An Interpolating Distance between Optimal Transport and Fisher-Rao metrics

Lenaic Chizat   Gabriel Peyré   Bernhard Schmitzer
François-Xavier Vialard*
CEREMADE, CNRS, Université Paris-Dauphine
INRIA, Project team Mokaplan

January 20, 2017

Abstract

This paper defines a new transport metric over the space of non-negative measures. This metric interpolates between the quadratic Wasserstein and the Fisher-Rao metrics and generalizes optimal transport to measures with different masses. It is defined as a generalization of the dynamical formulation of optimal transport of Benamou and Brenier, by introducing a source term in the continuity equation. The influence of this source term is measured using the Fisher-Rao metric, and is averaged with the transportation term. This gives rise to a convex variational problem defining the new metric. Our first contribution is a proof of the existence of geodesics (i.e. solutions to this variational problem). We then show that (generalized) optimal transport and Hellinger metrics are obtained as limiting cases of our metric. Our last theoretical contribution is a proof that geodesics between mixtures of sufficiently close Dirac measures are made of translating mixtures of Dirac masses. Lastly, we propose a numerical scheme making use of first order proximal splitting methods and we show an application of this new distance to image interpolation.

Keywords Unbalanced optimal transport · Wasserstein L^2 metric · Fisher-Rao metric · Positive Radon measures

1 Introduction

Optimal transport is an optimization problem which gives rise to a popular class of metrics between probability distributions. We refer to the monograph of Villani [31] for a detailed overview of optimal transport. We now describe in an informal way various formulations of transportation-like distances, and expose also our new metric. A mathematically rigorous treatment can be found in Section 2.

1.1 Static and dynamic optimal transport

In its original formulation by Monge, optimal transport takes into account a ground cost that measures the amount of work needed to transfer a measure towards another one. This formulation was later relaxed by Kantorovich [20] as a convex linear program, often referred to as a "static formulation". Given a *ground cost* $c : \Omega^2 \mapsto \mathbb{R}$, $\Omega \subset \mathbb{R}^d$ and two probability measures ρ_0 and ρ_1 on Ω, the central problem is that of minimizing

$$\int_{\Omega \times \Omega} c(x,y) \, \mathrm{d}\gamma(x,y) \tag{1}$$

over the set $\Gamma_{(\rho_0,\rho_1)}$ of coupling measures, that contains non-negative measures γ on Ω^2 admitting ρ_0 and ρ_1 as first and second marginals. More explicitly, for every measurable set $A \subset \Omega$, $\gamma(A, \Omega) = \rho_0(A)$ and $\gamma(\Omega, A) = \rho_1(A)$. When the *ground cost* is the distance $|x - y|^p$, the minimum induces a metric on the space of probability measures which is often referred to as the p-Wasserstein distance, denoted by W_p.

In a seminal paper, Benamou and Brenier [6] showed that this static formulation is equivalent to a "dynamical" formulation, where one seeks for a "geodesic path" $\rho(t, x)$ between the two densities which has a convex formulation. More precisely, they showed that the squared 2-Wasserstein is obtained by minimizing the kinetic energy

$$\int_0^1 \int_\Omega |v(t,x)|^2 \rho(t,x) \mathrm{d}x \mathrm{d}t$$

where ρ interpolates between ρ_0 and ρ_1 and v is a velocity field which advects the mass distribution ρ. They introduced the key change of variables via the momentum from classical mechanics, $(\rho, v) \mapsto (\rho, \omega) \overset{\text{def.}}{=} (\rho, \rho v)$, this amounts to minimizing the convex functional

$$\int_0^1 \int_\Omega \frac{|\omega(t,x)|^2}{\rho(t,x)} \, \mathrm{d}x \mathrm{d}t \tag{2}$$

over the couples (ρ, ω) satisfying the continuity equation

$$\partial_t \rho + \nabla \cdot \omega = 0, \quad \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1 \,. \tag{3}$$

In particular, their paper emphasized the Riemannian interpretation of this metric and paved the way for efficient numerical solvers.

The geometric nature of optimal transport makes this tool interesting to perform shape interpolation, especially in medical image analysis [17]. Even though the resulting interpolation is somewhat simplistic compared to methods based on diffeomorphic flows such as LDDMM [4], the fact that the problem is convex is interesting for numerical purposes and large scale applications. A notable restriction of optimal transport is however that it is only defined between measures having the same mass. This might be an issue for applications in medical imaging, where measures need not be normalized, and biophysical phenomena at stake typically involve some sort of mass creation or destruction. To date,

several generalizations of optimal transport have been proposed, with various goals in mind (see Section 1.3). However, there is currently no formulation for which geodesics can be interpreted as a joint displacement and modification of mass. In this article, we propose to tackle this issue by interpolating between the Wasserstein and the Fisher-Rao metric, which can be achieved using a convex dynamical formulation.

In the remainder of this section, we introduce our model and describe related work which generalizes optimal transport to unbalanced measures (nonnegative measures whose total mass may differ). In addition, we introduce in an informal manner a family of distances which shows the similarity between *a priori* unrelated models of transport between unbalanced measures. In Section 2, we prove the existence of geodesics and exhibit optimality conditions. Section 3 is devoted to the study of the effect of varying the interpolation factor and to the limit models. We detail in Section 4 explicit geodesics between atomic measures satisfying suitable properties. Finally, we describe a numerical scheme in Section 5 and use the distance for image interpolation experiments.

1.2 Presentation of the model and contributions

The dynamical formulation (2) suggests a natural way to relax the equality of mass constraint by introducing a source term ζ in the continuity equation (3) as follows

$$\partial_t \rho + \nabla \cdot \omega = \zeta, \quad \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1,$$

and adding a term penalizing ζ in the Benamou-Brenier functional, in the spirit of metamorphoses introduced in imaging in [30]. As for the penalization on ζ, we think it is natural to look for a Lagrangian which behaves like a Riemannian metric. Riemannian metrics are natural for interpolation purposes and for applications to imaging. Moreover, the standard notion of curvature is available in a Riemannian setting and provides important information on the underlying geometry, even in the case of infinite dimensional spaces.

To this end, we will use a direct extension of the Fisher-Rao Riemannian metric. This metric was introduced in 1945 by C. R. Rao, who applied for the first time differential geometry in the field of statistics [27]. More precisely, at a given probability density ρ and tangent vector $\delta\rho$, the Fisher-Rao metric tensor is $\mathrm{FR}(\rho)(\delta\rho, \delta\rho) = \int_\Omega \frac{(\delta\rho)^2}{\rho} \mathrm{d}x$. Note that $\delta\rho$ is a tangent vector at ρ to the space of probability densities and, therefore, its integral is zero. Usually in statistics, one often deals with parametric densities. Let us consider $\Theta : \mathbb{R}^n \to \mathrm{Prob}(\Omega)$ be a map from a space of parameters ($\mathbb{R}^n$) to the space of probability densities on Ω. The Fisher-Rao Riemannian metric on the space of parameters is $\Theta^* \mathrm{FR}$, the pull-back of the Riemannian metric FR by Θ. It is of interest to note that, on the space of probability densities, the Fisher-Rao metric has recently been proven to be the unique reparametrization invariant Riemannian metric, independently in [2] and [3]. Note that other extensions to the space of densities could have been considered by penalizing the total volume variation independently (see for instance [3]).

In this article, we will consider the metric tensor FR on the space of densities (of finite volume) but which do not integrate to 1, thus removing the constraint $\int_\Omega \delta\rho \, \mathrm{d}x = 0$. We will still call this metric tensor the Fisher-Rao (Riemannian) metric.

The distance induced by this tensor over all positive densities is called the Hellinger distance. More precisely, the name refers to the restriction of this induced distance to the space of probability densities, but analogously to the Fisher-Rao tensor, we decide to use the name for the distance over the full space of positive densities. It turns out to be simply the L^2 distance between the square roots of the densities. Indeed, between two smooth positive densities ρ_0 and ρ_1, this distance is defined as the minimum of

$$\int_0^1 \int_\Omega \frac{\zeta(t,x)^2}{\rho(t,x)} \, \mathrm{d}x \mathrm{d}t$$

over the couples (ρ, ζ) such that $\partial_t \rho = \zeta$, $\rho(0, \cdot) = \rho_0$ and $\rho(1, \cdot) = \rho_1$. Its geodesics are explicit and given by $\rho(t,x) = \left[t\sqrt{\rho_1(x)} + (1-t)\sqrt{\rho_0(x)} \right]^2$.

In this article, we propose to define a metric tensor on the space of all densities that interpolates between the Fisher-Rao tensor and the Wasserstein tensor. More precisely, we define the new metric tensor as the infimal convolution of the two metric tensors. Since this new metric is defined from a Riemannian point of view, we chose the name Fisher-Rao over Hellinger. Therefore, this new metric can be written as the minimization of

$$\int_0^1 \left[\int_\Omega \frac{|\omega(t,x)|^2}{2\rho(t,x)} \mathrm{d}x + \kappa^2 \int_\Omega \frac{\zeta(t,x)^2}{2\rho(t,x)} \mathrm{d}x \right] \mathrm{d}t \tag{4}$$

under the constraints

$$\partial_t \rho + \nabla \cdot \omega = \zeta, \quad \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1, \tag{5}$$

where $\kappa > 0$ is a parameter. The metric defined by this infimum as well as the minimizers are the central objects of this paper. While formula (4) only makes sense for positive smooth densities, we rigorously extend this definition on the space of non-negative measures in Section 2. This extension is well behaved since the functional that is minimized is convex and positively 1-homogeneous, two key properties which allow to extend it readily as a lower semi-continuous (l.s.c.) functional over measures [8].

Let us now give a physical interpretation of this model: just as the first term in the functional measures the kinetic energy of the interpolation $\|v\|_{L^2(\rho)}^2$ where v is the velocity field, the second term measures an energy of growth $\|g\|_{L^2(\rho)}^2$ where g denotes the rate of growth ζ/ρ which is a scalar field depending on t and x. In this way, our problem can be rewritten as the minimization of

$$\frac{1}{2} \int_0^1 \left[\int_\Omega |v(t,x)|^2 \rho(t,x) \mathrm{d}x + \kappa^2 \int_\Omega |g(t,x)|^2 \rho(t,x) \mathrm{d}x \right] \mathrm{d}t$$

under the constraints

$$\partial_t \rho = g\rho - \nabla \cdot (\rho v), \quad \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1 \,.$$

This formulation emphasizes the geometric nature of this model.

1.3 Previous works and connections

The idea of extending optimal transport to unbalanced measures is far from being new, and was already suggested by Kantorovitch in 1942 (see [16]) where he allows mass to be thrown away out of the (compact) domain. The geometry of the domain also plays a role in the more recent models of [15] where the boundary of the domain is taken as an infinite reserve of mass. In the past few years, some authors combined generalizations of optimal transport with numerical algorithms in order to tackle practical applications. In [5], the problem is formulated from an optimal control point of view where the *matching* term is the L^2 distance between densities. The dynamic problem remains convex and is solved numerically with an augmented Lagrangian algorithm. A quadratic penalization is also suggested in [23] but based, this time, on the dynamic formulation, with the model of metamorphoses [30] in mind: they use constraint (5) and add the term $\iint \zeta^2 \mathrm{d}x\mathrm{d}t$ in the Benamou-Brenier functional (along with a viscous dissipation term). From the perspective of computing image interpolations [22] suggests, instead of penalizing ζ in the dynamic functional, to fix it as a function of ρ, according to some *a priori* growth model . Then, if the growth model is well chosen, minimizing the Benamou-Brenier functional gives rise to meaningful interpolations. Note that one of our limit models (see Definition 6) falls into this category.

Now we introduce with more details the classes of distances introduced in [26, 25], along with partial optimal transport distances, as they share some similarities with our approach.

Partial transport and a dynamical formulation. Partial optimal transport, introduced in [9], consists in transferring a fixed amount of mass between two measures ρ_0 and ρ_1 as cheaply as possible. Letting $\Gamma_{\leq(\rho_0,\rho_1)}$ be the set of nonnegative measures on $\Omega \times \Omega$ whose left and right marginals are dominated by ρ_0 and ρ_1 respectively, this amounts to minimizing (1) under the constraints $\gamma \in \Gamma_{\leq(\rho_0,\rho_1)}$ and $\gamma(\Omega \times \Omega) = m$. What happens is rather intuitive: the plan is divided between an active set, where classical optimal transport is performed, and an inactive set. A theoretical study of this model is done in [9] and [14], where, for instance, results on the regularity of the boundary between the active and inactive sets are proved. For the quadratic cost, the Lagrangian formulation of this problem reads

$$\inf_{\gamma \in \Gamma_{\leq(\rho_0,\rho_1)}} \int_{\Omega \times \Omega} \left(|x-y|^2 - \lambda \right) \mathrm{d}\gamma(x,y). \qquad (6)$$

where λ is the Lagrange multiplier associated to the total mass constraint. It is shown in [9, Corollary 2.11] that there exists an optimal plan γ^* for (6), which is also a minimizer for the optimal partial transport problem between ρ_0 and ρ_1 with mass constraint $m = \gamma^*(\Omega \times \Omega)$. If ρ_0 and ρ_1 are absolutely continuous, m continuously increases from 0 to $\min\{\rho_0(\Omega), \rho_1(\Omega)\}$ as λ increases. It is clear that for optimality, the integrand must be nonpositive γ^*-a.e. which means that no mass is transported over a distance greater than $\sqrt{\lambda}$.

Independently, a class of distances interpolating between total variation and the W_p metrics has been introduced by [26, 25], taking their inspiration from the optimal control formulation of [7]. This distances are defined as the infimum of

$$W_p(\tilde{\rho}_0, \tilde{\rho}_1) + \kappa \cdot (|\rho_0 - \tilde{\rho}_0| + |\rho_1 - \tilde{\rho}_1|)$$

over $\tilde{\rho}_0, \tilde{\rho}_1$, non-negative measures on Ω, for $p \geq 1$. Now remark that, for fixed ρ_0 and ρ_1 and up to changing κ, the previous problem is equivalent to minimizing

$$W_p^p(\tilde{\rho}_0, \tilde{\rho}_1) + \kappa^p \cdot (|\rho_0 - \tilde{\rho}_0| + |\rho_1 - \tilde{\rho}_1|) \tag{7}$$

To our knowledge, the equivalence between this model and partial optimal transport has not been exposed yet although this allows an interesting dynamical formulation for the well studied partial transport problem. This is the object of the following Proposition.

Proposition 1.1 (Equivalence between partial OT and Piccoli-Rossi distances). *When choosing $2\kappa^2 = \lambda$ and $p = 2$, problems* (6) *and* (7) *are equivalent.*

Proof. First, it is clear that problem (7) is left unchanged when adding the domination constraints $\tilde{\rho}_0 \leq \rho_0$ and $\tilde{\rho}_1 \leq \rho_1$ ([25, Proposition 4]). This implies that $|\rho_i - \tilde{\rho}_i|$ is equal to $\rho_i(\Omega) - \tilde{\rho}_i(\Omega)$ in such a way that (7) can be rewritten as

$$\inf_{\gamma \in \Gamma_{\leq(\rho_0, \rho_1)}} \int_{\Omega \times \Omega} \left(|x - y|^2\right) \mathrm{d}\gamma(x, y) + \kappa^2 \left(\rho_0(\Omega) + \rho_1(\Omega) - 2\gamma(\Omega \times \Omega)\right)$$

This is clearly equivalent to the Lagrangian formulation (6). $\qquad \square$

The interest of this result is that it allows a dynamic formulation of the partial optimal transport problem. Indeed, it is shown in [25] that minimizing (7) is equivalent to minimizing the following dynamic functional

$$\int_0^1 \left[\int_\Omega \frac{|\omega(t, x)|^2}{\rho(t, x)} \mathrm{d}x + \kappa^2 \int_\Omega |\zeta(t, x)| \mathrm{d}x\right] \mathrm{d}t \tag{8}$$

subject to the continuity constraints with a source (5).

The case W_1. When chosing $p = 1$ in (7), one obtains another partial optimal transport problem, with $|\cdot|$ as a ground metric. This distance was already introduced by [18] in view of the study of Lipschitz spaces. Without providing a proof, it is quite clear that this distance is also obtained by minimizing

$$\int_0^1 \left[\int_\Omega |\omega(t, x)| \mathrm{d}x + \kappa \int_\Omega |\zeta(t, x)| \mathrm{d}x\right] \mathrm{d}t \tag{9}$$

over the set set of solutions to (5).

A unifying framework: homogeneity and convexity. We conclude this bibliographical section by suggesting a unifying framework for a class of generalizations of the W_p metrics. Consider the problem of minimizing

$$\int_0^1 \left[\frac{1}{p} \int_\Omega \frac{|\omega|^p}{\rho^{p-1}} \mathrm{d}x + \kappa^p \frac{1}{q} \int_\Omega \frac{|\zeta|^q}{\rho^{q-1}} \mathrm{d}x \right] \mathrm{d}t$$

over the set of solutions to (5). The functional penalizes the p-th power of the L_ρ^p norm of the velocity field and the q-th power of the L_ρ^q norm of the rate of growth, the latter being reparametrization invariant. Whatever $p, q \geq 1$, the above functional can be rigorously defined as a homogeneous, convex, l.s.c functional on measures (as in Section 2). Also, this problem can be seen as the dual problem of maximizing the variation of an energy

$$\int_\Omega \varphi(1, \cdot) \rho_1 - \int_\Omega \varphi(0, \cdot) \rho_0$$

over continuously differentiable potentials φ on $[0, 1] \times \Omega$ satisfying

$$\partial_t \varphi + \frac{p-1}{p} |\nabla \varphi|^{p/(p-1)} + \kappa^{-p/(q-1)} \frac{q-1}{q} |\varphi|^{q/(q-1)} \leq 0$$

if p and q are greater than 1. Otherwise, the constraint should be adapted as follows: if $p = 1$, remove the second term from the inequality constraint and add the constraint $|\nabla \varphi| \leq 1$. If $q = 1$, remove the third term and add the constraint $|\varphi| \leq \kappa$.

For $p = q = 1$, we recover (9) and for $p = 2$ and $q = 1$, we recover (8), implying the connections to partial optimal transport arising from our previous remarks. While there is an interesting decoupling effect between the growth term and the transport term when p or q is equal to 1, one should choose in general $p = q \geq 1$ so as to define (the p-th power of) a metric. In this article, we focus on the case $p = 2$, $q = 2$, which is the only Riemannian-like metric of this class, namely our interpolated distance between W_2 and Fisher-Rao.

1.4 Relation with [21]

After finalizing this paper, we became aware of the recent work of [21] which defines and studies the same metric. These two works were done simultaneously and independently. The approach of [21] is however quite different. While both works prove existence of geodesics, the proofs rely on different strategies, and in particular our construction is based on convex analysis. The use of tools from convex analysis allows us to prove a useful sufficient condition for the uniqueness of a geodesic. We study the limiting models (generalized transport and Fisher-Rao), through the introduction of the κ parameter. We prove existence and uniqueness of travelling Dirac solutions, while [21] provides an informal discussion. Lastly, we detail a numerical scheme and show applications to image interpolation. Let us stress that [21] provides many other contributions (such as a Riemannian calculus, topological properties and the definition of gradient flows) that we do not cover here.

1.5 Notations and preliminaries

Space of measures In the sequel, we consider a compact convex spatial domain $\Omega \subset \mathbb{R}^d$ with $d \in \mathbb{N}^*$. For X being either Ω or $[0,1] \times \Omega$, the set of Borel measures (resp. nonnegative Borel measures) on X is denoted by $\mathcal{M}(X)$ (resp. $\mathcal{M}_+(X)$). Endowed with the total variation norm

$$|\mu|(X) \overset{\text{def.}}{=} \sup \left\{ \int_X \varphi \, \mathrm{d}\mu : \varphi \in C(X), \, |\varphi|_\infty = 1 \right\},$$

$\mathcal{M}(X)$ is a Banach space. We identify this space with the topological dual of the space of real valued continuous functions $C(X)$ endowed with the sup norm topology. Another useful topology on $\mathcal{M}(X)$ is the weak* topology arising from this duality: a sequence of measures $(\mu_n)_{n \in \mathbb{N}}$ weak* converges towards $\mu \in \mathcal{M}(X)$ if and only if for all $\varphi \in C(X)$, $\lim_{n \to +\infty} \int_X \varphi \mathrm{d}\mu_n = \int_X \varphi \mathrm{d}\mu$. By Prokhorov's Theorem, any bounded subset of $\mathcal{M}(X)$ (for the total variation) is relatively sequentially compact for the weak* topology.

Convex analysis Let E and E^* be *topologically paired* vector spaces i.e. vector spaces assigned with locally convex Hausdorff topology such that all continuous linear functionals on each space can be identified with the elements of the other (e.g. $C(X)$ and $\mathcal{M}(X)$ with the strong and weak* topology respectively). The pairing between those spaces is the bilinear form $\langle \cdot, \cdot \rangle : E \times E^* \to \mathbb{R}$. The convex conjugate of a function $f : E \to \mathbb{R} \cup \{+\infty\}$ is defined for each $y \in E^*$ by

$$f^*(y) \overset{\text{def.}}{=} \sup_{x \in E} \langle x, y \rangle - f(x) \,.$$

The subdifferential operator is defined at a point $x \in E$ as

$$\partial f(x) \overset{\text{def.}}{=} \left\{ y \in E^*; f(x') - f(x) \geq \langle y, x' - x \rangle \text{ for all } x' \in E \right\}$$

and is empty if $f(x) = +\infty$. Those definitions admit their natural counterparts for functions defined on E^*.

Theorem 1.1 (Fenchel-Rockafellar [28])**.** *Let (E, E^*) and (F, F^*) be two couples of topologically paired spaces. Let $A : E \to F$ be a continuous linear operator and $A^* : F^* \to E^*$ its adjoint. Let f and g be l.s.c. and proper convex functions defined on E and F respectively, with values in $\mathbb{R} \cup \{+\infty\}$. If there exists $x \in E$ such that $f(x)$ is finite and g is continuous at Ax, then*

$$\sup_{x \in E} -f(-x) - g(Ax) = \min_{y^* \in F^*} f^*(A^* y^*) + g^*(y^*)$$

and the min *is attained. Moreover, there exists a maximizer $x \in E$ if and only if there exists $y^* \in F^*$ satisfying $Ax \in \partial g^*(y^*)$ and $A^* y^* \in \partial f(-x)$.*

Other notations

- $\mu \ll \nu$ means that the measure μ is absolutely continuous w.r.t. ν;
- for a (possibly vector valued) measure μ, $|\mu| \in \mathcal{M}_+(X)$ is its variation;
- for a positively 1-homogeneous (in short : homogeneous) function f, we denote by $f(\mu)$ the measure defined by $f(\mu)(A) = \int_A f(\mathrm{d}\mu/\mathrm{d}\lambda)\mathrm{d}\lambda$ for any measurable set A, with λ satisfying $|\mu| \ll \lambda$. The homogeneity assumption guarantees that this definition does not depend on λ;
- $T_\#\mu$ is the image measure of μ through the Borel map $T : X_1 \to X_2$, also called the pushforward measure. It is given by $(T_\#\mu)(A_2) = \mu(T^{-1}(A_2))$ for any Borel subset of X_2;
- δ_x is a Dirac measure of mass 1 located at the point x;
- $\iota_{\mathcal{C}}$ is the (convex) indicator function of a convex set $\mathcal{C}$ which takes the value 0 on $\mathcal{C}$ and $+\infty$ everywhere else;
- when μ is a measure on $[0,1] \times \Omega$ whose time marginal is absolutely continuous with respect to the Lebesgue measure on $[0,1]$, we denote by $(\mu_t)_{t \in [0,1]}$ the ($\mathrm{d}t$-a.e. uniquely determined) family of measures satisfying $\mu = \int_0^1 (\mu_t \otimes \delta_t)\mathrm{d}t$, where $\otimes$ denotes the product of measures. This is a consequence of the disintegration Theorem (see for instance [1, Theorem 5.3.1]) which we frequently use without explicitly mentioning it. In order to alleviate notations, we replace the integral expression of μ by $\mu = \mu_t \otimes \mathrm{d}t$.

2 Definition and existence of geodesics

In order to prove that the minimum of the variational problem is attained, rather than using the direct method of calculus of variation, we choose to take advantage of the convexity of the problem and prove it by Fenchel-Rockafellar duality. Indeed, we show that (a generalization of) problem (4) is obtained by considering the dual of a variational problem on the space of differentiable functions. This point of view allows to prove existence of geodesics and exhibit sufficient optimality conditions.

2.1 Definition of the interpolating distance

Consider the closed convex set

$$B_\kappa \overset{\text{def.}}{=} \left\{ (a, b, c) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} : a + \frac{1}{2}\left(|b|^2 + \frac{c^2}{\kappa^2} \right) \leq 0 \right\}$$

and its convex indicator function

$$\iota_{B_\kappa} : (a, b, c) \in \mathbb{R}^{d+2} \mapsto \begin{cases} 0 & \text{if } (a, b, c) \in B_\kappa \\ +\infty & \text{otherwise.} \end{cases}$$

We denote by f_κ the Fenchel-Legendre conjugate of ι_{B_κ}, i.e.

$$f_\kappa : (x, y, z) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \mapsto \begin{cases} \frac{|y|^2 + \kappa^2 z^2}{2x} & \text{if } x > 0, \\ 0 & \text{if } (x, |y|, z) = (0, 0, 0) \\ +\infty & \text{otherwise} \end{cases}$$

9

which is by construction proper, convex, lower semicontinuous (l.s.c.) and 1-homogeneous. One can check by direct computations that

$$\partial f_\kappa(x, y, z) = \begin{cases} \left(-\frac{|y|^2+\kappa^2 z^2}{2x^2}, \frac{y}{x}, \frac{\kappa^2 z}{x}\right) & \text{if } x > 0, \\ B_\kappa & \text{if } (x, |y|, z) = (0, 0, 0), \\ \emptyset & \text{otherwise.} \end{cases} \quad (10)$$

We denote $\mathcal{M} \stackrel{\text{def.}}{=} \mathcal{M}([0, 1] \times \Omega)$ and for $\mu = (\rho, \omega, \zeta) \in \mathcal{M} \times \mathcal{M}^d \times \mathcal{M}$ we define the convex functional

$$D_\kappa(\mu) \stackrel{\text{def.}}{=} \int_{[0,1]\times\Omega} f_\kappa\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda$$

where λ is any non-negative Borel measure satisfying $|\mu| \ll \lambda$. As f_κ is homogeneous, the definition of D_κ does not depend on the reference measure λ.

Definition 1 (Continuity equation with source). We denote by $\mathcal{CE}_0^1(\rho_0, \rho_1)$ the affine subset of $\mathcal{M} \times \mathcal{M}^d \times \mathcal{M}$ of triplets of measures $\mu = (\rho, \omega, \zeta)$ satisfying the continuity equation $\partial_t \rho + \nabla \cdot \omega = \zeta$ weakly, interpolating between ρ_0 and ρ_1 and satisfying homogeneous Neumann boundary conditions. This is equivalent to requiring

$$\int_0^1 \int_\Omega \partial_t \varphi \, \mathrm{d}\rho + \int_0^1 \int_\Omega \nabla\varphi \cdot \mathrm{d}\omega + \int_0^1 \int_\Omega \varphi \, \mathrm{d}\zeta = \int_\Omega \varphi(1, \cdot)\mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot)\mathrm{d}\rho_0 \quad (11)$$

for all $\varphi \in C^1([0, 1] \times \Omega)$.

Remarks.

- The set $\mathcal{CE}_0^1(\rho_0, \rho_1)$ is not empty: it contains for instance the linear interpolation $(\rho, 0, \partial_t \rho)$ with $\rho = (t\rho_1 + (1-t)\rho_0) \otimes \mathrm{d}t$. Moreover, if we assume that ρ_0 and ρ_1 have equal mass, then there exists $(\rho, \omega, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ such that $\zeta = 0$. This is a consequence of [1, Theorem 8.3.1].

- If we make the additional assumption that $\omega \ll \rho$ and $\zeta \ll \rho$ (which is satisfied as soon as $D_\kappa(\mu)$ is finite), we find that the time marginal of μ is absolutely continuous w.r.t. the Lebesgue measure, allowing to disintegrate it in time.

- Under the same assumptions, μ admits a continuous representative, i.e. it is $\mathrm{d}t-$a.e. equal to a curve which is weak* continuous in time.

- When ρ is disintegrable in time, the map $t \mapsto \rho_t(\Omega) = \int_\Omega \rho_t$ admits the distributional derivative $\rho_t'(\Omega) = \zeta_t(\Omega)$, for almost every $t \in [0, 1]$. Indeed, taking a test function φ constant in space in (11) gives

$$\int_0^1 \varphi'(t)\rho_t(\Omega)\mathrm{d}t + \int_0^1 \varphi(t)\zeta_t(\Omega)\mathrm{d}t = \varphi(1)\rho_1(\Omega) - \varphi(0)\rho_0(\Omega).$$

We are now in position to define the central object of this paper WF_κ, as a map $\mathcal{M}_+(\Omega) \times \mathcal{M}_+(\Omega) \to \mathbb{R}_+$.

Definition 2 (Interpolating distance). For $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)^2$, the (squared) interpolating distance is defined as

$$WF_\kappa^2(\rho_0, \rho_1) \stackrel{\text{def.}}{=} \inf_{\mu \in \mathcal{CE}_0^1(\rho_0, \rho_1)} D_\kappa(\mu) \,. \tag{$\mathcal{P}_\mathcal{M}$}$$

The following result shows that $WF_\kappa(\rho_0, \rho_1)$ is always finite.

Proposition 2.1 (Bound on the distance). *Let $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)^2$. The following bounds hold:*

$$WF_\kappa^2(\rho_0, \rho_1) \leq 2\kappa^2 |(\sqrt{\rho_1} - \sqrt{\rho_0})^2|(\Omega)$$
$$\leq 2\kappa^2 (\rho_0(\Omega) + \rho_1(\Omega)).$$

Proof. Consider the triplet of measures $\mu = (\rho, \omega, \zeta)$ with

$$\begin{cases} \rho = \left(t\sqrt{\rho_1} + (1-t)\sqrt{\rho_0}\right)^2 \otimes \mathrm{d}t \,, \\ \omega = 0 \,, \\ \zeta = 2(\sqrt{\rho_1} - \sqrt{\rho_0})(t\sqrt{\rho_1} + (1-t)\sqrt{\rho_0}) \otimes \mathrm{d}t, \end{cases}$$

and notice that it belongs to $\mathcal{CE}_0^1(\rho_0, \rho_1)$. Taking $\lambda \in \mathcal{M}_+$ such that $\mu \ll \lambda$, the first bound comes from

$$D_\kappa(\mu) = 2\kappa^2 \int_\Omega \left(\sqrt{\frac{\mathrm{d}\rho_1}{\mathrm{d}\lambda}} - \sqrt{\frac{\mathrm{d}\rho_0}{\mathrm{d}\lambda}} \right)^2 \mathrm{d}\lambda$$

and does not depend on λ by homogeneity, while the second bound is obtained in the worst case when the supports are disjoint. $\qquad\square$

Remark. Notice that the first bound is tight and it corresponds to the Hellinger distance, defined in Definition 7. It is the solution to ($\mathcal{P}_\mathcal{M}$) when adding the constraint $\omega = 0$ (see Theorem 3.1).

2.2 Existence of geodesics

We now prove existence of geodesics. The proof mainly uses the Fenchel-Rockafellar duality Theorem (see Theorem 1.1) and a duality result for integrals of convex functionals.

Definition 3 (Geodesics). Geodesics are measures $\rho \in \mathcal{M}_+$ for which there exists a momentum ω and a source ζ such that (ρ, ω, ζ) is a minimizer of ($\mathcal{P}_\mathcal{M}$).

Definition 4 (Primal problem). We introduce a variational problem on the space of differentiable functions which is defined as follows

$$\inf_{\varphi \in C^1([0,1] \times \Omega)} J(\varphi) \stackrel{\text{def.}}{=} \int_0^1 \int_\Omega \iota_{B_\kappa}(\partial_t \varphi, \nabla \varphi, \varphi) + \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0 - \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1 \,. \tag{$\mathcal{P}_{C^1}$}$$

Theorem 2.1 (Strong duality and existence of geodesics)**.** *Let $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)^2$. It holds*

$$WF_\kappa(\rho_0, \rho_1)^2 = - \inf_{\varphi \in C^1([0,1] \times \Omega)} J(\varphi)$$

and the infimum in $(\mathcal{P}_\mathcal{M})$ is attained.

Proof. First remark that $WF_\kappa(\rho_0, \rho_1)$ is necessarily finite from Proposition 2.1. Let us rewrite $(\mathcal{P}_{C^1})$ as

$$\inf_{\varphi \in C^1([0,1] \times \Omega)} F(A\varphi) + G(\varphi)$$

where

$$A : \varphi \mapsto (\partial_t \varphi, \nabla \varphi, \varphi),$$

$$F : (\alpha, \beta, \gamma) \mapsto \int_0^1 \int_\Omega \iota_{B_\kappa}(\alpha(t,x), \beta(t,x), \gamma(t,x)) \mathrm{d}x \mathrm{d}t,$$

$$\text{and} \quad G : \varphi \mapsto \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0 - \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1.$$

Remark that F and G are convex, proper and lower-semicontinuous. It is easy to find a function $\varphi \in C^1([0,1] \times \Omega)$ such that F is continuous at $A\varphi$, taking $(A\varphi)(t,x)$ in the interior of the closed set B_κ for all (t,x) is enough. The Fenchel-Rockafellar duality Theorem guarantees the following equality

$$\inf_{\varphi \in C^1([0,1] \times \Omega)} J(\varphi) = \max_{\mu \in \mathcal{M} \times \mathcal{M}^d \times \mathcal{M}} \{-F^*(\mu) - G^*(-A^*\mu)\}. \qquad (12)$$

In the second term, we recognize the generalized continuity constraint

$$
\begin{aligned}
G^*(-A^*\mu) &= \sup_{\varphi \in C^1([0,1] \times \Omega)} \left\{ -\langle A\varphi, \mu \rangle + \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0 \right\} \\
&= \begin{cases} 0 & \text{if } \mu \in \mathcal{CE}_0^1(\rho_0, \rho_1) \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

For the first term, as the Fenchel conjugate of ι_{B_κ} is f_κ, we can apply [29, Theorem 5]. This Theorem gives an integral representation for the conjugate of integral functionals by means of the recession function f^∞, defined as $f^\infty(z) \stackrel{\text{def.}}{=} \sup_t f(x + tz)/t$, for x such that $f(x) < +\infty$. More precisely,

$$F^*(\mu) = \int_0^1 \int_\Omega f_\kappa \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{L}} \right) \mathrm{d}\mathscr{L} + \int_0^1 \int_\Omega f_\kappa^\infty \left(\frac{\mathrm{d}\mu}{\mathrm{d}|\mu^S|} \right) \mathrm{d}|\mu^S|$$

where $\mathscr{L}$ is the Lebesgue measure, μ^S is any singular measure which dominates the singular part of μ and f_κ^∞ is f_κ itself by homogeneity. Finally we obtain $F^*(\mu) = D_\kappa(\mu)$, which allows to recognize the opposite of $(\mathcal{P}_\mathcal{M})$ in the right-hand side of (12). The fact that the minimum is attained is part of the Fenchel-Rockafellar Theorem. $\qquad \square$

The following Theorem shows that we have defined a metric and provides two useful formulas.

Theorem 2.2. *WF_κ defines a metric on $\mathcal{M}_+(\Omega)$. Moreover, we have the equivalent characterizations, by changing the time range*

$$WF_\kappa(\rho_0, \rho_1) = \left\{ \inf_{\mu \in \mathcal{CE}_0^\tau(\rho_0, \rho_1)} \tau \int_{[0,\tau] \times \Omega} f_\kappa \left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \right) \mathrm{d}\lambda \right\}^{\frac{1}{2}} \tag{13}$$

with $\tau > 0$, and by disintegrating μ in time:

$$WF_\kappa(\rho_0, \rho_1) = \inf_{\mu \in \mathcal{CE}_0^1(\rho_0, \rho_1)} \int_0^1 \left\{ \int_\Omega f_\kappa \left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\lambda_t} \right) \mathrm{d}\lambda_t \right\}^{\frac{1}{2}} \mathrm{d}t \tag{14}$$

where $(\lambda_t)_{t \in [0,1]}$ is such that for all $t \in [0,1]$, $\mu_t \ll \lambda_t$.

Remark. Curves $(\rho_t)_{t \in [0,1]}$ associated to minimizers of $(\mathcal{P}_\mathcal{M})$ have the *constant speed* property, i.e. $WF_\kappa(\rho_s, \rho_t) = |t - s| WF_\kappa(\rho_0, \rho_1)$ and this is not necessarily the case for minimizers of (14) (for more details see [13, Theorem 5.4] where the proof of a similar result is given).

Proof. We obtain characterization (13), by remarking that if $T : t \mapsto \tau \cdot t$ is a time scaling and $\mu = (\rho, \omega, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ then $\tilde{\mu} = (T_\# \rho, \frac{1}{\tau} T_\# \omega, \frac{1}{\tau} T_\# \zeta) \in \mathcal{CE}_0^\tau(\rho_0, \rho_1)$. For the proof of (14), the arguments of [13, Theorem 5.4] apply readily. Let us now prove that WF_κ defines a metric. By definition, it is non-negative and it is symmetric because the functional satisfies $D_\kappa(\rho, \omega, \zeta) = D_\kappa(\rho, -\omega, -\zeta)$ so time can be reversed leaving D_κ unchanged. Finally, the triangle inequality comes from characterization (14) and the fact that the continuity equation is stable by concatenation in time i.e. for $\tau \in]0, 1[$, if $\mu_1 \in \mathcal{CE}_0^\tau(\rho_0, \rho_\tau)$ and $\mu_2 \in \mathcal{CE}_\tau^1(\rho_\tau, \rho_1)$ then $\mu_1 1_{[0,\tau]} + \mu_2 1_{[\tau,1]} \in \mathcal{CE}_0^1(\rho_0, \rho_1)$. $\qquad\square$

Proposition 2.2 (Homogeneity by mass rescaling)**.** *Let $\alpha > 0$. If ρ is a geodesic for $WF_\kappa(\rho_0, \rho_1)$ then $\alpha\rho$ is a geodesic for $WF_\kappa(\alpha\rho_0, \alpha\rho_1)$ and*

$$WF_\kappa(\alpha\rho_0, \alpha\rho_1) = \sqrt{\alpha} WF_\kappa(\rho_0, \rho_1)$$

Proof. It is a direct consequence of the homogeneity of D_κ. $\qquad\square$

2.3 Sufficient optimality and uniqueness conditions

We now leverage tools from convex analysis in order to provide a useful condition ensuring uniqueness of geodesics. We use this condition to study travelling Dirac solutions in Section 4.2.

Lemma 2.1. *The subdifferential of D_κ at a point $\mu = (\rho, \omega, \zeta)$ such that $D_\kappa(\mu) < +\infty$ is*

$$\partial D_\kappa(\mu) = \left\{ (\alpha, \beta, \gamma) \in C([0,1] \times \Omega; B_\kappa) : \alpha + \frac{1}{2} \left(|\beta|^2 + \frac{\gamma^2}{\kappa^2} \right) = 0 - \rho \text{ a.e.,} \right.$$
$$\left. \beta\rho = \omega \text{ and } \gamma\rho = \kappa^2\zeta \right\}. \tag{15}$$

Remark. This set is exactly the set of continuous functions of the form (α, β, γ) which take their values in ∂f_κ (defined in (10)) for all $(t, x) \in [0, 1] \times \Omega$.

Proof. Take $\tilde\mu \in \mathcal{M}^{d+2}$ and choose $\rho^\perp \in \mathcal{M}_+$ such that ρ and $\rho^\perp$ are mutually singular and $\lambda \overset{\text{def.}}{=} \rho + \rho^\perp$ satisfies $|\mu| + |\tilde\mu| \ll \lambda$. We have

$$
\begin{aligned}
D_\kappa(\tilde\mu) - D_\kappa(\mu) &= \int_{[0,1]\times\Omega} \left[f_\kappa\left(\frac{\mathrm{d}\tilde\mu}{\mathrm{d}\lambda}\right) - f_\kappa\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right) \right] \mathrm{d}\lambda \\
&= \int_{[0,1]\times\Omega} \left[f_\kappa\left(\frac{\mathrm{d}\tilde\mu}{\mathrm{d}\lambda}\right) - f_\kappa\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right) \right] \mathrm{d}\rho + \int_{[0,1]\times\Omega} f_\kappa\left(\frac{\mathrm{d}\tilde\mu}{\mathrm{d}\lambda}\right) \mathrm{d}\rho^\perp \\
&\overset{(*)}{\geq} \int_{[0,1]\times\Omega} \langle (\alpha,\beta,\gamma), \frac{\mathrm{d}\tilde\mu}{\mathrm{d}\lambda} - \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \rangle \mathrm{d}\rho + \int_{[0,1]\times\Omega} \langle (\alpha,\beta,\gamma), \frac{\mathrm{d}\tilde\mu}{\mathrm{d}\lambda} \rangle \mathrm{d}\rho^\perp \\
&= \langle (\alpha,\beta,\gamma), \tilde\mu - \mu \rangle
\end{aligned}
$$

and inequality $(*)$ holds for every $\tilde\mu$ if and only if for all $(t, x) \in [0, 1] \times \Omega$, $(\alpha, \beta, \gamma)(t, x) \in \partial f_\kappa\left(\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(t, x)\right)$: this comes from the definition of the subdifferential and from the fact the $f_\kappa = \iota^*_{B_\kappa}$. □

Theorem 2.3 (Sufficient optimality and uniqueness condition). *Let $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)^2$. If $\mu = (\rho, \omega, \zeta) \in \mathcal{CE}^1_0(\rho_0, \rho_1)$ and there exists $\varphi \in C^1([0, T] \times \Omega)$ such that*

$$
(\partial_t \varphi, \nabla\varphi, \varphi) \in \partial D_\kappa(\mu)
$$

then ρ is a geodesic for $WF_\kappa(\rho_0, \rho_1)$. If moreover, $\nabla\varphi$ is Lipschitz, then ρ is the unique geodesic.

Definition 5 (Optimality certificate). From now on, a function φ as in Theorem 2.3 is referred to as an optimality certificate.

Proof. With the notations of the proof of Theorem 2.1, Fenchel-Rockafellar duality also gives the following result: φ is a solution to $(\mathcal{P}_{C^1})$ if and only if there exists μ such that $A\varphi \in \partial F^*(\mu)$ and $-A^*\mu \in \partial G(\varphi)$, in which case μ is a solution to $(\mathcal{P}_\mathcal{M})$. The first condtion is our hypothesis and second is satisfied since for all $\psi \in C^1([0, 1] \times \Omega)$,

$$
\langle -A^*\mu, \psi - \varphi \rangle = \langle \mu, A\phi \rangle - \langle \mu, A\psi \rangle = G(\psi) - G(\phi)
$$

as $\mu \in \mathcal{CE}^1_0(\rho_0, \rho_1)$. This shows that μ is a geodesic for $WF_\kappa(\rho_0, \rho_1)$.

For uniqueness, consider another minimizer $\tilde\mu = (\tilde\rho, \tilde\omega, \tilde\zeta) \in \mathcal{CE}^1_0(\rho_0, \rho_1)$. It holds

$$
D_\kappa(\tilde\mu) \geq \int_0^1 \int_\Omega \partial_t \varphi \mathrm{d}\tilde\rho + \int_0^1 \int_\Omega \nabla\varphi \mathrm{d}\tilde\omega + \int_0^1 \int_\Omega \varphi \mathrm{d}\tilde\zeta \tag{16}
$$

$$
= \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0 \tag{17}
$$

$$
= D_\kappa(\tilde\mu) \tag{18}
$$

where we used successively: the duality inequality for D_κ, the fact that $\tilde\mu \in \mathcal{CE}^1_0(\rho_0, \rho_1)$ and the optimality of φ.

14

So, the first inequality is an equality, and hence λ a.e., $(\partial_t\varphi, \nabla\varphi, \varphi)(t,x) \in \partial f_\kappa\left(\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\lambda}(t,x)\right)$ for λ such that $\tilde{\mu} \ll \lambda$. And thus, from the characterization of ∂f_κ, we have

$$\tilde{\mu} = (\tilde{\rho}, (\nabla\varphi)\tilde{\rho}, \frac{\varphi}{\kappa^2}\tilde{\rho}).$$

Thus, ρ and $\tilde{\rho}$ are both solutions of $\partial_t\rho + \nabla\cdot((\nabla\varphi)\rho) = \frac{\varphi}{\kappa^2}\rho$ with initial condition ρ_0. This equation has a unique solution if $\nabla\varphi$ is Lipschitz. $\qquad\square$

Remark (Relaxation). The above condition is not necessary: the optimality certificate φ is sometimes not as smooth as required by this condition. It is an interesting problem to study the space in which the infimum of $(\mathcal{P}_{C^1})$ is always attained. For instance, in the case of dynamical optimal transport, this space is that of Lipschitz functions [19]. In a different setting, for a class of distances which interpolate between optimal transport and Sobolev, [10] proves that the optimum is attained in $BV \cap L^\infty$, where BV is the set of functions with bounded variation. This question is non-trivial and beyond the scope of the present paper.

3   Interpolation properties

In this section we study the effect of varying κ on geodesics. This allows us to explain to which extent our metric really interpolates between optimal transport and Fisher-Rao.

3.1   Change of scale

Intuitively, the smaller κ, the finer the scale at which mass creation and removal intervene to make initial and final measures match. Accordingly, we show in the following Proposition that changing κ amounts to changing the scale of the problem.

Proposition 3.1 (Space rescaling). *Let $T : (t,x) \in \mathbb{R} \times \mathbb{R}^d \mapsto (t, s\,x)$ be the spatial scaling by a factor $s > 0$, and $\kappa \in]0, +\infty[$. If $\mu = (\rho, \omega, \zeta)$ is a minimizing triplet for the distance $WF_\kappa(\rho_0, \rho_1)$, then $\hat{\mu} = (T_\#\rho, s\,T_\#\omega, T_\#\zeta)$ is a minimizing triplet for the distance $WF_{s\,\kappa}(T_\#\rho_0, T_\#\rho_1)$.*

Proof. It is easy to check that $\hat{\mu} = (\hat{\rho}, \hat{\omega}, \hat{\zeta})$ satisfies the continuity equation with Neumann boundary conditions on $T(\Omega)$ between $T_\#\rho_A$ and $T_\#\rho_B$. Also, we can write

$$D_{s\kappa}(\hat{\mu}) = \int_{T([0,1]\times\Omega)} f_{s\kappa}\left(\frac{d\hat{\mu}}{d\hat{\lambda}}\right) \mathrm{d}\hat{\lambda}$$

where $\hat{\lambda} = T_\#\lambda$ and λ is any non-negative Borel measure satisfying $|\mu| \ll \lambda$. By denoting $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = (\rho^\lambda, \omega^\lambda, \zeta^\lambda)$ we have $\frac{d\hat{\mu}}{d\hat{\lambda}}(t,x) = (\rho^\lambda(t,x/s), s\omega^\lambda(t,x/s), \zeta^\lambda(t,x/s))$. Thus by the change of variables formula, we obtain

$$\begin{aligned} D_{s\kappa}(\hat{\mu}) &= \int_{[0,1]\times\Omega} f_{s\kappa}\left(\rho^\lambda, s\,\omega^\lambda, \zeta^\lambda\right) \mathrm{d}\lambda \\ &= s^2 D_\kappa(\mu). \end{aligned}$$

This shows that if μ minimizes D_κ then $\hat{\mu}$ minimizes $D_{s\kappa}$ and vice versa, hence Proposition 3.1. $\qquad\square$

Remark. The proof of Proposition 3.1 holds true as soon as f_κ is of the form $f_1(\rho, \omega, \zeta) + \kappa^p f_2(\rho, \zeta)$ where f_1 is p-homogeneous in ω and f_2 is independent of ω. Hence the choice of κ^p as an interpolation factor when introducing the various models in Section 1.3: in this way, κ can be interpreted as a spatial scale.

The following result is closely related and will be used later to prove uniqueness of geodesics between remote Dirac measures.

Proposition 3.2 (Monotonicity of the distance w.r.t. rescaling)**.** *Using the same notations as in the previous result, if $s < 1$, then $D_\kappa(\hat{\mu}) \leq D_\kappa(\mu)$ and thus $WF_\kappa(T_\#\rho_0, T_\#\rho_1) \leq WF_\kappa(\rho_0, \rho_1)$. Moreover, if $s < 1$ and the momentum associated to a geodesic between ρ_0 and ρ_1 is not zero, these inequalities are strict.*

Proof. We use the same change of variables as in the proof of Proposition 3.1 and remark that f_κ is strictly increasing w.r.t. its second variable. $\qquad\square$

3.2 Two variational problems on non-negative measures

The main Theorem of section 3.3 states that the geodesics of the quadratic Wasserstein and the Hellinger distances are both recovered (in a suitable sense) when letting the parameter κ go to $+\infty$ and 0, respectively. We start by defining two variational problems, before showing their connections to WF_κ in the next subsection.

Let us introduce

$$D_{BB} : (\rho, \omega) \mapsto D_\kappa(\rho, \omega, 0) \quad \text{and} \quad D_{FR} : (\rho, \zeta) \mapsto \frac{1}{\kappa^2} D_\kappa(\rho, 0, \zeta).$$

Those real valued functionals represent respectively the Benamou-Brenier and the Fisher-Rao terms in the interpolating functional D_κ. They do not depend on κ and are defined such that $D_\kappa(\rho, \omega, \zeta) = D_{BB}(\rho, \omega) + \kappa^2 D_{FR}(\rho, \zeta)$.

Definition 6 (Generalized Benamou-Brenier transport problem)**.** Given $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)$, we introduce the time dependent *rate of growth*

$$g : (t, x) \mapsto \begin{cases} 0 & \text{if } \rho_0(\Omega) = \rho_1(\Omega) \\ \frac{2}{t - t_0} & \text{otherwise} \end{cases} \tag{19}$$

with $t_0 = \sqrt{\rho_0(\Omega)} \left(\sqrt{\rho_0(\Omega)} - \sqrt{\rho_1(\Omega)} \right)^{-1}$. The generalized Benamou-Brenier transport problem is defined as

$$\inf_{\rho, \omega} \quad D_{BB}(\rho, \omega) \tag{20}$$

$$\text{subject to} \quad (\rho, \omega, g\rho) \in \mathcal{CE}_0^1(\rho_0, \rho_1) \,. \tag{21}$$

and we denote by $d_{gBB}(\rho_0, \rho_1)$ the square-root of the infimum.

Notice that d_{gBB} belongs to the class of models for unbalanced transport of [22] which prescribe the source term ζ as a function of ρ. In order to picture what is happening, consider $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)$ such that $\rho_0(\Omega) < \rho_1(\Omega)$: in that case, $t_0 < 0$ and the rate of growth g is positive, constant in space and decreasing in time. While d_{gBB} does not define a metric as it does not satisfy the triangle inequality, we prove below that it is closely related to the quadratic Wasserstein metric.

Proposition 3.3 (Relation between d_{gBB} and W_2). *Let ρ_0, $\rho_1 \in \mathcal{M}_+(\Omega)$. We introduce $\tilde{\rho}_0$, $\tilde{\rho}_1$ the rescaled measures whose masses are the geometric mean between $\rho_0(\Omega)$ and $\rho_1(\Omega)$, i.e. such that for $i \in \{0, 1\}$, $\tilde{\rho}_i(\Omega) = \sqrt{\rho_0(\Omega)\rho_1(\Omega)}$ and $\tilde{\rho}_i = \alpha_i \rho_i$ for some $\alpha_i \geq 0$. It holds*

$$d_{gBB}(\rho_0, \rho_1) = \frac{1}{2} W_2(\tilde{\rho}_0, \tilde{\rho}_1),$$

Moreover, the minimum in (20) *is attained and minimizers can be built explicitly from the geodesics of $W_2(\tilde{\rho}_0, \tilde{\rho}_1)$.*

Proof. If $\rho_0 = 0$ or $\rho_1 = 0$, the minimum is attained if and only if $\omega = 0$ and thus $d_{gBB}(\rho_0, \rho_1) = 0$. If $\rho_0(\Omega) = \rho_1(\Omega)$ then $g = 0$ and the conclusion follows directly. Otherwise, we have $t_0 \notin [0, 1]$ and we will reduce the problem to the case of a standard Benamou-Brenier problem by a rescaling and a change of variables in time, similarly as in [22, Proposition 7]. Consider

$$\mathcal{S} = \left\{ (\rho, \omega) : (\rho, \omega, g\rho) \in \mathcal{CE}_0^1(\rho_0, \rho_1), \ |\omega| \ll \rho \right\}$$

$$\text{and} \quad \mathcal{S}_0 = \left\{ (\rho, \omega) : (\rho, \omega, 0) \in \mathcal{CE}_0^1(\rho_0, \frac{\rho_0(\Omega)}{\rho_1(\Omega)}\rho_1), \ |\omega| \ll \rho \right\}.$$

Take $\mu = (\rho, \omega)$ in $\mathcal{S}$. It satisfies (weakly) the ordinary differential equation $\partial_t \rho_t(\Omega) = g(t)\rho_t(\Omega)$ and so $\rho_t(\Omega) = \left(\frac{t_0 - t}{t_0}\right)^2 \rho_0(\Omega)$ a.e. in $[0, 1]$. We define the rescaling

$$R : \mathcal{S} \mapsto \mathcal{S}_0,$$

$$R(\rho, \omega) = \left(\frac{t_0}{t_0 - t}\right)^2 (\rho, \omega).$$

which is a bijection as $t_0 \notin [0, 1]$. Writing $D_{BB}(R(\rho, \omega))$ yields a Benamou-Brenier functional with a time varying metric; we now counterbalance the latter by a change of variables in time. Consider the strictly increasing map $s : t \mapsto t\frac{t_0 - 1}{t_0 - t}$ which satisfies $s(0) = 0$, $s(1) = 1$ and $s'(t) = \frac{t_0 - 1}{t_0}\left(\frac{t_0}{t_0 - t}\right)^2 > 0$. Let $t = s^{-1}$ and introduce

$$T : \mathcal{S}_0 \mapsto \mathcal{S}_0$$

$$T(\rho, \omega) = (\rho \circ t(s), t'(s) \cdot \omega \circ t(s))$$

which is a bijection (see [1, Lemma 8.1.3]). The rescaling followed by the change of variables in time induces a bijection $T \circ R : \mathcal{S} \to \mathcal{S}_0$. Of course, the expression

for s has been chosen after an analysis of its effect on the D_{BB} functional. Take $(\tilde{\rho}, \tilde{\omega}) = R(\rho, \omega) \in \mathcal{S}_0$ and $(\hat{\rho}, \hat{\omega}) = T \circ R(\rho, \omega) \in \mathcal{S}_0$. All those measures can be disintegrated in time from the definition of $\mathcal{S}$ and $\mathcal{S}_0$. In order to alleviate notations, we introduce $f_{BB}(\rho_t, \omega_t) = \int_\Omega f_\kappa \left(\frac{d\rho}{d\lambda}, \frac{d\omega_t}{d\lambda_t}, 0 \right) d\lambda_t$, where $\lambda_t \in \mathcal{M}_+(\Omega)$ satisfies $|(\rho_t, \omega_t)| \ll \lambda_t$. We have

$$
\begin{aligned}
\int_0^1 f_{BB}(\hat{\rho}_s, \hat{\omega}_s) \mathrm{d}s &= \int_0^1 (\mathrm{t}' \circ \mathrm{s}(t))^2 f_{BB}(\tilde{\rho}_t, \tilde{\omega}_t) \mathrm{s}'(t) \mathrm{d}t \\
&= \int_0^1 \frac{f_{BB}(\tilde{\rho}_t, \tilde{\omega}_t)}{\mathrm{s}'(t)} \mathrm{d}t \\
&= \frac{t_0}{t_0 - 1} \int_0^1 \left(\frac{t_0 - t}{t_0} \right)^2 f_{BB}(\tilde{\rho}_t, \tilde{\omega}_t) \mathrm{d}t \\
&= \left(\frac{\rho_0(\Omega)}{\rho_1(\Omega)} \right)^{1/2} \int_0^1 f_{BB}(\rho, \omega) \mathrm{d}t
\end{aligned}
$$

Hence the relation

$$
D_{BB}(\rho, \omega) = \left(\frac{\rho_1(\Omega)}{\rho_0(\Omega)} \right)^{1/2} D_{BB}(T \circ R(\rho, \omega)),
$$

from which we deduce $d_{gBB}(\rho_0, \rho_1) = \frac{1}{2} \left(\frac{\rho_1(\Omega)}{\rho_0(\Omega)} \right)^{1/4} W_2(\rho_0, \frac{\rho_0(\Omega)}{\rho_1(\Omega)} \rho_1)$, as the minimum of the standard Benamou-Brenier problem is attained in $\mathcal{S}_0$. Finally, by remarking that $W_2(\alpha\rho_0, \alpha\rho_1) = \sqrt{\alpha} W_2(\rho_0, \rho_1)$ for $\alpha \geq 0$, we obtain the formulation of Proposition 3.3. $\qquad \square$

Definition 7 (Hellinger distance). Given $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)$, the Hellinger distance is defined as

$$
d_{FR}^2(\rho_0, \rho_1) \overset{\text{def.}}{=} \inf_{\rho, \zeta} \quad D_{FR}(\rho, \zeta)
$$
$$
\text{subject to} \quad (\rho, 0, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1).
$$

We now prove uniqueness of Hellinger geodesics and give their explicit expression. This result is well known in a positive, smooth setting.

Theorem 3.1. *The geodesics for the Hellinger distance are unique and have the explicit expression*

$$
\rho = (t\sqrt{\rho_1} + (1-t)\sqrt{\rho_0})^2 \otimes \mathrm{d}t. \tag{22}
$$

Proof. First notice that the expression (22) is not ambiguous since it is positively homogeneous as a function of (ρ_0, ρ_1). The fact that d_{FR} defines a metric is proven using the same arguments as in Theorem 2.2. As for the proof of uniqueness of geodesics, we organize the proof as follows. Using convex duality, we show that (22) is a geodesic among geodesics dominated at all times by some measure $\nu \in \mathcal{M}_+(\Omega)$: this allows to exhibit an isometric injection (the square root) into $L^2(\mathrm{d}\nu)$, from which we deduce uniqueness.

i. Duality. Written explicitly, the constraint in Definition 7 is equivalent to

$$\int_0^1 \int_\Omega \partial_t \varphi \mathrm{d}\rho + \int_0^1 \int_\Omega \varphi \mathrm{d}\zeta = \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0 \qquad (23)$$

for all test functions $\varphi \in C^1([0, 1] \times \Omega)$. Similarly as in Theorem 2.1, we obtain that the problem defining the Hellinger metric is the dual of

$$\sup_{\varphi \in C^1([0,1] \times \Omega)} \int_\Omega \varphi(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot) \mathrm{d}\rho_0$$

subject to, for all $(t, x) \in [0, 1] \times \Omega$,

$$(\partial_t \varphi(t, x), \varphi(t, x)) \in B_{FR} \quad \text{where} \quad B_{FR} = \left\{ (a, b) : a + \frac{1}{2} b^2 \le 0 \right\}.$$

By Fenchel-Rockafellar Theorem, we obtain the existence of minimizers and standard arguments based on reparametrization (see [13, Theorem 5.4] for instance) show that geodesics are constant speed and that d_{FR} defines a metric on $\mathcal{M}_+$.

ii. Value of d_{FR}. Let $\nu \in \mathcal{M}_+(\Omega)$ be any measure which dominates ρ_0 and ρ_1 and denote $\rho_0 = r_0 \nu$, $\rho_1 = r_1 \nu$. Assume for the moment that there exists $\kappa > 0$ such that ν a.e.,

$$\frac{1}{\kappa} \le \frac{r_0}{r_1} \le \kappa. \qquad (A1)$$

Consider $\mu = (\rho, \zeta)$ where ρ is defined as in (22), i.e.

$$\rho = (t\sqrt{r_1} + (1 - t)\sqrt{r_0})^2 \nu \otimes \mathrm{d}t$$
$$\zeta = 2(\sqrt{r_1} - \sqrt{r_0})(t\sqrt{r_1} - (1 - t)\sqrt{r_0})\nu \otimes \mathrm{d}t.$$

This couple satisfies the constraints (23) and $D_{FR}(\mu) = 2 \int_\Omega (\sqrt{\rho_1} - \sqrt{\rho_0})^2$. Now consider

$$\tilde{\varphi} \stackrel{\text{def.}}{=} \zeta/\rho = \begin{cases} 0 & \text{if } r_0(x) = r_1(x), \\ \frac{2}{t - t_0(x)} & \text{otherwise.} \end{cases}$$

with $t_0(x) = \sqrt{r_0} \left(\sqrt{r_0} - \sqrt{r_1} \right)^{-1}$. Under the assumption (A1), there exists $\epsilon > 0$ such that for all $x \in \Omega$, $t_0(x) \notin [-\epsilon, 1 + \epsilon]$ and thus, for all $t \in [0, 1]$, $\tilde{\varphi}(t, \cdot)$ is bounded and belongs to $L^1(\mathrm{d}\nu)$. Also, for all $x \in \Omega$, $\tilde{\varphi}(\cdot, x) \in C^1([0, 1])$. But, in general, $\tilde{\varphi} \notin C^1([0, 1] \times \Omega)$ as it lacks regularity w.r.t. the space variable.

We introduce thus its regularized version: $\varphi^\epsilon \stackrel{\text{def.}}{=} \eta^\epsilon * \tilde{\varphi}$, where $\eta^\epsilon(t, x) \stackrel{\text{def.}}{=} \epsilon^{-d}\alpha(x/\epsilon)$ with $\alpha \in C_c^\infty((-1/2, 1/2)^d)$, $\alpha \ge 0$, $\int \alpha = 1$ and α even. By convexity of B_{FR}, we have for all $(t, x) \in [0, 1] \times \Omega$, $(\partial_t \tilde{\varphi}, \tilde{\varphi})(t, x) \in B_{FR} \Rightarrow (\eta^\epsilon * \partial_t \tilde{\varphi}, \eta^\epsilon * \tilde{\varphi})(t, x) \in B_{FR} \Rightarrow (\partial_t \varphi^\epsilon, \varphi^\epsilon)(t, x) \in B_{FR}$. Thus

$$d_{FR}^2(\rho_0, \rho_1) \ge \lim_{\epsilon \to 0} \int_\Omega \varphi^\epsilon(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \varphi^\epsilon(0, \cdot) \mathrm{d}\rho_0$$

$$= \int_\Omega \tilde{\varphi}(1, \cdot) \mathrm{d}\rho_1 - \int_\Omega \tilde{\varphi}(0, \cdot) \mathrm{d}\rho_0$$

$$= 2 \int_\Omega (\sqrt{\rho_1} - \sqrt{\rho_0})^2.$$

As this value is also an upper bound, this shows that, under (A1), $d_{FR}^2(\rho_0, \rho_1) = 2\int_\Omega(\sqrt{\rho_1} - \sqrt{\rho_0})^2$. But this result remains true without any assumption on ρ_0 and ρ_1. Indeed, by introducing $\rho_1^\epsilon = \epsilon\rho_0 + (1-\epsilon)\rho_1$ and $\rho_0^\epsilon = \epsilon\rho_1 + (1-\epsilon)\rho_0$, the triangle inequality yields

$$d_{FR}(\rho_0^\epsilon, \rho_1^\epsilon) - d_{FR}(\rho_1^\epsilon, \rho_1) \leq d_{FR}(\rho_0^\epsilon, \rho_1) \leq d_{FR}(\rho_0^\epsilon, \rho_1^\epsilon) + d_{FR}(\rho_1^\epsilon, \rho_1)$$

and $\lim_{\epsilon\to 0} d_{FR}(\rho_1^\epsilon, \rho_1) = 0$ because we have a vanishing upper bound. Hence $d_{FR}^2(\rho_0^\epsilon, \rho_1) = 2\int_\Omega(\sqrt{\rho_1} - \sqrt{\rho_0^\epsilon})^2$. Repeating this operation for $\rho_0^\epsilon \to \rho_0$ we have that $d_{FR}^2(\rho_0, \rho_1) = 2\int_\Omega(\sqrt{\rho_1} - \sqrt{\rho_0})^2$ for $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$.

iii. Isometric injection and uniqueness. Finally, on the space $\mathcal{M}_\nu = \{\rho \in \mathcal{M}_+(\Omega) : \rho \ll \nu\}$, the map

$$\left|\begin{array}{ccc} (\mathcal{M}_\nu, d_{FR}) & \to & (L^2(\mathrm{d}\nu), \|\cdot\|_2) \\ \rho = \alpha\,\nu & \mapsto & \sqrt{2\alpha} \end{array}\right.$$

is an isometric injection. As geodesics in $L^2(\mathrm{d}\nu)$ are unique, uniqueness holds for geodesics in $\mathcal{M}_\nu$. Now let $(\rho_t)_{t\in[0,1]}$ be any geodesic in $\mathcal{M}_+(\Omega)$ (finiteness of the functional implies that ρ can be disintegrated in time). Fix a time $\tau \in [0, 1]$, and let $\nu' \in \mathcal{M}_+(\Omega)$ be a measure which dominates ρ_0, ρ_1, ρ_τ and λ. From the previous construction, and because geodesics are constant speed, one can construct (by concatenation and time reascaling) a geodesic which takes the value ρ_τ time τ and which is dominated by ν' at all times. By uniqueness of geodesics dominated by ν, this geodesic is actually dominated by ν (because there exists one such geodesic). This shows that there is no other geodesic than (22). □

The following Lemma motivates the expressions for d_{gBB} and d_{FR}. It tells that one obtains d_{gBB} by minimizing a "growth" and a "transport" problem successively (and vice-versa for and d_{FR}), as opposed to $W\!F$ where the minimization is performed "simultaneously".

Lemma 3.1 (Alternative characterizations). *For all $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)$ we have*

$$d_{gBB}(\rho_0, \rho_1)^2 = \min_{(\rho,\omega)\in\mathcal{A}_{FR}} D_{BB}(\rho, \omega)$$

$$d_{FR}(\rho_0, \rho_1)^2 = \min_{(\rho,\zeta)\in\mathcal{A}_{BB}} D_{FR}(\rho, \zeta)$$

where

$$\mathcal{A}_{FR} \stackrel{\text{def.}}{=} \operatorname*{argmin}_{(\rho,\omega,\zeta)\in\mathcal{CE}_0^1(\rho_0,\rho_1)} D_{FR}(\rho, \zeta) \quad and \quad \mathcal{A}_{BB} \stackrel{\text{def.}}{=} \operatorname*{argmin}_{(\rho,\omega,\zeta)\in\mathcal{CE}_0^1(\rho_0,\rho_1)} D_{BB}(\rho, \omega).$$

$$(24)$$

Proof. What we need to show is

$$\mathcal{A}_{BB} = \{(\rho, \zeta) : (\rho, 0, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)\}$$
$$\mathcal{A}_{FR} = \{(\rho, \omega) : (\rho, \omega, g\rho) \in \mathcal{CE}_0^1(\rho_0, \rho_1)\},$$

20

where the *rate of growth* g is defined in (19). The first equality is easy because $D_{BB}(\rho, \omega) = 0$ if and only if $\omega = 0$ a.e.

The second equality requires slightly more work. In the case when $\rho_0(\Omega) = \rho_1(\Omega)$, D_{FR} is minimized if and only if $\zeta = 0$ as previously. The case $\rho_0(\Omega) = 0$ or $\rho_1(\Omega) = 0$ is dealt with in Proposition 4.2 (case $\alpha = 0$). Otherwise, let us determine $\mathcal{A}_{FR}$ by adapting the optimality condition in Theorem 2.3. The argument is clearer if we consider D_{FR} as a function of $\mu = (\rho, \omega, \zeta)$ instead of just (ρ, ζ) as ω is a variable of the problem. We thus introduce

$$B_{FR} \stackrel{\text{def.}}{=} \left\{ (\alpha, \beta, \gamma) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} : \alpha + \frac{\gamma^2}{2} \leq 0, \ \beta = 0 \right\}$$

and it can be shown (similarly as in Section 2.3) that $\partial D_{FR}(\rho, \omega, \zeta)$ is equal to

$$\left\{ (\alpha, \beta, \gamma) \in C([0,T] \times \Omega; B_{FR}) : \alpha + \frac{\gamma^2}{2} = 0 - \ \rho \text{ a.e. and } \gamma\rho = \zeta \ \right\}.$$

Adapting the sufficient optimality condition given in Theorem 2.3, it holds that if $\mu \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ and if there exists $\varphi \in C^1([0,1] \times \Omega)$ satisfying $(\partial_t \varphi, \nabla \varphi, \varphi) \in \partial D_{FR}(\mu)$ i.e.

$$\begin{cases} \partial_t \varphi + \frac{1}{2}\varphi^2 \leq 0 \quad \text{(with equality } \rho \text{ a.e.)}, \\ \nabla \varphi = 0, \\ \varphi\rho = \zeta. \end{cases}$$

then $\mu \in \mathcal{A}_{FR}$. The equality ρ a.e. is actually an equality $\mathrm{d}t$ a.e. in time since ρ has a positive mass and φ is constant in space. All solutions are thus of the form $\varphi : (t,x) \mapsto \frac{2}{t-t_0}$. The integration constant t_0 is found by integrating the variations of mass. As $t_0 \notin [0,1]$, φ is well defined and is equal to g as introduced in (19). Now we prove that this condition is actually necessary: take $(\rho, \omega, \zeta) \in \mathcal{A}_{FR}$. It holds

$$D_{FR}(\rho, \omega, \zeta) \geq \int_0^1 \int_\Omega \partial_t g \rho + g\zeta = \int_\Omega g(1)\rho_1 - \int_\Omega g(0)\rho_0 = D_{FR}(\rho, \omega, \zeta)$$

where we used successively: duality, the fact that $(\rho, \omega, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ and the fact that the primal-dual gap vanishes at optimality. So, the first equality is an equality, and hence $(\partial_t g, g)(t,x) \in \partial f_1(\mathrm{d}\rho/\mathrm{d}|\rho|, 0, \mathrm{d}\zeta/\mathrm{d}|\rho|)(t,x)$ for $|\rho|$ a.e. $(t,x) \in [0,1] \times \Omega$, which implies that $\zeta = g\rho$. $\qquad\square$

3.3 Convergence of geodesics

A rigorous treatment of the *limit geodesics* requires some compactness for the set of geodesics when κ varies. That is why we need the following bounds, adapted from [13, Proposition 3.10].

Lemma 3.2. *Consider a triplet* $\mu = (\rho, \omega, \zeta) \in \mathcal{M}^{d+2}$ *such that* $D_{BB}(\rho, \omega) < +\infty$ *and* $D_{FR}(\rho, \zeta) < +\infty$. *For any non-negative Borel function* η *on* $[0,1] \times \Omega$ *we have*

$$\int_{[0,1]\times\Omega} \eta \, d|\omega| \leq \sqrt{2} D_{BB}(\rho, \omega)^{\frac{1}{2}} \left(\int_{[0,1]\times\Omega} \eta^2 \mathrm{d}\rho \right)^{\frac{1}{2}} \tag{25}$$

and

$$\int_{[0,1]\times\Omega} \eta d|\zeta| \leq \sqrt{2} D_{FR}(\rho,\zeta)^{\frac{1}{2}} \left(\int_{[0,1]\times\Omega} \eta^2 d\rho\right)^{\frac{1}{2}}. \tag{26}$$

Also, similar bounds can be written for μ_t (the disintegration of μ w.r.t. time) by integrating solely in space.

Proof. Take $\lambda \in \mathcal{M}_+$ such that $|\mu| \ll \lambda$ and decompose $d\mu$ as $(\rho^\lambda, \omega^\lambda, \zeta^\lambda)d\lambda$. As $D_{BB}(\rho,\omega) < +\infty$ and $D_{FR}(\rho,\zeta) < +\infty$, we have $\omega \ll \rho$ and $\zeta \ll \rho$. It follows

$$\int_{[0,1]\times\Omega} \eta d|\omega| = \int_{[0,1]\times\Omega} \left(2\eta^2 f_\kappa(\rho^\lambda, \omega^\lambda, 0)\rho^\lambda\right)^{\frac{1}{2}} d\lambda$$

$$\leq \sqrt{2}\left(\int_{[0,1]\times\Omega} \eta^2 \rho^\lambda d\lambda\right)^{\frac{1}{2}} \left(\int_{[0,1]\times\Omega} f_\kappa(\rho^\lambda, \omega^\lambda, 0)d\lambda\right)^{\frac{1}{2}}$$

by Cauchy-Schwartz inequality on the scalar product $\langle\cdot,\cdot\rangle_{L^2(d\lambda)}$. Inequality (25) follows and (26) is derived similarly. $\square$

In the following lemma, we derive a total variation bound on μ which only depends on $D_{BB}(\mu)$ and $D_{FR}(\mu)$.

Lemma 3.3 (Uniform bound). *Let $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$ and $M \in \mathbb{R}_+$. There exists $C \in \mathbb{R}_+$ satisfying $|\mu|([0,1]\times\Omega) < C$, for all $\mu = (\rho,\omega,\zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ such that $D_{BB}(\rho,\omega) < M$ and $D_{FR}(\rho,\zeta) < M$.*

Proof. We start by giving a bound on $\rho([0,1]\times\Omega) = \int_0^1 \rho_t(\Omega)dt$. Recall (from the remarks after Definition 1) that the map $t \mapsto \rho_t(\Omega)$ admits the distributional derivative $\rho_t'(\Omega) = \zeta_t(\Omega)$, for almost every $t \in [0,1]$. It follows, by Lemma 3.2

$$\rho_t'(\Omega) \leq |\zeta_t|(\Omega) \leq \sqrt{2D_{FR}(\rho_t, \zeta_t)\rho_t(\Omega)}$$

where $D_{FR}(\mu_t)$ denotes $\frac{1}{\kappa^2}\int_\Omega f_\kappa(\frac{d(\rho_t, 0, \zeta_t)}{d\lambda_t})d\lambda_t$ where λ_t is such that $\mu_t \ll \lambda_t$. By integrating in time and applying the Cauchy-Schwartz inequality

$$\rho_t(\Omega) - \rho_0(\Omega) \leq \sqrt{2D_{FR}(\rho,\zeta)}\sqrt{\rho([0,1]\times\Omega)}.$$

We integrate again in time to obtain

$$\rho([0,1]\times\Omega) \leq \rho_0(\Omega) + \sqrt{2D_{FR}(\rho,\zeta)}\sqrt{\rho([0,1]\times\Omega)}$$

which implies that $\rho([0,1]\times\Omega)$ is bounded by a constant depending only on $D_{FR}(\rho,\zeta)$ and $\rho_0(\Omega)$. The conclusion follows by bounding $|\zeta|([0,1]\times\Omega)$ and $|\omega|([0,1]\times\Omega)$ thanks to Lemma 3.2. $\square$

Let us now recall a well known property of weighted optimization problems.

Lemma 3.4 (Properties of weighted optimization problems). *Let f and g be two proper l.s.c. functions on a compact set $\mathcal{C}$ with values in $\mathbb{R} \cup \{+\infty\}$ such that $h = f + \kappa g$ admits a minimum in $\mathcal{C}$ for all $\kappa \in]0, \pm\infty[$. Let $(\kappa_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ and $(x_n)_{n\in\mathbb{N}}$ a sequence of minimizers of $h_n = f + \kappa_n g$. We introduce $\mathcal{A}(f) \overset{def.}{=} \operatorname{argmin}_{x\in\mathcal{C}} f(x)$ and $\mathcal{A}(g)$ similarly.*

1. *For all $n \in \mathbb{N}$, $x^f \in \mathcal{A}(f)$ and $x^g \in \mathcal{A}(g)$ we have $f(x_n) \leq f(x^g)$ and $g(x_n) \leq g(x^f)$,*

2. *If $\kappa_n \underset{n \to \infty}{\to} 0$, then $(x_n)_{n \in \mathbb{N}}$ admits an accumulation point in $\mathrm{argmin}_{\mathcal{A}(f)}\, g$,*

3. *If $\kappa_n \underset{n \to \infty}{\to} \infty$, then $(x_n)_{n \in \mathbb{N}}$ admits an accumulation point in $\mathrm{argmin}_{\mathcal{A}(g)}\, f$.*

Proof. Those results are derived using elementary inequality manipulations. $\square$

The following lemma gives total variation bounds on geodesics that are independent of κ, which will be helpful in order to explicit the limit models.

Lemma 3.5 (A relative compactness result). *Let ρ_0 and ρ_1 in $\mathcal{M}_+(\Omega)$ and consider $\mathcal{G}^{\rho_1}_{\rho_0} \overset{\text{def.}}{=} \bigcup_{\kappa>0} \arg\min D_\kappa(\rho_0, \rho_1)$. For all $\mu = (\rho, \omega, \zeta) \in \mathcal{G}^{\rho_1}_{\rho_0}$, the following bounds, independent of κ, hold*

$$D_{BB}(\rho, \omega) \leq d^2_{gBB}(\rho_0, \rho_1), \tag{27}$$
$$D_{FR}(\rho, \zeta) \leq d^2_{FR}(\rho_0, \rho_1). \tag{28}$$

Moreover, $\mathcal{G}^{\rho_1}_{\rho_0}$ is relatively compact.

Proof. Inequalities (27) and (28) are direct applications of Lemma 3.4-1, to the characterizations of d_{gBB} and d_{FR} given in Lemma 3.1. As these inequalities do not depend on κ, we can deduce from Lemma 3.3 a uniform TV bound on the elements of $\mathcal{G}^{\rho_1}_{\rho_0}$. More explicitly, there exists $C > 0$ such that for all $\mu \in \mathcal{G}^{\rho_1}_{\rho_0}$,

$$|\mu|([0,1] \times \Omega) < C$$

This implies that $\mathcal{G}^{\rho_1}_{\rho_0}$ is relatively compact for the weak* topology. $\square$

We can now state the main result of this section.

Theorem 3.2 (Limit models). *Let ρ_0 and ρ_1 in $\mathcal{M}_+(\Omega)$ and let $(\rho^n, \omega^n, \zeta^n)_{n \in \mathbb{N}}$ be a sequence of minimizers for the distance WF_{κ_n} between ρ_0 and ρ_1 with $(\kappa_n)_n \in \mathbb{R}^{\mathbb{N}}_+$.*
    A. If $\kappa_n \underset{n \to \infty}{\to} +\infty$ then $(\rho^n, \omega^n, \zeta^n)_{n \in \mathbb{N}}$ weak converges (up to a subsequence) to a minimizer for d_{gBB}.*
    B. If $\kappa_n \underset{n \to \infty}{\to} 0$ then $(\rho^n, \omega^n, \zeta^n)_{n \in \mathbb{N}}$ weak converges (up to a subsequence) to a minimizer for d_{FR}.*

Proof. Let us review the hypothesis of Lemma 3.4 . First, D_{BB} and D_{FR} are l.s.c. The existence of minimizers for D_κ has been shown in Theorem 2.1 for any positive value of κ. Moreover, as a consequence of Lemma 3.5, those minimizers are in a compact set—just take any compact set which contains the closure of the relatively compact set of all minimizers. All conditions are gathered for Lemma 3.4-(2,3) to be applied, hence the result. $\square$

4 Explicit geodesics

In this section, we give an explicit description of the behavior of geodesics in three cases: for an "inflating" measure, for the transport of one Dirac measure to another, and finally for the transport of multiple couples of Dirac measures. In order to establish explicit geodesics, we will first exhibit an ansatz and then prove the existence of an optimality certificate as defined in Theorem 2.3.

4.1 No transport case: inflating and deflating measures

The case when there is no transport is dealt with in the following Propositions.

Proposition 4.1 (A uniqueness result). *Let $(\rho_0, \rho_1) \in \mathcal{M}_+(\Omega)^2$. If $(\rho, 0, g\rho)$ is a minimizer of $(\mathcal{P}_\mathcal{M})$, where g is the homogeneous rate of growth defined in (19), then ρ is the* unique *geodesic.*

Proof. Assuming $(\rho, 0, g\rho)$ minimizes $(\mathcal{P}_\mathcal{M})$, we have $WF_\kappa^2(\rho_0, \rho_1) = \kappa^2 D_{FR}(\rho, \zeta)$. Moreover, $(\rho, 0, g\rho) \in \mathrm{argmin}_{\mu \in \mathcal{CE}_0^1(\rho_0, \rho_1)} D_{FR}(\rho, \zeta)$. Thus for any $\tilde{\mu} = (\tilde{\rho}, \tilde{\omega}, \tilde{\zeta}) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$, $D_\kappa(\tilde{\mu}) = D_{BB}(\tilde{\rho}, \tilde{\omega}) + \kappa^2 D_{FR}(\tilde{\rho}, \tilde{\zeta}) \geq D_{BB}(\rho, \omega) + WF_\kappa^2(\rho_0, \rho_1)$. This proves that for all minimizers, $\omega = 0$ and thus there is a unique geodesic which is the Fisher-Rao geodesic (see Theorem 3.1). $\square$

Proposition 4.2 (No transport case). *If $\rho_1 = \alpha\rho_0$ with $\alpha \geq 0$, then*

$$\rho = \left(t\sqrt{\alpha} + (1-t)\right)^2 \rho_0 \otimes \mathrm{d}t$$

is the unique *geodesic for WF_κ, for all $\kappa > 0$, and*

$$WF_\kappa(\rho_0, \alpha\rho_0) = \kappa|\sqrt{\alpha} - 1|\sqrt{2\rho_0(\Omega)}.$$

Proof. If $\alpha > 0$, consider the function defined on $[0,1] \times \bar{\Omega}$ by

$$\varphi(t, x) = \frac{2\kappa^2(\sqrt{\alpha} - 1)}{(t\sqrt{\alpha} + (1-t))},$$

and write the set $\partial D_\kappa(\mu)$:

$$\partial D_\kappa(\mu) = \left\{ (\alpha, \beta, \gamma) \in C([0,1] \times \Omega; B_\kappa) : \alpha + \frac{1}{2}\frac{\gamma^2}{\kappa^2} = 0 - \rho \text{ a.e. } \beta = 0 - \rho \text{ a.e.} \right.$$
$$\left. \text{and } \gamma = \frac{2\kappa^2(\sqrt{\alpha} - 1)}{(t\sqrt{\alpha} + (1-t))} - \rho \text{ a.e. } \right\}. \quad (29)$$

We can check that $(\partial_t\varphi, \nabla\varphi, \varphi) \in \partial^c D_\kappa(\mu)$ because, in particular, $\nabla\varphi = 0$ and $\partial_t\varphi + \varphi^2/(2\kappa^2) = 0$. Then, as $(\rho, \omega, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1)$ the ansatz is a geodesic. The distance is found by computing $D_\kappa(\mu)$.

Note that, if $\alpha = 0$ the certificate φ we built above is not defined at $t = 1$ and this approach is useless. However, by the triangle inequality, for any $\alpha > 0$,

$$WF_\kappa(\rho_0, \alpha\rho_0) \leq WF_\kappa(\rho_0, 0) + WF_\kappa(0, \alpha\rho_0).$$

By Proposition 2.1 we have, $WF_\kappa(0, \alpha\rho_0) \leq \kappa\sqrt{2\alpha\rho_0(\Omega)}$. We rearrange the inequalities to obtain

$$\kappa|\sqrt{\alpha} - 1|\sqrt{2\rho_0(\Omega)} - \kappa\sqrt{2\alpha\rho_0(\Omega)} \leq WF_\kappa(\rho_0, 0) \leq \kappa\sqrt{2\rho_0(\Omega)}.$$

As this inequality holds for all $\alpha > 0$, we obtain $WF_\kappa(\rho_0, 0) = \kappa\sqrt{2\rho_0(\Omega)}$ which is the value of the functional evaluated at the ansatz μ for $\alpha = 0$, so the proof that ρ is a geodesic is complete. For uniqueness, it only remains to remark that $\partial_t\rho = g\rho$ (with g defined in (19)) and then to apply Proposition 4.1. $\qquad \square$

4.2   Transport of one Dirac measure to another

We first solve the case of the geodesic between two Dirac measures and then generalize this result to configurations with more Dirac measures. In the case of two Dirac measures, we show that if the Dirac measures are closer than $\pi\kappa$, then the *unique* geodesic is a travelling Dirac solution. Beyond that distance, the Hellinger geodesic is the *unique* geodesic.

Theorem 4.1. *Consider two Dirac measures of mass h_0 and h_1 and location x_0 and x_1 respectively, i.e. $\rho_0 = h_0\delta_{x_0}$ and $\rho_1 = h_1\delta_{x_1}$. We distinguish three types of behaviors for geodesics of WF_κ:*

1. *Travelling Dirac. If $|x_1 - x_0| < \pi\kappa$ then the travelling Dirac measure $\rho = h(t)\delta_{x(t)} \otimes \mathrm{d}t$ implicitly defined by*

$$\begin{cases} x(0) = x_0 \\ h(t) = At^2 - 2Bt + h_0 \\ h(t)x'(t) = \omega_0 \end{cases}$$

   is the unique *geodesic, where ω_0, A and B are real constants depending on h_0, h_1 and $x_1 - x_0$ through explicit relations given in the proof below.*

2. *Cut Locus. If $|x_1 - x_0| = \pi\kappa$ then there are infinitely many geodesics. Some of them can be described in the following way: take $N \in \mathbb{N}$ and choose two N-tuples $(h_0^i)_{i=1,\dots,N}$ and $(h_1^i)_{i=1,\dots,N}$ of non-negative real numbers satisfying $h_0 = \sum_1^N h_0^i$ and $h_1 = \sum_{i=1}^N h_1^i$. A geodesic is given by $\rho = \sum_{i=1}^N \rho_i$ where ρ_i is either the travelling Dirac solution of point 1. or a Fisher-Rao geodesic between $\rho_0^i\delta_{x_0}$ and $\rho_1^i\delta_{x_1}$. Notice that a single travelling Dirac solution or the Hellinger geodesics are particular cases.*

3. *No transport. If $|x_1 - x_0| > \pi\kappa$ then the Hellinger geodesic*

$$\rho = \left[t^2h_1\delta_{x_1} + (1-t)^2h_0\delta_{x_0}\right] \otimes \mathrm{d}t$$

   is the unique *geodesic.*

Remark. If $h_0 = h_1 = h$ and $|x_1 - x_0| < \pi\kappa$ then the minimum mass of the geodesic is attained at $t = 1/2$ and its value is

$$\frac{h}{2}\left[1 + \cos\left(\frac{|x_1 - x_0|}{2\kappa}\right)\right]$$

which is never smaller than $h/2$.

Proof. The proof is divided into 4 parts, in the first three parts, Ω is a segment in $\mathbb{R}$. First we derive an ansatz for the geodesics when $|x_1 - x_0| < \pi\kappa$, and write the sufficient optimality conditions. Then comes a part with technical computations where we prove the existence of an optimality certificate. In a third part we study what happens at the cut locus and beyond. Finally, we extend the proof for $\Omega \in \mathbb{R}^d$, for any dimension d.

i. Ansatz and optimality conditions In this part, we consider the case where $|x_1 - x_0| < \pi\kappa$ and Ω is a closed segment of $\mathbb{R}$. Let us look for geodesics of the form of travelling Dirac, i.e

$$\rho = h(t)\delta_{x(t)} \otimes \mathrm{d}t \tag{30}$$

where h and x are functions from $\mathbb{R}$ to $\mathbb{R}$, assumed sufficiently smooth so that all oncoming expressions make sense. Moreover, we assume $h_0, h_1 \neq 0$ and $x_1 \neq x_0$ (those special cases are all dealt with in Proposition 4.2). Satisfying the continuity constraint imposes $\omega = h(t)x'(t)\delta_{x(t)} \otimes \mathrm{d}t$ and $\zeta = h'(t)\delta_{x(t)} \otimes \mathrm{d}t$, so the functional reads

$$D_\kappa(\rho, \omega, \zeta) = \int_0^1 \frac{1}{2}\left(x'(t)^2 h(t) + \kappa^2 \frac{h'(t)^2}{h(t)}\right) \mathrm{d}t.$$

The Euler-Lagrange conditions imply that minimizers among travelling Diracs should satisfy, for some $\omega_0 \in \mathbb{R}$,

$$\begin{cases} h(t)x'(t) = \omega_0 \\ 2h''(t)h(t) - h'(t)^2 = \omega_0^2/\kappa^2 \\ h(0), h(1), x(0), x(1) \text{ fixed.} \end{cases}$$

Solutions of the second order differential equation are of the form $h(t) = At^2 - 2Bt + h_0$ where A and B are given by the system

$$\begin{cases} Ah_0 - B^2 = \frac{\omega_0^2}{4\kappa^2} \\ A - 2B = h_1 - h_0 \end{cases}$$

This is a second order equation which admits two solutions

$$A = h_1 + h_0 - 2\epsilon\sqrt{h_0 h_1 - \frac{1}{4}\frac{\omega_0^2}{\kappa^2}} \quad \text{and} \quad B = h_0 - \epsilon\sqrt{h_0 h_1 - \frac{1}{4}\frac{\omega_0^2}{\kappa^2}}$$

where $\epsilon \in \{-1, +1\}$. In order to choose the value for ϵ, we plug the expressions in the functional

$$D_\kappa(\rho, \omega, \zeta) = \int_0^1 \frac{1}{2}\frac{\omega_0^2 + \kappa^2 h'(t)^2}{h(t)}\mathrm{d}t = 2\kappa^2 A \tag{31}$$

and conclude that $\epsilon = +1$ as otherwise, the functional is greater than the Fisher-Rao upper bound given in Proposition 2.1. Also, (31) shows that $A > 0$ since $WF_\kappa(\rho_0, \rho_1) > 0$. In order to determine ω_0, we integrate the speed to obtain

$$\frac{x_1 - x_0}{\omega_0} = \int_0^1 \frac{\mathrm{d}t}{h(t)} = \frac{2\kappa}{\omega_0}\int_{-\alpha}^{\beta} \frac{\mathrm{d}u}{1 + u^2} = \frac{2\kappa}{\omega_0}[\arctan(\beta) + \arctan(\alpha)]$$

with

$$\alpha \stackrel{\text{def.}}{=} \frac{2\kappa}{\omega_0} B \quad \text{and} \quad \beta \stackrel{\text{def.}}{=} \frac{2\kappa}{\omega_0}(A - B). \tag{32}$$

The expression for the sum of arctan depends on the sign of $1 - \alpha\beta$. Yet we find

$$1 - \alpha\beta = \frac{4\kappa^2}{\omega_0^2} A \sqrt{h_0 h_1 - \frac{\omega_0^2}{4\kappa^2}}$$

which is positive since $A > 0$. So we are in the case

$$\arctan \alpha + \arctan \beta = \arctan\left(\frac{\alpha + \beta}{1 - \alpha\beta}\right) \in [-\pi/2, \pi/2]$$

and solutions exist if and only if $|x_1 - x_0| < \pi\kappa$. We introduce $\tau \stackrel{\text{def.}}{=} \frac{\alpha + \beta}{1 - \alpha\beta}$ and by direct computations

$$\omega_0 = 2\kappa\tau\sqrt{\frac{h_0 h_1}{1 + \tau^2}} \quad \text{and} \quad \tau = \tan\left(\frac{x_1 - x_0}{2\kappa}\right).$$

Notice that we can rewrite A and B in a simpler way with the new parameter τ

$$A = h_1 + h_0 - 2\sqrt{\frac{h_0 h_1}{1 + \tau^2}} \geq 0 \quad \text{and} \quad B = h_0 - \sqrt{\frac{h_0 h_1}{1 + \tau^2}}$$

making it clear that A and B are well defined.

ii. Existence of a certificate So far, we have shown that travelling Diracs of the form (30) cannot be geodesics if $|x_1 - x_0| > \pi\kappa$. Otherwise, there is a unique candidate solution and this part is devoted to the construction of an optimality certificate in order to show that it is optimal over all measure valued trajectories and not only among travelling Dirac solutions. For simplicity, let us take $\kappa = 1$ from now on, without loss of generality due to Proposition 3.1. By Theorem 2.3, $\varphi \in C^1([0,1] \times \Omega)$ is an optimality certificate if for all $(t, x) \in [0,1] \times \Omega$

$$\partial_t \varphi + \frac{1}{2}\left((\partial_x \varphi)^2 + \varphi^2\right) \leq 0 \tag{33}$$

and on the support of ρ :

$$\begin{cases} \partial_x \varphi(x(t), t) = x'(t) & \text{(34a)} \\ \varphi(x(t), t) = \dfrac{h'(t)}{h(t)} & \text{(34b)} \\ \partial_t \varphi + \dfrac{1}{2}\left((\partial_x \varphi)^2 + \varphi^2\right) = 0. & \text{(34c)} \end{cases}$$

It is also a uniqueness certificate if $\nabla\varphi$ is Lipschitz. In view of the equations φ should satisfy, we suggest the following optimality certificate for the candidate (ρ, ω, ζ):

$$\varphi(t, x) = a(t)\cos(x - \theta) + b(t)$$

Inequality (33) gives, for all $(t, x) \in [0, 1] \times \Omega$:

$$\left(a'(t) + a(t)b(t)\right)\cos(x - \theta) + \left(b'(t) + \frac{1}{2}(a^2(t) + b^2(t))\right) \le 0$$

As equality has to be reached at all points of the form $(x(t), t)$ from condition (34c), the only way to satisfy this inequality is to have

$$\begin{cases} a' + ab = 0 \\ b' + \frac{1}{2}(a^2 + b^2) = 0. \end{cases}$$

The solutions to this differential system are of the form

$$\begin{cases} a(t) = \frac{1}{t-t_1} - \frac{1}{t-t_2} \\ b(t) = \frac{1}{t-t_1} + \frac{1}{t-t_2} \end{cases}$$

Now assume $t_1, t_2 \notin [0, 1]$, so that a and b are C^∞ on $[0, 1]$. This assumption will be checked below.

It remains to check conditions (34a) and (34b). Equation (34a) imposes $x'(t) = -a(t)\sin(x(t) - \theta)$. We first determine x implicitly from this equation before checking in a later step that it indeed corresponds to ω_0/h. For $k \in \mathbb{Z}$, the function $x : t \in [0, 1] \mapsto \theta + k\pi$ is a stationary solution. As a is smooth, the Picard-Lindelöf Theorem guarantees the uniqueness of the solution of any initial value problem, so solutions do not intersect. Since x is not constant, we have $\forall t \in [0, 1], x(t) \notin \{\theta + k\pi, k \in \mathbb{Z}\}$, and thus $\sin(x(t) - \theta)$ does not vanish in $[0, 1]$. Let us divide by $\sin(x(t) - \theta)$ and integrate equation (34a)

$$\frac{x'(t)}{\sin(x(t) - \theta)} = \frac{1}{t - t_2} - \frac{1}{t - t_1} \quad \Leftrightarrow \quad \frac{1}{2}\log\frac{1 - \cos(x(t) - \theta)}{1 + \cos(x(t) - \theta)} = \log\frac{1}{K}\left|\frac{t - t_2}{t - t_1}\right|$$

and thus

$$\cos(x(t) - \theta) = \frac{K^2(t - t_1)^2 - (t - t_2)^2}{K^2(t - t_1)^2 + (t - t_2)^2} \tag{35}$$

for some $K > 0$. The value of the integration constant K will be studied later on, but we already see that θ is determined given t_1 and t_2:

$$\theta = x(0) - \arccos\left(\frac{K^2 t_1^2 - t_2^2}{K^2 t_1^2 + t_2^2}\right) \bmod 2\pi$$

The last constraint we have is equality (34b): $a(t)\cos(x(t) - \theta) + b(t) = \frac{h'(t)}{h(t)}$. After several lines of re-arranging one finds this is equivalent to

$$2\frac{K^2(t - t_1) + (t - t_2)}{K^2(t - t_1)^2 + (t - t_2)^2} = 2\frac{At - B}{At^2 - 2Bt + h_0}.$$

By identification, there exists a factor $\lambda_2 > 0$ such that:

$$\begin{cases} A = \lambda_2(K^2 + 1) \\ B = \lambda_2(K^2 t_1 + t_2) \\ h_0 = \lambda_2(K^2 t_1^2 + t_2^2). \end{cases}$$

This can be symmetrized as

$$\begin{cases} A = \lambda_1 + \lambda_2 \\ B = \lambda_1 t_1 + \lambda_2 t_2 \\ h_0 = \lambda_1 t_1^2 + \lambda_2 t_2^2. \end{cases} \tag{36}$$

with $\lambda_1 > 0$, $\lambda_2 > 0$ (we imposed $\lambda_1 = \lambda_2 K^2$).

Also, note that gathering conditions (34a), (34b) and (34c) leads to the relation $2h'h - (h')^2 = (x')^2 h^2$ and thus, by the Euler-Lagrange conditions satisfied by h, $(x')^2 h^2 = \omega_0^2$. This proves that the function x implicitly defined by (35) is the same as that of the ansatz. To sum up, if a solution to system (36) satisfies the constraints

$$\begin{cases} \lambda_1 > 0 \\ \lambda_2 > 0 \\ (t_1, t_2) \notin [0, 1]^2 \end{cases}$$

then we have built an optimality certificate. System (36) leads to the following candidate solutions:

$$\begin{cases} t_1 = \frac{B}{A} + \varepsilon \frac{\omega_0}{2A} \frac{1}{K} \\ t_2 = \frac{B}{A} - \varepsilon \frac{\omega_0}{2A} K \\ \lambda_2 = A - \lambda_1 \end{cases}$$

where $K = \sqrt{\frac{\lambda_1}{\lambda_2}}$ and $\varepsilon \in \{-1, 1\}$. The sign of ε depends on direction of the movement. We choose say, $\varepsilon = 1$, the other case is similar. Let us look for a solution satisfying, e.g. $t_1 > 1$ and $t_2 < 0$. This leads to the conditions

$$K > \alpha \quad \text{and} \quad \frac{1}{K} > \beta$$

where α and β have been defined earlier in the proof in (32). It is not excluded that β or α be non-positive, in that case it is easy to find K satisfying both inequalities. Now, if $\alpha, \beta > 0$, the necessary and sufficient condition for a solution K to exist, is $\alpha\beta < 1$, which we have already checked. This concludes the proof for the case $|x_1 - x_0| < \kappa\pi$.

iii. The cut locus and beyond   Now we study what happens when $|x_1 - x_0| = \pi\kappa$, which is the cut locus distance. Introduce a Dirac measure $h_1 \delta_{x(s)}$ between $h_0 \delta_{x_0}$ and $h_1 \delta_{x_1}$ where $x(s) = x_0 + s(x_1 - x_0)$. It is not hard to check (see for instance equation (31)) that

$$WF_\kappa^2(h_0 \delta_{x_0}, h_1 \delta_{x(s)}) \underset{s \to 1}{\to} 2\kappa^2(h_0 + h_1)$$

and that

$$WF_\kappa^2(h_1 \delta_{x(s)}, h_1 \delta_{x_1}) \underset{s \to 1}{\to} 0.$$

Combining two triangle inequalities—for the upper and lower bounds—we obtain

$$WF_\kappa^2(h_0 \delta_{x_0}, h_1 \delta_{x_1}) = 2\kappa^2(h_0 + h_1).$$

This has important consequences. First, it proves, as expected, that the travelling Dirac solution is still a geodesic at the cut locus, with the trajectory $\rho_t = h(t)\delta_{x(t)}$ where

$$h(t) = h_0(t-1)^2 + h_1 t^2 \quad \text{and} \quad x'(t) = \frac{4h_0 h_1 \kappa}{h(t)}$$

since D_κ evaluated at this trajectory is worth $2\kappa^2(h_0 + h_1)$. Second, as D_κ linearly depends on the masses we can build infinitely many geodesics, examples are described in the Theorem. Finally, the cost at the cut locus equals the upper-bound of pure Hellinger (recall Proposition 2.1).

Thus, by monotonicity of the distance w.r.t. rescaling (see Proposition 3.2), the WF_κ distance between two Dirac measures for which $|x_1 - x_0| \geq \pi\kappa$ is $\kappa\sqrt{2(h_0 + h_1)}$ and a geodesic is

$$\rho = \left[h_1 t^2 \delta_{x_1} + h_0(1-t)^2 \delta_{x_0} \right] \otimes \mathrm{d}t \,.$$

For uniqueness, we again apply Proposition 3.2 for proving that, when $|x_1 - x_0| > \pi\kappa$, the component ω of minimizers is zero. Thus the only geodesic is the Hellinger geodesic, by Theorem 3.1.

iv. Extension to arbitrary dimension The extension of the previous results for $\Omega \in \mathbb{R}^d$, $d > 1$ is rather straightforward and we will do it in such a way so as to prepare the ground for the next Theorem. Recall that what we want to prove (and have proven for $d = 1$ so far) is that if $|x_1 - x_0| < \pi\kappa$, then the travelling Dirac solution (as described in the Theorem) is the unique geodesic. To show this, we choose $\kappa = 1$ without loss of generality and we introduce the function

$$\varphi(t,x) = \left(\frac{1}{t - t_1} - \frac{1}{t - t_2} \right) \cos(|x - \theta|) + \frac{1}{t - t_1} + \frac{1}{t - t_2} \tag{37}$$

where the constants t_1, t_2 and the vector θ can be found by the same method than above by considering the problem on the line $\mathrm{span}\{x_1 - x_0\}$. This is possible because cosine is an even function and $\theta \in \mathrm{span}\{x_1 - x_0\}$. Now remark that the derivative of $\cos(|\cdot|)$ is continuous at 0, thus $\varphi \in C^1([0,T] \times \Omega)$ and by construction satisfies all the conditions to be an optimality and uniqueness certificate and the result is shown. The behaviors at the cut locus and beyond extend straightforwardly to this case. $\square$

4.3 Transport of atomic measures

Geodesics between one couple of Dirac measures are so far quite well-understood. To go one step further in the comprehension of the behavior of our metric, we study the geodesics between two purely atomic measures ρ_0 and ρ_1. We show that under some geometrical conditions on the locations of the atoms, the behavior is similar to that of a pair of atoms. Those conditions require basically that the atoms be arranged by close pairs of Dirac measures, and the pairs should be far from each other.
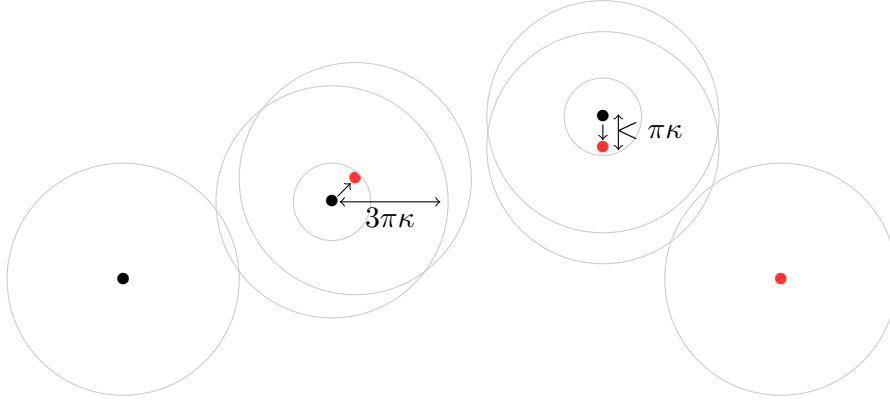
Figure 1: Configuration in dimension 2 satisfying hypotheses of Theorem 4.2. The initial ρ_0 and final ρ_1 measures are composed of the black and red Dirac measures, respectively. At least 1 couple of spheres of radius $3\pi\kappa$ do not intersect for each pair of systems. We show that in that case, the geodesic is the sum of two travelling Dirac solutions (for the pairs), an on-place decreasing Dirac measure (single black) and an on place increasing Dirac measure (single red).

Theorem 4.2. *Let $\rho_0 = \sum_{i=1}^{N} h_0^i \delta_{x_0^i}$ and $\rho_1 = \sum_{i=1}^{N} h_1^i \delta_{x_1^i}$ be the initial and final measures, where h_0^i or h_1^i may be zero. Also suppose that*

- $\forall i \in \{1, \dots N\}, |x_0^i - x_1^i| < \pi\kappa,$

- $\forall (i,j) \in \{1, \dots N\}^2, i \neq j \Rightarrow \max_{\alpha, \beta \in \{0,1\}} \{|x_\alpha^i - x_\beta^j|\} > 6\pi\kappa.$

Then $\rho = \sum_{i=1}^{N} \rho^i$, where for all $i \in \{1, \dots, N\}$, $\rho^i = h^i(t)\delta_{x^i(t)} \otimes \mathrm{d}t$ is the geodesic between $h_0^i \delta_{x_0^i}$ and $h_1^i \delta_{x_1^i}$ from Theorem 4.1, is a geodesic for WF_κ. This geodesic is unique *if all masses h_0^i, h_1^i are positive.*

Remarks.

- A finer analysis could easily lower the "security distance" of $6\pi\kappa$ between pairs of Dirac measures so that they can be considered independently. However, is is clear that when pairs of Dirac become really close, one has more complicated behaviors (cf. the limit model $\kappa \to \infty$ in Theorem 3.2 which corresponds to Diracs measures becoming infinitely close).

- In Figure 1 we give an example where the hypotheses of Theorem 4.2 are satisfied and in Figure 2 we give the shape of an optimality certificate as built in the proof.

- We only prove uniqueness of the geodesic when the certificate is non-degenerate. But it is likely that uniqueness hold in general under our hypotheses.

Proof. Again we choose $\kappa = 1$ without loss of generality. As for now, assume that all masses h_0^i, h_1^i are strictly positive, the case of vanishing masses, i.e.

31

"single" Dirac measures, will be fixed at the end of the proof. We aim at building an optimality certificate $\varphi \in C^1([0,1] \times \Omega)$ for the candidate geodesic of the Theorem. Hence φ should satisfy, for all $(t,x) \in [0,1] \times \Omega$,

$$\partial_t \varphi + \frac{1}{2} \left(|\nabla \varphi|^2 + \varphi^2 \right) \leq 0$$

and for all $t \in [0,1]$, for all $i \in \{1, \ldots, N\}$:

$$\begin{cases} \nabla \varphi(x^i(t), t) = (x^i)'(t) \\ \varphi(x^i(t), t) = (h^i)'(t)/h^i(t) \\ \partial_t \varphi + \frac{1}{2} \left(|\nabla \varphi|^2 + \varphi^2 \right) = 0 \quad \text{for points of the form } (t, x^i(t)). \end{cases}$$

We introduce for $i \in \{1, \ldots N\}$, the certificates φ^i for the geodesics between each couple of Dirac measures $(h_0^i \delta_{x_0^i}, h_1^i \delta_{x_1^i})$ as defined in equation (37). The three parameters describing φ^i are denoted t_1^i, t_2^i and θ^i. As θ^i is only defined up to translations, we decide from now on that θ^i is the unique such translation satisfying $\max\{|x_0^i - \theta^i|, |x_1^i - \theta^i|\} < \pi$, which exists under our assumption that $|x_0^i - x_1^i| < \pi$. Thus, from the hypotheses in the Theorem, we have that $\min_{i \neq j}\{|\theta^i - \theta^j|\} > 4\pi$. Now consider the *binding* functions, defined for $i \in \{1, \ldots, N\}$ by

$$\varphi_b^i(t,x) = -\frac{1}{t - t_2^i} \cos(|x - \theta^i|) + \frac{1}{t - t_2^i}.$$

Each φ_b^i is in phase with φ^i and satisfies

$$\varphi_b^i(t, \pi u + \theta^i) = \frac{2}{t - t_2^i} = \varphi^i(t, \pi u + \theta^i) \quad \text{and} \quad \varphi_b^i(t, 2\pi u + \theta^i) = 0$$

for all $u \in \mathbb{R}^d$, $|u| = 1$ and $t \in [0,1]$. Also they are solutions to the equation $\partial_t \varphi_b^i + \frac{1}{2} \left((\nabla \varphi_b^i)^2 + (\varphi_b^i)^2 \right) = 0$. Now, as the constant zero function is a solution of this p.d.e too, we introduce

$$\varphi(t,x) = \begin{cases} \varphi^i(t,x) & \text{if } |x - \theta^i| \leq \pi \\ \varphi_b^i(t,x) & \text{if } |x - \theta^i| \leq 2\pi \text{ and } |x - \theta^i| > \pi \\ 0 & \text{otherwise.} \end{cases}$$

Under the hypotheses of the Theorem, only one condition is met at a time so this function is well defined. It belongs to $C^1([0,1] \times \Omega)$ because the bindings are located where the gradients vanish, on the extrema of cosines. See Figure 2 to picture how the bindings are performed between the φ^is. By construction, all hypotheses are fulfilled so that φ be an optimality and uniqueness certificate.

Now if one or more of the h_0^i or h_1^i vanish, the same reasoning as in the proof of Theorem 4.1 allows to compute the distance between ρ_0 and ρ_1 (by an upper and lower bound, making the mass of vanishing Dirac measures tend to zero), and all that remains is to see that replacing the travelling Dirac solution by a Hellinger geodesic for single Dirac measures reaches that cost. $\qquad \square$
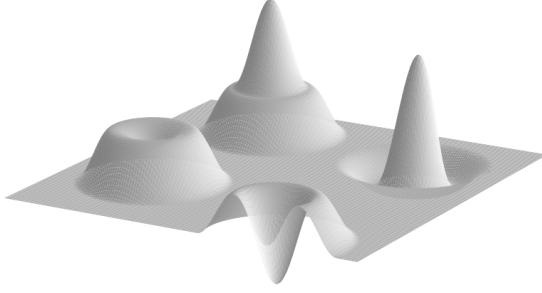
Figure 2: Example of an optimality certificate for 4 pairs of Dirac measures, in the case $d = 2$ at fixed t. The centers of the bumps are the θ^is. At the position $(t, x_i(t))$ of a travelling Dirac solution, the gradient gives its speed and the height gives its rate of growth.
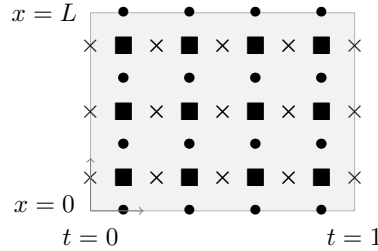

Figure 3: (■) centered and (×, ●) staggered grids. The grey rectangle is the space-time domain and here $N = 3$, $T = 4$.

5 Numerical results

This section presents a numerical implementation of the computation of geodesics for our metric. The source code to reproduce these results is available online[1].

5.1 Discretization

The numerical resolution of the dynamical optimal transport problem is often performed with first order methods using proximal splitting algorithms ([6], [24]). These methods extend with no difficulties to the case of transport with sources, as long as the proximal operator of the new functional can be easily computed. In order to compare a few models of optimal transport with source, we implemented a Douglas-Rachford proximal splitting algorithm on a staggered grid, in the footsteps of [24]. The introduction of the source term comes with a few adjustments so we describe the numerical method that we implemented, in the case of a transport problem in 1-D for simplicity (but the scheme works in arbitrary dimension).

Centered and staggered grids. We consider the space domain $[0, L]$ and the time domain $[0, 1]$. We denote N the number of dicretization points in space and

---

[1]https://github.com/lchizat/optimal-transport/

T the number of discretization points in time. The centered grid is an equidistant discretization of the interior of the space-time domain $[0, L] \times [0, 1]$, namely

$$\mathcal{G}_c = \left\{ \left(x_i = L\frac{i - 1/2}{N}, t_j = \frac{j - 1/2}{T} \right) : 1 \le i \le N, \, 1 \le j \le T \right\}$$

and the variable discretized on the centered grid are denoted

$$V = (\rho, \omega, \zeta) \in \mathcal{E}_c = (\mathbb{R}^{\mathcal{G}_c})^3.$$

Remark that the boundaries are not included in the centered grid. There are as many staggered grids as there are dimensions, defined as

$$\mathcal{G}_s^x = \left\{ (x_i = L\frac{i - 1}{N}, t_j = \frac{j - 1/2}{T}) : 1 \le i \le N + 1, \, 1 \le j \le T \right\},$$

$$\mathcal{G}_s^t = \left\{ (x_i = L\frac{i - 1/2}{N}, t_j = \frac{j - 1}{T}) : 1 \le i \le N, \, 1 \le j \le T + 1 \right\},$$

and the variables discretized on the staggered grid are denoted by

$$U = (\bar{\rho}, \bar{\omega}, \bar{\zeta}) \in \mathcal{E}_s = \mathbb{R}^{\mathcal{G}_s^t} \times \mathbb{R}^{\mathcal{G}_s^x} \times \mathbb{R}^{\mathcal{G}_c}.$$

The source ζ is discretized on the centered grid because there is no differentiation operator acting on this variable in the definition of the continuity constraint. Figure 3 gives an illustration of the centered and staggered grids.

Discrete continuity constraint. The discrete continuity constraint with boundary conditions is defined as the set

$$\mathcal{CE} = \left\{ U = (\bar{\rho}, \bar{\omega}, \bar{\zeta}) \in \mathcal{E}_s : AU = f_0, \, A = \left[\frac{\mathbf{div} - s_z}{s_b} \right], \, f_0 = \left[\frac{0}{b_0} \right] \right\} \quad (38)$$

where $s_z : U \mapsto \bar{\zeta}$, s_b selects the values of U on the boundaries of the staggered grids and b_0 is a vector containing ρ_0, ρ_1 and zeros for the Neumann boundary conditions. The divergence operator $\mathbf{div}$ only acts on the components $(\bar{\rho}, \bar{\omega})$ of U, takes its values on the centered grid and is defined for $U \in \mathcal{E}_s$ as

$$\mathbf{div}(U)_{i,j} = \frac{N}{L}(\bar{\omega}_{i+1,j} - \bar{\omega}_{i,j}) + T(\bar{\rho}_{i,j+1} - \bar{\rho}_{i,j}).$$

Discrete functional. First, we use Proposition 3.1 and bring ourselves back to the case $\kappa = 1$ by a space rescaling: this allows to obtain a simpler proximal operator. The discrete functional is defined, for $V \in \mathcal{E}_c$, as $D(V) \overset{\text{def.}}{=} \sum_{k \in \mathcal{G}_c} f(\rho_k, \omega_k, \zeta_k)$ with

$$f(\rho_k, \omega_k, \zeta_k) \overset{\text{def.}}{=} \frac{|\omega_k|^2 + \zeta_k}{2\rho_k}.$$

Interpolation operator. Since the continuity constraint and the functional D are not defined on the same grid, we need to keep two variables $(U, V) \in \mathcal{E}_s \times \mathcal{E}_c$ and to add an interpolation constraint between them, i.e. we require $V = \mathcal{I}(U)$ where I is the midpoint interpolation operator

$$I(U)_{i,j} = \left(\frac{\bar{r}_{i,j+1} + \bar{r}_{i,j}}{2}, \frac{\bar{m}_{i+1,j} + \bar{m}_{i,j}}{2}, \bar{z}_{i,j} \right).$$

Discrete optimization problem The discrete convex optimization problem that we aim to solve as an approximation of the continuous problem $(\mathcal{P_M})$ reads then

$$\min_{U,V} D(V) + \iota_{\mathcal{CE}}(U) + \iota_{\{V = \mathcal{I}(U)\}}(U, V).$$

5.2 Minimization algorithm

Proximal methods and the Douglas Rachford algorithm. A popular class of first order methods for solving non-smooth convex optimization problems, so-called proximal splitting schemes, replaces the explicit gradient descent step by an implicit descent step using the so-called proximal operator. For a proper, convex, l.s.c. function F defined on a Hilbert space $\mathcal{H}$ with values in $\mathbb{R} \cup +\infty$, the proximal operator is the single-valued map defined as

$$\text{Prox}_{\gamma F}(x) = \underset{\bar{x} \in \mathcal{H}}{\text{argmin}} \, \frac{1}{2} \|x - \bar{x}\|^2 + \gamma F(\bar{x}).$$

If F is smooth, the optimality condition imply that $\bar{x} \stackrel{\text{def.}}{=} \text{Prox}_{\gamma F}(x)$ should satisfy $\bar{x} = x - \gamma \nabla F(\bar{x})$, justifying the common interpretation of the proximal operator as an implicit gradient step. A minimizer of F can be found by successively applying the proximal operator, but for complicated functions without structure, computing the proximal operator can be as difficult as solving the whole minimization problem. However, if the function is the sum of simple terms, minimization can be performed by so-called *proximal splitting* methods. The one we implemented is the Douglas Rachford algorithm which allows to minimize the sum of two convex, proper, l.s.c. functions G_1 and G_2 with the following iterative scheme: choose two initial points $(z^{(0)}, w^{(0)}) \in \mathcal{H}^2$ and define the sequence

$$w^{(l+1)} = w^{(l)} + \alpha(\text{Prox}_{\gamma G_1}(2z^{(l)} - w^{(l)}) - z^{(l)}),$$
$$z^{(l+1)} = \text{Prox}_{\gamma G_2}(w^{(l+1)}).$$

If $0 < \alpha < 2$ and $\gamma > 0$, one can show thant $z^{(l)} \to z^*$ a minimizer, see [12]. Depending on which functions we take as G_1 and G_2, we obtain various algorithms; in our code we used

$$G_1(U, V) = \iota_{\mathcal{CE}}(U) + D(V) \quad \text{and} \quad G_2(U, V) = \iota_{V = \mathcal{I}(U)}(U, V).$$

For a detailed review of proximal splitting algorithms, we refer the reader to [11]. Let us now compute the proximal operators.

Computing $\mathrm{Prox}_{\gamma D}$. The proximal operator of the functional D can be computed in closed form. The following result is a direct adaptation of [24].

Proposition 5.1. *One has*

$$\forall V \in \mathcal{E}_c, \quad \mathrm{Prox}_{\gamma D}(V) = (\mathrm{Prox}_{\gamma f}(V_k))_{k \in \mathcal{G}_c}$$

where, for all $(\tilde{\rho}, \tilde{\omega}, \tilde{\zeta}) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$,

$$\mathrm{Prox}_{\gamma f}(\tilde{\rho}, \tilde{\omega}, \tilde{\zeta}) = \begin{cases} \left(\rho^*, \frac{\rho^* \tilde{\omega}}{\rho^* + \gamma}, \frac{\rho^* \tilde{\zeta}}{\rho^* + \gamma} \right) & \text{if } \rho^* > 0, \\ (0, 0, 0) & \text{otherwise,} \end{cases}$$

and ρ^* is the largest real root of the third order polynomial equation in X

$$P[X] = (X - \tilde{\rho})(X + \gamma)^2 - \frac{\gamma}{2}(|\tilde{\omega}|^2 + \tilde{\zeta}^2).$$

Computing $\mathrm{Proj}_{\mathcal{CE}}$. The proximal operator of $\iota_{\mathcal{CE}}$ is the projection on the affine set $\mathcal{CE}$, which can be computed for $U \in \mathcal{E}_s$ as

$$
\begin{align}
\mathcal{P}_{\mathcal{CE}}(U) &= U + A^*(AA^*)^{-1}(b_0 - AU) \tag{39} \\
&= \mathcal{P}_B(U) - (Id - I_b)(s_z^* - \mathbf{div}^*)S^{-1}(s_z - \mathbf{div})(\mathcal{P}_B(U)) \tag{40}
\end{align}
$$

where $S = (\mathbf{div} - s_z)(Id - s_b^* s_b)(-\nabla - s_z^*)$ is the Schur complement of AA^*, I_b is the identity on the boundaries on zero everywhere else and $\mathcal{P}_B$ is the projection on the boundary constraints. Given p, we can find u satisfying $Su = p$ by solving $\Delta u - u + p = 0$ on the centered grid with Neumann boundary conditions (on the staggered grids). In the Fourier domain, this equation reads

$$[1 + (2 - 2\cos(\pi m/N))N/L + (2 - 2\cos(\pi n/T))T] \hat{u}[m, n] = \hat{p}[m, n].$$

The DCT-II transform (with inverse DCT-III) which coefficients are given by

$$\hat{u}[k] = \sum_{n=1}^{N} u[n] \cos \left[\frac{\pi}{N}(n + \frac{1}{2})k \right]$$

is adapted to our boundary conditions.

Compute $\mathrm{Proj}_{V=\mathcal{I}(U)}$. The projection is given for all couples $(U_0, V_0) \in (\mathcal{E}_s, \mathcal{E}_c)$ by

$$\mathrm{Proj}_{V=\mathcal{I}(U)}(U_0, V_0) = (U^*, \mathcal{I}(U^*))$$

with $U^* = Q^{-1}(U_0 + \mathcal{I}^* V_0)$ and $Q = Id + \mathcal{I}^* \mathcal{I}$. As Q is a tridiagonal matrix, LU factorization techniques allow to efficiently invert the system.
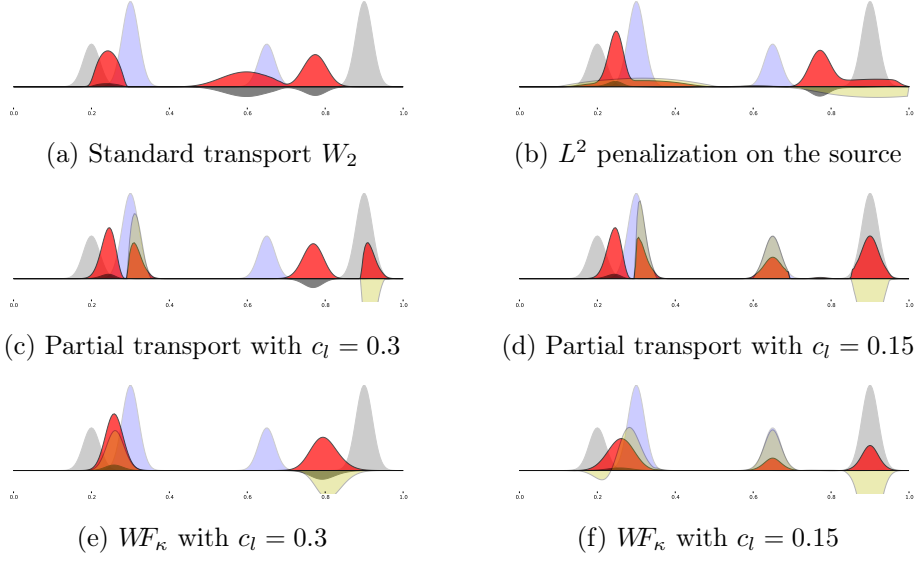
(a) Standard transport W_2

(b) L^2 penalization on the source

(c) Partial transport with $c_l = 0.3$

(d) Partial transport with $c_l = 0.15$

(e) WF_κ with $c_l = 0.3$

(f) WF_κ with $c_l = 0.15$

Figure 4: Comparison of interpolations for densities on the segment $[0, 1]$: ρ_0 (gray), ρ_1 (blue), $\rho_{t=1/2}$ (red), $\omega_{t=1/2}$ (black), $\zeta_{t=1/2}$ (yellow). We indicate by c_l the theoretical maximum distance a Dirac can travel, which is a function of κ.

5.3 Experiments

Transport of Gaussian bumps. Let us first explore a synthetic case where the initial and final measures have the same mass. This case is shown on Figure 4. The initial and the final measures are both composed of two Gaussian densities of mass 1 and 2 supported on the segment $\Omega = [0, 1]$. The modes of the bumps are located at, from left to right, 0.2, 0.3, 0.65 and 0.9. The problem is discretized with $N = 256$ samples in space and $T = 11$ samples in time. The peculiar location of the Gaussian bumps allows to highlight the variations of the behavior of the geodesics, which is highly dependent on the metric and on the parameter κ. For the W_2 geodesic (Figure 4a), the rightmost grey bump is split in half and a part of it travels to the left side of the graph, giving a interpolated measure which, for some applications, is not so natural. The effect of a non-homogeneous functional is visible on Figure 4b, where we penalized the L^2 norm of the source ζ in the functional. Non-homogeneity makes the density matter: in this case, the source is spread because low densities are favored. Consider now the partial transport geodesics: on Figure 4c, we fixed the maximum distance mass can travel to $2\kappa = 0.3$, while on Figure 4d, that distance is $2\kappa = 0.15$. We observe, as expected, that the domain $\Omega = [0, 1]$ is split in an active and an inactive set. Also note that as the TV cost is not strictly convex, geodesics are not unique and what happens in the inactive set depends on the initialization of the algorithm. Finally, Figures 4e and 4f display the geodesics at $t = 1/2$ for the WF_κ metric with the *cut locus* distances set, respectively, at $\pi\kappa = 0.3$ and $\pi\kappa = 0.15$. Note that there is a slight similarity between the configuration here and the hypotheses of Theorem 4.2, and the observed behavior can be expected if we extrapolate the
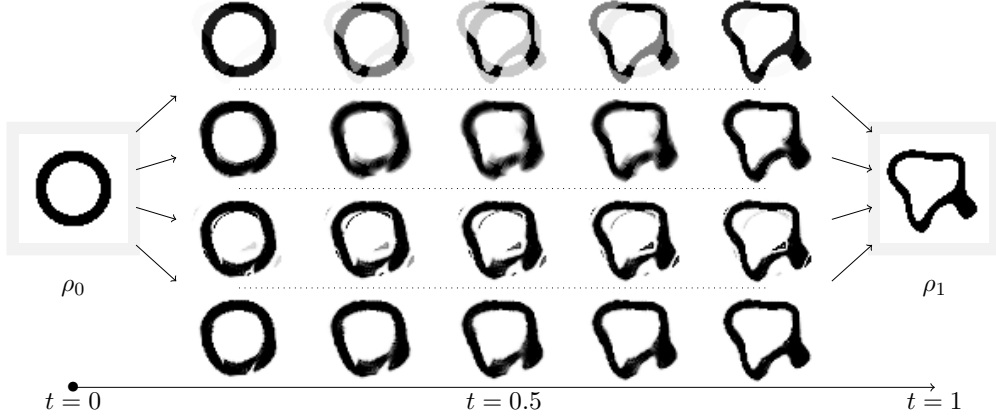
Figure 5: Geodesics between ρ_0 and ρ_1 for four metrics: Fisher-Rao (1st row), W_2 (2nd row), partial optimal transport (3rd row), WF_κ (4th row).

conclusions of that Theorem. In the first case, we obtain a geodesic consisting of two travelling bumps which inflate or deflate. While in the second configuration, only one bump travels as $\pi\kappa$ is now smaller than the distance between the two bumps at the right.

Synthetic 2D experiments. We now turn our attention to a synthetic example on the domain $\Omega = [0,1]^2$. The initial density ρ_0 is the indicator of the ring of center $(\frac{1}{2}, \frac{1}{2})$, internal diameter 0.5 and external diameter 0.7. The final density ρ_1 is the pushforward of ρ_1 by a random smooth map and thus $\rho_0(\Omega) = \rho_1(\Omega)$. The domain is discretized on a centered grid of 64×64 samples in space and $T = 12$ samples in time. We compare on Figure 5 the geodesics for four different metrics: d_{FR}, W_2, partial optimal transport (with $2\kappa = 0.2$) and WF_κ (with $\kappa\pi = 0.4$). Notice that the two first rows thus show the limit geodesics for WF_κ when κ tends to 0 or $+\infty$ (see Theorem 3.2).

Figure 6 helps to better understand the geodesics. On the top row, the velocity field shows that the W_2 geodesic transports a lot of mass to the bottom left protuberance. On the contrary, metrics allowing local variations of mass (partial transport and WF_κ), attenuate the component of the velocity field which is tangential to the ring, behavior which is more consistent with the intuitive solution. Finally, by looking at the source maps in the bottom row, we see the inactive sets of partial optimal transport with well defined edges and how this behavior strongly differs to that of WF_κ geodesics.

Figure 7 displays another experiment intended to illustrates how the geodesics behave when a lot of transport is involved. The measures ρ_0 and ρ_1 have same total mass, the domain $\Omega = [0,1]^2$ is discretized on a centered grid with 44×44 samples in space and $T = 12$ samples in time. We show the geodesics for W_2, partial optimal transport (with $2\kappa = 0.5$) and WF_κ (with $\kappa\pi = 0.5$).
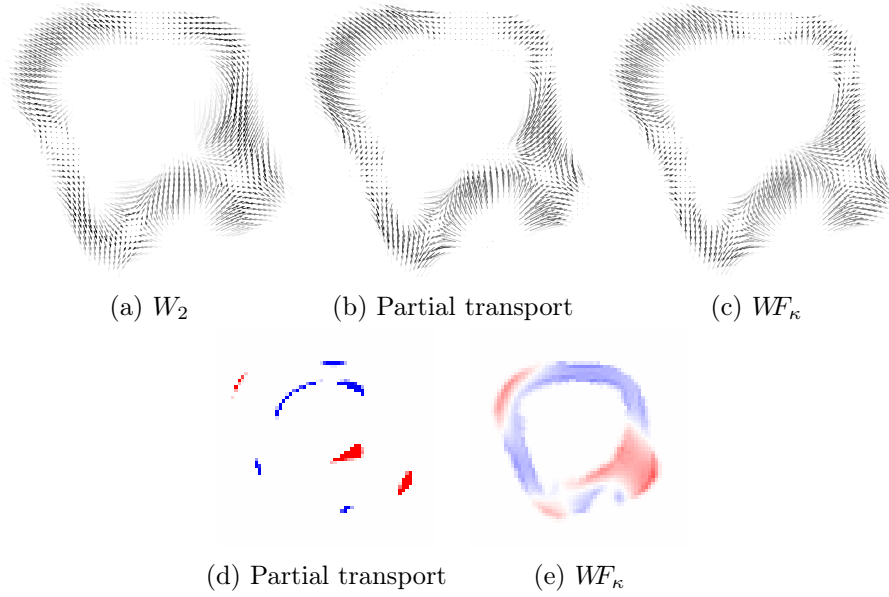
38

(a) W_2 (b) Partial transport   (c) WF_κ

(d) Partial transport   (e) WF_κ

Figure 6: First row: velocity field $v = \omega/\rho$ at time $t = 1/2$. The higher ρ, the darker the arrow. Second row: source ζ at time $t = 1/2$. Blue stands for negative density and red for positive.

Biological images interpolation. Interpolating between shapes with varying masses is an important motivation for introducing this metric and is the object of our third and last numerical experiment. The initial and final densities ρ_0, ρ_1 are extracted form two segmented images of the same young brain taken at different times. This example is rather challenging for image matching algorithms: matter is creased, folded and grows unevenly. There are 89×73 samples in space, 12 samples in time and the spatial domain is $\Omega = [0, 0.82] \times [0, 1]$. We computed the geodesics for W_2 and for WF_κ with $\pi\kappa = 0.2$. We show on Figure 8 the geodesics and on Figure 9 the velocity fields and the rate of growth a time $t = 1/2$.

Most of the tissue growth is located at the bottom of the domain. Consequently, the velocity field associated to the W_2 geodesic is dominated by top to bottom components, as seen on Figure 9a. To the contrary, the geodesic for the WF_κ metric manages to locally adapt the rate of growth, and the velocity field is far more consistent with the true evolution of the tissue. However, we retain some artifacts coming from optimal transport: some matter is teared off the brain lining and brought somewhere else to fill a need of mass. This behavior, inherent to the non-smooth nature of optimal transport plans, is clearly observed on Figure 8 near the bulges that appear on the right and left sides of the brain.

Finally, let us recall the reader that, at optimality, the rate of growth g is equal to the dual variable φ and thus, the vector field on Figure 9b is the gradient of the rate of growth on Figure 9c, which ranges in $[-0.05, 0.18]$.
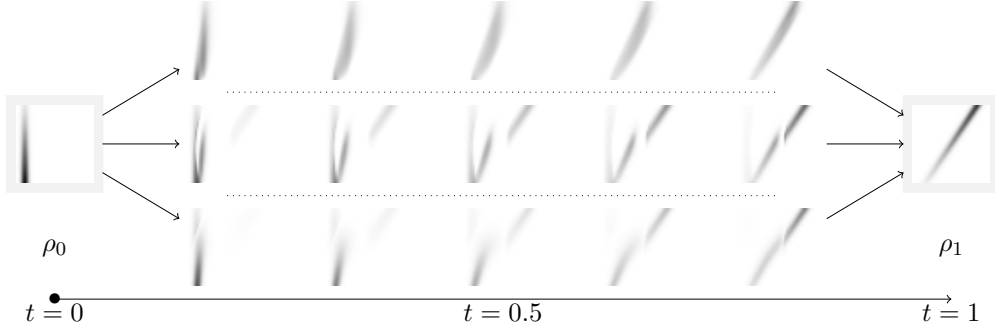
Figure 7: Geodesics between ρ_0 and ρ_1 for three metrics: W_2 (1st row), partial optimal transport (2nd row), WF_κ (3rd row). It is important to notice that ρ_0 is denser at the bottom left while ρ_1 is denser at the top right to better understand those geodesics.
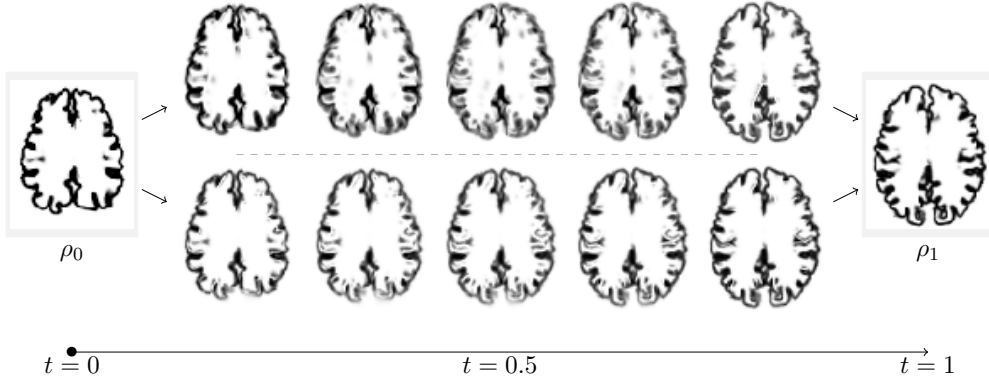

Figure 8: Geodesics between ρ_0 and ρ_1 for two metrics: W_2 between rescaled densities (top row), WF_κ metric with $\pi\kappa = 0.2$ (bottom row).

Conclusion

When looking for generalizations of the dynamic optimal transport problem to unbalanced measures, a distance interpolating between optimal transport and Fisher-Rao metrics appears naturally. It is a natural Riemannian-like metric deriving from a family of convex homogeneous dynamic minimization problems. By analyzing the effect of varying the interpolation factor, and by studying geodesics for atomic measures, we obtained theoretical clues on how the model behaves. These clues can be summarized by the two following interpretations: (i) the distance behaves as a spatially localized version of optimal transport, in the sense that mass whose supports are far away does not interact; (ii) it also behaves as a regularized version of Fisher-Rao, in the sense that, unlike Fisher-Rao, it is stable by perturbation of the supports. This theoretical behavior is confirmed by the numerical experiments.

(a) $v_{0.5}$ for W_2 (b) $v_{0.5}$ for WF_κ (c) $g_{0.5}$ for WF_κ
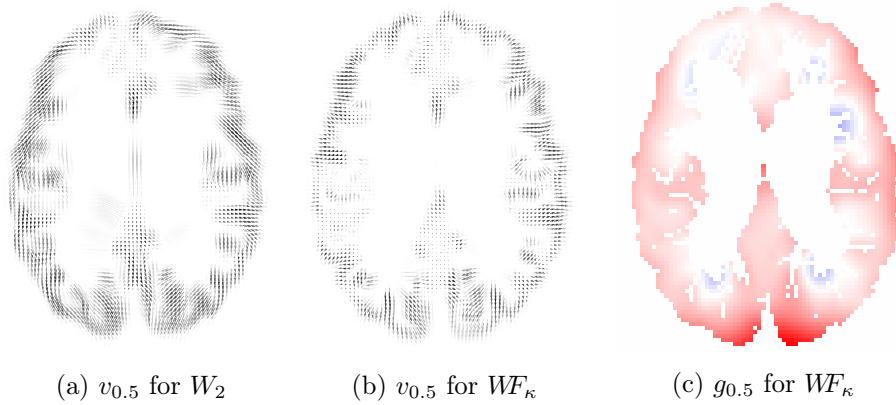
Figure 9: (a) and (b): velocity field $v = \omega/\rho$ for geodesics at time $t = 1/2$. The higher $\rho_{0.5}$, the darker the arrow. (c) rate of growth $g = \zeta/\rho$ at time $t = 1/2$. Blue stand for negative density and red for positive and g is set to zero when $\rho \approx 0$. Note that theoretically, (b) is the gradient field of (c).

Acknowledgements

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2008.

[2] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. Information geometry and sufficient statistics. *Probability Theory and Related Fields*, 162(1):327–364, 2015.

[3] M. Bauer, M. Bruveris, and P. W. Michor. Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.*, 48(3):499–506, 2016.

[4] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

[5] J.-D. Benamou. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(05):851–868, 2003.

[6] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[7] J.-D. Benamou and Y. Brenier. Mixed L2-Wasserstein optimal mapping between prescribed density functions. *Journal of Optimization Theory and Applications*, 111(2):255–271, 2001.

[8] G. Bouchitté and G. Buttazzo. New lower semicontinuity results for nonconvex functionals defined on measures. *Nonlinear Analysis: Theory, Methods & Applications*, 15(7):679–692, 1990.

[9] L. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampere obstacle problems. *Annals of mathematics*, 171(2):673–730, 2010.

[10] P. Cardaliaguet, G. Carlier, and B. Nazaret. Geodesics for a class of distances in the space of probability measures. *Calculus of Variations and Partial Differential Equations*, 48(3-4):395–420, 2013.

[11] P. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[12] P. L. Combettes and J.-C. Pesquet. A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):564–574, 2007.

[13] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. *Calculus of Variations and Partial Differential Equations*, 34(2):193–231, 2009.

[14] A. Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.

[15] A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with dirichlet boundary conditions. *Journal de mathématiques pures et appliquées*, 94(2):107–130, 2010.

[16] K. Guittet. Extended Kantorovich norms: a tool for optimization. Technical report, Tech. Rep. 4402, INRIA, 2002.

[17] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent. Optimal mass transport for registration and warping. *International Journal of computer vision*, 60(3):225–240, 2004.

[18] L. G. Hanin. Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces. *Proceedings of the American Mathematical Society*, 115(2):345–352, 1992.

[19] C. Jimenez. Dynamic formulation of optimal transport problems. *Journal of Convex Analysis*, 15(3):593, 2008.

[20] L. Kantorovich. On the transfer of masses (in russian). *Doklady Akademii Nauk*, 37(2):227–229, 1942.

[21] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. Technical report, Preprint, 2015.

[22] D. Lombardi and E. Maitre. Eulerian models and algorithms for unbalanced optimal transport. *ESAIM: M2AN*, 49(6):1717 – 1744, 2015.

[23] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM: M2AN*, 49(6):1745–1769, 2015.

[24] N. Papadakis, G. Peyré, and E. Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212–238, 2014.

[25] B. Piccoli and F. Rossi. On properties of the Generalized Wasserstein distance. arXiv:1304.7014, 2013.

[26] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[27] C. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.

[28] R. Rockafellar. Duality and stability in extremum problems involving convex functions. *Pacific Journal of Mathematics*, 21(1):167–187, 1967.

[29] R. Rockafellar. Integrals which are convex functionals. ii. *Pacific Journal of Mathematics*, 39(2):439–469, 1971.

[30] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005.

[31] C. Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.