Differential Geometric Heuristics for Riemannian Optimal Mass Transportation

Philippe Delanoë

Université de Nice Sophia–Antipolis, Laboratoire J.–A. Dieudonné, Parc Valrose 06108 NICE Cedex 2 FRANCE, Philippe.DELANOE@unice.fr

Summary. We give an account on Otto's geometrical heuristics for realizing, on a compact Riemannian manifold M, the L^2 Wasserstein distance restricted to smooth positive probability measures, as a Riemannian distance. The Hilbertian metric discovered by Otto is obtained as the base metric of a Riemannian submersion with total space, the group of diffeomorphisms of M equipped with the Arnol'd metric, and projection, the push-forward of a reference probability measure. The expression of the horizontal constant speed geodesics (time dependent optimal mass transportation maps) is derived using the Riemannian geometry of M as a guide.

Key words: optimal mass transportation; rearrangement of diffeomorphisms; Hilbertian manifold; Helmholtz splittings; Riemannian submersion; horizontal geodesics.
2000 MSC: 35B99; 35Q99; 58B20; 58D05; 58J05.

1 Optimal Mass Transportation Diffeomorphims

Let M be a compact connected n-dimensional manifold (all objects are C^∞ unless otherwise specified; so, a measure admits a smooth density in each chart). We may view measures as n-forms of odd type [dRh55], hence freely consider the pull-back measure $\phi^*\nu$ of a measure ν by a map $\phi : M \to M$. Pulling-back does not preserve the total mass: $\int_M d(\phi^*\nu) \not\equiv \int_M d\nu$, unless ϕ is a diffeomorphism. This is in contrast with the push-forward (also called *transport*) of a measure μ by a map $\phi : M \to M$, denoted by $\phi_\#\mu$, which may be defined (via the Riesz representation theorem [Rie09]) by:

$$\int_M u \; d\nu := \int_M (u \circ \phi) \; d\mu, \text{ with } \nu = \phi_\#\mu, \tag{1}$$

where u stands for an arbitrary continuous real function on M. Here, the measure $\phi_\#\mu$ is not necessarily smooth (even though μ and ϕ are), but it certainly is if ϕ is a diffeomorphism (if so, exercise: check that $\phi_\#\mu = (\phi^{-1})^*\mu$). In any case, the total mass is preserved (letting $u = 1$ in (1)). In the sequel, we normalize the total mass equal to 1, all maps from M to itself are diffeomorphisms

and we restrict the transport to (smooth) positive probability measures, the set of which we denote by Prob. The latter is a convex domain in an *affine* space modelled on the Fréchet space Mes_0 of measures with zero average on M. In particular, we will freely use the fact that the tangent bundle $T\,\mathrm{Prob}$ is trivial, equal to $\mathrm{Prob} \times \mathrm{Mes}_0$. As readily checked, the transport yields a *right action* on Prob of the group of diffeomorphisms of M.
From now on, we endow the manifold M with a Riemannian metric g:

Question Q: using the metric g and the above right action, how can one find good notions of distance and shortest path in Prob ?

Given measures $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$, optimal transport theory provides an answer which we now describe. First of all, a criterion of optimality is required. Following Brenier and McCann [Bre91, McC01], it is defined by choosing the *cost-function* given by: $\forall (p,q) \in M \times M,\ c(p,q) := \dfrac{1}{2} d_g^2(p,q)$, where d_g stands for the geodesic distance in M, and by looking for a minimizer of the total transport cost functional:

$$C_\mu(\phi) := \int_M \frac{1}{2} d_g^2(m, \phi(m))\, d\mu\ ,$$

among all Borel maps $\phi : M \to M$ satisfying $\phi_\# \mu = \nu$. Such a minimization problem is called a Monge's problem, after Gaspard Monge who was the first to consider such a problem, in the Euclidean space for the total work functional $\int_{\mathbb{R}^n} |x - \phi(x)| d\mu$ [Mon81].
An essential tool for solving a Monge's problem is the notion of c-convexity. Dropping temporarily smoothness, a real function f on M is called c-convex on M if it can be written $f = h^c$ for some real function h, where:

$$\forall m \in M, \quad h^c(m) := \sup_{p \in M} [-h(p) - c(m,p)],$$

and $c = \frac{1}{2} d_g^2$. If so, f is Lipschitz, thus differentiable outside a subset $S \subset M$ of zero Riemannian volume measure (Rademacher's theorem); moreover, the gradient $\boldsymbol{\nabla} f : M \setminus S \to TM$ is Borel measurable [McC01]. The map $h \mapsto h^c$ is often called the c-transform on M [CMS01] (thought of as a kind of Legendre transform) and a c-convex function $f = h^c$ satisfies the involution identity: $f = (f^c)^c$ [R-R98].
Setting $\exp : TM \to M$ for the Riemannian exponential map, we can now state the main result of the landmark paper [McC01]:

Theorem 1 (McCann). *Given* $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$*, there exists a* c*-convex function* $f : M \to \mathbb{R}$*, unique up to addition of a constant, which satisfies the equation:*

$$\exp(\boldsymbol{\nabla} f)_\# \mu = \nu\ . \tag{2}$$

Moreover, the Borel map $\exp(\boldsymbol{\nabla} f) : M \to M$ *is the unique minimizer for our Monge's problem (modulo discrepancies on a subset of zero* μ*-measure).*

The quantity

$$\mathrm{W}_2(\mu,\nu) := \sqrt{C_\mu(\exp(\nabla f))} \equiv \sqrt{\int_M \frac{1}{2}|\nabla f|^2 \, d\mu}$$

with μ, ν, f as in Theorem 1 defines a *distance* in Prob [Vil08, Chap.6] (see also [Vil03] and Theorem 3 below); let Prob_2 denote the completion of Prob for this distance. The complete metric space $(\mathrm{Prob}_2, \mathrm{W}_2)$ is called the L^2 Wasserstein space associated to (M, g), and W_2, the Wasserstein distance. The following starshapedness property holds:

Lemma 1 ([CMS01] Lemma 5.1). *For each $t \in [0,1]$, the function tf is c-convex on M if f is so.*

With Lemma 1 at hand, we infer from Theorem 1 that the path given by:

$$t \in [0,1] \to \mu_t := \exp(t\nabla f)_{\#}\mu \in \mathrm{Prob}_2 \tag{3}$$

is W_2-*minimizing* from $\mu_0 = \mu$ to $\mu_1 = \nu$. We thus have got an answer to Question Q, except for the smoothness of the measures μ_t for $t \in (0,1)$.
Indeed, the smoothness of the data (M, μ, ν, g) does not always imply that of the optimal transport map given by Theorem 1. Recently, this question has been intensively investigated (see [Vil08, Chap.12] and references therein). However, anytime the given measures μ and ν are close enough[1] in Prob, the c-convex solution of (2) must be smooth [Del04, Theorem 1]. By combining Theorem 1 with Theorem 5 of Appendix A below, we can state a result in the smooth category, namely:

Theorem 2. *Given $(\mu,\nu) \in \mathrm{Prob} \times \mathrm{Prob}$, assume the existence of a smooth solution f of the partial differential equation:*

$$\exp(\nabla f)^*\nu = \mu \ . \tag{4}$$

The function f must be c-convex on M and satisfy (2). Moreover, the path (3) ranges in Prob *and, for each $t \in [0,1]$, the map $\exp(t\nabla f)$ is a diffeomorphism.*

At this stage, the reader may not realize how natural, from the Riemannian geometric viewpoint, are the answers to Question Q given by the two preceding theorems. The goal of this paper is to convince ourselves that they are, indeed, completely natural. To do so, we give below an exhaustive account on the beautiful heuristics discovered by Félix Otto [Ott01] (see also [Lot08, K-L08]). We will proceed stepwise, in a pretty self-contained way[2], working mostly in the group of diffeomorphisms of M rather than in Prob, with elementary tools from (finite-dimensional) Riemannian geometry and Poisson's type equations. Hopefully, it will serve as a complement to John

[1] in Fréchet topology, of course (cf. supra)
[2] except for the last part of Appendix A

Lott's recent paper [Lot08] written in the spirit of infinite-dimensional calculations performed straight in Prob. It will also prepare the reader for further studies e.g. in the sub-Riemannian setting [K-L08].

Finally, as regards the geometry of equation (4), we would like to mention that (4) admits a (non-homogeneous) Monge–Ampère structure in Lychagin's sense [Lyc79] hence Lie solutions [Del08] which would deserve a deeper study.

Acknowledgment: I am grateful to Valentin Lychagin and Boris Kruglikov for inviting me at the Abel Symposium 2008, in the magnificent site of Tromsø. I would like to thank also a Referee for pointing out to me the reference [K-L08].

2 Geometry of the Group of Diffeomorphisms, after Arnol'd

Henceforth, we set Diff for the group of diffeomorphisms of the manifold M and, fixing $\lambda \in$ Prob, we single out the subgroup Diff_λ of diffeomorphisms which preserve the reference measure λ (pushing it to itself).

2.1 Rearrangement Classes

Let us consider the map $\mathcal{P}_\lambda : \mathrm{Diff} \to \mathrm{Prob}$ defined by

$$\forall \phi \in \mathrm{Diff}, \quad \mathcal{P}_\lambda(\phi) := \phi_\# \lambda \ .$$

It yields a partition of Diff into countersets:

$$\mathrm{c}_\mu := \{\phi \in \mathrm{Diff},\ \phi_\# \lambda = \mu\}, \quad \mu \in \mathcal{P}_\lambda(\mathrm{Diff}),$$

including the one which contains $\mathbb{I}$ (the identity of M), namely $\mathrm{c}_\lambda \equiv \mathrm{Diff}_\lambda$. Moreover, two diffeomorphisms ϕ and ψ lie in the same counterset c_μ if and only if:

$$\exists \xi \in \mathrm{Diff}_\lambda, \quad \phi = \psi \circ \xi \ . \tag{5}$$

In other words, letting Diff_λ act on Diff by right composition and considering (5) as an equivalence relation, we have for the quotient space:

$$\mathrm{Diff} \,/\, \mathrm{Diff}_\lambda = \{\mathrm{c}_\mu,\ \mu \in \mathcal{P}_\lambda(\mathrm{Diff})\};$$

each counterset c_μ may thus be viewed as a coset, called by Brenier [Bre91] a *rearrangement class*.

2.2 Tangent Bundle

For $t \in \mathbb{R}$ close to 0, let $t \mapsto \psi_t \in \mathrm{Diff}$ be a path satisfying $\psi_0 = \mathbb{I}$. On the one hand $\frac{d\psi_t}{dt}|_{t=0}$ lies in the tangent space $T_{\mathbb{I}}\,\mathrm{Diff}$, on the other hand we have:

$$\forall m \in M, \quad \frac{d\psi_t(m)}{dt}|_{t=0} \in T_m M \ .$$

So $T_{\mathbb{I}}\,\mathrm{Diff}$ coincides with the vector fields on M, a Fréchet space henceforth denoted by Vec.
Fixing an arbitrary $\phi \in \mathrm{Diff}$, let $t \mapsto \phi_t$ be a path satisfying $\phi_0 = \phi$. How can we view the tangent vector $\frac{d\phi_t}{dt}|_{t=0} \in T_\phi\,\mathrm{Diff}$? Sticking to the right composition, we may write $\phi_t = \psi_t \circ \phi$ with ψ_t as above, getting:

$$\frac{d\phi_t}{dt}|_{t=0} = \frac{d\psi_t}{dt}|_{t=0} \circ \phi \ .$$

We conclude:

$$T_\phi\,\mathrm{Diff} = \{V \circ \phi,\ V \in \mathrm{Vec}\} \ . \tag{6}$$

2.3 The Arnol'd Metric

Following Arnol'd [Arn66], let us define on the tangent bundle $T\,\mathrm{Diff}$ the following field of Hilbertian scalar products:

$$\forall \phi \in \mathrm{Diff},\ \forall (V,W) \in \mathrm{Vec}^2, \quad \langle V \circ \phi, W \circ \phi \rangle_\phi := \int_M \frac{1}{2} g(V \circ \phi, W \circ \phi)\ d\lambda \ .$$

Observing that

$$\phi \in \mathrm{c}_\mu \Longrightarrow \langle V \circ \phi, W \circ \phi \rangle_\phi = \int_M \frac{1}{2} g(V, W)\ d\mu \ ,$$

we infer that the Arnol'd metric is right-invariant along each rearrangement class $\mathrm{c}_\mu \in \mathrm{Diff} / \mathrm{Diff}_\lambda$ (originally, Arnol'd restricted it to Diff_λ with the idea that the resulting geodesics would describe the motion of an incompressible fluid in the manifold M, see [Arn66, E-M70]).
Using the Arnol'd metric, we can define the length of paths in Diff, hence a *distance* on Diff; let us denote it by d_A. Given $(\phi, \psi) \in \mathrm{Diff}^2$, we thus have:

$$\mathrm{d}_A(\phi, \psi) = \inf \int_0^1 \sqrt{\left\langle \frac{d\phi_t}{dt}, \frac{d\phi_t}{dt} \right\rangle_{\phi_t}}\ dt \ ,$$

where the infimum runs over all paths $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ such that $\phi_0 = \phi$ and $\phi_1 = \psi$.

3 The Riemannian Submersion $\mathcal{P}_\lambda : \mathrm{Diff} \to \mathrm{Prob}$, after Moser, Ebin–Marsden and Otto

With the view of improving the way of solving some nonlinear heat equations, Félix Otto (working in $\mathbb{R}^n$) [Ott01] advocated the use of a new gradient flow on Prob which he had the idea to construct with a metric inherited from the Arnol'd one via the projection $\mathcal{P}_\lambda$. In the present section, we implement the latter idea stepwise. The reader will find in [K-L08] a parallel theory outlined for the sub-Riemannian case.
We require notations: Funct will denote the Fréchet space of smooth real-valued functions on M, and for each $\mu \in \mathrm{Prob}$, Funct_0^μ will denote the subspace of functions $f \in \mathrm{Funct}$ such that $f\mu \in \mathrm{Mes}_0$. Auxiliary material for this section may be found in Appendix B.

3.1 The Submersion

The first step is Moser's famous result on volume forms [Mos65].

Proposition 1 (Moser). *The map* $\mathcal{P}_\lambda : \mathrm{Diff} \to \mathrm{Prob}$ *is onto.*

Proof. Following [Mos65, E-M70], let us construct a right-inverse for the map $\mathcal{P}_\lambda$. Given an arbitrary $\mu \in \mathrm{Prob}$, consider the linear interpolation path $t \in [0,1] \to \mu_t := t\lambda + (1-t)\mu \in \mathrm{Prob}$. By Corollary 5 of Appendix B, for each $t \in [0,1]$, there exists a unique $f_t \in \mathrm{Funct}_0^{\mu_t}$ solving the equation:

$$\mathrm{div}_{\mu_t}(\boldsymbol{\nabla} f_t)\ \mu_t = -\frac{d\mu_t}{dt} \ . \tag{7}$$

The map $t \mapsto f_t$ is smooth and, from (7), the flow $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ of the time-dependent vector field $\boldsymbol{\nabla} f_t$ on M satisfies:

$$\frac{d}{dt}\left(\phi_{t\#}\mu_t\right) = 0, \quad \phi_0 = \mathbb{I} \ .$$

We thus have $\phi_{t\#}\mu_t = \mu_0$ hence, in particular: $\mathcal{P}_\lambda(\phi_1) = \mu$ □

Let us denote by $\mathcal{M}_\lambda(\mu)$ the diffeomorphism ϕ_1 just constructed. The map $\mathcal{M}_\lambda : \mathrm{Prob} \to \mathrm{Diff}$ is a right-inverse for $\mathcal{P}_\lambda$ such that $\mathcal{M}_\lambda(\lambda) = \mathbb{I}$.

Corollary 1 (Ebin–Marsden). *The map* $\mathcal{P}_\lambda : \mathrm{Diff} \to \mathrm{Prob}$ *is a submersion.*

Proof. As observed in [E-M70], the map $\mathcal{M}_\lambda$ yields a *factorization* of Diff; specifically, setting for each $\phi \in \mathrm{Diff}$,

$$\mathrm{D}_\lambda(\phi) := \left[\mathcal{M}_\lambda\left(\mathcal{P}_\lambda(\phi)\right)\right]^{-1} \circ \phi \ ,$$

and recalling (5), we get a map $\mathrm{D}_\lambda : \mathrm{Diff} \to \mathrm{Diff}_\lambda$ such that the following factorization identically holds in Diff:

$$\phi \equiv \mathcal{M}_\lambda\left(\mathcal{P}_\lambda(\phi)\right) \circ \mathrm{D}_\lambda(\phi)\ .$$

In other words, as do Ebin and Marsden, we may declare that the map:

$$\phi \in \mathrm{Diff} \to (\mathcal{P}_\lambda(\phi), \mathrm{D}_\lambda(\phi)) \in \mathrm{Prob} \times \mathrm{Diff}_\lambda$$

is a diffeomorphism (global and onto). The latter makes the map $\mathcal{P}_\lambda$ read merely like a projection; so, indeed, it is a submersion □

Proposition 2. *The tangent map to $\mathcal{P}_\lambda$ is onto with direct kernel.*

Proof. A straightforward calculation, using (6) and Definition 1 of Appendix B, yields for the tangent map to $\mathcal{P}_\lambda$ the following important expression:

$$\forall \mu \in \mathrm{Prob},\ \forall \phi \in \mathrm{c}_\mu,\ \forall V \in \mathrm{Vec}, \quad T_\phi \mathcal{P}_\lambda(V \circ \phi) = \mathrm{div}_\mu(V)\mu\ . \tag{8}$$

Combining it with Corollaries 5 and 6 of Appendix B yields the proposition (for the notion of *direct* factor, see e.g. [Lan62]) □

3.2 Helmholtz Splitting

From Proposition 1 and Corollary 1, for each $\mu \in \mathrm{Prob}$, we have $\mathcal{P}_\lambda^{-1}(\mu) = \mathrm{c}_\mu$ and this fiber is a *submanifold* of Diff diffeomorphic to Diff_λ. Given $\phi \in \mathrm{c}_\mu$, let us identify the (so-called *vertical*) subspace $T_\phi \mathrm{c}_\mu$ of $T_\phi \mathrm{Diff}$. Pick a path $t \mapsto \phi_t \in \mathrm{c}_\mu$ with $\phi_0 = \phi$ and differentiate with respect to t at $t = 0$ the identity: $\mathcal{P}_\lambda(\phi_t) = \mu$. Recalling (8), we get the equation $\mathrm{div}_\mu(V) = 0$ satisfied by the vector field V such that $V \circ \phi = \frac{d\phi_t}{dt}|_{t=0}$. In other words, we have:

$$T_\phi \mathrm{c}_\mu = \{V \circ \phi,\ V \in \ker \mathrm{div}_\mu\}\ . \tag{9}$$

Regarding the orthogonal complement of $T_\phi \mathrm{c}_\mu$ in $T_\phi \mathrm{Diff}$ for the Arnol'd metric, the so-called *horizontal* subspace at ϕ, we can write from the definition of the Arnol'd metric and (9):

$$\forall W \in \mathrm{Vec},\ (W \circ \phi) \in T_\phi \mathrm{c}_\mu^\perp \iff \forall V \in \ker \mathrm{div}_\mu,\ \int_M g(V, W) d\mu = 0\ .$$

By Corollary 6 of Appendix B (Helmholtz decomposition), we conclude:

$$\forall W \in \mathrm{Vec},\ (W \circ \phi) \in T_\phi \mathrm{c}_\mu^\perp \iff \exists f \in \mathrm{Funct}_0^\mu,\ W = \boldsymbol{\nabla} f\ .$$

Setting $\mathcal{V}_\phi$ and $\mathcal{H}_\phi$ respectively for the vertical and horizontal tangent subspaces to Diff at ϕ, we may summarize the situation as follows:

Proposition 3. *At each $\phi \in \mathrm{c}_\mu$, the following splitting holds:*

$$T_\phi \mathrm{Diff} = \mathcal{V}_\phi \oplus \mathcal{H}_\phi\ ,$$

with the vertical subspace $\mathcal{V}_\phi = T_\phi \mathrm{c}_\mu$ given by (9) and the horizontal subspace, by:

$$\mathcal{H}_\phi = \{\boldsymbol{\nabla} f \circ \phi,\ f \in \mathrm{Funct}_0^\mu\}\ .$$

Moreover, the factors of the splitting are orthogonal for the Arnol'd metric and they vary smoothly with the diffeomorphism ϕ.

3.3 Horizontal lift

A path $t \mapsto \phi_t \in \mathrm{Diff}$ is called *horizontal* if: $\forall t$, $\dfrac{d\phi_t}{dt} \in \mathcal{H}_{\phi_t}$. It is the *horizontal lift* of a path $t \mapsto \mu_t \in \mathrm{Prob}$ if it is horizontal satisfying $\mu_t = \mathcal{P}_\lambda(\phi_t)$.

Proposition 4. *Each path $t \mapsto \mu_t \in \mathrm{Prob}$ admits a unique horizontal lift passing, at some time $t = t_0$, through a given diffeomorphism of $\mathrm{c}_{\mu_{t_0}}$.*

Proof. In order to prove the uniqueness, let $t \mapsto \phi_t \in \mathrm{Diff}$ and $t \mapsto \psi_t \in \mathrm{Diff}$ be two horizontal lifts of the same path $t \mapsto \mu_t \in \mathrm{Prob}$ with $\phi_{t_0} = \psi_{t_0}$. Set $\dot{\phi}_t = \boldsymbol{\nabla} f_t \circ \phi_t$ (resp. $\dot{\psi}_t = \boldsymbol{\nabla} h_t \circ \psi_t$), with f_t (resp. h_t) in $\mathrm{Funct}_0^{\mu_t}$, and differentiate with respect to t the equation $\mathcal{P}_\lambda(\phi_t) = \mathcal{P}_\lambda(\psi_t)$. Recalling (8), we get:

$$\mathrm{div}_{\mu_t}(\boldsymbol{\nabla}(f_t - h_t)) = 0 \ ,$$

hence $f_t = h_t$ by Theorem 6 (Appendix B). In particular, the time-dependent vector fields $\boldsymbol{\nabla} f_t$ and $\boldsymbol{\nabla} h_t$ have the same flow θ_t, so indeed:

$$\phi_t = \theta_{t-t_0} \circ \phi_{t_0} \equiv \theta_{t-t_0} \circ \psi_{t_0} = \psi_t \ .$$

As for the existence, given a path $t \mapsto \mu_t \in \mathrm{Prob}$ defined near $t = t_0$ and a diffeomorphism $\psi_0 \in \mathrm{c}_{\mu_{t_0}}$, Corollary 5 of Appendix B provides for each t a solution f_t of the equation:

$$\Delta_{\mu_t} f_t \ \mu_t = \frac{d\mu_t}{dt} \ . \tag{10}$$

From (8), the flow ϕ_t of the time-dependent vector field $\boldsymbol{\nabla} f_t$ is such that the path $t \mapsto \psi_t = \phi_{t-t_0} \circ \psi_0 \in \mathrm{Diff}$ is a horizontal lift of $t \mapsto \mu_t \in \mathrm{Prob}$ passing through ψ_0 at $t = t_0$, as required $\square$

We will sometimes call (10) the *horizontal lift* equation.

3.4 The Otto Metric

Following Otto [Ott01] (see also [Lot08]), for each $\mu \in \mathrm{Prob}$, we equip the tangent space $T_\mu \mathrm{Prob}$ with the Hilbertian scalar product such that, for each $\phi \in \mathrm{c}_\mu$, the restriction of the tangent map $T_\phi \mathcal{P}_\lambda$ to the horizontal subspace $\mathcal{H}_\phi$ is an *isometry*. Recalling (8), we see that it must be defined by[3]:

$$\forall (\nu, \nu') \in T_\mu \mathrm{Prob} \times T_\mu \mathrm{Prob}, \quad \langle \nu, \nu' \rangle_\mu := \frac{1}{2} \int_M g(\boldsymbol{\nabla} f, \boldsymbol{\nabla} f') d\mu \ , \tag{11}$$

with f given by:

$$\mathrm{div}_\mu(\boldsymbol{\nabla} f) \ \mu = \nu$$

(recalling Corollary 5 of Appendix B) and similarly for f' with ν'. By construction, when Prob (resp. Diff) is endowed with the Otto (resp. Arnol'd) metric, the map $\mathcal{P}_\lambda$ becomes a *Riemannian submersion* (see e.g. [C-E75, pp.65–68], [FIP04] and references therein).

[3] using (23), we also have: $\langle \nu, \nu' \rangle_\mu = \frac{1}{2} \int_M f d\nu' = \frac{1}{2} \int_M f' d\nu$

3.5 The L^2 Wasserstein Distance

Given $(\mu,\nu) \in \mathrm{Prob} \times \mathrm{Prob}$, recall that the L^2 Wasserstein distance $\mathrm{W}_2(\mu,\nu)$ is given by: $\mathrm{W}_2(\mu,\nu) = \inf \sqrt{C_\mu(\phi)}$ where the total cost functional C_μ was defined in Section 1 and the infimum is taken over all measurable maps $\phi : M \to M$ such that $\phi_\# \mu = \nu$.
Using the Otto metric, we can define in Prob the notion of arclength, hence an alternative distance (by the usual length infimum procedure) which we denote by d_O. The fundamental result of [Ott01, Lot08] is the following[4]:

Theorem 3 (Otto–Lott). *The Otto distance* d_O *coincides on* Prob *with the* L^2 *Wasserstein distance* W_2.

The inequality $\mathrm{W}_2 \leq \mathrm{d}_O$ is fairly straightforward to prove. For completeness, let us prove it here.
Pick a constant speed path $t \in [0,1] \to \mu_t \in \mathrm{Prob}$ with $\mu_0 = \mu, \mu_1 = \nu$, and let $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ be its horizontal lift, given by Proposition 4. From the definition of W_2, we have:

$$\mathrm{W}_2^2(\mu,\nu) \leq \int_M \frac{1}{2} d_g^2 \left[m, (\phi_1 \circ \phi_0^{-1})(m) \right] \, d\mu = \int_M \frac{1}{2} d_g^2(\phi_0(m), \phi_1(m)) \, d\lambda \, .$$

Moreover, recalling the definition of the Arnol'd metric, we may write:

$$\int_M \frac{1}{2} d_g^2(\phi_0(m), \phi_1(m)) \, d\lambda \leq \int_M \frac{1}{2} \left(\int_0^1 |\dot{\phi}_t(m)| dt \right)^2 \, d\lambda \leq \int_0^1 \langle \dot{\phi}_t, \dot{\phi}_t \rangle_{\phi_t} dt \, ,$$

where the latter inequality is derived by applying Schwarz inequality followed by Fubini theorem. From the definition of the Otto metric and since the path $t \in [0,1] \mapsto \mu_t \in \mathrm{Prob}$ has constant speed (we set L for its length), combining the above inequalities yields: $\mathrm{W}_2(\mu,\nu) \leq L$. Taking the infimum of the right-hand side over all (constant speed) paths in Prob going from μ to ν, we get the desired result $\square$
The reversed inequality is more tricky; it will be proved below (Corollary 3) in a different way than in [Lot08].

4 Geodesics

In this section, we will investigate the properties of the *horizontal geodesics* in the group Diff as total space of the Riemannian submersion precedingly defined.

[4] which implies that W_2 is, indeed, a distance

4.1 A Sufficient Condition for Geodesicity in Diff

What is a reasonable notion of shortest path $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ between two given diffeomorphisms ϕ_0 and ϕ_1 ? A naive guess prompts us, for each $m \in M$, to interpolate between the image points $\phi_0(m)$ and $\phi_1(m)$ by means of a constant speed minimizing geodesic in M (unique provided its end points are located close enough). It motivates the following condition:

Condition G: *for each $m \in M$, the path $t \in [0,1] \to \phi_t(m) \in M$ is minimizing with constant speed (MCS, for short).*

The next result is classical [E-M70]:

Proposition 5. *Let $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ be a path from ϕ_0 to ϕ_1. If it satisfies Condition G, it must be MCS in* Diff *for the Arnol'd metric.*

Proof. The constant speed (CS) property is trivial; let us focus on the minimizing one. For each CS path $t \in [0,1] \to \psi_t \in \mathrm{Diff}$ with $\psi_0 = \phi_0, \psi_1 = \phi_1$, Fubiny theorem yields for its length L_ψ the equality:

$$L_\psi^2 = \int_0^1 \langle \dot{\psi}_t, \dot{\psi}_t \rangle_{\psi_t} dt = \int_M \frac{1}{2} \left(\int_0^1 |\dot{\psi}_t(m)|^2 dt \right) d\lambda \ .$$

Schwarz inequality implies:

$$L_\psi^2 \geq \int_M \frac{1}{2} \left(\int_0^1 |\dot{\psi}_t(m)| dt \right)^2 d\lambda \geq \int_M \frac{1}{2} d_g^2(\phi_0(m), \phi_1(m)) \ d\lambda$$

and, taking the infimum of the left-hand side on such paths ψ_t, we conclude:

$$\mathrm{d}_A^2(\phi_0, \phi_1) \geq \int_M \frac{1}{2} d_g^2(\phi_0(m), \phi_1(m)) \ d\lambda \ .$$

But the squared length L_ϕ^2 of the path $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ (for the Arnol'd metric) is equal to the latter right-hand side, due to Condition G. In other words, we have $L_\phi \leq \mathrm{d}_A(\phi_0, \phi_1)$ therefore, indeed, the aforementioned path is minimizing $\square$

We defer to section 4.3 (Proposition 9) a proof, in the same spirit (avoiding to compute the Levi-Civita connection of the Arnol'd metric as in [E-M70, Theorem 9.1]), of a partial converse to Proposition 5.

4.2 Short Horizontal Segments

Throughout this section, we fix an arbitrary couple $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$ of distinct but *suitably close* probability measures, and a diffeomorphism $\phi \in \mathrm{c}_\mu$. We look for a horizontal path $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ starting from ϕ, such that its projection $\mu_t := \mathcal{P}_\lambda(\phi_t)$ satisfies $\mu_1 = \nu$ and realizes the distance $\mathrm{d}_O(\mu, \nu)$.

Choice of a Candidate Path

Since the path $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ should be minimizing, we assume that it satisfies Condition G. If so, we must have:

$$\exists V \in \mathrm{Vec}, \quad \phi_t = \exp(tV) \circ \phi \ .$$

Moreover, since the path is horizontal and $\dot{\phi}_0 = V \circ \phi$, we infer:

$$\exists f \in \mathrm{Funct}_0^\mu, \quad V = \boldsymbol{\nabla} f \ ,$$

with f unique. Several questions arise, namely: does there exists an actual candidate path:

$$t \in [0,1] \to \phi_t = \exp(t\boldsymbol{\nabla} f) \circ \phi \in \mathrm{Diff} \tag{12}$$

satisfying $\mathcal{P}_\lambda(\phi_1) = \nu$? Is that path horizontal ? Does it realize the Arnol'd distance between its end points ?
Let us focus for the moment on the first question and consider, near $u = 0$, the local map E_μ defined by:

$$u \in \mathrm{Funct}_0^\mu \to E_\mu(u) := \exp(\boldsymbol{\nabla} u)_\# \mu \in \mathrm{Prob} \ .$$

Proposition 6. *The map E_μ is a diffeomorphism of a neighborhood of 0 in* Funct_0^μ *to a neighborhood of μ in* Prob.

Proof. The linearization at 0 of the map E_μ is readily found equal to:

$$\forall v \in \mathrm{Funct}_0^\mu, \quad dE_\mu(0)(v) = \Delta_\mu v \ \mu \in \mathrm{Mes}_0 \ .$$

By Theorem 6 of Appendix B, the map $dE_\mu(0) : \mathrm{Funct}_0^\mu \to \mathrm{Mes}_0$ is an elliptic isomorphism. Recalling that Prob is a domain in an affine space modelled on Mes_0, the proposition follows from the *elliptic* inverse function theorem [Del90] □

Using the local diffeomorphism E_μ, for each $\nu \in \mathrm{Prob}$ close enough to μ, we let $f := E_\mu^{-1}(\nu)$ in the path (12) and verify that, indeed, it satisfies:

$$\mathcal{P}_\lambda(\phi_1) = (\exp(\boldsymbol{\nabla} f) \circ \phi)_\# \lambda = \exp(\boldsymbol{\nabla} f)_\# \mu = E_\mu(f) = \nu$$

as required, with $\exp(t\boldsymbol{\nabla} f) \in \mathrm{Diff}$ for each $t \in [0,1]$. The first question (local existence) is thus settled.

Remark 1. Since the function $f = E_\mu^{-1}(\nu) \in \mathrm{Funct}_0^\mu$ is small, it has the following property:

Property NC: *For each $m \in M$ and $t \in [0,1]$, the points m and $\exp_m(t\boldsymbol{\nabla}_m f)$ are not cut points of each other.*

The latter readily implies the existence, for each $t \in [0,1]$, of a *unique* vector field $Z_t(f) \in \mathrm{Vec}$ such that:

$$\exp(Z_t(f)) = [\exp(t\boldsymbol{\nabla} f)]^{-1} \quad \text{in Diff} \, . \tag{13}$$

Horizontality

Let us turn to the second question.

Proposition 7. *The path (12) is horizontal.*

Proof. We require two preliminary steps.
Step 1: from a Riemannian lemma (see Appendix C), we have:

$$\forall m \in M, \forall t \in [0,1], \forall v \in T_pM \text{ with } p = \exp_m(t\boldsymbol{\nabla}_m f),$$
$$g_p\left[d\exp_m(t\boldsymbol{\nabla}_m f)(\boldsymbol{\nabla}_m f), v\right] = g_m\left[\boldsymbol{\nabla}_m f, d\exp_p(Z_t(f)_p)(v)\right] \ ,$$

where the vector field $Z_t(f)$ is the one defined in the previous remark.

Step2: we have,

$$\forall t \in [0,1], \forall \xi \in \ker \mathrm{div}_{\mu_t} \text{ (with } \mu_t = \mathcal{P}_\lambda(\phi_t)),$$
$$d\exp_{\exp(t\boldsymbol{\nabla} f)}(Z_t(f))(\xi \circ \exp(t\boldsymbol{\nabla} f)) \in \ker \mathrm{div}_\mu \ .$$

Indeed, fix $t \in [0,1]$ and $\xi \in \ker \mathrm{div}_{\mu_t}$, set for short $Z_t(f) = Z_t$ and consider the vector field:

$$\zeta = d\exp_{\exp(t\boldsymbol{\nabla} f)}(Z_t)(\xi \circ \exp(t\boldsymbol{\nabla} f)) \in \mathrm{Vec}\,.$$

Let Ψ_τ be the flow of ξ and Θ_τ the composed map given by

$$\Theta_\tau := \exp(Z_t) \circ \Psi_\tau \circ \exp(t\boldsymbol{\nabla} f)\ .$$

The one-parameter map Θ_τ satisfies, on the one hand $\frac{d\Theta_\tau}{d\tau}|_{\tau=0} = \zeta$, on the other hand:

$$(\Theta_\tau)_{\#}\,\mu = \exp(Z_t)_{\#}\mu_t = \mu$$

since Ψ_τ preserves the measure μ_t. Therefore, indeed, we have:

$$\mathrm{div}_\mu(\zeta)\ \mu = \frac{d}{d\tau}\,(\Theta_\tau)_{\#}\,\mu|_{\tau=0} = 0 \quad \square$$

We are in position to prove Proposition 7. Fix $t \in [0,1]$ and set $\dot{\phi}_t = V_t \circ \phi_t$. From Helmholtz decomposition (Corollary 6 of Appendix B), it suffices to pick an arbitrary $\xi \in \ker \mathrm{div}_{\mu_t}$ and check that the integral $\int_M g(V_t,\xi)\ d\mu_t$ vanishes. We compute:

$$\int_M g(V_t,\xi)\,d\mu_t = \int_M g(\dot{\phi}_t, \xi \circ \phi_t)\,d\lambda = \int_M g(d\exp(t\boldsymbol{\nabla} f)(\boldsymbol{\nabla} f), \xi \circ \exp(t\boldsymbol{\nabla} f))\,d\mu,$$

hence, by Step 1: $\int_M g(V_t,\xi)\ d\mu_t = \int_M g(\boldsymbol{\nabla} f,\zeta)\ d\mu$, with ζ defined (from ξ and t) as in the proof of Step 2. Now Step 2 implies the desired vanishing $\square$

From Proposition 7 combined with Proposition 3, we immediately get:

Corollary 2. *For $\mu_t = \mathcal{P}_\lambda(\phi_t)$ with ϕ_t given by (12) and for each $t \in [0,1]$, there exists a unique $f_t \in \mathrm{Funct}_0^{\mu_t}$, depending smoothly on t, such that:*

$$\frac{d}{dt} exp(t\boldsymbol{\nabla} f) = \boldsymbol{\nabla} f_t \ \circ \ exp(t\boldsymbol{\nabla} f) \ . \tag{14}$$

The smoothness of $t \mapsto f_t$ follows from the, linear elliptic, horizontal lift equation (10) satisfied by f_t. For later use, we observe that $f_0 \equiv f$.

Minimization Property and Equality $\mathbf{W_2 = d_O}$

Finally, does our candidate path (12) realize the Arnol'd distance between its end points ? Having no direct grasp on the question, let us just go ahead with what we can prove and try to find the answer on the way – a typical scientific attitude[5]. Doing so, we will record the following result, established differently in [Lot08, Proposition 4.24].

Proposition 8. *Let $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ be given by (12). Up to addition of a function of t only, the path $t \in [0,1] \to f_t \in \mathrm{Funct}_0^{\mu_t}$ associated to it in Corollary 2 satisfies the equation:*

$$\frac{\partial f_t}{\partial t} + \frac{1}{2}|\boldsymbol{\nabla} f_t|^2 = 0 \ . \tag{15}$$

Moreover, the function f must be c-convex on M.

Proof. Since the path (12) satisfies Condition G, we have:

$$\forall m \in M, \quad \frac{\nabla}{dt}\left(\frac{d\phi_t}{dt}(m)\right) = 0$$

which shows that its (Eulerian) velocity field $V_t = \boldsymbol{\nabla} f_t$ satisfies the so-called [K-M07] inviscid Burgers equation:

$$\frac{\partial V_t}{\partial t} + \nabla_{V_t} V_t = 0 \ . \tag{16}$$

The latter yields for f_t the equation:

$$\boldsymbol{\nabla}\left(\frac{\partial f_t}{\partial t} + \frac{1}{2}|\boldsymbol{\nabla} f_t|^2\right) = 0$$

or else, equation (15) as claimed.
Regarding the c-convexity on M of the function f, let us stress that it is not new; it holds because $\exp(\boldsymbol{\nabla} f) \in \mathrm{Diff}$ [Del04, Proposition 2]. Alternatively, though, we will derive it now from (15), thus bringing to light how it originates from Condition G.

5 *street-lamp paradigm*, as René Thom used to call it

As indicated in [Lot08, Remark 4.27] (see also [Vil08]), the solution of (15) in $\mathrm{Funct}_0^{\mu_t}$ is equal to $f_t = \widetilde{f}_t - \int_M \widetilde{f}_t \, d\mu_t$ with $\widetilde{f}_t$ given by the Hopf–Lax–Oleinik formula:

$$\forall m \in M, \quad \widetilde{f}_t(m) = \inf_{p \in M} \left[f(p) + \frac{1}{2t} \, d_g^2(p,m) \right] .$$

For $t = 1$, we infer:

$$\forall m \in M, \quad -f_1(m) = \sup_{p \in M} \left[\int_M \widetilde{f}_1 \, d\nu - f(p) - \frac{1}{2} \, d_g^2(p,m) \right] ,$$

so the function $-f_1$ is c-convex on M. It is convenient to set $f^c := -f_1$ with a slight abuse of notation due to the Funct_0^ν normalization. The function f^c is easily seen to satisfy:

$$\nabla f^c \circ \exp(\nabla f) = -\frac{d}{dt} \exp(t \nabla f)|_{t=1}$$

hence also:

$$\exp(\nabla f^c) \circ \exp(\nabla f) = \mathbb{I} , \tag{17}$$

or else: $\nabla f^c = Z_1(f)$, with the auxiliary notation introduced in Remark 1. So we may repeat the arguments of this section with the reversed path:

$$t \in [0,1] \to \psi_t := \exp(t \nabla f^c) \circ \phi_1$$

instead of the original one (12), thus switching (μ, f) and (ν, f^c). Doing so, we set $\nu_t := \mathcal{P}_\lambda(\psi_t) \equiv \mu_{1-t}$, let $f_t^c \in \mathrm{Funct}_0^{\nu_t}$ be given by Corollary 2 and reach the conclusion that the function $-f_1^c \equiv f$ is, indeed, c-convex on M $\square$

Equation (15), with f_t solving the horizontal lift equation (10), may be viewed in Prob equipped with the Otto metric as the *geodesic equation* bearing on the path $t \mapsto \mu_t$ [0-V00, Ott01, Lot08].

To further specify how the functions f and f^c are related, let us establish a key result [McC01, Lemma 7] by means of an elementary (smooth) proof.

Lemma 2. *Set $F_f : M \times M \to \mathbb{R}$ for the auxiliary function given by:*

$$F_f(q_1, q_2) := f(q_1) + f^c(q_2) + \frac{1}{2} d_g^2(q_1, q_2) ,$$

and Σ_f, for the submanifold of $M \times M$ defined by:

$$\Sigma_f := \left\{ (q_1, q_2) \in M^2, \ q_2 = \exp_{q_1}(\nabla_{q_1} f) \right\} .$$

The function F_f is constant on Σ_f where it assumes a global minimum equal to $\int_M \widetilde{f}_1 \, d\nu$.

Proof. From the above expression of $-f_1 = f^c$, we know that $F_f \geq \int_M \widetilde{f_1}\, d\nu$. By a classical property of the function $\frac{1}{2}d_g^2$ (recalled in Appendix C), the function F_f satisfies $dF_f = 0$ at each point of Σ_f. In particular, it must be *constant* on the submanifold Σ_f since the latter is connected. To evaluate that constant, recalling:

$$f^c(q_2) = \sup_{q_1 \in M} \left[\int_M \widetilde{f_1}\, d\nu - f(q_1) - \frac{1}{2} d_g^2(q_1, q_2) \right] ,$$

we pick for q_2 a point $m \in M$ where f assumes its global *minimum* (such a point exists because M is compact). With $q_2 = m$, we observe that the (continuous) real function: $q_1 \in M \to -f(q_1) - \frac{1}{2}d_g^2(q_1, m)$ must assume a global *maximum* at $q_1 = m$. So $f^c(m) = \int_M \widetilde{f_1}\, d\nu - f(m)$ and $F_f = \int_M \widetilde{f_1}\, d\nu$ on Σ_f as claimed □

With Lemma 2 at hand, we can cope with the minimization question:

Corollary 3. *The path* $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ *given by (12) realizes the Arnol'd distance between its end points. Moreover,* $\mathrm{W}_2 = \mathrm{d}_O$.

Proof. Let $\psi : M \to M$ be a Borel map satisfying $\psi_{\#}\mu = \nu$. By Lemma 2, it satisfies $\int_M F_f(m, \psi(m))\, d\mu \geq \int_M \widetilde{f_1}\, d\nu$; since $f \in \mathrm{Funct}_0^\mu$ and $f^c \in \mathrm{Funct}_0^\nu$, we may readily rewrite this inequality as:

$$\int_M \frac{1}{2} d_g^2(m, \psi(m))\, d\mu \geq \int_M \widetilde{f_1}\, d\nu \ ,$$

with equality holding if $\psi = \exp(\boldsymbol{\nabla} f)$. From the latter, we infer:

$$\int_M \frac{1}{2} d_g^2(m, \psi(m))\, d\mu \geq \int_M \frac{1}{2} d_g^2(m, \exp_m(\boldsymbol{\nabla}_m f))\, d\mu \ .$$

Taking the infimum of the left-hand side over all measurable maps ψ such that $\psi_{\#}\mu = \nu$, we obtain:

$$\mathrm{W}_2(\mu, \nu) = \sqrt{\int_M \frac{1}{2} |\boldsymbol{\nabla} f|^2\, d\mu} \ .$$

The latter right-hand side is nothing but the Arnol'd length of the path (12) which, by Proposition 7, is horizontal; it thus coincides with the Otto length of the path $t \in [0,1] \to \mu_t = \mathcal{P}_\lambda(\phi_t) \in \mathrm{Prob}$. Recalling the inequality $\mathrm{W}_2 \leq \mathrm{d}_O$ proved above (after Theorem 3), we conclude that the path (12) is, indeed, minimizing for the Arnol'd distance and that we have:

$$\mathrm{d}_A(\phi, \phi_1) = \mathrm{d}_O(\mu, \nu) = \mathrm{W}_2(\mu, \nu) \ .$$

Now, the present proof shows that the equality $\mathrm{W}_2 = \mathrm{d}_O$ holds near the diagonal of $\mathrm{Prob} \times \mathrm{Prob}$. As in any length space, this result suffices to conclude that it holds everywhere □

4.3 Horizontal Segments in the Large

Dropping the closeness assumption on the assigned probability measures $(\mu, \nu) \in \text{Prob} \times \text{Prob}$, should we still consider the path (12) as a good candidate to solve the problem posed at the beginning of section 4.2 ? Deferring till Section 5 a tentative answer to the question, let us record global arguments in favour of such a path.

A Reinhart Lemma

It is a standard fact from Riemannian foliations theory, which goes back to [Rei59], that the horizontal distribution $\phi \in \text{Diff} \to \mathcal{H}_\phi$ is totally geodesic. Specifically, we have:

Lemma 3. *Let $t \in I \to \phi_t \in \text{Diff}$ be a geodesic (for the Arnol'd metric) defined on some interval $I \subset \mathbb{R}$. If, for some value of the parameter t, the velocity $\dot{\phi}_t = \frac{d\phi_t}{dt}$ is horizontal, it remains so for all $t \in I$.*

Proof. Let us argue by *connectedness* on the closed non-empty subset:

$$\mathcal{T} := \left\{ t \in I,\ \dot{\phi}_t \in \mathcal{H}_{\phi_t} \right\} .$$

We only have to prove that $\mathcal{T}$ is relatively open in the interval I. To do so, fix $T \in \mathcal{T}$ and set:

$$\dot{\phi_T} = \boldsymbol{\nabla} f_T \ \circ\ \phi_T \ .$$

By Proposition 7 combined with Corollary 3, for $\epsilon > 0$ small enough, the path

$$t \in (T-\epsilon, T+\epsilon) \cap I \to \psi_t := \exp\left((t-T)\boldsymbol{\nabla} f_T\right) \circ \phi_T \in \text{Diff}$$

is a horizontal geodesic. Since its position ψ_T and velocity $\dot{\psi_T}$ at time $t = T$ coincide with those of our original path $t \mapsto \phi_t$, both paths must coincide hence $(T-\epsilon, T+\epsilon) \cap I \subset \mathcal{T}$ as desired □

Necessity of Condition G

Let us provide a metric proof of the following partial converse to Proposition 5 (the full converse is proved differently in [E-M70]).

Proposition 9. *Any horizontal minimizing constant speed (HMCS, for short) geodesic for the Arnol'd metric must satisfy Condition G.*

Proof. Let the path $t \in [0,1] \to \phi_t \in \text{Diff}$ be HMCS. The constant speed assumption means that the total kinetic energy of the motion ϕ_t on the manifold M at time t, namely the quantity:

$$E(t) := \int_M \frac{1}{2} |\dot{\phi}_t|^2_{\phi_t} d\mu \ ,$$

is independent of $t \in [0,1]$. Its constancy implies that the squared Arnol'd distance:

$$\mathrm{d}_A^2(\phi_0, \phi_1) = \left(\int_0^1 \sqrt{\langle \dot{\phi}_t, \dot{\phi}_t \rangle_{\phi_t}} \; dt \right)^2$$

is equal to the total energy of the geodesic $t \in [0,1] \to \phi_t \in \mathrm{Diff}$, namely to $E = \int_0^1 E(t)dt$. Fubiny theorem thus provides:

$$\mathrm{d}_A^2(\phi_0, \phi_1) = \int_M \frac{1}{2} \left(\int_0^1 |\dot{\phi}_t(m)|^2_{\phi_t(m)} dt \right) d\mu$$

hence, by Schwarz inequality:

$$\mathrm{d}_A^2(\phi_0, \phi_1) \geq \int_M \frac{1}{2} \left(\int_0^1 |\dot{\phi}_t(m)|_{\phi_t(m)} dt \right)^2 d\mu \ . \tag{18}$$

Since $\displaystyle\int_0^1 |\dot{\phi}_t(m)|_{\phi_t(m)} dt$ is the length of the path $t \in [0,1] \to \phi_t(m) \in M$, we have identically:

$$\int_0^1 |\dot{\phi}_t(m)|_{\phi_t(m)} dt \geq d_g(\phi_0(m), \phi_1(m)) \ . \tag{19}$$

Combining the two inequalities yields:

$$\mathrm{d}_A^2(\phi_0, \phi_1) \geq \int_M \frac{1}{2} d_g^2(\phi_0(m), \phi_1(m)) \; d\mu \ ,$$

or else, setting $\mu_t := \mathcal{P}_\lambda(\phi_t)$ and $\psi := \phi_1 \circ \phi_0^{-1}$,

$$\mathrm{d}_A^2(\phi_0, \phi_1) \geq \int_M \frac{1}{2} d_g^2(m, \psi(m)) \; d\mu_0 \ .$$

Noting that $\psi_\# \mu_0 = \mu_1$ and recalling the second part of Corollary 3, we conclude:

$$\mathrm{d}_A^2(\phi_0, \phi_1) \geq \mathrm{d}_O^2(\mu_0, \mu_1) \ . \tag{20}$$

But the path $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ being minimizing horizontal, the projection $\mathcal{P}_\lambda$ restricted along it is an isometry, hence equality must hold in (20). It implies that it must also hold in (18) and (19). For each $m \in M$, equality in (19) forces the path $t \in [0,1] \mapsto \phi_t(m) \in M$ to be minimizing, while equality in (18) forces it to have constant speed □

5 Conclusion: Heuristical Statement

Back to the global question stated at the beginning of section 4.3, we are now in position to prove the following heuristical result:

Theorem 4. *Given* $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$*, the following properties are equivalent:*

(i) *there exists a constant speed geodesic* $t \in [0,1] \to \mu_t \in \mathrm{Prob}$ *with end points* $\mu_0 = \mu, \mu_1 = \nu$*, which is minimizing for the Otto metric;*
(ii) *there exists a unique function* $f \in \mathrm{Funct}_0^\mu$ *c-convex on* M *solving the transport equation (2);*
(iii) *there exists a unique function* $f \in \mathrm{Funct}_0^\mu$ *solving the Monge–Ampère equation (4);*
(iv) *there exists a unique function* $f \in \mathrm{Funct}_0^\mu$ *c-convex on* M *such that the path* $t \in [0,1] \to \mu_t := \exp(t\boldsymbol{\nabla} f)_\# \mu \in \mathrm{Prob}$ *is MCS with* $\exp(t\boldsymbol{\nabla} f) \in \mathrm{Diff}$ *and* $\mu_1 = \nu$.

Proof. The implication $(iv) \Rightarrow (ii)$ is trivial, while $(ii) \Rightarrow (iii)$ follows from Theorem 5 of Appendix A. The implication $(iii) \Rightarrow (i)$ holds by Theorem 2 combined with the remark which follows Lemma 1 (see (3)). We are thus left with proving that $(i) \Rightarrow (iv)$, which we now do.
Assume (i) and let $t \in [0,1] \to \phi_t \in \mathrm{Diff}$ be a horizontal lift of the path μ_t. By Proposition 9, it must satisfy Condition G hence, being horizontal at $t = 0$, it can be expressed as:

$$\phi_t = \exp(t\boldsymbol{\nabla} f) \circ \phi_0$$

for a unique $f \in \mathrm{Funct}_0^\mu$. The horizontality of this expression for all time is now guaranted by Lemma 3. By writing $\exp(t\boldsymbol{\nabla} f) = \phi_t \circ \phi_0^{-1}$, we see that $\exp(t\boldsymbol{\nabla} f) \in \mathrm{Diff}$ while, by the final statement of Proposition 8, we know that the function f is c-convex on M $\square$

A Jacobian Equation and Related Properties of Smooth Transport Maps

Given $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$ and a smooth map $\phi : M \to M$ pushing μ to ν, if ϕ is a diffeomorphism, it must satisfy the (pointwise) equation:

$$\phi^* \nu = \mu \ . \tag{21}$$

Indeed, making the change of variable $p \mapsto m = \phi(p)$ in the left-hand integral of (1), we get:

$$\int_M (u \circ \phi) \ d(\phi^* \nu) = \int_M (u \circ \phi) d\mu$$

which yields (21) since the function u is arbitrary. One often calls (21) the *Jacobian equation* of the (μ to ν) transport. Here, we wish to weaken the assumption on ϕ:

Proposition 10. *Assume only that the smooth map* $\phi : M \to M$ *pushing* μ *to* ν *is one-to-one. If so, it must still satisfy the Jacobian equation (21).*

Proof. Let E be the set of points of M at which the Jacobian equation is satisfied. If $m \in E$, (21) implies that m is not critical for the map ϕ since μ and ν nowhere vanish. The inverse function theorem implies the existence of a small enough ball B around m such that ϕ induces a *diffeomorphism* from B to its image Ω. Since ϕ is one-to-one on M pushing μ to ν, we have: $\mu(B) = \nu(\Omega)$. Restricting ϕ and μ to B, ν to Ω, we can argue as above and conclude that B lies in E. So the set E is both closed (since ϕ is smooth) and *open* in the manifold M. By connectedness, the proof is reduced to showing that E is non-empty.
We prove the latter by contradiction. If $E = \emptyset$, Sard theorem [Mil65] implies that the image set $\phi(M)$ has zero measure. Besides, it is closed, since M is compact. So we may pick a function u supported inside its *complement* (a dense open subset of M) with $\int_M u \, d\nu \neq 0$. With this choice of u in (1), we reach a contradiction □
Although the result just proved is certainly well-known, we did not find any simple proof of it in the literature (see e.g. [Vil08, Chapter 11] and references therein). It implies at once the following corollary:

Corollary 4. *Given* $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$ *and a smooth map* $\phi : M \to M$*, the following properties are equivalent:*

(i) ϕ *is one-to-one and satisfies* $\phi_\# \mu = \nu$*;*
(ii) ϕ *satisfies the Jacobian equation* $\phi^* \nu = \mu$*.*

In either case, the map ϕ *must be a diffeomorphism.*

Indeed, we know that $(i) \Rightarrow (ii)$ by Proposition 10, while (ii) implies that ϕ is a diffeomorphism [Del08, Lemma 4] and (i) follows by the above change of variable argument □

Regarding maps of the form $\phi = \exp(\boldsymbol{\nabla} f)$, we obtain:

Theorem 5. *Given* $(\mu, \nu) \in \mathrm{Prob} \times \mathrm{Prob}$ *and a smooth real function* f *on* M*, the following properties are equivalent:*

(i) the function f *is c-convex on* M *and satisfies* $\exp(\boldsymbol{\nabla} f)_\# \mu = \nu$*;*
(ii) the map $\exp(\boldsymbol{\nabla} f)$ *is one-to-one and satisfies* $\exp(\boldsymbol{\nabla} f)_\# \mu = \nu$*;*
(iii) the function f *satisfies the Monge–Ampère equation* $\exp(\boldsymbol{\nabla} f)^* \nu = \mu$*.*

If so, for each $t \in [0, 1]$*, the map* $\exp(t \boldsymbol{\nabla} f)$ *is a diffeomorphism and the function* tf *is c-convex on* M*.*

Proof. The equivalence between (ii) and (iii) holds by Corollary 4. One can readily infer (iii) from (i) relying on [CMS01, Theorem 4.2]. Assuming (iii), recalling that $\exp(\boldsymbol{\nabla} f)$ must be a diffeomorphism (cf. supra), the c-convexity of f is established in [Del04, Proposition 2]. The final statement of the theorem holds, assuming (iii), by [Del08, Theorem 2, Proposition 5 and Remark 6] (alternatively, the c-convexity of tf follows also from (i) and Lemma 1) □

Remark 2. The reader may get confused by the openings of [Del04, Del08] because, in both papers, we viewed smooth measures like n-forms and their push-forward by a diffeomorphism ϕ, like the (pointwise!) pull-back by ϕ^{-1} (which is, of course, *stronger* than the measure transport definition). So, for instance in [Del08, p.327–328], he should assume the c-convexity of the solution of the optimal transport equation (whereas, in a pointwise acceptance of that equation, it is a priori guaranted).

B The Helmholtz Decomposition of Vector Fields

Given a smooth positive measure μ on the compact manifold M (with μ taken in Prob to comply with the normalization of this paper), we can associate to it a differential operator, called the *divergence* (with respect to μ), as follows:

Definition 1. *The divergence operator* $\operatorname{div}_\mu : \mathrm{Vec} \to \mathrm{Funct}_0^\mu$ *is defined, for each vector field* V *on* M *with flow* ϕ_t*, by the formula:*

$$\operatorname{div}_\mu(V)\ \mu = \frac{d}{dt}\left(\phi_{t\#}\mu\right)_{t=0} \in \mathrm{Mes}_0\,.$$

Let us record the main properties of the divergence operator.

Proposition 11. *Given* $V \in \mathrm{Vec}$*, the measure* $\operatorname{div}(V)_\mu\ \mu$ *satisfies:*

$$\int_M f \operatorname{div}_\mu(V)\ d\mu \equiv \int_M df(V)\ d\mu\ , \tag{22}$$

for each $f \in \mathrm{Funct}$.

Proof. Fix $V \in \mathrm{Vec}$ and set ϕ_t for the flow of V. For each $f \in \mathrm{Funct}$ and each real t, Definition 1 yields:

$$\int_M f\ d(\phi_{t\#}\mu) = \int_M (f \circ \phi_t)\ d\mu\ ,$$

and (22) is obtained by differentiating both sides with respect to t at $t = 0$ □

From Proposition 11, using the Riemannian metric g, we can rewrite the identity (22) as:

$$\forall (V, f) \in \mathrm{Vec} \times \mathrm{Funct}_0^\mu, \quad \int_M f \operatorname{div}_\mu(V)\ d\mu = \int_M g(\boldsymbol{\nabla} f, V)\ d\mu\ , \tag{23}$$

which shows that, with $\boldsymbol{\nabla} : \mathrm{Funct} \to \mathrm{Vec}$ restricted to Funct_0^μ, the divergence and gradient operators are formally *adjoint* of each other with respect to the following L^2 scalar products:

$$\langle f, f' \rangle := \int_M f f' d\mu\ , \quad \langle V, V' \rangle := \int_M g(V, V') d\mu\ ,$$

defined in Funct_0^μ and Vec respectively. We will use the following key result:

Theorem 6. *The second order differential operator* $\Delta_\mu : \mathrm{Funct}_0^\mu \to \mathrm{Funct}_0^\mu$ *defined by:*

$$\forall f \in \mathrm{Funct}_0^\mu, \quad \Delta_\mu f := \mathrm{div}_\mu(\boldsymbol{\nabla} f) \ ,$$

is self-adjoint and elliptic. Moreover, it is an automorphism.

Let us call the operator Δ_μ the μ-Laplacian (when μ is the Lebesgue measure of the metric g, it coincides with the Laplacian of g). To see that the μ-Laplacian is one-to-one on Funct_0^μ, pick a function f in its kernel and infer from (23) with $V = \boldsymbol{\nabla} f$ that $\boldsymbol{\nabla} f = 0$ hence $f = 0$. The proof that Δ_μ is onto (thus an automorphism, by the open mapping theorem) relies on its self-adjointness (which holds by construction) combined with standard elliptic regularity theory and the Fredholm alternative. We skip the argument since it is lengthy but classical (see e.g. [G-T83][Bes87, Appendix]).

Corollary 5. *Assume* $\mu \in \mathrm{Prob}$ *and let* $\tilde{\mu} \in \mathrm{Mes}_0$. *There exists a unique* $f \in \mathrm{Funct}_0^\mu$ *solving the equation:*

$$\Delta_\mu f \ \mu = \tilde{\mu} \ .$$

Corollary 6 (Helmholtz decomposition). *The following splitting holds, with* L^2 *orthogonality (relative to* g *and* μ*) of its factors:*

$$\mathrm{Vec} = \mathrm{Im}\, \boldsymbol{\nabla} \ \oplus \ \ker \mathrm{div}_\mu \ .$$

The first corollary follows at once from Theorem 6. To prove Corollary 6, pick $V \in \mathrm{Vec}$ and use Theorem 6 to solve uniquely for $f \in \mathrm{Funct}_0^\mu$ the equation:

$$\Delta_\mu f = \mathrm{div}_\mu(V) \ .$$

The latter implies $(V - \boldsymbol{\nabla} f) \in \ker \mathrm{div}_\mu$ so the splitting, indeed, holds. The orthogonality of its factors now follows from (23) $\square$
Note that the weak form of the preceding equation, namely:

$$\int_M g(\boldsymbol{\nabla} f, \boldsymbol{\nabla} f') \ d\mu = \int_M g(V, \boldsymbol{\nabla} f') \ d\mu \ ,$$

can be solved just by the Riesz representation theorem applied in the completion of Funct_0^μ for the Hilbert scalar product defined by the left-hand side, since the linear form defined by the right-hand side is continuous (by Schwarz inequality).

C Complement to Gauss Lemma

In this appendix, we provide a result[6] of local Riemannian geometry, namely an adjointness property of the exponential map, which may be viewed as a complement to Gauss Lemma (see e.g. [C-E75, p.6] [doC92, pp.69–70]).

[6] which we could not find in the literature

Lemma 4. *Let (M,g) be a Riemannian manifold and $(m,p) \in M \times M$, a couple of points, not cut points of each other, which may be joined by a minimizing geodesic. Set $v \in T_mM$ and $v^c \in T_pM$ for the (unique) tangent vectors of smallest length such that:*

$$p = \exp_m(v) \text{ and } m = \exp_p(v^c).$$

For each couple of tangent vectors $(w,z) \in T_mM \times T_pM$, we have:

$$g_p\left(d\exp_m(v)(w), z\right) = g_m\left(w, d\exp_p(v^c)(z)\right).$$

In other words, the linear isomorphisms:

$$d\exp_m(v) : (T_mM, g_m) \to (T_pM, g_p)\ ,$$

$$d\exp_p(v^c) : (T_pM, g_p) \to (T_mM, g_m)$$

are adjoint *of each other.*

Proof. Setting $c = \frac{1}{2}d_g^2$, for short, the function c is smooth in a neighborhood $\mathcal{N}$ of (m,p) in $M \times M$; we denote by (q_1,q_2) the generic point of $\mathcal{N}$. As well known, for each $q \in M$ and $u \in T_qM$ of smallest length such that $(q, \exp_q(u)) \in \mathcal{N}$, we have (see e.g. [Jos95, p.256]):

$$-\left(d_{q_1}c\right)\left(q, \exp_q(u)\right)(.) \equiv g_q(u,.)\ .$$

Let us differentiate this identity with respect to $u \in T_qM$ and read the result at $(q,u) = (m,v)$. We get:

$$\forall w \in T_mM, \quad -\left(d^2_{q_1,q_2}c\right)(q,p)\left[., d\exp_m(v)(w)\right] = g_m(w,.)\ . \tag{24}$$

Similarly, for each $\tilde{q} \in M$ and $\tilde{u} \in T_{\tilde{q}}M$ of smallest length such that $(\exp_{\tilde{q}}(\tilde{u}), \tilde{q}) \in \mathcal{N}$, we have:

$$-\left(d_{q_2}c\right)\left(\exp_{\tilde{q}}(\tilde{u}), \tilde{q}\right)(.) \equiv g_{\tilde{q}}(\tilde{u},.)\ ,$$

which yields at $(\tilde{q},\tilde{u}) = (p, v^c)$:

$$\forall z \in T_pM, \quad -\left(d^2_{q_1,q_2}c\right)(q,p)\left[d\exp_p(v^c)(z), .\right] = g_p(z,.)\ . \tag{25}$$

Now, applying (24) to the vector $d\exp_p(v^c)(z) \in T_mM$ and (25), to the vector $d\exp_m(v)(w) \in T_pM$, produces the *same result*, due to the symmetry of the quadratic form $d^2_{q_1,q_2}c$ (Schwarz theorem). Identifying the resulting right-hand sides, we obtain the lemma $\square$

References

[Arn66] Arnol'd, V.: Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier, **16**, 319–361 (1966)

[Bes87] Besse, A.: Einstein Manifolds. Erg. Math., Vol. **10**, Springer–Verlag (1987)

[Bre91] Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Math., **44**, 375–417 (1991)

[doC92] do Carmo, M.: Riemannian Geometry. Birkhäuser, Boston (1992)

[C-E75] Cheeger, J., Ebin, D.: Comparison Theorems in Riemannian Geometry. North–Holland Mathematical Library, Vol. **9**, North–Holland Publishing Company, Amsterdam, Oxford (1975)

[CMS01] Cordero–Erausquin, D., McCann, R., Schmuckenschläger, M.: A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. Invent. math., **146**, 219–257 (2001)

[Del90] Delanoë, P.: Local inversion of elliptic problems on compact manifolds. Math. Japonica, **35**, 679–692 (1990)

[Del04] Delanoë, P.: Gradient rearrangement for diffeomorphisms of a compact manifold. Diff. Geom. Appl., **20**, 145–165 (2004)

[Del08] Delanoë, P.: Lie solutions of Riemannian transport equations on compact manifolds. Diff. Geom. Appl., **26**, 327–338 (2008)

[E-M70] Ebin, D., Marsden, J.: Groups of diffeomorphims and the motion of an incompressible fluid. Ann. Math., **92**, 102–163 (1970)

[FIP04] Falcitelli, M., Ianus, S., Pastore, A.-M.: Riemannian Submersions and Related Topics. World Scientific, Singapore (2004)

[G-T83] Gilbarg, D., Trudinger, N.: Elliptic Partial Differential Equations of Second Order. Second edition, Springer–Verlag, Berlin Heidelberg (1983)

[Jos95] Jost, J.: Riemannian Geometry and Geometric Analysis. Springer Universitext, Springer–Verlag, Berlin (1995)

[K-L08] Khesin, B., Lee, P.: A nonholonomic Moser theorem and optimal mass transport. Preprint, downloable at http://arxiv.org/abs/0802.1551 (2008)

[K-M07] Khesin, B., Misiołek, G.: Shock waves for the Burgers equation and curvatures of diffeomorphism groups. Proceed. Steklov Inst. Math., **259**, 73–81 (2007)

[Lan62] Lang, S.: Introduction to Differentiable Manifolds. John Wiley & Sons, New–York (1962)

[Lot08] Lott, J.: Some geometric calculations on Wasserstein space. Comm. Math. Phys., **277**, 423–437 (2008)

[Lyc79] Lychagin, V.: Contact geometry and nonlinear second order differential equations. Russian Math. Surv., **34**, 149–180 (1979)

[McC01] McCann, R.: Polar factorization of maps on Riemannian manifolds. Geom. Funct. Anal., **11**, 589–608 (2001)

[Mil65] Milnor, J.: Topology from the Differentiable Viewpoint. The University Press of Virginia, Charlottesville (1965)

[Mon81] Monge, G.: Mémoire sur la théorie des déblais et remblais. Mémoires de l'Acad. Royale Sci. Paris (1781).

[Mos65] Moser, J.: On the volume elements on a manifold. Trans. Amer. Math. Soc., **120**, 286–294 (1965)

[Ott01] Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Diff. Eq., **26**, 101–174 (2001)

[0-V00] Otto, F., Villani, C.: Generalization of an inequality by Talagrand, and links with the logarithmic Sobolev inequality. J. Funct. Anal., **173**, 361–400 (2000)

[R-R98] Rachev, S., Rüschendorf, L.: Mass Transportation Problems, Vol. I: Theory. Probability and its Applications, Springer–Verlag, New–York (1998)

[Rei59] Reinhart, B.: Foliated manifolds with bundle-like metrics. Ann. Math., **69**, 119–131 (1959)

[dRh55] de Rham, G.: Variétés différentiables. Hermann, Paris (1955)

[Rie09] Riesz, F.: Sur les opérations fonctionnelles linéaires. C. R. Acad. Sci. Paris, **149**, 974–977 (1909)

[Vil03] Villani, C.: Topics in Mass Transportation. Grad. Stud. Math., AMS, Providence R.I. (2003)

[Vil08] Villani, C.: Optimal Transport, Old and New. Grundlehren der mathematischen Wissenschaften **338**, Springer–Verlag (2009)