The Monge problem on non-compact manifolds

Alessio Figalli *
Scuola Normale Superiore
Pisa

Abstract

In this paper we prove the existence of an optimal transport map on non-compact manifolds for a large class of cost functions that includes the case $c(x,y) = d(x,y)$, under the only hypothesis that the source measure is absolutely continuous with respect to the volume measure. In particular, we assume compactness neither of the support of the source measure nor of that of the target measure.

1 Introduction

Monge transportation problem consists in moving a distribution of mass into another one in an optimal way, that is minimizing a given cost of transport. In a mathematical language, the problem can be stated as follows: given two probability distributions μ and ν, with respective support in measurable spaces X and Y, find a measurable map $T : X \to Y$ such that

$$T_\sharp \mu = \nu, \tag{1}$$

i.e.

$$\nu(A) = \mu\left(T^{-1}(A)\right) \quad \forall A \subset Y \text{ measurable},$$

and in such a way that T minimizes the transportation cost, that is

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\sharp \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\},$$

where $c : X \times Y \to \mathbb{R}$ is a given cost function. When condition (1) is satisfied, we say that T is a *transport map*, and if T minimizes also the cost we call it *optimal transport map*. The difficulties in solving such problem even in an Euclidean setting motived Kantorovich to find a relaxed formulation (see [11], [12]). He suggested to look for *plans* instead of transport maps, that is probability measures γ in $X \times Y$ whose marginals are μ and ν, i.e.

$$(\pi_X)_\sharp \gamma = \mu \quad \text{and} \quad (\pi_Y)_\sharp \gamma = \nu,$$

*a.figalli@sns.it

where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the canonical projections. Denoting by $\Pi(\mu, \nu)$ the class of plans, the new minimization problem becomes then the following:

$$\min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_X c(x,y)\, d\gamma(x,y) \right\}. \tag{2}$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan.* Using weak topologies, the existence of solutions to (2) becames simple under the assumption that X and Y are Polish spaces and c is lower semicontinuous (see [15, Chapter 1]). The connection between the formulation of Kantorovich and that of Monge can be seen by noticing that any transport map T induces the plan defined by $(Id \times T)_\sharp \mu$ which is concentrated on the graph ot T.

In a forthcoming paper with Fathi [8], we prove that, in the case $X = Y = M$ with M a smooth manifold, if μ gives zero mass to sets of dimension at most $n-1$ and the cost function is induced by a C^2 Lagrangian that verifies some reasonable assumptions, then the Monge problem has a unique solution and this coincide with the solution of the Kantorovich problem, which turns out to be unique. In particular, this result covers all the cases $c(x,y) = d(x,y)^p$ for $p > 1$, but not the limit case $p = 1$. So this paper arises as an extension to non-compact manifolds of the results of Bernard and Buffoni proved in [5], where the authors showed the existence of optimal transport maps for a large class of costs (that includes in particular the case $c(x,y) = d(x,y)$) on compact manifolds without boundary. The existence of an optimal transport maps in the case $c(x,y) = d(x,y)$ on non-compact manifolds has also been proved in [10] under the assumption of compactness of the supports of the two measures (see the references in [10] for earlier works in the same spirit). In the present paper we extend this result at the case of non-compactly supported measures and, more generally, we prove the existence of an optimal transport for a larger class of cost functions, which is the class of Mañé potentials associated to a supercritical Lagrangian, using in particolar results on weak KAM theory on non-compact manifolds (see [9]).

We remark that we do not assume, as usual in the standard theory of optimal trasportation, that the cost function is bounded by below. In fact such assumption would be quite nonnatural for a Mañé potential and, also in particular cases, it would not be simple to check its validity. So, in order to apply the standard duality result that gives us an optimal pair for the dual problem, the idea will be to add to our cost a null-Lagrangian, so that the cost becomes nonnegative and still satisfies the triangle inequality, and the minimization problem does not change (see Section 2.2).

1.1 The main result

Let M be a smooth $n-$dimensional manifold, g a complete Riemannian metric on M. We fix $L : TM \to \mathbb{R}$ a C^2 Lagrangian on M, that satisfies the following hypotheses:

(L1) **C^2-strict convexity:** $\forall (x,v) \in TM$, the second derivative along the fibers $\nabla_v^2 L(x,v)$ is positive strictly definite;

(L2) **uniform superlinearity:** for every $K \geq 0$ there exists a finite constant $C(K)$ such that

$$\forall (x,v) \in TM, \quad L(x,v) \geq K\|v\|_x + C(K),$$

where $\|\cdot\|_x$ is the norm on T_xM induced by g;

(L3) **uniform boundedness in the fibers:** for every $R \geq 0$, we have

$$A(R) := \sup_{x\in M} \{L(x,v) \mid \|v\|_x \leq R\} < +\infty.$$

We define the cost function

$$c_T(x,y) := \inf_{\gamma(0)=x,\ \gamma(T)=y} \int_0^T L(\gamma(t), \dot\gamma(t), t)\, dt.$$

The assumptions on the Lagrangian ensure that the inf in the definition of $c_T(x,y)$ is attained by a curve of class C^2. We now define the cost

$$c(x,y) := \inf_T c_T(x,y).$$

In the theory of Lagrangian Dynamics, this function is usually called Mañé potential. We now make the last assumption on L:

(L4) **supercriticality:** for each $x \neq y \in M$, we have $c(x,y) + c(y,x) > 0$.

This assumption ensures that also the inf in the definition of $c(x,y)$ is attained by a curve of class C^2. We will consider the Monge transportation problem for the cost c. Our main result is the following:

Theorem 1.1. *Assume that c is the cost function associated to a supercritical Lagrangian that satisfies all the assumption above. Suppose that*

$$\int_{M\times M} d(x,y)\, d\mu(x)\, d\nu(y) < +\infty,$$

where d is the distance associated to the Riemannian metric. If μ is absolutely continuous with respect to the volume measure, then there exists an optimal transport map $T : M \to M$ for the Monge transportation problem between μ and ν. This map turns out to be optimal for the Kantorovich problem. More precisely, the plan associated to this map is the unique minimizer of the secondary variational problem

$$\min \int_{M\times M} \sqrt{1 + (c(x,y) - U(y) + U(x))^2}\, d\gamma(x,y)$$

among all optimal plans for (2), where U is a strict subsolution of the Hamilton Jacobi equation (see Proposition 2.2).

We recall that the idea of using a secondary variational problem in order to select a "good" optimal plan was first used in [2] and refined in [3].

Remark 1.2. *We observe that, by the triangle inequality for the distance, the condition*

$$\int_{M\times M} d(x,y)\,d\mu(x)\,d\nu(y) < +\infty$$

is equivalent to the existence of a point $x_0 \in M$ such that

$$\int_M d(x,x_0)\,d\mu(x) < +\infty,$$

$$\int_M d(y,x_0)\,d\nu(y) < +\infty.$$

In fact, fixed $x_0, x_1 \in M$, since $d(x,x_0) - d(x_0,x_1) \le d(x,x_1) \le d(x,x_0) + d(x_0,x_1)$,

$$x \mapsto d(x,x_0) \quad \textit{is integrable if and only if} \quad x \mapsto d(x,x_1) \quad \textit{is integrable.}$$

In particular all Lipschitz functions on M are integrable with respect to both μ and ν.

We remark that the Lagrangian

$$L(x,v) = \frac{1 + \|v\|_x^2}{2}$$

satisfies all the hypotheses of the above theorem and, in this case, we obtain

$$c(x,y) = d(x,y).$$

2 Definitions and preliminary results

2.1 Preliminaries in Lagrangian Dynamics

We recall some results of Lagrangian Dynamics that will be useful in the sequel (see [6], [7], [13]) and that shows the naturality of the supercriticality assumption.

Proposition 2.1. *Let L be a Lagrangian that satisfies assumption (L1), (L2) and (L3). For $k \in \mathbb{R}$, let us define c_k the Mañé potential associated to the Lagrangian $L+k$. Then there exists a constant k_0 such that*

(i) for $k < k_0$, then $c_k \equiv -\infty$ and the Lagrangian is called subcritical;

(ii) for $k \ge k_0$, c_k is locally Lipschitz on $M \times M$ and satisfies the triangle inequality

$$c_k(x,z) \le c_k(x,y) + c_k(y,z) \quad \forall x,y,z \in M;$$

in addition $c_k(x,x) = 0\ \forall x \in M$;

(iii) for $k > k_0$, the Lagrangian L is supercritical, that is $c(x, y)+c(y, x) > 0$ for each $x \neq y \in M$.

The following proposition is a simple corollary of the results proved in [9]:

Proposition 2.2. *The Lagrangian L is supercritical if and only if there exist $\delta > 0$ and a C^∞ function U such that*

$$H(x, d_x U) \leq -\delta, \quad \forall x \in M,$$

or equivalently

$$L(x, v) - d_x U(v) \geq \delta, \quad \forall (x, v) \in TM,$$

where H is the Hamiltonian associated to the Lagrangian L, that is

$$\forall (x, p) \in T^*M, \quad H(x, p) := \sup_{v \in T_x M} \{\langle p, v\rangle - L(x, v)\}.$$

Proof. The value k_0 is the so called critical value of L, and is the smallest value for which there exists a global C^1 subsolution of

$$H(x, d_x u) = k$$

(under the assumptions made on the Lagrangian, this value exists and is unique). Then it suffices to apply the following approximation result, also proven in [9]:

Theorem 2.3. *If $u : M \to \mathbb{R}$ is locally Lipschitz, with its derivative $d_x u$ satisfying $H(x, d_x u) \leq k$ almost everywhere, then for each $\varepsilon > 0$ there exists a C^∞ function $u_\varepsilon : M \to \mathbb{R}$ such that $H(x, d_x u_\varepsilon) \leq k + \varepsilon$ and $|u(x) - u_\varepsilon(x)| \leq \varepsilon$ for each $x \in M$.*

In fact, if L is supercritical, then k_0 is strictly negative, and it suffices to use the theorem above with $\varepsilon = \frac{|k_0|}{2}$. On the other hand, the inverse implication follows by the caracterization of c_{k_0} made above. □

We observe that, in the case $L(x, v) = \frac{1}{2}\left(1 + \|v\|_x^2\right)$, it suffices to take $U \equiv 0$, $\delta = \frac{1}{2}$.

2.2 Duality and Kantorovich potential

It is well-known that a linear minimization problem with convex constraints, like (2), admits a dual formulation. Before stating the duality formula, we recall the definition of c-transform (see [2], [4], [5], [8], [15], [16]).

Definition 2.4 (**c-transform).** Given a pair of Borel functions $\varphi : X \to \mathbb{R} \cup \{-\infty\}$, $\psi : Y \to \mathbb{R} \cup \{+\infty\}$, with

$$\int_X |\varphi|\, d\mu < +\infty \quad \text{and} \quad \int_Y |\psi|\, d\nu < +\infty,$$

we say that ψ is the *c-transform* of φ if

$$\forall (x, y) \in X \times Y, \quad \psi(y) = \sup_{x \in M} \varphi(x) - c(x, y),$$

and we write $\psi = \varphi^c$.

Observe that, if $\psi = \varphi^c$, then one obviously has $\varphi(x) - \psi(y) \le c(x,y)$ for all $(x,y) \in X \times Y$. Integrating this inequality on the product $X \times Y$, one obtains

$$\begin{aligned}\forall \gamma \in \Pi(\mu,\nu), \qquad \int_X \varphi\, d\mu - \int_Y \psi\, d\nu &= \int_{X\times Y} (\varphi(x) - \psi(y))\, d\gamma(x,y) \\ &\le \int_{X\times Y} c(x,y)\, d\gamma(x,y).\end{aligned}$$

Theorem 2.5 (Duality formula). *Let X and Y be Polish spaces equipped with probability measures μ and ν respectively, $c : X \times Y \to \mathbb{R}$ a lower semicontinuous cost function bounded from below such that*

$$\int_{X\times Y} c(x,y)\, d\mu(x)\, d\nu(y) < +\infty.$$

Then

$$\min_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) \right\} = \max_{(\varphi,\psi)\in L^1(\mu)\times L^1(\nu)} \left\{ \int_X \varphi\, d\mu - \int_Y \psi\, d\nu \mid \psi = \varphi^c \right\}.$$

For a proof of this theorem see [1, Theorem 6.1.5], [2, Theorems 3.1 and 3.2], [16, Theorem 5.9].

We now consider a cost function $c(x,y)$ as in Theorem 1.1. By hypothesis (L3), c is Lipschitz. In fact, given $x, y \in M$, we consider a geodesic $\gamma_{x,y} : [0, d(x,y)] \to M$ from x to y with $\|\dot\gamma_{x,y}\| = 1$. Then

$$c(x,y) \le \int_0^{d(x,y)} L(\gamma_{x,y}(t), \dot\gamma_{x,y}(t))\, dt \le A(1) d(x,y),$$

and so we have

$$\begin{aligned}|c(x,y) - c(z,w)| &\le |c(x,y) - c(z,y)| + |c(z,y) - c(z,w)| \\ &\le \max\{|c(x,z)|, |c(z,x)|\} + \max\{|c(y,w)|, |c(w,y)|\} \\ &\le A(1)[d(x,z) + d(y,w)].\end{aligned}$$

Moreover c satisfies $c(x,x) \equiv 0$ and the triangle inequality

$$c(x,z) \le c(x,y) + c(y,z)$$

(see Proposition 2.1). Fix now $z \in M$ and consider the auxiliar cost

$$\overline{c}(x,y) := c(x,y) + a(y) - a(x),$$

with $a(x) := c(x,z)$. Obviously $\overline{c}$ still satisfies the triangle inequality. Moreover, since c is Lipschitz and satisfies the triangle inequality, we have

$$0 \le \overline{c}(x,y) \le c(x,y) + c(y,x) \le 2A(1) d(x,y). \tag{3}$$

Thus $\overline{c}(x,y)$ is integrable with respect to $\mu \otimes \nu$ if so it is $d(x,y)$, and in this case we can apply Theorem 2.5 to prove the following:

Theorem 2.6. *Given two probability measures μ and ν on M such that*

$$\int_{M\times M} d(x,y)\,d\mu(x)\,d\nu(y) < +\infty,$$

let $\overline{c}$ be a cost function as above. Then there exists a Lipschitz function $\overline{u}: M \to \mathbb{R}$ that satisfies

$$\overline{u}(y) - \overline{u}(x) \le \overline{c}(x,y) \quad \forall x,y \in M$$

and

$$\int_M \overline{u}\,d(\nu-\mu) = \int_{M\times M} \overline{c}\,d\gamma$$

for each γ optimal transport plan between μ and ν. In particular, this implies

$$\overline{u}(y) - \overline{u}(x) = \overline{c}(x,y) \quad \textit{for } \gamma\text{-a.e. } (x,y) \in M\times M,$$

that is

$$(\overline{u}-a)(y) - (\overline{u}-a)(x) = c(x,y) \quad \textit{for } \gamma\text{-a.e. } (x,y) \in M\times M,$$

The Lipschitz function $u := \overline{u} - a$ is called a Kantorovich potential.

Proof. Let $(\varphi,\psi) \in L^1(\mu)\times L^1(\nu)$ be a pair that realizes the maximum in the dual problem, with $\psi = \varphi^{\overline{c}}$. We want to prove that it suffices to take $\overline{u} = -\psi$.
Fix $x,y \in M$. Since $\psi = \varphi^{\overline{c}}$ and $\overline{c}$ satisfies the triangle inequality, we have

$$\psi(x) = \sup_{z\in M} \varphi(z) - \overline{c}(x,z) \le \sup_{z\in M} \psi(y) + \overline{c}(z,y) - \overline{c}(x,z) \le \psi(y) + \overline{c}(x,y).$$

Let now x_0 be a point of M such that $\psi(x_0) \in \mathbb{R}$ (such a point exists, being $\psi \in L^1$). Choosing in the inequality above first $x = x_0$ and after $y = x_0$, we obtain that ψ is finite everywhere. So we can subtract $\psi(y)$ to the two sides, obtaining

$$(-\psi)(y) - (-\psi)(x) \le \overline{c}(x,y).$$

Thus, if we define $\overline{u} := -\psi$, by (3) we have

$$\overline{u}(y) - \overline{u}(x) \le \overline{c}(x,y) \le 2A(1)d(x,y) \quad \forall x,y \in M.$$

This inequality tells us that $\overline{u}$ is $2A(1)$-Lipschitz, and so, by Remark 1.2, $\overline{u} \in L^1(\mu)\cap L^1(\nu)$. Observe now that

$$0 = \overline{c}(x,x) \ge \varphi(x) - \psi(x) \quad \Rightarrow \quad -\overline{u}(x) \ge \varphi(x).$$

Thus, if γ is an optimal transport plan between μ and ν, we have

$$\begin{aligned}\int_{M\times M} \overline{c}(x,y)\,d\gamma(x,y) &\ge \int_{M\times M} (\overline{u}(y) - \overline{u}(x))\,d\gamma(x,y)\\ &= \int_M \overline{u}\,d(\nu-\mu) \ge \int_M \varphi\,d\mu - \int_M \psi\,d\nu = \int_{M\times M} \overline{c}(x,y)\,d\gamma(x,y),\end{aligned}$$

as wanted. □

2.3 Calibrated curves

Fix a C^∞ function U and a $\delta > 0$ given by Proposition 2.2, and a Kantorovich potential u given by Theorem 2.6. Following [5], we recall some useful definitions.

Definition 2.7 (u-calibrated curve)**.** We say that a continuous piecewise differentiable curve $\gamma : I \to M$ is *u-calibrated* if

$$u(\gamma(t)) - u(\gamma(s)) = \int_s^t L(\gamma(\tau), \dot{\gamma}(\tau))\, d\tau = c(\gamma(s), \gamma(t)) \quad \forall s \leq t \in I,$$

where $I \subset \mathbb{R}$ is a nonempty interval of $\mathbb{R}$ (possibly a point). A u-calibrated curve is called *non-trivial* if I has non-empty interior.

Obviously a non-trivial u-calibrated curve is a minimizing extremal of L, and hence is of class C^2. In addition, we observe that each u-calibrated curve γ can be extended to a maximal one, that is a curve $\tilde{\gamma}$ that can't be extended on an interval that strictly contains I without losing the calibration property (this follows by the fact that, fixed the initials position and the velocity, the minimizer is unique; thus, if two u-calibrated curves locally coincide, they must coincide in the intersection of their domains of definition, and so one can use this fact to find an unique maximal extension of γ). We observe that, if γ is maximal, then I must be closed. In the sequel, also in the case $I = \mathbb{R}$, $I = [a, +\infty)$ or $(-\infty, b]$, for simplicity of notation we will always write the interval on which a maximal curve is defined as $[a, b]$.

Definition 2.8 (transport ray)**.** A *transport ray* is the image of a non-trivial u-calibrated curve.

In [7], it is proved that Kantorovich potentials are viscosity subsolutions of the Hamilton-Jacobi equation, that is equivalent to say that u is locally Lipschitz and satisfies

$$H(x, d_x u) \leq 0, \quad \text{for a.e. } x \in M,$$

or equivalently

$$L(x, v) - d_x u(v) \geq 0, \quad \text{for a.e. } x \in M, \ \forall v \in T_x M.$$

We recall that, if $\gamma : [a, b] \to \mathbb{R}$ is a u-calibrated curve, then for all $t \in (a, b)$ the function u is differentable at $\gamma(t)$ (see [7]). Then we have the following:

Lemma 2.9. *Let $\gamma : [a, b] \to \mathbb{R}$ be a u-calibrated curve. Then for all $t \in (a, b)$ the function u is differentable at $\gamma(t)$ and we have*

$$d_{\gamma(t)}(u - U)(\dot{\gamma}(t)) \geq \delta,$$

where U and δ are given by Proposition 2.2. This implies that γ is an embedding and transport rays are non-trivial embedded arcs.

Proof. As $\gamma(t)$ is u-calibrated, we have

$$u(\gamma(t)) - u(\gamma(s)) = \int_s^t L(\gamma(\tau), \dot{\gamma}(\tau))\, d\tau \quad \forall s \leq t, \ s, t \in [a, b],$$

that implies, recalling Proposition 2.2,

$$\frac{u(\gamma(t)) - u(\gamma(s))}{t-s} = \frac{1}{t-s}\int_s^t L\left(\gamma(\tau), \dot\gamma(\tau)\right)\, d\tau$$
$$\Rightarrow d_{\gamma(t)}u(\dot\gamma(t)) = L\left(\gamma(t), \dot\gamma(t)\right) \geq d_{\gamma(t)}U(\dot\gamma(t)) + \delta.$$

□

We now define the functions $\alpha : M \to \mathbb{R}$ and $\beta : M \to \mathbb{R}$ as follows:

- $\alpha(x)$ is the supremum of all times $T \geq 0$ such that there exists a u-calibrated curve $\gamma : [-T, 0] \to M$ such that $\gamma(0) = x$;
- $\beta(x)$ is the supremum of all times $T \geq 0$ such that there exists a u-calibrated curve $\gamma : [0, T] \to M$ such that $\gamma(0) = x$.

Lemma 2.10. *α and β are Borel functions.*

Proof. Let $K^i \subset M$ be a countable increasing sequence of compact set such that $\cup_i K^i = M$. Then we can define the auxiliary functions $\alpha_i(x)$ as the supremum of all times $T \geq 0$ such that there exists a u-calibrated curve $\gamma : [-T, 0] \to M$ such that $\gamma(0) = x$ and $\gamma(-T) \in K^i$. We will prove that α_i is upper semicontinuous for each i, and this will implies the measurability of α as $\alpha(x) = \sup_i \alpha_i(x)$ for all $x \in M$ (the case of β is analogous).
Fix $i \in \mathbb{N}$ and let $(x_j) \subset M$ be a sequence converging to a limit $x \in M$ such that $\alpha_i(x_j) \geq T$ for all j. Then we know that there exists a sequence $\gamma_j : [-T, 0] \to M$ of u-calibrated curves such that $\gamma_j(0) = x_j$ and $\gamma_j(-T) \in K^i$. As $\gamma_j(-T) \in K^i$, we know that there exists a constant A such that $\|\dot\gamma_j(0)\|_{\gamma_j(0)} \leq A$ for all j (see for example the Appendix in [8]). Then, taking a subsequence if necessary, we can assume that γ_j converges uniformly on $[-T, 0]$ to a curve $\gamma : [-T, 0] \to M$ which is still u-calibrated, as it is easy to see, and satisfies $\gamma(0) = x$, $\gamma(-T) \in K^i$. Then $\alpha_i(x) \geq T$. □

We now can define the following Borel sets:

Definition 2.11. We define the set $\mathcal{T}$ given by the union of all the transport rays as

$$\mathcal{T} := \{x \in M \mid \alpha(x) + \beta(x) > 0\}.$$

For $\varepsilon \geq 0$, we define the sets

$$\mathcal{T}_\varepsilon := \{x \in M \mid \alpha(x) > \varepsilon,\ \beta(x) > \varepsilon\}.$$

Clearly $\mathcal{T}_\varepsilon \subset \mathcal{T}$ for all $\varepsilon \geq 0$ and the set $\mathcal{E} := \mathcal{T} - \mathcal{T}_0$ is the set of ray ends.

We now recall the following:

Theorem 2.12. *The function u is differentiable at each point of $\mathcal{T}_0$. For each point $x \in \mathcal{T}_0$, there exists a unique maximal u-calibrated curve*

$$\gamma_x : [-\alpha(x), \beta(x)] \to M$$

such that $\gamma(0) = x$. This curve satisfies the relations

$$d_x u = \nabla_v L(x, \dot{\gamma}_x(0))$$

or equivalently

$$\dot{\gamma}_x(0) = \nabla_p H(x, d_x u).$$

For each $\varepsilon > 0$, the differential $x \mapsto d_x u$ is locally Lipschitz on $\mathcal{T}_\varepsilon$, or equivalently the map $x \mapsto \dot{\gamma}_x(0)$ is locally Lipschitz on $\mathcal{T}_\varepsilon$.

For a proof see [7].

Definition 2.13. For $x \in M$, we will denote by R_x the union of the transport rays containing x. We also denote

$$R_x^+ := \{y \in M \mid u(y) - u(x) = c(x, y)\}.$$

We observe that $R_x = \gamma_x([-\alpha(x), \beta(x)])$ for all $x \in \mathcal{T}_0$.

In order to conclude this section, we recall two results of [5].

Lemma 2.14. *We have*

$$R_x^+ = \begin{cases} \gamma_x([0, \beta(x)]) & \text{if } x \in \mathcal{T}_0, \\ \{x\} & \text{if } x \in M \setminus \mathcal{T}, \end{cases}$$

where γ_x is given by Theorem 2.12.

Proof. Let x be a point of $\mathcal{T}_0$. By the calibration property of γ_x, we have

$$u(\gamma_x(t)) - u(\gamma_x(0)) = c(\gamma_x(0), \gamma_x(t)) \quad \forall t \in [0, \beta(x)],$$

that is

$$\gamma_x(t) \in R_x^+ \quad \forall t \in [0, \beta(x)],$$

and so we have $\gamma_x([0, \beta(x)]) \subset R_x^+$. Conversely, let us fix $x \in M$, $y \in R_x^+$. Then we know that there exists a u-calibrated curve $\gamma : [0, T] \to M$ such that

$$\int_0^T L\left(\gamma(t), \dot{\gamma}(t)\right) dt = c(x, y) = u(y) - u(x), \quad \gamma(0) = x, \ \gamma(T) = y.$$

So, if $x \in \mathcal{T}_0$, by Theorem 2.12 $\gamma = \gamma_x|_{[0,T]}$ and hence $y = \gamma(T) = \gamma_x(T) \in \gamma_x([0, \beta(x)])$, while, if $x \notin \mathcal{T}$, there is no non-trivial u-calibrated curve and then we must have $y = x$ in the above discussion, that implies $R_x^+ = \{x\}$. □

Proposition 2.15. *A transport plan γ is optimal for the cost c if and only if it is supported on the closed set*

$$\bigcup_{x \in M} \{x\} \times R_x^+ = \{(x, y) \in M \times M \mid c(x, y) = u(y) - u(x)\}.$$

Proof. By Theorem 2.6, γ is optimal if and only if

$$\int_{M \times M} c(x, y)\, d\gamma(x, y) = \int_M u(x)\, d(\nu - \mu)(x) = \int_{M \times M} (u(y) - u(x))\, d\gamma(x, y).$$

The conclusion follows observing that $c(x, y) \geq u(y) - u(x)$ for all $x, y \in M$. □

3 Proof of the main theorem

The line of the proof is essentially the same as in [5], where the authors, using ideas of Lagrangian Dynamics, extend to a Riemannian setting the results obtained in the Euclidean case in [2].
As before, we fix a C^∞ function U and a $\delta > 0$ given by Proposition 2.2, and then fix a Kantorovich potential u given by Theorem 2.6 (see Sections 2.2, 2.3). We now define the second cost function

$$\tilde{c}(x, y) := \phi(c(x, y) - U(y) + U(x)),$$

with $\phi(t) := \sqrt{1 + t^2}$. Consider the secondary variational problem

$$\min_{\gamma \in \mathcal{O}} \int_{M \times M} \tilde{c}(x, y)\, d\gamma(x, y), \tag{4}$$

where $\mathcal{O}$ is the set of optimal transport plan, and select a minimizer γ_0 of this secondary variational problem. We now want to prove that it is supported on a graph.
The idea is the following: first one sees that the measure γ_0 is concentrated on a σ-compact set $\Gamma \subset \cup_{x \in M}\{x\} \times R_x^+$ which is $\tilde{c}$-cyclically monotone in a weak sense that we will define later in the proof. Then one considers the set on which the transport plan is not a graph, that is

$$\Lambda := \{x \in M \mid \#(\Gamma_x) \geq 2\},$$

where $\Gamma_x := \{y \in M \mid (x, y) \in \Gamma\}$ and $\#$ denotes the cardinality of the set. In this way, intersecting Λ with a transport ray R, thanks to the monotonicity of $\Lambda \cap R$ it is simple to see that $\Lambda \cap R$ is at most countable. Finally Theorem 2.12 allows us to parametrize the transport rays in a locally Lipschitz way. This and the fact that $\Lambda \cap R$ has zero $\mathscr{H}^1$-measure for each transport ray R (where $\mathscr{H}^k$ denote the k-dimensional Hausdorff measure) imply that Λ has null volume measure, and so $\mu(\Lambda) = 0$ as wanted.
We divide the proof in many steps, in order to make the overall strategy more clear.

Step 1: the construction of Γ.

First we observe that $\tilde{c}$ is integrable with respect to $\mu \otimes \nu$. Indeed, since ϕ has linear growth, it suffices to prove that

$$\underline{c}(x, y) := c(x, y) - U(y) + U(x)$$

is $\mu \otimes \nu$-integrable. The uniform boundedness in the fiber of $L(x, v)$ implies that the Hamiltonian $H(x, p)$ is uniformly superlinear. By this and the inequality $H(x, d_x U) \leq 0$, we get that the gradient of U is uniformly bounded, which implies that U is Lipschitz, that is

$$|U(y) - U(x)| \leq C d(x, y) \quad \forall x, y \in M.$$

So we have

$$0 \leq \underline{c}(x, y) \leq (C + A(1)) d(x, y),$$

and then $\underline{c}(x,y)$ is integrable with respect to $\mu \otimes \nu$, since so is $d(x,y)$ by assumption.
Let us consider the lower semicontinuous function $\zeta : M \times M \to [0,+\infty]$ given by

$$\zeta(x,y) = \begin{cases} \tilde{c}(x,y) & \text{if } u(y)-u(x) = c(x,y), \\ +\infty & \text{otherwise.} \end{cases}$$

Note that, as $\gamma_0(\{(x,y) \in M \times M \mid u(y)-u(x) = c(x,y)\}) = 1$,

$$\int_{M\times M} \zeta(x,y)\, d\gamma_0(x,y) = \int_{M\times M} \tilde{c}(x,y)\, d\gamma_0(x,y) < +\infty$$

and, thanks to Proposition 2.15, we have that γ_0 is a minimizer for the Kantorovich problem

$$\min_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_X \zeta(x,y)\, d\gamma(x,y) \right\}.$$

It is then a standard result that γ_0 is concentrated on a set $\tilde{\Gamma}$ that is ζ-ciclically monotone, that is if $\big((x_i,y_i)\big)_{1\le i\le l}$ is a finite family of points of $\tilde{\Gamma}$ and $\sigma(i)$ is a permutation, we have

$$\sum_{i=1}^{l} \zeta(x_i, y_{\sigma(i)}) \ge \sum_{i=1}^{l} \zeta(x_i, y_i)$$

(for a proof see, for example, [2, Theorem 3.2]). By the definition of ζ this implies the following monotonicity property of $\tilde{\Gamma}$:
if $\big((x_i,y_i)\big)_{1\le i\le l}$ is a finite family of points of $\tilde{\Gamma}$ and $\sigma(i)$ is a permutation such that $\big((x_i,y_{\sigma(i)})\big)_{1\le i\le l}$ is still contained in $\tilde{\Gamma}$, then

$$\sum_{i=1}^{l} \tilde{c}(x_i, y_{\sigma(i)}) \ge \sum_{i=1}^{l} \tilde{c}(x_i, y_i).$$

By inner regularity of the Borel measure γ_0, there exists a σ-compact subset $\Gamma \subset \tilde{\Gamma}$ on which γ_0 is concentrated. Obviously Γ is still monotone in the sense defined above.

Step 2: Λ is a Borel set and $\Lambda \subset \mathcal{T}$.

Now that we have constructed Γ, we define

$$\Lambda := \{x \in M \mid \#\Gamma_x \ge 2\},$$

where $\Gamma_x := \{y \in M \mid (x,y) \in \Gamma\}$. Let K^i be an countable increasing sequence of compact set such that $\Gamma = \cup_i K^i$ (we recall that Γ is σ-compact). For each $x \in M$, we consider the compact set $K^i_x := \{y \in M \mid (x,y) \in K^i\}$ and we define the upper semicontinuous function

$$\delta_i(x) := \mathrm{diam}(K^i_x),$$

where diam denotes the diameter of the set. Then $\delta(x) := \sup_i \delta_i(x) = \mathrm{diam}(\Gamma_x)$ is a Borel function and so

$$\Lambda = \{x \in M \mid \delta(x) > 0\}$$

is a Borel subset of M.
Let us now show that $\Lambda \subset \mathcal{T}$. If $x \notin \mathcal{T}$, then, by Lemma 2.14, $R^+_x = \{x\}$. Hence, as $\Gamma_x \subset R^+_x$ (see Proposition 2.15), $\Gamma_x \subset \{x\}$ and $x \notin \Lambda$.

Step 3: $\Lambda \cap R$ is at most countable for each transport ray R.

We fix a transport ray, that is the image of a non-trivial maximal u-calibrated curve $\gamma : [a, b] \to M$, and we consider the stricly increasing function $f : [a, b] \to \mathbb{R}$ defined by $f = (u - U) \circ \gamma$ (see Lemma 2.9). We observe that

$$\tilde{c}(\gamma(s), \gamma(t)) = \phi(f(t) - f(s)) \quad \forall s \leq t, \; s, t \in [a, b].$$

In view of the monotonicity of Γ we have

$$\phi(f(t) - f(s)) + \phi(f(t') - f(s')) \leq \phi(f(t') - f(s)) + \phi(f(t) - f(s'))$$

whenever $(\gamma(s), \gamma(t)) \in \Gamma$, $(\gamma(s'), \gamma(t')) \in \Gamma$, $s \leq t'$, $s' \leq t$. Now, following [2], we show the implication

$$(\gamma(s), \gamma(t)) \in \Gamma, \quad (\gamma(s'), \gamma(t')) \in \Gamma, \quad s < s' \quad \Rightarrow \quad t \leq t'. \tag{5}$$

Assume by contradiction that $t > t'$. Since $s \leq t$ and $s' \leq t'$, we have $s < s' \leq t' < t$. In this case, setting $a = f(s') - f(s)$, $b = f(t') - f(s')$, $c = f(t) - f(t')$, we have

$$\phi(a + b + c) + \phi(b) \leq \phi(a + b) + \phi(b + c).$$

On the other hand, since $c > 0$, the strictly convexity of ϕ gives

$$\phi(a + b + c) - \phi(b + c) > \phi(a + b) - \phi(b),$$

and therefore we have a contradiction. By (5), we obtain that the vertical sections Γ_x of Γ are ordered along a transport ray, i.e.

$$\forall y_1 \in \Gamma_{x_1}, \; \forall y_2 \in \Gamma_{x_2}, \quad y_1 \leq y_2 \quad \text{whenever } x_1 = \gamma(s_1), \; x_2 = \gamma(s_2), \; s_1 < s_2.$$

As a consequence, the set of all $x \in R$ such that Γ_x is not a singleton is at most countable, since, if for such x we consider I_x the smallest open interval such that $\gamma\left(\bar{I}_x\right) \supset \Gamma_x$, we obtain a family of pairwise disjoints open intervals of $\mathbb{R}$.

Step 4: covering the set Λ.

As $\Lambda \subset \mathcal{T}$, we will cover $\mathcal{T}$ with a countable family of, so called, transport beams.

Definition 3.1. We call *transport beam* a couple (B, χ) where $B \subset \mathbb{R}^n$ is a Borel subset and $\chi : B \to M$ is a locally Lipschitz map such that:

- there exist a bounded Borel set $\Omega \subset \mathbb{R}^{n-1}$ and two Borel functions $a < b : \Omega \to \mathbb{R}$ such that
$$B = \{(\omega, s) \in \Omega \times \mathbb{R} \mid a(\omega) \leq s \leq b(\omega)\} \subset \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R};$$
- for each $\omega \in \Omega$, the curve $\chi_\omega : [a(\omega), b(\omega)] \to M$ given by $\chi_\omega(s) = \chi(\omega, s)$ is u-calibrated.

We remark that we do not assume that χ is injective.

We now want to prove that there exists a countable family $(B_{j,k}, \chi_{j,k})_{j,k\in\mathbb{N}}$ of transport beams such that the images $\chi_{j,k}(B_{j,k})$ cover the set $\mathcal{T}$.
So let take $D \subset \mathbb{R}^{n-1}$ the closed unit ball and let $\psi_j : D \to M$, $j \in \mathbb{N}$, be a family of smooth embeddings such that, for each maximal u-calibrated curve $\gamma : [a, b] \to \mathbb{R}$, the embedded arc $\gamma((a, b))$ intersect transversally the image of ψ_j for some $j \in \mathbb{N}$. Indeed, in order to construct such embeddings, it suffices to take a countable atlas $(U_i, \theta_i)_{i\in N}$ such that $\theta_i(U_i) = B_2(0) \subset \mathbb{R}^n$ (where $B_r(0)$ denote the n-dimensional ball of radius r centered at the origin) and that satisfies $\cup_{i\in\mathbb{N}}\ \theta_i^{-1}(B_1(0)) = M$, and to consider the image by θ_i^{-1} of the countable family $(D_{l,q})_{1\le l\le n,\ q\in\mathbb{Q}\cap[-1,1]}$ of $(n-1)$-dimensional balls of radius 1 defined by

$$D_{l,q} := \Big\{(x_1, \dots, x_n) \mid x_l = q,\ \sum_{m\neq l} |x_m|^2 \le 1\Big\}.$$

For each $(j, k) \in \mathbb{N}^2$ let us consider the set $\Omega_{j,k} = D \cap \psi_j^{-1}(\mathcal{T}_{\frac{1}{k}})$. Let us define

$$a_{j,k}(\omega) := -\alpha \circ \psi_j : \Omega_{j,k} \to \mathbb{R},$$

$$b_{j,k}(\omega) := \beta \circ \psi_j : \Omega_{j,k} \to \mathbb{R},$$

where α and β were defined in section 2.3. We observe that, by Lemma 2.10, $a_{j,k}$ and $b_{j,k}$ are Borel functions. We can now define the Borel sets

$$B_{j,k} := \{(\omega, s) \in \Omega_{j,k} \times \mathbb{R} \mid a_{j,k}(\omega) \le s \le b_{j,k}(\omega)\}.$$

Finally, we define on $B_{j,k}$ the map

$$\chi_{j,k}(\omega, s) := \gamma_{\psi_j(\omega)}(s).$$

We now observe that $\chi_{j,k}$ is locally Lipschitz. In fact, we can write an extremal using the Euler-Lagrange flow $f_s : TM \to TM$, which is complete because of the energy conservation. Thanks to the hypotheses made on L, the map

$$(s, x, v) \mapsto f_s(x, v)$$

is of class C^1. As we have

$$\chi_{j,k}(\omega, s) := \pi_M \circ f_s(\psi_j(\omega), \dot{\gamma}_{\psi_j(\omega)}(0)),$$

where $\pi_M : TM \to M$ is the canonical projection, in view of Theorem 2.12 we deduce that this map is locally Lipschitz. It is clear that, for each transport ray R, there exist $j, k \in \mathbb{N}$ such that R is contained in $\chi_{j,k}(B_{j,k})$.

Step 5: $\mu(\Lambda) = 0$.

In order to conclude that $\mu(\Lambda) = 0$, it suffices to prove that, if (B, χ) is a transport beam, then the set $\Lambda \cap \chi(B)$ is negligible with respect to the volume measure.
We recall that, for each $\omega \in \Omega$, the curve χ_ω is a locally bilipschitz homeomorphism onto its image, and so, as we know that the set $\Lambda \cap \chi(\{\omega\} \times [a(\omega), b(\omega)])$ is countable, the set $\chi^{-1}(\Lambda) \cap B$ intersects each vertical line $\omega \times \mathbb{R}$ along a countable set, and so in particular has zero $\mathscr{H}^1$-measure. Then, by Fubini's theorem, $\chi^{-1}(\Lambda) \cap B$ has zero $\mathscr{H}^n$-measure in $\mathbb{R}^n$, and so, since locally Lipschitz maps send $\mathscr{H}^n$-null sets into $\mathscr{H}^n$-null sets, we get

$$\mathscr{H}^n(\Lambda \cap \chi(B)) \leq \mathscr{H}^n(\chi(\chi^{-1}(\Lambda) \cap B)) = 0,$$

that implies that $\Lambda \cap \chi(B)$ is negligible with respect to the volume measure.

Step 6: uniqueness.

We now prove that the transport plan selected with the secondary variational problem is unique. Let γ_0, γ_1 be two optimal transport plans, which are optimal also for the secondary variational problem. By what we proved above, we know that they are induced by two transport maps $t_0 : M \to M$ and $t_1 : M \to M$, respectively. Let us now consider $\overline{\gamma} := \frac{\gamma_0+\gamma_1}{2}$. By the linear structure of the two variational problems (2) and (4), $\overline{\gamma}$ is still optimal for both, and so it is induced by a transport map $\overline{t}$. This implies that both γ_0 and γ_1 are concentrated on the graph of $\overline{t}$, and so $t_0 = t_1$ μ-a.e.

Remark 3.2. *We observe that exactly this argument shows also that the set $\mathcal{E}$ of ray ends is negligible with respect to the volume measure.*

Acknowledgements: it's a pleasure to thank Albert Fathi for his suggestions and his support, and to gratefully acknowledge the hospitality of the École Normale Supérieure of Lyon, where this paper was written.

References

[1] L.Ambrosio, N.Gigli & G.Savaré: *Gradient flows in metric spaces and in the Wasserstein space of probability measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser, 2005.

[2] L.Ambrosio & A.Pratelli: *Existence and stability results in the L^1 theory of optimal transportation.* Lectures notes in Mathematics, **1813**, Springer Verlag, 2003, 123-160.

[3] L.Ambrosio, B.Kirchheim & A.Pratelli: *Existence of optimal transport maps for crystalline norms.* Duke Math. J., **125** (2004), no. 2, 207-241.

[4] P.Bernard & B.Buffoni: *Optimal mass transportation and Mather theory.* Preprint, 2004, to appear on the Journal of the European Mathematical Society.

[5] P.Bernard & B.Buffoni: *The Monge problem for supercritical Mañé potential on compact manifolds.* Preprint, 2005.

[6] G.Contreras, J.Delgado & R.Iturriaga: *Lagrangians flows: the dynamics of globally minimizing orbits. II.* Bol. Soc. Brasil. Mat. (N.S.), **28** (1997), no. 2, 155-196.

[7] A.Fathi: *Weak KAM theorems in Lagrangian Dynamics.* Book to appear.

[8] A.Fathi & A.Figalli: *Optimal transportation on manifolds.* Preprint.

[9] A.Fathi & E.Maderna: *Weak KAM theorem on noncompact manifolds.* Nonlin. Diff. Eq. and Appl., to appear.

[10] M.Feldman & R.J.McCann: *Monge's transport problem on a Riemannian manifold.* Trans. Amer. Math. Soc., **354** (2002), 1667-1697.

[11] L.V.Kantorovich: *On the transfer of masses.* Dokl. Akad. Nauk. SSSR, **37** (1942), 227-229.

[12] L.V.Kantorovich: *On a problem of Monge.* Uspekhi Mat.Nauk., **3** (1948), 225-226.

[13] R.Mañé: *Lagrangians flows: the dynamics of globally minimizing orbits.* Bol. Soc. Brasil. Mat. (N.S.), **28** (1997), no. 2, 141-153.

[14] S.T.Rachev & L. Ruschendorf: *Mass transportation problems.* Vol. I: Theory, Vol. II: Applications. Probability and ita applications, Springer, 1998.

[15] C.Villani: *Topics in optimal transportation.* Graduate Studies in Mathematics, **58** (2004), American Mathematical Society, Providence, RI.

[16] C.Villani: *Optimal transport, old and new.* Lecture notes, 2005 Saint-Flour summer school.