

THE L^p -FISHER-RAO METRIC AND AMARI-ČENCOV α -CONNECTIONS

MARTIN BAUER, ALICE LE BRIGANT, YUXIU LU, CY MAOR

ABSTRACT. We introduce a family of Finsler metrics, called the L^p -Fisher-Rao metrics F_p , for $p \in (1, \infty)$, which generalizes the classical Fisher-Rao metric F_2 , both on the space of densities $\text{Dens}_+(M)$ and probability densities $\text{Prob}(M)$. We then study their relations to the Amari-Čencov α -connections $\nabla^{(\alpha)}$ from information geometry: on $\text{Dens}_+(M)$, the geodesic equations of F_p and $\nabla^{(\alpha)}$ coincide, for $p = 2/(1 - \alpha)$. Both are pullbacks of canonical constructions on $L^p(M)$, in which geodesics are simply straight lines. In particular, this gives a new variational interpretation of α -geodesics as being energy minimizing curves. On $\text{Prob}(M)$, the F_p and $\nabla^{(\alpha)}$ geodesics can still be thought as pullbacks of natural operations on the unit sphere in $L^p(M)$, but in this case they no longer coincide unless $p = 2$. Using this transformation, we solve the geodesic equation of the α -connection by showing that the geodesics are pullbacks of projections of straight lines onto the unit sphere, and they always cease to exist after finite time when they leave the positive part of the sphere. This unveils the geometric structure of solutions to the generalized Proudman-Johnson equations, and generalizes them to higher dimensions. In addition, we calculate the associate tensors of F_p , and study their relation to $\nabla^{(\alpha)}$.

CONTENTS

1. Introduction	1
2. Spaces of densities and the Fisher-Rao metric	4
3. The L^p -Fisher-Rao metric and α -connections on the space of densities	5
4. The L^p -Fisher-Rao metric and α -connections on the space of probability densities	13
5. Summary of relations to known PDEs and metrics on diffeomorphism groups	19
6. A numerical comparison of geodesics on $\text{Dens}_+(M)$ and $\text{Prob}(M)$	20
7. Finite-dimensional geometry of parametric statistical models	21
Appendix A. Infinite dimensional Finsler geometry	26
References	28

1. INTRODUCTION

Information geometry is concerned with the study of spaces of probability densities as differentiable manifolds. Its first developments were mostly about the finite-dimensional geometry of parametric statistical models, for which the space of distributions can be identified with the parameter space. In 1945, Rao [37] showed that the Fisher information could be used to define a Riemannian metric on this space, and in 1982, Čencov [15] proved that it was the only metric invariant with respect to sufficient statistics, for families with finite sample spaces. The Fisher-Rao metric was also shown to induce well-known geometries on certain important statistical models, such as hyperbolic geometry on normal distributions [3].

Encompassing the Fisher-Rao metric, a richer geometric structure was introduced by Čencov [15] and Amari [2] on spaces of parametric probability distributions. The Amari-Čencov structure relies on a family of affine connections called the α -connections, denoted by $\nabla^{(\alpha)}$, that are dual with respect to the Fisher-Rao metric, and such that the 0-connection is the Levi-Civita connection.

The α -connections arise naturally as an interpolating family between the so-called exponential and mixture connections $\nabla^{(1)}$ and $\nabla^{(-1)}$, for which exponential and mixture families are (dually) flat manifolds. These geometric tools relate to natural information-theoretic quantities such as the Kullback-Leibler divergence, and have been used in statistical inference, e.g. to express conditions for existence of consistent and efficient estimators, or to obtain a purely geometric interpretation of the famous Expectation-Maximization (EM) algorithm in the presence of hidden variables [1].

In parallel, infinite-dimensional information geometry tools have also been developed in the non-parametric setting, although arguably to a lesser extent. The non-parametric Fisher-Rao metric was introduced by Friedrich in 1991 [21] on the space of all probability densities. He showed that it yields the historical Fisher information metric when restricted to finite-dimensional submanifolds representing parametric statistical models, and that the geometry is spherical with constant curvature $1/4$. More than two decades later, it was proved to be the only metric (up to a multiplicative factor) invariant with respect to the action of sufficient statistics, namely diffeomorphic change of the support, just like in the finite-dimensional case [4, 9]. In the infinite-dimensional setting, it is possible to work with diffeomorphisms of the support instead of the densities themselves, since the space of smooth densities on a compact manifold M with respect to a volume form λ can be obtained as the quotient $\text{Diff}(M)/\text{Diff}_\lambda(M)$ of diffeomorphisms modulo diffeomorphisms preserving λ . Using this representation Khesin, Lenells, Misiolek and Preston [26] have shown in 2013 that the Fisher-Rao metric can be obtained as the quotient of a right-invariant homogeneous Sobolev \dot{H}^1 -metric on $\text{Diff}(M)$, see also [33] and the recent overview article [27].

The Amari-Čencov structure induced by the α -connections also received interest in the non-parametric setting. Giblisco and Pistone [24] defined the exponential and mixture connections in this case, and showed that for $\alpha \in (-1, 1)$, the interpolating connections can be defined through a p -root mapping to an L^p sphere, for $p = \frac{2}{1-\alpha}$. Divergences and dualistic structures are investigated in the monograph of Ay, Jost, Lê and Schwachhöfer [5], although the α -connections themselves are not directly considered there in the infinite-dimensional setting. See also [35] for a definition of the α -divergences and α -connections in a Hilbert manifold settings. In [30], Lenells and Misiolek study the α -connections on diffeomorphisms and relate their geodesic equations to a well-known equation, the generalized Proudman–Johnson equation. Very recently, three authors of the present paper showed that these Proudman–Johnson equations, on the real line, could alternatively be seen as the geodesic equations of right-invariant Finsler metrics on the diffeomorphism group [11], which were first introduced in [18]. This led to making a first link between α -connections and a family of Finsler metrics, which we investigate further here.

1.1. Main contributions. The aim of the present paper is three-fold. First, to introduce and study the L^p -Fisher-Rao metrics on (probability) densities

$$F_p(a) := F_p(\mu, a) = \left(\int \left| \frac{a}{\mu} \right|^p \mu \right)^{\frac{1}{p}},$$

for $p \in (1, \infty)$ and any density μ and tangent vector a . Note, that is a family of Finsler metrics that coincides with the Fisher-Rao metric when $p = 2$. Second, to give a precise and rigorous review of the Amari-Čencov α -connections in the infinite-dimensional setting, a new variational formulation of their corresponding geodesics, and explicit solution formulas for them. Finally, to make links between the two, distinguishing between the space of densities, the space of probability densities, and parametric statistical models.

Next we will describe the main contributions in more details: we study the L^p -Fisher-Rao geometry of (probability) densities through a mapping to the set of positive functions,

$$\Phi_p(\mu) = \left(\frac{\mu}{\lambda} \right)^{1/p},$$

where λ is some background probability measure. Just like the Fisher-Rao metric is the pullback of the standard L^2 -metric via the square-root transform [26, 13, 22], we show that the L^p -Fisher-Rao metric is the pullback of the L^p -norm via the mapping Φ_p , that we call by analogy the *p-root transform* (Theorems 3.12 and 4.10). The L^p -Fisher-Rao geometry on the space of densities is therefore that of a flat space, as described in Corollary 3.13, and on the space of probability densities that of the L^p -sphere (Theorem 4.10). The *p-root transform* (for $p = \frac{2}{1-\alpha}$) also presents an alternative way to define the α -connections as pullbacks of the trivial connection of the vector space of functions (Theorems 3.12 and 4.10), as first shown by Gibilisco and Pistone [24] for probability distributions, albeit with a slightly different construction. The geometric differences between these constructions for the L^p -Fisher-Rao metric and the α -connections, which we systematically study in this paper, are summarized in Figure 1.

Towards this aim, we show that the geodesic equations of F_p and $\nabla^{(\alpha)}$ coincide on $\text{Dens}_+(M)$ (for $\alpha = 1 - 2/p$), but not on $\text{Prob}(M)$ (see Theorems 3.3, 3.7, 4.2 and 4.4); similarly, on $\text{Dens}_+(M)$ the Chern connection induced by F_p coincides with the α -connection, while this no longer holds on $\text{Prob}(M)$ (Theorem 3.10 and Remark 4.8). This provides the novel variational formulation of these α -connection geodesics.

We further use the *p-root transform* to obtain explicit solution formulas for α -geodesics on densities and on probability densities: for densities, we show in Corollary 3.13 that geodesics are pullbacks of straight lines in L^p space, whereas for probability densities we show in Theorem 4.11 that they are pullbacks of projections of straight lines in L^p onto the L^p -sphere. In the latter case the projection involves a time rescaling that is obtained as a solution of an ordinary differential equation. Similar solutions of the geodesic equation of the α -connection were obtained for finite sample space [5, pp. 50-51]. In the infinite-dimensional case with a one-dimensional base manifold M , it gives an explicit solution (modulo a solution to an ODE) of the generalized Proudman-Johnson equation, for a certain range of parameters, and to the generalization to higher-dimensional base manifolds by Lenells and Misiólek [30]. There, they proved the complete integrability of these equations for the flat case $\alpha = \pm 1$ by providing an explicit solution formula. Similarly, the integrability for the case $\alpha = 0$ was shown in [26]. Our results can thus be interpreted as complete integrability of the α -geodesic equation for the whole range $\alpha \in (-1, 1)$.

The results in the one-dimensional situation are in correspondence with the analysis of [29, 38], where a similar *p-root transform* was used to study the generalized Proudman-Johnson equation. In these articles it was used as an ad-hoc simplification of some auxiliary equations; here we expose the geometry behind it, which also simplifies some of the authors' calculations, and generalize it to higher dimensions. These connections are summarized in Section 5.

Throughout this paper we work in the smooth category, i.e., all densities are assumed to be smooth, and the underlying space M is assumed to be a smooth manifold. This is mainly in order to avoid some technicalities, and most results work in much lower regularity. For example, for all results not involving the action of $\text{Diff}(M)$, the underlying space M can be simply a measurable space, and in many cases densities only need to be integrable.

1.2. Outline. The rest of the paper is organized as follows. We start by describing some background on spaces of densities and the Fisher-Rao metric in Section 2. Then we investigate the geometries induced by the α -connections and the L^p -Fisher-Rao metrics as well as their links, on the space of smooth densities in Section 3 and on the space of probability densities in Section 4. In Section 5 we discuss the relations of the various geodesic equations obtained in Sections 3-4 to some known PDEs, as well as the relation between the L^p -Fisher-Rao metric to Finsler metrics on diffeomorphism groups. The different notions of geodesics are compared numerically on an example in Section 6. Finally, we consider the finite-dimensional setting of parametric statistical models in Section 7, illustrated by the special case of normal distributions. In Appendix A we present a short overview of infinite-dimensional Finsler geometry.

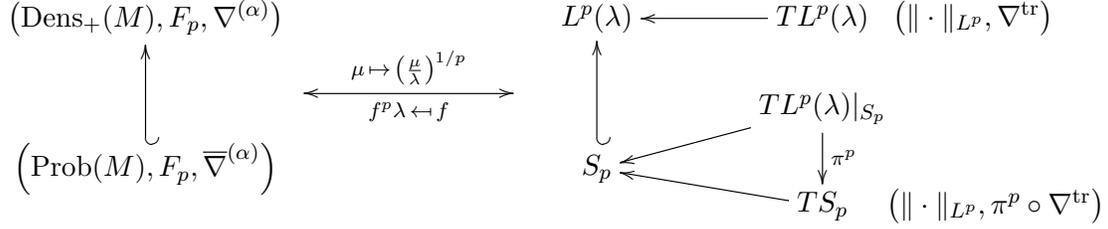


FIGURE 1. Geometric structures on $\text{Dens}_+(M)$ and $\text{Prob}(M)$ via the p -root transform: The map $\mu \mapsto (\frac{\mu}{\lambda})^{1/p}$ maps $\text{Dens}_+(M)$ to (a subset of) $L^p(\lambda)$, and $\text{Prob}(M)$ to its unit sphere S_p . On $L^p(\lambda)$ there is the natural Finsler metric $\|\cdot\|_{L^p}$ and the trivial connection ∇^{tr} of a vector space, the geodesics of both are straight lines. Their pullback via the p -root map yield (up to a constant) the L^p -Fisher-Rao metric F_p and the α -connection $\nabla^{(\alpha)}$ on $\text{Dens}_+(M)$, whose geodesic equations coincide. The metric $\|\cdot\|_{L^p}$ naturally restricts to S_p . The connection ∇^{tr} induces a connection on S_p via the natural projection $\pi^p : TL^p(\lambda)|_{S_p} \rightarrow TS_p$. The geodesics of these induced metric and connection differ. Their pullbacks via the p -root map yield (up to a constant) F_p and the α -connection $\bar{\nabla}^{(\alpha)}$ on $\text{Prob}(M)$.

Acknowledgements. Parts of this work are contained in the PhD-thesis of the third author [31]. MB was partially supported by NSF grants DMS-1912037 and DMS-1953244 and by FWF grant FWF-P 35813-N. The first two authors acknowledge support of the Institut Henri Poincaré (IHP, UAR 839 CNRS-Sorbonne Université), and LabEx CARMIN (ANR- 10-LABX-59-01). CM was partially supported by ISF grant 1269/19.

Data availability statement. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

2. SPACES OF DENSITIES AND THE FISHER-RAO METRIC

In all of this article let M be a closed manifold of dimension $\dim(M) < \infty$. We denote by $\text{Dens}_+(M)$ the space of smooth positive densities and by $\text{Prob}(M)$ the subspace of smooth probability densities, i.e.,

$$\begin{aligned} \text{Dens}_+(M) &:= \{\mu \in \Omega^n(M) : \mu > 0\} \\ \text{Prob}(M) &:= \left\{ \mu \in \text{Dens}_+(M) : \int \mu = 1 \right\}. \end{aligned}$$

Since $\text{Dens}_+(M)$ is an open subset of the Fréchet space $\Omega^n(M)$ it carries the structure of a Fréchet manifold with tangent space $T_\mu \text{Dens}(M) = \Omega^n(M)$. Similarly, as a linear subspace of a Fréchet manifold, the space of probability densities is a Fréchet manifold, where the tangent space is given by

$$T_\mu \text{Prob}(M) = \left\{ a \in \Omega^n(M) : \int a = 0 \right\}.$$

On both the space of densities and probability densities we can consider the pushforward action of the diffeomorphism group $\text{Diff}(M)$. On $\text{Dens}_+(M)$ it is given by

$$(1) \quad \text{Diff}(M) \times \text{Dens}_+(M) \ni (\varphi, \mu) \mapsto \varphi_* \mu \in \text{Dens}_+(M)$$

and, since the pushforward by a diffeomorphism is volume preserving, this action restricts to an action on the space of probability densities. By a result of Moser [34] this action is transitive, which

allows us to identify the space of probability densities with the quotient

$$(2) \quad \text{Prob}(M) \equiv \text{Diff}(M)/\text{Diff}_\lambda(M),$$

where $\text{Diff}_\lambda(M)$ is the group of volume preserving diffeomorphisms of some fixed probability density λ . Thus, constructions (metrics, connections, geodesics) on $\text{Prob}(M)$ can be pulled back to $\text{Diff}(M)$ via the map $\varphi \mapsto \varphi_*\lambda$.

For $a \in \Omega^n(M)$ and $\mu \in \text{Dens}_+(M)$, we denote by $\frac{a}{\mu}$ the Radon-Nikodym derivative of a with respect to μ . In particular, the map $\mu \mapsto \frac{\mu}{\lambda}$ allows us to identify $\text{Dens}_+(M)$ with positive smooth functions on M , and $\text{Prob}(M)$ with the positive smooth functions that integrate to one. For the proof of the local wellposedness results in Sections 3 and 4 we will also need the Sobolev completions of these spaces, which can be defined using their Radon-Nikodym derivative w.r.t. to λ , i.e., for $k > \dim(M)/2$ we consider

$$\begin{aligned} \text{Dens}_+^k(M) &:= \left\{ \mu : \frac{\mu}{\lambda} \in H^k(M), \text{ and } \mu > 0 \right\} \\ \text{Prob}^k(M) &:= \left\{ \mu \in \text{Dens}_+^k(M) : \int \mu = 1 \right\}. \end{aligned}$$

Note, that the assumption $k > \dim(M)/2$ is necessary to make sense of the positivity condition.

A central object in information geometry is the Fisher-Rao metric, which we introduce now:

Definition 2.1 (Fisher-Rao metric). Given $\mu \in \text{Dens}_+(M)$ and $a, b \in T_\mu \text{Dens}_+(M)$ the Fisher-Rao metric on $\text{Dens}_+(M)$ is given by

$$(3) \quad G_\mu^{\text{FR}}(a, b) = \int \frac{a}{\mu} \frac{b}{\mu} \mu.$$

Via restriction G^{FR} induces a Riemannian metric on $\text{Prob}(M)$, which we denote by the same letter.

3. THE L^p -FISHER-RAO METRIC AND α -CONNECTIONS ON THE SPACE OF DENSITIES

In this section we will introduce the L^p -Fisher-Rao metric on the space of densities, which will allow us to obtain a new interpretation of the family of α -connections.

3.1. The Amari-Čencov α -connections on $\text{Dens}_+(M)$. First we will introduce the family of α -connections on the space $\text{Dens}_+(M)$. In the finite-dimensional case, i.e., when M is a finite set, the below definitions coincide with the classical ones, see e.g. [2, 4].

Definition 3.1 (α -divergence). For $\alpha \in (-1, 1)$, define the α -divergence $D^{(\alpha)} : \text{Dens}_+(M) \times \text{Dens}_+(M) \rightarrow \mathbb{R}$, by

$$D^{(\alpha)}(\mu||\nu) = p \int_M \nu + p^* \int_M \mu - p^* p \int_M \left(\frac{\mu}{\lambda}\right)^{1/p} \left(\frac{\nu}{\lambda}\right)^{1/p^*} \lambda,$$

where $p = \frac{2}{1-\alpha}$ and $p^* = \frac{2}{1+\alpha}$ is its Hölder conjugate.

Using Hölder inequality, it follows that $D^{(\alpha)}$ is non-negative and vanishes if and only if $\mu = \nu$. Furthermore, a straightforward calculation shows that the negative of its second derivative defines a positive bilinear form, which is exactly the Fisher-Rao metric, i.e.,

$$-\partial_\mu \partial_\nu D^{(\alpha)}(\mu||\nu)|_{\nu=\mu}[a, b] = \int_M \frac{a}{\mu} \frac{b}{\mu} \mu = G_\mu^{\text{FR}}(a, b), \quad a, b \in T_\mu \text{Dens}_+(M).$$

Here ∂_μ and ∂_ν refer to derivatives with respect to the μ and ν variables, respectively. Thus, for any $\alpha \in (-1, 1)$, $D^{(\alpha)}$ is a divergence in the sense of [4, Section 4.4], and induces a connection $\nabla^{(\alpha)}$ on $\text{Dens}_+(M)$ via the relation

$$(4) \quad G_\mu^{\text{FR}}(\nabla_a^{(\alpha)} b, c) = -\partial_\mu(\partial_\mu \partial_\nu D^{(\alpha)}(\mu||\nu)[b, c])[a]|_{\nu=\mu} = \int_M \frac{Db \cdot a}{\mu} \frac{c}{\mu} \mu - \frac{1}{p^*} \int_M \frac{a}{\mu} \frac{b}{\mu} \frac{c}{\mu} \mu,$$

where $a, b, c \in T_\mu \text{Dens}_+(M)$. Since $\text{Dens}_+(M)$ is a Fréchet manifold, G^{FR} is merely a weak Riemannian metric, and as such, (4) does not necessarily define $\nabla^{(\alpha)}$ uniquely. However, in our case it does, yielding the following formulae:

Lemma 3.2 (α -connection). *For any $\alpha \in (-1, 1)$ the α -connections $\nabla^{(\alpha)}$ on $\text{Dens}_+(M)$ are given by*

$$(5) \quad \nabla_a^{(\alpha)} b = Db.a - \frac{1}{p^*} \frac{a}{\mu} b, \quad a, b \in T_\mu \text{Dens}_+(M), \quad p^* = \frac{2}{1 + \alpha}.$$

Here $Db.a|_\mu := D_\mu b(a_\mu)$ denotes the directional derivative of the vector field b in the direction given by a_μ .

The easiest way to read this lemma (and similar formulae below) is to consider again the identification of densities and positive functions via $\mu \mapsto \mu/\lambda$.

Proof. This follows directly from formula (4). □

In the following result we will study the local wellposedness of the corresponding geodesic equations. Therefore we will first consider these equations on a Banach space of Sobolev densities, where it will be easy to obtain the local wellposedness using the theorem of Picard-Lindelöf. The result in the smooth category will then follow from an Ebin-Marsden type no-loss-no-gain result [19]:

Theorem 3.3. *A path $\mu : [0, 1] \rightarrow \text{Dens}_+(M)$ is a geodesic with respect to $\nabla^{(\alpha)}$ if*

$$(6) \quad \mu_{tt} = \frac{1}{p^*} \frac{\mu_t}{\mu} \mu_t.$$

For any $k > \dim(M)/2$ the geodesic equations are locally wellposed on the space of Sobolev densities $\text{Dens}_+^k(M)$, i.e., given initial conditions $\mu(0) \in \text{Dens}_+^k(M)$, $\mu_t(0) \in T_{\mu(0)} \text{Dens}_+^k(M)$ there exists an unique solution to equation (6) defined on a maximal interval of existence $[0, T)$. The maximal interval of existence is uniform in the Sobolev order k and thus the local wellposedness continues to hold in the limit, i.e., on the space of smooth densities $\text{Dens}_+(M)$.

Proof. The formula for the geodesic equation follows directly from Lemma 3.2. To show the local well-posedness we view the geodesic equation (6) as a flow equation on $T \text{Dens}_+^k(M)$. Therefore let $F(\mu_t)$ denote the right hand side of the geodesic equation, i.e.,

$$(7) \quad F(\mu, \mu_t) = \mu^{-1} \mu_t^2$$

where we use the identification of $\text{Dens}_+^k(M)$ with the space of positive, Sobolev functions $H_+^k(M)$ and $T_\mu \text{Dens}_+^k(M)$ with all of $H^k(M)$. Using the Sobolev module properties and the positivity of μ it follows that F is a smooth map from $H_+^k(M) \times H^k(M)$ and thus the local well-posedness follows by the theorem of Picard-Lindelöf. Next, we observe that F is equivariant under the action of the diffeomorphism group $\text{Diff}(M)(M)$, i.e., $F(\varphi^* \mu, \varphi^* \mu_t) = \varphi^* F(\mu, \mu_t)$. Thus the result on the uniformness of the maximal interval of existence follows by an adaption of the Ebin-Marsden no-loss-no-gain theorem [19, Lemma 12.2] to the present setting, i.e., the diffeomorphism group acting on densities. This can be achieved by following the proof in [7], where the no-loss-no-gain result has been extended to the action of diffeomorphisms on the space of all Riemannian metrics. The key ingredient for this result is the fact that, in a chart, Lie derivatives along coordinate vector fields coincide with ordinary derivatives. □

3.2. The L^p -Fisher-Rao metric. Next we introduce the main object of the present article, the L^p -Fisher-Rao (Finsler) metric on the space $\text{Dens}_+(M)$:

Definition 3.4. Given $\mu \in \text{Dens}_+(M)$ and $a \in T_\mu \text{Dens}_+(M)$ we define the L^p -Fisher-Rao metric via:

$$(8) \quad F_p(a) := F_p(\mu, a) = \left(\int \left| \frac{a}{\mu} \right|^p \mu \right)^{\frac{1}{p}}.$$

Remark 3.5. It is easy to see that the L^p -Fisher-Rao metric satisfies the axioms of a Finsler metric, as defined in Definition A.1, for any $p \in (1, \infty)$. We will, however, see in Lemma 3.8, that it is not strongly convex if $p \neq 2$.

First we will show, that the family of L^p -Fisher-Rao metrics shares an important property with the Fisher-Rao metric: they are invariant under the action of the diffeomorphism group as defined in (1).

Lemma 3.6. *For any $p \in (1, \infty)$, the L^p -Fisher-Rao metric on the space of $\text{Dens}_+(M)$ is invariant under the action of the diffeomorphism group $\text{Diff}(M)$, i.e.,*

$$(9) \quad F_p(\mu, a) = F_p(\varphi_*\mu, \varphi_*a), \quad a \in T_\mu \text{Dens}_+(M), \varphi \in \text{Diff}(M).$$

Proof. This result follows by direct computation using the transformation formula for integrals. \square

Next we calculate the geodesic equations of this family of Finsler metrics on $\text{Dens}_+(M)$.

Theorem 3.7 (Geodesic equation on $\text{Dens}_+(M)$). *For any $p \in (1, \infty)$, the geodesic equation of the L^p -Fisher-Rao metric on the space of densities $\text{Dens}_+(M)$ is given by*

$$(10) \quad \frac{d}{dt} \left(\frac{\mu_t}{\mu} \right) + \frac{1}{p} \left(\frac{\mu_t}{\mu} \right)^2 = 0,$$

which coincides with the geodesic equation of the α -connection for $\alpha = 1 - \frac{2}{p}$. Thus the local well-posedness result of Theorem 3.3 also hold for the geodesic equation of the L^p -Fisher-Rao metric.

Proof. The length functional of the L^p -Fisher-Rao metric on $\text{Dens}_+(M)$ is given by

$$L(\mu) = \int_0^1 \left(\int \left| \frac{\mu_t}{\mu} \right|^p \mu \right)^{\frac{1}{p}} dt,$$

where $\mu : [0, 1] \rightarrow \text{Dens}_+(M)$ such that $\mu(0) = \mu_0$, $\mu(1) = \mu_1$ and where μ_t denotes its (time) derivative. A geodesic is a path that locally minimizes the length functional; since L is invariant to reparametrization, we can restrict ourselves to paths of constant speed. By the Hölder inequality, it follows that constant speed geodesics are equivalently the local minimizers of the q -energy

$$E_q(\mu) = \frac{1}{q} \int_0^1 \left(\int \left| \frac{\mu_t}{\mu} \right|^p \mu \right)^{\frac{q}{p}} dt,$$

for any $q > 1$. In our case the most convenient choice is to consider the q -Energy with $q = p$. The corresponding energy functional reads as

$$E_p(\mu) = \frac{1}{p} \int_0^1 \int \left| \frac{\mu_t}{\mu} \right|^p \mu dt.$$

Calculating the variation of the p -energy functional in direction $\delta\mu$ leads to

$$(11) \quad \begin{aligned} \delta E_p(\mu)(\delta\mu) &= \frac{1}{p} \int_0^1 \int p \left| \frac{\mu_t}{\mu} \right|^{p-2} \frac{\mu_t}{\mu} \delta\mu_t - (p-1) \left| \frac{\mu_t}{\mu} \right|^p \delta\mu \, d\lambda \, dt \\ &= -\frac{1}{p} \int_0^1 \int \left(p \frac{d}{dt} \left(\left| \frac{\mu_t}{\mu} \right|^{p-2} \frac{\mu_t}{\mu} \right) + (p-1) \left| \frac{\mu_t}{\mu} \right|^p \right) \delta\mu \, d\lambda \, dt, \end{aligned}$$

where we used integration by parts in time t and that the variational direction vanishes at the end points, i.e., $\delta\mu(0) = \delta\mu(1) = 0$. From here we can immediately read off the geodesic equation

$$p \frac{d}{dt} \left(\left| \frac{\mu_t}{\mu} \right|^{p-2} \frac{\mu_t}{\mu} \right) + (p-1) \left| \frac{\mu_t}{\mu} \right|^p = 0.$$

which can be simplified to the desired formula. That this equation coincides with the geodesic equation of the α -connection can be seen by comparing it to the equation of Theorem 3.3. \square

Next we will study the Finslerian geometry induced by the L^p -Fisher-Rao metric (see Appendix A for a short overview of the main definitions). We will see in the next Lemma, that the L^p -Fisher-Rao metric is, in general, not strongly convex and thus some of the calculations in this and the next sections have to be understood formally.

Lemma 3.8 (The Hessian matrix). *Let $\mu \in \text{Dens}_+(M)$ and $\nu, a, b \in T_\mu \text{Dens}_+(M)$. The Hessian matrix g^ν of the squared L^p -Fisher-Rao metric at ν is given by*

$$(12) \quad g_\mu^\nu(a, b) = (p-1)I(\nu, \nu)^{\frac{2}{p}-1}I(a, b) - (p-2)I(\nu, \nu)^{\frac{2}{p}-2}I(\nu, a)I(\nu, b).$$

where

$$(13) \quad I(a, b) := I_\mu^\nu(a, b) := \int \left| \frac{\nu}{\mu} \right|^{p-2} \frac{a}{\mu} \frac{b}{\mu} \mu.$$

If ν is nowhere zero than g^ν is positive definite and thus a Riemannian metric. If ν vanishes on an open set $\mathcal{U} \subset M$ then, for $p > 2$, g^ν is degenerate as it vanishes for all $a, b \in T_\mu \text{Dens}_+(M)$ with support contained in \mathcal{U} , and for $p < 2$ it is not well-defined.

Proof. We introduce the notations

$$\tilde{\omega} = \tilde{\omega}(r, s) := \frac{\nu}{\mu} + r \frac{a}{\mu} + s \frac{b}{\mu},$$

$$\omega := |\tilde{\omega}(r, s)| = \left| \frac{\nu}{\mu} + r \frac{a}{\mu} + s \frac{b}{\mu} \right|.$$

To compute the Hessian matrix of $F_p^2(\mu, \nu)$ we need to calculate the second derivative in r and s of $F_p^2(\mu, \omega)$. We have

$$\partial_r F_p^2(\mu, \omega) = 2 \left(\int \omega^p \mu \right)^{2/p-1} \int \omega^{p-1} \partial_r \omega \mu = 2 \left(\int \omega^p \mu \right)^{2/p-1} \int \omega^{p-1} \text{sgn}(\tilde{\omega}) \frac{a}{\mu} \mu.$$

For the second derivative we get

$$\begin{aligned} \partial_s \partial_r F_p^2(\mu, \omega) &= 2(2-p) \left(\int \omega^p \mu \right)^{2/p-2} \int \omega^{p-1} \text{sgn}(\tilde{\omega}) \frac{a}{\mu} \mu \int \omega^{p-1} \text{sgn}(\tilde{\omega}) \frac{b}{\mu} \mu \\ &\quad + 2(p-1) \left(\int \omega^p \mu \right)^{2/p-1} \int \omega^{p-2} \frac{a}{\mu} \frac{b}{\mu} \mu. \end{aligned}$$

Evaluating at $r = s = 0$ yields the desired formula for g_μ^ν .

For $\nu \neq 0$ we can use the Cauchy-Schwarz inequality to prove the positive-definiteness of the Hessian:

$$\left(\int \left| \frac{\nu}{\mu} \right|^{p-1} \left| \frac{a}{\mu} \right| \mu \right)^2 \leq \int \left| \frac{\nu}{\mu} \right|^p \mu \int \left| \frac{\nu}{\mu} \right|^{p-2} \left(\frac{a}{\mu} \right)^2 \mu.$$

Then we get the inequality

$$g^\nu(a, a) \geq \left(\int \left| \frac{\nu}{\mu} \right|^p \mu \right)^{2/p-1} \int \left| \frac{\nu}{\mu} \right|^{p-2} \left(\frac{a}{\mu} \right)^2 \mu.$$

Thus for ν being a nowhere vanishing vector field, $g^\nu(a, a) = 0$ implies that $\frac{a}{\mu} = 0$. \square

Lemma 3.9 (The Cartan tensor). *Let $\mu \in \text{Dens}_+(M)$ and $\nu, a, b, c \in T_\mu \text{Dens}_+(M)$. The Cartan tensor of the L^p -Fisher-Rao metric is given by*

$$(14) \quad \begin{aligned} C^\nu(a, b, c) = & \frac{1}{2}(p-1)(p-2)I(\nu, \nu)^{\frac{2}{p}-3} \left(2I(\nu, a)I(\nu, b)I(\nu, c) \right. \\ & - I(\nu, \nu)I(\nu, a)I(b, c) - I(\nu, \nu)I(\nu, b)I(c, a) \\ & \left. - I(\nu, \nu)I(\nu, c)I(a, b) + I(\nu, \nu)^2 J(\nu, a, b, c) \right), \end{aligned}$$

where

$$(15) \quad \begin{aligned} I(a, b) &:= I_\mu^\nu(a, b) := \int \left| \frac{\nu}{\mu} \right|^{p-2} \frac{a}{\mu} \frac{b}{\mu} \mu, \\ J(a, b, c) &:= J_\mu^\nu(a, b, c) := \int \left| \frac{\nu}{\mu} \right|^{p-4} \frac{a}{\mu} \frac{b}{\mu} \frac{c}{\mu} \mu. \end{aligned}$$

Proof. This formula can be derived similarly as the formula for the Hessian by computing

$$C^\nu(a, b, c) = \partial_r \partial_s \partial_t |_{r=s=t=0} F_p(\mu, \tilde{\omega})$$

where $\tilde{\omega}(r, s, t) = \nu + ra + sb + tc$. □

3.3. The α -connection as Chern connection of the L^p -Fisher-Rao metric. Next we will show that the Chern connection associated to the L^p -Fisher-Rao metric on $\text{Dens}_+(M)$ is an α -connection, when two entries are taken to be the same.

Theorem 3.10 (The Chern connection on $\text{Dens}_+(M)$). *Let $\alpha = 1 - \frac{2}{p}$. For every nowhere vanishing vector field ν on $\text{Dens}_+(M)$ and any $a \in T_\mu \text{Dens}_+(M)$, we have*

$$(16) \quad \nabla_a^\nu \nu = \nabla_a^{(\alpha)} \nu,$$

where ∇^ν is the Chern connection induced by the L^p -Fisher-Rao metric and $\nabla^{(\alpha)}$ is the α -connection on $\text{Dens}_+(M)$ defined by (5).

Proof. Formula (16) defines the Chern connection if and only if it verifies the generalized Koszul formula (see Lemma A.8)

$$\begin{aligned} 2g^\nu(\nabla_a^\nu \nu, b) = & ag^\nu(\nu, b) + \nu g^\nu(b, a) - bg^\nu(a, \nu) + g^\nu([a, \nu], b) - g^\nu([\nu, b], a) + g^\nu([b, a], \nu) \\ & - 2C^\nu(\nabla_a^\nu \nu, \nu, b) - 2C^\nu(\nabla_\nu^\nu \nu, b, a) + 2C^\nu(\nabla_\nu^\nu b, a, \nu). \end{aligned}$$

Since the Cartan tensor verifies $C^\nu(\nu, \cdot, \cdot) = 0$ this formula reduces to

$$(17) \quad \begin{aligned} 2g^\nu(\nabla_a^\nu \nu, b) = & ag^\nu(\nu, b) + \nu g^\nu(b, a) - bg^\nu(a, \nu) \\ & + g^\nu([a, \nu], b) - g^\nu([\nu, b], a) + g^\nu([b, a], \nu) - 2C^\nu(\nabla_\nu^\nu \nu, b, a). \end{aligned}$$

To compute the first terms of the right hand-side of this equality, we will need

$$cI(a, b) = I(Da.c, b) + I(a, Db.c) - (p-1)K(a, b, c) + (p-2)J(\nu, a, b, D\nu.c),$$

where I and J are defined by (15), and

$$K(a, b, c) := K_\mu^\nu(a, b, c) = \int \left| \frac{\nu}{\mu} \right|^{p-2} \frac{a}{\mu} \frac{b}{\mu} \frac{c}{\mu} \mu.$$

Using this we get

$$\begin{aligned} ag^\nu(\nu, b) &= -\frac{p-2}{p}I(\nu, \nu)^{2/p-2}I(\nu, b) (pI(\nu, D\nu.a) - (p-1)K(\nu, \nu, a)) \\ &\quad + I(\nu, \nu)^{2/p-1} (I(\nu, Db.a) - (p-1)K(\nu, a, b) + (p-1)I(b, D\nu.a)), \\ bg^\nu(\nu, a) &= -\frac{p-2}{p}I(\nu, \nu)^{2/p-2}I(\nu, a) (pI(\nu, D\nu.b) - (p-1)K(\nu, \nu, b)) \\ &\quad + I(\nu, \nu)^{2/p-1} (I(\nu, Da.b) - (p-1)K(\nu, a, b) + (p-1)I(a, D\nu.b)), \end{aligned}$$

and

$$\begin{aligned} \nu g^\nu(a, b) &= -\frac{(p-1)(p-2)}{p}I(\nu, \nu)^{2/p-2}I(a, b) (pI(\nu, D\nu.\nu) - (p-1)K(\nu, \nu, \nu)) \\ &\quad + (p-1)I(\nu, \nu)^{2/p-1} (I(Da.\nu, b) + I(a, Db.\nu) - (p-1)K(\nu, a, b) + (p-2)J(\nu, a, b, D\nu.\nu)) \\ &\quad + \frac{2(p-1)(p-2)}{p}I(\nu, \nu)^{2/p-3}I(\nu, a)I(\nu, b) (pI(\nu, D\nu.\nu) - (p-1)K(\nu, \nu, \nu)) \\ &\quad - (p-2)I(\nu, \nu)^{2/p-2}I(\nu, b) (I(\nu, Da.\nu) - (p-1)K(\nu, \nu, a) + (p-1)I(a, D\nu.\nu)) \\ &\quad - (p-2)I(\nu, \nu)^{2/p-2}I(\nu, a) (I(\nu, Db.\nu) - (p-1)K(\nu, \nu, b) + (p-1)I(b, D\nu.\nu)). \end{aligned}$$

The following terms of the right hand-side of the generalized Koszul formula (17) are given by

$$\begin{aligned} g^\nu([a, \nu], b) &= (p-1)I(\nu, \nu)^{2/p-1}I(D\nu.a - Da.\nu, b) \\ &\quad - (p-2)I(\nu, \nu)^{2/p-2}I(\nu, D\nu.a - Da.\nu)I(\nu, b), \\ g^\nu([\nu, b], a) &= (p-1)I(\nu, \nu)^{2/p-1}I(Db.\nu - D\nu.b, a) \\ &\quad - (p-2)I(\nu, \nu)^{2/p-2}I(\nu, Db.\nu - D\nu.b)I(\nu, a), \\ g^\nu([b, a], \nu) &= I(\nu, \nu)^{2/p-1}I(\nu, Da.b - Db.a). \end{aligned}$$

Finally there remains to compute the two terms involving the Chern connection, i.e. the term on the left hand-side and the last term of the right hand-side. With the chosen value of α , we have

$$\nabla_a^\nu \nu = D\nu.a - \frac{p-1}{p} \frac{a \nu}{\mu \mu} \mu,$$

and so

$$\begin{aligned} I(\nabla_a^\nu \nu, b) &= I(D\nu.a, b) - \frac{p-1}{p}K(\nu, a, b) \\ J(\nu, a, b, \nabla_\nu^\nu \nu) &= J(\nu, a, b, D\nu.\nu) - \frac{p}{p-1}K(\nu, a, b). \end{aligned}$$

This yields, using (14),

$$\begin{aligned}
2C^\nu(\nabla_\nu^\nu \nu, b, a) &= (p-1)(p-2)I(\nu, \nu)^{2/p-3} \\
&\cdot \left[2I(\nu, a)I(\nu, b) \left(I(\nu, D\nu.\nu) - \frac{p-1}{p}K(\nu, \nu, \nu) \right) \right. \\
&\quad - I(\nu, \nu)I(\nu, a) \left(I(b, D\nu.\nu) - \frac{p-1}{p}K(\nu, \nu, b) \right) \\
&\quad - I(\nu, \nu)I(\nu, b) \left(I(a, D\nu.\nu) - \frac{p-1}{p}K(\nu, \nu, a) \right) \\
&\quad - I(\nu, \nu)I(a, b) \left(I(\nu, D\nu.\nu) - \frac{p-1}{p}K(\nu, \nu, \nu) \right) \\
&\quad \left. + I(\nu, \nu)^2 \left(J(\nu, a, b, D\nu.\nu) - \frac{p-1}{p}K(\nu, a, b) \right) \right].
\end{aligned}$$

Putting all the terms together yields the left hand-side of the generalized Koszul formula (17), i.e.

$$\begin{aligned}
2g^\nu(\nabla_a^\nu \nu, b) &= 2(p-1)I(\nu, \nu)^{2/p-1} \left(I(D\nu.a, b) - \frac{p-1}{p}K(\nu, a, b) \right) \\
&\quad - 2(p-2)I(\nu, \nu)^{2/p-2}I(\nu, b) \left(I(D\nu.a, \nu) - \frac{p-1}{p}K(\nu, \nu, a) \right).
\end{aligned}$$

□

As a direct consequence of the above characterization of the α -connections as a Chern connection we obtain that these connections have an interpretation as describing energy minimizing curves:

Corollary 3.11. *Let $\alpha \in (-1, 1)$. Geodesic curves of the α -connection describe locally minimizing curves of the $\frac{2}{1-\alpha}$ -Energy*

$$E_{\frac{2}{1-\alpha}}(\mu) = \frac{1-\alpha}{2} \int_0^1 \int \left| \frac{\mu_t}{\mu} \right|^{\frac{2}{1-\alpha}} \mu dt.$$

3.4. The p -root transform. Next, we will isometrically map the space of densities to a simpler space, which will allow us to obtain explicit expressions for solutions to the geodesic equation; we call this construction, which is a direct generalization of the square-root transform for the Fisher-Rao metric, the p -root transform. At the same time the p -root transform presents an alternative way to define the α -connection. This has been first proposed by Gibilisco and Pistone [24], who considered this construction specifically for the space $\text{Prob}(M)$ albeit with slightly different notations and a different identification of a tangent vector with a function.

Theorem 3.12. *Endow the space $C^\infty(M)$ of smooth functions with the standard L^p -norm and with the trivial vector space connection ∇^{tr} , i.e., for two vector fields $\xi, \eta : C^\infty(M) \rightarrow C^\infty(M)$,*

$$\nabla_\xi^{tr} \eta = D\eta.\xi.$$

Let $\alpha \in (-1, 1)$ and, as before, denote $p = \frac{2}{1-\alpha}$. Define the map $\Phi_p : \text{Dens}_+(M) \rightarrow C^\infty(M)$ by

$$(18) \quad \Phi_p(\mu) = \left(\frac{\mu}{\lambda} \right)^{1/p}.$$

We have:

- (a) The image $\Phi_p(\text{Dens}_+(M))$ is the set of all positive functions in $C^\infty(M)$.
- (b) The mapping Φ_p is an isometric embedding, where $\text{Dens}_+(M)$ is equipped with a multiple of the L^p -Fisher-Rao metric and where $C^\infty(M)$ is viewed as a vector space equipped with the standard L^p -norm.

(c) The pullback of $\Phi_p^* \nabla^{tr}$ coincides with $\nabla^{(\alpha)}$ up to a constant depending only on the footpoint:

$$(\Phi_p^* \nabla^{tr})_a b|_\mu = \frac{1}{p} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \nabla_a^{(\alpha)} b|_\mu, \quad \mu \in \text{Dens}_+(M), a, b \in \mathfrak{X}(\text{Dens}_+(M))$$

In particular, the geodesics of $\Phi_p^* \nabla^{tr}$ and $\nabla^{(\alpha)}$ coincide.

Note that geodesics of the trivial connection on a vector space are always straight lines; in particular, this proposition allows us to obtain geodesics of the L^p -Fisher-Rao metric (of the α -connection, resp.) by pulling-back straight lines in $C^\infty(M)$ using Φ_p . We will use this in Corollary 3.13 below to explicitly describe the resulting formulas on $\text{Dens}_+(M)$. First we present the proof of the above theorem, which is a fairly straightforward calculation:

Proof of Theorem 3.12. The characterization of the image of Φ_p follows directly from the definition of $\text{Dens}_+(M)$. To show item (b) we calculate for $\mu \in \text{Dens}_+(M)$ and $a \in T_\mu \text{Dens}_+(M)$ the differential of Φ_p :

$$D_\mu \Phi_p(a) = \frac{1}{p} \frac{a}{\lambda} \left(\frac{\mu}{\lambda} \right)^{1/p-1}.$$

Therefore the pullback of the L^p -norm via the embedding Φ_p is given by

$$\|D_\mu \Phi_p(a)\|_{L^p} = \frac{1}{p} \left(\int_M |D_\mu \Phi_p(a)|^p d\lambda \right)^{1/p} = \frac{1}{p} \left(\int_M \left| \frac{a}{\mu} \right|^p \mu \right)^{1/p} = \frac{1}{p} F_p(\mu, a),$$

which implies that the embedding Φ_p is indeed an isometry.

Similarly we calculate for item (c)

$$\begin{aligned} (\Phi_p^* \nabla^{tr})_a b|_\mu &= (T\Phi_p)^{-1} \nabla_{T\Phi_p(a)}^{tr} T\Phi_p(b)|_{\Phi_p(\mu)} \\ &= (T\Phi_p)^{-1} \left(\frac{1}{p^2} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \left(\nabla_{a/\lambda}^{tr} \left(\frac{\nu}{\lambda} \right)^{\frac{1}{p}-1} \frac{b}{\lambda} \right)_{\mu=\nu} \right) \\ &= \frac{1}{p} \left(\nabla_{a/\lambda}^{tr} \left(\frac{\nu}{\lambda} \right)^{\frac{1}{p}-1} \frac{b}{\lambda} \right)_{\mu=\nu} \lambda \\ &= \frac{1}{p} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \left(D(b/dx) \cdot (a/dx) + \left(\frac{1}{p} - 1 \right) \left(\frac{\mu}{\lambda} \right)^{-1} \frac{a}{\lambda} \frac{b}{\lambda} \right) \lambda \\ &= \frac{1}{p} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \left(Db \cdot a + \left(\frac{1}{p} - 1 \right) \frac{a}{\mu} b \right) = \frac{1}{p} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \overline{\nabla}_a^{(\alpha)} b|_\mu. \end{aligned}$$

□

The above theorem allows us to explicitly solve for geodesics on $\text{Dens}_+(M)$, which in turn leads to a proof of metric and geodesic incompleteness of the L^p -Fisher-Rao metric for any $p > 1$. By the equivalence of geodesics for the α -connections and for the L^p -Fisher-Rao metric the formulas for geodesics also hold for the former. In the finite dimensional setting this solution formula (via the p -root mapping) for the α -geodesics is known albeit without any geometric interpretation, cf. [5, Page 50].

Corollary 3.13 (The geometry of the L^p -Fisher-Rao metric). *For any $p > 1$ we have the following statements:*

(a) The space $\text{Dens}_+(M)$ equipped with the L^p -Fisher-Rao metric (the α -connection resp.) is geodesically convex and, even more, there exists an explicit formula for all minimizing

geodesics: given any $\mu_0, \mu_1 \in \text{Dens}_+(M)$ the unique geodesic $\mu : [0, 1] \rightarrow \text{Dens}_+(M)$ connecting μ_0 to μ_1 is given by

$$\mu(t) = \left(t \sqrt[p]{\frac{\mu_1}{\lambda}} + (1-t) \sqrt[p]{\frac{\mu_0}{\lambda}} \right)^p \lambda.$$

(b) *Given any $\mu_0, \mu_1 \in \text{Dens}_+(M)$ the geodesic distance of the L^p -Fisher-Rao metric is given by*

$$d(\mu_0, \mu_1) = \left(\int_M \left| \sqrt[p]{\frac{\mu_1}{\lambda}} - \sqrt[p]{\frac{\mu_0}{\lambda}} \right| \lambda \right)^{1/p}$$

In particular, the geodesic distance of the L^p -Fisher-Rao metric on $\text{Dens}_+(M)$ is non-degenerate.

(c) *For any initial conditions $\mu_0 \in \text{Dens}_+(M)$ and $a \in T_\mu \text{Dens}_+(M)$ the unique L^p -Fisher-Rao geodesic (α -connection geodesic, resp.) $\mu : [0, T) \rightarrow \text{Dens}_+(M)$ defined on its maximal interval of existence $[0, T)$ is given by*

$$\mu(t) = \left(\sqrt[p]{\frac{\mu_0}{\lambda}} + t \frac{a}{\lambda} \left(\frac{\mu}{\lambda} \right)^{1/p-1} \right)^p \lambda.$$

The geodesic $\mu(t)$ exists for all time t , i.e., $T = \infty$, if and only if $\frac{a}{\lambda}(x) \geq 0$ for all $x \in M$. Thus the space $\text{Dens}_+(M)$ equipped with the L^p -Fisher-Rao metric is geodesically incomplete since the solution to the geodesic equation (10) leaves the space in finite time for any initial condition with $\frac{a}{\lambda}(x) < 0$ for some x .

(d) *The space $\text{Dens}_+(M)$ equipped with the geodesic distance of the L^p -Fisher-Rao metric is metrically incomplete.*

(e) *The metric completion of the space $\text{Dens}_+(M)$ with respect to the geodesic distance of the L^p -Fisher-Rao metric is the space of all non-negative L^1 -densities:*

$$\text{Dens}_{L^1}(M) = \left\{ \mu : \frac{\mu}{\lambda} \in L^1(M), \frac{\mu}{\lambda} \geq 0 \text{ a.e.} \right\}$$

Proof. Statements (a)–(d) follow directly from the isometry of Theorem 3.12, the fact that geodesics on the vector space $(C^\infty(M), L^p)$ are straight lines and the characterization of the image of Φ_p as an open, convex subset of $C^\infty(M)$. To see the statement regarding the metric completion we observe that the metric completion of the image is exactly the set of a.e. non-negative L^p -functions and thus the statement on the metric completion follows by applying Φ_p^{-1} . \square

4. THE L^p -FISHER-RAO METRIC AND α -CONNECTIONS ON THE SPACE OF PROBABILITY DENSITIES

The L^p -Fisher-Rao metric F_p and the α -divergence D^α define, via restriction, corresponding objects on $\text{Prob}(M)$, which we study in this section. In particular, we will see that $\text{Prob}(M)$ equipped with the L^p -Fisher-Rao metric corresponds geometrically to an infinite dimensional L^p -sphere. In addition we will see that the equivalence to the α -connection, that has been established for the space of all densities in the previous section, does not hold on the space of probability densities. Consequently we obtain three different notions of p -geodesics on this space:

- (1) geodesics of the restriction of the L^p -Fisher-Rao metric to $\text{Prob}(M)$;
- (2) geodesics of the α -connections on $\text{Prob}(M)$;
- (3) projections of L^p -Fisher-Rao geodesic curves (or equivalently, the α -connection ones) on $\text{Dens}_+(M)$.

In addition, if we allow to leave the space of probability densities, we obtain a fourth notion:

- (4) L^p -Fisher-Rao geodesics in $\text{Dens}_+(M)$. In analogy to the L^2 case, the induced geodesic distance between probability densities defines an L^p version of the Hellinger distance.

We will show that (2) and (3) coincide, thereby providing an explicit formula for α -geodesics on $\text{Prob}(M)$. For a graphic summary of these constructions we refer to Figure 1. In the next section we will compare the remaining three notions of geodesics numerically.

4.1. The Amari-Čencov α -connections on $\text{Prob}(M)$. The restriction of the α -divergences D^α to the space $\text{Prob}(M)$ induces again a family of α -connections, which we will denote by $\overline{\nabla}^{(\alpha)}$. Note, that this connection is not simply the restriction of the α -connections on $\text{Dens}_+(M)$, which is the reason for choosing a different notation for it. We start by deriving an explicit formula for the α -connections on $\text{Prob}(M)$:

Lemma 4.1. *For any $\alpha \in (-1, 1)$ the α -connections $\overline{\nabla}^{(\alpha)}$ on $\text{Prob}(M)$ are given by*

$$(19) \quad \overline{\nabla}_a^{(\alpha)} b = Db.a - \frac{1}{p^*} \left(\frac{a}{\mu} b - \left(\int_M \frac{a}{\mu} \frac{b}{\mu} \mu \right) \mu \right).$$

Thus, the connection $\overline{\nabla}_a^{(\alpha)} b$ on $\text{Prob}(M)$ is the projection of $\nabla_a^{(\alpha)} b$ with respect to the Fisher-Rao metric G^{FR} .

For finite sample spaces this result is well-known (e.g., [5, Section 2.5.2]); in infinite dimensions formula (19) agrees with the formula (22) in [30], under the identification of $\text{Prob}(M) = \text{Diff}(M)/\text{Diff}_\lambda(M)$.

Proof. To derive the formula for the α -connection $\overline{\nabla}^{(\alpha)}$ we calculate the second derivative of the restriction of D^α , which is given again by formula (4) with the only difference being that $a, b, c \in T_\mu \text{Prob}(M)$. Thus we have determined $\overline{\nabla}^{(\alpha)}$ up to a function in the G^{FR} orthogonal complement of $T_\mu \text{Prob}(M)$, which are exactly the constant multiples of μ . Thus the formula follows by ensuring that $\overline{\nabla}_a^{(\alpha)} b \in \text{Prob}(M)$. This argument also proves that $\overline{\nabla}_a^{(\alpha)} b$ is the Fisher-Rao projection of $\nabla_a^{(\alpha)} b$. \square

Theorem 4.2. *A path $\mu : [0, 1] \rightarrow \text{Prob}(M)$ is a geodesic with respect to $\nabla^{(\alpha)}$ if*

$$\mu_{tt} - \frac{1}{p^*} \mu^{-1} \mu_t^2 = -\frac{1}{p^*} \left(\int \left(\frac{\mu_t}{\mu} \right)^2 \mu \right) \mu.$$

For any $k > \dim(M)/2$ the geodesic equations are locally wellposed on the space of Sobolev probability densities $\text{Prob}^k(M)$, i.e., given initial conditions $\mu(0) \in \text{Prob}^k(M)$, $\mu_t(0) \in T_{\mu(0)} \text{Prob}^k(M)$ there exists a unique solution to equation (6) defined on a maximal interval of existence $[0, T)$. The maximal interval of existence is uniform in the Sobolev order k and thus the local wellposedness continues to hold in the limit, i.e., on the space of smooth, probability densities $\text{Prob}(M)$.

Proof. The proof of the local wellposedness follows exactly as in Theorem 3.3. \square

4.2. The L^p -Fisher-Rao metric on $\text{Prob}(M)$. Next, we study the restriction of the L^p -Fisher-Rao metric to the space $\text{Prob}(M)$.

Remark 4.3 (Čencov's theorem). Note that Lemma 3.6 on the invariance of the L^p -Fisher-Rao metric continues to hold on the space $\text{Prob}(M)$. For the Riemannian case and $\dim(M) > 1$ Čencov's theorem states that the Fisher-Rao metric is the only Riemannian metric on $\text{Prob}(M)$ that is invariant under the action of the diffeomorphism group $\text{Diff}(M)$, cf. [16, 4, 9]. In the Finslerian case there is a significant amount of additional flexibility, and one can indeed construct metrics beyond the L^p -Fisher-Rao metric that satisfy this property. In future work it would be interesting to obtain a complete characterization of all such Finsler metrics.

We start by computing the geodesic equation of the (restriction) of the L^p -Fisher-Rao metric F_p on $\text{Prob}(M)$:

Theorem 4.4 (Geodesic equation on $\text{Prob}(M)$). *For any $p \in (1, \infty)$, the geodesic equation of the L^p -Fisher-Rao metric on the space of densities $\text{Prob}(M)$ is given by*

$$(20) \quad \left| \frac{\mu_t}{\mu} \right|^{p-2} \frac{d}{dt} \left(\frac{\mu_t}{\mu} \right) + \frac{1}{p} \left| \frac{\mu_t}{\mu} \right|^p = C(t)$$

where $C(t)$ is a constant depending only on time t , that is chosen such that $\int_M \mu(t) = 1$.

This equation coincides with the geodesic equation of the α -connection if and only if $p = 2$ ($\alpha = 0$, resp.).

Remark 4.5 (Existence of solutions). In the previous section we showed that the geodesic equation of the α -connections is locally wellposed on the space $\text{Prob}(M)$. One would be tempted to expect a similar result for the geodesic equation of the L^p -Fisher-Rao metric; recall that this statement was true on the space $\text{Dens}_+(M)$. It turns out that the above equation is analytically much worse-behaved: the problem arises from the vanishing of the quantity $\frac{\mu_t}{\mu}$ which leads to singularities of the geodesic equation. As a consequence we conjecture that the geodesic equation does not admit any classical solutions. This behavior can also be observed in the numerical simulations (Figure 2), where the obtained (approximate) solutions show a singular behavior.

Proof of Theorem 4.4. To derive this equation, we proceed as for the geodesic equation on the space $\text{Dens}_+(M)$. We then obtain again

$$\delta E_p(\mu)(\delta\mu) = -\frac{1}{p} \int_0^1 \int \left(p \frac{d}{dt} \left(\left| \frac{\mu_t}{\mu} \right|^{p-2} \frac{\mu_t}{\mu} \right) + (p-1) \left| \frac{\mu_t}{\mu} \right|^p \right) \delta\mu \, d\lambda \, dt := -\frac{1}{p} \int_0^1 \int \Psi \, \delta\mu \, d\lambda \, dt$$

for the variation of the p -Energy with the only difference being that $\delta\mu$ now has to integrate to zero. Thus we do not get that $\Psi = 0$ as we had on the space $\text{Dens}_+(M)$, but only that Ψ has to be orthogonal to all such $\delta\mu$. This is equivalent to Ψ being a constant for each fixed time t , which is determined by the condition that $\int_M \mu(t) = 1$. \square

The above result suggests that the equivalence between the α -connection and the Chern-connection of the L^p -Fisher-Rao metric cannot hold in this setting. We will make this formal in the following theorem:

Theorem 4.6 (The Chern connection on $\text{Prob}(M)$). *For a vector field ν on $\text{Prob}(M)$ the Chern connection is given by, for all $a \in T_\mu \text{Prob}(M)$,*

$$(21) \quad \bar{\nabla}_a^\nu \nu = D\nu \cdot a - \frac{p-1}{p} \frac{a \nu}{\mu \mu} \mu + k_1 \left| \frac{\nu}{\mu} \right|^{-p} \frac{a \nu}{\mu} + k_2 \left| \frac{\nu}{\mu} \right|^{2-p} \mu,$$

with the constants

$$k_1(\nu) := -\frac{(p-1)(p-2)}{2p} \left(\int \left| \frac{\nu}{\mu} \right|^2 \mu \right) / \left(\int \left| \frac{\nu}{\mu} \right|^{2-p} \mu \right)$$

$$k_2(\nu, a) := \frac{p-1}{p} \left(\int \frac{a \nu}{\mu \mu} \mu \right) / \left(\int \left| \frac{\nu}{\mu} \right|^{2-p} \mu \right) + \frac{(p-1)(p-2)}{2p} \left(\int \left| \frac{\nu}{\mu} \right|^{-p} \frac{a \nu}{\mu} \int \left| \frac{\nu}{\mu} \right|^2 \mu \right) / \left(\int \left| \frac{\nu}{\mu} \right|^{2-p} \mu \right)^2.$$

Remark 4.7. As any vector field $\nu \in T_\mu \text{Prob}(M)$ has zeros the above formula has to be taken with caution and should be understood formally only.

Remark 4.8. In particular, when all entries are the same, the Chern connection on $\text{Prob}(M)$ is the orthogonal projection of the α -connection $\nabla^{(\alpha)}$ on $\text{Dens}_+(M)$, for $\alpha = 1 - \frac{2}{p}$, with respect to g^ν , the Riemannian metric (12) induced by the L^p -Fisher-Rao metric

$$(22) \quad \bar{\nabla}_\nu^\nu \nu = \text{Proj}^\nu \left(\nabla_\nu^{(\alpha)} \nu \right) = D\nu \cdot \nu - \frac{1}{p^*} \frac{\nu \nu}{\mu \mu} \mu + k \left| \frac{\nu}{\mu} \right|^{2-p} \mu$$

where p^* is the Hölder conjugate of p and

$$k(\nu) := \frac{1}{p^*} \left(\int \left| \frac{\nu}{\mu} \right|^2 \mu \right) / \left(\int \left| \frac{\nu}{\mu} \right|^{2-p} \mu \right).$$

Indeed, the correction term $k \left| \frac{\nu}{\mu} \right|^{2-p} \mu$ is orthogonal to $T \text{Prob}(M)$ and makes the integral zero.

Proof of Theorem 4.6. We start by noticing that, since $D\nu(a)$ integrates to zero, the integral of the right hand-side of (21) is zero and so it defines a tangent vector of $\text{Prob}(M)$. The formula (21) defines the Chern connection if and only if it verifies the generalized Koszul formula (17). Letting $\alpha = 1 - \frac{2}{p}$ and $\nabla^{(\alpha)}$ be the corresponding α -connection on $\text{Dens}_+(M)$, we can decompose the candidate for the Chern connection as

$$\bar{\nabla}_a^\nu \nu = \nabla_a^{(\alpha)} \nu + k_1 \left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \nu + k_2 \left| \frac{\nu}{\mu} \right|^{2-p} \mu.$$

Since $\nabla^{(\alpha)}$ is the Chern connection on $\text{Dens}_+(M)$ for this choice of α , the candidate (21) verifies the generalized Koszul formula if and only if

(23)

$$\begin{aligned} 0 &= 2g^\nu \left(k_1(\nu) \left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \nu + k_2(\nu, a) \left| \frac{\nu}{\mu} \right|^{2-p} \mu, b \right) + 2C^\nu \left(k_1(\nu) \left| \frac{\nu}{\mu} \right|^{-p} \frac{\nu}{\mu} + k_2(\nu, \nu) \left| \frac{\nu}{\mu} \right|^{2-p} \mu, b, a \right) \\ &= 2k_1(\nu) g^\nu \left(\left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \nu, b \right) + 2k_2(\nu, a) g^\nu \left(\left| \frac{\nu}{\mu} \right|^{2-p} \mu, b \right) + 2(k_1(\nu) + k_2(\nu, \nu)) C^\nu \left(\left| \frac{\nu}{\mu} \right|^{2-p} \mu, b, a \right). \end{aligned}$$

Noticing that, for all $b \in T_\mu \text{Prob}(M)$, $I(|\frac{\nu}{\mu}|^{2-p} \mu, b) = \int b = 0$, we see from (12) that

$$g^\nu \left(\left| \frac{\nu}{\mu} \right|^{2-p} \mu, b \right) = 0.$$

This also means that all terms in the Cartan tensor (14) but one vanish, leaving

$$\begin{aligned} 2C^\nu \left(\left| \frac{\nu}{\mu} \right|^{2-p} \mu, b, a \right) &= (p-1)(p-2) I(\nu, \nu)^{\frac{2}{p}-1} J \left(\nu, \left| \frac{\nu}{\mu} \right|^{2-p} \mu, b, a \right) \\ &= (p-1)(p-2) I(\nu, \nu)^{\frac{2}{p}-1} \int \left(\frac{\nu}{\mu} \right)^{-1} \frac{a}{\mu} \frac{b}{\mu} \mu \end{aligned}$$

Finally there remains to compute

$$\begin{aligned} g^\nu \left(\left| \frac{\mu}{\nu} \right|^p \frac{a}{\mu} \nu, b \right) &= (p-1) I(\nu, \nu)^{\frac{2}{p}-1} \int \left| \frac{\mu}{\nu} \right|^{p-2} \left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \frac{\nu}{\mu} \frac{b}{\mu} \mu \\ &\quad - (p-2) I(\nu, \nu)^{\frac{2}{p}-2} I(\nu, b) \int \left| \frac{\nu}{\mu} \right|^{p-2} \frac{\nu}{\mu} \left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \frac{\nu}{\mu} \mu \\ &= (p-1) I(\nu, \nu)^{\frac{2}{p}-1} \int \left(\frac{\nu}{\mu} \right)^{-1} \frac{a}{\mu} \frac{b}{\mu} \mu. \end{aligned}$$

Putting all these together, and noticing that $k_2(\nu, \nu) = -\frac{p}{p-2}k_1(\nu)$, we obtain

$$\begin{aligned} & 2g^\nu \left(k_1(\nu) \left| \frac{\nu}{\mu} \right|^{-p} \frac{a}{\mu} \nu + k_2(\nu, a) \left| \frac{\nu}{\mu} \right|^{2-p} \mu, b \right) + 2C^\nu \left(k_1(\nu) \left| \frac{\nu}{\mu} \right|^{-p} \frac{\nu}{\mu} + k_2(\nu, \nu) \left| \frac{\nu}{\mu} \right|^{2-p} \mu, b, a \right) \\ &= (pk_1(\nu) + (p-2)k_2(\nu))(p-1)I(\nu, \nu)^{\frac{2}{p}-1} \int \left(\frac{\nu}{\mu} \right)^{-1} \frac{a}{\mu} \frac{b}{\mu} \mu \\ &= 0, \end{aligned}$$

and so condition (23) is satisfied. \square

4.3. The p -root transform on $\text{Prob}(M)$. In the previous section we have seen that the α -connection and the L^p -Fisher-Rao metric induce different geodesics on the space $\text{Prob}(M)$. In this section we will investigate the geometric reasons behind this, by connecting both of these objects to the p -root transform. In order to state this result we will need to define an appropriate connection on the sphere

$$S_p := \{f \in C^\infty(M) : \|f\|_{L^p} = 1\},$$

as the image of $\text{Prob}(M)$ under Φ_p is in this set. To this end, we define:

Definition 4.9 (p -projection and p -connection). The p -projection map $\pi^p : TC^\infty|_{S_p} \rightarrow TS_p$ is defined by

$$\pi_f^p(\xi) = \xi - \left(\int_M \xi f |f|^{p-2} d\lambda \right) f, \quad f \in S_p, \xi \in TC^\infty|_{S_p}.$$

The induced p -connection on S_p is defined by

$$\nabla_\xi^p \eta = \pi^p(\nabla_\xi^{\text{tr}} \eta).$$

Note, that π^p is the projection with respect to the splitting $T_f C^\infty = T_f S_p \oplus \text{span}\{f\}$. The geodesic equation $\nabla_\gamma^p \dot{\gamma} = 0$ can therefore be written as:

$$(24) \quad \begin{cases} \dot{\gamma} \parallel \gamma \\ \int_M \gamma^p d\lambda = 1 \end{cases}$$

Note that from a metric point of view, this splitting is natural since $f \in T_f C^\infty$ is the unique direction from which straight lines (i.e., geodesics in C^∞) emanating from f gets the fastest away from S_p with respect to the L^p norm (since for $p \in (1, \infty)$ the space L^p is strictly convex). Similarly, $\pi_f^p(\xi)$ satisfies $\|\xi - \pi_f^p(\xi)\|_{L^p} = \text{dist}_{L^p}(\xi, T_f S_p)$. For a more general viewpoint on projections on a sphere in uniformly convex Banach spaces whose dual is also uniformly convex, see [23] and [22, Prop. 2].

We are now able to formulate the analogous statement of Theorem 3.12, which will demonstrate the geometric differences between the α -connections and the L^p -Fisher-Rao metric:

Theorem 4.10. Let $\alpha \in (-1, 1)$ and, as before, denote $p = \frac{2}{1-\alpha}$. Consider the restriction of the map Φ_p , as defined in (18), to the space $\text{Prob}(M)$. We have:

- (a) The image $\Phi_p(\text{Prob}(M))$ is the set of all positive functions in the L^p -sphere S_p .
- (b) The mapping Φ_p is an isometric embedding, where $\text{Prob}(M)$ is equipped with a multiple of the L^p -Fisher-Rao metric and where S_p is equipped with the restriction of the standard L^p -norm.
- (c) The pullback of $\Phi_p^* \nabla^p$ to $\text{Prob}(M)$ coincides with the connection $\overline{\nabla}^{(\alpha)}$ up to a constant depending only on the footpoint:

$$(\Phi_p^* \nabla^p)_a b|_\mu = \frac{1}{p} \left(\frac{\mu}{\lambda} \right)^{\frac{1}{p}-1} \overline{\nabla}_a^{(\alpha)} b|_\mu, \quad \mu \in \text{Prob}(M), a, b \in \mathfrak{X}(\text{Prob}(M)).$$

In particular, the geodesics of $\Phi_p^* \nabla^p$ and $\overline{\nabla}^{(\alpha)}$ coincide.

Proof. The proof follows by the same calculation as the proof of Theorem 3.12. \square

On S_p , geodesics are no longer straight lines, and we do not have an explicit solution for the geodesic equations of either the α -connection or the L^p -Fisher-Rao metric. However, by projecting straight lines on the sphere and rescaling time, one can obtain geodesics for the α -connection (cf. [4, Section 2.5.2] where this result has been shown in the finite dimensional situation):

Theorem 4.11. *Let $f \in S_p$ and $\xi \in T_f S_p$. Let $I \subset \mathbb{R}$ be an interval containing 0, and let $\tau : I \rightarrow \mathbb{R}$ satisfy the ODE*

$$\begin{aligned}\ddot{\tau}(t) &= 2 \frac{\int_M |f + \tau(t)\xi|^{p-2} (f + \tau(t)\xi)\xi \, d\lambda}{\int_M |f + \tau(t)\xi|^p \, d\lambda} \dot{\tau}(t)^2 \\ \tau(0) &= 0 \\ \dot{\tau}(0) &= 1\end{aligned}$$

Then $\gamma : I \rightarrow S_p$ defined by

$$\gamma(t) = \frac{f + \tau(t)\xi}{\|f + \tau(t)\xi\|_{L^p}}$$

is a geodesic of ∇^p , with initial condition $\gamma(0) = f$, $\dot{\gamma}(0) = \xi$.

A boundary value problem between $f, g \in S_p$ can be similarly addressed by putting $\xi = g - f$ and $I = [0, 1]$, and replacing the initial conditions for τ by the boundary conditions $\tau(0) = 0$, $\tau(1) = 1$.

Geodesics of $\bar{\nabla}^{(\alpha)}$ are obtained by pulling back these geodesics using Φ_p . They all cease to exist (i.e., leave the space $\text{Prob}(M)$) after finite time. Since the geodesic equation is locally well-posed (Proposition 4.2), this procedure induces all the α -connection geodesics, i.e., the exponential map of $\bar{\nabla}^{(\alpha)}$.

Proof. Using (24), we need to show that $\ddot{\gamma} \parallel \gamma$; all the other assumptions are satisfied by construction. We have

$$\ddot{\gamma} = \ddot{\tau}(t) \|f + \tau(t)\xi\|_{L^p}^{-1} \xi + 2\dot{\tau}(t) \frac{d}{dt} \|f + \tau(t)\xi\|_{L^p}^{-1} \xi + \frac{d^2}{dt^2} \|f + \tau(t)\xi\|_{L^p}^{-1} (f + \tau(t)\xi).$$

The last addend is clearly parallel to γ . Hence it is sufficient to require that

$$\ddot{\tau}(t) \|f + \tau(t)\xi\|_{L^p}^{-1} + 2\dot{\tau}(t) \frac{d}{dt} \|f + \tau(t)\xi\|_{L^p}^{-1} = 0,$$

which is equivalent to the wanted ODE.

In order to prove that the pullback of the solutions leaves $\text{Prob}(M)$ after a finite time, we need to show that $\gamma(t)$ stops being positive, i.e., that for some $t > 0$, $f(x) + \tau(t)\xi(x) \leq 0$ for some $x \in M$. From the equivariance under the action of $\text{Diff}(M)$, cf. Remark 4.3, it is sufficient to consider the case $f \equiv 1$ (which corresponds to $\mu(0) = \lambda$). In this case ξ is a non-zero function satisfying $\int_M \xi \, \lambda = 0$, and thus in particular $\xi(x) < 0$ for some x . Therefore, in order to prove that $1 + \tau(t)\xi(x) \leq 0$ for some t , it is sufficient to prove that τ is unbounded as $t \rightarrow \infty$. Note that we can write the equation for τ as

$$(25) \quad \ddot{\tau}(t) = 2 \left(1 - \frac{1}{\int_M |1 + \tau(t)\xi|^p \, d\lambda} \right) \frac{\dot{\tau}(t)^2}{\tau}.$$

Now, since $s \mapsto 1 + s\xi$ is a tangent line to the unit sphere at $f = 1$ in the strictly convex space L^p , it follows that $\|1 + s\xi\|_{L^p} \geq 1$, and equality holds if and only if $s = 0$. Thus, the term in the parentheses in (25) is non-negative, and vanishes if and only if $\tau(t) = 0$. Since we also have that $\tau(0) = 0$ and $\dot{\tau}(0) = 1$, it follows that $\ddot{\tau}(t) > 0$ for $t \in (0, t_0)$ for some t_0 small enough, and thus for any positive t . It follows therefore that $\tau > t$ for all $t > 0$, and in particular, it is unbounded. \square

Remark 4.12. In fact, the estimate $\tau > t$ implies that $1 + \tau(t)\xi$ hits zero at some point for the first time at $t_* < \frac{1}{-\min \xi}$. Pulling back to $\text{Prob}(M)$, we obtain that a geodesic from λ with initial condition $a \in T_\lambda \text{Prob}(M)$ blows up at time

$$(26) \quad t_* < \frac{p}{-\min(a/\lambda)}.$$

In principle, better estimates on the blowup can be obtained by more careful analysis of (25). The estimate (26) is exactly the estimate obtained in [29, Formula (78)] (there, the parameter a is equivalent to $-1 - \frac{2}{p}$ in our notation).

Example 4.13 (Fisher-Rao geodesics). For the case $p = 2$, assuming that ξ is a unit vector (which is, by definition, perpendicular to f), we obtain that the ODE takes a simpler form

$$\ddot{\tau} = \frac{2\tau}{1 + \tau^2} \dot{\tau}^2,$$

whose solution is $\tau(t) = \tan t$, yielding the known solution of the Fisher–Rao geodesics [26, Remark 4.4].

5. SUMMARY OF RELATIONS TO KNOWN PDES AND METRICS ON DIFFEOMORPHISM GROUPS

We now summarize how the L^p -Fisher-Rao metric relates to (degenerate) right-invariant Finsler metric on the group of diffeomorphisms, in a similar spirit as in [26] who studied this for the L^2 -case. Furthermore, we will see how the geodesics equations described in this paper relate to other previously studied equations in hydrodynamics and mathematical physics:

- On the diffeomorphism group of a closed manifold M one can consider the family of, right-invariant (degenerate) $W^{1,p}$ -Finsler metrics of the form

$$\tilde{F}_p(\varphi, X \circ \varphi) = \left(\int_M |\text{div}(X)|^p d\lambda \right)^{1/p}, \quad X \in T_\varphi \text{Diff}(M).$$

These metrics were useful for proving that the diameter of $\text{Diff}(M)$ with respect to some critical Sobolev Riemannian metrics is infinite [12]. Note that the kernel of the Finsler metric \tilde{F}_p consists exactly of all divergence free vector fields, and thus \tilde{F}_p is only a “true” Finsler metric on the quotient space $\text{Diff}(M)/\text{Diff}_\lambda(M)$. The relation to the L^p -Fisher-Rao metric, as studied in the present article, becomes clear by considering the mapping $\varphi \mapsto \text{Jac}(\varphi)\lambda$, which gives rise to an isometry

$$(\text{Diff}(M)/\text{Diff}_\lambda(M), \tilde{F}_p) \rightarrow (\text{Prob}(M), F_p).$$

Note, that this result is a direct generalization of the case $p = 2$ treated in Khesin et al. [26]. For this case Modin [33] constructed an extension of the metric \tilde{F}_2 to obtain a non-degenerate, right invariant Riemannian metric on the full group of diffeomorphisms $\text{Diff}(M)$, that still descends to the Fisher-Rao metric F_2 on $\text{Prob}(M)$. In future work it would be interesting to consider a similar extension for the case $p \neq 2$.

- Similarly, the α -connections on $\text{Prob}(M)$ can be pulled back to $\text{Diff}(M)/\text{Diff}_\lambda(M)$; the corresponding geodesic equation (which is equivalent to the one in Theorem 4.2) was first considered in [30]. Theorem 4.11 shows their integrability and finite-time blowup.
- For the special case $M = S^1$, where the group of volume preserving diffeomorphisms is given by the group of rotations $\text{Rot}(S^1)$, the α -connections on $\text{Prob}(S^1)$ can thus be pulled back to $\text{Diff}(S^1)/\text{Rot}(S^1)$, where the associated geodesic equation, when presented on the Lie algebra, is the generalized periodic inviscid Proudman–Johnson equation

$$u_{txx} + (2 - \alpha)u_x u_{xx} + uu_{xxx} = 0,$$

as was first shown in [30]. See [38, 29] and the references therein for analysis of this equation, also beyond the range $\alpha \in (-1, 1)$.

- Similarly, the L^p -Fisher–Rao metric on $\text{Prob}(S^1)$ can be considered as a Finsler metric on $\text{Diff}(S^1)/\text{Rot}(S^1)$. The resulting geodesic equation is the periodic r -Hunter–Saxton equation for $r = 1/p$, as considered in [18, 11]. As shown in this paper, this is not the same equation as the one of the α -connections on $\text{Prob}(S^1)$ (i.e., the generalized periodic invicid Proudman–Johnson equation), unlike what we erroneously stated in [11].
- For $M = \mathbb{R}$, the geodesic equations of α -connections (equiv., of the L^p -Fisher–Rao metric) on $\text{Dens}(\mathbb{R})$ can be considered as equations of an appropriate subgroup of $\text{Diff}(\mathbb{R})$, defined in [11]. The resulting equation is the generalized non-periodic invicid Proudman–Johnson equation, or equivalently, the non-periodic r -Hunter–Saxton equation (for $r = 1/p$) [18]. Moreover, the metric \tilde{F}_p described above on this subgroup of $\text{Diff}(\mathbb{R})$ yields a similar isometry to $(\text{Dens}_+(M)(\mathbb{R}), F_p)$, as follows from [11]. It is interesting whether $(\text{Dens}_+(M)(\mathbb{R}), F_p)$ can be similarly interpreted on compact manifolds as well, maybe in a similar way to the ”simple unbalanced optimal transport” extension, introduced recently in [28].

6. A NUMERICAL COMPARISON OF GEODESICS ON $\text{Dens}_+(M)$ AND $\text{Prob}(M)$

In this section we aim to numerically compare the different notions of geodesics that we have encountered in this article. Given two probability densities we consider three notions of geodesics:

- (1) The geodesic for the L^p -Fisher–Rao metric and the α -connection on $\text{Dens}_+(M)$, which is simply obtained as the pullback by the p -root transform Φ_p of the straight line in L^p . This geodesic leaves the space $\text{Prob}(M)$.
- (2) The geodesic for the α -connection on $\text{Prob}(M)$, which is the pullback by the p -root transform of the projection of the straight line on the L^p sphere, as described in Theorem 4.11.
- (3) The geodesic for the L^p -Fisher–Rao geodesic on $\text{Prob}(M)$, which is the pullback by the p -root transform of the geodesic of the L^p -metric restricted to the L^p -sphere.

Specifically we consider the example of probability densities on the one-dimensional base space $M = [0, 1]$. Note, that we have an explicit formula for the first two notions of geodesics (geodesics on $\text{Dens}_+(M)$ and α -connection geodesics on $\text{Prob}(M)$), but that the calculation of the L^p -Fisher–Rao geodesic between two probability distributions μ_0 and μ_1 requires us to solve an optimization problem: the geodesic boundary value problem on the L^p -sphere. Namely, we minimize the p -energy for the L^p metric on smooth functions

$$(27) \quad E_p(f) = \frac{1}{p} \int_0^1 \int |f_t|^p \, d\lambda \, dt,$$

where $f : [0, 1] \rightarrow C^\infty(M)$ is a path constrained to belong to the L^p -sphere, such that $f(0) = \Phi_p(\mu_0)$, $f(1) = \Phi_p(\mu_1)$ and f_t denotes its time derivative. This is equivalent to minimizing the length functional, as explained in the proof of Theorem 3.7. We then obtain the wanted geodesic $\mu : [0, 1] \rightarrow \text{Prob}(M)$ by applying Φ_p^{-1} .

In Figure 2 we show the three types of geodesics obtained for different values of p ($p = 2, 3, 5, 10$ from top to bottom), and the corresponding values of $\alpha = 1 - 2/p$. The constrained minimization of (27) was performed in Python using the Sequential Least Squares Programming (SLSQP) method provided by the Scipy minimization solver, with a discretization of 30 time points and 100 sampling points, in a straightforward implementation that was not aimed for computational efficiency. As expected, the L^p -Fisher–Rao metric and the α -connection yield different geodesics on $\text{Prob}(M)$, except for the special case $p = 2$ corresponding to the Fisher–Rao metric and its Levi-Civita connection.

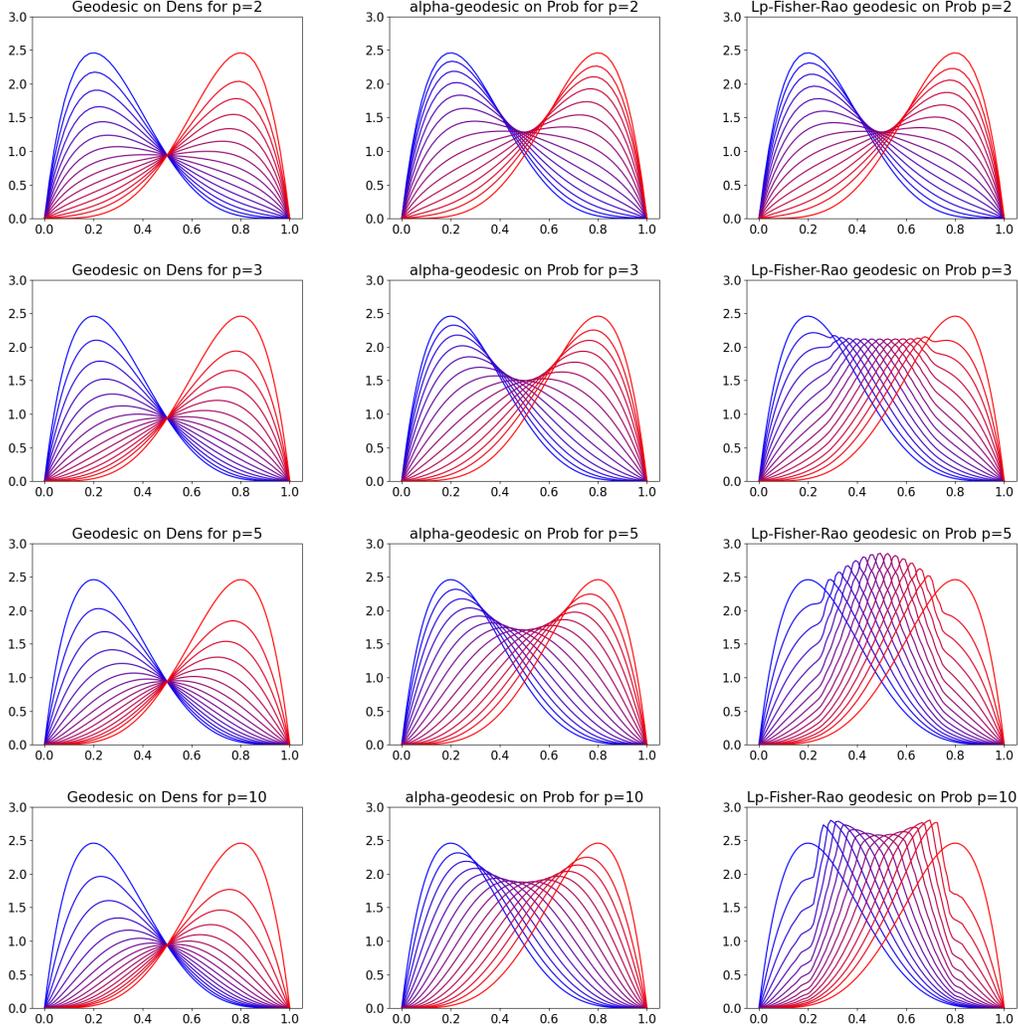


FIGURE 2. Different notions of geodesics between two probability distributions on $[0, 1]$, for $p = 2, 3, 5, 10$ from top to bottom, and corresponding values of $\alpha = 1 - 2/p$. On the left: geodesics of $\text{Dens}_+(M)$ for the L^p -Fisher-Rao metric and the corresponding α -connection. In the middle: α -geodesics on $\text{Prob}(M)$. On the right: L^p -Fisher-Rao geodesics on $\text{Prob}(M)$. The last two notions coincide only for $p = 2$.

7. FINITE-DIMENSIONAL GEOMETRY OF PARAMETRIC STATISTICAL MODELS

In this section we make the link with the finite-dimensional setting of parametric statistical models. Let us consider a finite-dimensional submanifold of $\text{Prob}(\mathbb{R}^n)$ corresponding to a family of probability distributions on \mathbb{R}^n that are absolutely continuous with respect to the Lebesgue measure, and whose densities are parametrized by a parameter θ belonging to an open subset Θ of \mathbb{R}^d :

$$\mathcal{P}_\Theta = \{\mu(dx) = f(x, \theta) dx : \theta \in \Theta\} \subset \text{Prob}(\mathbb{R}^n).$$

Here $x \in \mathbb{R}^n$ is the sample variable and dx denotes the Lebesgue measure on \mathbb{R}^n . Then a tangent vector of \mathcal{P}_Θ at a given $\mu = f(\cdot, \theta)dx$ is of the form $a = \frac{d}{dt}\big|_{t=0} \mu_t$, where $\mu_t = f(\cdot, \theta_t)dx$ with $t \mapsto \theta_t$

a curve in Θ such that $\theta_0 = \theta$ and $\dot{\theta}_0 = u \in T_\theta\Theta$. Thus the tangent space at μ is

$$(28) \quad \begin{aligned} T_\mu\mathcal{P}_\Theta &= \{a = \langle \nabla_\theta f, u \rangle dx : u \in T_\theta\Theta \simeq \mathbb{R}^d\} \\ &= \text{span}\{e_1, \dots, e_d\}, \end{aligned}$$

where $e_i = \frac{\partial f}{\partial \theta^i} dx$. Here ∇_θ denotes the gradient with respect to θ and $\langle \cdot, \cdot \rangle$ the Euclidean scalar product on \mathbb{R}^d . In all the sequel, we identify $\mathcal{P}_\Theta \simeq \Theta$ and $T_\mu\mathcal{P}_\Theta \simeq T_\theta\Theta \simeq \mathbb{R}^d$ via the one-to-one maps

$$(29) \quad \begin{aligned} \phi : \Theta &\rightarrow \mathcal{P}_\Theta, \quad \theta \mapsto f(\cdot, \theta) dx, \\ (\phi_*)_\theta : T_\theta\Theta &\rightarrow T_{\phi(\theta)}\mathcal{P}_\Theta, \quad u \mapsto \langle \nabla_\theta f, u \rangle dx. \end{aligned}$$

7.1. The Fisher-Rao metric and the α -connection. The Fisher-Rao metric on the parameter space Θ is the Riemannian metric whose metric matrix is the Fisher information matrix

$$G(\theta) = \mathbb{E} \left[\nabla_\theta \ell(X, \theta) \nabla_\theta \ell(X, \theta)^\top \right].$$

Here \mathbb{E} denotes the expectation taken with respect to the random variable X of density $f(\cdot, \theta)$, and $\ell(x, \theta) = \log f(x, \theta)$ is the log-likelihood.

Definition 7.1. Given $\theta \in \Theta$ and $u, v \in T_\theta\Theta \simeq \mathbb{R}^d$, the Fisher-Rao metric is

$$G_\theta^{\text{FR}}(u, v) = u^\top G(\theta) v = \mathbb{E} [\langle \nabla_\theta \ell, u \rangle \langle \nabla_\theta \ell, v \rangle],$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product on \mathbb{R}^d .

The Fisher-Rao metric on the parameter space Θ is the pullback of the Fisher-Rao metric on the infinite-dimensional space $\text{Prob}(\mathbb{R}^n)$ by the bijection ϕ defined by (29), i.e. for any $\theta \in \Theta$ and $u, v \in T_\theta\Theta$,

$$G_{\phi(\theta)}^{\text{FR}}(\phi_*u, \phi_*v) = G_\theta^{\text{FR}}(u, v),$$

and so both are denoted the same way.

Just like in the infinite-dimensional setting, the α -connection on the parameter space can be defined using the α -divergence.

Definition 7.2. The α -connection on Θ is defined by its Christoffel symbols of the first kind ([39], Eqn 2.9)

$$\tilde{\Gamma}_{ij,k}^{(\alpha)} \Big|_\theta = G_\theta^{\text{FR}}(\nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_k) = - \frac{\partial^3}{\partial u^i \partial u^j \partial v^k} D^{(\alpha)}(\theta + u, \theta + v) \Big|_{u=v=0}$$

where

$$D^{(\alpha)}(\theta, \theta') = \frac{4}{1 - \alpha^2} \left(1 - \int f(x, \theta)^{\frac{1-\alpha}{2}} f(x, \theta')^{\frac{1+\alpha}{2}} dx \right)$$

is the α -divergence. This yields the following formula in local coordinates, where ∂_i denotes $\partial/\partial\theta^i$,

$$(30) \quad \tilde{\Gamma}_{ij,k}^{(\alpha)} = \mathbb{E} \left[\left(\partial_i \partial_j \ell + \frac{1-\alpha}{2} \partial_i \ell \partial_j \ell \right) \partial_k \ell \right].$$

The following result is well-known in the literature, and stated e.g. in [2] for spaces of probability distributions on a finite set.

Theorem 7.3 (α -connection on Θ). *For any $u, v \in T_\theta\Theta$, we have*

$$\tilde{\nabla}_u^{(\alpha)} v = \text{Proj}^{\text{FR}} \left(\bar{\nabla}_{\phi_*u}^{(\alpha)} \phi_*v \right),$$

where $\tilde{\nabla}^{(\alpha)}$ and $\bar{\nabla}^{(\alpha)}$ denote the α -connections on $\mathcal{P}_\Theta \simeq \Theta$ and $\text{Prob}(\mathbb{R}^n)$ respectively, and $\text{Proj}^{\text{FR}} : T_{\phi(\theta)}\text{Prob}(\mathbb{R}^n) \rightarrow T_\theta\Theta$ is the orthogonal projection with respect to the Fisher-Rao metric.

Proof. First notice that at any $\mu = \phi(\theta) \in \mathcal{P}_\Theta$, the orthogonal projection of a tangent vector $a \in T_\mu \text{Prob}(\mathbb{R}^n)$ onto $T_\theta \Theta$ with respect to the Fisher-Rao metric G^{FR} is given by

$$(31) \quad \text{Proj}^{\text{FR}}(a) = G^{ij} G^{\text{FR}}(a, e_j) \partial_i,$$

where $(G^{ij})_{ij}$ is the inverse of the Fisher matrix. Indeed, the tangent space $T_\mu \mathcal{P}_\Theta$ is a d -dimensional vector space spanned by the tangent vectors $e_i = \partial_i f dx$ for $i = 1, \dots, d$, and so the orthogonal projection of $a \in T_\mu \text{Prob}(\mathbb{R}^n)$ onto $T_\mu \mathcal{P}_\Theta$ is given by $u^i e_i$ where for $j = 1, \dots, d$,

$$G^{\text{FR}}(a - u^i e_i, e_j) = 0 \quad \text{i.e.} \quad G^{\text{FR}}(a, e_j) = u^i G^{\text{FR}}(e_i, e_j) = u^i G_{ij}.$$

The α -connection on $\text{Prob}(\mathbb{R}^n)$ is given by

$$(\bar{\nabla}_a^{(\alpha)} b)_\mu = D_\mu b(a) - \frac{1 + \alpha}{2} \left(\frac{a}{\mu} \frac{b}{\mu} \mu - \left(\int \frac{a}{\mu} \frac{b}{\mu} \mu \right) \mu \right),$$

where $D_\mu b(a)$ is the directional derivative of the vector field b in the direction of the vector a_μ . Let ∂_i denote partial derivative with respect to θ^i for all $i = 1, \dots, d$. For vector fields on the finite-dimensional manifold \mathcal{P}_Θ ,

$$a = \phi_* u = u^i e_i, \quad b = \phi_* v = v^j e_j,$$

and at $\mu = \phi(\theta)$, we get since $e_i = \partial_i f dx$,

$$\begin{aligned} D_\mu b(a) &= D_\theta v(u) = \partial_i (\partial_j f v^j) u^i dx = (\partial_i \partial_j f u^i v^j + \partial_j f u^i \partial_i v^j) dx \\ &= ((\partial_i \partial_j \ell + \partial_i \ell \partial_j \ell) u^i v^j + \partial_j \ell u^i \partial_i v^j) f dx, \end{aligned}$$

where in the last equality we used the equality $\partial_i \partial_j \ell = \partial_i \partial_j f / f - \partial_i \ell \partial_j \ell$. Since

$$\frac{a}{\mu} \frac{b}{\mu} \mu = \partial_i \ell \partial_j \ell u^i v^j f dx,$$

we obtain $\bar{\nabla}_a^{(\alpha)} b = h dx$ where

$$h/f = \partial_j \ell u^i \partial_i v^j + \left(\partial_i \partial_j \ell + \frac{1 - \alpha}{2} \partial_i \ell \partial_j \ell \right) u^i v^j - \frac{1 + \alpha}{2} G^{\text{FR}}(a, b).$$

Remembering that $G^{\text{FR}}(h dx, k dx) = \mathbb{E}(hk/f^2)$ and since $\mathbb{E}(\partial_m \ell) = 0$, we obtain using (30),

$$\begin{aligned} G^{\text{FR}}(\bar{\nabla}_a^{(\alpha)} b, e_m) &= \mathbb{E}[\partial_j \ell \partial_m \ell] u^i \partial_i v^j + \mathbb{E} \left[\left(\partial_i \partial_j \ell + \frac{1 - \alpha}{2} \partial_i \ell \partial_j \ell \right) \partial_m \ell \right] u^i v^j \\ &= G_{jm} u^i \partial_i v^j + \tilde{\Gamma}_{ij,m}^{(\alpha)} u^i v^j. \end{aligned}$$

Finally, using (31), we see that

$$\text{Proj}^{\text{FR}}(\bar{\nabla}_a^{(\alpha)} b) = G^{km} G^{\text{FR}}(\bar{\nabla}_a^{(\alpha)} b, e_m) \partial_k = \left(u^i \partial_i v^k + G^{km} \tilde{\Gamma}_{ij,m}^{(\alpha)} u^i v^j \right) \partial_k = \tilde{\nabla}_u^{(\alpha)} v,$$

which concludes the proof. \square

7.2. The L^p -Fisher-Rao metric. We now introduce a finite-dimensional version of the Finsler L^p -Fisher-Rao metric.

Definition 7.4. Given $\theta \in \Theta$ and $v \in T_\theta \Theta$ we define the L^p -Fisher-Rao metric on Θ as

$$(32) \quad F_p(\theta, v) = (\mathbb{E} |\langle \nabla_\theta \ell(X, \theta), v \rangle|^p)^{1/p}.$$

Here $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product on \mathbb{R}^d , \mathbb{E} denotes the expectation taken with respect to the random variable X of density $f(\cdot, \theta)$, and $\ell(x, \theta) = \log f(x, \theta)$ is the log-likelihood.

The metric (32) on the parameter space Θ coincides with the Finsler metric induced on \mathcal{P}_Θ by the L^p -Fisher-Rao metric (8) through the identification $\mathcal{P}_\Theta \simeq \Theta$, which is why they are denoted the same way. Indeed, for any $(\theta, v) \in T\Theta$,

$$F_p(\phi(\theta), \phi_*v) = \int \left| \frac{\langle \nabla_\theta f(x, \theta), v \rangle}{f(x, \theta)} \right|^p f(x, \theta) dx = \mathbb{E} |\langle \nabla_\theta \ell(X, \theta), v \rangle|^p = F_p(\theta, v).$$

Lemma 7.5 (Induced Chern connection on Θ). *The Chern connection associated to the L^p -Fisher-Rao metric on Θ is given by*

$$(33) \quad (\tilde{\nabla}_u^v v)^m = u(v^m) + (g^v)^{mk} (g^{\phi_*v}(\omega(u, v), e_k) + C^{\phi_*v}(\omega(v, v), \phi_*u, e_k) - (g^v)^{ij} g^{\phi_*v}(\omega(v, v), e_i) C^{\phi_*v}(\phi_*u, e_j, e_k)),$$

where g and C respectively denote the Riemannian metric (12) and Cartan tensor (14) induced by the L^p -Fisher-Rao metric, $(g^v)_{ij} = g^{\phi_*v}(e_i, e_j)$ and $(e_i)_i$ are the basis vectors (28) of $T_\mu \mathcal{P}_\Theta$ and

$$(34) \quad \omega(u, v) = \omega_{ij} u^i v^j \quad \text{with} \quad \omega_{ij} = \left(\partial_i \partial_j \ell + \frac{1}{p} \partial_i \ell \partial_j \ell \right) f dx.$$

Proof. Let $a = \phi_*u$, $\nu = \phi_*v$, $\alpha = 1 - 2/p$, and $\nabla^{(\alpha)}$ be the α -connection on $\text{Dens}_+(\mathbb{R}^n)$. Similarly to the orthogonal projection with respect to the Fisher-Rao metric (31), the orthogonal projection on $T\Theta$ with respect to g^ν is given by

$$(35) \quad \text{Proj}^\nu(a) = (g^\nu)^{ij} g^\nu(a, e_j) \partial_i.$$

Let us denote

$$(\nabla_a^{(\alpha)} \nu)^\top := \phi_* \text{Proj}^\nu \left(\nabla_a^{(\alpha)} \nu \right) = (g^\nu)^{ij} g^\nu(a, e_j) e_i, \quad (\nabla_a^{(\alpha)} \nu)^\perp := (\nabla_a^{(\alpha)} \nu) - (\nabla_a^{(\alpha)} \nu)^\top.$$

We define the connection $\tilde{\nabla}$ by

$$(36) \quad \phi_*(\tilde{\nabla}_u v) := (\nabla_a^{(\alpha)} \nu)^\top + (g^\nu)^{mk} C^\nu \left((\nabla_\nu^{(\alpha)} \nu)^\perp, a, e_k \right) e_m.$$

Let us show that $\tilde{\nabla}$ is the Chern connection $\tilde{\nabla}^v$ on Θ , by showing once again that it verifies the generalized Koszul formula (40). Using the notations $g^v(u, w) = g^{\phi_*v}(\phi_*u, \phi_*w)$, $C^v(u, w, z) = C^{\phi_*v}(\phi_*u, \phi_*w, \phi_*z)$ and the fact that $C^v(v, \cdot, \cdot) = 0$, the generalized Koszul formula can be written

$$2g^v(\tilde{\nabla}_u v, w) = ug^v(v, w) + vg^v(w, u) - wg^v(u, v) + g^v([u, v], w) - g^v([v, w], u) + g^v([w, u], v) - 2C^v(\nabla_v v, u, w).$$

Recalling that $\nabla^{(\alpha)}$ is the Chern connection on $\text{Dens}_+(M)$ and noticing that $\phi_*(\nabla_v v) = (\nabla_\nu^{(\alpha)} \nu)^\top$, the previous equation is satisfied if and only if

$$g^\nu \left((g^\nu)^{ij} C^\nu \left((\nabla_\nu^{(\alpha)} \nu)^\perp, a, e_j \right) e_i, \phi_*w \right) = C^\nu \left((\nabla_\nu^{(\alpha)} \nu)^\perp, a, \phi_*w \right),$$

which is easily checked to be true using the fact that $g^\nu(e_i, e_j) = (g^\nu)_{ij}$. To obtain the desired formula for $\tilde{\nabla}^v$, we write the α -connection in coordinates, through the same computations as in the proof of Theorem 7.3

$$\nabla_a^{(\alpha)} \nu = D\nu \cdot a - \frac{1 + \alpha}{2} \frac{a \nu}{\mu} \mu = u^i \partial_i v^j e_j + \left(\partial_i \partial_j \ell + \frac{1 - \alpha}{2} \partial_i \ell \partial_j \ell \right) u^i v^j f dx = u(v^j) e_j + \omega(u, v).$$

Using (35) we obtain

$$\begin{aligned} (\nabla_a^{(\alpha)} \nu)^\top &= u(v^m) e_m + (g^v)^{mk} g^\nu(\omega(u, v), e_k) e_m \\ (\nabla_\nu^{(\alpha)} \nu)^\perp &= w(v, v) - (g^v)^{ij} g^\nu(\omega(v, v), e_i) e_j \end{aligned}$$

which injected into (36) gives the desired result. \square

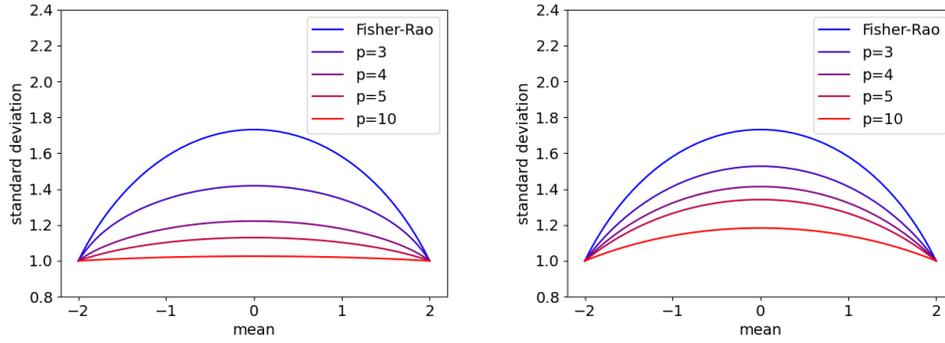


FIGURE 3. Comparison of the geodesics between two normal distributions shown in the parameter space for the L^p -Fisher-Rao metric (left) and for the α -connection (right), for different values of p and the corresponding values of $\alpha = 1 - 2/p$. The geodesics all start at the normal distribution of parameters $m_0 = -2$, $\sigma_0 = 1$, and end at the normal distribution of parameters $m_1 = 2$, $\sigma_1 = \sigma_0 = 1$.

Remark 7.6. Like in infinite dimensions (see Remark 4.8), when all entries are the same, the Chern connection on $\mathcal{P}_\Theta \simeq \Theta$ is the orthogonal projection of the α -connection $\nabla^{(\alpha)}$ on $\text{Dens}_+(\mathbb{R}^n)$, for $\alpha = 1 - \frac{2}{p}$, with respect to g^{ϕ_*v} , the Riemannian metric (12) induced by the L^p -Fisher-Rao metric

$$(37) \quad \tilde{\nabla}_v^v v = \text{Proj}^{\phi_*v} \left(\nabla_{\psi_*v}^{(\alpha)} \phi_*v \right) = \left(v(v^m) + (g^v)^{mk} g^{\phi_*v}(\omega(v, v), e_k) \right) \partial_m.$$

Theorem 7.7 (Geodesic equation on Θ). *The geodesic equation of the L^p -Fisher-Rao metric on the space \mathcal{P}_Θ is given by*

$$(38) \quad \ddot{\theta}^m + (g^{\dot{\theta}})^{mk} g^{\phi_*\dot{\theta}}(\omega(\dot{\theta}, \dot{\theta}), e_k) = 0,$$

where $(e_i)_i$ are the basis vectors (28) and ω is defined by (34).

Proof. This results directly from Lemma A.9 in Appendix A and writing $\tilde{\nabla}_{\dot{\theta}}^{\dot{\theta}} \dot{\theta} = 0$ in local coordinates using (37). \square

Example 7.8 (Normal distributions). Let us consider the example of univariate normal distributions, parametrized by mean and standard deviation. The parameter space is the upper half-plane $\Theta = \mathbb{R} \times \mathbb{R}_+^*$, and the Fisher-Rao metric, after a change of coordinates, is the hyperbolic metric of the Poincaré half-plane. The family of Riemannian metrics induced by the Finsler L^p -Fisher-Rao metric are also Poincaré metrics: for any $\theta = (m, \sigma) \in \Theta$ and $v \in \mathbb{R}^2$, the metric matrix is given by

$$g_{(m,\sigma)}^v = \frac{1}{\sigma^2} g_0^v$$

where g_0^v does not depend on m and σ . In order to compute geodesics for the L^p -Fisher-Rao metric, one can solve the geodesic equation (38), using the following densities with respect to a given $\mu(dx) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-m)^2}{2\sigma^2}) dx$: the basis vectors of the tangent plane $T_\mu \mathcal{P}_\Theta$ are given by

$$\frac{e_1}{\mu} = \frac{1}{\sigma} z, \quad \frac{e_2}{\mu} = \frac{1}{\sigma} (z^2 - 1) \quad \text{with} \quad z := \frac{x - m}{\sigma}$$

and for a given curve $t \mapsto \theta(t) = (m(t), \sigma(t))$,

$$\begin{aligned} \frac{\dot{\theta}}{\mu} &= \frac{1}{\sigma}(\dot{m}z + \dot{\sigma}(z^2 - 1)) \\ \frac{\omega(\dot{\theta}, \dot{\theta})}{\mu} &= \frac{1}{\sigma^2} \left((-1 + \frac{1}{p}z^2)\dot{m}^2 + (1 - 3z^2 + \frac{1}{p}(z^2 - 1)^2)\dot{\sigma}^2 + 2(-2z + \frac{1}{p}z(z^2 - 1))\dot{m}\dot{\sigma} \right). \end{aligned}$$

The L^p -Fisher-Rao geodesic can be compared to the solutions of the geodesic equation of the α -connection for $\alpha = 1 - 2/p$:

$$\ddot{m} - 2\frac{1+\alpha}{\sigma}\dot{m}\dot{\sigma} = 0, \quad \ddot{\sigma} + \frac{1-\alpha}{2\sigma}\dot{m}^2 - \frac{1+2\alpha}{\sigma}\dot{\sigma}^2 = 0.$$

In both cases, we solve the geodesic ODE with boundary constraints in Python for a discretization of 50 time steps, using the dedicated function in Scipy¹, which implements a fourth order collocation algorithm. We plot in Figure 7.8 the L^p -Fisher-Rao geodesics for several values of p as well as the α -geodesics for the corresponding values of α . As expected, these geodesics do not coincide, except for $p = 2$, where we retrieve the Fisher-Rao metric.

APPENDIX A. INFINITE DIMENSIONAL FINSLER GEOMETRY

In this appendix we will present several key definitions of Finsler geometry in the infinite dimensional setting. We will base our definitions on their counterparts from classical finite dimensional Finsler geometry, see eg. [6, 17, 36].

In the following let \mathcal{M} be an infinite dimensional, Fréchet manifold with tangent bundle $T\mathcal{M}$.

Definition A.1 (Finsler structure). A Finsler structure on \mathcal{M} is a function $F : T\mathcal{M} \rightarrow [0, \infty)$, that is smooth on the complement of the zero section of $T\mathcal{M}$ and satisfies for all $x \in \mathcal{M}$ and $X, Y \in T_x\mathcal{M}$

- (a) $F(\lambda Y) = \lambda F(Y)$ for all $\lambda > 0$;
- (b) $F(Y) \geq 0$ and $F(Y) = 0$ if and only if $Y = 0$.
- (c) $F(X + Y) \leq F(X) + F(Y)$.

The Finsler norm F is called strongly convex if we have in addition

- (d) For any $0 \neq V \in T_x\mathcal{M}$ the Hessian matrix g^V of F^2 at V exists and is positive definite, where

$$g^V(X, Y) := \frac{1}{2} \frac{\partial^2}{\partial s \partial t} [F^2(V + sX + tY)]_{s=t=0}.$$

Remark A.2. It can be shown that the strong convexity condition (d) implies the subadditivity condition (c) and several modern textbooks require strong convexity instead of subadditivity in the definition of a Finsler metric as this allows to develop several concepts of Riemannian geometry in the Finslerian setting. We choose to not assume this stronger condition as our main example, the L^p -Fisher-Rao metric, is not strongly convex.

Remark A.3 (Weak and strong Finsler metrics). For each $x \in \mathcal{M}$ the Finsler metric F induces a topology on $T_x\mathcal{M}$ and in finite dimensions this topology coincides with the original manifold topology. In infinite dimensions this is not the case and we will distinguish between two different types of Finsler metrics: strong Finsler metric, for which F_x induces the locally convex topology on $T_x\mathcal{M}$ and weak Finsler metrics, where the induced topology can be weaker than the locally convex topology. If \mathcal{M} is not a Banach manifold then any Finsler metric on \mathcal{M} can only be a weak Finsler metric.

Similarly as a Riemannian metric a Finsler structure F on a manifold \mathcal{M} defines a length structure on the set of piece wise smooth curves and thus one can define a corresponding path length distance:

¹https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_bvp.html

Definition A.4. Let $c : [a, b] \rightarrow \mathcal{M}$ be a piece wise smooth curve. The length of c with respect to F is defined as

$$L_F(c) := \int_a^b F(c(t), \dot{c}(t)) dt.$$

For any pair of points $x, y \in \mathcal{M}$ we consider the induced geodesic distance function

$$d_F(p, q) := \inf_c L_F(c),$$

where the infimum is calculated over the set of a piece wise smooth curves that connect x to y . Similar as in Riemannian geometry one can show that minimizing the length is equivalent to minimizing the energy, which is defined as

$$(39) \quad E_F(c) := \int_a^b F^2(c(t), \dot{c}(t)) dt.$$

Remark A.5 (Vanishing Geodesic distance). It is easy to see that the geodesic distance functions is symmetric and satisfies the triangle inequality. In general, for weak Finsler metrics, it does not satisfy the non-degeneracy property – $d_F(x, y) = 0$ if and only if $x = y$ for Finsler metrics. Indeed, even in the Riemannian case, several examples have been encountered where the geodesic distance can be degenerate or even vanishes identically, see eg. [20, 32, 10, 25].

Next we will introduce two important concepts from Finsler geometry: the Cartan tensor, which was introduced by E. Cartan [14] to evaluate the differences between Finsler metrics and Riemannian metrics, and the Chern connection, which is a generalization of the Levi-Civita connection on a Finsler manifold.

Note, that the definition of these two objects requires that the Finsler metric is strongly convex. As the L^p -Fisher-Rao metric, studied in the following sections, will not satisfy this property several of the calculations in these parts have to be taken with caution and should be thus understood only formally.

Definition A.6 (Cartan Tensor and Chern connection). Let (\mathcal{M}, F) be a Finsler manifold, where F is assumed to satisfy the strong convexity assumption. For any nonzero tangent vector $V \in T_x \mathcal{M}$, the Cartan tensor is defined as the symmetric trilinear form

$$C^V(X, Y, Z) := \frac{1}{4} \frac{\partial^3}{\partial s \partial t \partial r} [F^2(V + sX + tY + rZ)]_{s=t=r=0},$$

and the Chern connection, if it exists, is the unique affine, torsion-free connection ∇^V that is almost metric, that is for vector fields X, Y, Z we have

$$Xg^V(Y, Z) = g^V(\nabla_X^V Y, Z) + g^V(Y, \nabla_X^V Z) + 2C^V(\nabla_X^V V, Y, Z).$$

Remark A.7. In the above definition of the Chern-connection we have added the assumption on it's existence. This is additional assumption is not necessary in finite dimensions, but is entirely an infinite dimensional phenomenon, see eg. [8] where the authors studied a Riemannian metric on a group of diffeomorphisms such that the corresponding Levi-Civita connection does not exist.

The next Lemma, which will be of importance when we show the equivalence of the Chern-connection of the L^p -Fisher-Rao metric and the α -connection on $\text{Dens}(M)$, provides a generalized Koszul-formula for the Chern-connection:

Lemma A.8. *Let (\mathcal{M}, F) be a Finsler manifold, where F is assumed to satisfy the strong convexity assumption. For every non-zero vector field $V \in T_x \mathcal{M}$, the Chern connection, if it exists, satisfies*

the generalized Koszul formula

$$(40) \quad \begin{aligned} 2g^V(\nabla_X^V Y, Z) &= Xg^V(Y, Z) + Yg^V(Z, X) - Zg^V(X, Y) \\ &+ g^V([X, Y], Z) - g^V([Y, Z], X) + g^V([Z, X], Y) \\ &- 2C^V(\nabla_X^V V, Y, Z) - 2C^V(\nabla_Y^V V, Z, X) + 2C^V(\nabla_Z^V V, X, Y) \end{aligned}$$

Proof. The proof of this result follows exactly as in the finite dimensional situation, see eg. [17]. \square

The next results shows that the Chern-connection, similarly to the Levi-Civita connection in Riemannian geometry, describes the locally minimizing curves.

Lemma A.9. *Let (\mathcal{M}, F) be a Finsler manifold, where F is assumed to satisfy the strong convexity assumption. Assume in addition that the Chern connection ∇ exists. Then the critical points of the energy functional (39) are describe by the geodesic equation*

$$(41) \quad \nabla_{c_t}^{c_t} c_t = 0.$$

Proof. Assuming the existence of the Chern connection, this follows exactly as in finite dimensions, see eg. [17, 36]. \square

REFERENCES

- [1] S. Amari. *Information geometry and its applications*, volume 194. Springer, 2016.
- [2] S. Amari and H. Nagaoka. *Methods of information geometry*, volume 191. American Mathematical Soc., 2000.
- [3] C. Atkinson and A. F. Mitchell. Rao’s distance measure. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 345–365, 1981.
- [4] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. Information geometry and sufficient statistics. *Probability Theory and Related Fields*, 162:327–364, 2015.
- [5] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. *Information geometry*, volume 64. Springer, 2017.
- [6] D. Bao, S.-S. Chern, and Z. Shen. *An introduction to Riemann-Finsler geometry*, volume 200. Springer Science & Business Media, 2000.
- [7] M. Bauer, M. Bruveris, P. Harms, and P. W. Michor. Smooth perturbations of the functional calculus and applications to Riemannian geometry on spaces of metrics. *Communications in Mathematical Physics*, 389(2):899–931, 2022.
- [8] M. Bauer, M. Bruveris, and P. W. Michor. Homogeneous Sobolev metric of order one on diffeomorphism groups on real line. *Journal of Nonlinear Science*, 24:769–808, 2014.
- [9] M. Bauer, M. Bruveris, and P. W. Michor. Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bulletin of the London Mathematical Society*, 48(3):499–506, 2016.
- [10] M. Bauer, P. Harms, and S. C. Preston. Vanishing distance phenomena and the geometric approach to SQG. *Archive for Rational Mechanics and Analysis*, 235:1445–1466, 2020.
- [11] M. Bauer, Y. Lu, and C. Maor. A geometric view on the generalized Proudman–Johnson and r-Hunter–Saxton equations. *Journal of Nonlinear Science*, 32(1):17, 2022.
- [12] M. Bauer and C. Maor. Can we run to infinity? the diameter of the diffeomorphism group with respect to right-invariant sobolev metrics. *Calculus of Variations and Partial Differential Equations*, 60:1–35, 2021.
- [13] M. Bruveris and P. W. Michor. Geometry of the Fisher–Rao metric on the space of smooth densities on a compact manifold. *Mathematische Nachrichten*, 292(3):511–523, 2019.
- [14] É. Cartan. Les espaces de Finsler. *Actualités*, 79, 1934.
- [15] N. N. Cencov. Statistical decision rules and optimal inference. transl. math. *Monographs, American Mathematical Society, Providence, RI*, 1982.
- [16] N. N. Cencov. *Statistical decision rules and optimal inference*. Number 53. American Mathematical Soc., 2000.
- [17] S.-S. Chern and Z. Shen. *Riemann-Finsler geometry*, volume 6. World Scientific Publishing Company, 2005.
- [18] C. J. Cotter, J. Deasy, and T. Pryer. The r -Hunter–Saxton equation, smooth and singular solutions and their approximation. *Nonlinearity*, 33(12):7016, 2020.
- [19] D. G. Ebin and J. Marsden. Groups of diffeomorphisms and the motion of an incompressible fluid. *Annals of Mathematics*, pages 102–163, 1970.
- [20] Y. Eliashberg and L. Polterovich. Bi-invariant metrics on the group of Hamiltonian diffeomorphisms. *Internat. J. Math*, 4(5):727–738, 1993.
- [21] T. Friedrich. Die Fisher-information und symplektische strukturen. *Mathematische Nachrichten*, 153(1):273–296, 1991.

- [22] P. Gibilisco. l^p unit spheres and the α -geometries: Questions and perspectives. *Entropy*, 22(12):1409, 2020.
- [23] P. Gibilisco and T. Isola. Connections on statistical manifolds of density operators by geometry of noncommutative l^p -spaces. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 2(01):169–178, 1999.
- [24] P. Gibilisco and G. Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1(02):325–347, 1998.
- [25] R. L. Jerrard and C. Maor. Vanishing geodesic distance for right-invariant sobolev metrics on diffeomorphism groups. *Annals of Global Analysis and Geometry*, 55:631–656, 2019.
- [26] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geometric and Functional Analysis*, 23(1):334–366, 2013.
- [27] B. Khesin and G. Misiołek. Information geometry of diffeomorphism groups.
- [28] B. Khesin, K. Modin, and L. Volk. Simple unbalanced optimal transport. *arXiv preprint arXiv:2307.05703*, 2023.
- [29] F. Kogelbauer. On the global well-posedness of the inviscid generalized Proudman–Johnson equation using flow map arguments. *Journal of Differential Equations*, 268(3):1050–1080, 2020.
- [30] J. Lenells and G. Misiołek. Amari–Chentsov connections and their geodesics on homogeneous spaces of diffeomorphism groups. *Journal of Mathematical Sciences*, 196(2):144–151, 2014.
- [31] Y. Lu. *The L^p -Fisher-Rao metric and information geometry*. PhD thesis, Florida State University, 2023.
- [32] P. W. Michor and D. Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Documenta Mathematica*, 10:217–245, 2005.
- [33] K. Modin. Generalized Hunter–Saxton equations, optimal information transport, and factorization of diffeomorphisms. *The Journal of Geometric Analysis*, 25:1306–1334, 2015.
- [34] J. Moser. On the volume elements on a manifold. *Transactions of the American Mathematical Society*, 120(2):286–294, 1965.
- [35] N. J. Newton. An infinite-dimensional statistical manifold modelled on Hilbert space. *Journal of Functional Analysis*, 263(6):1661–1681, 2012.
- [36] H.-B. Rademacher. Nonreversible Finsler metrics of positive flag curvature. *A sampler of Riemann-Finsler geometry*, 50:261–302, 2004.
- [37] C. Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37, 01 1945.
- [38] A. Sarria and R. Saxton. Blow-up of solutions to the generalized inviscid Proudman–Johnson equation. *Journal of Mathematical Fluid Mechanics*, 15(3):493–523, 2013.
- [39] J. Zhang. Divergence function, duality, and convex analysis. *Neural computation*, 16(1):159–195, 2004.

M. BAUER: FLORIDA STATE UNIVERSITY AND UNIVERSITY OF VIENNA; A. LE BRIGANT: SAMM, UNIVERSITÉ PARIS 1; Y. LU: FLORIDA STATE UNIVERSITY; C. MAOR: EINSTEIN INSTITUTE OF MATHEMATICS, THE HEBREW UNIVERSITY OF JERUSALEM

Email address: bauer@math.fsu.edu, alice.le-brigant@univ-paris1.fr.com, yl18f@fsu.edu, cy.maor@mail.huji.ac.il