless, it does not happen that $r_p(q) = \mathcal{D}_p(q) + \mathcal{D}_p^*(q)$. To overcome this difficulty, we introduced two more functions, namely the *Phi-functions* in this paper. In this way, we provided a split of $r_p(q)$ in terms of both the sums, $\varphi_p + \mathcal{D}_p^*$ and $\varphi_p^* + \mathcal{D}_p$. A further relevance of these *Phi-functions* consists in supplying local decompositions of Π and Π^* in terms of $\mathrm{grad}\varphi_p$ and $\mathrm{grad}\varphi_p^*$, respectively.

We proved that both the functions, the canonical divergence $\mathcal{D}$ and the Phi-function φ generate the dual geometry of a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$. Very surprisingly, we showed that they coincide up to a suitable monotonic function. This nice result led us to address our investigation to the symmetry of $\mathcal{D}(p, q)$ as stated by Eguchi in [9]. We were able to prove that $\mathcal{D}(q, p) = \Upsilon(\mathcal{D}^*(p, q))$, where Υ is a suitable monotonic function.

In addition, we proved that $\mathcal{D}(p, q)$ reduces to the divergence introduced by Ay and Amari when M is self dual, when M is dually flat and when M satisfies a property that is the statistical geometric analogue of the concept of symmetric spaces in Riemannian geometry. The last class of statistical manifold drove our study to the approach proposed by Henmi and Kobayashi in [11] and then we described a close connection between our divergence and the one of Henmi and Kobayashi.

Several examples of divergences can be found in literature arising from a wide range of physical sciences. In [8] a divergence is defined as the solution of the Hamilton-Jacobi problem associated with a canonical Lagrangian defined in TM. In [17] by resorting to the dual structure of the Hamiltonian and Lagrangian formulation of mechanics in $\mathrm{T}^*\mathrm{M}$ and TM, it is established that the divergence function agrees with the exact discrete Lagrangian up to third order if and only if

M is a Hessian manifold. In this manuscript, we explored the intrinsic geometric structure of a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ when only the torsion freeness of ∇ and ∇^* is required. This investigation identified appropriate vector fields upon which it is based a very natural definition of divergence and its dual function. The present approach also paves the way for investigating the geometric structure of submanifolds of M in order to generalize the Pythagorean Theorem and then provide a deeper understanding of the projection theorem. This will constitute the study of forthcoming investigation.

---

[1] Amari, S.I.: Information Geometry and Its Applications, 1st edn. Springer Publishing Company (2016)

[2] Amari, S.I., Ay, N.: Standard divergence in manifold of dual affine connections. In: F. Nielsen, F. Barbaresco (eds.) Geometric Science of Information, pp. 320–325. Springer International Publishing, Cham (2015)

[3] Amari, S.I., Cichock, A.: Information geometry of divergence functions. Bulletin of the Polish Academy of Sciences: Technical Sciences **58**(1), 183–195 (2010)

[4] Amari, S.I., Nagaoka, H.: Methods of Information Geometry, Translations of Mathematical monographs, vol. 191. Oxford University Press (2000)

[5] Ay, N., Amari, S.I.: A novel approach to canonical divergences within information geometry. Entropy **17**, 8111–8129 (2015)

[6] Ay, N., Jost, J., Le, H.V., Schwachhöfer, L.: Information Geometry, 1st edn. Springer International Publishing (2017)

[7] Cheng, T.P.: Relativity, Gravitation and Cosmology: A basic introduction, 2nd edn. Oxford University Press (2016)

[8] Ciaglia, F.M., Di Cosmo, F., Felice, D., Mancini, S., Marmo, G., Pérez-Pardo, J.M.: Hamilton-jacobi approach to potential functions in information geometry. Journal of Mathematical Physics **58**(6), 063506 (2017)

[9] Eguchi, S.: Second order efficiency of minimum contrast estimators in a curved exponential family. Ann. Statist. **11**(3), 793–803 (1983)

[10] Helgason, S.: Differential Geometry, Lie Groups, and Symmetric Spaces, vol. 80 (1978)

[11] Henmi, M., Kobayashi, R.: Hooke's law in statistical manifolds and divergences. Nagoya Math. J. **159**, 1–24 (2000)

[12] Katanaev, M.O.: Normal coordinates in affine geometry. Lobachevskii Journal of Mathematics **39**(3), 464–476 (2018)

[13] Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)

[14] Kurose, T.: On the divergences of 1-conformally flat statistical manifolds. Tohoku Math. J. (2) **46**(3), 427–433 (1994)

[15] Lauritzen, S.L.: Differential geometry in statistical inference. Lecture Notes-Monograph Series **10**, 163–218 (1987)

[16] Lee, J.M.: Riemannian Manifolds: An introduction to Curvature, vol. 176, 1st edn. Springer-Verlag New York (1997)

[17] Leok, M., Zhang, J.: Connecting information geometry and geometric mechanics. Entropy **19**(10) (2017)

[18] Matsuzoe, H.: On realization of conformally-projectively flat statistical manifolds and the divergences. Hokkaido Math. J. **27**(2), 409–421 (1998)

[19] 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. **23**(2), 327–332 (1993)

[20] Merkulov, S., Schwachhöfer, L.: Classification of irreducible holonomies of torsion-free affine connections. Annals of Mathematics **150**(1), 77–149 (1999)

[21] Milnor, J.: Morse theory. Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J. (1963)

[22] Presnov, E.: Global decomposition of vector field on riemannian manifolds along natural coordinates. Reports on Mathematical Physics **62**(3), 273 – 282 (2008)

[23] Rani, R.: On parallel transport and curvature (2009)

[24] Yang, D.: Holonomy equals curvature

[25] Zhang, J.: A note on curvature of connections of a statistical manifold. Annals of the Institute of Statistical Mathematics **59**(1), 161–170 (2007)

APPENDIX A: DIFFERENTIAL GEOMETRY OF STATISTICAL MANIFOLDS

In this section we review useful tools of differential geometry of statistical manifolds mainly focusing on the curvature tensor. We describe classes of statistical manifolds in terms of curvature tensor features of them and give description of *conjugate symmetric* and *dually flat* statistical manifolds. For a more detailed presentation we refer to [15], [1] and [6]. A statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is the datum of a C^∞ manifold M, a metric tensor g and two affine connections ∇ and ∇^* such that Eq. (1) holds true. Let us recall that an affine connection ∇ on M is a linear connection on the tangent bundle TM,

$$\nabla : \mathcal{T}(\mathrm{M}) \times \mathcal{T}(\mathrm{M}) \to \mathcal{T}(\mathrm{M}), \quad (X, Y) \mapsto \nabla_X Y \ ,$$

such that

$$\nabla_{fX_1+gX_2}X = f\nabla_{X_1}X + g\nabla_{X_2}X \quad \forall X_1, X_2, X \in \mathcal{T}(\mathrm{M}) \text{ and } f, g \in C^\infty(\mathrm{M})$$
$$\nabla_X(aX_1 + bX_2) = a\nabla_X X_1 + b\nabla_X X_2 \quad \forall X_1, X_2, X \in \mathcal{T}(\mathrm{M}) \text{ and } a, b \in \mathbb{R}$$
$$\nabla_X(fY) = f\nabla_X Y + X(f)Y \qquad \forall X, Y \in \mathcal{T}(\mathrm{M}) \text{ and } f \in C^\infty(\mathrm{M}) .$$

Roughly speaking, an affine connection is directional derivative of vector fields. In particular, $\nabla_X Y$ is the change of Y in the direction of X. The rule for comparing vectors in two distinct tangent spaces $\mathrm{T}_p\mathrm{M}$ and $\mathrm{T}_q\mathrm{M}$ is established by the notion of parallel transport.

Let us now introduce such a notion by relying on a smooth curve $\gamma : [0,1] \to \mathrm{M}$ of M. A vector field along γ is a smooth map $V : [0,1] \to \mathrm{TM}$ such that $V(t) \in \mathrm{T}_{\gamma(t)}\mathrm{M}$ for all $t \in [0,1]$. Let $\mathcal{T}(\gamma)$ the space of all vector fields along γ, then the *covariant derivative* $\nabla_t : \mathcal{T}(\gamma) \to \mathcal{T}(\gamma)$ of $V \in \mathcal{T}(\gamma)$ along γ is defined in terms of the connection ∇ as $\nabla_t V(t) := \nabla_{\dot{\gamma}(t)}\widetilde{V}$, where $\widetilde{V}$ is the extension of V to $\mathcal{T}(\mathrm{M})$. A vector field $V \in \mathcal{T}(\gamma)$ is said to be *parallel* along γ with respect to ∇ if $\nabla_t V(t) \equiv 0$ for all $t \in [0,1]$. Therefore, a basic result in Calculus allows us to consider the isomorphism

$$\mathrm{P}_\gamma : \mathrm{T}_{\gamma(t_0)}\mathrm{M} \to \mathrm{T}_{\gamma(t)}\mathrm{M}, \quad V \mapsto \mathrm{P}_\gamma(V) \tag{A1}$$

where $\mathrm{P}_\gamma(V) := V(t)$ and $V \in \mathcal{T}(\gamma)$ is the unique parallel vector along γ such that $V(t_0) \equiv V$. Likewise, we have the parallel transport with respect to the ∇^*-connection,

$$\mathrm{P}^*_\gamma : \mathrm{T}_{\gamma(t_0)}\mathrm{M} \to \mathrm{T}_{\gamma(t)}\mathrm{M}, \quad V \mapsto \mathrm{P}^*_\gamma(V) . \tag{A2}$$

All the concepts that we are from here describing can be naturally passed to the ∇^* connection; so for the sake of simplicity, we only refer to the ∇ connection.

The expression of ∇ connection in local coordinates ξ^i_p at $p \in \mathrm{M}$ is given in terms of the local basis $\{\partial_i\}_p$ ($\partial_i = \partial/\partial\xi^i_p$) of the tangent space $\mathrm{T}_p\mathrm{M}$ by means of the Christoffel's symbols Γ^k_{ij},

$$\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k,$$

where we adopted Einstein's summation convention according to which whenever an index appears in an expression as upper and lower index, we sum over that index. The same happens for ∇^* connection, i.e. $\nabla^*_{\partial_i}\partial_j = {}^*\Gamma^k_{ij}\partial_k$.

By relying on local coordinates $\{\xi\}$, we can also give the local expression of the parallel transport P. Consider $X \in \mathcal{T}(\gamma)$, then we have $X(t) = X^i(t)\partial_i(t)$, where $\{\partial_i(t)\}$ is a local frame of the tangent space $\mathrm{T}_{\gamma(t)}\mathrm{M}$. Then, we have that

$$\frac{\mathrm{d}X^k(t)}{\mathrm{d}t} + \Gamma^k_{ij}(\gamma(t))\dot{\gamma}^i(t)X^j(t) = 0 . \tag{A3}$$

It is clear from Eq. (A3) that whenever we specify one initial condition $X^i(0) = X^i_p \in \mathrm{T}_p\mathrm{M}$ we get one solution of the differential equation and then we can define the isomorphism (A1).

A *geodesic* of ∇ is a curve with parallel tangent vector field,

$$\nabla_t \dot{\gamma} \equiv 0 \ , \tag{A4}$$

and in local coordinates it reads as

$$\ddot{\gamma}^k + \Gamma^k_{ij}\dot{\gamma}^i\dot{\gamma}^j = 0 \ . \tag{A5}$$

For all $p \in \mathrm{M}$ and $X_p \in \mathrm{T}_p\mathrm{M}$ there is a unique geodesic γ_{X_p} such that,

$$\gamma_{X_p}(0) = p, \text{ and } \dot{\gamma}_{X_p}(0) = X_p. \tag{A6}$$

Hence, by defining for $X_p \in \mathrm{T}_p\mathrm{M}$,

$$\exp_p(X_p) := \gamma_{X_p}(1), \tag{A7}$$

we obtain the *exponential map* at p. The exponential map is in general well-defined at least in a neighborhood of zero in $\mathrm{T}_p\mathrm{M}$ and, moreover, can be globally defined in special cases.

1. Conjugate Symmetric Statistical Manifolds

To the affine connection ∇ we can associate two tensors, the *torsion* and the *curvature*. They are given by

$$\mathrm{Tor}(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] \tag{A8}$$

$$\mathcal{R}(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, \tag{A9}$$

where $X, Y, Z \in \mathcal{T}(\mathrm{M})$ and $[X,Y] = XY - YX$ is the Lie bracket of X and Y.

Analogously, we can associate two tensors to the dual connection: the torsion tensor and the curvature tensor of ∇^*,

$$\mathrm{Tor}^*(X,Y) = \nabla^*_X Y - \nabla^*_Y X - [X,Y] \tag{A10}$$

$$\mathcal{R}^*(X,Y)Z = \nabla^*_X\nabla^*_Y Z - \nabla^*_Y\nabla^*_X Z - \nabla^*_{[X,Y]}Z \ . \tag{A11}$$

Then $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is called a statistical manifold when both the connections ∇ and ∇^* are *torsion free*, i.e Tor $\equiv 0$ and Tor$^* \equiv 0$. From this, it follows that the curvature $\mathcal{R}$ satisfies the first Bianchi identity,

$$\mathcal{R}(X,Y)Z + \mathcal{R}(Y,Z)X + \mathcal{R}(Z,X)Y = 0 \tag{A12}$$

for all $X, Y, Z \in \mathcal{T}(\mathrm{M})$. The same holds true for the curvature tensor $\mathcal{R}^*$.

Given the metric structure on M, we can also consider the Riemann curvature tensor of ∇ that is defined as follows

$$\mathrm{R}(X,Y,Z,W) := \mathrm{g}\left(\mathcal{R}_{XY}Z, W\right) \ . \tag{A13}$$

From definition (A9) it immediately follows that $\mathrm{R}(X,Y,Z,W) = -\mathrm{R}(Y,X,Z,W)$ and in particular $\mathrm{R}(X,X,Z,W) = 0$. Moreover, from the first Bianchi identity (A12) we have that

$$\mathrm{R}(X,Y,Z,W) + \mathrm{R}(Y,Z,X,W) + \mathrm{R}(Z,X,Y,W) = 0 \ .$$

Analogously we can define the Riemann curvature tensor of ∇^*,

$$\mathrm{R}^*(X,Y,Z,W) := \mathrm{g}\left(\mathcal{R}^*_{XY}Z, W\right) \ , \tag{A14}$$

and same equalities as R hold true as well.

Consider now Riemann curvature tensors R and R* both together. We have the following result [15],

Proposition A.1. *If* R *is the Riemann curvature tensor of* ∇ *and* R* *the one of* ∇^* *we have that*

$$\mathrm{R}(X,Y,Z,W) = -\mathrm{R}^*(X,Y,W,Z) \ . \tag{A15}$$

Proof. By considering X, Y as part of an orthonormal frame on the tangent bundle TM we can assume that $[X,Y] = 0$. Owing to this consideration we have that

$$\begin{aligned}
&XY\mathrm{g}\left(Z,W\right) = X\left(Y\mathrm{g}(Z,W)\right) \\
&= X\left(\mathrm{g}\left(\nabla_Y Z, W\right) + \mathrm{g}\left(Z, \nabla^*_Y W\right)\right) \\
&= \mathrm{g}\left(\nabla_X\nabla_Y Z, W\right) + \mathrm{g}\left(\nabla_Y Z, \nabla^*_X W\right) + \mathrm{g}\left(\nabla_X Z, \nabla^*_Y W\right) + \mathrm{g}\left(Z, \nabla^*_X\nabla^*_Y W\right)
\end{aligned}$$

By alternating X and Y we arrive at

$$\begin{aligned}
&0 = [X,Y]\mathrm{g}(Z,W) = XY\mathrm{g}(Z,W) - YX\mathrm{g}(Z,W) \\
&= \mathrm{R}(X,Y,Z,W) + \mathrm{R}^*(X,Y,Z,W)
\end{aligned}$$ □

As direct consequence we have

Corollary A.1. *The following conditions are equivalent,*

1. $\mathrm{R} \equiv \mathrm{R}^*$

2. $\mathrm{R}(X, Y, Z, W) = -\mathrm{R}(X, Y, W, Z)$

From the second condition in Cor. A.1 we trivially have that

$$\mathrm{g}\left(\mathcal{R}(X, Y)Z, Z\right) \equiv 0 \quad \text{for all } X, Y, Z \in \mathcal{T}(\mathrm{M}) \,. \tag{A16}$$

Another consequence of Cor. A.1 is that ∇ is flat if and only if ∇^* is flat. Let us now briefly discuss about the second condition of Cor. A.1, or equivalently the Eq. (A16), and see for which classes of statistical manifolds it holds true.

Given the dual structure $(\mathrm{g}, \nabla, \nabla^*)$, we can obtain the Levi-Civita connection as follows [1],

$$\overline{\nabla} := \frac{1}{2}\left(\nabla + \nabla^*\right). \tag{A17}$$

In addition, we can define a totally symmetric cubic tensor T [15],

$$T(X, Y, Z) := \mathrm{g}\left(\widetilde{T}(X, Y), Z\right), \quad \text{where } \widetilde{T}(X, Y) := \nabla_X Y - \nabla^*_X Y \,. \tag{A18}$$

Let us now define a 1-parameter family of α-connections on M as follows,

$$\overset{\alpha}{\nabla}_X Y := \overline{\nabla}_X Y - \frac{1}{2}\widetilde{T}(X, Y). \tag{A19}$$

From the torsion-freeness of the statistical manifold $\mathcal{S}$ and the symmetry of $\widetilde{T}$ we immediately have that

$$\left(\overset{\alpha}{\nabla}\right)^* = \overset{-\alpha}{\nabla} \quad \text{and} \quad \overset{1}{\nabla} = \nabla, \; \overset{-1}{\nabla} = \nabla^* \,. \tag{A20}$$

We say that a statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is *conjugate symmetric* if for all α the curvature tensor $\overset{\alpha}{\mathrm{R}}$ fulfils the following relation,

$$\overset{\alpha}{\mathrm{R}} \equiv \overset{-\alpha}{\mathrm{R}} \,. \tag{A21}$$

Therefore, by means of Cor. A.1 we have

Proposition A.2. *Sufficient conditions for a statistical* $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ *being conjugate symmetric are*

1. *There exists* $\alpha \neq 0$ *such that* $\overset{\alpha}{\mathrm{R}} \equiv \overset{-\alpha}{\mathrm{R}}$.

2. There exists $\alpha \neq 0$ *such that* $\overset{\alpha}{\mathrm{R}} \equiv 0$*, i.e.* $\mathcal{S}$ *is* α*-flat.*

Finally, in a conjugate symmetric manifold, the Riemann curvature tensor satisfies all the identities as the Riemann curvature tensor of the Levi-Civita connection, i.e.

$$\overset{\alpha}{\mathrm{R}}(X,Y,Z,W) = -\overset{\alpha}{\mathrm{R}}(Y,X,Z,W); \tag{A22}$$

$$\overset{\alpha}{\mathrm{R}}(X,Y,Z,W) + \overset{\alpha}{\mathrm{R}}(Y,Z,X,W) + \overset{\alpha}{\mathrm{R}}(Z,X,Y,W) = 0; \tag{A23}$$

$$\overset{\alpha}{\mathrm{R}}(X,Y,Z,W) = -\overset{\alpha}{\mathrm{R}}(X,Y,W,Z); \tag{A24}$$

$$\overset{\alpha}{\mathrm{R}}(X,Y,Z,W) = \overset{\alpha}{\mathrm{R}}(Z,W,X,Y). \tag{A25}$$

The statistical manifold $\mathcal{S} = (\mathrm{M}, \mathrm{g}, \nabla, \nabla^*)$ is called *dually flat* if $\mathrm{R} \equiv 0 \equiv \mathrm{R}^*$. Then, according to Eq. (A21) and Pro. A.2 we can say that a dually flat manifold is conjugate symmetric. In this particular case, there exists α_0 such that $\overset{\alpha_0}{\mathrm{R}} \equiv 0$ and then the statistical manifold $\mathcal{S}$ is often referred as equivalent to *dually flat* manifold [1]. In this particular case, we can rely on two sets of local coordinates $\{\theta^i\}$ and $\{\eta_i\}$ such that

$$\Gamma_{ijk}(\theta) = 0, \quad \text{and} \quad \overset{*}{\Gamma}_{ijk}(\eta) = 0 \,.$$

Here, Γ_{ijk} and $\overset{*}{\Gamma}_{ijk}$ are the connection symbols of ∇ and ∇^*, respectively. In local coordinates they are expressed by

$$\Gamma_{ijk} = \mathrm{g}_{il}\Gamma^l_{jk}, \quad \overset{*}{\Gamma}_{ijk} = \mathrm{g}_{il}{}^*\Gamma^l_{jk} \,, \tag{A26}$$

where Γ^l_{jk} and ${}^*\Gamma_{jk}$ are the Christoffel's symbols of ∇ and ∇^*, respectively. Additionally, if we consider the tangent vectors $\{\partial_i\}$ and $\{\partial^i\}$ of the local coordinates $\{\theta^i\}$ and $\{\eta_i\}$ we have that

$$\mathrm{g}_{ij}\partial_i\partial^j = \delta^j_i, \tag{A27}$$

meaning that these tangent vectors are reciprocal orthogonal with respect to the metric tensor g.

2. Parallel transport and curvature tensor

Now we describe the connection between the parallel transport and the curvature tensor of the connection ∇. Obviously, the same is for ∇^* connection. Roughly speaking, parallel transport along a loop Σ based at $p \in \mathrm{M}$ provides the Lie group of rotations $\mathrm{P}_\Sigma : \mathrm{T}_p\mathrm{M} \to \mathrm{T}_p\mathrm{M}$. This is called the *holonomy group* of ∇ at p. Then, the Lie algebra of it is spanned by the curvature tensor of ∇.

Given $p \in \mathrm{M}$, let

$$\mathcal{L}_p := \{\Sigma : [0,1] \to \mathrm{M} \mid \Sigma(0) = \Sigma(1) = p\} \tag{A28}$$

be the set of piecewise smooth loop based on p and assume that M is simply connected. Then, each $\Sigma \in \mathcal{L}_p$ is homotopic to the trivial loop.

Therefore the *holonomy* of ∇ at $p \in \mathrm{M}$ is defined as the subset of $\mathrm{Aut}(\mathrm{T}_p\mathrm{M})$, i.e. the automorphisms of $\mathrm{T}_p\mathrm{M}$,

$$\mathrm{Hol}_p := \{\mathrm{P}_\Sigma \in \mathrm{Aut}(\mathrm{T}_p\mathrm{M}) \mid \Sigma \in \mathcal{L}_p\} \ . \tag{A29}$$

Basics properties of Hol_p are listed in the following proposition.

Proposition A.3. *The following basic properties of* Hol_p *hold true:*

1. Hol_p *is a closed Lie subgroup of* $\mathrm{Aut}(\mathrm{T}_p\mathrm{M})$ *and its Lie algebra* $\mathfrak{hol}_p \subset \mathrm{End}(\mathrm{T}_p\mathrm{M})$ *is called the holonomy algebra at* p*.*

2. *Given* $\Sigma' : [0,1] \to \mathrm{M}$ *such that* $\Sigma'(0) = p$ *and* $\Sigma'(1) = q$*. Let* $P_{\Sigma'} : \mathrm{T}_p\mathrm{M} \to \mathrm{T}_q\mathrm{M}$ *the parallel transport along* Σ'*. Then*

$$P_{\Sigma'} \circ \mathrm{Hol}_p \circ {P_{\Sigma'}}^{-1} = \mathrm{Hol}_q.$$

From the second property in the latter Proposition, it follows that the holonomy groups are independent of the base point.

Since ∇ is torsion free, the *Ambrose-Singer Holonomy Theorem* [20] supply a very remarkable connection between the curvature tensor $\mathcal{R}$ and the holonomy algebra $\mathfrak{hol}_p(\nabla)$ of ∇. It states that $\mathfrak{hol}_p(\nabla)$ is generated by operators $\mathcal{R}_\Sigma(x,y) := \mathrm{P}_\alpha \circ \mathcal{R}({\mathrm{P}_\Sigma}^{-1}x, {\mathrm{P}_\Sigma}^{-1}y) \circ {\mathrm{P}_\Sigma}^{-1}$,

$$\mathfrak{hol}_p = \langle \{(\mathcal{R})_\Sigma(x,y) \mid x, y \in \mathrm{T}_p\mathrm{M}, \Sigma \text{ a loop at } p\}\rangle. \tag{A30}$$

Eq. (A30) shows that $\mathfrak{hol}_p(\nabla)$ is the vector subspace of $\mathrm{End}(\mathrm{T}_p\mathrm{M})$ spanned by the endomorphisms $\mathcal{R}_\Sigma(x,y)$. Thus, $\mathcal{R}$ determines $\mathfrak{hol}_p(\nabla)$ and, hence $\mathrm{Hol}(\nabla)$. Therefore, if we consider the case of a flat manifold we have that $\mathcal{R} \equiv 0$. Then $\mathfrak{hol}_p(\nabla) = 0$, from which it follows that $\mathrm{Hol}(\nabla) = Id$.

For the purpose of the present manuscript, the previous theoretical setting for highlighting connection between holonomy and curvature tensor is performed into the following result.

Lemma A.1. *Let B be a smooth closed 2-disk such that $p \in \partial B$ and B is foliated by connecting ∇-geodesics segment starting from p. Then*

$$\mathrm{P}_\Sigma Z_p - Z_p = \int_B \frac{\mathrm{P}\left(\mathcal{R}(X,Y)Z\right)}{\|X \wedge Y\|} dA, \tag{A31}$$

where

- *dA is the surface area measure on B induced by the Riemannian metric tensor g on M.*
- *X and Y are linearly independent vector fields on B.*
- *$\Sigma : \mathrm{I} \to \partial B$ is a parametrization of ∂B such that $\Sigma(0) = \Sigma(1) = p$ and, given any inward pointing vector $X \in \mathrm{T}_pB$, the orientation of $(\dot{\Sigma}, X)$ is the same as (X, Y).*
- *$Z_p \in \mathrm{T}_p\mathrm{M}$ and Z is defined by parallel translating Z_p first along the parametrized curve Σ and then, for each $0 \leq s \leq 1$, along the unique ∇-geodesic segment going from $\Sigma(s) \in \partial B$ to B.*
- *P is parallel translation from each point in B to p along the unique ∇-geodesic segment joining them.*

Proof. The proof of this result is provided in [24]. However, for both the sake of completeness and its usefulness we report it here.

Consider a map $H : [0,1] \times [0,1] \to B$ such that $H(1,t) = \Sigma(t)$ and $H(\cdot,t)$ is the ∇-geodesic connecting p to $\Sigma(t)$, for all $t \in [0,1]$. Let us denote

$$S(s,t) = \frac{\partial H}{\partial s}, \quad T(s,t) = \frac{\partial H}{\partial t}$$

then we have $[S,T] = 0$ since ∇ is torsion free. Let us observe that T is the Jacobi vector field along each geodesic $H(\cdot,t)$. Define now

$$J = T - \frac{\mathrm{g}\left(S,T\right)}{\|S\|^2} S.$$

Clearly J is orthogonal to T. Then we have that

$$dA = \|T \wedge S\| ds dt = \|S\| \|T\| ds dt.$$

Let $\{e_i\} \subset \mathrm{T}_p\mathrm{M}$ be an orthonormal frame and extend it by parallel transport along each ∇ geodesic $H(\cdot,t)$. In particular we have that

$$\nabla_t e_i(0,t) = 0, \quad \nabla_s e_i(s,t) = 0,$$

for all $(s,t) \in [0,1] \times [0,1]$. In addition, we also have that

$$\nabla_t Z(1,t) = 0, \quad \nabla_s Z(s,t) = 0,$$

for all $(s,t) \in [0,1] \times [0,1]$. Let us now note that $Z_p = Z(0,0) = Z(1,0)$ and that $\mathrm{P}_\Sigma Z_p = Z(1,1) = Z(0,1)$. Then we have that

$$\begin{aligned}
\langle e_i(p), (\mathrm{P}_\Sigma Z_p - Z_p)\rangle_p &= \langle e_i(0,1), Z(0,1)\rangle_p - \langle e_i(0,0), Z(0,0)\rangle_p \\
&= \int_0^1 \partial_t \langle e_i(0,t), Z(0,t)\rangle dt \\
&= \int_0^1 \langle e_i, \nabla_t Z(0,t)\rangle dt \\
&= \int_0^1 \left[\langle e_i, \nabla_t Z(1,t)\rangle - \int_0^1 \partial_s \langle e_i, \nabla_t Z(s,t)\rangle ds\right] dt \\
&= -\int_0^1 \int_0^1 \langle e_i, \nabla_s \nabla_t Z(s,t)\rangle dsdt \\
&= -\int_0^1 \int_0^1 \langle e_i, \mathcal{R}(S,T) Z(s,t)\rangle dsdt \\
&= -\int_0^1 \int_0^1 \langle e_i, \mathcal{R}(S,J) Z(s,t)\rangle dsdt \\
&= -\int_0^1 \int_0^1 \langle e_i, \mathcal{R}(\sigma,\tau) Z(s,t)\rangle \|S\| \|T\| dsdt \\
&= \int_0^1 \frac{\langle e_i, \mathcal{R}(X,Y)Z\rangle}{\|X \wedge Y\|} dA,
\end{aligned}$$

where $\sigma = S/\|S\|$, $\tau = T/\|T\|$ form an orthonormal frame on B. Finally the result follows from

$$\mathrm{P}_\Sigma \left(\mathcal{R}(X,Y)Z\right) = \sum_i e_i(0,0) \langle e_i(s,t), \mathcal{R}(X,Y)Z\rangle \ .$$