Information Geometry
https://doi.org/10.1007/s41884-024-00149-w

RESEARCH

Information geometry of the Otto metric

Nihat Ay[1,2,3]

Received: 29 June 2024 / Revised: 14 September 2024 / Accepted: 23 September 2024

Abstract
We introduce the dual of the mixture connection with respect to the Otto metric which represents a new kind of exponential connection. This provides a dual structure consisting of the mixture connection, the Otto metric as a Riemannian metric, and the new exponential connection. We derive the geodesic equation of this exponential connection, which coincides with the Kolmogorov forward equation of a gradient flow. We then derive the canonical contrast function of the introduced dual structure.

Keywords Wsserstein geometry · Otto metric · Exponential connection · Dual structure · Canonical contrast function

1 Introduction

In recent years, information geometry has been greatly extended in a direction that is generally referred to as "Wasserstein geometry". The need for this important direction arises from the fact that classical information-geometric structures such as the Fisher-Rao metric or the Kullback–Leibler divergence do not take into account the geometry of the underlying sample space. The incorporation of the sample space geometry into information geometry is expected to give rise to many improvements within application fields, in particular in data science and machine learning [1–3].

The very first article of the journal Information Geometry (INGE), [4], written by Amari et al., is dealing with the geometry of optimal transportation plans for

This article is dedicated to Shun-ichi Amari on the occasion of his 88th birthday.

Communicated by Hiroshi Matsuzoe.

✉ Nihat Ay
nihat.ay@tuhh.de

1 Institute for Data Science Foundations, Hamburg University of Technology, Blohmstraße 15, 21073 Hamburg, Germany

2 Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

3 Faculty of Mathematics and Computer Science, Leipzig University, Augustusplatz 10, 04109 Leipzig, Germany

Published online: 06 November 2024

a general cost function. Their family, which turns out to be an exponential family, is studied from the perspective of classical information geometry by means of the entropy-relaxed optimal transportation problem where the Cuturi function plays a central role. In the same first volume of INGE, Wong [5] defines a family of divergences based on optimal transport that parallels the family of α-divergences from classical information geometry [6]. He proves that the induced geometries of the corresponding statistical models are dually projectively flat with constant sectional curvatures and that a generalised Pythagorean theorem holds true. In [7], Malagó et al. provide an extensive study of the L^2-Wasserstein Riemannian geometry of Gaussian densities. In particular, they derive explicit matrix representations of the corresponding inner product, the Levi-Civita connection, the exponential map, the Riemannian geodesic, and the analogue to the natural gradient [8]. Around the same time, Li and Montúfar [3] pursue a similar idea of the natural gradient via optimal transport on weighted graphs within the context of a finite sample space. They derive the L^2-Wasserstein Riemannian geometry for a statistical model in the corresponding finite-dimensional simplex. This study is then extended in [9] from finite to continuous sample spaces. A detailed review, including further contributions to INGE, is given in the introduction [10] to the INGE special issue "Information Geometry on Optimal Transport".

The general field of Wasserstein geometry, without reference to information geometry, has a long history with a major corpus of fundamental results. For a review of the field and its literature we refer to the comprehensive standard textbooks of Villani [11, 12]. When it comes to bridging between Wasserstein geometry and information geometry one faces various challenges. Information geometry originates from parametric statistics and utilises methods from differential, in particular Riemannian, geometry. It is dealing with the geometry of parametrised statistical models where the parametrisation is interpreted as a coordinate system. Wasserstein geometry, on the other hand, studies the non-parametric geometry of the full set of probability measures without assuming the structure of a differentiable manifold. Despite this fact, a formal application of the Riemannian calculus turns out to be extremely far-reaching. This has been demonstrated by Otto in his seminal work [13], building on and extending the work [14]. Derivations with this so-called "Otto calculus" are often revealed as exact calculations when restricted to an appropriate manifold of probability measures thereby still modelling the non-parametric setting. A classical choice is given by the density manifold [15], which is a smooth infinite-dimensional Fréchet manifold in the sense of, for instance, [16]. This choice provides the setting of the present article. An alternative approach is based on the formalism of a Banach manifold modelled on Orlicz spaces [17, 18]. Quite generally, the Otto calculus opens up the possibility of a fruitful exchange of concepts, formalisms, and methods between the fields of Wasserstein geometry and information geometry, which is documented by the above short review of recent works within this exchange.

This article focusses on the generalisation of the following fundamental facts from information geometry: 1. the exponential connection is the dual of the mixture connection with respect to the Fisher-Rao metric, 2. both connections are flat, giving rise to a dually flat structure, 3. the dually flat structure is coupled with a canonical contrast function, the Kullback–Leibler divergence. In this article, we replace the Fisher-Rao metric by the so-called Otto metric and provide results related to these three statements for that replacement. In particular, we show that the dual of the mixture connection with respect to the Otto metric, a new kind of exponential connection, leads to the Kolmogorov forward equation as geodesic equation.

This article is organised as follows. In Sections 2 und 3, we first introduce the density manifold, related basic spaces of study, and review some facts about vector fields and flows. In Section 4, we present the Otto metric and characterise it in terms of its gradients. In Sections 5, we introduce the dual of the mixture connection, a new kind of exponential connection, and then derive the geodesic equation and its solution. In Section 6, we finally provide the canonical contrast function and its dual associated with the new exponential connection.

2 Preliminary definitions

We consider a finite-dimensional, compact, connected Riemannian C^∞-manifold (M, g) without boundary. By μ_g we denote the measure determined by the volume form of g. As M is compact, we may assume that μ_g is a probability measure. The vector space $C^\infty(M)$ of real-valued functions on M that are k times continuously differentiable for all $k = 0, 1, \ldots,$ is a graded Fréchet space, where the grading is given by the sequence of the supremum norms of derivatives up to kth order, $k = 0, 1, \ldots$.

The main set we are going to study from the information-geometric perspective is the set of all probability measures on M with smooth positive density with respect to μ_g:

$$P_+^\infty(M) \;:=\; \left\{ f\mu_\mathrm{g} \;:\; f \in C^\infty(M),\; f > 0,\; \int_M f\, d\mu_\mathrm{g} = 1 \right\}.$$

It is a convex subset of the vector space

$$S^\infty(M) \;:=\; \left\{ f\mu_\mathrm{g} \;:\; f \in C^\infty(M) \right\}$$

of signed measures on M with smooth density with respect to μ_g. Note that $S^\infty(M)$ inherits the Fréchet space structure from $C^\infty(M)$ and is therefore complete. With the standard total variation as a norm of a signed measure, which corresponds to the $L^1(\mu_\mathrm{g})$-norm on $C^\infty(M)$ (see [19]), $S^\infty(M)$ is not complete. We have the dual pairing

$$\psi \,:\, C^\infty(M) \times S^\infty(M) \;\to\; \mathbb{R}, \qquad (f, \mu) \;\mapsto\; \int_M f\, d\mu, \tag{1}$$

which allows us to identify every $f \in C^\infty(M)$ with the continuous linear form $\psi_f := \psi(f, \cdot)$ on $S^\infty(M)$ and every $\mu \in S^\infty(M)$ with the continuous linear form $\psi_\mu := \psi(\cdot, \mu)$ on $C^\infty(M)$. Consider now the closed linear subspace

$$S_0^\infty(M) \;:=\; \left\{ f\mu_\mathrm{g} \;:\; f \in C^\infty(M),\; \int_M f\, d\mu_\mathrm{g} = 0 \right\}$$

of $S^\infty(M)$. With that, $P_+^\infty(M)$ is an open convex subset of the closed affine subspace $\mu_g + S_0^\infty(M)$ of $S^\infty(M)$. In order to translate the dual pairing (1) to a pairing for measures in the subspace $S_0^\infty(M)$, it is required to identify functions in $C^\infty(M)$ whenever they differ by a constant function. This yields the induced dual pairing

$$\psi : C^\infty(M)/\mathbb{R} \times S_0^\infty(M) \to \mathbb{R}, \qquad (f + \mathbb{R}, \mu) \mapsto \int_M f \, d\mu. \tag{2}$$

Given a class $f + \mathbb{R} \in C^\infty(M)/\mathbb{R}$, we can naturally identify it with the continuous linear form $\psi_{f+\mathbb{R}} := \psi(f + \mathbb{R}, \cdot)$ on $S_0^\infty(M)$. Correspondingly, we can identify any measure $\mu \in S_0^\infty(M)$ with the continuous linear form $\psi_\mu := \psi(\cdot, \mu)$ on $C^\infty(M)/\mathbb{R}$.

The set $P_+^\infty(M)$ can be equipped with the structure of an infinite-dimensional differentiable Fréchet manifold (see [15]). For $\mu \in P_+^\infty(M)$, the tangent and (continuous) cotangent spaces are given by

$$T_\mu P_+^\infty(M) = \{\mu\} \times S_0^\infty(M), \quad T_\mu^* P_+^\infty(M) = \{\mu\} \times C^\infty(M)/\mathbb{R},$$

which follows from the dual pairing (2). Throughout this article, we simplify the notation and omit the base point μ whenever appropriate. In particular, we also write $T_\mu P_+^\infty(M) = S_0^\infty(M)$ and $T_\mu^* P_+^\infty(M) = C^\infty(M)/\mathbb{R}$. For the corresponding tangent and cotangent bundles we obtain

$$T P_+^\infty(M) = P_+^\infty(M) \times S_0^\infty(M), \quad T^* P_+^\infty(M) = P_+^\infty(M) \times C^\infty(M)/\mathbb{R}.$$

Note that, for two measures $\mu, \nu \in P_+^\infty(M)$, we can trivially identify $T_\mu P_+^\infty(M)$ with $T_\nu P_+^\infty(M)$ and $T_\mu^* P_+^\infty(M)$ with $T_\nu^* P_+^\infty(M)$ by mapping $(\mu, f \mu_g)$ to $(\nu, f \mu_g)$ and $(\mu, f + \mathbb{R})$ to $(\nu, f + \mathbb{R})$, respectively. As we will see in Sect. 5, this will give rise to two corresponding affine connections, the mixture and exponential connection. Here, the mixture connection does not require any further specification, while the exponential connection depends on the way we identify the individual tangent spaces with the respective cotangent spaces. This is typically done with an inner product as an additional structure on the manifold. Depending on the particular inner product we obtain different exponential connections. For the classical Fisher-Rao inner product we have the classical exponential connection as the dual of the mixture connection. In this article, we replace the Fisher-Rao inner product by the Otto inner product [13] and study the corresponding exponential connection as the dual of the mixture connection.

While the assumptions on the sample space M made in this article follow the setting of classical works, such as [15, 20, 21], they are quite restrictive in view of some application domains. In particular, it is desirable to incorporate non-compact sample spaces. However, the modest aim of this article is to highlight fundamental structures and related results within a convenient and non-technical setting. We expect that analogous results will hold within more general settings, which gives rise to further research.

3 Vector fields and flows

We denote the set of smooth vector fields on M by $\mathcal{T}(M)$. Given a vector field $X \in \mathcal{T}(M)$, we consider the corresponding differential equation for a curve c_x in M that goes through x at time $t = 0$:

$$\dot{c}_x \;=\; X(c_x), \quad c_x(0) = x. \tag{3}$$

Under the assumptions of this article, there exists a unique smooth curve $c_x : \mathbb{R} \to M$ satisfying (3). With the collection of the curves $c_x, x \in M$, we can define for all $t \in \mathbb{R}$,

$$\varphi_t : \; M \;\to\; M, \quad x \;\mapsto\; \varphi_t(x) := c_x(t).$$

Each φ_t is a C^∞-diffeomorphism of M, and the following holds:

$$\varphi_{s+t} \;=\; \varphi_t \circ \varphi_s, \qquad s, t \in \mathbb{R}.$$

Thus, the family φ_t, $t \in \mathbb{R}$, is a one-dimensional subgroup of the diffeomorphism group of M which we associate with the vector field $X \in \mathcal{T}(M)$. We refer to it as the *flow generated by* X, or simply the *flow of* X.

In what follows, we study how the flow of a vector field acts on a probability measure $\mu \in P_+^\infty(M)$. For that we consider the densities $\rho_t \in C^\infty(M)$ of the corresponding push-forward probability measures $\mu_t := (\varphi_t)_*(\mu)$ with respect to μ, that is

$$\rho(x, t) \;:=\; \rho_t(x) \;:=\; \frac{d\mu_t}{d\mu}(x), \quad x \in M.$$

We now consider the infinitesimal decrease of the probability density in x at time $t = 0$ due to the flow of X. More formally, we set

$$\mathrm{div}_\mu(X)(x) \;:=\; -\left.\frac{\partial}{\partial t}\right|_{t=0} \rho(x, t), \tag{4}$$

which defines the *divergence* of X in μ,

$$\mathrm{div}_\mu : \;\mathcal{T}(M) \;\to\; C^\infty(M), \quad X \;\mapsto\; \mathrm{div}_\mu(X).$$

We have the following characterising equation for the divergence:

$$\int_M Xf\, d\mu \;=\; -\int_M f\, \mathrm{div}_\mu(X)\, d\mu, \qquad \text{for all } f \in C^\infty(M). \tag{5}$$

To verify (5), consider $f \in C^\infty(M)$:

$$\int_M (Xf)(x)\, \mu(dx) = \int_M \left.\frac{d}{dt}\right|_{t=0} (f \circ \varphi_t)(x)\, \mu(dx)$$

$$
\begin{aligned}
&= \frac{d}{dt}\Big|_{t=0} \int_M (f \circ \varphi_t)(x)\, \mu(dx) \\
&= \frac{d}{dt}\Big|_{t=0} \int_M f(x)\, \mu_t(dx) \\
&\quad \text{(by the general transformation rule for integrals)} \\
&= \frac{d}{dt}\Big|_{t=0} \int_M f(x)\, \rho(x,t)\, \mu(dx) \\
&= \int_M f(x) \left\{ \frac{\partial}{\partial t}\Big|_{t=0} \rho(x,t) \right\} \mu(dx) \\
&= -\int_M f(x)\, \mathrm{div}_\mu(X)(x)\, \mu(dx). \qquad \text{(by(4))}
\end{aligned}
$$

We now list a few basic properties of the divergence which we will use throughout this article.

Proposition 1 *Let X be a smooth vector field in $\mathcal{T}(M)$, and let μ be a probability measure in $P_+^\infty(M)$. Then:*
(a) *For any smooth function f on M, we have the Leibniz rule*[1]

$$
\mathrm{div}_\mu(fX) \; = \; \langle \mathrm{grad}\, f, X\rangle + f\, \mathrm{div}_\mu(X). \tag{6}
$$

(b) *For any probability measure $\nu = \rho\mu$ on M with $\rho \in C^\infty(M)$, $\rho > 0$, $\mathrm{div}_\nu(X)$ is determined by $\mathrm{div}_\mu(X)$ according to the following transformation rules:*

$$
\mathrm{div}_\nu(X) = \frac{1}{\rho}\, \mathrm{div}_\mu(\rho X) \tag{7}
$$

$$
= X(\ln \rho) + \mathrm{div}_\mu(X). \tag{8}
$$

Proof For a function $g \in C^\infty(M)$, we have

$$
\begin{aligned}
-\int_M fg\, \mathrm{div}_\mu(X)\, d\mu &= \int_M X(fg)\, d\mu \\
&= \int_M \{(Xf)g + f(Xg)\}\, d\mu \\
&= \int_M \langle \mathrm{grad}\, f, X\rangle g\, d\mu + \int (fX)g\, d\mu \\
&= \int_M g\langle \mathrm{grad}\, f, X\rangle\, d\mu - \int_M g\, \mathrm{div}_\mu(fX)\, d\mu.
\end{aligned}
$$

This implies

$$
\int_M g\, \mathrm{div}_\mu(fX)\, d\mu = \int_M g\langle \mathrm{grad}\, f, X\rangle\, d\mu + \int_M fg\, \mathrm{div}_\mu(X)\, d\mu
$$

[1] Throughout this article, we use the common notation $\langle X, Y\rangle$ for $\mathrm{g}(X, Y)$.

$$= \int_M g\left(\langle \text{grad}\, f, X\rangle + f\, \text{div}_\mu(X)\right)\, d\mu.$$

Given that this equality has to hold for all g, we obtain the rule (6). Now we come to the transformation rule (7):

$$\begin{aligned}\int_M Xf\, d\nu &= \int_M (Xf)\rho\, d\mu \\ &= \int_M (\rho X) f\, d\mu \\ &= -\int_M f\, \text{div}_\mu(\rho X)\, d\mu \\ &= -\int_M f\left\{\frac{1}{\rho}\text{div}_\mu(\rho X)\right\} d\nu.\end{aligned}$$

Finally, we prove (8):

$$\begin{aligned}\text{div}_{\rho\mu}(X) &= \frac{1}{\rho}\text{div}_\mu(\rho X) && \text{(by(7))} \\ &= \frac{1}{\rho}\left(\langle \text{grad}\, \rho, X\rangle + \rho\, \text{div}_\mu(X)\right) && \text{(by(6))} \\ &= \left\langle \frac{1}{\rho}\text{grad}\, \rho, X\right\rangle + \text{div}_\mu(X) \\ &= \langle \text{grad}(\ln \rho), X\rangle + \text{div}_\mu(X) \\ &= \frac{\partial \ln \rho}{\partial X} + \text{div}_\mu(X).\end{aligned}$$

□

The following proposition provides an explicit formula for the divergence of X in μ, by transforming the corresponding standard formula for the divergence of X in μ_{g}.

Proposition 2 *Let μ be a probability measure in $P_+^\infty(M)$ with density ρ with respect to the measure μ_{g}. Furthermore, let $\sqrt{\mathrm{g}}$ denote the determinant of the first fundamental form of* g *with respect to some local coordinates $(x^1, \ldots, x^n)$ with $n := \dim M$. Then for a vector field $X \in \mathcal{T}(M)$, written in the same local coordinates as $X^i \frac{\partial}{\partial x^i}$, we have*

$$\text{div}_\mu(X) = \frac{1}{\rho\sqrt{\mathrm{g}}}\frac{\partial}{\partial x^i}\left(\rho\sqrt{\mathrm{g}}\, X^i\right) \tag{9}$$

$$= \frac{1}{\sqrt{\mathrm{g}}}\frac{\partial}{\partial x^i}\left(\sqrt{\mathrm{g}}\, X^i\right) + \langle \text{grad}\, \ln \rho, X\rangle. \tag{10}$$

Proof From Eq. (7) we know

$$\text{div}_\mu(X) = \frac{1}{\rho}\text{div}_{\mu_{\mathrm{g}}}(\rho X)\,. \tag{11}$$

The formula for the divergence with respect to μ_g is given as follows (see, for instance, [22]):

$$\mathrm{div}_{\mu_g}(Z) = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i}\left(\sqrt{g}\, Z^i\right), \tag{12}$$

where $Z = Z^i \frac{\partial}{\partial x^i}$. We simply set $Z = \rho X$ in (12) and then use Eq. (11) to show (9). The equality (10) then follows by the Leibniz rule. □

We now come to the *Kolmogorov forward equation* which describes the evolution of $\mu_t = \rho_t \mu$ for all $t \in \mathbb{R}$. For $f \in C^\infty(M)$, we have

$$\begin{aligned}
\int_M f \frac{d}{dt} \ln \rho_t \, d\mu_t &= \int_M f \frac{d}{dt} \rho_t \, d\mu = \frac{d}{dt} \int f \rho_t \, d\mu = \frac{d}{dt} \int f \circ \varphi_t \, d\mu \\
&= \int \frac{d}{dt} f \circ \varphi_t \, d\mu = \int X f(\varphi_t) \, d\mu = \int X(f) \rho_t \, d\mu \\
&= \int X f \, d\mu_t = - \int f \, \mathrm{div}_{\mu_t}(X) \, d\mu_t .
\end{aligned}$$

As this holds for all $f \in C^\infty(M)$, we obtain

$$\frac{d}{dt} \ln \rho_t = -\mathrm{div}_{\mu_t}(X)$$

or, equivalently,

$$\begin{aligned}
\frac{d}{dt} \rho_t &= -\mathrm{div}_{\mu_t}(X)(x)\, \rho_t \\
&= -\mathrm{div}_{\mu}(\rho_t X). \qquad \text{(by(7))}
\end{aligned} \tag{13}$$

This is the Kolmogorov forward equation for a deterministic drift (without diffusion) given by the vector field X. Below we will interpret this equation as the geodesic equation of a particular affine connection, the dual of the mixture connection with respect to the Otto metric which we introduce in Sect. 4.

We conclude this section with some facts about the decomposition of a vector field in terms of a gradient field and a divergence-free field. For a probability measure $\mu \in \mathcal{P}_+^\infty(M)$, consider the $L^2(\mu)$-product of smooth vector fields $X, Y \in \mathcal{T}(M)$, given by

$$\langle\langle X, Y \rangle\rangle_\mu := \int_M \langle X, Y \rangle \, d\mu . \tag{14}$$

The vector space $\mathcal{T}(M)$ equipped with this inner product is a pre-Hilbert space (it is not complete). Now consider the linear map

$$\mathrm{div}_\mu : \mathcal{T}(M) \to C^\infty(M), \quad X \mapsto \mathrm{div}_\mu(X).$$

Its image is given by

$$C_0^\infty(M,\mu) \;:=\; \left\{ f \in C^\infty(M) \;:\; \int_M f\, d\mu = 0 \right\}$$

and its kernel

$$\mathcal{K}_\mu(M) \;:=\; \ker\left(\mathrm{div}_\mu\right) \;=\; \left\{ X \in \mathcal{T}(M) \;:\; \mathrm{div}_\mu(X) = 0 \right\}$$

consists of the *divergence-free* vector fields with respect to μ. For such a vector field X and a gradient field $Y = \mathrm{grad}\, f$ with some $f \in C^\infty(M)$, we have

$$\langle\!\langle X, Y \rangle\!\rangle_\mu \;=\; \int_M \langle X, \mathrm{grad}\, f\rangle\, d\mu \;=\; \int_M Xf\, d\mu \;=\; -\int_M f\, \mathrm{div}_\mu(X)\, d\mu \;=\; 0.$$

Thus, the vector space of gradient fields,

$$\mathcal{G}(M) \;:=\; \left\{ X \in \mathcal{T}(M) \;:\; \text{There exists} f \in C^\infty(M) \text{with} X = \mathrm{grad}\, f \right\},$$

is orthogonal to the kernel of div_μ. It follows from the general theory that we even have

$$\mathcal{T}(M) \;=\; \mathcal{K}_\mu(M) \oplus \mathcal{G}(M),$$

so that every vector field $Y \in \mathcal{T}(M)$ has the orthogonal decomposition

$$Y = X + \mathrm{grad}\, f,$$

where $\mathrm{div}_\mu(X) = 0$ (see, for instance, [15]). Here, X and $\mathrm{grad}\, f$ are clearly unique whereas f is unique up to a constant. This decomposition is referred to as the *Helmholtz-Hodge decomposition*. The restriction of div_μ to the gradient fields $\mathcal{G}(M)$,

$$\mathcal{G}(M) \;\to\; C^\infty(M), \quad X \;\mapsto\; \mathrm{div}_\mu(X),$$

has the same image as div_μ. On the other hand, the space $\mathcal{G}(M)$ is "parametrised" by functions in $C^\infty(M)$. Altogether, we obtain the *Laplace operator*

$$\Delta_\mu \,:\; C^\infty(M) \;\to\; C^\infty(M), \quad f \;\mapsto\; \Delta_\mu(f) \;:=\; \mathrm{div}_\mu(\mathrm{grad}\, f).$$

The image of this operator is given by $C_0^\infty(M,\mu)$ and the kernel simply consists of all constant functions on M. Thus, we can construct the following isomorphism for which we use, by abuse of nation, the same symbol:

$$\Delta_\mu \,:\; C^\infty(M)/\mathbb{R} \;\to\; C_0^\infty(M,\mu), \quad f + \mathbb{R} \;\mapsto\; \Delta_\mu(f + \mathbb{R}) \;:=\; \Delta_\mu(f). \qquad (15)$$

This operator plays an important role when it comes to the identification of the tangent space of $P_+^\infty(M)$ with its cotangent space based on the Otto metric. This is the subject of the next section.

4 The Otto metric

In this section, we are going to introduce the Otto metric. To this end, we first recapitulate how the Fisher-Rao metric can be derived from the following inner product on $C^\infty(M)/\mathbb{R}$:

$$\langle\!\langle f + \mathbb{R}, g + \mathbb{R} \rangle\!\rangle_\mu^{\mathrm{FR}} \;:=\; \mathrm{Cov}_\mu(f, g) \;=\; \int_M \big(f - \mathbb{E}_\mu(f)\big)\, \big(g - \mathbb{E}_\mu(g)\big)\, d\mu. \tag{16}$$

Note that this inner product is obtained from the $L^2(\mu)$-product

$$\int_M f g \, d\mu \tag{17}$$

of functions f, g by adjustment to classes $f + \mathbb{R}$, $g + \mathbb{R}$ of functions. The product (16) is essentially the Fisher-Rao product (this is why we use the abbreviation "FR"), but defined on the cotangent space $C^\infty(M)/\mathbb{R}$ rather than the tangent space of $P_+^\infty(M)$. We are now going to translate it to the tangent space in a straight forward manner. First of all, it allows us to assign to each function class $f + \mathbb{R} \in C^\infty(M)/\mathbb{R}$ the continuous linear form $\langle\!\langle f + \mathbb{R}, \cdot \rangle\!\rangle_\mu^{\mathrm{FR}}$ on $C^\infty(M)/\mathbb{R}$. With the dual pairing (2), we can identify this linear form with the measure $(f - \mathbb{E}_\mu(f))\mu \in S_0^\infty(M)$ which has density $(f - \mathbb{E}_\mu(f))$ with respect to the base measure μ. Altogether, we obtain the isomorphism

$$\phi_\mu :\; C^\infty(M)/\mathbb{R} \;\to\; S_0^\infty(M), \quad f + \mathbb{R} \;\mapsto\; \big(f - \mathbb{E}_\mu(f)\big)\, \mu, \tag{18}$$

with inverse given by the Radon-Nikodym derivative, that is

$$\phi_\mu^{-1} :\; S_0^\infty(M) \;\to\; C^\infty(M)/\mathbb{R}, \quad a \;\mapsto\; \frac{da}{d\mu} + \mathbb{R}.$$

We finally use the map ϕ_μ to translate the inner product (16) to $S_0^\infty(M)$:

$$\langle\!\langle a, b \rangle\!\rangle_\mu^{\mathrm{FR}} \;:=\; \left\langle\!\!\left\langle \phi_\mu^{-1}(a), \phi_\mu^{-1}(b) \right\rangle\!\!\right\rangle_\mu^{\mathrm{FR}} \;=\; \int_M \frac{da}{d\mu} \cdot \frac{db}{d\mu}\, d\mu. \tag{19}$$

This is now the Fisher-Rao inner product defined on $S_0^\infty(M)$, the tangent space of $P_+^\infty(M)$ in μ.

We are now going to iterate this derivation based on the $L^2(\mu)$-product (14) of vector fields instead of the $L^2(\mu)$-product (17) of functions. We restrict attention to gradient fields and obtain the following inner product on $C^\infty(M)/\mathbb{R}$ as replacement of (16):

$$\langle\langle f+\mathbb{R}, g+\mathbb{R}\rangle\rangle_\mu^{\mathrm{O}} \;:=\; \langle\langle \operatorname{grad} f, \operatorname{grad} g\rangle\rangle_\mu \;=\; \int_M \langle \operatorname{grad} f, \operatorname{grad} g\rangle \, d\mu. \tag{20}$$

This is basically the Otto metric (this is why we use the abbreviation "O"), but again defined on the cotangent space rather than the tangent space of $P_+^\infty(M)$. As in the case of the Fisher-Rao metric, we are now going to translate (20) to the tangent space by following the steps of the above derivation. First of all, we assign to each equivalence class $f+\mathbb{R}$ the continuous linear form $\langle\langle f+\mathbb{R}, \cdot\rangle\rangle_\mu$ on $C^\infty(M)/\mathbb{R}$. With the dual pairing (2), we can interpret this linear form as the measure $-\Delta_\mu(f)\mu = -\operatorname{div}_\mu(\operatorname{grad} f)\mu \in S_0^\infty(M)$. To see this, we analyse how this measure acts on a class $g+\mathbb{R} \in C^\infty/\mathbb{R}$:

$$\begin{aligned}
-\int_M g\,\Delta_\mu(f)\,d\mu &= -\int_M g\operatorname{div}_\mu(\operatorname{grad} f)\,d\mu \\
&= \int_M dg(\operatorname{grad} f)\,d\mu \\
&= \langle\langle \operatorname{grad} g, \operatorname{grad} f\rangle\rangle_\mu \\
&= \langle\langle f+\mathbb{R}, g+\mathbb{R}\rangle\rangle_\mu^{\mathrm{O}}.
\end{aligned}$$

Thus, the measure $-\Delta_\mu(f)\mu$ acts on $g+\mathbb{R}$ in the same way as $\langle\langle f+\mathbb{R}, \cdot\rangle\rangle_\mu$. It plays the role of the measure $(f-\mathbb{E}_\mu(f))\mu$ in the context of the Fisher-Rao metric. We finally obtain the following identification of the cotangent space of $P_+^\infty(M)$ with its tangent space:

$$\phi_\mu : \; C^\infty(M)/\mathbb{R} \;\to\; S_0^\infty(M), \quad f+\mathbb{R} \;\mapsto\; -\Delta_\mu(f)\mu, \tag{21}$$

with inverse given by the Laplace operator (15), that is

$$\phi_\mu^{-1} : \; S_0^\infty(M) \;\to\; C^\infty(M)/\mathbb{R}, \quad a \;\mapsto\; -\Delta_\mu^{-1}\left(\frac{da}{d\mu}\right).$$

This identification is analogous to but fundamentally different from the identification (18). Even though we are going to use the same symbol, ϕ_μ, for both identifications, it will be clear from the context which one we mean.

With the isomorphism (21), we can now interpret the inner product (20) as an inner product on the tangent space of $P_+^\infty(M)$, leading to the *Otto metric*. For $a, b \in S_0^\infty(M)$, we simply set

$$\langle\langle a, b\rangle\rangle_\mu^{\mathrm{O}} \;:=\; \left\langle\!\!\left\langle \phi_\mu^{-1}(a), \phi_\mu^{-1}(b) \right\rangle\!\!\right\rangle_\mu^{\mathrm{O}}. \tag{22}$$

For $a = \Delta_\mu(f)$ and $b = \Delta_\mu(g)$, we obtain

$$\langle\langle a, b \rangle\rangle_\mu^{\mathrm{O}} = \int_M \langle \operatorname{grad} f, \operatorname{grad} g \rangle \, d\mu. \tag{23}$$

The Otto metric (22) defines a weak Riemannian structure on $\mathcal{P}_+^\infty(M)$ where the term "weak" refers to the fact that $\mathcal{S}_0^\infty(M)$ is not complete with respect to $\langle\langle \cdot, \cdot \rangle\rangle_\mu^{\mathrm{O}}$.

Remark 1 The identification (21) is derived in [15] (Theorem 3.5). It also plays a fundamental role in Otto's seminal work [13] in which he considers the sample space $\mathbb{R}^N$. This corresponds to the Riemannian manifold (M, g) of the present article. On the one hand, $\mathbb{R}^N$, equipped with the standard inner product, is a very special case of a Riemannian manifold. On the other hand, its non-compactness is a desirable property excluded from this article. In a somewhat different notation, Otto compares the identification of a class $f + \mathbb{R}$ with $-\nabla^2 f$, which he refers to as the "traditional approach", with the identification of that function class with $-\nabla \cdot (\rho \nabla f)$, his "new approach" (Section 1.3 of [13]). Here, ρ denotes the density of a probability measure with respect to the Lebesgue measure λ on $\mathbb{R}^N$, the analogue of our measure μ_{g}. Reformulating these expressions in terms of our notation, we have the following correspondences:

$$\begin{aligned} -\nabla^2 f \quad &\leftrightarrow \quad -\operatorname{div}_{\mu_{\mathrm{g}}}(\operatorname{grad} f) = -\Delta_{\mu_{\mathrm{g}}}(f), \\ -\nabla \cdot (\rho \nabla f) \quad &\leftrightarrow \quad -\operatorname{div}_{\mu_{\mathrm{g}}}(\rho \operatorname{grad} f) = -\operatorname{div}_{\rho\mu_{\mathrm{g}}}(\operatorname{grad} f)\, \rho = -\Delta_{\rho\mu_{\mathrm{g}}}(f)\rho. \end{aligned} \tag{24}$$

Thus, the identification (21) coincides with Otto's identification of $f + \mathbb{R}$ with (24), expressed in terms of the densities with respect to the fixed reference measures λ and μ_{g}.

Given a smooth function $F : \mathcal{P}_+^\infty(M) \to \mathbb{R}$, we can evaluate the gradient of F with respect to the Fisher-Rao metric as well as the Otto metric. We denote these gradients by $\operatorname{grad}^{\mathrm{FR}} F$ and $\operatorname{grad}^{\mathrm{O}} F$, respectively. The following theorem deals with the gradients of particularly simple functions that are defined in terms of corresponding functions on the sample space M. More precisely, for $f \in C^\infty(M)$, we consider the mean value function

$$\mathbb{E}(f) : \mathcal{P}_+^\infty(M) \to \mathbb{R}, \quad \mu \mapsto \mathbb{E}_\mu(f) = \int_M f \, d\mu.$$

Theorem 1 *For all $f \in C^\infty(M)$, we have*

$$\operatorname{grad}_\mu^{\mathrm{FR}} \mathbb{E}(f) = (f - \mathbb{E}_\mu(f))\, \mu, \tag{25}$$

$$\operatorname{grad}_\mu^{\mathrm{O}} \mathbb{E}(f) = -\Delta_\mu(f)\, \mu. \tag{26}$$

Furthermore, the Fisher-Rao metric and the Otto metric are uniquely characterised by their respective properties (25) and (26).

Proof We first consider the differential of $\mathbb{E}(f)$. For $a \in S_0^\infty(M)$, we have

$$\begin{aligned} d_\mu \mathbb{E}(f)(a) &= \mathbb{E}(f)(\mu + a) - \mathbb{E}(f)(\mu) \\ &= \mathbb{E}_{\mu+a}(f) - \mathbb{E}_\mu(f) \\ &= \int_M f\, d\mu + \int_M f\, da - \int_M f\, d\mu \\ &= \int_M f\, da. \end{aligned} \tag{27}$$

In order to prove (25) we have to show

$$d_\mu \mathbb{E}(f)(a) \;=\; \langle\!\langle (f - \mathbb{E}_\mu(f))\mu, a \rangle\!\rangle_\mu^{\mathrm{FR}}, \qquad \text{for all } a \in S_0^\infty(M).$$

This is obvious from the following simple calculation:

$$\begin{aligned} d_\mu \mathbb{E}(f)(a) &= \int_M f\, da && (\text{by}(27)) \\ &= \int_M (f - \mathbb{E}_\mu(f))\, da \\ &= \int_M (f - \mathbb{E}_\mu(f)) \cdot \frac{da}{d\mu}\, d\mu \\ &= \langle\!\langle (f - \mathbb{E}_\mu(f))\mu, a \rangle\!\rangle_\mu^{\mathrm{FR}}. && (\text{by}(19)) \end{aligned}$$

For proving (26), we have to show that

$$d_\mu \mathbb{E}(f)(a) \;=\; -\langle\!\langle \Delta_\mu(f)\, \mu, a \rangle\!\rangle_\mu^{\mathrm{O}}, \qquad \text{for all } a \in S_0^\infty(M).$$

With $a = \Delta_\mu(g)\mu$ for some $g \in C^\infty(M)$, we have

$$\begin{aligned} d_\mu \mathbb{E}(f)(a) &= d_\mu \mathbb{E}(f)(\Delta_\mu(g)\mu) \\ &= \int_M f\, \Delta_\mu(g)\, d\mu && (\text{by}(27)) \\ &= \int_M f\, \mathrm{div}_\mu(\mathrm{grad}\, g)\, d\mu \\ &= -\int_M \langle \mathrm{grad}\, f, \mathrm{grad}\, g \rangle\, d\mu && (\text{by}(5)) \\ &= -\langle\!\langle \Delta_\mu(f)\mu, \Delta_\mu(g)\mu \rangle\!\rangle_\mu^{\mathrm{O}} && (\text{by}(23)) \\ &= -\langle\!\langle \Delta_\mu(f)\mu, a \rangle\!\rangle_\mu^{\mathrm{O}}. \end{aligned}$$

The uniqueness of the Fisher-Rao metric as well as the Otto metric follows from Proposition 3 below. □

Proposition 3 *Let* $\langle\!\langle \cdot, \cdot \rangle\!\rangle^{(1)}$ *and* $\langle\!\langle \cdot, \cdot \rangle\!\rangle^{(2)}$ *be two Riemannian metrics on* $P_+^\infty(M)$. *Then:*

$$\mathrm{grad}^{(1)}\, \mathbb{E}(f) \;=\; \mathrm{grad}^{(2)}\, \mathbb{E}(f), \quad \textit{for all} f \in C^\infty(M) \quad \Rightarrow \quad \langle\!\langle \cdot, \cdot \rangle\!\rangle^{(1)} = \langle\!\langle \cdot, \cdot \rangle\!\rangle^{(2)}.$$

Proof Consider a point $\mu \in P_+^\infty(M)$ and two vectors $a, b \in S_0^\infty(M)$. We can represent a as a gradient of $\mathbb{E}(f)$ with a function $f \in C^\infty(M)$, say with respect to the metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle^{(1)}$, that is

$$a \;=\; \mathrm{grad}_\mu^{(1)}\, \mathbb{E}(f).$$

This implies

$$\begin{aligned}
\langle\!\langle a, b \rangle\!\rangle_\mu^{(1)} &= \left\langle\!\!\left\langle \mathrm{grad}_\mu^{(1)}\, \mathbb{E}(f), b \right\rangle\!\!\right\rangle_\mu^{(1)} \\
&= d_\mu \mathbb{E}(f)(b) \\
&= \left\langle\!\!\left\langle \mathrm{grad}_\mu^{(2)}\, \mathbb{E}(f), b \right\rangle\!\!\right\rangle_\mu^{(2)} \\
&= \left\langle\!\!\left\langle \mathrm{grad}_\mu^{(1)}\, \mathbb{E}(f), b \right\rangle\!\!\right\rangle_\mu^{(2)} \\
&= \langle\!\langle a, b \rangle\!\rangle_\mu^{(2)}.
\end{aligned}$$

□

5 Dual structure with the Otto metric

In addition to a Riemannian metric, such as the Fisher-Rao metric or the Otto metric, information geometry is based on a pair of affine connections that are dual with respect to that metric. Of particular importance is the case where the two affine connections are flat. In that case we talk about a dually flat structure. The classical dually flat structure is given by the Fisher-Rao metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle^{\mathrm{FR}}$, the mixture connection $\nabla^{(\mathrm{m})}$ and its dual, the exponential connection $\nabla^{(\mathrm{e})}$. Here, the duality is expressed by

$$A\, \langle\!\langle B, C \rangle\!\rangle^{\mathrm{FR}} \;=\; \left\langle\!\!\left\langle \nabla_A^{(\mathrm{m})} B, C \right\rangle\!\!\right\rangle^{\mathrm{FR}} + \left\langle\!\!\left\langle A, \nabla_B^{(\mathrm{e})} C \right\rangle\!\!\right\rangle^{\mathrm{FR}}, \tag{28}$$

where A, B, C are smooth vector fields on $P_+^\infty(M)$. The duality condition (28) can be stated in terms of the parallel transports $\Pi_\gamma^{(\mathrm{m})}$ and $\Pi_\gamma^{(\mathrm{e})}$ of vectors along a curve γ in $P_+^\infty(M)$ with initial point μ and endpoint ν. In our setting, these transports only depend on μ and ν and not on the way these two points are connected. Therefore, we can write $\Pi_{\mu,\nu}^{(\mathrm{m})}$ for $\Pi_\gamma^{(\mathrm{m})}$ and $\Pi_{\mu,\nu}^{(\mathrm{e})}$ for $\Pi_\gamma^{(\mathrm{e})}$. In this notation, (28) translates to

$$\left\langle\!\!\left\langle \Pi_{\mu,\nu}^{(\mathrm{m})} a, \Pi_{\mu,\nu}^{(\mathrm{e})} b \right\rangle\!\!\right\rangle_\nu^{\mathrm{FR}} \;=\; \langle\!\langle a, b \rangle\!\rangle_\mu^{\mathrm{FR}}, \tag{29}$$

where a, b are vectors in $T_\mu P_+^\infty(M)$. The mixture connection, which we refer to as m-connection, uniquely determines the exponential connection as its dual with respect to the Fisher-Rao metric. In this article, we will introduce the dual of the mixture connection with respect to the Otto metric. This will also be termed an exponential connection. In order to distinguish it from the classical exponential connection, we use the symbol "e_1" for it and replace the standard symbol "e" for the classical version by "e_0". Thus, we will refer to two distinct exponential connections, the e_0-connection and the e_1-connection. Before we come to the e_1-connection, we first recapitulate how the classical e_0-connection can be derived, thereby following the strategy presented in [19].

We begin with the mixture connection. Its parallel transport from μ to ν is simply given by the following trivial identification:

$$\Pi^{(\mathrm{m})}_{\mu,\nu} : T_\mu P_+^\infty(M) \to T_\nu P_+^\infty(M), \qquad (\mu, a) \mapsto (\nu, a). \tag{30}$$

In order to derive the geodesic curve $\gamma : t \mapsto \gamma(t)$ with respect to the m-connection, we motivate its equation using terminology from physics. Starting with an initial "position" and an initial "velocity", say $\gamma(0) = \mu$ and $\dot\gamma(0) = a$, it is uniquely characterised by the requirement that the "acceleration" of γ vanishes for all t. This condition can be expressed in terms of the parallel transport as follows:

$$\dot\gamma(t) = \Pi^{(\mathrm{m})}_{\mu,\gamma(t)} a, \tag{31}$$

which is equivalent to $\dot\gamma(t) = a$. Thus, we obtain the m-geodesic $t \mapsto \gamma(t) = \mu + ta$, where we require $\gamma(t) \in P_+^\infty(M)$ for all t. This is satisfied for an open time interval that contains $t = 0$.

Now we come to the e_0-connection. In essence, the parallel transport of the e_0-connection is as simple as (30) when expressed as a map between the respective cotangent spaces:

$$\Pi^{(\mathrm{e}_0)}_{\mu,\nu} : T^*_\mu P_+^\infty(M) \to T^*_\nu P_+^\infty(M), \qquad (\mu, f + \mathbb{R}) \mapsto (\nu, f + \mathbb{R}). \tag{32}$$

In order to define the parallel transport of the e_0-connection as a map between the corresponding tangent spaces, we use the identification (18) of the tangent space of $P_+^\infty(M)$, that is $S_0^\infty(M)$, with its cotangent space $C^\infty(M)/\mathbb{R}$ based on the Fisher-Rao metric in μ and ν, respectively. We first go from the tangent space to the dual space with ϕ_μ^{-1}, then transport there trivially according to (32), and finally go back to the tangent space with ϕ_ν. More precisely, we consider

$$\Pi^{(\mathrm{e}_0)}_{\mu,\nu} : T_\mu P_+^\infty(M) \to T_\nu P_+^\infty(M), \quad (\mu, a) \mapsto (\nu, (\phi_\nu \circ \phi_\mu^{-1})(a)), \tag{33}$$

where

$$(\phi_\nu \circ \phi_\mu^{-1})(a) = \left(\frac{da}{d\mu} - \mathbb{E}_\nu\left(\frac{da}{d\mu}\right)\right)\nu. \tag{34}$$

In order to derive the e_0-geodesic, we follow the above reasoning and simply replace in the condition (31) "m" by "e_0". Using (33) and (34), this leads to the following equation for the e_0-geodesic:

$$\dot{\gamma}(t) \;=\; \Pi^{(e_0)}_{\mu,\gamma(t)} a \;=\; \left(\frac{da}{d\mu} - \mathbb{E}_{\gamma(t)}\left(\frac{da}{d\mu}\right)\right)\gamma(t), \tag{35}$$

with $\gamma(0) = \mu$ and $\dot{\gamma}(0) = a$. It can be easily verified that the following curve solves this equation:

$$\gamma : \mathbb{R} \;\to\; P_+^\infty(M), \quad t \;\mapsto\; \exp\left(t\frac{da}{d\mu} - \ln\left(\int_M \exp\left(t\frac{da}{d\mu}\right)d\mu\right)\right)\mu. \tag{36}$$

This is the classical e_0-geodesic, which is well-known in information geometry.

After having introduced the m-connection and the classical e_0-connection we now verify their duality relation (29) with respect to the Fisher-Rao metric:

$$\begin{aligned}
\left\langle\!\!\left\langle \Pi^{(m)}_{\mu,\nu} a, \Pi^{(e_0)}_{\mu,\nu} b \right\rangle\!\!\right\rangle^{\mathrm{FR}}_\nu &= \left\langle\!\!\left\langle a, \left(\frac{db}{d\mu} - \mathbb{E}_\nu\left(\frac{db}{d\mu}\right)\right)\nu \right\rangle\!\!\right\rangle^{\mathrm{FR}}_\nu \\
&= \int_M \frac{da}{d\nu}\left(\frac{db}{d\mu} - \mathbb{E}_\nu\left(\frac{db}{d\mu}\right)\right) d\nu \\
&= \int_M \frac{da}{d\nu}\frac{db}{d\mu}\, d\nu - \mathbb{E}_\nu\left(\frac{db}{d\mu}\right) \underbrace{\int_M \frac{da}{d\nu}\, d\nu}_{=0} \\
&= \int_M \frac{da}{d\mu}\frac{db}{d\mu}\, d\mu \\
&= \langle\!\langle a, b\rangle\!\rangle^{\mathrm{FR}}_\mu .
\end{aligned}$$

A core contribution of this article is the introduction of a new exponential connection, the dual of the m-connection with respect to the Otto metric which we call the e_1-connection. The derivation of this connection follows the same scheme that we applied for the e_0-connection. We start again with the trivial map (32) as the definition of the parallel transport of the e_1-connection expressed as a map between cotangent spaces. At this point, nothing is new, and we can simply replace in (32) the symbol e_0 by e_1 without changing the definition of this map. A difference emerges when we formulate this parallel transport as a map between tangent spaces. To do that, we now use the isomorphism based on the Otto metric, that is (21). With that we define the parallel transport

$$\Pi^{(e_1)}_{\mu,\nu} : \; T_\mu P_+^\infty(M) \;\to\; T_\nu P_+^\infty(M), \quad (\mu, a) \;\mapsto\; (\nu, (\phi_\nu \circ \phi_\mu^{-1})(a)),$$

where

$$(\phi_\nu \circ \phi_\mu^{-1})(a) \;=\; \left(\Delta_\nu \circ \Delta_\mu^{-1}\right)\left(\frac{da}{d\mu}\right)\nu. \tag{37}$$

Now let us consider the analogue of the geodesic equation (35) for the e_1-connection:

$$\dot{\gamma}(t) \;=\; \Pi^{(\mathrm{e}_1)}_{\mu,\gamma(t)} a \;=\; \left(\Delta_{\gamma(t)} \circ \Delta_\mu^{-1}\right)\left(\frac{da}{d\mu}\right)\gamma(t), \tag{38}$$

with $\gamma(0) = \mu$ and $\dot{\gamma}(0) = a$. The following theorem specifies the solution of this equation.

Theorem 2 *Let $\mu \in P_+^\infty(M)$, $a \in S_0^\infty(M)$. To express the e_1-geodesic, consider $f \in C^\infty(M)$ that satisfies $a = -\Delta_\mu(f)\mu$ and the flow φ_t, $t \in \mathbb{R}$, generated by the gradient field* grad f. *Then the curve*

$$\gamma : \mathbb{R} \to P_+^\infty(M), \qquad t \mapsto \gamma(t) := (\varphi_t)_*(\mu), \tag{39}$$

is the e_1-geodesic satisfying $\gamma(0) = \mu$ and $\dot{\gamma}(0) = a$.

Proof The e_1-geodesic satisfies

$$\begin{aligned}\dot{\gamma}(t) &= (\Delta_{\gamma(t)} \circ \Delta_\mu^{-1})\left(-\Delta_\mu(f)\right)\gamma(t) \qquad (\text{by}(38))\\ &= -\Delta_{\gamma(t)}(f)\,\gamma(t).\end{aligned}$$

With $\gamma(t) = \rho_t \mu$ and (7), this equation translates to

$$\frac{d}{dt}\,\rho_t = -\mathrm{div}_\mu\,(\,\rho_t \,\mathrm{grad}\, f)$$

and therefore coincides with the Kolmogorov forward equation (13) which determines γ uniquely. □

As in the classical setting where the exponential connection is the dual of the mixture connection with respect to the Fisher-Rao metric, we can also prove duality in the context of the Otto metric.

Theorem 3 *The* e_1*-connection is the dual of the* m*-connection with respect to the Otto metric, that is*

$$\left\langle\!\!\left\langle \Pi^{(\mathrm{m})}_{\mu,\nu} a,\, \Pi^{(\mathrm{e}_1)}_{\mu,\nu} b \right\rangle\!\!\right\rangle^{\mathrm{O}}_\nu \;=\; \langle\!\langle a, b\rangle\!\rangle^{\mathrm{O}}_\mu, \tag{40}$$

for all vectors $a, b \in T_\mu P_+^\infty(M)$.

Proof In order to prove the duality relation (40) consider two vectors a and b in $T_\mu P_+^\infty(M)$ with the representations $a = \Delta_\mu(f)\,\mu$ and $b = \Delta_\mu(g)\,\mu$. We now transport these vectors to $T_\nu P_+^\infty(M)$ with the m- and the e_1-connection, respectively, and evaluate the inner product:

$$\left\langle\!\!\left\langle \Pi^{(\mathrm{m})}_{\mu,\nu} a,\, \Pi^{(\mathrm{e}_1)}_{\mu,\nu} b \right\rangle\!\!\right\rangle^{\mathrm{O}}_\nu = \left\langle\!\!\left\langle \Pi^{(\mathrm{m})}_{\mu,\nu}(\Delta_\mu(f)\,\mu),\, \Pi^{(\mathrm{e}_1)}_{\mu,\nu}(\Delta_\mu(g)\,\mu) \right\rangle\!\!\right\rangle^{\mathrm{O}}_\nu$$

$$
\begin{aligned}
&= \left\langle\!\left\langle \Delta_\mu(f)\,\mu,\, \Delta_\nu(g)\,\nu \right\rangle\!\right\rangle_\nu^{\mathrm{O}} && \text{(by(37))}\\
&= \left\langle\!\left\langle \mathrm{div}_\mu(\mathrm{grad}\, f)\,\mu,\, \Delta_\nu(g)\,\nu \right\rangle\!\right\rangle_\nu^{\mathrm{O}}\\
&= \left\langle\!\!\left\langle \frac{1}{\frac{d\mu}{d\nu}} \mathrm{div}_\nu \left(\frac{d\mu}{d\nu}\, \mathrm{grad}\, f \right) \mu,\, \Delta_\nu(g)\,\nu \right\rangle\!\!\right\rangle_\nu^{\mathrm{O}} && \text{(by(7))}\\
&= \left\langle\!\!\left\langle \mathrm{div}_\nu \left(\frac{d\mu}{d\nu}\, \mathrm{grad}\, f \right) \nu,\, \mathrm{div}_\nu\,(\mathrm{grad}\, g)\,\nu \right\rangle\!\!\right\rangle_\nu^{\mathrm{O}}
\end{aligned}
$$

With the decomposition

$$
\frac{d\mu}{d\nu}\, \mathrm{grad}\, f \;=\; X + \mathrm{grad}\, \tilde{f}, \qquad \text{with } \mathrm{div}_\nu(X) = 0,
$$

this translates to

$$
\begin{aligned}
&\left\langle\!\left\langle \mathrm{div}_\nu \left(\mathrm{grad}\, \tilde{f}\right) \nu,\, \mathrm{div}_\nu\,(\mathrm{grad}\, g)\,\nu \right\rangle\!\right\rangle_\nu^{\mathrm{O}}\\
&\quad = \int_M \left\langle \mathrm{grad}\, \tilde{f},\, \mathrm{grad}\, g \right\rangle d\nu && \text{(by(23))}\\
&\quad = \int_M \left\langle \frac{d\mu}{d\nu}\, \mathrm{grad}\, f - X,\, \mathrm{grad}\, g \right\rangle d\nu\\
&\quad = \int_M \frac{d\mu}{d\nu}\, \langle \mathrm{grad}\, f,\, \mathrm{grad}\, g \rangle\, d\nu - \underbrace{\int_M \langle X,\, \mathrm{grad}\, g \rangle\, d\nu}_{=0}\\
&\quad = \int_M \langle \mathrm{grad}\, f,\, \mathrm{grad}\, g \rangle\; d\mu\\
&\quad = \langle\!\langle a, b \rangle\!\rangle_\mu^{\mathrm{O}}. && \text{(by(23))}
\end{aligned}
$$

□

We can make the following observations which follow from the general theory [6, 19]. As the parallel transport along a curve with respect to the e_1-connection does not depend on the curve itself but only on the initial and endpoint, the curvature tensor based on the infinitesimal formulation of the e_1-connection, $\nabla^{(\mathrm{e}_1)}$, vanishes. This follows also from the fact that the e_1-connection is dual to the m-connection which has vanishing curvature and torsion tensors. This duality, however, is not sufficient for proving torsion-freeness of the e_1-connection. We conjecture that also the torsion tensor vanishes so that we have a dually flat structure at hand.

We are now going to relate the m- and the e_1-connection to the Levi-Civita connection of the Otto metric. In order to do that, we first recapitulate the classical setting where the dual structure is defined with respect to the Fisher-Rao metric. There, one interpolates between $\nabla^{(\mathrm{m})}$ and $\nabla^{(\mathrm{e}_0)}$ by a parameter $\alpha \in [-1, +1]$, leading to the family of α-connections,

$$\nabla^{(e_0,\alpha)} := \frac{1-\alpha}{2}\,\nabla^{(m)} + \frac{1+\alpha}{2}\,\nabla^{(e_0)}. \tag{41}$$

Here, the Levi-Civita connection of the Fisher-Rao metric is given for $\alpha = 0$. We recover the m-connection with $\alpha = -1$ and the e_0-connection with $\alpha = +1$. Furthermore, $\nabla^{(e_0,-\alpha)}$ and $\nabla^{(e_0,\alpha)}$ are dual with respect to the Fisher-Rao metric, that is

$$A\,\langle\langle B, C\rangle\rangle^{\mathrm{FR}} = \left\langle\!\!\left\langle \nabla_A^{(e_0,-\alpha)} B, C \right\rangle\!\!\right\rangle^{\mathrm{FR}} + \left\langle\!\!\left\langle A, \nabla_B^{(e_0,\alpha)} C \right\rangle\!\!\right\rangle^{\mathrm{FR}}, \tag{42}$$

where A, B, C are smooth vector fields on $P_+^\infty(M)$. Obviously, the equality (42) generalises (28). A similar situation holds for the new dual structure that is proposed in this article. We can define a new family of α-connections in the same way by just replacing the e_0-connection in (41) by the e_1-connection. For $\alpha \in [-1, +1]$, we then have

$$\nabla^{(e_1,\alpha)} := \frac{1-\alpha}{2}\,\nabla^{(m)} + \frac{1+\alpha}{2}\,\nabla^{(e_1)}. \tag{43}$$

In this family, the Levi-Civita connection is again given for $\alpha = 0$. However, this time it coincides with the Levi-Civita connection of the Otto metric instead of the Fisher-Rao metric. For $\alpha = -1$, we recover the m-connection and for $\alpha = +1$ we have the newly defined exponential connection, that is the e_1-connection. If the e_1-connection turns out to be torsion-free, we also have duality of $\nabla^{(e_1,-\alpha)}$ and $\nabla^{(e_1,\alpha)}$, this time with respect to the Otto metric:

$$A\,\langle\langle B, C\rangle\rangle^{\mathrm{O}} = \left\langle\!\!\left\langle \nabla_A^{(e_1,-\alpha)} B, C \right\rangle\!\!\right\rangle^{\mathrm{O}} + \left\langle\!\!\left\langle A, \nabla_B^{(e_1,\alpha)} C \right\rangle\!\!\right\rangle^{\mathrm{O}},$$

where A, B, C are smooth vector fields on $P_+^\infty(M)$.

6 Canonical contrast function

With a dually flat structure, we can define a canonical contrast function as proposed, for instance, in [6].[2] A more general canonical contrast function, defined for a Riemannian manifold with an affine connection that is typically different from the Levi-Civita connection and not necessarily flat, has been proposed in [23] and further developed in [24] (see also the related proposal [25]). In this section, we derive the canonical contrast function associated with the Otto metric and the e_1-connection. To outline our approach, we first review the classical setting of the Fisher-Rao metric and the

[2] In information geometry, the term "divergence" or "divergence function" is more commonly used than "contrast function". However, given that we are using the term "divergence" already with a different meaning, we will consistently refer to contrast functions.

corresponding dual of the m-connection, the e_0-connection. Given a pair $(\mu, \nu) \in P_+^\infty(M) \times P_+^\infty(M)$, let $\gamma : [0, 1] \to P_+^\infty(M)$ be the e_0-geodesic with $\gamma(0) = \mu$ and $\gamma(1) = \nu$. With that, we define

$$D^{(e_0)}(\mu\|\nu) \;:=\; \int_0^1 t\, \|\dot\gamma(t)\|^2_{\gamma(t)}\, dt. \tag{44}$$

Here, the norm on the RHS of (44) is meant to be with respect to the Fisher-Rao metric. The geodesic (36) can be restricted to the unit interval and adjusted to satisfy $\gamma(0) = \mu$ and $\gamma(1) = \nu$:

$$\gamma : \; [0, 1] \;\to\; P_+^\infty(M), \quad t \;\mapsto\; \frac{\left(\frac{d\nu}{d\mu}\right)^t}{\int_M \left(\frac{d\nu}{d\mu}\right)^t d\mu}\, \mu. \tag{45}$$

Using this geodesic in (44), we obtain

$$D^{(e_0)}(\mu\|\nu) \;:=\; \int_M \left(\frac{d\nu}{d\mu}\right) \ln\left(\frac{d\nu}{d\mu}\right) d\mu,$$

which is the well-known Kullback–Leibler contrast function. The dual contrast function can be obtained by simply switching the arguments.

We now come to the contrast function with respect to the Otto metric and the e_1-connection. It is evaluated according to (44) where we replace "e_0" by "e_1" and reinterpret the norm on the RHS to be defined with respect to the Otto metric. Consider the e_1-geodesic $\gamma : [0, 1] \to P_+^\infty(M)$ connecting μ with ν, that is

$$\gamma(t) \;=\; (\varphi_t)_*(\mu), \qquad (\varphi_1)_*(\mu) \;=\; \nu,$$

where φ_t is the flow generated by a gradient field grad $f \in \mathcal{T}(M)$ (see (39)). For that curve we have

$$\dot\gamma(0) \;=\; -\Delta_\mu(f)\,\mu, \qquad \dot\gamma(t) \;=\; -\Delta_{\gamma(t)}(f)\,\gamma(t).$$

For the squared norm of $\dot\gamma(t)$ we obtain

$$\begin{aligned}
\|\dot\gamma(t)\|^2_{\gamma(t)} &= \int_M \langle \mathrm{grad}_x\, f, \mathrm{grad}_x\, f\rangle_x\, \gamma(t)(dx)\\
&= \int_M \langle \mathrm{grad}_x\, f, \mathrm{grad}_x\, f\rangle_x\, (\varphi_t)_*(\mu)(dx)\\
&= \int_M \langle \mathrm{grad}_{\varphi_t(x)}\, f, \mathrm{grad}_{\varphi_t(x)}\, f\rangle_{\varphi_t(x)}\, \mu(dx)\\
&\quad \text{(by the general transformation rule for integrals)}\\
&= \int_M \langle \mathrm{grad}_{\varphi_t(x)}\, f, \dot\varphi_t(x)\rangle_{\varphi_t(x)}\, \mu(dx)
\end{aligned}$$

$$
\begin{aligned}
&= \int_M df_{\varphi_t(x)}\,(\dot{\varphi}_t(x))\;\;\mu(dx) \\
&= \int_M \frac{\partial}{\partial t}\, f(\varphi_t(x))\,\mu(dx). \qquad (46)
\end{aligned}
$$

The canonical contrast function (44) can now be evaluated as:

$$
\begin{aligned}
D^{(\mathrm{e}_1)}(\mu\|\nu) &= \int_0^1 t\,\|\dot{\gamma}(t)\|^2_{\gamma(t)}\,dt \\
&= \int_0^1 t \int_M \frac{\partial}{\partial t}\, f(\varphi_t(x))\,\mu(dx)\,dt \qquad (\text{by}(46)) \\
&= \int_M \int_0^1 t\,\frac{\partial}{\partial t}\, f(\varphi_t(x))\,dt\,\mu(dx) \\
&\quad \text{(change of integration order)} \\
&= \int_M \int_0^1 \left(\frac{\partial}{\partial t}\, t\, f(\varphi_t(x)) - f(\varphi_t(x))\right) dt\,\mu(dx) \\
&\quad \text{(integration by parts)} \\
&= \int_M \left(f(\varphi_1(x)) - \int_0^1 f(\varphi_t(x))\,dt \right) \mu(dx) \\
&= \int_M \int_0^1 \Big(f(\varphi_1(x)) - f(\varphi_t(x)) \Big)\,dt\,\mu(dx) \\
&= \int_M \int_0^1 \Big(f \circ \varphi_1 - f \circ \varphi_t \Big)\,dt\,d\mu. \qquad (47)
\end{aligned}
$$

There are various ways to define the dual contrast function. Here, we restrict attention to the dual obtained by switching the arguments. For that we consider the following equality

$$
D^{(\mathrm{e}_1)}(\mu\|\nu) + D^{(\mathrm{e}_1)}(\nu\|\mu) \;=\; \int_0^1 \|\dot{\gamma}(t)\|^2_{\gamma(t)}\,dt.
$$

Thus, the sum of the contrast function and its dual coincides with twice the energy of the e_1-geodesic connecting μ with ν. It can be evaluated as

$$
\begin{aligned}
\int_0^1 \|\dot{\gamma}(t)\|^2_{\gamma(t)}\,dt &= \int_0^1 \int_M \frac{\partial}{\partial t}\, f(\varphi_t(x))\,\mu(dx)\,dt \qquad (\text{by}(46)) \\
&= \int_M \int_0^1 \frac{\partial}{\partial t}\, f(\varphi_t(x))\,dt\,\mu(dx) \\
&\quad \text{(change of integration order)} \\
&= \int_M (f(\varphi_1(x)) - f(\varphi_0(x)))\,\mu(dx) \\
&= \int_M \Big(f \circ \varphi_1 - f \circ \varphi_0 \Big)\,d\mu. \qquad (48)
\end{aligned}
$$

With that, we obtain

$$\begin{aligned} D^{(\mathrm{e}_1)}(\nu\|\mu) &= \int_0^1 \|\dot{\gamma}(t)\|^2_{\gamma(t)}\, dt - D^{(\mathrm{e}_1)}(\mu\|\nu) \\ &= \int_M \Big(f\circ\varphi_1 - f\circ\varphi_0\Big)\, d\mu - \int_M \int_0^1 \Big(f\circ\varphi_1 - f\circ\varphi_t\Big)\, dt\, d\mu \\ &\quad \text{(by (48) and (47))} \\ &= \int_M \int_0^1 \Big(f\circ\varphi_t - f\circ\varphi_0\Big)\, dt\, d\mu. \end{aligned}$$

In case of a dually flat structure, $D^{(\mathrm{e}_1)}(\nu\|\mu)$ will coincide with $D^{(\mathrm{m})}(\mu\|\nu)$ computed according to definition (44) with γ being the m-geodesic.

The formula (44) for the canonical contrast function can also be applied to the α-connection (43). This leads to

$$D^{(\alpha)}(\mu\|\nu) \;:=\; \int_0^1 t\, \|\dot{\gamma}(t)\|^2_{\gamma(t)}\, dt, \tag{49}$$

where on the RHS of (49) the norm is meant to be with respect to the Otto metric and γ denotes the geodesic of the α-connection (43) with $\gamma(0) = \mu$ and $\gamma(1) = \nu$. For $\alpha = +1$, we recover $D^{(\mathrm{e}_1)}$. The case $\alpha = 0$ is particularly interesting. In that case, we have the geodesic γ of the Levi-Civita connection of the Otto metric. The term $\|\dot{\gamma}(t)\|_{\gamma(t)}$ is then independent of t and reduces to the Riemannian distance which coincides with the Wasserstein distance $W_2(\mu,\nu)$. More precisely, denoting by $\Pi(\mu,\nu)$ the set of probability measures π on the product $M\times M$ with first marginal equal to μ and second marginal equal to ν, we have

$$W_2(\mu,\nu) \;=\; \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{M\times M} d^2(x,y)\, \pi(dx,dy)\right)^{\frac{1}{2}}, \tag{50}$$

where d is the Riemannian distance on M induced by g. Thus, we obtain

$$D^{(0)}(\mu\|\nu) \;=\; \frac{1}{2}\, W_2(\mu,\nu)^2. \tag{51}$$

Observe that the contrast function (51) explicitly depends on the sample space distance d, and thereby on the Riemannian metric g, via the Wasserstein distance (50). This is different from the corresponding contrast function with respect to the Fisher-Rao metric, which is related to the Hellinger distance and does not have this dependence on the sample space geometry. A fundamental research direction of Wasserstein geometry is concerned with the coupling of the curvature of M with the curvature of $P_+^\infty(M)$ equipped with the Otto metric [12, 13, 21]. Clearly, we do not have such a coupling for the Fisher-Rao metric. Instead, $P_+^\infty(M)$ has a spherical geometry with constant sectional positive curvature [26].

Various authors have introduced contrast functions related to Wasserstein geometry. These works are based on foundations that are different from the one of this article. The idea of a Bregman contrast function provides a particularly important foundational approach, for instance in the works [5, 27]. In case of the e_1-connection being torsion-free, the contrast function introduced in this article will be of Bregman type [23]. Studying the interrelations of proposed contrast functions in the context of Wasserstein geometry constitutes an important direction of research.

Acknowledgements The author would like to thank the two anonymous reviewers for their insightful comments. These comments contributed to a significant improvement of the article.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availibility statement Not applicable.

Declarations

Conflict of interest The author declares that he is the Editor-in-Chief of the journal. He was not involved in the peer review or handling of the manuscript. There is no other potential conflict of interest to declare.

References

1. Peyré, G., Cuturi, M.: Computational optimal transport: With applications to data science. Found. Trends Mach. Learn. **11**(5–6), 355–607 (2019)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning. PMLR, vol. 70, pp. 214–223 (2017)
3. Li, W., Montúfar, G.: Natural gradient via optimal transport. Inf. Geom. **1**, 181–214 (2018)
4. Amari, S., Karakida, R., Oizumi, M.: Information geometry connecting Wasserstein distance and Kullback-Leibler divergence via the entropy-relaxed transportation problem. Inf. Geom. **1**, 3–37 (2018)
5. Wong, T.L.: Logarithmic divergences from optimal transport and Rényi geometry. Inf. Geom. **1**, 39–78 (2018)
6. Amari, S., Nagaoka, H.: Methods of Information Geometry, vol. 191. American Mathematical Soc, Providence, RI (2000)
7. Malagó, L., Montrucchio, L., Pistone, G.: Wasserstein Riemannian geometry of Gaussian densities. Inf. Geom. **1**, 137–179 (2018)
8. Amari, S.: Natural gradient works efficiently in learning. Neural Comput. **10**(2), 251–276 (1998). https://doi.org/10.1162/089976698300017746
9. Chen, Y., Li, W.: Optimal transport natural gradient for statistical manifolds with continuous sample space. Inf. Geom. **3**, 1–32 (2020)
10. Khan, G., Zhang, J.: When optimal transport meets information geometry. Inf. Geom. **5**, 47–78 (2022)
11. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI (2003)

12. Villani, C.: Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften, vol. 338. Springer, Berlin (2009)
13. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**(1–2), 101–174 (2001)
14. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
15. Lafferty, J.D.: The density manifold and configuration space quantization. Trans. Am. Math. Soc. **305**(2), 699–741 (1988)
16. Kriegl, A., Michor, P.: The Convenient Setting of Global Analysis. Mathematical Surveys and Monographs, vol. 53. Amer. Math. Soc., Providence, RI (1997)
17. Pistone, G.: Information geometry of smooth densities on the Gaussian space: Poincaré inequalities. In: Signals and Communication Technology. Springer, Tokyo (2021)
18. Pistone, G.: Affine statistical bundle modeled on a Gaussian Orlicz-Sobolev space. Inf. Geom. **7**(S1), 109–130 (2022)
19. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information Geometry, vol. 64. Springer, Cham (2017)
20. Moser, J.: On the volume elements on a manifold. Trans. Am. Math. Soc. **120**(2), 286 (1965)
21. Lott, J.: Some geometric calculations on Wasserstein space. Commun. Math. Phys. **277**(2), 423–437 (2008)
22. Jost, J.: Riemannian Geometry and Geometric Analysis, 7th edn. Springer, Berlin (2017)
23. Ay, N., Amari, S.: A novel approach to canonical divergences within information geometry. Entropy **17**(12), 8111–8129 (2015). https://doi.org/10.3390/e17127866
24. Felice, D., Ay, N.: Towards a canonical divergence within information geometry. Inf. Geom. **4**, 65–130 (2021)
25. Henmi, M., Kobayashi, R.: Hooke's law in statistical manifolds and divergences. Nagoya Math. J. **159**, 1–24 (2000). https://doi.org/10.1017/S002776300000739X
26. Friedrich, T.: Die Fisher-Information und symplektische Strukturen. Math. Nachr. **152**, 273–296 (1991)
27. Li, W.: Transport information Bregman divergences. Inf. Geom. **4**, 435–470 (2021)