

**Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig**

**Dynamical Systems induced by
Canonical Divergence in dually flat
manifolds**

by

Domenico Felice and Nihat Ay

Preprint no.: 103

2018



Dynamical Systems induced by Canonical Divergence in dually flat manifolds

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*

The principles of classical mechanics have shown that the inertial quality of mass is characterized by the kinetic energy. This, in turn, establishes the connection between geometry and mechanics. We aim to exploit such a fundamental principle for information geometry entering the realm of mechanics. According to the modification of curve energy stated by Amari and Nagaoka for a smooth manifold M endowed with a dual structure (g, ∇, ∇^*) , we consider ∇ and ∇^* kinetic energies. Then, we prove that a recently introduced canonical divergence and its dual function coincide with Hamilton principal functions associated with suitable Lagrangian functions when (M, g, ∇, ∇^*) is dually flat. Corresponding dynamical systems are studied and the tangent dynamics is outlined in terms of the Riemannian gradient of the canonical divergence. Solutions of such dynamics are proved to be ∇ and ∇^* geodesics connecting any two points sufficiently close to each other. Application to the standard Gaussian model is also investigated.

PACS numbers: Classical differential geometry (02.40.Hw), Riemannian geometries (02.40.Ky), Lagrangian and Hamiltonian approach (11.10.Ef).

I. INTRODUCTION

A. The role of geometry in classical mechanics

The Riemannian geometry of a couple (M, g) is based on one single differential quantity called the *line element* ds . This is defined in terms of the metric tensor g by the following expression,

$$ds^2 = \sum_{i,j=1}^n g_{ij}(\boldsymbol{\xi}) d\xi^i \otimes d\xi^j, \quad (1)$$

*Electronic address: felice@mis.mpg.de

†Electronic address: nay@mis.mpg.de

and it allows the development of a complete geometry on M . Here, $\{\xi\}$ is any system of local coordinates in the smooth manifold M and n denotes the dimension of M . In general, the coefficients g_{ij} of the metric tensor are not constant over M . However, when the geometry is Euclidean, the line element can be re-written by $ds^2 = g_{ij} d\xi^i \otimes d\xi^j$, with $g_{ij} = \delta_{ij}$ given by the Kronecker delta symbol and the Einstein's notation is adopted: from here on, whenever an index is repeated as sub and superscript in a product, it represents summation over the range of the index.

Geometry enters the realm of mechanics in connection with the *inertia of mass* which is characterized by the kinetic energy [12]. The kinetic energy K of a single particle of mass 1, is given in the n -dimensional Euclidean space by $K = \frac{1}{2} \sum_{i=1}^n v_i^2$, where $\mathbf{v} = (v_1, \dots, v_n)$ is the velocity of the particle defined by

$$\|\mathbf{v}\|^2 = \frac{(d\xi^1)^2 + \dots + (d\xi^n)^2}{dt^2},$$

and $\xi = (\xi^1, \dots, \xi^n)$ denotes the coordinates of the particle. Therefore, the line element in the n -dimensional space is defined by the following relation

$$ds^2 = 2Kdt^2$$

which establishes the connection between geometry and mechanics. The inertial quality of mass is expressed on the left-hand side of Newton's law in the form of *mass times acceleration* [12]. The first law of dynamics states that if the vector sum of forces acting on the particle is zero, then the velocity of the particle is constant [2]. If this is the case, the trajectory of the particle of mass 1 with velocity $\mathbf{v} = \xi_q - \xi_p$ is the straight line

$$\xi(t) = \xi_p + t\mathbf{v}, \quad t \in [0, 1]$$

where ξ_p and ξ_q are the local coordinates at p and q , respectively, for any $p, q \in M$.

The extension of these concepts to the general Riemannian geometry can be expressed through the Lagrangian formulation of mechanics [12] by requiring some *kinematical conditions* between the coordinates. We may then restrict the free movability of the particle by forcing it to stay along an arbitrary curve $\gamma(t)$. Hence, the kinetic energy along $\gamma(t)$ ($t \in [0, 1]$) is defined as

$$K = \mathcal{L}_g := \frac{1}{2} \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} = \frac{1}{2} \|\dot{\gamma}(t)\|_{\gamma(t)}^2, \quad (2)$$

where $\langle \cdot, \cdot \rangle_{\gamma(t)}$ denotes the inner product and $\|\cdot\|_{\gamma(t)}$ is the norm, both induced by g on $\gamma(t)$. The Hamilton's principle requires that the time-integral of the *Lagrangian function* \mathcal{L}_g shall be

stationary. Therefore, by defining the *energy of the curve* γ as

$$E(\gamma) := \int_0^1 \mathcal{L}_g(\gamma(t), \dot{\gamma}(t)) dt, \quad (3)$$

the Hamilton's principle is formulated by $\delta E(\gamma) = 0$, where δE is induced by infinitesimal variation $\delta\gamma$ of the trajectory under the constraints that $\delta\gamma(0) = \delta\gamma(1) = 0$. Working with the local coordinates $(\xi^1(\gamma(t)), \dots, \xi^n(\gamma(t)))$, this principle, also called *the principle of the least action*, yields the Euler-Lagrange equations associated with \mathcal{L}_g [19],

$$\frac{d}{dt} \frac{\partial \mathcal{L}_g}{\partial \dot{\xi}^i} - \frac{\partial \mathcal{L}_g}{\partial \xi^i} = 0, \quad (4)$$

and it selects the definite path “chosen by nature as the actual path of the motion” [12]. Then, the Euler-Lagrange equations for the energy E are

$$\ddot{\xi}^i(t) + \bar{\Gamma}_{jk}^i \dot{\xi}^j(t) \dot{\xi}^k(t) = 0, \quad i = 1, \dots, n, \quad \gamma(0) = p, \gamma(1) = q \quad (5)$$

with $\bar{\Gamma}_{jk}^i$ denoting the Christoffel's symbols of the Levi-Civita connection $\bar{\nabla}_{\text{LC}}$ and we used the abbreviation $\dot{\xi}^i(t) = \frac{d}{dt} (\xi^i(\gamma(t)))$ (see, e.g. [11]). The evaluation of the energy $E(\gamma_c)$ at the $\bar{\nabla}_{\text{LC}}$ -geodesic γ_c , i.e. the solution of Eq. (5), gives a two points function

$$S(p, q) = E(\gamma_c) \quad (6)$$

which is known in literature as the *Hamilton principal function* [19] associated with the Lagrangian \mathcal{L}_g . We will discover later in the paper the important role played by the Hamilton principal function for describing dynamical systems in a generalization of the Riemannian geometry. Therefore, given the Lagrangian \mathcal{L}_g the Hamilton's principle asserts that the actual motion realized in nature is that particular motion for which the energy E assumes its smallest value [12]. This principle is strengthened by the geometric interpretation of $\bar{\nabla}_{\text{LC}}$ -geodesics. Indeed, a classical result in Riemannian geometry states that the path between $\gamma(0) = p$ and $\gamma(1) = q$ with minimum length is the $\bar{\nabla}_{\text{LC}}$ -geodesic parametrized with respect to the length arc [14]. This establishes a close connection between mechanics and Riemannian geometry.

Information Geometry is a generalization of Riemannian geometry where not only Levi-Civita geodesics are considered. This poses the problem on interpreting more general geodesics in mechanistic terms. This paper is addressing this problem by introducing a generalization of the kinetic energy. The following section summarizes the main results of the paper.

B. Outline of the main results

In Information Geometry (IG) [1], the geometry on a smooth manifold M is induced by a dual structure (g, ∇, ∇^*) , where ∇ and ∇^* are linear connections on the tangent bundle TM such that

$$X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad \forall X, Y, Z \in \mathcal{T}(M), \quad (7)$$

and $\mathcal{T}(M)$ denotes the space of vector fields on M . Hereafter, we assume that both the connections, ∇ and ∇^* , are torsion-free and we refer to $\mathcal{S} = (M, g, \nabla, \nabla^*)$ as *statistical manifold* [4]. The complete information of the geometric structure of \mathcal{S} is encoded in a distance-like function $\mathcal{D} : M \times M \rightarrow \mathbb{R}^+$ with $\mathcal{D}(p, q) = 0$ iff $p = q$. Such \mathcal{D} is usually referred to as *divergence function* on M and allows to recover the dual structure (g, ∇, ∇^*) in the following way [7]:

$$\begin{aligned} g_{ij}(p) &= -\partial_i \partial_j' \mathcal{D}(\xi_p, \xi_q)|_{p=q}, \\ \Gamma_{ijk}(p) &= -\partial_i \partial_j \partial_k' \mathcal{D}(\xi_p, \xi_q)|_{p=q}, \quad \Gamma_{ijk}^*(p) = -\partial_i' \partial_j' \partial_k \mathcal{D}(\xi_p, \xi_q)|_{p=q}, \end{aligned} \quad (8)$$

where $\Gamma_{ijk} = g(\nabla_{\partial_i} \partial_j, \partial_k)$, $\Gamma_{ijk}^* = g(\nabla_{\partial_i}^* \partial_j, \partial_k)$ are the symbols of the dual connections ∇ and ∇^* , respectively. Here, $\partial_i = \frac{\partial}{\partial \xi_p^i}$, $\partial_i' = \frac{\partial}{\partial \xi_q^i}$ and $\{\xi_p := (\xi_p^1, \dots, \xi_p^n)\}$ and $\{\xi_q := (\xi_q^1, \dots, \xi_q^n)\}$ are local coordinate systems at p and q , respectively.

The concept of divergence function has been exploited in [1] to supply a definition of energy curve in IG. More precisely, let (M, g, ∇, ∇^*) be a statistical manifold, the energy of any arbitrary path $\gamma : [0, 1] \rightarrow M$ is given by

$$D(\gamma) := D_\gamma(\gamma(1), \gamma(0)) := \int \int_{0 \leq s \leq t \leq 1} g_\gamma(s) \frac{\mu(t)}{\mu(s)} ds dt, \quad \mu(s) := e^{\int_0^s \Gamma_\gamma(s) ds} \quad (9)$$

where g_γ and Γ_γ are the projection coefficients to γ of g and ∇ , respectively. The energy $D(\gamma)$ is usually referred as the (g, ∇) -divergence of the curve γ from $\gamma(1)$ to $\gamma(0)$. Very remarkably, $D(\gamma)$ does not depend on the parametrization $t \mapsto \gamma(t)$ of γ but its orientation, and the (g, ∇^*) -divergence $D^*(\gamma)$ of γ coincides with the (g, ∇) -divergence of the reversly oriented curve [1]. Hence, it turns out that $D^*(\gamma) = D_\gamma(\gamma(0), \gamma(1))$.

Recently, a divergence function has been proposed by using the geodesic integration of the inverse exponential map [3]. Such a divergence, named henceforth by *canonical divergence*, assumes the following expression,

$$\mathcal{D}(p, q) = \int_0^1 t \langle \dot{\sigma}(t), \dot{\sigma}(t) \rangle_{\sigma(t)} dt, \quad (10)$$

where σ is the ∇ -geodesic connecting $p = \sigma(0)$ and $q = \sigma(1)$. In this manuscript, we show that for any arbitrary curve $\tilde{\gamma}(\tau)$ we can consider the re-parametrized curve $\gamma(t(\tau)) = \tilde{\gamma}(\tau)$ such that

$$D^*(\gamma) = \int_0^1 t \|\dot{\gamma}(t)\|^2 dt = \mathcal{D}_\gamma(\gamma(0), \gamma(1)) .$$

This amounts to require that the acceleration of γ has no tangential component and then the function

$$\mathcal{L}(t, \gamma, \dot{\gamma}) = \frac{t}{2} \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} \quad (11)$$

can be understood as the kinetic energy of γ in IG. In this way, according to Eq. (3), $\mathcal{D}_\gamma(\gamma(0), \gamma(1))$ turns out to be twice the energy of the curve γ . The target of this manuscript is to provide a close relationship between mechanics and geometry in IG by analogy with classical mechanics and Riemannian geometry. Specifically, we aim to exploit the canonical divergence (10) as playing the key role of this connection. We mainly focus our investigation on *dually flat manifolds* (see Appendix A for more details on this class of statistical manifolds). Then we supply the Euler-Lagrange equations for the energy \mathcal{D}_γ and prove that the solution is given by an unparametrized ∇ -geodesic. Therefore, by noticing that unparametrized ∇ -geodesics are the integral curves of $\text{grad}_{\gamma(t)} \mathcal{D}_p(\cdot)$, we prove that $\mathcal{D}(p, q)$ coincides with the *Hamilton principal function* associated with the Lagrangian function \mathcal{L} . To be more precise, by defining the energy of the curve γ in analogy with Eq. (3) as

$$\mathcal{D}_\gamma(p, q) = \int_0^1 \mathcal{L}(t, \gamma(t), \dot{\gamma}(t)) dt, \quad \gamma(0) = p, \gamma(1) = q ,$$

we succeed to prove that the canonical divergence \mathcal{D} is a two point function on M such that

$$\mathcal{D}(p, q) = \mathcal{D}_{\gamma_c}(p, q) , \quad (12)$$

where γ_c is the unparametrized ∇ -geodesic from p to q . In [6] the Hamilton principal function associated with a suitably chosen Lagrangian function is showed to be a potential function for the dual structure (g, ∇, ∇^*) on a smooth manifold M. In this manuscript we take the opposite avenue and prove that the canonical divergence proposed in [3], which is a potential function for the dual structure (g, ∇, ∇^*) , turns out to be the Hamilton principal function associated with the Lagrangian function (11).

The relevance of the result stated by Eq. (12) can be evaluated in the Hamiltonian description of the mechanics [12]. So, we address our investigation to the Hamiltonian formulation by defining a new function, i.e. the Hamiltonian of the system, through the Legendre transform [19]. More

specifically, we define the conjugate momentum ζ to ξ by means of the Lagrangian \mathcal{L} , i.e. $\zeta = \partial\mathcal{L}/\partial\dot{\xi}$, and then the Hamiltonian H associated with \mathcal{L} is given by

$$H(t, \xi(t), \zeta(t)) := \zeta_i \xi^i - \mathcal{L}(t, \xi(t), \dot{\xi}(t)) ,$$

where $\zeta = (\zeta_1, \dots, \zeta_n)$. In this context, the equations of motion are given by the so-called Hamilton's equations for the dynamics,

$$\dot{\xi}^i = \frac{\partial H}{\partial \zeta_i}, \quad \dot{\zeta}_i = -\frac{\partial H}{\partial \xi^i}, \quad i = 1, \dots, n , \quad (13)$$

which substitute the Euler-Lagrange equations (4). Hamilton's equations consist of $2n$ first-order differential equations, while Lagrange's equations consist of n second-order equations. However, Hamilton's equations usually do not reduce the difficulty of finding explicit solutions. The issue of finding transformations in the space of coordinates $\{(\xi, \zeta)\}$ which preserve the structure of Eq. (13) and simplify it to a form in which the Hamilton's equations become directly integrable takes place in the Hamilton-Jacobi description of mechanics [12]. The crucial point of this theory is that those transformations are completely characterized by knowing one single function S , the generating function of the transformation. This function, commonly known as Hamilton principal function, is the solution of the Hamilton-Jacobi equation

$$H\left(t, \xi^i, \frac{\partial S}{\partial \xi^i}\right) + \frac{\partial S}{\partial t} = 0, \quad i = 1, \dots, n$$

where H is the Hamiltonian function [19]. This equation establishes a close connection between the Hamilton description of mechanics and the Hamilton-Jacobi theory. To see the relation between the Lagrangian formulation and the Hamilton-Jacobi theory, let firstly observe that in the most general case, $S \equiv S(\xi, t)$ is a function of coordinates ξ and time t . Then, let assume that the transformation induced by S ,

$$\zeta_i = \frac{\partial S}{\partial \xi^i}, \quad i = 1, \dots, n$$

yields a solution of the equations of motion. By taking the total derivative of S and exploiting the Hamilton-Jacobi equation, we obtain

$$\frac{dS}{dt} = \sum_i \frac{\partial S}{\partial \xi^i} \dot{\xi}^i + \frac{\partial S}{\partial t} = \zeta_i \dot{\xi}^i - H = \mathcal{L} .$$

Finally, by integrating with respect to t we have that

$$S = \int \mathcal{L}(t, \xi_c, \zeta_c(\xi, \dot{\xi})) dt ,$$

where (ξ_c, ζ_c) is the solution of the equations of motion [19]. This proves that the Hamilton principal function is equal to the time integral of the Lagrangian function evaluated upon the solution of equations of motion.

In this article, we aim to employ the canonical divergence (10) as Hamilton principal function for simplifying the Hamilton's equations of motion. In particular, we mainly focus on the equations expressed by $\dot{\xi}^i = \frac{\partial H}{\partial \zeta_i}$, $i = 1, \dots, n$ which are referred in literature to as the tangent dynamics [5]. Thus, we show that these equations are differentiable gradient systems [18]. Finally, we succeed to prove that such dynamics is held by ∇ -geodesic. On the contrary, owing to the duality of the structure (g, ∇, ∇^*) we obtain the same result for the dual \mathcal{D}^* of the canonical divergence (10). In this case, the trajectory of the tangent dynamics is given by the ∇^* -geodesic.

As application of the theoretical approach so far described, we consider the standard Gaussian model. We then determine a nice physical interpretation of the tangent dynamics outlined in terms of the canonical divergence gradient system: the ∇ -geodesic corresponds to the well-known Uhlenbeck-Ornstein process which describes the probability that a free particle in Brownian motion after a time t has a velocity lying between v and $v + dv$, when it started at $t = 0$ with the velocity v_0 [20].

Usually, statistical manifolds are used to model families of probability distributions [4]. A pioneering work on the relationship between dynamical systems and probability distributions was established by Nakamura in [16]. Here, gradient systems on the manifolds of Gaussian and multinomial distributions are shown to be completely integrable Hamiltonian systems. This result has been generalized by Fujiwara and Amari in [9]. Here, the authors showed that the dualistic gradient flow can be characterized as completely integrable system and proved that solutions are unparametrized ∇ and ∇^* geodesics. In the present article, we take a different approach. Specifically, by minimizing the energy curve \mathcal{D}_γ we may interpret the canonical divergence \mathcal{D} as the Hamilton principal function associated with the Lagrangian \mathcal{L} . Then, we exploit the properties of the Hamilton principal function in the Hamiltonian description of the mechanics and show that the tangent dynamics is a gradient system. Finally, we determine the solution of such a dynamics that is the ∇ -geodesic connecting any two points p, q sufficiently close to each other.

The layout of the article is as follows. In Section II we show that the function \mathcal{D}_γ can be understood as modification of curve energy in Information Geometry. In Section III we prove that the canonical divergence \mathcal{D} is a Hamilton principal function associated with \mathcal{L} and we show that the tangent dynamics induced by \mathcal{L} is given by the gradient flow induced by \mathcal{D} . We then prove that dynamical paths between any two points sufficiently close to each other are ∇ or ∇^* geodesics.

Finally, we apply the methods so far described to the standard Gaussian model and compare them to the results obtained in [9]. In Section IV we draw some conclusions by outlining the results obtained in this work and discussing possible extensions. Finally, in Appendix A we present the basics of Information Geometry and we describe the canonical divergence introduced in [3].

II. KINETIC ENERGY AND INFORMATION GEOMETRY

Let $\mathcal{S} = (M, g, \nabla, \nabla^*)$ be a statistical manifold and $\gamma : [a, b] \rightarrow M$ ($a < b$) be a smooth curve in M connecting $p = \gamma(a)$ and $q = \gamma(b)$. The dual structure on γ is then induced from (g, ∇, ∇^*) by projection [1]. Therefore, the coefficients of g_γ , ∇_γ and ∇_γ^* corresponding to g , ∇ and ∇^* are given by [1],

$$g_\gamma(t) = g_{ij}(\gamma(t)) \dot{\gamma}^i \dot{\gamma}^j \quad (14)$$

$$\Gamma_\gamma(t) = \{ \dot{\gamma}^i(t) \dot{\gamma}^j(t) \Gamma_{kij}(t) + \ddot{\gamma}^j(t) g_{jk}(t) \} \dot{\gamma}^k(t) / g_\gamma(t) \quad (15)$$

$$\Gamma_\gamma^*(t) = \{ \dot{\gamma}^i(t) \dot{\gamma}^j(t) \Gamma_{kij}^*(t) + \ddot{\gamma}^j(t) g_{jk}(t) \} \dot{\gamma}^k(t) / g_\gamma(t) . \quad (16)$$

Since $(\gamma, g_\gamma, \nabla_\gamma, \nabla_\gamma^*)$ is 1-dimensional, it is dually flat. Now, Nagaoka and Amari introduced on dually flat manifolds a canonical divergence D of Bregman type (see Appendix A, Eq. (A10)). In particular, this divergence is defined on γ . It assumes the following form [1],

$$D(\gamma) := D_\gamma(\gamma(b), \gamma(a)) = \int \int_{a \leq s \leq t \leq b} g_\gamma(s) \frac{\mu(t)}{\mu(s)} ds dt, \quad \mu(t) := e^{\int_a^t \Gamma_\gamma(s) ds} \quad (17)$$

and we call it (g, ∇) -canonical divergence of the curve γ . Very remarkably, D_γ does not depend on the parametrization $t \mapsto \gamma(t)$ but its orientation and the (g, ∇^*) -divergence D^* of γ coincides with the (g, ∇) -divergence of the reversely oriented curve [1].

Remark II.1. *The statistical manifold $\mathcal{S} = (M, g, \nabla, \nabla^*)$ is self-dual when $\nabla = \nabla^*$. In this case, it turns out to be a Riemannian manifold endowed with the Levi-Civita connection $\bar{\nabla}_{LC} = \frac{1}{2}(\nabla + \nabla^*)$. Therefore, the projection of $\bar{\nabla}_{LC}$ to an arbitrary curve $\gamma : [a, b] \rightarrow M$ is given by $\bar{\nabla}_\gamma = \bar{\Gamma}_\gamma(t) \dot{\gamma}$, where*

$$\bar{\Gamma}_\gamma(t) = \frac{\langle (\bar{\nabla}_{LC})_{\dot{\gamma}} \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}} = \frac{1}{2} \frac{\frac{d}{dt} g_\gamma(t)}{g_\gamma(t)},$$

and we used the well-known equality $\langle (\bar{\nabla}_{LC})_{\dot{\gamma}} \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)} = \frac{1}{2} \frac{d}{dt} \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}$ [14]. Let us observe that the factor $\frac{\mu(t)}{\mu(s)}$ in the integral of Eq. (17) can be re-written as follows,

$$\mu(s, t) := \frac{\mu(t)}{\mu(s)} = e^{\int_s^t \bar{\Gamma}_\gamma(\tau) d\tau} .$$

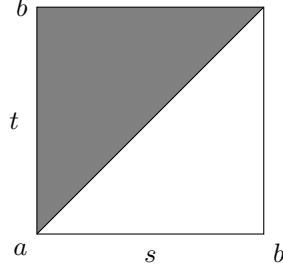


FIG. 1: The domain of the definite integral in Eq. (17) is highlighted in grey color.

By plugging $\bar{\Gamma}_\gamma$ into the μ factor of Eq. (17), we obtain

$$\mu(s, t) = e^{\int_s^t \bar{\Gamma}_\gamma(\tau) d\tau} = e^{\ln\left(\frac{\sqrt{g_\gamma(t)}}{\sqrt{g_\gamma(s)}}\right)} = \frac{\sqrt{g_\gamma(t)}}{\sqrt{g_\gamma(s)}}.$$

Consequently, from Eq.(17) we obtain

$$D(\gamma) = \int \int_{a \leq s \leq t \leq b} \sqrt{g_\gamma(s)} \sqrt{g_\gamma(t)} ds dt.$$

We may now observe from Fig. 1 that the area of the integration domain in Eq. (17) is one half the area of the rectangle $[a, b] \times [a, b]$. Therefore, by the symmetry properties of g_γ we obtain that

$$D(\gamma) = \frac{1}{2} \left(\int_a^b \sqrt{g_\gamma(s)} ds \right) \left(\int_a^b \sqrt{g_\gamma(t)} dt \right) = \frac{1}{2} \left(\int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} dt \right)^2$$

which proves that $D(\gamma)$ is one half the square of the length of γ . This suggests to consider $D(\gamma)$ as a modification of energy curve within Information Geometry.

Remark II.2. In order to show that $D(\gamma)$ does not depend on the parametrization of the curve, let us define $\tilde{\gamma}(\tau) = \gamma(t(\tau))$. Then, we have that

$$\dot{\gamma} = \frac{1}{t'(\tau)} \dot{\tilde{\gamma}}, \quad \ddot{\gamma} = \frac{1}{(t'(\tau))^2} \left(\ddot{\tilde{\gamma}} - \frac{t''(\tau)}{t'(\tau)} \dot{\tilde{\gamma}} \right).$$

By plugging these expressions into Eq. (14) and Eq. (15), we obtain

$$\begin{aligned} g_\gamma(t(\tau)) &= g_{\tilde{\gamma}}(\tau) \frac{1}{(t'(\tau))^2} \\ \Gamma_\gamma(t(\tau)) &= \left\{ \dot{\tilde{\gamma}}^i(\tau) \dot{\tilde{\gamma}}^j(\tau) \Gamma_{ijk}(\tilde{\gamma}(\tau)) + \ddot{\tilde{\gamma}}^j(\tau) g_{jk}(\tilde{\gamma}(\tau)) \right\} \frac{\dot{\tilde{\gamma}}^k(\tau)}{t'(\tau) g_{\tilde{\gamma}}(\tau)} \\ &\quad - \frac{t''(\tau)}{(t'(\tau))^2} \frac{\dot{\tilde{\gamma}}^j \dot{\tilde{\gamma}}^k g_{jk}(\tilde{\gamma}(\tau))}{g_{\tilde{\gamma}}(\tau)} \\ &= \Gamma_{\tilde{\gamma}}(\tau) \frac{1}{t'(\tau)} - \frac{t''(\tau)}{(t'(\tau))^2}. \end{aligned}$$

Consider now the factor $\frac{\mu(t)}{\mu(s)}$ in the integral of Eq. (17), which is given by,

$$\mu(s(\tau), t(\tau)) = \frac{\mu(t(\tau))}{\mu(s(\tau))} = e^{\int_{s(\tau)}^{t(\tau)} \Gamma_\gamma(\omega(\tau)) d\omega} .$$

Then, the integral $\int_{s(\tau)}^{t(\tau)} \Gamma_\gamma(\omega(\tau)) d\omega$ can be computed by

$$\begin{aligned} \int_{s(\tau)}^{t(\tau)} \Gamma_\gamma(\omega(\tau)) d\omega &= \int_{s(\tau)}^{t(\tau)} \Gamma_{\tilde{\gamma}}(\tau) \frac{1}{\omega'(\tau)} d\omega - \int_{s(\tau)}^{t(\tau)} \frac{\omega''(\tau)}{(\omega'(\tau))^2} d\omega \\ &= \int_{\tau^{-1}(s)}^{\tau^{-1}(t)} \Gamma_{\tilde{\gamma}}(\tau) d\tau - \int_{\tau^{-1}(s)}^{\tau^{-1}(t)} \frac{\omega''(\tau)}{\omega'(\tau)} d\tau , \end{aligned}$$

where we adopted the change of variable rule $d\omega = \omega'(\tau) d\tau$. The second integral of the latter equality can be computed as follows,

$$\int_{s^{-1}(\tau)}^{t^{-1}(\tau)} \frac{\omega''(\tau)}{\omega'(\tau)} d\tau = \ln [\omega'(\tau)] \Big|_{s^{-1}(\tau)}^{t^{-1}(\tau)} = \ln \frac{s'(\tau)}{t'(\tau)} .$$

By collecting all these results we then perform the last calculation

$$\begin{aligned} D(\gamma) &= \int \int_{a \leq s \leq t \leq b} g_\gamma(s) \frac{\mu(t)}{\mu(s)} ds dt \\ &= \int \int_{a \leq s(\tau) \leq t(\tau) \leq b} \frac{g_{\tilde{\gamma}}(\tau)}{(s'(\tau))^2} \frac{s'(\tau)}{t'(\tau)} \frac{\tilde{\mu}(t(\tau))}{\tilde{\mu}(s(\tau))} t'(\tau) d\tau s'(\tau) d\tau \\ &= \int \int_{\tau^{-1}(a) \leq \tau^{-1}(s) \leq \tau^{-1}(t) \leq \tau^{-1}(b)} g_{\tilde{\gamma}}(\tau) \frac{\tilde{\mu}(t(\tau))}{\tilde{\mu}(s(\tau))} d\tau d\tau = D(\tilde{\gamma}) , \end{aligned}$$

where $\mu(t(\tau)) = e^{\int_a^{\tau^{-1}(t)} \Gamma_{\tilde{\gamma}}(\tau) d\tau}$.

From here on, we assume that the statistical manifold (M, g, ∇, ∇^*) is dually flat (for more details see Appendix A). Consider an arbitrary curve $\gamma : [0, 1] \rightarrow M$ such that $\gamma(0) = p$ and $\gamma(1) = q$. The intrinsic geometry $(\gamma, g_\gamma, \nabla_\gamma, \nabla_\gamma^*)$ of γ is inherited from (g, ∇, ∇^*) through the Eqs. (14), (15), (16). We aim to prove that, whenever γ is ∇_γ or ∇_γ^* geodesic, the energy curve $D^*(\gamma)$ reduces to the form

$$\mathcal{D}_\gamma(p, q) = \int_0^1 t \langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)} dt \quad (18)$$

which is assumed by the canonical divergence introduced by Ay and Amari in [3] for an arbitrary curve γ . This result amounts to require some kinematical conditions which pave the way for a nice interpretation of \mathcal{D}_γ as the time integral of twice the kinetic energy along γ . To be more precise, consider firstly the general form of $D^*(\gamma)$,

$$D^*(\gamma) = \int \int_{0 \leq s \leq t \leq 1} g_\gamma(s) e^{\int_s^t \Gamma_\gamma^*(\tau) d\tau} ds dt , \quad (19)$$

where $\Gamma_\gamma^*(t)$ is given by Eq. (16). Since we assumed that (M, g, ∇, ∇^*) is dually flat, we can work with a ∇^* -affine system of coordinates $\{\xi^*\}$ and then write

$$\gamma(t) = (\xi^{*1}(\gamma(t)), \dots, \xi^{*n}(\gamma(t))) . \quad (20)$$

Therefore, we can see from Eq. (16) that Γ_γ^* can be written as follows:

$$\Gamma_\gamma^*(t) = \frac{\langle \ddot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}}{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} . \quad (21)$$

We can now decompose the acceleration $\ddot{\gamma}$ by a tangential component along γ plus a normal direction [14]:

$$\ddot{\gamma} = (\nabla_\gamma^*)_{\dot{\gamma}} \dot{\gamma} + II^*(\dot{\gamma}, \dot{\gamma}) , \quad (22)$$

where II^* denotes the second fundamental form of γ with respect to ∇^* . Owing to the intrinsic geometry of γ , we can find a re-parametrization $\tau_{\nabla_\gamma^*}(t)$ such that $\gamma(\tau_{\nabla_\gamma^*}(t))$ is the ∇_γ^* -geodesic from p to q . Hence, we have that $(\nabla_\gamma^*)_{\dot{\gamma}} \dot{\gamma} = 0$. From Eq. (22), we can see then that $\ddot{\gamma}$ is normal to the trajectory of γ . Therefore, from Eq. (21) we have that $\Gamma_\gamma^* \equiv 0$ and consequently $D^*(\gamma)$ assumes the following form,

$$\begin{aligned} D^*(\gamma) &= \int_0^1 t g_\gamma(\tau_{\nabla_\gamma^*}(t)) dt \\ &= \int_0^1 t \left\langle \frac{d(\gamma \circ \tau_{\nabla_\gamma^*})}{dt}(t), \frac{d(\gamma \circ \tau_{\nabla_\gamma^*})}{dt}(t) \right\rangle_{\gamma(t)} dt \\ &= \mathcal{D}_\gamma(p, q) . \end{aligned}$$

This result suggests the following physical interpretation: on one side $(\nabla_\gamma^*)_{\dot{\gamma}} \dot{\gamma} = 0$ says that the virtual particle of mass 1 subjected to the acceleration (22) has constant velocity in the direction of $\dot{\gamma}$; on the other side, $\ddot{\gamma} = II^*(\dot{\gamma}, \dot{\gamma})$ implies that the sum of all force fields acting on the virtual particle are orthogonal to the trajectory of γ . Hence, we may think of some *holonomic constraints* forcing the virtual particle as moving along the curve γ . In this way, we can say that \mathcal{D}_γ is the action functional [12] in the ambient space (M, g, ∇^*) measured by twice the following kinetic energy,

$$\mathcal{L}(\xi^*, \dot{\xi}^*, t) = \frac{t}{2} \dot{\xi}^{*T} G(\xi^*) \dot{\xi}^* . \quad (23)$$

Here $\xi^{*T} = (\xi^{*1}, \dots, \xi^{*n})$ and $\dot{\xi}^{*T} = (\dot{\xi}^{*1}, \dots, \dot{\xi}^{*n})$ are the generalized coordinates and the generalized velocities, respectively. Moreover, $G(\xi^*) = (g_{ij}(\xi^*))_{i,j}$, where g_{ij} are the components of the metric tensor g .

On the other hand, we may also work with ∇ -affine coordinates $\{\xi(t)\}$ and then write

$$\gamma(t) = (\xi^1(\gamma(t)), \dots, \xi^n(\gamma(t))) . \quad (24)$$

In this case, the coefficient of ∇_γ^* obtained by projecting ∇^* to γ is given by

$$\Gamma_\gamma^*(t) = \frac{\langle \nabla_{\dot{\gamma}}^* \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}} .$$

However, thanks to the duality (7) of the geometric structure (g, ∇, ∇^*) we can rewrite Γ_γ^* as follows,

$$\Gamma_\gamma^*(t) = \frac{\frac{d}{dt} \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}} - \frac{\langle \ddot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}} ,$$

where we used $\nabla_{\dot{\gamma}} \dot{\gamma} = \ddot{\gamma}$ as $\gamma(t)$ is written in ∇ -affine coordinates. Again, the acceleration of γ can be given in terms of a tangential component and a normal direction to the trajectory of γ :

$$\ddot{\gamma} = (\nabla_\gamma)_{\dot{\gamma}} \dot{\gamma} + II(\dot{\gamma}, \dot{\gamma}) ,$$

where II is the second fundamental form of γ with respect to ∇ . In analogy to the previous case, we can choose a re-parametrization $\tau_{\nabla_\gamma}(t)$ such that $\gamma(\tau_{\nabla_\gamma}(t))$ is the ∇_γ -geodesic from p to q . Hence, we have that $(\nabla_\gamma)_{\dot{\gamma}} \dot{\gamma} = 0$. This implies that $\Gamma_\gamma^*(t) = \frac{\frac{d}{dt} \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}$. Consequently, we have that

$$\int_s^t \Gamma_\gamma^*(\tau) d\tau = \int_s^t \frac{\frac{d}{dt} \langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}}{\langle \dot{\gamma}, \dot{\gamma} \rangle_{\gamma(t)}} d\tau = \log \left(\frac{g_\gamma(t)}{g_\gamma(s)} \right) .$$

By plugging it in Eq. (19) we then obtain that

$$\begin{aligned} D^*(\gamma) &= \int \int_{0 \leq s \leq t \leq 1} g_\gamma(s) e^{\log \left(\frac{g_\gamma(t)}{g_\gamma(s)} \right)} ds dt \\ &= \int_0^1 t g_\gamma(t) dt = \int_0^1 t \left\langle \frac{d(\gamma \circ \tau_{\nabla_\gamma})}{dt}(t), \frac{d(\gamma \circ \tau_{\nabla_\gamma})}{dt}(t) \right\rangle_{\gamma(t)} dt \\ &= \mathcal{D}_\gamma(p, q) . \end{aligned}$$

Therefore, we may interpret \mathcal{D}_γ as the action functional in the ambient space (M, g, ∇) measured by twice the following kinetic energy,

$$\mathcal{L}^*(t, \xi, \dot{\xi}) = \frac{t}{2} \langle \dot{\xi}(t), \dot{\xi}(t) \rangle_{\xi(t)} , \quad (25)$$

where $\{\xi(t)\}$ denotes a ∇ -affine coordinate system.

To sum up, we have showed that the function \mathcal{D}_γ introduced in [3] may be interpreted as the energy of an arbitrary curve γ in the framework of Information Geometry. In particular, it is the

time integral of twice the kinetic energy \mathcal{L} in the ambient space (M, g, ∇^*) and the time integral of twice the kinetic energy \mathcal{L}^* in the ambient space (M, g, ∇) . In the next section, we shall prove that the dynamics associated with \mathcal{L} and \mathcal{L}^* are unparametrized ∇ -geodesics and unparametrized ∇^* -geodesics, respectively.

III. DYNAMICS IN DUALY FLAT MANIFOLDS

Let (M, g, ∇, ∇^*) be a dually flat statistical manifold and $\gamma : [0, 1] \rightarrow M$ an arbitrary path such that $\gamma(0) = p$ and $\gamma(1) = q$. So far, we have introduced two kinetic energies along γ ; the one in the ambient space (M, g, ∇^*) is given by

$$K^* = \frac{t}{2} \left\langle \frac{d(\gamma \circ \tau_{\nabla^*})}{dt}(t), \frac{d(\gamma \circ \tau_{\nabla^*})}{dt}(t) \right\rangle_{\gamma(t)}, \quad (26)$$

where $\tau_{\nabla^*}(t)$ is a reparametrization of t such that $(\nabla_{\tilde{\gamma}}^*) \dot{\tilde{\gamma}} = 0$ with $\tilde{\gamma} = \gamma \circ \tau_{\nabla^*}$. The second kinetic energy is defined in the ambient space (M, g, ∇) and it is given by

$$K = \frac{t}{2} \left\langle \frac{d(\gamma \circ \tau_{\nabla})}{dt}(t), \frac{d(\gamma \circ \tau_{\nabla})}{dt}(t) \right\rangle_{\gamma(t)}, \quad (27)$$

where $\tau_{\nabla}(t)$ is a reparametrization of t such that $(\nabla_{\hat{\gamma}}) \dot{\hat{\gamma}} = 0$ with $\hat{\gamma} = \gamma \circ \tau_{\nabla}$.

Working with the ∇^* -affine coordinates (20) and the ∇ -affine coordinates (24), we can exploit the tools of the Lagrangian formulation of mechanics [19] for supplying the Euler-Lagrange equations for the energy curve $\mathcal{D}_\gamma(p, q)$ within both the ambient spaces, (M, g, ∇^*) and (M, g, ∇) .

A. Lagrangian approach to mechanics in dually flat manifolds

In this section we aim to prove that

$$\mathcal{D}(p, q) = \min\{\mathcal{D}_\gamma(p, q) \mid \gamma : [0, 1] \rightarrow M, \gamma(0) = p, \gamma(1) = q\}$$

in the ambient space (M, g, ∇^*) . In order to carry out this result, we rely on the following representation of the canonical divergence introduced in [3],

$$\mathcal{D}(p, q) = \int_0^1 \langle \dot{\sigma}_t(1), \dot{\sigma}(t) \rangle_{\sigma(t)} dt, \quad (28)$$

where $\sigma_t(s)$ ($0 \leq s \leq 1$) is the ∇ -geodesic from p to $\sigma(t)$. The importance of this representation relies on the statement claimed by Theorem A.1, namely $\dot{\sigma}_t(1) = \text{grad}_t \mathcal{D}_p(\cdot)$ (see Section A 2 of the Appendix A for more details).

Before claiming the next result we need the following definition,

Definition III.1. A curve $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$ is an unparametrized ∇ -geodesic if under suitable parametrization $t(\tau)$ it obeys the following geodesic equation

$$\ddot{\gamma}^h(t(\tau)) + \Gamma_{ij}^h(\gamma(t(\tau))) \dot{\gamma}^i(t(\tau)) \dot{\gamma}^j(t(\tau)) = 0 . \quad (29)$$

We are now ready to prove that the canonical divergence \mathcal{D} is a Hamilton principal function associated with the Lagrangian function (23), i.e. the time integral of \mathcal{L} evaluated at the solution of the Euler-Lagrange equations. More precisely, we are going to show that the solution of the Euler-Lagrange equations associated with \mathcal{L} is an unparametrized ∇ -geodesic $\xi_c(t)$ from p to q such that $t \dot{\xi}_c(t) = \dot{\sigma}(\tau(t))$ with σ being the ∇ -geodesic from p to q . Then, by recalling that the integral curves of $\text{grad}_t \mathcal{D}_p(\cdot)$ are ∇ -geodesic starting from p , we succeed to prove the following result.

Theorem III.1. Let (M, g, ∇, ∇^*) be a dually flat manifold and $p, q \in M$ sufficiently close to each other. Let \mathcal{L} be the Lagrangian function (23) set up in the (M, g, ∇^*) space. Then, the canonical divergence \mathcal{D} coincides with the Hamilton principle function associated with \mathcal{L} . In particular, we have that

$$\mathcal{D}(p, q) = \int_0^1 \mathcal{L}(t, \xi_c, \dot{\xi}_c) dt , \quad (30)$$

where ξ_c is an unparametrized ∇ -geodesic from p to q .

Proof. In order to prove our claim, we need to solve the Euler-Lagrange equations of \mathcal{L} . For this reason, we firstly write these equations in the local ∇^* -affine coordinates (20) (for the reason of readability, we drop out the symbol “*”). Therefore, the Lagrangian function (23) reads as follows,

$$\mathcal{L}(t, \xi, \dot{\xi}) = \frac{t}{2} g_{ij} \dot{\xi}^i \dot{\xi}^j .$$

Then, by noticing that g_{ij} depends only on ξ and not on $\dot{\xi}$ we obtain

$$\frac{\partial \mathcal{L}}{\partial \dot{\xi}^k} = \frac{t}{2} \partial_k g_{ij} \dot{\xi}^i \dot{\xi}^j \quad (31)$$

and

$$\frac{\partial \mathcal{L}}{\partial \xi^k} = \frac{t}{2} (g_{ik} \dot{\xi}^i + g_{kj} \dot{\xi}^j) . \quad (32)$$

Then, it follows that

$$\begin{aligned} \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\xi}^k} &= \frac{g_{ik} \dot{\xi}^i + g_{jk} \dot{\xi}^j}{2} + \frac{t}{2} (g_{ik} \ddot{\xi}^i + g_{kj} \ddot{\xi}^j) + \frac{t}{2} (\partial_r g_{ik} \dot{\xi}^i \dot{\xi}^r + \partial_r g_{kj} \dot{\xi}^j \dot{\xi}^r) \\ &= t g_{ik} \ddot{\xi}^i + g_{ik} \dot{\xi}^i + \frac{t}{2} (\partial_r g_{ik} \dot{\xi}^i \dot{\xi}^r + \partial_r g_{kj} \dot{\xi}^j \dot{\xi}^r) , \end{aligned} \quad (33)$$

where we used the symmetry $g_{ik} = g_{ki}$. Now, we can plug Eqs. (31), (33) in Eq. (4) and we arrive at

$$t \left(2 g_{ik} \ddot{\xi}^i + \partial_j g_{ik} \dot{\xi}^i \dot{\xi}^j + \partial_i g_{kj} \dot{\xi}^j \dot{\xi}^i - \partial_k g_{ij} \dot{\xi}^i \dot{\xi}^j \right) + 2 g_{ik} \dot{\xi}^i = 0,$$

where we changed r to j and i in the second and third terms, respectively. Now, since ξ is ∇^* -affine, we have that $\partial_j g_{ik} = \Gamma_{jki}$. In addition this is 3-symmetric tensor [1]. Therefore, multiplying by the inverse g^{hk} of the metric tensor g_{hk} , we obtain

$$\frac{t}{2} \left(2 \ddot{\xi}^h + \Gamma_{ij}^h \dot{\xi}^i \dot{\xi}^j \right) + \dot{\xi}^h = 0, \quad h = 1, \dots, n$$

where the Christoffel symbols Γ_{ij}^h of ∇ are given by $\Gamma_{ij}^h = g^{hk} \partial_k g_{ij}$ and the Einstein's notation is adopted. Finally, we get the Euler-Lagrange equations associated with \mathcal{L} ,

$$t \ddot{\xi}^h + \frac{t}{2} \Gamma_{ij}^h \dot{\xi}^i \dot{\xi}^j + \dot{\xi}^h = 0, \quad h = 1, \dots, n. \quad (34)$$

We may now observe that $t \ddot{\xi} + \dot{\xi} = \frac{d}{dt} (t \dot{\xi})$. In this way, Eq. (34) becomes

$$\frac{d}{dt} V^h + \Gamma_{ij}^h(\xi(t)) V_t^i \frac{\dot{\xi}^j}{2} = 0,$$

where we defined $V(t) := t \dot{\xi}(t)$. This implies that $V(t)$ is ∇ -parallel along the curve $\xi(t) =: \tilde{\xi}(2t)$. By noticing that $V(t) = 2t \dot{\tilde{\xi}}$ we can see that $V(t)$ is proportional to $\dot{\tilde{\xi}}$. Moreover, it is ∇ -parallel along $\tilde{\xi}$. Then we can conclude that $\xi(t)$ is an unparametrized ∇ -geodesic from p to q . Indeed, by choosing a re-parametrization $t(\tau)$ such that $\frac{d}{d\tau}(t(\tau)) = 2t$, in the new parametrization the velocity vector

$$\dot{\sigma}(\tau) = 2t \dot{\tilde{\xi}}(t) = V(t)$$

is ∇ -parallel. To prove this, we can carry out the following calculation:

$$\begin{aligned} \frac{d}{d\tau} \dot{\sigma}^h(\tau) &= \frac{dt(\tau)}{d\tau} \frac{d}{dt} \left(2t \dot{\tilde{\xi}}^h \right) = 2t \frac{d}{dt} \left(t \dot{\xi}^h \right) \\ &= 2t \left(\dot{\xi}^h(t) + \ddot{\xi}^h(t) \right) = 2t \left(-\frac{t}{2} \Gamma_{ij}^h \dot{\xi}^i \dot{\xi}^j \right) \\ &= -\Gamma_{ij}^h(\sigma(\tau)) \dot{\sigma}^i(\tau) \dot{\sigma}^j(\tau). \end{aligned}$$

This proves that the curve $\sigma(\tau) := \tilde{\xi}(t(\tau))$ is the ∇ -geodesic from p to q .

In order to conclude, let $\xi_c(t)$ be the unparametrized ∇ -geodesic which solves Eq. (34). Then, the curve $\sigma(\tau) = \xi_c(t(\tau))$ is the ∇ -geodesic from p to q . According to the result given by Theorem A.1 in the Appendix A, we know that the ∇ -geodesic σ_τ being part of the ∇ -geodesic σ from p

to q and corresponding to the interval $(\tau(0), \tau(t)]$ coincides with $\text{grad}_t \mathcal{D}_p(\xi_c(t))$. This implies that $\dot{\sigma} = \frac{\text{grad}_\tau \mathcal{D}_p(\cdot)}{\tau}$. Therefore, by naming $S(p, q)$ the evaluation of $\int_0^1 \mathcal{L}(\xi, \dot{\xi}) dt$ at the critical path $\xi_c(t)$, we obtain that

$$\begin{aligned} S(p, q) &= \int_0^1 \mathcal{L}(\xi_c, \dot{\xi}_c) dt = \int_0^1 \langle t \dot{\xi}_c, \frac{\dot{\xi}_c}{2} \rangle_{\xi_c(t)} dt \\ &= \int_0^1 \langle \dot{\sigma}(\tau(t)), \dot{\xi}_c(t) \rangle_{\tilde{\xi}_c(t)} dt = \int_0^1 \langle \text{grad}_{\tau(t)} \mathcal{D}_p(\cdot), \frac{\dot{\xi}_c(t)}{\tau(t)} \rangle_{\sigma(\tau(t))} dt \\ &= \mathcal{D}(p, q) . \end{aligned}$$

This proves claim (30). \square

The dual function of \mathcal{D} has been also introduced in [3]. Likewise \mathcal{D} , the dual canonical divergence \mathcal{D}^* can be viewed as the path integral of a specific vector field. More precisely, we have that

$$\mathcal{D}^*(p, q) = \int_0^1 \langle \dot{\sigma}_t^*(1), \dot{\sigma}^*(t) \rangle_{\sigma(t)} dt , \quad (35)$$

where $\sigma_t^*(s)$ ($0 \leq s \leq 1$) is the ∇^* -geodesic from p to $\sigma^*(t)$ and σ^* is the ∇^* -geodesic from p to q . Again, the vector field $\dot{\sigma}_t^*(1)$ is a gradient vector field along $\sigma^*(t)$, i.e. $\dot{\sigma}_t^*(1) = \text{grad}_t \mathcal{D}_p^*(\cdot)$. This plays a key role for interpreting \mathcal{D}^* as the Hamilton principal function associated with the Lagrangian function (25) in the (M, g, ∇) ambient space.

The Euler-Lagrange equations of \mathcal{L}^* are given by

$$t \ddot{\xi}^h + \frac{t}{2} \Gamma_{ij}^{*h} \dot{\xi}^i \dot{\xi}^j + \dot{\xi}^h = 0 , \quad h = 1, \dots, n , \quad (36)$$

where Γ_{ij}^{*h} are the Christoffel's symbols of the ∇^* -connection. The solution of Eq. (36) is an unparametrized ∇^* -geodesic $\xi_c(t)$ such that $t \dot{\xi}_c(t) = \dot{\sigma}^*(\tau)$ is the velocity vector of the ∇^* -geodesic σ^* from p to q . Therefore, by applying the same methods of Theorem III.1 we can prove that the dual canonical divergence \mathcal{D}^* given by Eq. (35) is a Hamilton principal function associated with the Lagrangian function \mathcal{L}^* . In particular, we obtain that

$$\mathcal{D}^*(p, q) = \min\{\mathcal{D}_\gamma(p, q) \mid \gamma : [0, 1] \rightarrow M, \gamma(0) = p, \gamma(1) = q\}$$

in the ambient space (M, g, ∇) .

Remark III.1. *By using the Lagrangian formalism, Theorem III.1 yields $\mathcal{D}^*(\gamma) = \mathcal{D}(p, q)$ when γ is a ∇ -geodesic. This result is confirmed by the geometric viewpoint as claimed in [1]. Here, the authors showed indeed that the (g, ∇) -divergence of γ coincides with the canonical divergence of Bregman type between $\gamma(1)$ and $\gamma(0)$ whenever γ is either ∇ -geodesic or ∇^* -geodesic.*

B. Hamiltonian approach to mechanics in dually flat manifolds

The Lagrangian formulation of mechanics [19] on the configuration manifold M is defined on the tangent bundle TM in terms of the Lagrangian $\mathcal{L} : TM \rightarrow \mathbb{R}$. Then, by considering a system of local coordinates $\boldsymbol{\xi} = \{\xi^1, \dots, \xi^n\}$ on M , we can define an action integral S which is a functional over the set of differentiable path $\boldsymbol{\xi} : [0, 1] \rightarrow M$ with fixed endpoints,

$$S(\boldsymbol{\xi}) = \int_0^1 \mathcal{L}(t, \boldsymbol{\xi}(t), \dot{\boldsymbol{\xi}}(t)) dt. \quad (37)$$

The evaluation of S at the solution of the Euler-Lagrange equations $\boldsymbol{\xi}_c$ gives a two-point function [6],

$$S(p, q) = S(\boldsymbol{\xi}_c), \quad p = \boldsymbol{\xi}(0) \text{ and } q = \boldsymbol{\xi}(1) \quad (38)$$

which is known in literature as the Hamilton principal function [19]. In the previous section, we proved that $\mathcal{D}(p, q) = \int_0^1 \mathcal{L}(t, \boldsymbol{\xi}_c, \dot{\boldsymbol{\xi}}_c) dt$ when \mathcal{L} is given by Eq. (23) and $\boldsymbol{\xi}_c$ is an unparametrized ∇ -geodesic from p to q . In addition, we also proved that $\mathcal{D}^*(p, q) = \int_0^1 \mathcal{L}^*(t, \boldsymbol{\xi}_c, \dot{\boldsymbol{\xi}}_c) dt$ when \mathcal{L}^* is given by Eq. (25) and $\boldsymbol{\xi}_c$ is an unparametrized ∇^* -geodesic from p to q .

The Hamilton principal function is the generating function of a canonical transformation in the phase space of the system [5]. More precisely, the dynamics on the phase space is described by the Hamiltonian formulation of mechanics [19]. This is defined on the cotangent bundle T^*M and it takes place by replacing the generalized velocity $\dot{\boldsymbol{\xi}}$ with the generalized momentum $\boldsymbol{\zeta}$. Specifically, the Legendre transform $\mathcal{FL} : TM \rightarrow T^*M$ relates the tangent bundle and the cotangent bundle as follows,

$$(\boldsymbol{\xi}, \dot{\boldsymbol{\xi}}) \mapsto (\boldsymbol{\xi}^i, \boldsymbol{\zeta}) = (\boldsymbol{\xi}, d_{\dot{\boldsymbol{\xi}}} \mathcal{L}), \quad d_{\dot{\boldsymbol{\xi}}} \mathcal{L} : TM \rightarrow \mathbb{R}, \quad (39)$$

where $d_{\dot{\boldsymbol{\xi}}} \mathcal{L}$ denotes the differential of \mathcal{L} at $\dot{\boldsymbol{\xi}}$. Owing to the metric tensor g and the regularity of \mathcal{L} , the differential $d_{\dot{\boldsymbol{\xi}}} \mathcal{L}$ is given by

$$\begin{aligned} d_{\dot{\boldsymbol{\xi}}} \mathcal{L}(Y) &= \langle \text{grad}_{\dot{\boldsymbol{\xi}}} \mathcal{L}, Y \rangle_{\boldsymbol{\xi}(t)} = \langle g^{ij} \frac{\partial \mathcal{L}}{\partial \dot{\xi}^i} \partial_j, Y^k \partial_k \rangle_{\boldsymbol{\xi}(t)} \\ &= g_{jk} g^{ij} \frac{\partial \mathcal{L}}{\partial \dot{\xi}^i} Y^k = \frac{\partial \mathcal{L}}{\partial \dot{\xi}^i} Y^i, \end{aligned}$$

where we used the canonical identification $TTM \simeq TM$ and the well-known expression $\text{grad} f = g^{ij} \frac{\partial f}{\partial \xi^i} \partial_j$, with g^{ij} the components of the inverse of g and $\partial_i = \frac{\partial}{\partial \xi^i}$. The momentum conjugate to $\boldsymbol{\xi}$ is then given by

$$\zeta_i = \frac{\partial \mathcal{L}}{\partial \dot{\xi}^i}, \quad i = 1, \dots, n.$$

In this way a Hamiltonian function can be defined by

$$H : T^*M \rightarrow \mathbb{R}, \quad H(t, \boldsymbol{\xi}, \boldsymbol{\zeta}) = \zeta_i \dot{\xi}^i - \mathcal{L}(t, \boldsymbol{\xi}, \dot{\boldsymbol{\xi}}), \quad (40)$$

where $\dot{\boldsymbol{\xi}} \equiv \dot{\boldsymbol{\xi}}(\boldsymbol{\xi}, \boldsymbol{\zeta})$ is now function of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$. Then, the Euler-Lagrange equations transform into Hamilton's equations,

$$\frac{d\xi^i}{dt} = \frac{\partial H}{\partial \zeta_i}, \quad \frac{d\zeta_i}{dt} = -\frac{\partial H}{\partial \xi^i}. \quad (41)$$

The canonical transformation, i.e. the one that preserves the structure of Eq. (41), induced by the Hamilton principal function is given by

$$\frac{\partial S}{\partial \xi^i} = \zeta_i, \quad i = 1, \dots, n. \quad (42)$$

Therefore, having in hands the canonical divergence \mathcal{D} which we proved is the Hamilton principal function associated with the Lagrangian function (23), we can try to simplify the Hamilton equations and hopefully solve it in order to describe dynamical systems in a general dually flat statistical manifold. To this aim, let us perform the following computation

$$\begin{aligned} \zeta_i &= \frac{\partial \mathcal{L}(t, \boldsymbol{\xi}, \dot{\boldsymbol{\xi}})}{\partial \dot{\xi}^i} = \frac{\partial}{\partial \dot{\xi}^i} \left(\frac{t}{2} g_{ij}(\boldsymbol{\xi}) \dot{\xi}^i \dot{\xi}^j \right) \\ &= t g_{ij}(\boldsymbol{\xi}) \dot{\xi}^j. \end{aligned} \quad (43)$$

Therefore, the fiber Legendre transform $\mathcal{FL} : TM \rightarrow T^*M$ and its inverse can be written as follows,

$$\boldsymbol{\zeta}^T = t G(\boldsymbol{\xi}) \dot{\boldsymbol{\xi}} \quad (44)$$

$$\dot{\boldsymbol{\xi}} = \frac{1}{t} G^{-1}(\boldsymbol{\xi}) \boldsymbol{\zeta}^T, \quad (45)$$

where $G(\boldsymbol{\xi}) = (g_{ij}(\boldsymbol{\xi}))_{i,j}$ and G^{-1} denotes the inverse matrix of G . At this point, from Eq. (40) we obtain the Hamiltonian function of the system that reads as follows,

$$\begin{aligned} H(t, \boldsymbol{\xi}, \boldsymbol{\zeta}) &= \frac{1}{2t} g^{ij}(\boldsymbol{\xi}) \zeta_i \zeta_j \\ &= \frac{1}{2t} \boldsymbol{\zeta} G^{-1}(\boldsymbol{\xi}) \boldsymbol{\zeta}^T. \end{aligned} \quad (46)$$

From Eq. (41) we can compute the Hamilton equations of the dynamics,

$$\dot{\xi}^i = \frac{1}{t} g^{ij}(\boldsymbol{\xi}) \zeta_j \quad (47)$$

$$\dot{\zeta}_i = \frac{1}{2t} \partial_i g^{jk}(\boldsymbol{\xi}) \zeta_j \zeta_k. \quad (48)$$

Eq. (47) is usually referred to as the tangent dynamics associated with the Lagrangian \mathcal{L} . The next result claims that such a dynamics can be understood in terms of gradient systems.

Theorem III.2. *Let (M, g, ∇, ∇^*) be a dually flat statistical manifold and consider the canonical divergence $\mathcal{D}_p(q)$ as Hamilton characteristic function associated with \mathcal{L} . Then the equation $\dot{\xi}^i = \frac{\partial H}{\partial \zeta_i}$, $i = 1, \dots, n$ becomes the following dynamical system*

$$\dot{\xi}(t) = \frac{1}{t} \text{grad}_t \mathcal{D}_p. \quad (49)$$

Proof. According to Eq. (42), we have that $\zeta_j = \frac{\partial \mathcal{D}_p}{\partial \xi^j}$. Then, from Eq. (47) we obtain that

$$\dot{\xi}^i(t) = \frac{1}{t} g^{ij} \zeta_j(t) = \frac{1}{t} (\text{grad}_t \mathcal{D}_p)^i,$$

where we used $\text{grad} f = g^{ij} \frac{\partial_i f}{\partial \xi^i} \partial_j$, with f any suitable smooth function and the sum over i, j is intended. This proves Eq. (49). \square

We are now ready to determine the trajectories of the tangent dynamics described by Eq. (49).

Theorem III.3. *Let (M, g, ∇, ∇^*) be a dually flat statistical manifold and $p, q \in M$ sufficiently close to each other. Then the tangent dynamics $\dot{\xi} = \frac{1}{t} \text{grad}_t \mathcal{D}_p$ evolves along the ∇ -geodesic connecting p and q .*

Proof. We know that $\text{grad}_t \mathcal{D}_p = \dot{\sigma}_t(1)$, where $\sigma_t(s)$ is a ∇ -geodesic starting from p (see Theorem A.1 in the Appendix A). Consider now the ∇ -geodesic $\sigma(t)$ from p to q . Then, $\sigma_t(s)$ ($0 \leq s \leq 1$) can be defined by $\sigma_t(s) := \sigma(st)$. This implies that $\dot{\sigma}(t) = \frac{\dot{\sigma}_t(1)}{t}$ and then $\sigma(t)$ solves Eq. (49). \square

Remark III.2. *By stepping back to the proof of Theorem III.1, we can see that the solution of the Euler-Lagrange equations associated with the Lagrangian function (23) is an unparametrized ∇ -geodesic from p to q . In [9], the authors proved that the solution of the gradient system*

$$\dot{\eta}^i = -g^{ij} \partial_j D(\eta)$$

converges to q from p along an unparametrized ∇ -geodesic. Here, η is the ∇^ -affine coordinates of the point p and $D(\eta)$ is the canonical divergence introduced by Amari and Nagaoka (see (A10) in the Appendix A) from q to $p(\theta)$.*

On the contrary, we obtained Eq. (49) by means of the Hamiltonian formalism. Thanks to this approach we succeed to prove that the solution of Eq. (49) is the ∇ -geodesic from p to q .

Remark III.3. *Let $U \subset M$ be an open set such that there exists an homeomorphism $\varphi : U \rightarrow \mathbb{R}^n$ with $\varphi(p) = (\xi^1, \dots, \xi^n)$ for all $p \in U$. A gradient system is a system of differential equations of the form [18]*

$$\dot{\xi} = -\text{grad} V(\xi),$$

where $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is a C^∞ function on $\varphi(U)$. The system of differential equations given by Eq. (47) and Eq. (48) is instead usually referred to as Hamiltonian system [5]. This system defines the local flow of the Hamiltonian vector field $X_H \in \mathcal{T}(\mathbb{T}^*\mathbb{M})$ which is defined by

$$X_H = \left(\frac{\partial H}{\partial \zeta}, -\frac{\partial H}{\partial \xi} \right).$$

The Hamilton equations can be put in the following form

$$\begin{pmatrix} \dot{\xi} \\ \dot{\zeta} \end{pmatrix} = J \text{grad } H, \quad J = \begin{pmatrix} 0 & \text{Id}_n \\ -\text{Id}_n & 0 \end{pmatrix}$$

where $J \text{grad } H = X_H$ is known as the Hamilton symplectic gradient [5]. This form of the Hamilton equations highlights a symplectic structure. Indeed, by the vector X_H we may define a symplectic form on $\mathbb{T}\mathbb{T}^*\mathbb{M}$ as

$$\omega(X_H, Y) := dH(Y) = \sum_i \left(\frac{\partial H}{\partial \xi^i} \frac{\partial Y}{\partial \zeta_i} - \frac{\partial H}{\partial \zeta_i} \frac{\partial Y}{\partial \xi^i} \right), \quad Y \in \mathbb{T}\mathbb{T}^*\mathbb{M}.$$

To sum up, a gradient system underlies a Riemannian structure, while a Hamiltonian system carries a symplectic structure. Our observation suggests that symplectic structures play important role also in Information Geometry (see related work [15]).

Theorem III.2 selects one of the Hamilton equations, namely Eq. (47), and shows that it can be interpreted as gradient system as the canonical divergence \mathcal{D} is a function of the coordinates $\{\xi\}$.

The dual canonical divergence $\mathcal{D}^*(p, q)$ is a Hamilton principal function associated with the Lagrangian (25). Therefore, we can apply the same arguments as above in order to obtain a Hamiltonian function H^* . In this case, the tangent dynamics is given by the following differential equation

$$\dot{\xi} = \frac{1}{t} \text{grad}_t \mathcal{D}_p^*, \quad (50)$$

where ξ is any path in \mathbb{M} from p to q and \mathcal{D}_p^* is given by Eq. (35). Finally, we succeed to determine the tangent trajectories of the dynamics (50) induced by the Lagrangian function \mathcal{L}^* .

Theorem III.4. *Let $(\mathbb{M}, g, \nabla, \nabla^*)$ be a dually flat statistical manifold and $p, q \in \mathbb{M}$. Then the tangent dynamics $\dot{\xi} = \frac{1}{t} \text{grad}_t \mathcal{D}_p^*$ evolves along the ∇^* -geodesic connecting p and q .*

Proof. The dual version of Theorem A.1 (see Appendix A) holds true, as well. In particular the gradient of the divergence $\mathcal{D}_p^*(\cdot)$ is given by ∇^* -geodesics. Therefore, by using the same methods as in the proof of Theorem III.3 we get the desired result. \square

C. Example of Theorem III.3

Let us consider the family of Gaussian probability distributions with mean μ and variance σ^2 :

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad (51)$$

where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ are the local coordinates of $M = \{p \equiv p(x; \mu, \sigma)\}$. The information geometry of this statistical model is given in terms of the Fisher metric g^F , whose components are computed by

$$g_{ij}^F = \int_{\mathbb{R}} p(x; \mu, \sigma) \partial_i \ln(p(x; \mu, \sigma)) \partial_j \ln(p(x; \mu, \sigma)) dx \quad (52)$$

and in terms of the exponential connection $\nabla = \nabla^e$ and the mixture connection $\nabla^* = \nabla^m$. The connection symbols are given by the following relations,

$$\Gamma_{ij,k}^e = \int_{\mathbb{R}} p(x; \mu, \sigma) \partial_i \partial_j \ln(p(x; \mu, \sigma)) \partial_k \ln(p(x; \mu, \sigma)) dx, \quad (53)$$

$$\Gamma_{ij,k}^m = \int_{\mathbb{R}} p(x; \mu, \sigma) (\partial_i \partial_j \ln(p(x; \mu, \sigma)) + \partial_i \ln(p(x; \mu, \sigma)) \partial_j \ln(p(x; \mu, \sigma))) \partial_k \ln(p(x; \mu, \sigma)) dx,$$

where $i, j \in \{1, 2\}$ and $\partial_1 = \frac{\partial}{\partial \mu}$ and $\partial_2 = \frac{\partial}{\partial \sigma}$.

Consider now two probability distributions of this family, namely $p \equiv p_1(x; \mu_1, \sigma_1)$ and $q \equiv p_2(x; \mu_2, \sigma_2)$. The tangent dynamics (49) can be obtained by the gradient system induced by the canonical divergence \mathcal{D} . We now take this avenue for studying the tangent dynamics within the standard Gaussian model. In order to compute $\mathcal{D}(p_1, p_2)$ we exploit Eq. (28). In particular, we look for the e -geodesic σ^e from (μ_1, σ_1) to (μ_2, σ_2) and the 1-parameter family of e -geodesics σ_t^e from (μ_1, σ_1) to $\sigma^e(t) = (\mu(t), \sigma(t))$.

According to Eq. (52), the Fisher metric of (51) is given by

$$g^F = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \quad (54)$$

Whereas, the non-zero connection symbols of ∇^e and ∇^m are obtained by Eq. (53) and they read as follows,

$$\begin{aligned} \Gamma_{121}^e = \Gamma_{211}^e = -\frac{2}{\sigma^3}, \quad \Gamma_{222}^e = -\frac{6}{\sigma^3} \\ \Gamma_{112}^m = \frac{2}{\sigma^3}, \quad \Gamma_{222}^m = \frac{2}{\sigma^3}, \end{aligned} \quad (55)$$

respectively. By denoting g^{ij} the components of the inverse of g^F , the Christoffel symbols are computed by $\Gamma_{ij}^k = g^{ij}\Gamma_{ij,k}$, where the Einstein's notation is adopted. Finally we arrive at the geodesic equations of ∇^e ,

$$\begin{cases} \ddot{\mu}(t) - \frac{4\dot{\mu}(t)\dot{\sigma}(t)}{\sigma(t)} = 0 \\ \ddot{\sigma}(t) - \frac{3\dot{\sigma}^2(t)}{\sigma(t)} = 0 \end{cases}. \quad (56)$$

At this point we can look for the ∇^e -geodesic $\sigma^e(t)$ from (μ_1, σ_1) to (μ_2, σ_2) . By assuming $0 < \sigma_2 < \sigma_1$, we obtain the following expression,

$$\sigma^e(t) = \left(\frac{\mu_1\sigma_2^2(1-t) + \mu_2\sigma_1^2 t}{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}, \frac{\sigma_1\sigma_2}{\sqrt{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}} \right) \quad (57)$$

and then

$$\dot{\sigma}^e(t) = \frac{\sigma_1\sigma_2}{(\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t)^2} \left((\mu_2 - \mu_1)\sigma_1\sigma_2, \frac{\sigma_2^2 - \sigma_1^2}{\sqrt{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}} \right). \quad (58)$$

Moreover, by requiring the ending points to be (μ_1, σ_1) and $\sigma^e(t)$, the 1-parameter ∇^e -geodesics are then given by

$$\sigma_t^e(s) = \left(\frac{\mu_2\sigma_1^2 ts + \mu_1\sigma_2^2(1-ts)}{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)ts}, \frac{\sigma_1\sigma_2}{\sqrt{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)ts}} \right), \quad s, t \in [0, 1]. \quad (59)$$

From Eq. (59) we have that

$$\dot{\sigma}_t^e(1) = \frac{\sigma_1\sigma_2 t}{(\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t)^2} \left((\mu_2 - \mu_1)\sigma_1\sigma_2, \frac{\sigma_2^2 - \sigma_1^2}{\sqrt{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}} \right). \quad (60)$$

The last element we need for evaluating the integral (28) is the metric tensor g^F computed along the exponential geodesic $\sigma^e(t)$. Then, we have that

$$g(\sigma^e(t)) = \begin{pmatrix} \frac{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}{\sigma_1^2\sigma_2^2} & 0 \\ 0 & 2 \frac{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}{\sigma_1^2\sigma_2^2} \end{pmatrix}$$

By performing the integral (28) we finally arrive at

$$\mathcal{D}(p_1, p_2) = \frac{(\mu_1 - \mu_2)^2}{2\sigma_1^2} + \frac{\sigma_2^2}{2\sigma_1^2} + \ln\left(\frac{\sigma_1}{\sigma_2}\right) - 1,$$

which proves that $\mathcal{D}(p_1, p_2)$ is the K-L divergence between two standard normal distributions.

Let us now address our investigation to the tangent dynamics described by Eq. (49). In the particular case under consideration, the computation of $\text{grad } \mathcal{D}_{p_1}$ leads to the following expression

$$\text{grad } \mathcal{D}_{p_1} = \frac{(\mu(t) - \mu_1)\sigma^2(t)}{\sigma_1^2} \partial_1 + \frac{\sigma(t)}{2} \left(\frac{\sigma^2(t)}{\sigma_1^2} - 1 \right) \partial_2, \quad (61)$$

where $\partial_1 = \frac{\partial}{\partial \mu}$ and $\partial_2 = \frac{\partial}{\partial \sigma}$. Consequently, the tangent dynamics (49) of the model (51) is described by the following ordinary differential equations,

$$\begin{cases} \dot{\mu} = \frac{\sigma^2(t)}{t\sigma_1^2} (\mu(t) - \mu_1), & \mu(0) = \mu_1, \quad t \in (0, 1] \\ \dot{\sigma} = \frac{\sigma(t)}{2t} \left(\frac{\sigma^2(t)}{\sigma_1^2} - 1 \right), & \sigma(0) = \sigma_1, \quad t \in (0, 1] \end{cases} \quad (62)$$

This system can be easily integrated. By assuming $\sigma_1 > \sigma_2$ and recalling the positivity of $\sigma(t)$ we obtain

$$\sigma^e(t) = (\mu(t), \sigma(t)) = \begin{cases} \mu(t) = \frac{\mu_1 \sigma_2^2 - t(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2)}{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}, & t \in [0, 1] \\ \sigma(t) = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_2^2 + (\sigma_1^2 - \sigma_2^2)t}}, & t \in [0, 1] \end{cases} \quad (63)$$

A slight comparison between Eq. (57) and Eq. (63) immediately leads to the conclusion that the latter curve is the ∇^e -geodesic between p_1 and p_2 . However, we can rely on the ∇^e -affine coordinates $\boldsymbol{\theta} = (\theta^1, \theta^2)$, where

$$\theta^1 = \frac{\mu}{\sigma^2}, \quad \theta^2 = \frac{1}{2\sigma^2}$$

and then we have that (63) reduces to the straight line from p_1 to p_2 ,

$$\boldsymbol{\theta}(t) = \begin{cases} \theta^1(t) = \theta_{p_1}^1 + t(\theta_{p_2}^1 - \theta_{p_1}^1), & t \in [0, 1] \\ \theta^2(t) = \theta_{p_1}^2 + t(\theta_{p_2}^2 - \theta_{p_1}^2), & t \in [0, 1] \end{cases}, \quad (64)$$

where $\theta_{p_1}^i$ and $\theta_{p_2}^i$ are the ∇^e -affine coordinates of points p_1 and p_2 , respectively.

In [17], the authors proved that the Uhlenbeck-Ornstein (UO) process characterized by the Gaussian distribution

$$\mu = \mu_0 + v(s - s_0), \quad \sigma^2 = 2D(s - s_0) \quad (65)$$

is a geodesic with respect to ∇^e -connection. Here, μ_0 is the starting point at $s = s_0$ and v, D are usually referred as the drift coefficient and the diffusion coefficient, respectively. In addition, the authors also supplied a universal relation between the "Newtonian" time s and the affine time t which is given by [17], $s - s_0 = \frac{b}{t + a}$, for suitable constants $a, b \in \mathbb{R}$. In order to prove that the

∇ -geodesic solution of the tangent dynamics (62) can be seen as a UO process, we can consider a rescaling of the time t such that

$$\theta^i(t) = \theta^i(0) + \frac{\tau}{t + \tau} (\theta_q^i - \theta^i(0))$$

and we assume that q is the uniform distribution, i.e. $\theta_q^2 = 0$. Therefore, by setting

$$\mu_0 = \frac{\theta^1(0) - \theta_q^1}{2\theta^2(0)}, \quad v = \frac{\theta^1(0)}{2\tau\theta^2(0)}, \quad D = \frac{1}{4\tau\theta^2(0)},$$

we obtain

$$\mu(t) = \mu_0 + v(t + \tau), \quad \sigma^2(t) = 2D(t + \tau)$$

which is the UO process as it is figured out in [9], up to some constants.

IV. CONCLUSIONS

Riemannian geometry enters the realm of mechanics in connection with the inertial quality of mass and principles of classical mechanics have shown that the inertia of mass is characterized by the kinetic energy [12]. Starting from this fundamental statement, in this article we established a relationship between Information Geometry (IG) [1] and classical mechanics [2] by means of the divergence function \mathcal{D} recently introduced by Ay and Amari in [3]. Given a statistical manifold (M, g, ∇, ∇^*) , we have firstly considered the energy of an arbitrary curve γ proposed in [1]. Then, we proved that such a general energy reduces to the function \mathcal{D}_γ , which consequently is interpreted as the time integral of twice the kinetic energy (26) in the ambient space (M, g, ∇^*) and the time integral of twice the kinetic energy (27) in the ambient space (M, g, ∇) . In particular, by focusing our investigation on dually flat manifolds, we applied the Lagrangian formalism to both the Lagrangian functions, \mathcal{L} given by Eq. (23) and \mathcal{L}^* given by Eq. (25). In one case, the solution of the Euler-Lagrange equations is an unparametrized ∇ -geodesic; while in the other case, the Euler-Lagrange equations are solved by an unparametrized ∇^* -geodesic. These results allowed us to show that both functions, the canonical divergence \mathcal{D} and its dual function \mathcal{D}^* , coincide with the Hamilton principal function of \mathcal{L} and the Hamilton principal function of \mathcal{L}^* , respectively.

By means of these Hamilton principal functions, we moved to the Hamiltonian description of the dynamics in a general dually flat manifold. Then, we succeeded to describe the Hamiltonian equations of dynamics in term of the gradient of \mathcal{D} as well as in terms of the gradient of \mathcal{D}^* . In particular, we showed that the tangent dynamics is a gradient system. According to the particular

form of the Hamiltonian function (40), we proved Theorem III.3 which claims that the solution of the tangent dynamics is the ∇ -geodesic.

Finally, we applied Theorem III.3 to the standard Gaussian distribution (51) and proved that the canonical divergence \mathcal{D} is the Kullback-Leibler divergence. In addition we gave an interpretation of the tangent dynamics (49) in terms of the Uhlenbeck-Ornstein process which describes the probability that a free particle in Brownian motion has a given velocity v after a time t .

The general case is still not properly understood. More precisely, given a general statistical manifold (M, g, ∇, ∇^*) the geometric theory put forward in [8] would suggest to use the following action functionals

$$\begin{aligned} L(\gamma) &= \int_0^1 \langle \dot{\gamma}(t), \Pi_t(p) \rangle_{\gamma(t)} dt, & \Pi_t(p) &= P_{\sigma_t^*} \dot{\sigma}_t(0), \\ L^*(\gamma) &= \int_0^1 \langle \dot{\gamma}(t), \Pi_t^*(p) \rangle_{\gamma(t)} dt, & \Pi_t^*(p) &= P_{\sigma_t^*}^* \dot{\sigma}_t^*(0), \end{aligned}$$

where σ_t and σ_t^* are ∇ and ∇^* geodesic, respectively, connecting p to $\gamma(t)$ and P, P^* denote the ∇ and ∇^* parallel transports, respectively. Here, the gradient flow of surfaces of constant L is obtained by projecting $\Pi_t(p)$ along the ∇^* -geodesic from p to q . Whereas, the gradient flow of surfaces of constant L^* is obtained by projecting $\Pi_t^*(p)$ along the ∇ -geodesic from p to q . However, it remains unclear how to use the Lagrangian formalism in this context.

-
- [1] S.-I. Amari and H. Nagaoka. Methods of Information Geometry, volume 191 of Translations of Mathematical monographs. Oxford University Press, 2000.
 - [2] V. I. Arnol'd. Mathematical Methods of Classical Mechanics. Springer, Berlin, 1978.
 - [3] N. Ay and S.-I. Amari. A novel approach to canonical divergences within information geometry. Entropy, 17:8111–8129, 2015.
 - [4] N. Ay, J. Jost, H. Van Le, and L. Schwachhöfer. Information Geometry. Springer International Publishing, 1st edition, 2017.
 - [5] J.F. Cariñena, I. Ibort, G. Marmo, and G. Morandi. Geometry from Dynamics, Classical and Quantum. Springer-Berlin, 2015.
 - [6] F. M. Ciaglia, F. Di Cosmo, D. Felice, S. Mancini, G. Marmo, and J. M. Pérez-Pardo. Hamilton-jacobi approach to potential functions in information geometry. Journal of Mathematical Physics, 58(6):063506, 2017.
 - [7] S. Eguchi. Geometry of minimum contrast. Hiroshima Math. J., 22(3):631–647, 1992.
 - [8] D. Felice and N. Ay. Towards a canonical divergence within information geometry. arXiv:1806.11363 [math.DG], 2018.

- [9] A. Fujiwara and S.-I. Amari. Gradient systems in view of information geometry. Physica D: Nonlinear Phenomena, 80(3):317 – 327, 1995.
- [10] M. Henmi and R. Kobayashi. Hooke’s law in statistical manifolds and divergences. Nagoya Math. J., 159:1–24, 2000.
- [11] J. Jost. Riemannian Geometry and Geometric Analysis. Univesitext. Springer International Publishing, seventh edition, 2017.
- [12] C. Lanczos. The Variational Principles of Mechanics. University of Toronto Press, Toronto, 1952.
- [13] S. L. Lauritzen. Differential geometry in statistical inference. Lecture Notes-Monograph Series, 10:163–218, 1987.
- [14] J. M. Lee. Riemannian Manifolds: An introduction to Curvature, volume 176. Springer-Verlag New York, 1st edition, 1997.
- [15] M. Leok and J. Zhang. Connecting information geometry and geometric mechanics. Entropy, 19(10), 2017.
- [16] Y. Nakamura. Completely integrable gradient systems on the manifolds of gaussian and multinomial distributions. Japan Journal of Industrial and Applied Mathematics, 10(2):179, Jun 1993.
- [17] T. Obata, H. Hara, and K. Endo. Differential geometry of nonequilibrium processes. Phys. Rev. A, 45:6997–7001, May 1992.
- [18] S. Smale. Differentiable dynamical systems. Bull. Amer. Math. Soc., 73(6):747–817, 11 1967.
- [19] T. T. Taylor. Mechanics : classical and quantum. Pergamon Press Oxford ; New York, 1st ed. edition, 1976.
- [20] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. Phys. Rev., 36:823–841, Sep 1930.

APPENDIX A: INFORMATION GEOMETRY

In what follows, we present the main information geometric concepts of relevance to our work. For a more detailed description, we refer to Refs. [1] and [4].

A *statistical manifold* is the quadruple (M, g, ∇, ∇^*) , where g is a Riemannian metric tensor and ∇ and ∇^* are torsion free linear connections on TM coupled by Eq. (7) [4].

To the affine connections ∇ and ∇^* we can associate two curvature tensors,

$$\mathcal{R}(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z, \quad (\text{A1})$$

$$\mathcal{R}^*(X, Y)Z = \nabla_X^* \nabla_Y^* Z - \nabla_Y^* \nabla_X^* Z - \nabla_{[X, Y]}^* Z. \quad (\text{A2})$$

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

Given the metric structure on M , we can also consider the Riemann curvature tensors of ∇ and

∇^* that are defined as follows

$$R(X, Y, Z, W) := g(\mathcal{R}(XY)Z, W) \quad (\text{A3})$$

$$R^*(X, Y, Z, W) := g(\mathcal{R}^*(XY)Z, W) . \quad (\text{A4})$$

The relation between R and R^* is given by [13]

$$R(X, Y, Z, W) = -R^*(X, Y, W, Z) . \quad (\text{A5})$$

For a general statistical manifold (M, g, ∇, ∇^*) it is not true that

$$R(X, Y, Z, W) = -R(X, Y, W, Z) . \quad (\text{A6})$$

This condition identifies a class of statistical manifolds called *conjugate symmetric* [13]. Very remarkably, for a conjugate symmetric statistical manifold the following property holds true:

$$R(X, Y, Z, W) = R^*(X, Y, Z, W), \quad \forall X, Y, Z, W \in \mathcal{T}(M) . \quad (\text{A7})$$

Finally, we call (M, g, ∇, ∇^*) a *dually flat* statistical manifold if it is conjugate symmetric and $R \equiv 0$.

Dually flat manifolds are widely studied by Amari and Nagaoka [1]. In particular, given a dually flat manifold (M, g, ∇, ∇^*) , there exist affine coordinate systems $\theta = \{\theta^1, \dots, \theta^n\}$ with respect to ∇ and $\eta = \{\eta_1, \dots, \eta_n\}$ with respect to ∇^* such that

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_j^i, \quad (\text{A8})$$

where δ_j^i denotes the delta-Dirac function. These coordinate systems are called *dual coordinate systems*. In this case, there exist two functions $\Psi, \Phi : M \rightarrow \mathbb{R}$ such that $d\Psi = \eta_i d\theta^i$ and $d\Phi = \theta_i d\eta^i$. The correspondence between θ and η is given in terms of the functions Ψ and Φ by means of the following *Legendre transformation*

$$\Phi + \Psi = \theta^i \eta_i, \quad (\text{A9})$$

where the summation over $i = 1, \dots, n$ is intended. Through Φ and Ψ a function $D : M \times M \rightarrow \mathbb{R}$ can be defined by

$$D(p, q) := \Phi(p) + \Psi(q) - \eta_i(p)\theta^i(q), \quad \forall p, q \in M, \quad (\text{A10})$$

and it is called the *canonical divergence* on M of Bregman type [1].

1. Canonical Divergence within Information Geometry

In this section we describe the canonical divergence introduced in [3] by Ay and Amari. It 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. A 1). Then, the difference between p and q in M is supplied by the exponential map of the connection ∇ . In particular, assuming that $p \in U_q$ and $U_q \subset M$ is a ∇ -geodesic neighborhood of q , the difference vector from q to p is defined as (see (B) of Fig. A 1)

$$X_q(p) := X(q, p) := \exp_q^{-1}(p) = \dot{\gamma}_{q,p}(0) , \quad (\text{A11})$$

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

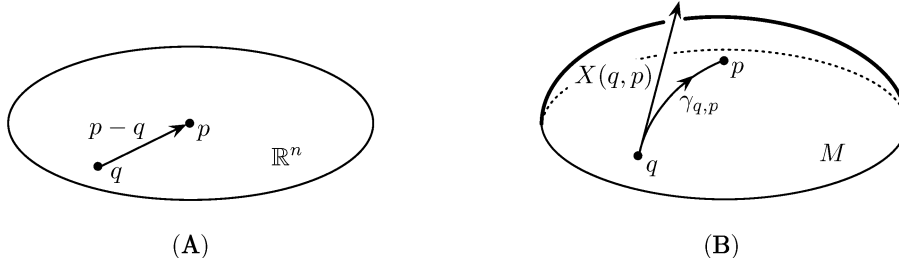


FIG. 2: On the left, (A) illustrates the difference vector $p - q$ in the linear vector space \mathbb{R}^n ; whereas, in (B) we can see the difference vector $X(q, p) = \dot{\gamma}_{q,p}(0)$ in M as the inverse of the exponential map at q (This Figure comes from [3]).

Therefore, the divergence proposed by Ay and Amari in [3] is defined as the path integral

$$\mathcal{D}_\gamma(p, q) := \int_0^1 \langle X_t(p), \dot{\gamma}(t) \rangle_{\gamma(t)} dt , \quad (\text{A12})$$

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. (A12), $X_t(p)$ is the vector field along $\gamma(t)$ given by Eq. (A11) as follows,

$$X_t(p) = X(\gamma(t), p) = \exp_{\gamma(t)}^{-1}(p) . \quad (\text{A13})$$

After elementary computation Eq. (A12) reduces to [3],

$$\mathcal{D}_\gamma(p, q) = \int_0^1 t \|\dot{\gamma}_{p,q}(t)\|^2 dt, \quad (\text{A14})$$

where $\gamma_{p,q}(t)$ is the ∇ -geodesic from p to q . If we consider definition (A12) for general path γ then $\mathcal{D}_\gamma(p, q)$ will be depending on γ . On the contrary, if the vector field $X_t(p)$ is integrable, then $\mathcal{D}_\gamma(p, q) =: \mathcal{D}(p, q)$ turns out to be independent of the path from q to p .

2. Gradient fields in dually flat manifolds

Consider a dually flat manifold (M, g, ∇, ∇^*) . In order to prove that (A 1) is a gradient vector field, we recall that the canonical divergence $\mathcal{D}_p(q)$ assumes the following form,

$$\mathcal{D}_p(q) = \int_0^1 \langle \dot{\sigma}_t(1), \dot{\sigma}(t) \rangle_{\sigma(t)} dt,$$

where σ is the ∇ -geodesic from p to q and σ_t is the ∇ -geodesic from p to $\sigma(t)$. Let us observe that $X_t(p) = -\dot{\sigma}_t(1)$. We shall show that $\dot{\sigma}_t(1) = \text{grad}_t \mathcal{D}_p$. To accomplish this aim, we firstly need the following result. This is well-known in literature (see Ref. [1]). However, here we propose an alternative derivation.

Proposition A.1. *Let (M, g, ∇, ∇^*) be a dually flat statistical manifold. Then each ∇ -geodesic from p is perpendicular to the level-hypersurfaces of the function $\mathcal{D}_p(\cdot)$.*

Proof. Let $\Sigma(\tau, t)$ ($-\varepsilon < \tau < \varepsilon$, $0 \leq t \leq 1$) be a variation of ∇ -geodesics such that

$$\begin{cases} \Sigma(0, t) = \sigma(t) \\ \Sigma(\tau, 0) = p \quad \forall \tau \\ \langle V, T \rangle(1) = 0 \end{cases}$$

where $V := \partial_\tau \Sigma(0, t)$ is the ∇ -Jacobi vector field and $T := \partial_t \Sigma(\tau, t)$ is the velocity vector along $\sigma(t)$. Consider now a ∇ -geodesic $\sigma_{\tau,t}(s)$ such that $\sigma_{\tau,t}(0) = p$ and $\sigma_{\tau,t}(1) = \Sigma(\tau, t)$. By defining $S := \partial_s \sigma_{\tau,t}(1)$, we immediately get the proportionality relation $S = t T$. In this setting, we have then

$$\begin{aligned} \langle \text{grad } \mathcal{D}_p(\cdot), \partial_\tau \Sigma(0, 1) \rangle &= \left. \frac{d}{d\tau} \right|_{\tau=0} \mathcal{D}_p(\Sigma(\tau, 1)) = \int_0^1 V \langle T, S \rangle dt \\ &= \int_0^1 (\langle \nabla_V^* T, S \rangle + \langle T, \nabla_V S \rangle) dt \\ &= \int_0^1 (\langle \nabla_T^* V, S \rangle + \langle T, \nabla_V S \rangle) dt \\ &= \int_0^1 (T \langle V, S \rangle - \langle V, \nabla_T S \rangle + \langle T, \nabla_V S \rangle) dt, \end{aligned}$$

where we used the torsion freeness of ∇^* and the dual property (7) of ∇ and ∇^* . By recalling that $S = t T$ we can observe that

$$\begin{aligned}\langle V, \nabla_T tT \rangle &= \langle V, T + \nabla_T T \rangle = \langle V, T \rangle \\ \langle T, \nabla_V tT \rangle &= \langle T, 0 + \nabla_V T \rangle,\end{aligned}$$

where the first line is obtained because σ is ∇ -geodesic and the second line follows by taking the derivative with respect to τ as indicated by the definition of V . Therefore, we have that

$$\langle \text{grad } \mathcal{D}_p(\cdot), \partial_\tau \Sigma(0, 1) \rangle = [\langle V, T \rangle]_0^1 - \int_0^1 \langle V, T \rangle dt + \int_0^1 t \langle T, \nabla_T V \rangle dt ,$$

where we used the torsion freeness of ∇ .

From classical Riemannian Geometry we know that V is a Jacobi field of a ∇ -geodesic variation if and only if [14]

$$D_t^2 V + \mathcal{R}(V, \dot{\sigma})\dot{\sigma} = 0 , \quad (\text{A15})$$

where \mathcal{R} denotes the curvature tensor of ∇ given by Eq. (A1). Since M is ∇ -flat, Eq. (A15) reduces to $D_t^2 V = 0$. This implies that, if we set $W = \nabla_T V$ then W is ∇ -parallel along $\sigma(t)$, i.e. $D_t W = 0$.

Let us now assume that $V(0) = 0$ and $D_t V(0) = W(0)$. Because of ∇ -flatness the solution of $D_t^2 V = 0$ is given by $V(t) = t W(t)$. Therefore, we have that

$$\begin{aligned}\langle \text{grad } \mathcal{D}_p(\cdot), \partial_\tau \Sigma(0, 1) \rangle &= [\langle V, T \rangle]_0^1 - \int_0^1 \langle V, T \rangle dt + \int_0^1 t \langle T, \nabla_T V \rangle dt \\ &= - \int_0^1 (\langle t W, T \rangle - t \langle T, \nabla_T tW \rangle) dt \\ &= - \int_0^1 (t \langle W, T \rangle - t \langle T, \nabla_T W \rangle - t \langle W, T \rangle) dt \\ &= 0 .\end{aligned}$$

This proves the desired result. □

Remark A.1. *The dual version of Proposition A.1 holds true, as well. More precisely, given a dually flat statistical manifold (M, g, ∇, ∇^*) the dual function of the canonical divergence reads as follows,*

$$\mathcal{D}_p^*(q) = \int_0^1 \langle \dot{\sigma}_t^*(1), \dot{\sigma}^*(t) \rangle_{\sigma^*(t)} dt,$$

where σ^* is the ∇^* -geodesic from p to q and σ_t^* is the ∇^* -geodesic from p to $\sigma^*(t)$. Then, each ∇^* -geodesic from p is perpendicular to the level-hypersurfaces of the function $\mathcal{D}_p^*(\cdot)$.

Actually, this result has been obtained by Henmi and Kobayashi in a context more general than dually flatness [10]. They have considered the claim of Proposition A.1 under the conditions

$$\nabla R = 0, \quad R(X, Y, Y, Y) = 0 \quad \forall X, Y \in \mathcal{T}(M),$$

which are obviously implied by dually flatness.

Theorem A.1. *Let (M, g, ∇, ∇^*) be a dually flat manifold. Then, we have that*

$$\dot{\sigma}_t(1) = \text{grad}_{\sigma(t)} \mathcal{D}_p, \quad (\text{A16})$$

where $\sigma(t)$ is the ∇ -geodesic between $p, q \in M$ and $\sigma_t(s)$ is the ∇ -geodesic from p to $\gamma(t)$.

Proof. Consider a ∇ -geodesic $\sigma(t)$ such that $\sigma(0) = p$ and $\sigma(1) = q$. Then, for all $t \in [0, 1]$ consider the ∇ -geodesic $\sigma_t(s)$ such that $\sigma_t(0) = p$ and $\sigma_t(1) = \sigma(t)$. We have that

$$\mathcal{D}_p(q) = \int_0^1 \langle \dot{\sigma}_t(1), \dot{\sigma}(t) \rangle_{\sigma(t)} dt.$$

Consider now a variation $q_\tau = \sigma_1(1 + \tau)$ of the endpoint q . By considering the ∇ -geodesic variation $\Sigma(\tau, t)$ such that $\Sigma(\tau, 0) = p$ and $\Sigma(\tau, 1) = q_\tau$ we can apply methods of proposition (A.1) and then we have that

$$\begin{aligned} \langle \text{grad } \mathcal{D}_p(\cdot), \left. \frac{d}{d\tau} \right|_{\tau=0} q_\tau \rangle &= \left. \frac{d}{d\tau} \right|_{\tau=0} \mathcal{D}_p(\cdot) \\ &= \langle \dot{\sigma}_1(1), \left. \frac{d}{d\tau} \right|_{\tau=0} q_\tau \rangle. \end{aligned}$$

From Proposition (A.1) we know that ∇ -geodesics are perpendicular to level-hypersurfaces $\mathcal{D}_p(\cdot)$. This implies

$$\text{grad } \mathcal{D}_p(q) = \dot{\sigma}_1(1).$$

In particular, we have that $\text{grad}_t \mathcal{D}_p = \text{grad } \mathcal{D}_p(\sigma(t)) = \dot{\sigma}_t(1)$. □