METHODS AND APPLICATIONS OF ANALYSIS.
Vol. 18, No. 2, pp. 127–158, June 2011
001

ON THE INVERSE IMPLICATION OF BRENIER-MCCANN THEOREMS AND THE STRUCTURE OF $(\mathscr{P}_2(M), W_2)$*

NICOLA GIGLI†

Abstract. We do three things. First, we characterize the class of measures $\mu \in \mathscr{P}_2(M)$ such that for any other $\nu \in \mathscr{P}_2(M)$ there exists a unique optimal transport plan, and this plan is induced by a map. Second, we study the tangent space at any measure and we identify the class of measures for which the tangent space is an Hilbert space. Third, we prove that these two classes of measures coincide. This answers a question recently raised by Villani. Our results concerning the tangent space can be extended to the case of Alexandrov spaces.

Key words. Optimal transport map, tangent space.

AMS subject classifications. 49Q20.

Introduction. Among the several papers devoted to the study of mass transportation problems, two can certainly be called cornerstones of the theory: the work of Brenier [6] (together with the generalization to the case of Riemannian manifolds due to McCann [25]) where existence, uniqueness and structure of the optimal transport map is established, and the work of Otto [28], where the Riemannian structure of the space $(\mathscr{P}_2(M), W_2)$ is described.

The theory has been deeply studied in the past years. A topic which became suddenly clear, in particular for what concerns the Riemannian structure of the space of measures, is the fact that there are 'good' measures (like absolutely continuous ones) near which the Riemannian structure behaves nicely, and 'bad' measures (like deltas) at which such structure degenerates. The precise borderline between these two kind of measures was up to now not completely understood, and the question of finding the 'right' structure of the space $(\mathscr{P}_2(M), W_2)$ was also recently posed in Villani's monograph [33].

The problem of the gray area between 'good' measures and 'bad' ones appears also in Brenier-McCann theorems. Indeed, the typical statement of such theorem is: Assume that $\mu, \nu \in \mathscr{P}_2(M)$ are such that μ gives 0 mass to $\dim(M) - 1$ dimensional sets, then there exists a unique optimal transport plan, and such plan is induced by a map (where a structural characterization of the map in terms of Kantorovich potential is also given). Now, the point is that the assumption made on μ, although clearly sufficient to get the conclusion, is not necessary. Given the fundamental importance of the Brenier-McCann theorems, it is natural to look for the sharp hypothesis in their statement.

The aim of this paper is to clarify the situation, our main results being:
- the characterization of those measures to which Brenier-McCann theorem applies (Propositions 2.4 and 2.10),
- the identification of the tangent space at any measure μ (theorem 5.5),
- the proof of the fact that the class of measures for which the tangent space is an Hilbert space coincides with the class of measures to which Brenier-McCann theorem applies (corollary 6.6). Also, in this case the tangent space is naturally identified with the well known 'space of gradients'.

*Received November 19, 2009; accepted for publication March 14, 2011.
†Department of Mathematics, University of Nice, Valrose 06108 Nice Cedex 02, France (gigli@unice.fr). Partially financed by KAM Faible, ANR-07-BLAN-0361.

From a purely geometric perspective, some of our results apply also to the case of Alexandrov spaces with curvature bounded from below. In particular, the description of the tangent space at a certain measure provided by theorem 3.4 is a sharper statement than the analogous appeared in [27].

I would like to thank Luigi Ambrosio and Shin-ichi Ohta for valuable comments at the early stage of development of this work.

1. Preliminaries and notation. M is a fixed smooth, connected Riemannian manifold without boundary, d its Riemannian distance.

The natural set to endow with the Wasserstein distance is the set $\mathscr{P}_2(M)$ of Borel probability measures with bounded second moment:

$$\mathscr{P}_2(M) := \left\{ \mu \in \mathscr{P}(M) \; : \; \int \mathrm{d}^2(x, x_0) d\mu(x) < \infty \quad \forall x_0 \in M \right\}.$$

The set $\mathcal{P}_c(M) \subset \mathscr{P}_2(M)$ is the set of Borel probability measures with compact support.

Recall that for any couple of topological spaces X, Y, any Borel probability measure μ on X and any Borel map $f : X \to Y$, the push forward $f_\# \mu$ of μ through f is the Borel probability measure on Y defined by

$$f_\# \mu(E) := \mu\big(f^{-1}(E)\big), \qquad \forall \text{ Borel sets E} \subset \text{Y}.$$

The Wasserstein distance W_2 on $\mathscr{P}_2(M)$ is defined by

$$W_2(\mu, \nu) := \sqrt{\inf \int \mathrm{d}^2(x, y) d\boldsymbol{\gamma}(x, y)},$$

where the infimum is taken in the set $\mathscr{Adm}(\mu, \nu)$ of *admissible plans* $\boldsymbol{\gamma}$ from μ to ν, i.e. among all the probability measures on M^2 satisfying $\pi^1_\# \boldsymbol{\gamma} = \mu$ and $\pi^2_\# \boldsymbol{\gamma} = \nu$, where π^1 and π^2 are the projections onto the first and second coordinate respectively. The quantity $\int \mathrm{d}^2(x, y) d\boldsymbol{\gamma}(x, y)$ is called the *cost* of the plan $\boldsymbol{\gamma}$. A plan which realizes the infimum is called *optimal* and the set of optimal plans for a given couple (μ, ν) of measures will be indicated by $\mathscr{Opt}(\mu, \nu)$. A plan is said to be *induced by a map*, if it is of the form $(Id, T)_\# \mu$ for some measurable map T. This is the same as to say that $\boldsymbol{\gamma}$ is concentrated on the graph of T.

It is well known that the function W_2 is a distance on $\mathscr{P}_2(M)$ and that the space $(\mathscr{P}_2(M), W_2)$ is Polish; we skip the proof this fact: the interested reader may study the question in detail on, for instance, [33, Chapter 6].

A central question of the theory is: when do we know that there is only one optimal transport plan, and that this plan is induced by a map? The answer to this question is given by Brenier-McCann theorems (Brenier's theorem concerns with the Euclidean case, while McCann's one generalizes to the case of Riemannian manifolds), and the proof is essentially divided in the following steps, each of which has its independent interest. Here and in the following we will put $c(x, y) := \frac{\mathrm{d}^2(x,y)}{2}$. One proves that:

- a plan is optimal if and only if its support is c-cyclically monotone (definition 1.1 and theorem 1.2 below),
- a set $\Gamma \subset M^2$ is c-cyclically monotone if and only if there is a c-concave function φ (definition 1.4) such that Γ is contained in the graph $\partial^{c+} \varphi$ of the c-superdifferential of φ (definition 1.5 and theorem 1.6 below),

- $\partial^{c+}\varphi \subset \exp(-\partial^{+}\varphi)$, where $\partial^{+}\varphi$ is the superdifferential of φ (Theorem 1.8 below).

We will skip most of the proofs, as these are well known statements in the theory.

Definition 1.1 (c-cyclical monotonicity). *A subset K of $M \times M$ is c-cyclically monotone if for every $n \in \mathbb{N}$, every $(x_i, y_i) \in K$, $i = 0, \dots, n-1$, and every permutation σ of $\{0, \dots, n-1\}$ it holds:*

$$\sum_{i=0}^{n-1} c(x_i, y_i) \leq \sum_{i=0}^{n-1} c(x_i, y_{\sigma(i)}).$$

Theorem 1.2. *Let $\mu, \nu \in \mathscr{P}_2(M)$. A plan $\boldsymbol{\gamma} \in \mathcal{A}dm(\mu, \nu)$ is optimal if and only if its support* $\operatorname{supp}(\boldsymbol{\gamma})$ *is a c-cyclically monotone set.*

Definition 1.3 (c_+ transform). *Let $\psi : M \to \mathbb{R} \cup \{-\infty\}$. The function ψ^{c+} : $M \to \mathbb{R} \cup \{-\infty\}$ is defined as*

$$\psi^{c+}(x) := \inf_{y \in M} \left(c(x, y) - \psi(y) \right),$$

Observe that we have the following trivial inequality:

$$\psi(x) + \psi^{c+}(y) \leq c(x, y), \qquad \forall x, y \in M.$$

Definition 1.4 ($c-$concavity). *We say that $\varphi : M \to \mathbb{R} \cup \{-\infty\}$ is $c-$concave if it is not identically $-\infty$ and there exists $\psi : M \to \mathbb{R} \cup \{-\infty\}$ such that*

$$\varphi = \psi^{c+}.$$

Definition 1.5 ($c-$superdifferential). *Let $\varphi : M \to \mathbb{R} \cup \{-\infty\}$ be a $c-$concave function. Its $c-$superdifferential $\partial^{c+}\varphi \subset M^2$ is defined as*

$$\partial^{c+}\varphi := \left\{ (x, y) \ : \ \varphi(x) + \varphi^{c+}(y) = c(x, y) \right\},$$

and the $c-$superdifferential $\partial^{c+}\varphi(x)$ at a point $x \in M$ is the set of y such that $(x, y) \in \partial^{c+}\varphi$.

Theorem 1.6. *Let $\Gamma \subset M^2$. Then Γ is c-cyclically monotone if and only if $\Gamma \subset \partial^{c+}\varphi$ for some c-concave function φ.*

Given $\mu, \nu \in \mathscr{P}_2(M)$, we will say that a c-concave function φ is a Kantorovich potential for the couple (μ, ν) if $\partial^{c+}\varphi$ contains the support of any optimal plan from μ to ν. It is well known that a Kantorovich potential always exists.

Remark 1.7. In case the two given measures μ, ν have compact support, there exists a $c-$concave Kantorovich potential φ of the form

$$\varphi(x) = \inf_{y \in K} c(x, y) - \psi(y),$$

for some function $\psi : M \to \mathbb{R} \cup \{\pm\infty\}$, where K is a compact set which contains the supports of μ and ν.

In particular, this potential is locally semiconcave. □

It is important to underline that the $c-$superdifferential at a certain point x is made of points on the manifold, and not of tangent vectors. However there is a strict link between the $c-$superdifferential and the usual superdifferential, as the following proposition shows: this link was exploited in the setting of optimal transport by McCann in [25]. The same argument used by McCann was already known to Cabré who used it in an earlier work on elliptic equation on manifolds ([7]).

THEOREM 1.8 (Cabré-McCann). *Let $\varphi : M \to \mathbb{R} \cup \{-\infty\}$ be a c-concave function and $x, y \in M$ such that $y \in \partial^{c+}\varphi(x)$. Then $\exp_x^{-1}(y) \subset -\partial^+\varphi(x)$. Conversely, if φ is differentiable at x and $\nabla\varphi(x) = \mathrm{v}$, then $y := \exp_x(-\mathrm{v})$ is the unique point in $\partial^{c+}\varphi(x)$.*

REMARK 1.9. The converse implication in this theorem is false if one does not assume φ to be differentiable at x: i.e. it is *not* true in general that $\mathrm{v} \in \partial^+\varphi(x)$ implies $\exp_x(-\mathrm{v}) \in \partial^{c+}\varphi(x)$. The question is related to the so called *regularity* of the cost function. A sufficient condition for this regularity is the satisfaction of the Ma-Trudinger-Wang condition (see [22]). We won't stress this point further, the interested reader may look at [33], chapter 12. □

Theorems 1.2, 1.6 and 1.8 allow to understand when the optimal plan is unique and induced by a map and to characterize this map.

THEOREM 1.10 (Brenier-McCann). *Let $\mu, \nu \in \mathcal{P}_c(M)$ and assume that μ is absolutely continuous. Then there exists a unique optimal plan from μ to ν and this plan is induced by the map $\exp(-\nabla\varphi)$, where φ is a Kantorovich potential for μ, ν.*

Proof. By remark 1.7 we know that there exists a Kantorovich potential φ which is semiconcave in some open set Ω containing the supports of both μ and ν. By a classical result of convex analysis, φ is a.e. differentiable in Ω w.r.t. the volume measure. Thus, by the hypothesis on μ, it is also $\mu-$a.e. differentiable. By theorem 1.6 we know that every optimal plan $\boldsymbol{\gamma}$ from μ to ν must be concentrated on $\partial^{c+}\varphi$. By proposition 1.8 and what we said on the differentiability of φ we get that for μ-a.e. x there is only on $y \in M$ such that $(x, y) \in \mathrm{supp}(\boldsymbol{\gamma})$, and that this y is identified by $y = \exp_x(-\nabla\varphi(x))$. Which is the thesis. □

By TM we intend the tangent bundle of M, which will always be endowed with the Sasaki metric d_* (see e.g. [9] Chapter 3 exercise 2). In particular, it makes sense to speak about the metric space $(\mathscr{P}_2(TM), W_2)$, where here W_2 is the quadratic Wasserstein distance built over the distance d_*. We will denote by $\mathscr{P}_2(TM)_\mu \subset \mathscr{P}_2(TM)$, $\mu \in \mathscr{P}_2(M)$, the set of plans $\boldsymbol{\gamma}$ such that $\pi^M_\#\boldsymbol{\gamma} = \mu$, where $\pi^M : TM \to M$ is the natural projection. This is the same as the set of plans $\boldsymbol{\gamma} \in \mathscr{P}(TM)$ satisfying

$$\pi^M_\#\boldsymbol{\gamma} = \mu,$$
$$\int |\mathrm{v}|^2 d\boldsymbol{\gamma}(x, \mathrm{v}) < \infty.$$

The *exponential* $\mathbf{exp}_\mu(\boldsymbol{\gamma})$ of a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu$ is defined as

$$\mathbf{exp}_\mu(\boldsymbol{\gamma}) := (\exp)_\#\boldsymbol{\gamma},$$

it is immediate to verify that $\mathbf{exp}_\mu(\boldsymbol{\gamma}) \in \mathscr{P}_2(M)$, whenever $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)$. See also the appendix of [4] and Chapter 7 of [17] for the properties of the exponential map[1].

[1] We call this map exponential because of theorem 1.11, but it should be noted carefully that exponentiation does not produce geodesics in general.

The right inverse $\mathbf{exp}_\mu^{-1} : \mathscr{P}_2(M) \to \mathscr{P}_2(TM)_\mu$ of the exponential map is defined as

$$\mathbf{exp}_\mu^{-1}(\nu) := \Big\{\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu \ : \ \mathbf{exp}_\mu(\boldsymbol{\gamma}) = \nu, \ \int |\mathrm{v}|^2 d\boldsymbol{\gamma}(x, \mathrm{v}) = W^2(\mu, \nu)\Big\},$$

or, which is the same, $\mathbf{exp}_\mu^{-1}(\nu)$ is the set of those plans $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)$ such that $(\pi^M, \exp)_\# \boldsymbol{\gamma}$ is an optimal plan from μ to ν and $\int |\mathrm{v}|^2 d\boldsymbol{\gamma}(x, \mathrm{v}) = W^2(\mu, \nu)$ (notice that the second condition is not implied by the first one if on M some points have non empty cut locus). The map $\mathbf{exp}_\mu^{-1}$ is not really the inverse of $\mathbf{exp}_\mu$, as only optimal plans are taken in consideration: while this may sound confusing, the choice has been made so that for any $\boldsymbol{\gamma} \in \mathbf{exp}_\mu^{-1}(\mu)$, the curve $t \mapsto \mathbf{exp}_\mu(t \cdot \boldsymbol{\gamma})$ is a geodesic connecting μ to ν, see below.

Observe that plans in $\mathbf{exp}_\mu^{-1}(\nu)$ carry more information about optimal coupling from μ to ν then those in $\mathcal{O}pt(\mu, \nu)$: indeed the latter one only specify from where to where the mass is moved, while the former ones also specify which geodesic is chosen in this movement. The following statement collects the main properties of geodesics in $(\mathscr{P}_2(M), W_2)$ which we will need:

Theorem 1.11 (Geodesics in $\mathscr{P}_2(M)$). *A curve (μ_t) is a constant speed geodesic on $[0, 1]$ from μ to ν if and only if there exists a plan $\boldsymbol{\gamma} \in \mathbf{exp}_\mu^{-1}(\nu)$ such that:*

$$\mu_t = \exp_{\pi^M}(t\pi^1)_\# \boldsymbol{\gamma}, \tag{1.1}$$

π^1 being the map which associates to $(x, \mathrm{v}) \in TM$ the vector $\mathrm{v} \in T_xM$. The plan $\boldsymbol{\gamma}$ is uniquely identified by the geodesic. Moreover, for any $t \in (0, 1)$ there exists a unique optimal plan from μ to μ_t. Finally, two different geodesics from μ to ν cannot intersect at intermediate times.

Introducing the notion of *rescalation* of a plan:

$$\lambda \cdot \boldsymbol{\gamma} := (\pi^M, \lambda\pi^1)_\# \boldsymbol{\gamma}, \qquad \forall \lambda \in \mathbb{R}, \ \boldsymbol{\gamma} \in \mathscr{P}(TM),$$

equation (1.1) takes the more appealing form

$$\mu_t := \mathbf{exp}_\mu(t \cdot \boldsymbol{\gamma}), \qquad \forall t \in [0, 1].$$

From the above theorem we get the following statement about constant speed geodesics starting from μ:

Proposition 1.12. *Let $\mu \in \mathscr{P}_2(M)$ and (μ_t) a constant speed geodesic starting from μ and defined on some right neighborhood of 0, say $[0, a]$. Then there exists a unique plan $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu$ such that*

$$\mu_t = \mathbf{exp}_\mu(t \cdot \boldsymbol{\gamma}), \qquad \forall t \in [0, a].$$

Moreover, two constant speed geodesics (μ_t) and $(\tilde{\mu}_t)$ defined on $[0, a]$ and $[0, \tilde{a}]$ coincide on $[0, \min\{a, \tilde{a}\}]$ if and only if the associated plans are the same.

Proof. Uniqueness follows from the uniqueness part of the theorem above. For the existence, just reparametrize the curve by defining $\tilde{\mu}_t := \mu_{at}$ and observe that $(\tilde{\mu}_t)$ is a constant speed geodesic defined on $[0, 1]$. By the above theorem, we know that there exists a unique plan $\tilde{\boldsymbol{\gamma}} \in \mathscr{P}_2(TM)_\mu$ such that $\tilde{\mu}_t = \mathbf{exp}_\mu(t \cdot \tilde{\boldsymbol{\gamma}})$. The conclusion follows by defining $\boldsymbol{\gamma} := \frac{1}{a} \cdot \tilde{\boldsymbol{\gamma}}$. The rest is obvious. □

In the following we will need to work with couplings of plans in $\mathscr{P}_2(TM)_\mu$, in order to do so, it is better to introduce some notation. By T^2M we intend the set defined as

$$T^2M := \Big\{(x, \mathrm{v}_1, \mathrm{v}_2) \ : \ \mathrm{v}_1, \mathrm{v}_2 \in T_xM\Big\},$$

and we endow this set with the distance d_{*2} defined by

$$\mathrm{d}^2_{*2}\big((x, \mathrm{v}_1, \mathrm{v}_2), (y, \mathrm{w}_1, \mathrm{w}_2)\big) = \mathrm{d}^2_*\big((x, \mathrm{v}_1), (y, \mathrm{w}_1)\big) + \mathrm{d}^2_*\big((x, \mathrm{v}_2), (y, \mathrm{w}_2)\big).$$

The space $(\mathscr{P}_2(T^2M), W_2)$ is then naturally build over (T^2M, d_{*2}). The three natural projections π^M, π^1, π^2 are defined as

$$\begin{aligned}
\pi^M(x, \mathrm{v}_1, \mathrm{v}_2) &= x \in M,\\
\pi^1(x, \mathrm{v}_1, \mathrm{v}_2) &= \mathrm{v}_1 \in T_xM,\\
\pi^2(x, \mathrm{v}_1, \mathrm{v}_2) &= \mathrm{v}_2 \in T_xM.
\end{aligned}$$

A plan $\boldsymbol{\alpha} \in \mathscr{P}_2(T^2M)$ will be called an admissible coupling for $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathscr{P}_2(TM)_\mu$ if:

$$\begin{aligned}
(\pi^M, \pi^1)_\# \boldsymbol{\alpha} &= \boldsymbol{\gamma}_1,\\
(\pi^M, \pi^2)_\# \boldsymbol{\alpha} &= \boldsymbol{\gamma}_2,
\end{aligned}$$

in this case we write $\boldsymbol{\alpha} \in \mathcal{Adm}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$.

The following characterization of compactness is well known, we skip the proof (see, e.g. [4, Remark 5.2.3.]).

Proposition 1.13 (Stability of optimality and compactness). *Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathscr{P}_2(M)$ and $\mathcal{K} \subset \mathscr{P}_2(TM)$ be defined as*

$$\mathcal{B} := \bigcup_{\mu \in \mathcal{A}_1,\, \nu \in \mathcal{A}_2} \mathbf{exp}_\mu^{-1}(\nu).$$

Then $\mathcal{B}$ is a compact subset of $(\mathscr{P}_2(TM), W_2)$ if and only if $\mathcal{A}_1, \mathcal{A}_2$ are compact subsets of $(\mathscr{P}_2(M), W_2)$.

Also, for $\mathcal{B}_1, \mathcal{B}_2 \subset \mathscr{P}_2(TM)$, consider the set $\mathcal{C} \subset \mathscr{P}_2(T^2M)$ defined as

$$\mathcal{C} := \bigcup_{\mu \in \mathscr{P}_2(M)} \ \bigcup_{\substack{\boldsymbol{\gamma}_1 \in \mathcal{B}_1 \cap \mathscr{P}_2(TM)_\mu \\ \boldsymbol{\gamma}_2 \in \mathcal{B}_2 \cap \mathscr{P}_2(TM)_\mu}} \mathcal{Adm}_\mu(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2).$$

Then $\mathcal{C}$ is a compact subset of $(\mathscr{P}_2(T^2M), W_2)$ if and only if $\mathcal{B}_1, \mathcal{B}_2$ are compact subsets of $(\mathscr{P}_2(TM), W_2)$.

2. Sharp hypothesis in Brenier-McCann theorems.

From the proof of theorem 1.10 it is clear that the problem of understanding for which μ we have existence and uniqueness of optimal map is strongly related to the problem of convex analysis 'how it is made the set of non differentiability points of a convex function?'. The answer to the latter question is given by a theorem of Zajícek. To state his result, we need to give the following definition:

Definition 2.1 ($c-c$ hypersurfaces in $\mathbb{R}^d$). *A set $E \subset \mathbb{R}^d$ is a $c-c$ hypersurface if, up to a permutation of the indexes, there exist two convex functions $f, g : \mathbb{R}^{d-1} \to \mathbb{R}$ such that E is the graph of $f - g$, i.e.*

$$E = \Big\{(y,t) \in \mathbb{R}^d \ : \ t = f(y) - g(y)\Big\}.$$

The following theorem is proven in [34]:

Theorem 2.2 (Zajícek). *Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then the set of points where φ is not differentiable is contained in the union of countably many $c-c$ hypersurfaces.*

Conversely, if a set $E \subset \mathbb{R}^d$ can be covered by countably many $c-c$ hypersurfaces, then there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ which is not differentiable at all the points in E.

The interest of this theorem, for our purpose, is that the set of points is completely characterized by covering with $c-c$ hypersurfaces (while other related results concern covering up to $\mathcal{H}^{d-1}$ null sets).

From the theorem of Zajícek, the characterization we were looking for comes immediately, at least for the case $M = \mathbb{R}^d$.

Definition 2.3 (Regular measures on $\mathbb{R}^d$). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. We say that μ is regular if it gives 0 mass to any $c-c$ hypersurface.*

Proposition 2.4 (Sharp hypothesis on Brenier's theorem). *Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then for every $\nu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists only one optimal plan from μ to ν and this plan is induced by a map from μ if and only if μ is regular.*

Proof. We start with the 'if' part. By theorem 1.6 we know that there exists a c-concave function φ such that any optimal plan from μ to ν is concentrated on the graph of the $c-$superdifferential of φ, where here $c(x,y) = \frac{|x-y|^2}{2}$. On $\mathbb{R}^d$ any c-concave function φ satisfies: $x \mapsto \varphi(x) - \frac{|x|^2}{2}$ is concave and the graph of the c-superdifferential of φ is the same as minus-the graph of the superdifferential of $x \mapsto \varphi(x) - \frac{|x|^2}{2}$. Thus by Zajícek theorem and the hypothesis on μ we know that the set of points x such that $\partial^{c+}\varphi(x)$ contains more than one point is μ-negligible. Therefore the disintegration of any optimal plan w.r.t. the projection onto the μ coordinate has to be a delta for μ-a.e. point: this means that the optimal plan is unique and induced by a map.

We turn to the 'only if': we argue by contradiction. Suppose that there is a $c-c$ hypersurface E such that $\mu(E) > 0$. Then, again by Zajícek theorem, there exists a concave function $\varphi : \mathbb{R}^d \to \mathbb{R}$ whose set of points of non differentiability contains E and thus has μ-positive measure. For some ball $B \subset \mathbb{R}^d$, it holds $\mu(B \cap E) > 0$, now pick a smooth cut-off function χ identically 1 on a neighborhood of B, with compact support and values in $[0,1]$. The computation of the distributional Hessian of $\varphi\chi$ shows that for some $a > 0$ the function $\tilde{\varphi}(x) := \varphi(x)\chi(x) - a|x|^2$ is concave. By construction, the set of non differentiability points of $\tilde{\varphi}$ contained in B coincides with the set of non differentiability points of φ contained in B, and thus has μ-positive measure.

Define the two maps $T_1, T_2 : \mathbb{R}^d \to \mathbb{R}^d$ as:

$$\begin{aligned} T_1(x) &:= \text{the element of smallest norm in } \partial^+\tilde{\varphi}(x), \\ T_2(x) &:= \text{the element of biggest norm in } \partial^+\tilde{\varphi}(x). \end{aligned}$$

Since $\tilde{\varphi}$ coincides with $-a|x|^2$ outside a compact set, $T_1(x) = T_2(x) = -2ax$ outside a compact set. In particular, $T_1, T_2 \in L^2_\mu$. Also, $\mu(\{T_1 \neq T_2\}) > 0$. By construction, the plan

$$\frac{(Id, -T_1)_\# \mu + (Id, -T_2)_\# \mu}{2},$$

is c-cyclically monotone and not induced by a map. □

Remark 2.5. The fact that 'μ gives 0 mass to $c-c$ hypersurfaces' is a sufficient assumption to get uniqueness of the optimal plan, and the fact that this plan is induced by a map, was already noticed by Gangbo and McCann in [13]. This was a sharpening of the previous observation, due to McCann [23], that it is sufficient to assume 'μ gives 0 mass to $d-1$ rectifiable surfaces' (while the original version of Brenier's theorem requires the absolute continuity of μ). □

The case of generic manifolds is analogous, the only thing we have to take care of, is that there is no complete analogy between c-concave functions and semiconcave functions.

Definition 2.6 ($c-c$ hypersurfaces in M). *A set $E \subset M$ is a $c-c$ hypersurface if it can be covered by coordinate charts on each of which it can be covered by a countable number of $c-c$ hypersurfaces on $\mathbb{R}^d$.*

Remark 2.7 (Independence on the chart). A set $E \subset \mathbb{R}^d$ can be covered by a countable number of $c-c$ hypersurface if and only if the same is true for $\phi(E)$, where ϕ is a smooth diffeomorphism of $\mathbb{R}^d$.

This can be seen either by a direct application of the definition, or - in a heavier way - by calling into play Zajícek's theorem and observing that:

- the composition of a convex function φ with a smooth diffeomorphism ϕ is locally semiconvex
- the set of points of non differentiability of $\varphi \circ \phi$ is precisely

$$\phi(\{\text{points of non differentiability of } \varphi\})$$

- the conclusions of Zajícek's theorem are unchanged if 'convex' is replaced by 'semiconvex', because adding a multiple of $|x|^2$ does not change the set of non differentiability points
- the conclusions of Zajícek's theorem are unchanged if 'semiconvex' is replaced by 'locally semiconvex', because a set can be covered by a countable union of $c-c$ hypersurfaces if and only if it can be locally covered by a countable union of $c-c$ hypersurfaces.

□

Definition 2.8 (Regular measures on M). *Let $\mu \in \mathscr{P}_2(M)$. We say that μ is regular if it gives 0 mass to any $c-c$ hypersurface.*

We will use the following lemma:

Lemma 2.9. *Let $\varphi : M \to \mathbb{R}$ be a $-\lambda$-concave function (i.e. $\nabla^2 \varphi \leq \lambda Id$ in the sense of distributions) with compact support. Then for $\varepsilon > 0$ sufficiently small the function $\varepsilon\varphi$ is c-concave and it holds $\mathrm{v} \in \partial^+(\varepsilon\varphi)(x)$ if and only if $\exp_x(-\mathrm{v}) \in \partial^{c+}(\varepsilon\varphi)(x)$.*

Proof. We make the following claim: there exists $\varepsilon > 0$ such that for every $x_0 \in M$ and every $\mathrm{v} \in \partial^+\varphi(x_0)$ the function

$$x \mapsto \varepsilon\varphi(x) - \frac{\mathrm{d}^2(x, \exp_{x_0}(-\varepsilon \mathrm{v}))}{2}$$

has a global maximum at $x = x_0$. The claim is obvious if x_0 lies outside the support of φ. Also, since φ is semiconcave, real valued and compactly supported, it is Lipschitz. Thus for $\varepsilon > 0$ sufficiently small we have that for any $x_0 \in M$ and $\mathrm{v} \in \partial^+\varphi(x_0)$, the vector $\varepsilon\mathrm{v}$ has norm smaller than the injectivity radius at x_0. Therefore with this choice of ε we have $\nabla\frac{1}{2}\mathrm{d}^2(\cdot, \exp_{x_0}(-\varepsilon\mathrm{v}))(x_0) = \varepsilon\mathrm{v}$ and hence

$$0 \in \partial^+\left(\varepsilon\varphi(\cdot) - \frac{\mathrm{d}^2(\cdot, \exp_{x_0}(-\varepsilon\mathrm{v}))}{2}\right)(x_0) \tag{2.1}$$

Use the smoothness of M and the compactness of $\mathrm{supp}(\varphi)$ to find $r > 0$ such that

$$\frac{\mathrm{d}^2(\cdot,\cdot)}{2} : \Big\{(x, y) : d(x, y) < r, \mathrm{d}(x, \mathrm{supp}(\varphi)) < r, \mathrm{d}(y, \mathrm{supp}(\varphi)) < r\Big\} \to \mathbb{R}$$

is C^∞ and satisfies $\nabla^2\mathrm{d}^2(\cdot, y)/2 \geq cId$, $c > 0$, for any y s.t. $\mathrm{d}(y, \mathrm{supp}(\varphi)) < r$. Using again the fact that φ is Lipschitz, for $\varepsilon_0 > 0$ sufficiently small it holds $\varepsilon_0|\mathrm{v}| < r/3$ for any $\mathrm{v} \in \partial^+\varphi(x)$ and any $x \in M$. Also, since φ is bounded, up to decreasing the value of ε_0 we can assume that

$$\varepsilon_0|\varphi(x)| \leq \frac{r^2}{12}.$$

Fix $x_0 \in M$, $\mathrm{v} \in \partial^+\varphi(x_0)$ and let $y_0 := \exp_{x_0}(-\varepsilon_0\mathrm{v})$. We claim that for ε_0 chosen as above, the maximum of $\varepsilon_0\varphi - \mathrm{d}^2(\cdot, y_0)/2$, cannot lie outside $B_r(x_0)$. Indeed if $\mathrm{d}(x, x_0) \geq r$ we have $\mathrm{d}(x, y_0) > 2r/3$ and thus:

$$\varepsilon_0\varphi(x) - \frac{\mathrm{d}^2(x, y_0)}{2} < \frac{r^2}{12} - \frac{2r^2}{9} = -\frac{r^2}{12} - \frac{r^2}{18} \leq \varepsilon_0\varphi(x_0) - \frac{\mathrm{d}^2(x_0, y_0)}{2}.$$

Thus the maximum must lie in $B_r(x_0)$. Recall that in this ball, the function $\mathrm{d}^2(\cdot, y_0)$ is C^∞ and satisfies $\nabla^2(\mathrm{d}^2(\cdot, y_0)/2) \geq cId$, thus it holds

$$\nabla^2\left(\varepsilon_0\varphi(\cdot) - \frac{\mathrm{d}^2(\cdot, y_0)}{2}\right) \leq (\varepsilon_0\lambda - c)Id,$$

where $\lambda \in \mathbb{R}$ is such that $\nabla^2\varphi \leq \lambda Id$ on the whole M. Thus decreasing if necessary the value of ε_0 we can assume that

$$\nabla^2\left(\varepsilon_0\varphi(\cdot) - \frac{\mathrm{d}^2(\cdot, y_0)}{2}\right) < 0 \qquad \text{on } B_r(x_0),$$

which implies that $\varepsilon_0\varphi(\cdot) - \mathrm{d}^2(\cdot, y_0)/2$ admits a unique point $x \in B_r(x_0)$ such that $0 \in \partial^+(\varphi - \mathrm{d}^2(\cdot, y_0)/2)(x)$. Since by (2.1) this is true for x_0, it must hold $x = x_0$, which therefore implies that x_0 is the unique global maximum of $\varphi - \mathrm{d}^2(\cdot, y_0)/2$. By the arbitrariness of x_0, v the claim is proved.

To complete the proof, we define the function $\psi : M \to \mathbb{R} \cup \{-\infty\}$ in the following way: if $y = \exp_{x_0}(-\mathrm{v}_0)$ for some $\mathrm{v}_0 \in \partial^+\varphi(x_0)$ we put

$$\psi(y) := \inf_z \frac{\mathrm{d}^2(z, y)}{2} - \varphi(z),$$

otherwise $\psi(y) = -\infty$. By definition we have

$$\varphi(x) \leq \frac{\mathrm{d}^2(x, y)}{2} - \psi(y), \qquad \forall x, y \in M,$$

and the claim proved ensures that if $y = \exp_{x_0}(-\mathrm{v}_0)$ for some $\mathrm{v}_0 \in \partial^+\varphi(x_0)$, the infimum in the definition of $\psi(y)$ is realized at $z = x_0$, that is

$$\varphi(x_0) = \frac{\mathrm{d}^2(x_0, y)}{2} - \psi(y).$$

Since for any $x_0 \in M$ there exists $\mathrm{v} \in \partial^+\varphi(x_0)$, we proved the c-concavity of φ. Following the same lines one easily see that for $y \in \exp_x(-\partial^+\varphi(x))$ it holds

$$\varphi(x) = \frac{\mathrm{d}^2(x, y)}{2} - \varphi^c(y),$$

that is: $\exp_x(\partial^+\varphi(x)) \subset \partial^c\varphi(x)$. Since the other inclusion was proven in Theorem 1.8, the proof is completed. □

Proposition 2.10 (Sharp hypothesis on McCann's theorem). *Let $\mu \in \mathscr{P}_2(M)$. Then for every $\nu \in \mathscr{P}_2(M)$ there exists only one optimal plan from μ to ν and this plan is induced by a map from μ if and only if μ is regular.*

Proof. We start with the 'if' part. We claim that we can assume that μ and ν have compact support. Indeed, let $\boldsymbol{\gamma}$ be an optimal plan from μ to ν and $n \mapsto K_n \subset M$ be an increasing sequence of compact sets. Then $\boldsymbol{\gamma}$ is induced by a map if and only if each of the restrictions $\boldsymbol{\gamma}|_{K_n \times K_n}$ is induced by a map. Also, observe that if each optimal plan is induced by a map, then the optimal plan is unique, as if there were two different optimal maps T, S from μ to ν, the plan $\frac{1}{2}\big((Id, T)_\#\mu + (Id, S)_\#\mu\big)$ would be optimal and not induced by a map.

So we assume $\mu, \nu \in \mathcal{P}_c(M)$. Then by remark 1.7 we can find a Kantorovich potential φ which is semiconcave. Looking φ in charts, we have that it is a locally semiconcave function. Thus by Zajícek theorem we know that the set of points of non differentiability of φ is μ-negligible. The conclusion follows as in the case $M = \mathbb{R}^d$.

Now we turn to the 'only if'. Suppose that $\mu(E) > 0$ for some $c - c$ hypersurface E. Then we can find an open set $\Omega \subset M$ diffeomorphic to $\mathbb{R}^d$ and a compact set $K \subset \Omega$ such that E is a $c - c$ hypersurface in Ω (having identified Ω to $\mathbb{R}^d$), and $\mu(E \cap K) > 0$. By Zajícek theorem, we can find a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ which is not differentiable at any point in $E \cap K \subset \Omega \approx \mathbb{R}^d$. Pick a $C_c^\infty(\Omega)$ cut off function which is identically 1 on K and observe that the function $\chi\varphi : M \to \mathbb{R}$ defined by $\chi\varphi = 0$ outside Ω, is semiconvex. Now apply lemma 2.9 to find $\varepsilon > 0$ such that $\varepsilon\chi\varphi$ is c-concave and $\mathrm{v} = \nabla(\varepsilon\chi\varphi)(x)$ if and only if $\exp_x(-\mathrm{v}) \in \partial^{c+}(\varepsilon\chi\varphi)(x)$. The conclusion follows as for the case $M = \mathbb{R}^d$. □

Remark 2.11. Here we just proved that the optimal plan is unique and induced by a map, but of course one may ask whether such a map can be recovered by taking the gradient of a c-concave function. This is actually the case: we postpone the proof of this fact (which slightly generalizes what is proven in [12]) to the appendix. □

Remark 2.12. I do not know whether, given a geodesic (μ_t) such that μ_0 is regular, it holds 'μ_t is regular for any $t < 1$' or not. □

3. The abstract tangent space. In this section we study from a purely metric perspective the tangent space of $(\mathscr{P}_2(M), W_2)$ at a certain measure μ. We will stick to the case of the Wasserstein space built over a Riemannian manifold, but actually all of what we are going to say here is valid in the setting of metric space with Alexandrov curvature bounded from below (see remark 3.5).

In this section we will assume that the manifold M is compact.

Let $r_{\min} > 0$ be the injectivity radius of M, i.e. the largest r such that $t \mapsto \exp_x(t\mathrm{v})$ is the unique minimizing geodesic between $x \in M$ and $\exp_x(\mathrm{v})$ for any $x \in M$ and any $\mathrm{v} \in T_xM$ with $|\mathrm{v}| \leq r$. The fact that $r_{\min}$ is positive is ensured by the compactness of M.

We do not want to do a general discussion about tangent spaces of metric spaces, we just recall that if a certain geodesic space (X, d) is 'sufficiently well behaved' near a certain point $x_0 \in X$, then we can define the *angle* $\theta(\gamma, \tilde{\gamma}) \in [0, \pi]$ between two constant speed geodesics $(\gamma(t)), (\tilde{\gamma}(t))$ starting from x_0 and defined in some right neighborhood of 0 by:

$$\cos\big(\theta(\gamma,\tilde{\gamma})\big) := \lim_{t,s\downarrow 0} \frac{d^2(\gamma(t),x_0) + d^2(\tilde{\gamma}(s),x_0) - d^2(\gamma(t),\tilde{\gamma}(s))}{2d(\gamma(t),x_0)d(\tilde{\gamma}(s),x_0)},$$

where of course the problem is in proving that the joint limit exists (and typically it does not). Assume that the angle always exists, and let Dir_{x_0} be the set of constant speed geodesics starting from x_0, defined on some right neighborhood of 0, where we identify two of them if they coincide near 0. Then one can define the distance $D(\gamma, \tilde{\gamma})$ between $\gamma, \tilde{\gamma} \in \mathrm{Dir}_{x_0}$ by the formula:

$$D^2(\gamma,\tilde{\gamma}) := \frac{1}{2}\Big(|\gamma'|^2 + |\tilde{\gamma}'|^2 - |\gamma'||\tilde{\gamma}'|\cos\big(\theta(\gamma,\tilde{\gamma})\big)\Big) = \lim_{t\downarrow 0}\frac{d^2(\gamma(t),\tilde{\gamma}(t))}{t^2},$$

where $|\gamma'|, |\tilde{\gamma}'|$ are the metric speed of $\gamma, \tilde{\gamma}$ respectively. The abstract tangent space at x_0 is then defined as the completion of Dir_{x_0} w.r.t. the distance D.

We want to apply this construction to the space $(\mathscr{P}(M), W_2)$, where M is a compact Riemannian manifold. We start with the following concavity estimate. This statement is not new (see for instance Theorem 4.6.1 of [19] for a sharper result), but we preferred to provide the simple proof for completeness.

PROPOSITION 3.1. *Let M be a compact smooth Riemannian manifold. Then there exists a constant $C < +\infty$ such that for any $x \in M$, $\mathrm{v}, \mathrm{w} \in T_xM$ and $T, S > 0$ such that $T|\mathrm{v}|, S|\mathrm{w}| < r_{\min}$ it holds*

$$\begin{aligned} &\mathrm{d}^2\Big(\exp_x(t\mathrm{v}), \exp_x(S\mathrm{w})\Big) \\ &\geq \left(1 - \frac{t}{T}\right) S^2|\mathrm{w}|^2 + \frac{t}{T}\mathrm{d}^2\Big(\exp_x(T\mathrm{v}), \exp_x(S\mathrm{w})\Big) - t(T-t)\Big(|\mathrm{v}|^2 + SC\Big), \end{aligned} \tag{3.1}$$

for any $0 \leq t \leq T$.

Proof. Fix $c > 0$ and let $K \subset T^2M \times \mathbb{R}^2$ be the compact set defined by

$$K := \Big\{(x, \mathrm{v}, \mathrm{w}, T, S) \ : \ |\mathrm{v}|, |\mathrm{w}| < c, \ T, S < \frac{r_{\min}}{c}\Big\},$$

and consider the function $Rem(x,\mathrm{v},\mathrm{w},T,S):K\to\mathbb{R}$ given by

$$(x,\mathrm{v},\mathrm{w},T,S)\ \mapsto\ \mathrm{d}^2\Big(\exp_x(T\mathrm{v}),\exp_x(S\mathrm{w})\Big)-T^2|\mathrm{v}|^2-S^2|\mathrm{w}|^2.$$

By the definition of K, it is obvious that Rem is a C^∞ function. Also, we know that

$$Rem(x,\mathrm{v},\mathrm{w},T,0)=Rem(x,\mathrm{v},\mathrm{w},0,S)=0,$$

therefore it is possible to write

$$Rem(x,\mathrm{v},\mathrm{w},T,S)=TS\,Rem'(x,\mathrm{v},\mathrm{w},T,S),$$

for some C^∞ function $Rem':K\to\mathbb{R}$.

Define the constant $C\in\mathbb{R}$ as

$$C:=\sup_K\left\{\left|\frac{d}{dT}Rem'(x,\mathrm{v},\mathrm{w},T,S)\right|+\frac{1}{2}\left|T\frac{d^2}{dT^2}Rem'(x,\mathrm{v},\mathrm{w},T,S)\right|\right\}<+\infty,$$

and observe that by a simple scaling argument C does not depend on c. Now fix $(x,\mathrm{v},\mathrm{w},T,S)\in K$ and let $f:[0,1]\to\mathbb{R}$ be defined by

$$f(\lambda):=\mathrm{d}^2\Big(\exp_x(\lambda T\mathrm{v}),\exp_x(S\mathrm{w})\Big)=\lambda^2T^2|\mathrm{v}|^2+S^2|\mathrm{w}|^2+\lambda TS\,Rem'(x,\mathrm{v},\mathrm{w},\lambda T,S),$$

and observe that

$$\begin{aligned}\frac{d^2}{d\lambda^2}f(\lambda)&=2T^2|\mathrm{v}|^2+2T^2S\frac{d}{dT}Rem'(x,\mathrm{v},\mathrm{w},T,S)+\lambda T^3S\frac{d^2}{dT^2}Rem'(x,\mathrm{v},\mathrm{w},T,S)\\&\le 2T^2\Big(|\mathrm{v}|^2+SC\Big).\end{aligned}$$

This bound implies the inequality

$$f(\lambda)\ge(1-\lambda)f(0)+\lambda f(1)-\lambda(1-\lambda)T^2\Big(|\mathrm{v}|^2+SC\Big),$$

which gives the conclusion by putting $\lambda=\frac{t}{T}$ ∎

It is known that inequalities like (3.1) are inherited by the quadratic Wasserstein space (see e.g. inequality 7.3.1 [4] and proposition 3.1. of [27]). In our case we have:

Proposition 3.2. *Let $\mu\in\mathscr{P}(M)$ and (μ_t), (ν_s) be two constant speed geodesics starting from μ and defined on some right neighborhood of 0, say $[0,a]$. Then for $T,S<\frac{r_{\min}a}{\mathrm{Diam}(M)}$ it holds*

$$\begin{aligned}W_2^2(\mu_t,\nu_S)&\ge\left(1-\frac{t}{T}\right)W_2^2(\mu,\nu_S)\\&+\frac{t}{T}W_2^2(\mu_T,\nu_S)-t(T-t)\left(\frac{W_2^2(\mu,\mu_T)}{T^2}+SC\right),\qquad\forall t<T\end{aligned}\tag{3.2}$$

where C is the constant provided by proposition 3.1.

Proof. Given that the arguments we are going to use are pretty well known, we will be a bit sloppy in the proof. We know by proposition 1.12 that there exists plans $\boldsymbol{\gamma},\boldsymbol{\eta}\in\mathscr{P}_2(TM)$ such that

$$\begin{aligned}\mu_t&=\exp_{\pi^M}(t\pi^1)_\#\boldsymbol{\gamma},\\\nu_s&=\exp_{\pi^M}(s\pi^1)_\#\boldsymbol{\eta},\end{aligned}$$

for any $t, s < a$. In particular, for $\boldsymbol{\gamma}$-a.e. x, v, the curve $t \mapsto \exp_x(t\mathrm{v})$ is a globally minimizing geodesic in $[0, a]$: therefore $|\mathrm{v}| < \frac{\mathrm{Diam}(M)}{a}$ for $\boldsymbol{\gamma}$-a.e. (x, v). Similarly for $\boldsymbol{\eta}$.

Now fix $T, S < \frac{r_{\min} a}{\mathrm{Diam}(M)}$ and choose $t < T$. Arguing as in the proof of proposition 3.1. of [27], it is possible to show the existence of a plan $\boldsymbol{\alpha} \in \mathscr{P}(T^2M)$ satisfying:

$$(\pi^M, \pi^1)_\# \boldsymbol{\alpha} = \boldsymbol{\gamma},$$
$$\big(\exp_{\pi^M}(t\pi^1), \exp_{\pi^M}(S\pi^2)\big)_\# \boldsymbol{\alpha} \in \mathcal{O}pt(\mu_t, \nu_S)$$

Now pick $(x, \mathrm{v}, \mathrm{w}) \in \mathrm{supp}(\boldsymbol{\alpha})$ and observe that since $T|\mathrm{v}|, S|\mathrm{w}| < r_{\min}$ we may apply proposition 3.1 and get

$$\begin{aligned} &\mathrm{d}^2(\exp_x(t\mathrm{v}), \exp_x(S\mathrm{w})) \\ \geq &\left(1 - \frac{t}{T}\right) S^2|\mathrm{w}|^2 + \frac{t}{T}\mathrm{d}^2(\exp_x(t\mathrm{v}), \exp_x(S\mathrm{w})) - t(T-t)(|\mathrm{v}|^2 + SC). \end{aligned}$$

Integrating this inequality w.r.t. $\boldsymbol{\alpha}$ and observing that it holds

$$(\pi^M, \exp_{\pi^M}(S\pi^2))_\# \boldsymbol{\alpha} \in \mathcal{A}dm(\mu, \nu_S),$$
$$\big(\exp_{\pi^M}(T\pi^1), \exp_{\pi^M}(S\pi^2)\big)_\# \boldsymbol{\alpha} \in \mathcal{A}dm(\mu_T, \nu_S),$$

we get the conclusion. □

Inequality (3.2) is the key tool which allows the proof of existence of the angle between geodesics. We will use the following simple lemma:

Lemma 3.3. *Let F be a real valued function defined on an open set of the kind $(0, a)^2 \subset \mathbb{R}^2$, for some $a > 0$. Assume that F satisfies: for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\left.\begin{matrix} t \leq T \leq \delta \\ s \leq S \leq \delta \end{matrix}\right\} \quad \Rightarrow \quad F(t, s) \leq F(T, S) + \varepsilon. \tag{3.3}$$

Then there exists the joint limit $\lim_{t,s\downarrow 0} F(t, s)$.

Proof. Let $L := \underline{\lim}_{t,s\downarrow 0} F(t, s)$. Fix $\varepsilon > 0$ and find δ so that (3.3) is true. Also, find $T, S < \delta$ such that $F(T, S) \leq L + \varepsilon$. Conclude observing that the inequality

$$\left.\begin{matrix} t \leq T \\ s \leq S \end{matrix}\right\} \quad \Rightarrow \quad F(t, s) \leq F(T, S) + \varepsilon \leq L + 2\varepsilon,$$

implies that $L \geq \overline{\lim}_{t,s\downarrow 0} F(t, s)$. □

Theorem 3.4 (Existence of angles between geodesics). *Let M be a compact Riemannian manifold, $\mu \in \mathscr{P}(M)$ and $(\mu_t), (\nu_s)$ two constant speed geodesics starting from μ and defined on some right neighborhood of 0, say $[0, a]$. Then there exists the joint limit of*

$$F(t, s) := \frac{W_2^2(\mu_t, \mu) + W_2^2(\nu_s, \mu) - W_2^2(\mu_t, \nu_s)}{2W_2(\mu_t, \mu)W_2(\nu_s, \mu)},$$

as $t, s \downarrow 0$.

(for a comparison with the analogous statement appeared in [27], see the remark below)

Proof. Fix $T, S < \delta < \frac{r_{\min} a}{\mathrm{Diam}(M)}$, choose $t < T$, $s < S$ and apply twice inequality (3.2) to get

$$W_2^2(\mu_t,\nu_s) \geq \left(1-\frac{t}{T}\right) W_2^2(\nu_s,\mu) + \frac{t}{T} W_2^2(\mu_T,\nu_s) - \frac{t}{T}\left(1-\frac{t}{T}\right) W_2^2(\mu_T,\mu) - t(T-t)sC,$$
$$W_2^2(\mu_T,\nu_s) \geq \left(1-\frac{s}{S}\right) W_2^2(\mu_T,\mu) + \frac{s}{S} W_2^2(\mu_T,\nu_S) - \frac{s}{S}\left(1-\frac{s}{S}\right) W_2^2(\nu_S,\mu) - s(S-s)TC.$$

Plugging together these inequalities we obtain

$$\begin{aligned} W_2^2(\mu_t,\nu_s) \geq \frac{ts}{TS} W_2^2(\mu_T,\nu_S) &+ \frac{s}{S}\left(\frac{s}{S}-\frac{t}{T}\right) W_2^2(\nu_S,\mu) + \frac{t}{T}\left(\frac{t}{T}-\frac{s}{S}\right) W_2^2(\mu_T,\mu) \\ &- Cts(T-t+S-s). \end{aligned} \tag{3.4}$$

Then by short calculations we deduce:

$$F(t,s) \leq F(T,S) + CTS(T-t+S-s) \leq F(T,S) + 2\delta^3 C.$$

The conclusion follows by lemma 3.3. □

Remark 3.5 (The conclusion holds also in bounded Alexandrov spaces). The key result which allows the proof of the existence of angles between geodesics in $\mathscr{P}(M)$, is inequality (3.1). Such an inequality concerns the behavior of the distance function along an edge of the geodesic triangle $x, \exp_x(T\mathrm{v}), \exp_x(S\mathrm{w})$ and the constants $C, r_{\min}$ which come into play are the only things which depend on the manifold we are working on.

Now suppose we are working on a bounded Alexandrov space (X,d) with curvature bounded below by K. Say that its diameter is D and that, without loss of generality, $K<0$. Consider as reference manifold with constant curvature the Hyperbolic plane with curvature K: in this case $r_{\min} = +\infty$. Now, the plane is not compact, obviously, but since X is bounded, any time we have a geodesic triangle in X, we may consider a reference triangle in some fixed ball of radius $2D$ inside the Hyperbolic plane. The validity of (3.1) and the definition of Alexandrov space ensures that for any $x \in X$, and any two constant speed globally minimizing geodesics $\gamma, \tilde\gamma$ starting from x and defined, say, on [0,1], we have

$$d^2(\gamma(t),\tilde\gamma(S)) \geq \left(1-\frac{t}{T}\right) d^2(x,\tilde\gamma(S)) + \frac{t}{T} d^2(\gamma(T),\tilde\gamma(S)) - t(T-t)\Big(\frac{d^2(x,\gamma(T))}{T^2} + SC\Big),$$

where the constant C depends only on K and D. Once we have this inequality, we can proceed as above and prove first its analogue on the Wasserstein space, and then the existence of angles between geodesics.

This is a generalization of Theorem 3.6. of [27], where it was firstly proven the existence of the limit

$$\lim_{\varepsilon \downarrow 0} \frac{W_2(\mu_{\varepsilon t},\nu_{\varepsilon s})}{\varepsilon} =: f(t,s),$$

and then that the value of

$$\frac{t^2|\dot\mu_t|^2 + s^2|\dot\nu_s| - f^2(t,s)}{2ts},$$

is independent on t, s (here $|\dot{\mu}_t|$ and $|\dot{\nu}_s|$ stand for the metric speed of the geodesics (μ_t), (ν_s) respectively). $\square$

An immediate consequence of theorem 3.4 is the following statement:

PROPOSITION 3.6. *Let $\mu \in \mathscr{P}(M)$ and (μ_t) $(\tilde{\mu}_t)$ be two constant speed geodesics starting from μ and defined on some right neighborhood of 0. Then there exists the limit*

$$D\big((\mu_t), (\tilde{\mu}_t)\big) := \lim_{t \downarrow 0} \frac{W_2(\mu_t, \tilde{\mu}_t)}{t}.$$

and this limit defines a distance on the space of directions Dir_μ (i.e. the set of constant speed geodesics starting from μ, where we identify two geodesics which coincide in some right neighborhood of 0).

Proof. Straightforward. $\square$

Therefore, the following definition is meaningful:

DEFINITION 3.7 (The Abstract Tangent Space). *Let M be a compact Riemannian manifold and $\mu \in \mathscr{P}(M)$. Then the Abstract Tangent Space $\mathrm{AbstrTan}_\mu$ is defined as*

$$\mathrm{AbstrTan}_\mu := \overline{\mathrm{Dir}_\mu}^D,$$

where of course by closure w.r.t. D we intend the abstract completion.

4. Directional derivative of the squared distance. We just proved that there exists an abstract notion of tangent space. Our goal now is to provide a concrete representation of such space. The argument we are going to use is based on the precise calculation of the directional derivative of the squared Wasserstein distance. Such a formula is already known for the Wasserstein space built over an Euclidean space (see [4] proposition 7.3.6.[2]). The generalization to the case of manifolds is pretty straightforward. As in the previous section, we are going to assume that M is compact.

Let us recall that if M is a compact Riemannian manifold, then there exists a constant $C > 0$ such that

$$\mathrm{d}^2(x, \gamma(t)) \geq (1-t)\mathrm{d}^2(x, \gamma(0)) + t\mathrm{d}^2(x, \gamma(1)) - Ct(1-t)\mathrm{d}^2(\gamma(0), \gamma(1)), \tag{4.1}$$

where $\gamma(t)$ is any minimizing constant speed geodesic on $[0, 1]$ (see e.g. Lemma 3.3 of [27] for the proof of this inequality in the setting of Alexandrov spaces).

As already observed by Ohta [27], the Wasserstein space built over a space satisfying an inequality like (4.1) satisfies the same kind of inequality:

PROPOSITION 4.1. *Let M be a compact Riemannian manifold and let C be the optimal constant in* (4.1). *Then the space $(\mathscr{P}_2(M), W_2)$ satisfies*

$$W_2^2(\mu_t, \sigma) \geq (1-t)W_2^2(\mu_0, \sigma) + tW_2^2(\mu_1, \sigma) - Ct(1-t)W_2^2(\mu_0, \mu_1),$$

[2]observe that the formula is one of the key tools used to build a solid analysis of the properties of geodesically convex functionals, thus the generalization of the results of [4] to the case of manifolds has to pass through this formula. In particular, to prove that in a minimizer of $\mu \mapsto F(\mu) + \frac{1}{2\tau}W_2^2(\mu, \mu_0)$ the subdifferential is non empty, a directional derivation is needed

where (μ_t) *is a constant speed minimizing geodesic and* σ *is a generic element of* $\mathscr{P}_2(M)$.

Proof. Same as proposition 3.1. of [27]. $\square$

Recall that given $x, y \in M$ and $\mathrm{v} \in T_xM$, the derivative of $t \mapsto \frac{1}{2}\mathrm{d}^2(\exp_x(t\mathrm{v}), y)$ is given by

$$\frac{d}{dt}|_{t=0}\frac{1}{2}\mathrm{d}^2(\exp_x(t\mathrm{v}), y) = -\sup_{\substack{\mathrm{w}\in\exp_x^{-1}(y),\\ |w|=\mathrm{d}(x,y)}} \langle \mathrm{v}, \mathrm{w}\rangle\,, \tag{4.2}$$

and that the supremum is always achieved.

Theorem 4.2 (Directional derivative of squared distance). *Let* M *be compact,* $\mu, \sigma \in \mathscr{P}(M)$, $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu$ *and define* $\mu_t := \mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma})$. *Then it holds*

$$\frac{d}{dt}|_{t=0^+}\frac{1}{2}W_2^2(\mu_t, \sigma) = -\sup\int \langle \mathrm{v}_1, \mathrm{v}_2\rangle\, d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2),$$

where the supremum is taken among all $\boldsymbol{\alpha} \in \mathscr{P}_2(T^2M)$ *such that*

$$\begin{aligned}&(\pi^M, \pi^1)_\#\boldsymbol{\alpha} = \boldsymbol{\gamma},\\ &(\pi^M, \pi^2)_\#\boldsymbol{\alpha} \in \mathbf{exp}_\mu^{-1}(\sigma).\end{aligned} \tag{4.3}$$

Observe that there are no assumptions on $\boldsymbol{\gamma}$, therefore the curve (μ_t) may not be a geodesic.

Proof. Observe that from proposition 4.1 we know that the map $t \mapsto \frac{1}{2}W_2^2(\mu_t, \sigma)$ is semiconcave, therefore the right derivative at 0 exists.

We start with $\leq$. Choose $\boldsymbol{\alpha}$ satisfying (4.3) and observe that

$$\Big(\exp_{\pi^M}(t\pi^1), \exp_{\pi^M}(\pi^2)\Big)_\#\boldsymbol{\alpha} \in \mathscr{Adm}(\mu_t, \sigma),$$

therefore we have

$$\begin{aligned}&\lim_{t\downarrow 0}\frac{W_2^2(\mu_t, \sigma) - W_2^2(\mu, \sigma)}{t}\\ &\leq \lim_{t\downarrow 0}\frac{1}{t}\left(\int \mathrm{d}^2\big(\exp_x(t\mathrm{v}_1), \exp_x(\mathrm{v}_2)\big) - |\mathrm{v}_2|^2 d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2)\right).\end{aligned} \tag{4.4}$$

Now recalling inequality (4.1) we get

$$\begin{aligned}|\mathrm{v}_2|^2 + \mathrm{d}^2(\exp_x(\mathrm{v}_1), \exp_x(\mathrm{v}_2)) - C(1-t)|\mathrm{v}_1|^2 &\leq \frac{1}{t}\left(\mathrm{d}^2\big(\exp_x(t\mathrm{v}_1), \exp_x(\mathrm{v}_2)\big) - |\mathrm{v}_2|^2\right)\\ &\leq t|\mathrm{v}_1|^2 + 2|\mathrm{v}_2||\mathrm{v}_1|,\end{aligned}$$

from which we get the a uniform domination in t of the integrand in the right hand side of (4.4). Therefore we can pass the limit inside the integral and, from formula (4.2), obtain

$$\begin{aligned}\lim_{t\downarrow 0}\frac{W_2^2(\mu_t, \sigma) - W_2^2(\mu, \sigma)}{t} &\leq \int \frac{d}{dt}|_{t=0^+}\mathrm{d}^2\big(\exp_x(t\mathrm{v}_1), \exp_x(\mathrm{v}_2)\big) d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2)\\ &\leq -2\int \langle \mathrm{v}_1, \mathrm{v}_2\rangle\, d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2).\end{aligned}$$

Now we pass to the opposite inequality. Fix t_0 and let $\boldsymbol{\gamma}_{t_0} \in \mathscr{P}_2(TM)_{\mu_{t_0}}$ be defined by

$$\boldsymbol{\gamma}_{t_0} := \Big(\exp_{\pi^M}(t_0\pi^1), \tau(\pi^1)\Big)_{\#}\boldsymbol{\gamma},$$

where τ is the parallel transport map from x to $\exp_x(t_0\mathrm{v})$ along the geodesic $t \mapsto \exp_x(t\mathrm{v})$. In particular observe that

$$\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}) = \mathbf{exp}_{\mu_{t_0}}\big((t-t_0)\cdot\boldsymbol{\gamma}_{t_0}\big).$$

Now choose $\boldsymbol{\alpha}_{t_0}$ such that

$$\begin{aligned}
&\pi^{0,1}_{\#}\boldsymbol{\alpha}_{t_0} = \boldsymbol{\gamma}_{t_0},\\
&\pi^{0,2}_{\#}\boldsymbol{\alpha}_{t_0} \in \mathbf{exp}^{-1}_{\mu_{t_0}}(\sigma),
\end{aligned}$$

and let $\mathcal{T} : T^2M \to TM$ be the map defined by

$$\begin{aligned}
\frac{d}{dt}|_{t=0}\mathrm{d}^2\big(\exp_x(t\mathrm{v}_1), \exp_x(\mathrm{v}_2)\big) &= -2\left\langle \mathrm{v}_1, T_{t_0}(x,\mathrm{v}_1,\mathrm{v}_2)\right\rangle,\\
\exp_x\big(T_{t_0}(x,\mathrm{v}_1,\mathrm{v}_2)\big) &= \exp_x(\mathrm{v}_2),\\
|T_{t_0}(x,\mathrm{v}_1,\mathrm{v}_2)| &= |\mathrm{v}_2|
\end{aligned}$$

i.e. $\mathcal{T}$ identifies the element of T_xM which realizes the derivative in formula (4.2) with $y := \exp_x(\mathrm{v}_2)$. Define $\tilde{\boldsymbol{\alpha}}_{t_0}$ as

$$\tilde{\boldsymbol{\alpha}}_{t_0} := \Big(\pi^M, \pi^1, \mathcal{T}\Big)_{\#}\boldsymbol{\alpha}_{t_0},$$

and observe that from the definition of $\mathcal{T}$ we have

$$\begin{aligned}
\exp_{\pi^M}(\pi^2)_{\#}\tilde{\boldsymbol{\alpha}}_{t_0} &= \exp_{\pi^M}(\pi^2)_{\#}\boldsymbol{\alpha}_{t_0} = \sigma,\\
\|\mathrm{v}_2\|_{\tilde{\boldsymbol{\alpha}}_{t_0}} &= \|\mathrm{v}_2\|_{\boldsymbol{\alpha}_{t_0}} = W_2(\mu_{t_0},\sigma),
\end{aligned}$$

so that $(\pi^M,\pi^2)_{\#}\tilde{\boldsymbol{\alpha}}_{t_0} \in \mathbf{exp}^{-1}_{\mu_{t_0}}(\sigma)$.

Now argue as in the first part of the proof to get

$$\begin{aligned}
\varlimsup_{t\uparrow t_0}\frac{W_2^2(\mu_t,\sigma) - W_2^2(\mu_{t_0},\sigma)}{t-t_0} &\le -2\int \langle \mathrm{v}_1,\mathrm{v}_2\rangle\, d\tilde{\boldsymbol{\alpha}}_{t_0}(x,\mathrm{v}_1,\mathrm{v}_2)\\
&\le \varliminf_{t\downarrow t_0}\frac{W_2^2(\mu_t,\sigma) - W_2^2(\mu_{t_0},\sigma)}{t-t_0}.
\end{aligned}$$

By the semiconcavity of $t \mapsto W_2^2(\mu_t,\sigma)$ we know

$$\begin{aligned}
\frac{d}{dt}|_{t=0^+}W_2^2(\mu_t,\sigma) &= \lim_{t_0\to 0}\frac{d}{dt}|_{t=t_0^+}W_2^2(\mu_t,\sigma) = \lim_{t_0\to 0}\frac{d}{dt}|_{t=t_0^-}W_2^2(\mu_t,\sigma)\\
&= -2\lim_{t_0\to 0}\int \langle \mathrm{v}_1,\mathrm{v}_2\rangle\, d\tilde{\boldsymbol{\alpha}}_{t_0}(x,\mathrm{v}_1,\mathrm{v}_2).
\end{aligned} \tag{4.5}$$

Observe that $(\pi^M,\pi^1)_{\#}\tilde{\boldsymbol{\alpha}}_{t_0} = \boldsymbol{\gamma}_{t_0} \to \boldsymbol{\gamma}$ in $\mathscr{P}_2(TM)$ and that the family $\{(\pi^M,\pi^2)_{\#}\tilde{\boldsymbol{\alpha}}_{t_0}\}_{t_0\in(0,1]}$ is relatively compact by proposition 1.13. Therefore by proposition 1.13 again we know that the family $\{\tilde{\boldsymbol{\alpha}}_{t_0}\}_{t_0\in(0,1]}$ is relatively compact as well in $\mathscr{P}_2(T^2M)$. Therefore there is a sequence $t^n \downarrow 0$ such that $\tilde{\boldsymbol{\alpha}}_{t^n}$ converges to some $\tilde{\boldsymbol{\alpha}}$ in $\mathscr{P}_2(T^2M)$. By the continuity of (π^M, π^1) and (π^M, π^2) we know that

$$\begin{aligned}&(\pi^M, \pi^1)_\# \tilde{\boldsymbol{\alpha}} = \boldsymbol{\gamma},\\ &(\pi^M, \pi^2)_\# \tilde{\boldsymbol{\alpha}} \in \mathbf{exp}_\mu^{-1}(\sigma).\end{aligned}$$

Since the function $\langle \cdot, \cdot \rangle : T^2M \to \mathbb{R}$ given by $(x, \mathrm{v}_1, \mathrm{v}_2) \to \langle \mathrm{v}_1, \mathrm{v}_2 \rangle$ has quadratic growth, the convergence in $\mathscr{P}_2(T^2M)$ gives

$$\lim_{t^n \to 0} \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\tilde{\boldsymbol{\alpha}}_{t^n}(x, \mathrm{v}_1, \mathrm{v}_2) = \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\tilde{\boldsymbol{\alpha}}(x, \mathrm{v}_1, \mathrm{v}_2),$$

therefore the conclusion follows from equation (4.5). □

5. The Geometric tangent space. In this section we will use the formula for the directional derivative of the squared distance to provide an explicit representation of the tangent space.

We know that the abstract tangent space $\mathrm{AbstrTan}_\mu$ is defined as the completion w.r.t. to the distance D of the set of constant speed geodesics Dir_μ emanating from μ. Now, proposition 1.12 tells that to each (equivalence class of) geodesic $(\mu_t) \in \mathrm{Dir}_\mu$ is canonically associated a unique plan $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu$ via the formula

$$\mu_t = (\exp_{\pi^M}(t\pi^1))_\# \boldsymbol{\gamma}, \qquad \forall t \ll 1. \tag{5.1}$$

The point we want to address here is to understand whether the distance D between geodesics can be read - hopefully in a simple way - in terms of the plans associated.

To understand how this distance between plans should look like, observe that if we have two plans induced by vector fields, then the arguments introduced in [28] suggest that their distance should be the distance between the corresponding vector fields in L^2_μ. A natural way to generalize this distance to the case of general plans is via the following Wasserstein-like definition:

Definition 5.1 (The distance W_μ). *Let $\mu \in \mathscr{P}_2(M)$ and $\boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(TM)_\mu$. Then $W_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ is defined by*

$$W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}) := \int W_2^2(\boldsymbol{\gamma}_x, \boldsymbol{\eta}_x) d\mu(x),$$

where $(\boldsymbol{\gamma}_x)$ and $(\boldsymbol{\eta}_x)$ are the disintegration w.r.t. the projection π^M.

Before studying the relationship between the function W_μ and the geometry of $(\mathscr{P}_2(M), W_2)$, we briefly discuss the main properties of the space $(\mathscr{P}_2(TM)_\mu, W_\mu)$.

Recall that given $\boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(TM)_\mu$ we defined the set $\mathcal{Adm}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ of *admissible* couplings between them as the set of plans $\boldsymbol{\alpha} \in \mathscr{P}(T^2M)$ such that

$$\begin{aligned}&(\pi^M, \pi^1)_\# \boldsymbol{\alpha} = \boldsymbol{\gamma},\\ &(\pi^M, \pi^2)_\# \boldsymbol{\alpha} = \boldsymbol{\eta}.\end{aligned}$$

Now let the *cost* of a plan $\boldsymbol{\alpha} \in \mathscr{P}(T^2M)$ be defined as

$$\int |\mathrm{v}_1 - \mathrm{v}_2|^2 d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2).$$

PROPOSITION 5.2 (Basic properties of W_μ). *Let $\mu \in \mathscr{P}_2(M)$, $\boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(TM)_\mu$. Then*

$$W_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}) = \inf_{\boldsymbol{\alpha} \in \mathcal{Adm}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})} \int |\mathrm{v}_1 - \mathrm{v}_2|^2 d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2). \tag{5.2}$$

Also, the infimum is always achieved. The function $W_\mu : [\mathscr{P}_2(TM)_\mu]^2 \to \mathbb{R}^+$ is a distance and the space $(\mathscr{P}_2(TM)_\mu, W_\mu)$ is complete and separable.

Proof. We start by proving $\leq$ in (5.2). Consider an admissible plan $\boldsymbol{\alpha} \in \mathcal{Adm}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$ and its disintegration $(\boldsymbol{\alpha}_x)$ w.r.t. the projection π^M. It is clear that $\boldsymbol{\alpha}_x \in \mathcal{Adm}(\boldsymbol{\gamma}_x, \boldsymbol{\eta}_x)$ for μ-a.e. x. Therefore it holds

$$\begin{aligned} W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}) = \int W_2^2(\boldsymbol{\gamma}_x, \boldsymbol{\eta}_x) d\mu(x) &\leq \int \int |\mathrm{v}_1 - \mathrm{v}_2|^2 d\boldsymbol{\alpha}_x(\mathrm{v}_1, \mathrm{v}_2) d\mu(x) \\ &= \int |\mathrm{v}_1 - \mathrm{v}_2|^2 d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2). \end{aligned}$$

The opposite inequality follows by a measurable selection argument: basically, choose for μ-a.e. x an optimal plan $\boldsymbol{\alpha}_x \in \mathcal{Opt}(\boldsymbol{\gamma}_x, \boldsymbol{\eta}_x)$ and then define $\boldsymbol{\alpha} \in \mathscr{P}(T^2M)$ by $d\boldsymbol{\alpha} := d\mu(x) \times d\boldsymbol{\alpha}_x$ (we omit the technical details). The same argument shows that the infimum is achieved.

Completeness and separability now follow as in the classical Wasserstein case. □

We will call a plan $\boldsymbol{\alpha}$ which realizes the minimum in (5.2) an *optimal* plan from $\boldsymbol{\gamma}$ to $\boldsymbol{\eta}$, and we write $\boldsymbol{\alpha} \in \mathcal{Opt}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})$. Also, we introduce the following notation

norm of a plan:	$\|\boldsymbol{\gamma}\|_\mu := \sqrt{\int \|\mathrm{v}\|^2 d\boldsymbol{\gamma}(x, \mathrm{v})}, \quad \forall \boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu,$
scalar product of 2 plans:	$\langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu := \frac{1}{2}\Big(\|\boldsymbol{\gamma}\|_\mu^2 + \|\boldsymbol{\eta}\|_\mu^2 - W_\mu^2(\boldsymbol{\gamma}, \boldsymbol{\eta}) \Big) \quad \forall \boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(TM)_\mu,$

Also, recall that we defined the rescaling of a plan as

$$\lambda \cdot \boldsymbol{\gamma} := (\pi^M, \lambda \pi^1)_\# \boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu, \qquad \forall \boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu, \, \lambda \in \mathbb{R}.$$

It is immediate to verify that

$$\begin{aligned} \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu &= \sup_{\boldsymbol{\alpha} \in \mathcal{Adm}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta})} \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2) \\ &= \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2), \qquad \forall \boldsymbol{\alpha} \in \mathcal{Opt}_\mu(\boldsymbol{\gamma}, \boldsymbol{\eta}), \end{aligned}$$

and that

$$\begin{aligned} \|\lambda \cdot \boldsymbol{\gamma}\|_\mu &= |\lambda| \|\boldsymbol{\gamma}\|_\mu, \qquad \forall \boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu, \, \lambda \in \mathbb{R}, \\ \Big\langle \lambda \cdot \boldsymbol{\gamma}, \tilde{\lambda} \cdot \boldsymbol{\eta} \Big\rangle_\mu &= \lambda \tilde{\lambda} \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu, \qquad \forall \boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathscr{P}_2(TM)_\mu, \, \lambda, \tilde{\lambda} > 0. \end{aligned}$$

REMARK 5.3. In general it is *not* true that $\langle \lambda \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu = \lambda \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu$ for negative values of λ. One always has the inequality

$$\langle \lambda \cdot \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu \geq \lambda \langle \boldsymbol{\gamma}, \boldsymbol{\eta} \rangle_\mu, \qquad \lambda < 0$$

and this inequality can very well be strict. See also remark 5.8 and the proof of 6.6. □

With the notation just introduced, the formula for the directional derivative of the squared Wasserstein distance reads as:

$$\begin{aligned}(5.3)\qquad &\frac{d}{dt}\frac{1}{2}W_2^2(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\sigma)\\ &= -\sup_{\boldsymbol{\eta}\in\mathbf{exp}_\mu^{-1}(\sigma)}\langle\boldsymbol{\gamma},\boldsymbol{\eta}\rangle_\mu\,,\qquad \forall\boldsymbol{\gamma}\in\mathscr{P}_2(TM)_\mu,\ \sigma\in\mathscr{P}(M).\end{aligned}$$

Observe the formal analogy with equation (4.2).

Let us now define the set $\mathbf{Dir}_\mu\subset\mathscr{P}_2(TM)_\mu$ as

$$\mathbf{Dir}_\mu := \Big\{\boldsymbol{\gamma}\in\mathscr{P}_2(TM)_\mu\ :\ t\mapsto(\exp_{\pi^M}(t\pi^1))_\#\boldsymbol{\gamma}\\ \text{is a geodesic in some right neighborhood of } 0\Big\},$$

and:

Definition 5.4 (The Geometric tangent space). *Let $\mu\in\mathscr{P}_2(M)$. The Geometric tangent space $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ at μ is defined as the closure of $\mathbf{Dir}_\mu$ w.r.t. the distance W_μ.*

Observe that being $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ a closed subspace of a separable and complete metric space, it is separable and complete as well.

Notice that if $\boldsymbol{\gamma}\in\mathbf{Dir}$, then the norm $\|\boldsymbol{\gamma}\|_\mu$ coincides with the metric speed of the geodesic $t\mapsto\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma})$ defined on some right neighborhood of 0.

We are now ready to prove one of our main results:

Theorem 5.5 (Representation of abstract tangent space). *Let M be a compact Riemannian manifold and $\mu\in\mathscr{P}(M)$. Consider the natural bijection $\mathbf{Dir}_\mu\mapsto\mathrm{Dir}_\mu$ which associate to a plan $\boldsymbol{\gamma}\in\mathbf{Dir}_\mu$ the (equivalence class of the) curve $t\mapsto\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma})$. Then this bijection is an isometry, which therefore extends to a canonical isometry between $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ and* $\mathrm{AbstrTan}_\mu$.

Proof. The fact that the map considered is a bijection follows from proposition 1.12 and the definition of $\mathbf{Dir}_\mu$. Thus all we need to prove is that this map is an isometry. By the definition of distance on Dir_μ and of scalar product on $\mathscr{P}_2(TM)_\mu$, our thesis is equivalent to

$$\lim_{t,s\downarrow0}\frac{t^2\|\boldsymbol{\gamma}\|_\mu^2+s^2\|\boldsymbol{\eta}\|_\mu^2-W_2^2\big(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\mathbf{exp}_\mu(s\cdot\boldsymbol{\eta})\big)}{2ts}=\langle\boldsymbol{\gamma},\boldsymbol{\eta}\rangle_\mu\,,\qquad\forall\boldsymbol{\gamma},\boldsymbol{\eta}\in\mathbf{Dir}_\mu.$$

By theorem 3.4, we know that the joint limit on the left hand side of this equation exists. In particular, its value is unchanged if we first take the limit w.r.t. s and then w.r.t. t. Since for t sufficiently small we have $t^2\|\boldsymbol{\gamma}\|_\mu^2=W_2^2(\mu,\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}))$, we have

$$\begin{aligned}&\lim_{t,s\downarrow0}\frac{t^2\|\boldsymbol{\gamma}\|_\mu^2+s^2\|\boldsymbol{\eta}\|_\mu^2-W_2^2\big(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\mathbf{exp}_\mu(s\cdot\boldsymbol{\eta})\big)}{2ts}\\ &\qquad\qquad= -\lim_{t\downarrow0}\frac{1}{2t}\frac{d}{ds}_{|s=0^+}W_2^2\big(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\mathbf{exp}_\mu(s\cdot\boldsymbol{\eta})\big).\end{aligned}$$

Now we call into play the formula for the directional derivative of the squared Wasserstein distance. Observe that for t sufficiently small theorem 1.11 ensures that the plan $t\cdot\boldsymbol{\gamma}$ is the unique element of $\mathbf{exp}_\mu^{-1}(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}))$, therefore we have

$$-\lim_{t\downarrow 0}\frac{1}{2t}\frac{d}{ds}|_{s=0^+}W_2^2\big(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\mathbf{exp}_\mu(s\cdot\boldsymbol{\eta})\big)=\lim_{t\downarrow 0}\frac{1}{t}\left\langle t\cdot\boldsymbol{\gamma},\boldsymbol{\eta}\right\rangle_\mu=\left\langle\boldsymbol{\gamma},\boldsymbol{\eta}\right\rangle_\mu,$$

and the proof is complete. □

Corollary 5.6. *Let M be a compact Riemannian manifold, $\mu\in\mathscr{P}(M)$ and $\boldsymbol{\gamma},\boldsymbol{\eta}\in\mathbf{Tan}_\mu(\mathscr{P}_2(M))$. Then it holds*

$$W_\mu(\boldsymbol{\gamma},\boldsymbol{\eta})=\lim_{t\downarrow 0}\frac{W_2\big(\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma}),\mathbf{exp}_\mu(t\cdot\boldsymbol{\eta})\big)}{t}.$$

Proof. The previous proof shows that the result is true if $\boldsymbol{\gamma},\boldsymbol{\eta}\in\mathbf{Dir}_\mu$. The conclusion follows by a simple approximation argument, we omit the details. □

Remark 5.7 (On the topology of $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$). Easy examples show that the topology induced by the distance W_μ is *stronger* that the one of $(\mathscr{P}_2(TM),W_2)$. □

Example 5.8 (Weird behavior of tangent plans). Suppose $M=\mathbb{R}$, let $\mu:=\delta_0$ and consider the plan

$$\boldsymbol{\gamma}:=\frac{1}{2}(\delta_{0,1}+\delta_{0,-1})\in\mathscr{P}_2(TM)_\mu.$$

Since $\mu_t:=\mathbf{exp}_\mu(t\cdot\boldsymbol{\gamma})=\frac{1}{2}(\delta_t+\delta_{-t})$ it is immediate to verify that $\boldsymbol{\gamma}\in\mathbf{Dir}_\mu$ (actually, in this situation $\mathscr{P}_2(TM)_\mu$ coincides with $\mathbf{Dir}_\mu$). Along the curve (μ_t) the mass initially in 0 is split: half goes to the left and half to the right. Now suppose we want to move from μ in the 'opposite' direction than the one indicated by $\boldsymbol{\gamma}$. It is easy to be convinced that this means that the mass which was moving to the right now has to move to the left and viceversa, or, which is the same, that we have to consider the plan $-1\cdot\boldsymbol{\gamma}$ and then the curve $t\mapsto\mathbf{exp}_\mu(t\cdot(-1\cdot\boldsymbol{\gamma}))$.

Now, the point is that $-1\cdot\boldsymbol{\gamma}=\boldsymbol{\gamma}$ (!). Therefore in this case 'to move back is the same as to move forward'. In particular, it holds $\left\langle -1\cdot\boldsymbol{\gamma},\boldsymbol{\gamma}\right\rangle_\mu=\left\langle\boldsymbol{\gamma},\boldsymbol{\gamma}\right\rangle_\mu$ (which is a concrete example of strict inequality in remark 5.3). We will see in corollary 6.6 that this kind of behavior in some sense characterizes tangent plans which are not induced by maps. □

6. Relation between $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ and the 'space of gradients'. In the previous section we proved that the tangent space of $\mathscr{P}(M)$ at a measure μ is always given by the space $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$. A natural question which arises is then whether this space is an Hilbert space, and whether this Hilbert space may by identified with the well known 'space of gradients' $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$ defined as

$$\mathrm{Tan}_\mu(\mathscr{P}_2(M)):=\overline{\left\{\nabla\varphi\ :\ \varphi\in C_c^\infty(M)\right\}}^{L^2_\mu}.$$

Observe that we proved that $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ coincides with the abstract notion of tangent space whenever the manifold M is compact. Still, the definition of $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$

makes sense also without such compactness assumption: in this section we drop it, and deal with a generic Riemannian manifold M.

Observe that there is a natural embedding $\iota_\mu : L^2_\mu \to \mathscr{P}_2(TM)_\mu$ given by

$$\iota_\mu(v) := (Id, v)_\# \mu,$$

and this embedding is also an isometry. As usual, we will say that a plan $\boldsymbol{\gamma} \in \mathscr{P}_2(TM)_\mu$ is induced by a map, if $\boldsymbol{\gamma} = \iota_\mu(v)$ for some $v \in L^2_\mu$. A natural right inverse of ι_μ is the barycentric projection defined by:

$$\mathscr{B}(\boldsymbol{\gamma})(x) := \int \mathrm{v} d\boldsymbol{\gamma}_x(\mathrm{v}),$$

where $\{\boldsymbol{\gamma}_x\}_{x\in M}$ is the disintegration of $\boldsymbol{\gamma}$ w.r.t. the projection π^M. The barycentric projection is characterized by the equality

$$\int \langle u(x), \mathrm{v}\rangle \, d\boldsymbol{\gamma}(x,\mathrm{v}) = \int \left\langle u(x), \int \mathrm{v} d\boldsymbol{\gamma}_x(\mathrm{v}) \right\rangle d\pi^M_\# \boldsymbol{\gamma}(x) = \int \langle u, \mathscr{B}(\boldsymbol{\gamma})\rangle \, d\mu, \quad \forall u \in L^2_\mu.$$

Notice that $\mathscr{B} : \mathscr{P}_2(TM)_\mu \to L^2_\mu$ is 1-Lipschitz, indeed for $\boldsymbol{\alpha} \in \mathcal{O}pt_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ it holds

$$\begin{aligned}
\int |\mathscr{B}(\boldsymbol{\gamma}) - \mathscr{B}(\boldsymbol{\gamma}')|^2 d\mu &= \int \left| \int \mathrm{v} d\boldsymbol{\gamma}_x(\mathrm{v}) - \int \mathrm{v} d\boldsymbol{\gamma}'_x(\mathrm{v}) \right|^2 d\mu(x) \\
&= \int \left| \int \mathrm{v} - \mathrm{v}' d\boldsymbol{\alpha}_x(\mathrm{v}, \mathrm{v}') \right|^2 d\mu(x) \\
&\leq \int |\mathrm{v} - \mathrm{v}'|^2 d\boldsymbol{\alpha}(x, \mathrm{v}, \mathrm{v}') = W^2_\mu(\boldsymbol{\gamma}, \boldsymbol{\gamma}')
\end{aligned}$$

The two main results of this section are given in corollaries 6.4 and 6.6. In corollary 6.4 we exploit the relation between $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$ and other natural sets of 'potential tangent' maps, in corollary 6.6 we prove that $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is an Hilbert space if and only if μ is regular, and in this case it coincides, via the embedding ι_μ, with $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$.

We start with the following simple statement:

Proposition 6.1. *Let $\mu \in \mathscr{P}_2(M)$. Then $\iota_\mu(\mathrm{Tan}_\mu(\mathscr{P}_2(M))) \subset \mathbf{Tan}_\mu(\mathscr{P}_2(M))$.*

Proof. By density, it is enough to show that $\iota_\mu(\nabla\varphi) \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$. This is true because of lemma 2.9. □

Now we want to prove that $\mathscr{B}\Big(\mathbf{Tan}_\mu(\mathscr{P}_2(M))\Big) \subset \mathrm{Tan}_\mu(\mathscr{P}_2(M))$. From the technical point of view, this will be the key enabler from which we will get our results. We will use the following lemma, which will end to be a particular case of proposition 6.3, but needs to be proved apart. Observe that we are going to use estimates concerning the regularization by convolution: the proof of these estimates can be found in the appendix.

We will write $\|v\|_\mu$ for the norm of the vector field $v \in L^2_\mu$.

Lemma 6.2. *Let $\mu \in \mathscr{P}_2(M)$ be a measure absolutely continuous w.r.t. the volume measure and $\nu \in \mathscr{P}_2(M)$. Let $v \in L^2_\mu$ be the (unique) vector field such that $\exp(v)$ is the unique optimal transport map from μ to ν provided by McCann theorem. Then $v \in \mathrm{Tan}_\mu(\mathscr{P}_2(M))$.*

Proof. Let $n \mapsto K_n \subset M$ be an increasing sequence of compact sets such that $M = \cup_n K_n$ and, for every $n \in \mathbb{N}$, $\chi_n \in C_c^\infty(M)$ be a cut off function satisfying $0 \le \chi_n \le 1$, $\chi_n|_{K_n} \equiv 1$ and $|\nabla \chi_n(x)| \le 1$ for every $x \in M$.

Let $\varphi : M \to \mathbb{R}$ be a c-concave potential for the couple (μ, ν). Assume for a moment that $K = \mathrm{supp}(\nu)$ is compact. Then from remark 1.7 we know that a choice of φ is given by

$$\varphi(x) := \inf_{y \in K} \frac{\mathrm{d}^2(x,y)}{2} + f(y),$$

for an appropriate f, so that φ is locally semiconcave and hence differentiable μ-a.e.. We know that $v = -\nabla\varphi$ belongs to L^2_μ. The fact that y is taken among the elements of a compact set, implies that $\varphi\chi_n$ is a Lipschitz function and (recalling proposition 8.6) for any family of mollifiers ρ^ε defined as in the appendix and ε sufficiently small, we have

$$\|\nabla(\varphi\chi_n * \rho^\varepsilon) - \nabla(\varphi\chi_n) * \rho^\varepsilon\|_\mu \to 0$$

as ε goes to 0 (actually, we will prove 8.6 only for smooth φ and not for Lipschitz ones - the generalization is straightforward, we omit the details). This, together with the (obvious) fact that $\|\nabla(\varphi\chi_n) * \rho^\varepsilon - \nabla(\varphi\chi_n)\|_\mu \to 0$ as $\varepsilon \to 0$, gives

$$\|\nabla(\varphi\chi_n * \rho^\varepsilon) - \nabla(\varphi\chi_n)\|_\mu \to 0,$$

as $\varepsilon \to 0$. Since $\varphi\chi_n * \rho^\varepsilon \in C_c^\infty(M)$, we have that $v_n = \nabla(\varphi\chi_n) \in \mathrm{Tan}_\mu(\mathscr{P}_2(M))$. Letting n go to $+\infty$ and using the dominated convergence theorem we deduce $v \in \mathrm{Tan}_\mu(\mathscr{P}_2(M))$.

The generalization to the case in which ν has not compact support follows by approximation and a stability of optimality argument. □

Let the normal space $\mathrm{Tan}_\mu^\perp(\mathscr{P}_2(M))$ be the orthogonal complement of $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$ in L^2_μ and $\mathrm{P}_\mu : L^2_\mu \to \mathrm{Tan}_\mu(\mathscr{P}_2(M))$ be the orthogonal projection.

PROPOSITION 6.3. *Let $\mu \in \mathscr{P}_2(M)$. Then $\mathscr{B}(\mathbf{Tan}_\mu(\mathscr{P}_2(M))) = \mathrm{Tan}_\mu(\mathscr{P}_2(M))$.*

Proof. Proposition 6.1 gives the inclusion $\supset$, so we only need to prove the other one.

Fix $\mu \in \mathscr{P}_2(M)$ and $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$. By density and positive 1-homogeneity we may assume that $\boldsymbol{\gamma} \in \mathbf{Dir}_\mu$. We claim that up to a further rescaling, we can also assume that $\boldsymbol{\gamma}$ is the unique element of $\mathbf{exp}_\mu^{-1}(\mathbf{exp}_\mu(\boldsymbol{\gamma}))$. Indeed, assume that $t \mapsto \mu_t := \mathbf{exp}_\mu(t \cdot \boldsymbol{\gamma})$ is a geodesic on $[0,2]$, and assume that there is a geodesic $(\tilde{\mu}_t)$ from μ to μ_1 different from $[0,1] \ni t \mapsto \mu_t$. Then following first $[0,1] \ni t \mapsto \tilde{\mu}_t$ and then $[1,2] \ni t \mapsto \mu_t$ we could build a geodesic from μ to μ_2 which intersect (μ_t) at intermediate times. This is impossible by Theorem 1.11.

Observe that the thesis is equivalent to

$$\int \langle w(x), \mathrm{v}\rangle \, d\boldsymbol{\gamma}(x, \mathrm{v}) = 0, \qquad \forall w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(M)).$$

Thus fix $w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(M))$. Choose a family of mollifiers ρ^ε and define

$$\begin{aligned} \mu^\varepsilon &:= \mu * \rho^\varepsilon, \\ w^\varepsilon &:= \frac{(w\mu) * \rho^\varepsilon}{\mu^\varepsilon}, \end{aligned}$$

where we are identifying the measure μ^ε with its density w.r.t. the volume measure.

Choose $\varphi \in C_c^\infty(M)$ and observe that it holds

$$\begin{aligned}\int \langle w^\varepsilon, \nabla\varphi\rangle\, d\mu^\varepsilon &= \int \langle (w\mu) * \rho^\varepsilon(x), \nabla\varphi(x)\rangle\, d\mathrm{vol}(x)\\ &= \int \langle w, (\nabla\varphi) * \rho^\varepsilon\rangle\, d\mu + \mathrm{Rem}\\ &= \int \langle w, \nabla(\varphi * \rho^\varepsilon)\rangle\, d\mu + \mathrm{Rem} + \mathrm{Rem}'\\ &= \mathrm{Rem} + \mathrm{Rem}'\end{aligned}$$

since $w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(M))$. The reminder terms Rem and Rem' can be bounded by using propositions 8.5 and 8.6 to obtain:

$$\left|\mathrm{Rem} + \mathrm{Rem}'\right| \le c_\varepsilon \|w\|_\mu \|\nabla\varphi\|_{\mu^\varepsilon},$$

where c_ε goes to 0 with ε. Therefore we proved

$$\|\mathrm{P}_{\mu^\varepsilon}(w^\varepsilon)\|_{\mu^\varepsilon} \le c_\varepsilon \|w\|_\mu. \tag{6.1}$$

Now let $v^\varepsilon \in L^2_{\mu^\varepsilon}$ be the unique vector field such that $\|v^\varepsilon\|_{\mu^\varepsilon} = W_2(\mu^\varepsilon, \nu)$ and the optimal transport map from μ^ε to ν is given by $\exp(v^\varepsilon)$. In other words, the plans $\boldsymbol{\gamma}^\varepsilon := \iota_{\mu^\varepsilon}(v^\varepsilon)$ are the unique elements of $\mathbf{exp}_{\mu^\varepsilon}^{-1}(\nu)$. By the stability of optimality, the uniqueness assumption on $\boldsymbol{\gamma}$ and the uniform bound on $\int |\mathrm{v}|^2 d\boldsymbol{\gamma}^\varepsilon(x, \mathrm{v}) = W_2^2(\mu^\varepsilon, \nu)$ it is immediate to verify that the following passage to the limit holds:

$$\lim_{\varepsilon\to 0} \int \langle \xi(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}^\varepsilon(x, \mathrm{v}) = \int \langle \xi(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}(x, \mathrm{v}), \qquad \forall \xi \in \mathcal{V}(M).$$

Also, it is easy to check that from the validity of the such limit and the convergence of the w^ε to w (we skip the details), that

$$\lim_{\varepsilon\to 0} \int \langle w^\varepsilon(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}^\varepsilon(x, \mathrm{v}) = \int \langle w(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}(x, \mathrm{v}).$$

Using the previous proposition, we know that $v^\varepsilon \in \mathrm{Tan}_{\mu^\varepsilon}(\mathscr{P}_2(M))$, therefore from (6.1) we obtain

$$\left|\int \langle w^\varepsilon(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}^\varepsilon(x, \mathrm{v})\right| = \left|\langle w^\varepsilon, v^\varepsilon\rangle_{\mu^\varepsilon}\right| \le c_\varepsilon \|w\|_\mu \|v^\varepsilon\|_{\mu^\varepsilon} = c_\varepsilon \|w\|_\mu W_2(\mu^\varepsilon, \nu).$$

Passing to the limit in ε we obtain

$$\int \langle w(x), \mathrm{v}\rangle\, d\boldsymbol{\gamma}(x, \mathrm{v}) = 0.$$

By the arbitrariness of $w \in \mathrm{Tan}_\mu^\perp(\mathscr{P}_2(M))$ we got the thesis. □

Corollary 6.4 (Characterization of tangent maps). *Let $\mu \in \mathscr{P}_2(M)$. The following three sets are equal:*

Tangent vector fields:	$\mathrm{Tan}_\mu(\mathscr{P}_2(M))$,
Closure of vector fields which are optimal in a right neighborhood of 0:	$\overline{\left\{v \in L^2_\mu \;:\; \exists \varepsilon > 0 \textit{ s.t. } (Id, \exp(tv))_\# \mu \textit{ is optimal for } t \le \varepsilon\right\}}^{L^2_\mu}$,
Vector fields which induce tangent plans:	$\left\{v \in L^2_\mu \;:\; \iota_\mu(v) \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))\right\}$,

Proof. By lemma 2.9 we know that the first set is included in the second, while by definition of $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ as closure of Dir_μ we know that the second is included in the third. To conclude, pick $v \in L^2_\mu$ such that $\iota_\mu(v) \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$ and observe that by proposition 6.3 above we have $\mathscr{B}(\iota_\mu(v)) \in \mathrm{Tan}_\mu(\mathscr{P}_2(M))$. Now just observe that $\mathscr{B}(\iota_\mu(v)) = v$. □

Remark 6.5. Observe that a priori the third of the spaces above could be strictly bigger than the second one, as it may be the case that a certain plan in $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ induced by a map cannot be approximated by plans in $\mathbf{Dir}_\mu$ induced by maps.

Also, observe that a priori both the second and the third of the spaces above could be just cones, rather than vector spaces.

What the corollary says, is that these kind of complications do not occur regardless of any assumption on μ or on the manifold. □

Corollary 6.6 (The tangent space is an Hilbert space if and only if μ is regular). *The tangent space $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is an Hilbert space if and only if μ is regular. In this case $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is canonically identified to $\mathrm{Tan}_\mu(\mathscr{P}_2(M))$ via the map ι_μ.*

Proof. We start with *if.* Assume that μ is regular. Then, since all the optimal plans are induced by maps, the space $\mathbf{Dir}_\mu$ is canonically identified, via ι_μ, to the set

$$\Big\{v \in L^2_\mu \ : \ \exists \varepsilon > 0 \ s.t. \ (Id, \exp(\varepsilon v))_\# \mu \text{ is optimal for } t \le \varepsilon\Big\}.$$

Since $W^2_\mu(\iota_\mu(v), \iota_\mu(w)) = \int |v-w|^2 d\mu$, the closure of $\mathbf{Dir}_\mu$ w.r.t. W_μ is identified to the closure of the space above w.r.t. the distance L^2_μ. By the corollary above, we get the claim.

Now we turn to the *only if.* Assume that $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is an Hilbert space and choose $\boldsymbol{\gamma} \in \mathbf{Dir}_\mu \subset \mathbf{Tan}_\mu(\mathscr{P}_2(M))$. Since $\mathbf{Tan}_\mu(\mathscr{P}_2(M))$ is an Hilbert space, it must hold $-1 \cdot \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$ and

$$\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu = - \langle \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu .$$

Define $\boldsymbol{\alpha} := (\pi^M, \pi^1, \pi^1)_\# \boldsymbol{\gamma} \in \mathscr{P}_2(T^2M)$ and $\tilde{\boldsymbol{\alpha}} := (\pi^M, -\pi^1, \pi^1)_\# \boldsymbol{\alpha}$. It is obvious that $\boldsymbol{\alpha}$ is the unique element in $\mathcal{O}pt(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ and that $\tilde{\boldsymbol{\alpha}} \in \mathcal{A}dm(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma})$. Since

$$\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu \ge \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\tilde{\boldsymbol{\alpha}}(x, \mathrm{v}_1, \mathrm{v}_2) = - \int \langle \mathrm{v}_1, \mathrm{v}_2 \rangle \, d\boldsymbol{\alpha}(x, \mathrm{v}_1, \mathrm{v}_2) = - \langle \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu ,$$

it holds $\langle -1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu = - \langle \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_\mu$ if and only if $\tilde{\boldsymbol{\alpha}} \in \mathcal{O}pt(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma})$. By proposition 5.2 and its proof, we know that $\tilde{\boldsymbol{\alpha}} \in \mathcal{O}pt(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma})$ if and only if $\tilde{\boldsymbol{\alpha}}_x \in \mathcal{O}pt((-1 \cdot \boldsymbol{\gamma})_x, \boldsymbol{\gamma}_x)$ for μ-a.e. x, where as usual the subscript x stands for the disintegration w.r.t. the projection onto M. For μ-a.e. x, the plan $\tilde{\boldsymbol{\alpha}}_x$ is induced by the map $\mathrm{v} \mapsto -\mathrm{v}$; it is clear that such a map is cyclically monotone if and only if it is defined on only 1 point. This means that $\tilde{\boldsymbol{\alpha}} \in \mathcal{O}pt(-1 \cdot \boldsymbol{\gamma}, \boldsymbol{\gamma})$ if and only if $\boldsymbol{\gamma}$ is induced by a map. By the arbitrariety of $\boldsymbol{\gamma}$ in $\mathbf{Dir}_\mu$, we deduce that all the plans in $\mathbf{Dir}_\mu$ are induced by a map. By proposition 2.10 this means that μ is regular. The result follows. □

Remark 6.7. In the 1-dimensional case, the set of regular measures coincides with the set of measures having no atoms. Its complement has 0 mass w.r.t. the entropic measure $\mathbb{P}^\beta$ built by vonRenesse-Sturm for any $\beta > 0$ in [30]. This tells that the natural measures in $\mathscr{P}_2(\mathbb{R})$ are concentrated on 'nice' measures, where 'nice' is intended w.r.t. the Riemannian point of view. From a purely formal point of view, this fact has some analogies with the well known statement 'the set of points in an Alexandrov space with curvature bonded from below whose tangent space is not an Euclidean space has 0 volume measure'. From this perspective, it would be interesting to know whether non regular measures have 0 mass also w.r.t. the Gibbs-like measures built by Sturm in [31] in dimension bigger that 1. □

REMARK 6.8. It can be proved that for any $\mu \in \mathscr{P}_2(M)$ and any $\boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$, it holds $-1 \cdot \boldsymbol{\gamma} \in \mathbf{Tan}_\mu(\mathscr{P}_2(M))$. Also, there is natural notion of (multivaled) sum of plans in $\mathscr{P}_2(TM)_\mu$. This is why we preferred to keep the wording'Geometric Tangent space', rather than 'Geometric Tangent cone'. For details on this for the case $M = \mathbb{R}^d$ see [15], the generalization to the case of manifolds presents no difficulties. □

7. Appendix A - On the regularity of the Kantorovich potential. We proved in theorem 2.10 that if a measure $\mu \in \mathscr{P}_2(M)$ gives 0 mass to $c{-}c$ hypersurfaces on M, then for every other measure $\nu \in \mathscr{P}_2(M)$ there exists and is unique the optimal transport map from μ to ν. A natural question is then whether this map can be recovered by exponentiation of (minus) the gradient of a $c-$concave Kantorovich potential.

In the recent paper [12] it is proven the following result:

THEOREM 7.1. *Let φ be a Kantorovich potential for some couple of measures $\mu, \nu \in \mathscr{P}_2(M)$, set $D := \{\varphi > -\infty\}$ and let Ω be the interior of D. Then φ is locally semiconcave in Ω, $\partial^{c+}\varphi(x)$ is non-empty for any x in Ω, and $\partial^{c+}\varphi$ is locally bounded in Ω. Moreover, $D \setminus \Omega$ is $(n-1)$-rectifiable (n being the dimension of M).*

Since we know that μ is concentrated on D, this result almost answer the question, in the sense that if μ does not charge $n-1$ rectifiable sets, then the optimal map can be recovered as $\exp(-\nabla\varphi)$, as discussed in [12].

However, we saw that the condition for the map to exist is that μ gives 0 mass only to $c-c$ hypersurfaces, and not necessarily to all $n-1$ rectifiable sets. Therefore it would be better to know that, in the notation of the theorem above, the set $D \setminus \Omega$ is a $c-c$ hypersurface. This is the case. Actually, it can be proved that $D \setminus \Omega$ can be locally covered by graphs of semiconvex functions (and it is trivial that such graphs are $c-c$ hypersurfaces):

PROPOSITION 7.2. *With the same notation of the above theorem, the set $D \setminus \Omega$ can be covered by charts on each of which it can be covered by a countable number of graphs of semiconvex functions.*

Proof. We will prove that for any $\overline{x} \in D \setminus \Omega$ and any $r > 0$ there exists an open ball B of radius r disjoint from $D \setminus \Omega$ such that $\overline{x} \in \overline{B}$. This, by standard rectifiability results, will imply the thesis.

Fix $\overline{x} \in D \setminus \Omega$, $r > 0$ and find a sequence $(x_n) \subset M \setminus D$ converging to $\overline{x}$. Recall that since φ is $c-$concave, it can be written as

$$\varphi(x) = \inf_{y \in M} \frac{\mathrm{d}^2(x,y)}{2} - \psi(y),$$

for a suitable $\psi : M \to \mathbb{R} \cup \{-\infty\}$. We know by assumption that $\varphi(x_n) = -\infty$, thus for any $n \in \mathbb{N}$ we can find $y_n \in M$ such that

$$\frac{\mathrm{d}^2(x_n, y_n)}{2} - \psi(y_n) \leq -n, \qquad \forall n \in \mathbb{N}.$$

This implies $\psi(y_n) \to \infty$. Also, since

$$-\infty < \varphi(\overline{x}) \leq \frac{\mathrm{d}^2(\overline{x}, y_n)}{2} - \psi(y_n),$$

we get that $\mathrm{d}^2(x, y_n) \to +\infty$ as well. Now for any $n \in \mathbb{N}$ choose a geodesic parametrized by arc length connecting x_n to y_n, call it γ_n. Define the open ball $B_n := B_r(\gamma_n(r))$ (since $\mathrm{d}(x_n, y_n) \to \infty$, $\gamma_n(r)$ is eventually well defined - and so is B_n).

We claim that $\varphi(x) \leq -n$ for any $x \in B_n$. Indeed, for such x it holds

$$\mathrm{d}(x, y_n) \leq \mathrm{d}(x, \gamma_n(r)) + \mathrm{d}(\gamma_n(r), y_n) \leq r + \mathrm{d}(x_n, y_n) - r = \mathrm{d}(x_n, y_n),$$

and therefore

$$\varphi(x) \leq \frac{\mathrm{d}^2(x, y_n)}{2} - \phi(y_n) \leq \frac{\mathrm{d}^2(x_n, y_n)}{2} - \phi(y_n) \leq -n.$$

By compactness, some subsequence of $(\gamma_n(r))$ converges to some $z \in M$ satisfying $\mathrm{d}(\overline{x}, z) = r$. Since it holds

$$B_r(z) \subset \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} B_r\big(\gamma_k(r)\big),$$

we have $\varphi(x) = -\infty$ for every $x \in B_r(z)$. This shows that $B_r(z)$ is disjoint from D. A fortiori it is disjoint from $D \setminus \Omega$. □

REMARK 7.3. This proposition comes from a collaboration with A. Figalli. Actually, its proof is just a variant of the proof of the step 1 of the main theorem in [12]. Unfortunately, we realized that the result in [12] could have been improved in the way presented above, only when it was too late to modify our paper. We therefore agreed to include this new approach in this paper. □

Adding everything up, we proved the following statement:

THEOREM 7.4. *Let $\mu, \nu \in \mathscr{P}_2(M)$ and assume that μ is regular. Then there exists a unique optimal plan, this plan is induced by a map, and this map may be written as* $\exp(-\nabla\varphi)$ *for any c-concave Kantorovich potential for the couple* (μ, ν).

Proof. Straightforward. □

8. Appendix B - reminders on the convolution on manifolds. Here we recall some basic notions regarding how it is possible to define a convolution operator for functions and vector fields on manifolds.

Throughout this section, $\rho^\varepsilon : \mathbb{R}^{\dim M} \to [0, +\infty)$ will be a family of radial functions which are smooth, supported in $B_\varepsilon(0)$ and satisfying $\int \rho = 1$. Also, we will denote by $D(x, y) : TM \to \mathbb{R}$ the Jacobian determinant of $\exp_x^{-1}(y)$. It is clear that for every compact set $K \subset M$ there exists a constants $C(K), \tilde{C}(K) > 0$ such that it holds:

- for any $\varepsilon < C(K)$, the function $D(x,y)$ is well defined, smooth and satisfies $\frac{1}{2} \leq D(x,y) \leq 2$ in the set $\{(x,y) : x \in K, \mathrm{d}(x,y) < \varepsilon\}$,
- for any $\varepsilon < C(K)$ and for every x such that $\mathrm{d}(x,K) < \varepsilon$ the radius of injectivity of the exponential map at x is bigger than ε
- for any $\varepsilon < C(K)$, $x \in K$ and y such that $\mathrm{d}(x,y) < \varepsilon$ it holds: $|D(x,y) - D(y,x)| < \varepsilon \tilde{C}(K) D(x,y)$.

Given a compact set $K \subset M$, we will denote by $K^\varepsilon \subset M$ the compact set of those x such that $\mathrm{d}(x,K) \leq \varepsilon$.

Definition 8.1 (Convolution of functions). *Let $f : M \to \mathbb{R}$ be an integrable function with compact support and $\varepsilon < C(\mathrm{supp}(f))$. The convolution $f * \rho : M \to \mathbb{R}$ is the function defined by:*

$$(f * \rho^\varepsilon)(x) = \int_{T_x M} f(\exp_x(\mathrm{v}))\rho^\varepsilon(\mathrm{v})d\mathrm{v} = \int_M f(y)\rho^\varepsilon\big(\exp_x^{-1}(y)\big)D(x,y)d\mathrm{vol}(y), \tag{8.1}$$

Note that due to the fact that ρ^ε is radially symmetric, there is no ambiguity in the identification of $T_x M$ with $\mathbb{R}^{\dim M}$ (which is the domain of ρ^ε) in the formula above. It is immediate to verify that $f * \rho^\varepsilon \in C_c^\infty(M)$.

The convolution of a measure is defined analogously:

Definition 8.2. *Let $\mu \in \mathcal{P}_c(M)$, $\varepsilon < C(\mathrm{supp}(\mu))$ and $f : M \to \mathbb{R}$ such that $\|f\|_\mu < \infty$. Then $(f\mu) * \rho^\varepsilon \in C_c^\infty(M)$is defined as*

$$(f\mu) * \rho^\varepsilon(x) := \int_M f(y)\rho\big(\exp_x^{-1}(y)\big)D(x,y)d\mu(y).$$

The convolution of vector fields and vector valued measures is defined analogously.

Definition 8.3 (Convolution of vector fields). *Let v be a vector field on M with compact support and with integrable norm and $\varepsilon < C(\mathrm{supp}(v))$. The convolution $v * \rho^\varepsilon \in \mathcal{V}(M)$ is the vector field defined as*

$$(v * \rho^\varepsilon)(x) := \int T_y^x(v(y))\rho^\varepsilon\big(\exp_x^{-1}(y)\big)D(x,y)d\mathrm{vol}(y),$$

where $T_y^x : T_y M \to T_x M$ is the parallel transport map along the unique geodesic connecting y to x.

Similarly, if $\mu \in \mathcal{P}_c(M)$, $v \in L^2_\mu$ and $\varepsilon < C(\mathrm{supp}(\mu))$, $(v\mu) * \rho^\varepsilon \in \mathcal{V}(M)$ is the vector field on M defined by

$$\big((v\mu) * \rho^\varepsilon\big)(x) := \int T_y^x(v(y))\rho^\varepsilon\big(\exp_x^{-1}(y)\big)D(x,y)d\mu(y).$$

An important inequality is given in the following proposition, which is the analogous of the similar result which hold in the case $M = \mathbb{R}^d$ and was proved in [4] (lemma [8.1.10]).

Proposition 8.4. *Let $\mu \in \mathcal{P}_c(M)$, $v \in L^2_\mu$ and $\varepsilon < C(\mathrm{supp}(\mu))$ sufficiently small. Then it holds*

$$\int \left| \frac{(v\mu) * \rho^\varepsilon(x)}{\mu * \rho^\varepsilon(x)} \right|^2 \mu * \rho^\varepsilon(x)dx \leq \int |v(x)|^2 d\mu(x).$$

Similarly for functions.

Proof. The proof of lemma[8.1.10] in [4] never uses the fact that the underlying space is $\mathbb{R}^d$ rather than a generic Riemannian manifold, thus the conclusion follows by the same arguments used there. We omit the details. $\square$

The convolution on a manifold has the same smoothening and convergence properties that are valid on $\mathbb{R}^d$: this means that $f * \rho^\varepsilon \in C_c^\infty(M)$ as soon as ρ^ε is C^∞, and that $\varphi * \rho^\varepsilon \to \varphi$ in L^2_μ for any $\varphi \in C_c^\infty(M)$ and any $\mu \in \mathscr{P}_2(M)$. Similarly for vector fields.

What one needs to care about, is the lack of commutativity in many operations that are usually done with convolutions. This non commutativity may be estimated in terms of bounds on the curvature of M, as we are going to show in the following propositions.

Proposition 8.5. *Let $\mu \in \mathcal{P}_c(M)$, $v \in L^2_\mu$, $\xi \in \mathcal{V}(M)$ and $\varepsilon < C(\operatorname{supp}(\mu))$. Then it holds*

$$\left| \int \langle (v\mu) * \rho^\varepsilon, \xi \rangle \, d\mathrm{vol} - \int \langle v, \xi * \rho^\varepsilon \rangle \, d\mu \right| \le \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \|v\|_\mu \|\xi\|_{\mu * \rho^\varepsilon}.$$

Proof. By definition we have

$$\int \langle (v\mu) * \rho^\varepsilon(x), \xi \rangle \, d\mathrm{vol}(x) = \iint \left\langle T_y^x(v(y)), \xi(x) \right\rangle \rho^\varepsilon\big(\exp_x^{-1}(y)\big) D(x,y) d\mathrm{vol}(x) d\mu(y),$$
$$\int \langle v(y), \xi * \rho^\varepsilon(y) \rangle \, d\mu(y) = \iint \langle v(y), T_x^y(\xi(x)) \rangle \, \rho^\varepsilon\big(\exp_y^{-1}(x)\big) D(y,x) d\mathrm{vol}(x) d\mu(y).$$

Now observe that $|\exp_x^{-1}(y)| = |\exp_y^{-1}(x)|$, and thus $\rho^\varepsilon\big(\exp_x^{-1}(y)\big) = \rho^\varepsilon\big(\exp_y^{-1}(x)\big)$, and recall that in the domain of our integrals it holds $|D(y,x) - D(x,y)| \le \varepsilon \tilde{C}(\operatorname{supp}(\mu)) D(x,y)$ to get

$$\begin{aligned}
&\left| \int \langle (v\mu) * \rho^\varepsilon, \xi \rangle \, d\mathrm{vol} - \int \langle v, \xi * \rho^\varepsilon \rangle \, d\mu \right| \\
&\qquad \le \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \iint |v(y)||\xi(x)| \rho^\varepsilon\big(\exp_x^{-1}(y)\big) D(x,y) d\mathrm{vol}(x) d\mu(y) \\
&\qquad = \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \int (|v|\mu) * \rho^\varepsilon(x) |\xi(x)| d\mathrm{vol}(x) \\
&\qquad = \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \int \frac{(|v|\mu) * \rho^\varepsilon(x)}{\mu * \rho^\varepsilon(x)} |\xi(x)| \mu * \rho^\varepsilon(x) d\mathrm{vol}(x) \\
&\qquad \le \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \left\| \frac{(|v|\mu) * \rho^\varepsilon(x)}{\mu * \rho^\varepsilon(x)} \right\|_{\mu * \rho^\varepsilon} \|\xi\|_{\mu * \rho^\varepsilon} \\
&\qquad \le \varepsilon \tilde{C}(\operatorname{supp}(\mu)) \|v\|_\mu \|\xi\|_{\mu * \rho^\varepsilon}.
\end{aligned}$$

$\square$

Proposition 8.6. *Let $\varphi \in C_c^\infty(M)$, $\mu \in \mathcal{P}_c(M)$ and $\varepsilon < C(\operatorname{supp}(\mu))$. Then it holds:*

$$\|\nabla(\varphi * \rho^\varepsilon) - (\nabla\varphi) * \rho^\varepsilon\|_\mu \le C_\varepsilon \|\nabla\varphi\|_{\mu * \rho^\varepsilon},$$

where C_ε is given by

$$C_\varepsilon := \Big(1+\varepsilon\tilde{C}\big(\mathrm{supp}(\mu)^\varepsilon\big)\Big)\bigg(\cosh\Big(\varepsilon\sqrt{\mathcal{C}\big(\mathrm{supp}(\mu)^{2\varepsilon}\big)}\Big)-1\bigg),$$

and in particular goes to 0 with ε (here $(\mathrm{supp}(\mu))^\varepsilon$ is the ε-neighborhood of $\mathrm{supp}(\mu)$).

Proof. A direct computation of the derivative of $\varphi * \rho^\varepsilon$ along the direction $u \in T_xM$ gives:

$$\nabla(\varphi*\rho^\varepsilon)(x)\cdot u = \int \big\langle \nabla\varphi\big(\exp_x(v)\big), J_v(u)\big\rangle\, \rho^\varepsilon(v)dv, \tag{8.2}$$

where $J_v(u) \in T_{\exp_x(v)}$ is the value at $t = 1$ of the Jacobi field j_t defined along the geodesic $t \mapsto \exp_x(tv)$ with initial conditions $j_0 = u$ and $j_0' = 0$.

We want to bound the distance between $J_v(u)$ and $T_x^y(u)$. To this aim, let $tr_t \in T_{\exp_x(tv)M}$ be the parallel transport of u along the geodesic $t \mapsto \exp_x(tv)$ and observe that from $\frac{d}{dt}|j_t - tr_t|^2 = 2\langle j_t', j_t - tr_t\rangle$ we get

$$\frac{d}{dt}|j_t - tr_t| \le |j_t'|,$$

thus we have

$$\begin{aligned}
|j_t'| \le \int_0^t |j_s''|ds = \int_0^t |R(v,j_s)v|ds &\le \mathcal{C}\big(\mathrm{supp}(\mu)^{2\varepsilon}\big)|v|^2\int_0^t |j_s|ds\\
&\le \sqrt{\mathcal{C}\big(\mathrm{supp}(\mu)^{2\varepsilon}\big)}|v||u|\sinh\Big(\sqrt{\mathcal{C}\big(\mathrm{supp}(\mu)^{2\varepsilon}\big)}|v|t\Big),
\end{aligned}$$

and therefore

$$|J_v(u) - T_x^y(u)| = |j_1 - tr_1| \le \int_0^1 |j_t'|dt \le |u|\bigg(\cosh\Big(\sqrt{\mathcal{C}\big(\mathrm{supp}(\mu)^{2\varepsilon}\big)}|v|\Big)-1\bigg).$$

Using this inequality with equation (8.2), we obtain

$$\big(\nabla(\varphi*\rho^\varepsilon)(x) - (\nabla\varphi)*\rho^\varepsilon(x)\big)\cdot u \le C_\varepsilon|u|\int \Big|\nabla\varphi\big(\exp_x(v)\big)\Big|\rho^\varepsilon(v)dv,$$

which is equivalent to

$$\big|\nabla(\varphi*\rho^\varepsilon) - (\nabla\varphi)*\rho^\varepsilon\big|(x) \le C_\varepsilon\int \Big|\nabla\varphi\big(\exp_x(v)\big)\Big|\rho^\varepsilon(v)dv.$$

Taking the squares and integrating we get

$$\begin{aligned}
\|\nabla(\varphi*\rho^\varepsilon) - (\nabla\varphi)*\rho^\varepsilon\|_\mu^2 &\le C_\varepsilon^2\bigg(\iint \Big|\nabla\varphi\big(\exp_x(v)\big)\Big|\rho^\varepsilon(v)dv\,d\mu(x)\bigg)^2\\
&\le C_\varepsilon^2\iint |\nabla\varphi\big(\exp_x(v)\big)\Big|^2\rho^\varepsilon(v)dv\,d\mu(x)\\
&= C_\varepsilon^2\iint |\nabla\varphi(y)|^2\rho^\varepsilon\big(\exp_x^{-1}(y)\big)D(x,y)d\mathrm{vol}(y)d\mu(x)\\
&\le (1+\varepsilon\tilde{C})C_\varepsilon^2\iint |\nabla\varphi(y)|^2\rho^\varepsilon\big(\exp_x^{-1}(y)\big)D(y,x)d\mu(x)d\mathrm{vol}(y)\\
&= (1+\varepsilon\tilde{C})C_\varepsilon^2\int |\nabla\varphi(y)|^2\mu*\rho^\varepsilon(y)d\mathrm{vol}(y)\\
&= (1+\varepsilon\tilde{C})C_\varepsilon^2\|\nabla\varphi\|_{\mu*\rho^\varepsilon}^2,
\end{aligned}$$

where $\tilde{C} = \tilde{C}(\mathrm{supp}(\mu)^{\varepsilon})$. □

REFERENCES

[1] G. Alberti, *On the structure of singular sets of convex functions*, Calc. Var. and Part. Diff. Eq., 2:1 (1994), pp. 17–27

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[3] L. Ambrosio and N. Gigli, *Construction of parallel transport in the Wasserstein space*, Methods and Applications of Analysis, 15:1 (2008), pp. 1–30.

[4] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in spaces of probability measures*, Birkäuser, 2005.

[5] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

[6] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions.*, Comm. Pure Appl. Math. 44, 4 (1991), pp. 375–417.

[7] X. Cabré, *Nondivergent elliptic equations on manifolds with nonegative curvature*, Comm. on Pure and Applied Math., 50 (1998), pp. 623–665.

[8] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal., 179 (2006), pp. 217–263.

[9] M. P. do Carmo, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[10] L.C. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137 (1999), pp. viii+66.

[11] A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds*, Israel J. Math., 175 (2010), pp. 1–9.

[12] A. Figalli and N. Gigli, *Local semiconvexity of Kantorovich potential on non-compact manifold*, accepted paper, ESAIM COCV.

[13] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math. 177, 2 (1996), pp. 113–161.

[14] W. Gangbo, T. Nguyen, and A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Methods and Applications of Analysis, 15:2 (2008), pp. 155–184.

[15] N. Gigli, *On the geometry of the space of measures in $\mathbb{R}^d$ endowed with the quadratic optimal transportation distance*, PhD thesis, Scuola Normale Superiore, Pisa, 2008. Available at: http://cvgmt.sns.it

[16] N. Gigli, *on Hölder continuity in time of the optimal transport map towards measures along a curve*, accepted paper at Proceedings of the Edinburgh Mathematical Society, available at: http://cvgmt.sns.it

[17] N. Gigli, *Second order calculus on $(P_2(M), W_2)$*, accepted paper at Memoirs of the AMS, available at: http://cvgmt.sns.it

[18] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17 (electronic).

[19] J. Jost, *Riemannian Geometry and Geometric Analysis* (1st edition), Springer, 1991.

[20] J. Lott, *Some geometric calculations in the wasserstein space*, Comm. Math. Phys., 277:2 (2008), pp. 423–437. (2007).

[21] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2), 169:3 (2009), pp. 903–991.

[22] X.-N. Ma, N. S. Trudinger, and X.-J. Wang, *Regularity of potential functions of the optimal transportation problem*, Arch. Ration. Mech. Anal. 177, 2 (2005), pp .151–183.

[23] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. 80, 2 (1995), pp. 309–323.

[24] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128 (1997), pp. 153–179.

[25] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal. 11, 3 (2001), pp. 589–608.

[26] L. Natile and G. Savaré, *A Wasserstein approach to the one-dimensional sticky particle system*, SIAM J. Math. Anal., 41:4 (2009), pp. 1340–1365.

[27] S.-I. Ohta, *Gradient flows on wasserstein spaces over compact alexandrov spaces*, Amer. J. Math., 131 (2009), pp. 475–516.

[28] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

[29] G. Savaré, *Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds*, C. R. Math. Acad. Sci. Paris, 345 (2007), pp. 151–154.

[30] K.-T. Sturm, M.K. von Renesse, *Entropic Measure and Wasserstein Diffusion*, Ann. Probab., 37:3 (2009), pp. 1114–1191.

[31] K.-T. Sturm, *Entropic measure on multidimensional spaces*, Stochastic analysis, random fields and applications VI. Progress in Probability. Birkhuser 2010.

[32] K.-T. Sturm, *On the geometry of metric measure spaces. I-II*, Acta Math. 196, 1 (2006), pp. 65–177.

[33] C. Villani, *Optimal transport, old and new*, Springer Verlag, 2008.

[34] L. Zajíček, *On the differentiability of convex functions in finite and infinite dimensional spaces*, Czechoslovak Math. J., 29 (104) (1979), pp. 340–348.