OPTIMAL TRANSPORTATION
ON NON-COMPACT MANIFOLDS

BY

ALBERT FATHI

UMPA, ENS Lyon, 46 Allée d'Italie
69007 Lyon, France
e-mail: albert.fathi@umpa.ens-lyon.fr

AND

ALESSIO FIGALLI

Scuola Normale Superiore, Piazza dei Cavalieri 7
56100 Pisa, Italy
e-mail: a.figalli@sns.it

ABSTRACT

In this work, we show how to obtain for non-compact manifolds the results that have already been done for Monge Transport Problem for costs coming from Tonelli Lagrangians on compact manifolds. In particular, the already known results for a cost of the type $d^r, r > 1$, where d is the Riemannian distance of a complete Riemannian manifold, hold without any curvature restriction.

1. Introduction

Monge transportation problem is more than 200 hundred years old; see [34]. It has generated a huge amount of work. It is impossible for us to give all the relevant references to previous works. There are now several books and surveys that can help the reader through the literature, see for example, the books [12, 4, 37, 41, 42] and the surveys [2, 18, 26]. Both authors benefited a lot from all these sources.

Originally Monge wanted to move rubble (déblais) in 3-space to build up a mound or fortification (remblais) minimizing the cost. Now if the rubble consist of masses $m_1, \ldots, m_n$ at locations $\{x_1, \ldots x_n\}$, one should move them into another set of positions $\{y_1, \ldots, y_n\}$ by minimizing the traveled distance taking into accounts the weights. Therefore, one should try to minimize

$$(1) \qquad\qquad \sum_{i=1}^{n} m_i d(x_i, T(x_i)),$$

over all bijections $T : \{x_1, \ldots, x_n\} \to \{y_1, \ldots, y_n\}$, where d is the usual Euclidean distance on 3-space.

Nowadays, one would be more interested in minimizing the energy cost rather than the traveled distance. Therefore, one would try rather to minimize

$$(2) \qquad\qquad \sum_{i=1}^{n} m_i d^2(x_i, T(x_i)).$$

Of course, one would like to generalize to continuous rather than just deal with discrete distributions of matter. Therefore Monge transportation problem is now stated in the following general form: given two probability measures μ and ν, defined on the measurable spaces X and Y, find a measurable map $T : X \to Y$ with

$$(3) \qquad\qquad T_\sharp \mu = \nu,$$

i.e.

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

and in such a way that T minimize the transportation cost. This last condition means

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

where $c : X \times Y \to \mathbb{R}$ is some given cost function, and the minimum is taken over all measurable maps $S : X \to Y$ with $S_\sharp \mu = \nu$. When condition (3) is satisfied, we say that T is a **transport map**, and if T also minimizes the cost we call it an **optimal transport map**.

It is easy to build examples where the Monge problem is ill-posed simply because there is no transport map: this happens, for instance, when μ is a Dirac mass while ν is not. This means that one needs some restrictions on the measures μ and ν.

Even in Euclidean spaces, and the cost c is equal to the Euclidean distance or its square, the problem of the existence of an optimal transport map is far from being trivial. Due to the strict convexity of the square of the Euclidean distance, case (2) above is simpler to deal with than case (1). The reader should consult the books and surveys given above to have a better view of the history of the subject, in particular Villani's second book on the subject [42]. However for the case where the cost is a distance, like in (1), one should at least cite the works of Sudakov [39], Evans-Gangbo [19], Feldman-McCann [23], Caffarelli-Feldman-McCann [11], Ambrosio-Pratelli [5], and Bernard-Buffoni [8]. For the case where the cost is the square of the Euclidean or of a Riemannian distance, like in (2), one should at least cite the works of Knott-Smith [31], Brenier [9], Rachev-Rüschendorf [36], Gangbo-McCann [27], McCann [33], and Bernard-Buffoni [7].

Our work is related to the case where the cost behaves like a square of a Riemannian distance. It is strongly inspired by the work of Bernard-Buffoni [7]. In fact, we prove the non-compact version of this last work adapting some techniques that were first used in the Euclidean case in [4] by Ambrosio,Gigli, and Savaré. We show that the Monge transport problem can be solved for the square distance on any complete Riemannian manifold *without any assumption on the compactness or curvature*, with the usual restriction on the measures. Most of the arguments in this work are well-known to specialists, at least in the compact case, but they have not yet been assembled and adapted to the case we treat. Of course, there is a strong overlap with some of the results that appear in [42]. For the case where the cost behaves like the distance of a complete non-compact Riemannian manifold see the work [24] of the second author.

We will prove a generalization of the following theorem (see Theorems 4.2 and 4.3):

THEOREM 1.1: *Suppose that M is a connected complete Riemannian manifold, whose Riemannian distance is denoted by d. Suppose that $r > 1$. If μ and ν are probability (Borel) measures on M, with μ absolutely continuous with respect to Lebesgue measure, and*

$$\int_M d^r(x, x_0) \, d\mu(x) < \infty \quad \text{and} \quad \int_M d^r(x, x_0) \, d\nu(x) < \infty$$

for some $x_0 \in M$, then we can find a transport map $T : M \to M$, with $T_\sharp \mu = \nu$,

which is optimal for the cost d^r on $M \times M$. Moreover, the map T is uniquely determined μ-almost everywhere.

We recall that a measure on a smooth manifold is absolutely continuous with respect to the Lebesgue measure if its representatives in charts are absolutely continuous with respect to Lebesgue measure. Again we note that there is no restriction on the curvature of M in the theorem above.

The paper is structured as follows: in Section 2, we recall some known results on the general theory of the optimal transport problem, and we introduce some useful definitions. Then, in Section 3, we give very general results for the existence and the uniqueness of optimal transport maps (Theorems 3.1 and 3.2, and Complement 3.4). In Section 4, the above results are applied to the case of costs functions coming from (weak) Tonelli Lagrangians (Theorems 4.2 and 4.3). In Section 5, we study the so called "dispacement interpolation," showing a countably Lipschitz regularity for the transport map starting from an intermidiate time (Theorem 5.1). Finally, in the appendix, we collect all the tecnical results about semi-concave functions and Tonelli Lagrangians used in our proofs.

2. Background and some definitions

A major advance towards solving the Monge transport problem is due to Kantorovich. He proposed in [29], [30] a notion of weak solution of the transport problem. 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.

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

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 set of plans, the new minimization problem then

becomes

$$(4) \qquad C(\mu, \nu) = \min_{\gamma \in \Pi(\mu, \nu)} \left\{ \int_{M \times M} c(x, y) \, d\gamma(x, y) \right\}.$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an **optimal plan**. Due to the linearity of the constraint $\gamma \in \Pi(\mu, \nu)$, weak topologies can be used to provide existence of solutions to (4): this happens for instance whenever X and Y are Polish spaces and c is lower semicontinuous (see [37], [41, Proposition 2.1] or [42]). We can see the connection between the formulation of Kantorovich and that of Monge if we notice that any transport map T induces the plan defined by $(\mathrm{Id}_X \tilde{\times} T)_\sharp \mu$ which is concentrated on the graph of T, where the map $\mathrm{Id}_X \tilde{\times} T : X \to X \times Y$ is defined by

$$\mathrm{Id}_X \tilde{\times} T(x) = (x, T(x)).$$

It is well-known that a linear minimization problem with convex constraints, like (4), admits a dual formulation. Before stating the duality formula, we make some definitions similar to that of the weak KAM theory (see [21]):

Definition 2.1 (c-subsolution): We say that a pair of functions $\varphi : X \to \mathbb{R} \cup \{+\infty\}$, $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ is a c-**subsolution** if

$$\psi(y) - \varphi(x) \leq c(x, y) \quad \forall (x, y) \in X \times Y.$$

Observe that when c is measurable and bounded below, and (φ, ψ) is a c-subsolution with $\varphi \in L^1(\mu), \psi \in L^1(\nu)$, then

$$\int_Y \psi \, d\nu - \int_X \varphi \, d\mu = \int_{X \times Y} (\psi(y) - \varphi(x)) \, d\gamma(x, y)$$
$$\leq \int_{X \times Y} c(x, y) \, d\gamma(x, y)$$

for all $\gamma \in \Pi(\mu, \nu)$. If moreover $\int_{X \times Y} c(x, y) \, d\gamma < +\infty$, and

$$\int_{X \times Y} (\psi(y) - \varphi(x)) \, d\gamma(x, y) = \int_{X \times Y} c(x, y) \, d\gamma(x, y),$$

then one would obtain the following equality:

$$\psi(y) - \varphi(x) = c(x, y) \quad \text{for } \gamma\text{-a.e. } (x, y)$$

(without any measurability or integrability assumptions on (φ, ψ), this is just a formal computation).

Definition 2.2 (Calibration): Given an optimal plan γ, we say that a c-sub-solution (φ, ψ) is (c, γ)-**calibrated** if φ and ψ are Borel measurable, and

$$\psi(y) - \varphi(x) = c(x, y) \quad \text{for } \gamma\text{-a.e. } (x, y).$$

THEOREM 2.3 (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 the infimum in the Kantorovitch problem (4) is finite. Then a transport plan $\gamma \in \Pi(\mu, \nu)$ is optimal if and only if there exists a (c, γ)-calibrated subsolution (φ, ψ).*

For a proof of this theorem see [38] and [42, Theorem 5.10 (ii)].

In this work we study Monge's problem on manifolds for a large class of cost functions induced by Lagrangians like in [7], where the authors consider the case of compact manifolds. We generalize their result to arbitrary non-compact manifolds.

Following the general scheme of proof, we will first prove a result on more general costs, see Theorem 3.2. In this general result, the fact that the target space for the Monge transport is a manifold is not necessary. So we will assume that only the source space (for the Monge transport map) is a manifold.

Let M be a n-dimensional manifold (Hausdorff and with a countable basis), N a Polish space, $c : M \times N \to \mathbb{R}$ a cost function, μ and ν two probability measures on M and N respectively. We want to prove existence and uniqueness of an optimal transport map $T : M \to N$, under some reasonable hypotheses on c and μ.

One of the conditions on the cost c is given in the following definition:

Definition 2.4 (Twist Condition): For a given cost function $c(x, y)$, we define the **skew left Legendre transform** as the partial map

$$\Lambda_c^l : M \times N \to T^*M,$$

$$\Lambda_c^l(x, y) = (x, \frac{\partial c}{\partial x}(x, y)),$$

and its definition on the domain is

$$\mathcal{D}(\Lambda_c^l) = \left\{ (x, y) \in M \times N : \frac{\partial c}{\partial x}(x, y) \text{ exists} \right\}.$$

Moreover, we say that c satisfies the **left twist condition** if Λ_c^l is injective on $\mathcal{D}(\Lambda_c^l)$.

If N is also a manifold, one can similarly define the **skew right Legendre transform**

$$\Lambda_c^r : M \times N \to T^* N$$

by $\Lambda_c^r(x, y) = (y, \frac{\partial c}{\partial y}(x, y))$. The map Λ_c^r is defined on the domain $\mathcal{D}(\Lambda_c^r) = \{(x, y) \in M \times N : \frac{\partial c}{\partial y}(x, y)$ exists$\}$. We say that c satisfies the **right twist condition** if Λ_c^r is injective on $\mathcal{D}(\Lambda_c^r)$.

The usefulness of these definitions will be clear in the Section 4, in which we will treat the case where $M = N$ and the cost is induced by a Lagrangian. This condition has already appeared in the subject. This fact (explicitly or not) was known to Gangbo (oral communication) and Villani (see [41, page 90]) among others. It is used in [7], since it is always satisfied for a cost coming from a Lagrangian, as we will see below. We borrow the terminology "twist condition" from the theory of Dynamical Systems: if $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto h(x, y)$ is C^2, one says that h satisfies the twist condition if there exists a constant $\alpha > 0$ such that $\frac{\partial^2 h}{\partial x \partial y} \geq \alpha$ everywhere. In that case both maps $\Lambda_h^l : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (x, \partial h / \partial x(x, y))$ and $\Lambda_h^r : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (y, \partial h / \partial y(x, y))$ are C^1 diffeomorphisms. The twist map $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ associated with h is determined by $f(x_1, v_1) = (x_2, v_2)$, where $v_1 = -\partial h / \partial x(x_1, x_2), v_2 = \partial h / \partial y(x_1, x_2)$, which means $f(x_1, v_1) = \Lambda_h^r \circ [\Lambda_h^l]^{-1}(x_1, -v_1)$, see [32] or [25].

Now we recall some useful measure-theoretical facts that we will need in the sequel.

LEMMA 2.5: *Let M be an n-dimensional manifold, N be a Polish space, and let $c : M \times N \to \mathbb{R}$ be a measurable function such that $x \mapsto c(x, y)$ is continuous for any $y \in N$. Then the set*

$$\left\{ (x, y) : \frac{\partial c}{\partial x}(x, y) \text{ exists} \right\} \quad \text{is Borel measurable.}$$

Moreover, $(x, y) \mapsto \frac{\partial c}{\partial x}(x, y)$ is a Borel function on that set.

Proof. This is a standard result in measure theory. We give here just a sketch of the proof.

By the locality of the statement, using charts we can assume $M = \mathbb{R}^n$. Let $T_k : \mathbb{R}^n \to \mathbb{R}^n$ be a dense countable family of linear maps. For any $j, k \in \mathbb{N}$, we

consider the Borel function

$$L_{j,k}(x, y) := \sup_{|h| \in (0, \frac{1}{j})} \frac{|c(x + h, y) - c(x, y) - T_k(h)|}{|h|}$$

$$= \sup_{|h| \in (0, \frac{1}{j}), h \in \mathbb{Q}^n} \frac{|c(x + h, y) - c(x, y) - T_k(h)|}{|h|},$$

where in the second equality we used the continuity of $x \mapsto c(x, y)$. Then it is not difficult to show that the set of points where $\frac{\partial c}{\partial x}(x, y)$ exists can be written as

$$\{(x, y) : \inf_j \inf_k L_{j,k}(x, y) = 0\},$$

which is clearly a Borel set.

To show that $x \mapsto \frac{\partial c}{\partial x}(x, y)$ is Borel, it suffices to note that the partial derivatives

$$\frac{\partial c}{\partial x_i}(x, y) = \lim_{\ell \to \infty} \frac{c(x_1, \ldots, x_i + 1/\ell, \ldots, x_n, y) - c(x_1, \ldots, x_i, \ldots, x_n, y)}{1/\ell}$$

are countable limits of continuous functions, and hence are Borel measurable. ■

Therefore, by the above lemma, $\mathcal{D}(\Lambda_c^l)$ is a Borel set. Moreover if we assume that c satisfies the left twist condition (that is, Λ_c^l is injective on $\mathcal{D}(\Lambda_c^l)$), then one can define

$$(\Lambda_c^l)^{-1} : T^*M \supset \Lambda_c^l(\mathcal{D}(\Lambda_c^l)) \to \mathcal{D}(\Lambda_c^l) \subset M \times N.$$

Then, by the injectivity assumption, one has that $\Lambda_c^l(\mathcal{D}(\Lambda_c^l))$ is still a Borel set, and $(\Lambda_c^l)^{-1}$ is a Borel map (see [17, Proposition 8.3.5 and Theorem 8.3.7], [22]). We can so extend $(\Lambda_c^l)^{-1}$ as a Borel map on the whole T^*M as

$$\Lambda_c^{l,inv}(x, p) = \begin{cases} (\Lambda_c^l)^{-1}(x, p) & \text{if } p \in T_x^*M \cap \Lambda_c^l(\mathcal{D}(\Lambda_c^l)), \\ (x, \bar{y}) & \text{if } p \in T_x^*M \setminus \Lambda_c^l(\mathcal{D}(\Lambda_c^l)), \end{cases}$$

where $\bar{y}$ is an arbitrary point, but fixed point, in N.

3. The main result

In order to have general results of existence and uniqueness of transport maps which are sufficiently flexible so that they can also be used in other situations, and to well show where measure-theoretic problems enter in the proof of the

existence of the transport map, we will first give a general result where no measures are present (see Appendix A.3 for the definition of locally semi-concave function and Appendix A.7 for the definition of countably $(n-1)$-Lipschitz set).

THEOREM 3.1: *Let M be a smooth (second countable) manifold, and let N be a Polish space. Assume that the cost $c : M \times N \to \mathbb{R}$ is Borel measurable, bounded from below, and satisfies the following conditions:*

(i) *the family of maps $x \mapsto c(x,y) = c_y(x)$ is locally semi-concave in x locally uniformly in y,*

(ii) *the cost c satisfies the left twist condition.*

Let (φ, ψ) be a c-subsolution, and consider the set $G_{(\varphi,\psi)} \subset M \times N$ given by

$$G_{(\varphi,\psi)} = \{(x,y) \in M \times N : \psi(y) - \varphi(x) = c(x,y)\}.$$

We can find a Borel countably $(n-1)$-Lipschitz set $E \subset M$ and a Borel measurable map $T : M \to N$ such that

$$G_{(\varphi,\psi)} \subset \mathrm{Graph}(T) \cup \pi_M^{-1}(E),$$

where $\pi_M : M \times N \to M$ is the canonical projection, and $\mathrm{Graph}(T) = \{(x, T(x)) : x \in M\}$ is the graph of T.

In other words, if we define $P = \pi_M\big(G_{(\varphi,\psi)}\big) \subset M$ the part of $G_{(\varphi,\psi)}$ which is above $P \setminus E$ is contained a Borel graph.

More precisely, we will prove that there exist an increasing sequence of locally semi-convex functions $\varphi_n : M \to \mathbb{R}$, with $\varphi \geq \varphi_{n+1} \geq \varphi_n$ on M, and an increasing sequence of Borel subsets C_n such that

- *For $x \in C_n$, the derivative $d_x\varphi_n$ exists, $\varphi_{n+1}(x) = \varphi_n(x)$ and $d_x\varphi_{n+1} = d_x\varphi_n$.*
- *If we set $C = \bigcup_n C_n$, there exists a Borel countably $(n-1)$-Lipschitz set $E \subset M$ such that $P \setminus E \subset C$.*

Moreover, the Borel map $T : M \to N$ is such that

- *For every $x \in C_n$, we have*

$$(x, T(x)) = \Lambda_c^{l,inv}(x, -d_x\varphi_n),$$

  where $\Lambda_c^{l,inv}$ is the extension of the inverse of Λ_c^l defined at the end of Section 2.

- If $x \in P \cap C_n \setminus E$, then the partial derivative $\frac{\partial c}{\partial x}(x, T(x))$ exists (i.e. $(x, T(x)) \in \mathcal{D}(\Lambda_c^l)$), and

$$\frac{\partial c}{\partial x}(x, T(x)) = -d_x \varphi_n.$$

In particular, if $x \in P \cap C_n \setminus E$, we have

$$(x, T(x)) \in \mathcal{D}(\Lambda_c^l) \quad \text{and} \quad \Lambda_c^l(x, T(x)) = (x, -d_x \varphi_n).$$

Therefore, thanks to the twist condition, the map T is uniquely defined on $P \setminus E \subset C$.

The existence and uniqueness of a transport map is then a simple consequence of the above theorem.

THEOREM 3.2: *Let M be a smooth (second countable) manifold, let N be a Polish space, and consider μ and ν (Borel) probability measures on M and N, respectively. Assume that the cost $c : M \times N \to \mathbb{R}$ is lower semicontinuous and bounded from below. Assume moreover that the following conditions hold:*

(i) *the family of maps $x \mapsto c(x, y) = c_y(x)$ is locally semi-concave in x locally uniformly in y,*
(ii) *the cost c satisfies the left twist condition,*
(iii) *the measure μ gives zero mass to countably $(n - 1)$-Lipschitz sets,*
(iv) *the infimum in the Kantorovitch problem (4) is finite.*

Then there exists a Borel map $T : M \to N$, which is an optimal transport map from μ to ν for the cost c. Morover, the map T is unique μ-a.e., and any plan $\gamma_c \in \Pi(\mu, \nu)$ optimal for the cost c is concentrated on the graph of T.

More precisely, if (φ, ψ) is a (c, γ_c)-calibrating pair, with the notation of Theorem 3.1, there exists an increasing sequence of Borel subsets B_n, with $\mu(\bigcup_n B_n) = 1$, such that the map T is uniquely defined on $B = \bigcup_n B_n$ via

$$\frac{\partial c}{\partial x}(x, T(x)) = -d_x \varphi_n \quad \text{on } B_n,$$

and any optimal plan $\gamma \in \Pi(\mu, \nu)$ is concentrated on the graph of that map T.

We remark that condition (iv) is trivially satisfied if

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

However we needed to state the above theorem in this more general form in order to apply it in Section 5 (see Remark 5.2).

Proof of Theorem 3.2. Let $\gamma_c \in \Pi(\mu, \nu)$ be an optimal plan. By Theorem 2.3 there exists a (c, γ)-calibrated pair (φ, ψ). Consider the set

$$G = G_{(\varphi, \psi)} = \{(x, y) \in M \times N : \psi(y) - \varphi(x) = c(x, y)\}.$$

Since both M and N are Polish and both maps φ and ψ are Borel, the subset G is a Borel subset of $M \times N$. Observe that, by the definition of (c, γ_c)-calibrated pair, we have $\gamma_c(G) = 1$.

By Theorem 3.1 there exists a Borel countably $(n-1)$-Lipschitz set E such that $G \setminus (\pi_M)^{-1}(E)$ is contained in the graph of a Borel map T. This implies that

$$B = \pi_M \left(G \setminus (\pi_M)^{-1}(E) \right) = \pi_M(G) \setminus E \subset M$$

is a Borel set, since it coincides with $(\mathrm{Id}_M \,\tilde{\times} T)^{-1} \left(G \setminus (\pi_M)^{-1}(E) \right)$ and the map $x \mapsto \mathrm{Id}_M \,\tilde{\times} T(x) = (x, T(x))$ is Borel measurable.

Thus, recalling that the first marginal of γ_c is μ, by assumption (iii) we get $\gamma_c((\pi_M)^{-1}(E)) = \mu(E) = 0$. Therefore $\gamma_c(G \setminus (\pi_M)^{-1}(E)) = 1$, so that γ_c is concentrated on the graph of T, which gives the existence of an optimal transport map. Note now that $\mu(B) = \gamma_c(\pi^{-1}(B)) \geq \gamma_c(G \setminus (\pi_M)^{-1}(E)) = 1$. Therefore $\mu(B) = 1$. Since $B = P \setminus E$, where $P = \pi_M(G)$, using the Borel set C_n provided by Theorem 3.1, it follows that $B_n = P \cap C_n \setminus E = D \cap C_n$ is a Borel set with $B = \bigcup_n B_n$. The end of Theorem 3.1 shows that T is indeed uniquely defined on B as said in the statement.

Let us now prove the uniqueness of the transport map μ-a.e. If S is another optimal transport map, consider the measures $\gamma_T = (\mathrm{Id}_M \times T)_{\#} \mu$ and $\gamma_S = (\mathrm{Id}_M \times S)_{\#} \mu$. The measure $\bar{\gamma} = \frac{1}{2}(\gamma_T + \gamma_S) \in \Pi(\mu, \nu)$ is still an optimal plan, and therefore must be concentrated on a graph. This implies that $S = T$ μ-a.e., and thus T is the unique optimal transport map. Finally, since any optimal $\gamma \in \Pi(\mu, \nu)$ is concentrated on a graph, we also deduce that any optimal plan is concentrated on the graph of T. ∎

Proof of Theorem 3.1. By definition of c-subsolution, we have $\varphi > -\infty$ everywhere on M, and $\psi < +\infty$ everywhere on N. Therefore, if we define $W_n = \{\psi \leq n\}$, we have $W_n \subset W_{n+1}$, and $\bigcup_n W_n = N$. Since, by hypothesis (i), $c(x, y) = c_y(x)$ is locally semi-concave in x locally uniformly in y, for each $y \in N$ there exist a neighborhood V_y of y such that the family of functions $(c(\cdot, z))_{z \in V_y}$ is locally uniformly semi-concave. Since N is separable, there exists a countable family of points $(y_k)_{k \in \mathbb{N}}$ such that $\bigcup_k V_{y_k} = N$. We now consider

the sequence of subsets $(V_n)_{n \in \mathbb{N}} \subset N$ defined as

$$V_n = W_n \cap \Big(\bigcup_{1 \leq k \leq n} V_{y_k} \Big).$$

We have $V_n \subset V_{n+1}$. Define $\varphi_n : M \to N$ by

$$\varphi_n(x) = \sup_{y \in V_n} \psi(y) - c(x, y) = \max_{1 \leq k \leq n} \Big(\sup_{y \in W_n \cap V_{y_k}} \psi(y) - c(x, y) \Big).$$

Since $\psi \leq n$ on K_n, and $-c$ is bounded from above, we see that φ_n is bounded from above. Therefore, by hypothesis (i), the family of functions $(\psi(y) - c(\cdot, y))_{y \in W_n \cap V_{y_k}}$ is locally uniformly semi-convex and bounded from above. Thus, by Theorem A.4 and Proposition A.11 in the Appendix, the function φ_n is locally semi-convex. Since $\psi(y) - \varphi(x) \leq c(x, y)$ with equality on $G_{(\varphi, \psi)}$, and $V_n \subset V_{n+1}$, we clearly have

$$\varphi_n \leq \varphi_{n+1} \leq \varphi \quad \text{everywhere on } M.$$

A key observation is now the following:

$$\varphi|_{P_n} = \varphi_n|_{P_n},$$

where $P_n = \pi_M(G_{(\varphi, \psi)} \cap (M \times V_n))$. In fact, if $x \in P_n$, by the definition of P_n we know that there exists a point $y_x \in V_n$ such that $(x, y_x) \in G_{(\varphi, \psi)}$. By the definition of $G_{(\varphi, \psi)}$, this implies

$$\varphi(x) = \psi(y_x) - c(x, y_x) \leq \varphi_n(x) \leq \varphi(x).$$

Since φ_n is locally semi-convex, by Theorem A.8 in the Appendix applied to $-\varphi_n$, it is differentiable on a Borel subset F_n such that its complement F_n^c is a Borel countably $(n-1)$-Lipschitz set. Let us then define $F = \bigcap_n F_n$. The complement $E = F^c = \bigcup_n F_n^c$ is also a Borel countably $(n-1)$-Lipschitz set. We now define the Borel set

$$C_n = F \cap \{x \in M : \varphi_k(x) = \varphi_n(x) \ \forall k \geq n\}.$$

We observe that $C_n \supset P_n \cap F$.

Now we prove that $G_{(\varphi, \psi)} \cap ((P_n \cap F) \times V_n)$ is contained in a graph.

To prove this assertion, fix $x \in P_n \cap F$. By the definition of P_n, and what was stated above, there exists $y_x \in V_n$ such that

$$\varphi(x) = \varphi_n(x) = \psi(y_x) - c(x, y_x).$$

Since $x \in F$, the map $z \mapsto \varphi_n(z) - \psi(y_x)$ is differentiable at x. Moreover, by condition (i), the map $z \mapsto -c(z, y_x) = -c_{y_x}(z)$ is locally semi-convex and, by the definition of φ_n, for every $z \in M$, we have $\varphi_n(z) - \psi(y_x) \geq -c(z, y_x)$, with equality at $z = x$. These facts taken together imply that $\frac{\partial c}{\partial x}(x, y_x)$ exists and is equal to $-d_x\varphi_n$. In fact, working in a chart around x, since $c_{y_x} = c(\cdot, y_x)$ is locally semi-concave, by Definition A.3 of a locally semi-concave function, there exists linear map l_x such that

$$c(z, y_x) \leq c(x, y_x) + l_x(z - x) + o(|z - x|),$$

for z in a neighborhood of x. Using also that φ_n is differentiable at x, we get

$$\begin{aligned}
\varphi_n(x) - \psi(y_x) &+ d_x\varphi_n(z - x) + o(|z - x|) \\
&= \varphi_n(z) - \psi(y_x) \\
&\geq -c(z, y_x) \\
&\geq -c(x, y_x) - l_c(z - x) + o(|z - x|) \\
&= \varphi_n(x) - \psi(y_x) - l_c(z - x) + o(|z - x|).
\end{aligned}$$

This implies that $l_c = -d_x\varphi_n$, and that c_{y_x} is differentiable at x with differential at x equal to $-d_x\varphi_n$. Setting now $G_x = \{y \in N : \varphi(x) - \psi(y) = c(x, y)\}$, we have just shown that $\{x\} \times (G_x \cap V_n) \subset \mathcal{D}(\Lambda_c^l)$ for each $x \in C_n$, and also $\frac{\partial c}{\partial x}(x, y) = -d_x\varphi_n$, for every $y \in G_x \cap V_n$. Recalling now that, by hypothesis (ii), the cost c satisfies the left twist condition, we obtain that $G_x \cap V_n$ is reduced to a single element which is uniquely characterized by the equality

$$\frac{\partial c}{\partial x}(x, y_x) = -d_x\varphi_n.$$

So we have proved that $G \cap (M \times V_n)$ is the graph over $P_n \cap F$ of the map T defined uniquely, thanks to the left twist condition, by

$$\frac{\partial c}{\partial x}(x, T(x)) = -d_x\varphi_n$$

(observe that, since $\varphi_n \leq \varphi_k$ for $k \geq n$ with equality on P_n, we have $d_x\varphi_n|_{P_n} = d_x\varphi_k|_{P_n}$ for $k \geq n$). Since $P_{n+1} \supset P_n$, and $V_n \subset V_{n+1} \nearrow N$, we can conclude that $G_{(\varphi, \psi)}$ is a graph over $\bigcup_n P_n \cap F = P \cap F$ (where $P = \pi_M(G_{(\varphi, \psi)}) = \bigcup_n P_n$).

Observe that, for the moment, we do not know that T is a Borel map, since P_n is not a priori Borel. Note first that by definition of $B_n \subset B_{n+1}$, we have $\varphi_n = \varphi_{n+1}$ on B_n, and they are both differentiable at every point of B_n. Since

$\varphi_n \le \varphi_{n+1}$ everywhere, by the same argument as above we get $d_x\varphi_n = d_x\varphi_{n+1}$ for $x \in B_n$. Thus, setting $B = \bigcup_n B_n$, we can extend T to M by

$$T(x) = \begin{cases} \pi_N \Lambda_c^{l,inv}(x, -d_x\varphi_n) & \text{on } B_n, \\ \bar{y} & \text{on } M \setminus B, \end{cases}$$

where $\pi_N : M \times N \to N$ is the canonical projection, $\Lambda_c^{l,inv}$ is the Borel extension of $(\Lambda_c^l)^{-1}$ defined after Lemma 2.5, and $\bar{y}$ is an arbitrary but fixed point in N. Obviously, the map T thus defined is Borel measurable and extends the map T already defined on $P \setminus E$. ∎

In the case where μ is absolutely continuous with respect to Lebesgue measure we can give a complement to our main theorem. In order to state it, we need the following definition, see [4, Definition 5.5.1, page 129]:

Definition 3.3 (Approximate differential): We say that $f : M \to \mathbb{R}$ has an **approximate differential** at $x \in M$ if there exists a function $h : M \to \mathbb{R}$ differentiable at x such that the set $\{f = h\}$ has density 1 at x with respect to the Lebesgue measure (this just means that the density is 1 in charts). In this case, the approximate value of f at x is defined as $\tilde{f}(x) = h(x)$, and the approximate differential of f at x is defined as $\tilde{d}_x f = d_x h$. It is not difficult to show that this definition makes sense. In fact, both $h(x)$ and $d_x h$ do not depend on the choice of h, provided x is a density point of the set $\{f = h\}$.

Another characterization of the approximate value $\tilde{f}(x)$ and of the approximate differential $\tilde{d}_x f$ is given, in charts, saying that the sets

$$\left\{ y : \frac{|f(y) - \tilde{f}(x) - \tilde{d}_x f(y - x)|}{|y - x|} > \varepsilon \right\}$$

have density 0 at x for each $\varepsilon > 0$ with respect to Lebesgue measure. This last definition is the one systematically used in [22]. On the other hand, for the purpose of this paper, Definition 3.3 is more convenient.

The set of points $x \in M$ where the approximate derivative $\tilde{d}_x f$ exists is measurable; moreover, the map $x \mapsto \tilde{d}_x f$ is also measurable, see [22, Theorem 3.1.4, page 214].

COMPLEMENT 3.4: *Under the hypothesis of Theorem 3.2, if we assume that μ is absolutely continuous with respect to Lebesgue measure (this is stronger than condition (iii) of Theorem 3.2), then for any calibrated pair (φ, ψ), the function*

φ is approximatively differentiable μ-a.e., and the optimal transport map T is uniquely determined μ-a.e., thanks to the twist condition, by

$$\frac{\partial c}{\partial x}(x, T(x)) = -\tilde{d}_x\varphi,$$

where $\tilde{d}_x\varphi$ is the approximate differential of φ at x. Moreover, there exists a Borel subset $A \subset M$ of full μ measure such that $\tilde{d}_x\varphi$ exists on A, the map $x \mapsto \tilde{d}_x\varphi$ is Borel measurable on A, and $\frac{\partial c}{\partial x}(x, T(x))$ exists for $x \in A$ (i.e. $(x, T(x)) \in \mathcal{D}(\Lambda_c^l)$).

Proof. We will use the notations and the proof of Theorems 3.1 and 3.2. We denote by $\tilde{A}_n \subset B_n$ the set of $x \in B_n$ which are density points for B_n with respect to some measure λ whose measure class in charts is that of Lebesgue (for example, one can take λ as the Riemannian measure associated to a Riemannian metric). By Lebesgue's Density Theorem $\lambda(B_n \setminus \tilde{A}_n) = 0$. Since μ is absolutely continuous with respect to Lebesgue measure, we have $\mu(\tilde{A}_n) = \mu(B_n)$, and therefore $\tilde{A} = \bigcup_n \tilde{A}_n$ is of full μ-measure, since $\mu(B_n) \nearrow \mu(B) = 1$. Moreover, since $\varphi = \varphi_n$ on B_n and φ_n is differentiable at each point of B_n, the function φ is approximatively differentiable at each point of $\tilde{A}_n$ with $\tilde{d}_x\varphi = d_x\varphi_n$.

The last part of this complement on measurability follows of course from [22, Theorem 3.1.4, page 214]. But in this case, we can give a direct simple proof. We choose $A_n \subset \tilde{A}_n$ Borel measurable with $\mu(\tilde{A}_n \setminus A_n) = 0$. We set $A = \bigcup_n A_n$. The set A is of full μ measure. Moreover, for every $x \in A_n$, the approximate differential $\tilde{d}_x\varphi$ exists and is equal to $d_x\varphi_n$. Thus it suffices to show that the map $x \mapsto d_x\varphi_n$ is Borel measurable, and this follows as in Lemma 2.5. ∎

Remark 3.5: We observe that the key steps in the results of this section, can be summarized in the following way. Let $\varphi : M \to \mathbb{R} \cup \{+\infty\}$ be a c-convex function, that is

$$\varphi(x) = \sup_{y \in N} \psi(y) - c(x, y)$$

for certain $\psi : N \to \mathbb{R} \cup \{-\infty\}$, and assume $\varphi \not\equiv +\infty$. Then, if c satisfies hypothesis (i) of Theorem 3.1, we can prove the following:

  (1) There exists a non-decreasing sequence φ_n of locally semi-convex functions such that $\bigcup_n \{\varphi = \varphi_n\} = M$.

From this we deduce:

(2a) There exists a Borel section $P(x)$ of the cotangent bundle T^*M and a $(n-1)$-Lipschitz set $E \subset M$ such that, if $x \notin E$, $P(x)$ is the only possible sub-differential of φ at x.

(2b) The function φ is approximately differentiable a.e. with respect to the Lebesgue measure.

By these facts, the duality theorem and the semi-concavity of the cost, we can deduce the existence and uniqueness of an optimal transport (under the general assumptions of Theorem 3.2 we use (2a), while under the assumption that the source measure is absolutely continuous with respect to the Lebesgue measure we use (2b)).

4. Costs obtained from Lagrangians

Now that we have proved Theorem 3.2, we want to observe that the hypotheses are satisfied by a large class of cost functions.

We will consider first the case of a Tonelli Lagrangian L on a connected manifold (see Definition B.4 of the Appendix for the definition of a Tonelli Lagrangian). For $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ associated to L is given by

$$c_{t,L}(x,y) = \inf_{\gamma} \int_0^t L(\gamma(s), \dot{\gamma}(s)) \, ds,$$

where the infimum is taken over all the continuous piecewise C^1 curves $\gamma :$ $[0,t] \to M$, with $\gamma(0) = x$, and $\gamma(t) = y$ (see Definition B.18 of the appendix).

PROPOSITION 4.1: *If $L : TM \to \mathbb{R}$ is a Tonelli Lagrangian on the connected manifold M, then, for $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ associated to the Lagrangian L is continuous, bounded from below, and satisfies conditions* (i) *and* (ii) *of Theorem 3.2.*

Proof. Since L is a Tonelli Lagrangian, observe that L is bounded below by C, where C is the constant given in condition (c) of Definition B.4. Hence the cost $c_{t,L}$ is bounded below by tC. By Theorem B.19 of the appendix, the cost $c_{t,L}$ is locally semi-concave, and therefore continuous. Moreover, we can now apply Proposition A.17 of the appendix to conclude that $c_{t,L}$ satisfies condition (i) of Theorem 3.2.

The twist condition (ii) of Theorem 3.2 for $c_{t,L}$ follows from Lemma B.22 and Proposition B.23. ∎

For costs coming from Tonelli Lagrangians, applying the main Theorem 3.2, and its Complement 3.4 we obtain

THEOREM 4.2: *Let L be a Tonelli Lagrangian on the connected manifold M. Fix $t > 0$, μ, ν a pair of probability measure on M, with μ giving measure zero to countably $(n-1)$-Lipschitz sets, and assume that the infimum in the Kantorovitch problem (4) with cost $c_{t,L}$ is finite. Then there exists a uniquely μ-almost everywhere defined transport map $T : M \to M$ from μ to ν which is optimal for the cost $c_{t,L}$. Moreover, any plan $\gamma \in \Pi(\mu, \nu)$, which is optimal for the cost $c_{t,L}$, verifies $\gamma(\mathrm{Graph}(T)) = 1$.*

If μ is absolutely continuous with respect to Lebesgue measure, and (φ, ψ) is a $c_{t,L}$-calibrated subsolution for (μ, ν), then we can find a Borel set B of full μ measure, such that the approximate differential $\tilde{d}_x \varphi$ of φ at x is defined for $x \in B$, the map $x \mapsto \tilde{d}_x \varphi$ is Borel measurable on B, and the transport map T is defined on B (hence μ-almost everywhere) by

$$T(x) = \pi^* \phi_t^H(x, \tilde{d}_x \varphi),$$

where $\pi^* : T^*M \to M$ is the canonical projection, and ϕ_t^H is the Hamiltonian flow of the Hamiltonian H associated to L.

We can also give the following description for T valid on B (hence μ-almost everywhere):

$$T(x) = \pi \phi_t^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)),$$

where ϕ_t^L is the Euler-Lagrange flow of L, and $x \to \widetilde{\mathrm{grad}}_x^L(\varphi)$ is the measurable vector field on M defined on B by

$$\frac{\partial L}{\partial v}(x, \widetilde{\mathrm{grad}}_x^L(\varphi)) = \tilde{d}_x \varphi.$$

Moreover, for every $x \in B$, there is a unique L-minimizer $\gamma : [0,t] \to M$, with $\gamma(0) = x, \gamma(t) = T(x)$, and the curve γ is given by $\gamma(s) = \pi \phi_s^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi))$, for $0 \le s \le t$.

Proof. The first part is a consequence of Proposition 4.1 and Theorem 3.2. When μ is absolutely continuous with respect to Lebesgue measure, we can apply Complement 3.4 to obtain a Borel subset $A \subset M$ of full μ measure such that, for every $x \in A$, we have $(x, T(x)) \in \mathcal{D}(\Lambda_{c_{t,L}}^l)$ and

$$\frac{\partial c_{t,L}}{\partial x}(x, T(x)) = -\tilde{d}_x \varphi.$$

By Lemma B.22 and Proposition B.23, if $(x, y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$, then there is a unique L-minimizer $\gamma : [0, t] \to M$, with $\gamma(0) = x, \gamma(t) = y$, and this minimizer is of the form $\gamma(s) = \pi \phi^L_s(x, v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_x M$ is uniquely determined by the equation

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, v).$$

Therefore $T(x) = \pi \phi^L_t(x, \widetilde{\mathrm{grad}}^L_x(\varphi))$, where $\widetilde{\mathrm{grad}}^L_x(\varphi)$ is uniquely determined by

$$\frac{\partial L}{\partial v}(x, \widetilde{\mathrm{grad}}^L_x(\varphi)) = -\frac{\partial c_{t,L}}{\partial x}(x, T(x)) = \tilde{d}_x \varphi,$$

which is precisely the second description of T. The first description of T follows from the second one, once we observe that

$$\mathscr{L}(x, \widetilde{\mathrm{grad}}^L_x(\varphi))) = (x, \frac{\partial L}{\partial v}(x, \widetilde{\mathrm{grad}}^L_x(\varphi)) = (x, \tilde{d}_x \varphi)$$

$$\phi^H_t = \mathscr{L} \circ \phi^L_t \circ \mathscr{L}^{-1}$$

$$\pi^* \circ \mathscr{L} = \pi,$$

where $\mathscr{L} : TM \to T^* M$ is the global Legendre Transform, see Definition B.8 of the appendix. ∎

We now turn to the proof of Theorem 1.1, which is not a consequence of Theorem 4.2 since the cost d^r with $r > 1$ does not come from a Tonelli Lagrangian for $r \neq 2$.

THEOREM 4.3: *Suppose that the connected manifold M is endowed with a Riemannian metric g which is complete. Denote by d the Riemannian distance. If $r > 1$, and μ and ν are probability (Borel) measures on M, where μ gives measure zero to countably $(n-1)$-Lipschitz sets, and*

$$\int_M d^r(x, x_0) \, d\mu(x) < \infty \quad \text{and} \quad \int_M d^r(x, x_0) \, d\nu(x) < \infty$$

for some given $x_0 \in M$, then we can find a transport map $T : M \to M$, with $T_\sharp \mu = \nu$, which is optimal for the cost d^r on $M \times M$. Moreover, the map T is uniquely determined μ-almost everywhere.

If μ is absolutely continuous with respect to Lebesgue measure, and (φ, ψ) is a calibrated subsolution for the cost $d^r(x, y)$ and the pair of measures (μ, ν), then the approximate differential $\tilde{d}_x \varphi$ of φ at x is defined μ-almost everywhere,

and the transport map T is defined μ-almost everywhere by

$$T(x) = \exp_x \left(\frac{\widetilde{\mathrm{grad}}^g_x(\varphi)}{r^{1/(r-1)} \|\widetilde{\mathrm{grad}}^g_x(\varphi)\|_x^{(r-2)/(r-1)}} \right),$$

where the approximate Riemannian gradient $\widetilde{\mathrm{grad}}^g(\varphi)$ of φ is defined by

$$g_x(\widetilde{\mathrm{grad}}^g_x(\varphi), \cdot) = \tilde{d}_x \varphi,$$

and $\exp : TM \to M$ is the exponential map of g on TM, which is globally defined since M is complete.

Proof. We first remark that

$$d^r(x,y) \leq [d(x,x_0) + d(x_0,y)]^r \leq [2 \max(d(x,x_0), d(x_0,y))]^r$$
$$\leq 2^r [d(x,x_0)^r + d(y,x_0)^r].$$

Therefore

$$\int_{M \times M} d^r(x,y)\, d\mu(x) d\nu(y) \leq \int_{M \times M} 2^r [d(x,x_0)^r + d(y,x_0)^r]\, d\mu(x) d\nu(y)$$
$$= 2^r \int_M d^r(x,x_0)\, d\mu(x) + 2^r \int_M d^r(y,x_0)\, d\nu(y)$$
$$< \infty,$$

and thus the infimum in the Kantorovitch problem (4) with cost d^r is finite.

By Example B.5, the Lagrangian $L_{r,g}(x,v) = \|v\|_x^r = g_x(v,v)^{r/2}$ is a weak Tonelli Lagrangian. By Proposition B.24, the non-negative and continuous cost $d^r(x,y)$ is precisely the cost $c_{1,L_{r,g}}$. Therefore, this cost is locally semi-concave by Theorem B.19. By Proposition A.17, this implies that $d^r(x,y)$ satisfies condition (i) of Theorem 3.2. The fact that the cost $d^r(x,y)$ satisfies the left twist condition (ii) of Theorem 3.2 follows from Proposition B.24. Therefore, there is an optimal transport map T.

If the measure μ is absolutely continuous with respect to Lebesgue measure, and (φ, ψ) is a calibrated subsolution for the cost $d^r(x,y)$ and the pair of measures (μ, ν), then by Complement 3.4, for μ-almost every x, we have $(x, T(x)) \in \mathcal{D}(\Lambda^l_{c_{1,L_{r,g}}})$, and

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x, T(x)) = -\tilde{d}_x \varphi.$$

Since $(x, T(x))$ is in $\mathcal{D}(\Lambda^l_{c_{1,L_{r,g}}})$, it follows from Proposition B.24 that $T(x) = \pi \phi_1^g(x, v_x)$, where $\pi : TM \to M$ is the canonical projection, the flow ϕ_t^g is the

geodesic flow of g on TM, and $v_x \in T_x M$ is determined by

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x, T(x)) = -\frac{\partial L_{r,g}}{\partial v}(x, v_x),$$

or, given the equality above, by

$$\frac{\partial L_{r,g}}{\partial v}(x, v_x) = \tilde{d}_x \varphi.$$

Now the vertical derivative of $L_{r,g}$ is computed in Example B.5

$$\frac{\partial L_{r,g}}{\partial v}(x, v) = r\|v\|_x^{r-2} g_x(v, \cdot).$$

Hence $v_x \in T_x M$ is determined by

$$r\|v_x\|_x^{r-2} g_x(v_x, \cdot) = \tilde{d}_x \varphi = g_x(\widetilde{\mathrm{grad}}_x^g(\varphi), \cdot).$$

This gives the equality

$$r\|v_x\|_x^{r-2} v_x = \widetilde{\mathrm{grad}}_x^g(\varphi),$$

from which we easily get

$$v_x = \frac{\widetilde{\mathrm{grad}}_x^g(\varphi)}{r^{1/(r-1)}\|\widetilde{\mathrm{grad}}_x^g(\varphi)\|_x^{(r-2)/(r-1)}}.$$

Therefore

$$T(x) = \pi\phi_t^g(x, \frac{\widetilde{\mathrm{grad}}_x^g(\varphi)}{r^{1/(r-1)}\|\widetilde{\mathrm{grad}}_x^g(\varphi)\|_x^{(r-2)/(r-1)}}).$$

By definition of the exponential map $\exp : TM \to M$, we have $\exp_x(v) = \pi\phi_t^g(x, v)$, and the formula for $T(x)$ follows. ∎

5. The interpolation and its absolute continuity

For a cost $c_{t,L}$ coming from a Tonelli Lagrangian L, Theorem 4.2 shows not only that we have an optimal transport map T but also that this map is obtained by following an extremal for time t. We can therefore interpolate the optimal transport by maps T_s where we stop at intermediary times $s \in [0, t]$. We will show in this section that these maps are also optimal transport maps for costs coming from the same Lagrangian. Let us give now precise definitions.

For the rest of this section, we consider L a Tonelli Lagrangian on the connected manifold M. We fix $t > 0$ and μ_0 and μ_t two probability measures on

M, with μ_0 absolutely continuous with respect to Lebesgue measure, and such that

$$\min_{\gamma \in \Pi(\mu_0, \mu_t)} \left\{ \int_{M \times M} c_{t,L}(x,y) \, d\gamma(x,y) \right\} < +\infty.$$

We call T_t the optimal transport map given by Theorem 4.2 for $(c_{t,L}, \mu_0, \mu_t)$. We denote (φ, ψ) a fixed $(c_{t,L}, \gamma_t)$-calibrated pair. Therefore $\gamma_t = (\mathrm{Id}_M \, \tilde{\times} T_t)_\# \mu_0$ is the unique optimal plan from μ_0 to μ_t. By Theorem 4.2, we can find a Borel subset $B \subset M$ such that:

- the subset B is of full μ_0 measure;
- the approximate differential $\tilde{d}_x \varphi$ exists for every $x \in B$, and is Borel measurable on B;
- the map T_t is defined at every $x \in B$, and we have

$$T_t(x) = \pi \phi_t^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)),$$

  where ϕ_L^t is the Euler-Lagrange flow of L, $\pi : TM \to M$ is the canonical projection, and the Lagrangian approximate gradient $x \mapsto \widetilde{\mathrm{grad}}_x^L(\varphi))$ is defined by

$$\frac{\partial L}{\partial v}(x, \widetilde{\mathrm{grad}}_x^L(\varphi)) = \tilde{d}_x \varphi;$$

- for every $x \in B$, the partial derivative $\frac{\partial c}{\partial x}(x, T_t(x))$ exists, and is uniquely defined by

$$\frac{\partial c}{\partial x}(x, T_t(x)) = -\tilde{d}_x \varphi;$$

- for every $x \in B$ there exists a unique L-minimizer $\gamma_x : [0, t] \to M$ between x and $T_t(x)$. This L-minimizer γ_x is given by

$$\gamma_x(s) = \pi \phi_s^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)) \quad \forall s \in [0, t];$$

- for every $x \in B$, we have

$$\psi(T_t(x)) - \varphi(x) = c_{t,L}(x, T_t(x)).$$

For $s \in [0, t]$, we can therefore define an interpolation $T_s : M \to M$ defined on B, and hence μ_0 almost everywhere, by

$$T_s(x) = \gamma_x(s) = \pi \phi_s^L(x, \widetilde{\mathrm{grad}}_x^L(\varphi)).$$

Each map T_s is Borel measurable. In fact, since the global Legendre transform is a homeomorphism and the approximate differential is Borel measurable, the

Lagrangian approximate gradient $\widetilde{\mathrm{grad}}^L(\varphi)$ is itself Borel measurable. More-over, the map $\pi\phi_s^L : TM \to M$ is continuous, and thus T_s is Borel measurable. We can therefore define the probability measure $\mu_s = T_{s\#}\mu_0$ on M, i.e., the measure μ_s is the image of μ_0 under the Borel measurable map T_s.

THEOREM 5.1: *Under the hypotheses above, the maps T_s satisfy the following properties:*

   (i) *For every $s \in]0, t[$, the map T_s is the (unique) optimal transport map for the cost $c_{s,L}$ and the pair of measures (μ_0, μ_s).*
   (ii) *For every $s \in]0, t[$, the map $T_s : B \to M$ is injective. Moreover, if we define $\bar{c}_{s,L}(x, y) = c_{s,L}(y, x)$, the inverse map $T_s^{-1} : T_s(B) \to B$ is the (unique) optimal transport map for the cost $\bar{c}_{s,L}$ and the pair of measures (μ_s, μ_0), and it is countably Lipschitz (i.e., there exist a Borel countable partition of M such that T_s^{-1} is Lipschitz on each set).*
   (iii) *For every $s \in]0, t[$, the measure $\mu_s = T_{s\#}\mu_0$ is absolutely continuous with respect to Lebesgue measure.*
   (iv) *For every $s \in]0, t[$, the composition $\hat{T}_s = T_t T_s^{-1}$ is the (unique) optimal transport map for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t), and it is countably Lipschitz.*

Proof. Fix $s \in]0, t[$. It is not difficult to see, from the definition of $c_{t,L}$, that

$$(1) \qquad c_{t,L}(x, y) \leq c_{s,L}(x, y) + c_{t-s,L}(y, z), \quad \forall x, y, z \in M,$$

and even that

$$c_{t,l}(x, z) = \inf_{y \in M} c_{s,L}(x, y) + c_{t-s,L}(y, z) \quad \forall x, z \in M.$$

If $\gamma : [a, b] \to M$ is an L-minimizer, the restriction $\gamma|_{[c,d]}$ to a subinterval $[c, d] \subset [a, b]$ is also an L-minimizer. In particular, we obtain

$$c_{b-a,L}(\gamma(a), \gamma(b)) = c_{s-a,L}(\gamma(a), \gamma(s)) + c_{b-s,L}(\gamma(s), \gamma(b)) \quad \forall s \in]a, b[.$$

Applying this to the L-minimizer γ_x, we get

$$(2) \qquad c_{t,L}(x, T_t(x)) = c_{s,L}(x, T_s(x)) + c_{t-s,L}(T_s(x), T_t(x)) \quad \forall x \in B.$$

We define for every $s \in]0, t[$, two probability measures $\gamma_s, \hat{\gamma}_s$ on $M \times M$, by

$$\gamma_s = (\mathrm{Id}_M \,\tilde{\times}\, T_s)_{\#}\mu_0 \quad \text{and} \quad \hat{\gamma}_s = (T_s \,\tilde{\times}\, T_t)_{\#}\mu_0,$$

where $\mathrm{Id}_M \tilde{\times} T_s$ and $T_s \tilde{\times} T_t$ are the maps from the subset B of full μ_0 measure to $M \times M$ defined by

$$(\mathrm{Id}_M \tilde{\times} T_s)(x) = (x, T_s(x)), \quad (T_s \tilde{\times} T_t)(x) = (T_s(x), T_t(x)).$$

Note that the marginals of γ_s are (μ_0, μ_s), and those of $\hat{\gamma}_s$ are (μ_s, μ_t). We claim that $c_{s,L}(x, y)$ is integrable for γ_s and $\hat{\gamma}_{t-s}$. In fact, we have $C = \inf_{TM} L > -\infty$, hence $c_{r,L} \geq Cr$. Therefore, equality (2) gives

$$[c_{t,L}(x, T_t(x)) - Ct] = [c_{s,L}(x, T_s(x)) - Cs] + [c_{t-s,L}(T_s(x), T_t(x)) - C(t-s)]$$
$$\forall x \in B.$$

Since the functions between brackets are all non-negative, we can integrate this equality with respect to μ_0 to obtain

$$\int_{M \times M} [c_{t,L}(x, y) - Ct] \, d\gamma_t$$
$$= \int_{M \times M} [c_{s,L}(x, y) - Cs] \, d\gamma_s + \int_{M \times M} [c_{t-s,L}(x, y) - C(t-s)] \, d\hat{\gamma}_s.$$

But all numbers involved in the equality above are non-negative, all measures are probability measures, and the cost $c_{t,L}$ is γ_t integrable since γ_t is an optimal plan for $(c_{t,L}, \mu_0, \mu_t)$, and the optimal cost of $(c_{t,L}, \mu_0, \mu_t)$ is finite. Therefore, we obtain that $c_{s,L}$ is γ_s-integrable, and $c_{t-s,L}$ is $\hat{\gamma}_s$-integrable.

Since by definition of a calibrating pair we have $\varphi > -\infty$ and $\psi < +\infty$ everywhere on M, we can find an increasing sequence of compact subsets $K_n \subset M$ with $\bigcup_n K_n = M$, and we consider $V_n = K_n \cap \{\varphi \geq -n\}$, $V'_n = K_n \cap \{\psi \leq n\}$, so that $\bigcup_n V_n = \bigcup_n V'_n = M$.

We define the functions $\varphi_s^n, \psi_s^n : M \to \mathbb{R}$ by

$$\psi_s^n(z) = \inf_{\tilde{z} \in V_n} \varphi(\tilde{z}) + c_{s,L}(\tilde{z}, z), \quad \varphi_s^n(z) = \sup_{\tilde{z} \in V'_n} \psi(\tilde{z}) - c_{t-s,L}(z, \tilde{z}),$$

where (φ, ψ) is the fixed $c_{t,L}$-calibrated pair. Note that ψ_s^n is bounded from below by $-n + t \inf_{TM} L > -\infty$. Moreover, the family of functions $(\varphi(\tilde{z}) + c_{s,L}(\tilde{z}, \cdot))_{\tilde{z} \in V'_n}$ is locally uniformly semi-concave with a linear modulus, since this is the case for the family of functions $(c_{s,L}(\tilde{z}, \cdot))_{\tilde{z} \in V'_n}$, by Theorem B.19 and Proposition A.17. It follows from Proposition A.16 that ψ_s^n is semi-concave with a linear modulus. A similar argument proves that $-\varphi_s^n$ is semi-concave with a linear modulus. Note also that, since V_n and V'_n are both increasing

sequences, we have $\psi_s^n \geq \psi_s^{n+1}$ and $\varphi_s^{n+1} \geq \varphi_s^n$, for every n. Therefore, we can define φ_s (resp., ψ_s) as the pointwise limit of the sequence φ_s^n

Using the fact that (φ, ψ) is a $c_{t,L}$-subsolution, and inequality (1) above, we obtain

$$\psi(y) - c_{t-s,L}(z,y) \leq \varphi(x) + c_{s,L}(x,z) \quad \forall x,y,z \in M.$$

Therefore we obtain for $x \in V_n, y \in V_n', z \in M$

(3) $\psi(y) - c_{t-s,L}(z,y) \leq \varphi_s^n(z) \leq \varphi_s(z) \leq \psi_s(z) \leq \psi_s^n(z) \leq \varphi(x) + c_{s,L}(x,z).$

Inequality (3) above yields

(4) $\psi(y) - c_{t-s,L}(z,y) \leq \varphi_s(z) \leq \psi_s(z) \leq \varphi(x) + c_{s,L}(x,z) \quad \forall x,y,z \in M.$

In particular, the pair (φ, ψ_s) is a $c_{s,L}$-subsolution, and the pair (φ_s, ψ) is a $c_{t-s,L}$-subsolution. Moreover, $\varphi, \psi_s, \varphi_s$ and ψ are all Borel measurable.

We now define

$$B_n = B \cap V_n \cap T_t^{-1}(V_n'),$$

so that $\bigcup_n B_n = B$ has full μ_0-measure.

If $x \in B_n$, it satisfies $x \in V_n$ and $T_t(x) \in V_n'$. From Inequality (3) above, we obtain

$$\psi(T_t(x)) - c_{t-s,L}(T_s(x), T_t(x)) \leq \varphi_s^n(T_s(x)) \leq \varphi_s(T_s(x)) \leq \psi_s(T_s(x))$$
$$\leq \psi_s^n(T_s(x)) \leq \varphi(x) + c_{s,L}(x, T_s(x))$$

Since $B_n \subset B$, for $x \in B_n$, we have $\psi(T_t(x)) - \varphi(x) = c_{t,L}(x, T_t(x))$. Combining this with equality (2), we conclude that the two extreme terms in the inequality above are equal. Hence, for every $x \in B_n$, we have

$$\psi(T_t(x)) - c_{t-s,L}(T_s(x), T_t(x)) = \varphi_s^n(T_s(x)) = \varphi_s(T_s(x)) = \psi_s(T_s(x))$$
(5) $$= \psi_s^n(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)).$$

In particular, we get

$$\psi_s(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)) \quad \forall x \in B,$$

or equivalently

$$\psi_s(y) - \varphi(x) = c_{s,L}(x,y) \quad \text{for } \gamma_s\text{-a.e. } (x,y).$$

Since (φ, ψ_s) is a (Borel) $c_{s,L}$-subsolution, it follows that the pair (φ, ψ_s) is $(c_{s,L}, \gamma_s)$-calibrated. Therefore, by Theorem 2.3 we get that $\gamma_s = (\mathrm{Id}_M \tilde{\times} T_s)_{\#}\mu_0$ is an optimal plan for $(c_{s,L}, \mu_0, \mu_s)$. Moreover, since $c_{s,L}$ is γ_s-integrable, the

infimum in the Kantorovitch problem (4) in Theorem 3.2 with cost $c_{s,L}$ is finite, and therefore there exists a unique optimal transport plan. This proves (i).

Note for further reference that a similar argument, using the equality

$$\psi(T_t(x)) = \varphi_s(T_s(x)) + c_{t-s,L}(T_s(x), T_t(x)), \quad \forall x \in B,$$

which follows from equation (5) above, shows that the measure $\hat{\gamma}_s = (T_s \tilde{\times} T_t)_{\#}\mu_0$ is an optimal plan for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t).

We now want to prove (ii). Since B is the increasing union of $B_n = B \cap V_n \cap T_t^{-1}(V_n')$, it suffices to prove that T_s is injective on B_n and that the restriction $T_s^{-1}|_{T_s(B_n)}$ is locally Lipschitz on $T_s(B_n)$.

Since $B_n \subset V_n$, by Inequality (3) above we have

$$(6) \qquad \varphi_s^n(y) \leq \psi_s^n(y) \leq \varphi(x) + c_{s,L}(x,y) \quad \forall x \in B_n, \forall y \in M.$$

Moreover, by Equality (5) above

$$(7) \qquad \varphi_s^n(T_s(x)) = \psi_s^n(T_s(x)) = \varphi(x) + c_{s,L}(x, T_s(x)) \quad \forall x \in B_n.$$

In particular, we have $\varphi_s^n \leq \psi_s^n$ everywhere with equality at every point of $T_s(B_n)$. As we have said above, both functions ψ_s^n and $-\varphi_s^n$ are locally semi-concave with a linear modulus. It follows, from Theorem A.19, that both derivatives $d_z\varphi_s^n, d_z\psi_s^n$ exist and are equal for $z \in T_s(B_n)$. Moreover, the map $z \mapsto d_z\varphi_s^n = d_z\psi_s^n$ is locally Lipschitz on $T_s(B_n)$. Note that we also get from (6) and (7) above that for a fixed $x \in B_n$, we have $\varphi_s^n \leq \varphi(x) + c_{s,L}(x, \cdot)$ everywhere with equality at $T_s(x)$. Since φ_n is semi-convex, using that $c_{s,L}(x, \cdot)$ is semi-concave, again by Theorem A.19, we obtain that the partial derivative $\frac{\partial c_{s,L}}{\partial y}(x, T_s(x))$ of $c_{s,L}$ with respect to the second variable exists and is equal to $d_{T_s(x)}\varphi_s^n = d_{T_s(x)}\psi_s^n$. Since $\gamma_x : [0,t] \to M$ is an L-minimizer with $\gamma_x(0) = x$ and $\gamma_x(s) = T_s(x)$, it follows from Corollary B.20 that

$$d_{T_s(x)}\psi_s^n = \frac{\partial c_{s,L}}{\partial y}(x, T_s(x)) = \frac{\partial L}{\partial v}(\gamma_x(s), \dot{\gamma}_x(s)).$$

Since γ_x is an L-minimizer, its speed curve is an orbit of the Euler-Lagrange flow, and therefore

$$(T_s(x), d_{T_s(x)}\psi_s^n) = \mathscr{L}((\gamma_x(s), \dot{\gamma}_x(s))) = \mathscr{L}\phi_s^L(\gamma_x(0), \dot{\gamma}_x(0))$$

and

$$x = \pi\phi_{-s}^L\mathscr{L}^{-1}(T_s(x), d_{T_s(x)}\psi_s^n).$$

It follows that T_s is injective on B_n with inverse given by the map $\theta_n : T_s(B_n) \to B_n$ defined, for $z \in T_s(B_n)$, by

$$\theta_n(z) = \pi\phi_{-s}^L \mathscr{L}^{-1}(z, d_z\psi_s^n).$$

Note that the map θ_n is locally Lipschitz on $T_s(B_n)$, since this is the case for $z \mapsto d_z\psi_s^n$, and both maps $\phi_{-s}^L, \mathscr{L}^{-1}$ are C^1, since L is a Tonelli Lagrangian. An analogous argument proves the countably Lipshitz regularity of $\hat{T}_s = T_t T_s^{-1}$ in part (iv). Finally the optimality of T_s^{-1} simply follows from

$$\min_{\gamma \in \Pi(\mu_s, \mu_0)} \left\{ \int_{M \times M} \bar{c}_{s,L}(x, y) \, d\gamma(x, y) \right\} = \min_{\gamma \in \Pi(\mu_0, \mu_s)} \left\{ \int_{M \times M} c_{s,L}(x, y) \, d\gamma(x, y) \right\}$$
$$= \int_M c_{s,L}(x, T_s(x)) \, d\mu_0(x)$$
$$= \int_M \bar{c}_{s,L}(y, T_s^{-1}(y)) \, d\mu_s(y).$$

Part (iii) of the theorem follows from part (ii). In fact, if $A \subset M$ is Lebesgue negligible, the image $T_s^{-1}(T_s(B) \cap A)$ is also Lebesgue negligible, since T_s^{-1} is countably Lipschitz on $T_s(B)$, and therefore sends Lebesgue negligible subsets to Lebesgue negligible subsets. It remains to note, using that B is of full μ_0-measure, that $\mu_s(A) = T_{s\#}\mu_0(A) = \mu_0(T_s^{-1}(T_s(B) \cap A)) = 0$.

To prove part (iv), we already know that $\hat{\gamma}_s = (T_s \tilde{\times} T_t)_\# \mu_0$ is an optimal plan for the cost $c_{t-s,L}$ and the measures (μ_s, μ_t). Since the Borel set B is of full μ_0-measure, and $T_s : B \to T_s(B)$ is a bijective Borel measurable map, we obtain that T_s^{-1} is a Borel map, and $\mu_0 = T_{s\#}^{-1}\mu_s$. It follows that

$$\hat{\gamma}_s = (\mathrm{Id}_M \tilde{\times} T_t T_s^{-1})_\# \mu_s.$$

Therefore, the composition $T_t T_s^{-1}$ is an optimal transport map for the cost $c_{t-s,L}$ and the pair of measures (μ_s, μ_t), and it is the only one since $c_{t-s,L}$ is $\hat{\gamma}_s$-integrable and μ_s is absolutely continuous with respect to the Lebesgue measure. $\blacksquare$

Remark 5.2: We observe that, in proving the uniqueness statement in parts (i) and (iv) of the above theorem, we needed the full generality of Theorem 4.2. Indeed, assuming

$$\int_{M \times M} c_{t,L}(x, y) \, d\mu_0(x)d\mu_t(y) < +\infty,$$

there is a priori no reason for which the two integrals

$$\int_{M \times M} c_{s,L}(x,y)\, d\mu_0(x)d\mu_s(y), \quad \int_{M \times M} c_{t-s,L}(x,y)\, d\mu_s(x)d\mu_t(y)$$

would have to be finite. So the existence and uniqueness of a transport map in Theorem 3.2 under the integrability assumption on c with respect to $\mu \otimes \nu$ instead of assumption (iv) would not have been enough to obtain Theorem 5.1.

Remark 5.3: We remark that, if both μ_0 and μ_t are not assumed to be absolutely continuous, and therefore no optimal transport map necessarily exists, one can still define an "optimal" interpolation $(\mu_s)_{0 \le s \le t}$ between μ_0 and μ_t using some measurable selection theorem; see [42, Chapter 7]. Then, adapting our proof, one still obtains that, for any $s \in (0,t)$, there exists a unique optimal transport map S_s for $(\bar{c}_{s,L}, \mu_s, \mu_0)$ (resp., a unique optimal transport map $\hat{S}_s$ for $(c_{t-s,L}, \mu_s, \mu_t)$), and this map is countably Lipschitz.

We also observe that, if the manifold is compact, our proof shows that the above maps are globally Lipschitz; see [7].

Appendix A. Semi-concave functions

We give the definition of semi-concave function and recall their principal properties. The main reference on semi-concave functions is the book [13].

We first recall the definition of a modulus (of continuity).

Definition A.1 (Modulus): A **modulus** ω is a continuous non-decreasing function $\omega : [0, +\infty) \to [0, +\infty)$ such that $\omega(0) = 0$.

We will say that a modulus is **linear** if it is of the form $\omega(t) = kt$, where $k \ge 0$ is some fixed constant.

We will need the notion of superdifferential. We define it in an intrinsic way on a manifold.

Definition A.2 (Superdifferential): Let $f : M \to \mathbb{R}$ be a function. We say that $p \in T_x^* M$ is a **superdifferential** of f at $x \in M$, and we write $p \in D^+ f(x)$, if there exists a function $g : V \to \mathbb{R}$, defined on some open subset $U \subset M$ containing x, such that $g \ge f$, $g(x) = f(x)$, and g is differentiable at x with $d_x g = p$.

We now give the definition of a semi-concave function on an open subset of a Euclidean space.

Definition A.3 (Semi-concavity): Let $U \subset \mathbb{R}^n$ open. A function $f : U \to \mathbb{R}$ is said to be **semi-concave** in U with modulus ω (equivalently ω-**semi-concave**) if, for each $x \in U$, we have

$$f(y) - f(x) \leq \langle l_x, y - x \rangle + \|y - x\| \omega(\|y - x\|)$$

for a certain linear form $l_x : \mathbb{R}^n \to \mathbb{R}$.

Note that necessarily $l_x \in D^+ f(x)$. Moreover we say that $f : U \to \mathbb{R}$ is **locally semi-concave** if, for each $x \in U$, there exists an open neighborhood of x in which f is semi-concave for a certain modulus.

We will say that the function $f : U \to \mathbb{R}$ is locally semi-concave with a linear modulus if, for each $x \in U$, we can find an open neighborhood V_x such that the restriction $f|_{V_x}$ is ω-semi-concave, with ω a linear modulus.

PROPOSITION A.4: 1) *Suppose* $f_i : U \to \mathbb{R}, i = 1, \ldots, k$ *is* ω_i-*semi-concave, where* U *is an open subset of* $\mathbb{R}^n$. *Then we have:*

   (i) *for any* $\alpha_1, \ldots, \alpha_k \geq 0$, *the functions* $\sum_{i=1}^{k} \alpha_i f_i$ *is* $(\sum_{i=1}^{k} \alpha_i \omega_i)$-*semi-concave on* U.
   (ii) *the function* $\min_{i=1}^{k} f_i$ *is* $(\max_{i=1}^{k} \omega_i)$-*semi-concave.*

2) *Any* C^1 *function is locally semi-concave.*

Proof. The proof of 1)(i) is obvious. For the proof of (ii), we fix $x \in U$ and find $i_0 \in \{1, \ldots, k\}$ such that $\min_{i=1}^{k} f_i(x) = f_{i_0}(x)$. Since f_{i_0} is ω_{i_0}-semi-concave, we can find a linear map $l_x : \mathbb{R}^n \to \mathbb{R}$ such that

$$f_{i_0}(y) - f_{i_0}(x) \leq l_x(y - x) + \|y - x\| \omega_{i_0}(\|y - x\|) \quad \forall y \in U.$$

It clearly follows that

$$\min_{i=1}^{k} f_i(y) - \min_{i=1}^{k} f_i(x) \leq l_x(y - x) + \|y - x\| \max_{i=1}^{k} \omega_i(\|y - x\|) \quad \forall y \in U.$$

To prove 2), consider an open convex subset C with $\bar{C}$ compact and contained in U. By compactness of $\bar{C}$ and continuity of $x \mapsto d_x f$, we can find a modulus ω, which is a modulus of continuity for the map $x \mapsto d_x f$ on C. The Mean Value Formula in integral form

$$f(y) - f(x) = \int_0^1 d_{tx+(1-t)y} f(y - x) \, dt,$$

which is valid for every $y, x \in C$ implies that

$$f(y) - f(x) \leq d_x f(y - x) + \|y - x\|\omega(\|y - x\|) \quad \forall x, y \in U.$$

Therefore f is ω-semi-concave in the open subset C. ∎

We now state and prove the first important consequences of the definition of semi-concavity.

LEMMA A.5: *Suppose U is an open subset of $\mathbb{R}^n$. Let $f : U \to \mathbb{R}$ be an ω-semi-concave function. Then we have:*

(i) *for every compact subset $K \subset U$, we can find a constant A such that for every $x \in K$, and every linear form l_x on $\mathbb{R}^n$ satisfying*

$$\forall y \in U, \quad f(y) - f(x) \leq \langle l_x, y - x \rangle + \|y - x\|\omega(\|y - x\|),$$

we have $\|l_x\| \leq A$;

(ii) *the function f is locally Lipschitz.*

Proof. From the definition, it follows that a semi-concave function is locally bounded from above. We now show that f is also locally bounded from below. Fix a (compact) cube C contained in U and let $\{y_1, \ldots, y_{2^n}\}$ be the vertices of the cube. Then, for each $x \in C$, we can write $x = \sum_i \alpha_i y_i$, with $\sum_i \alpha_i = 1$. By the semi-concavity of f we have, for each $i = 1, \ldots, 2^n$,

$$f(y_i) - f(x) \leq \langle l_x, y_i - x \rangle + \|y_i - x\|\omega(\|y_i - x\|);$$

multiplying by α_i and summing over i, we get

$$\sum_i \alpha_i f(y_i) \leq f(x) + \sum_i \alpha_i \|y_i - x\|\omega(\|y_i - x\|) \leq f(x) + B,$$

with $B = D_C \omega(D_C)$, where D_C is the diameter of the compact cube C. It follows that

$$f(x) \geq \min_i f(y_i) - B \quad \forall x \in C.$$

We now know that f is locally bounded. Using this fact, it is not difficult to show (i). In fact, suppose that the closed ball $\bar{B}(x_0, 2r), r < +\infty$, is contained in U. For $x \in \bar{B}(x_0, r)$, we have $x - rv \in \bar{B}(x_0, 2r) \subset U$ for each $v \in \mathbb{R}^n$ with $\|v\| = 1$, and therefore

$$f(x - rv) - f(x) \leq \langle l_x, -rv \rangle + \|-rv\|\omega(\|-rv\|) = -r\langle l_x, v \rangle + r\omega(r).$$

Since, by the compactness of $\bar{B}(x_0, 2r)$, we already know that

$$\tilde{B} = \sup_{z \in \bar{B}(x_0, 2r)} |f(z)|$$

is finite, we have

$$\langle l_x, v \rangle \leq \frac{f(x) - f(x - rv)}{r} + \omega(r) \leq \frac{2\tilde{B}}{r} + \omega(r).$$

It follows that, for $x \in \bar{B}(x_0, r)$,

$$\|l_x\| \leq \frac{2\tilde{B}}{r} + \omega(r).$$

Since the compact set $K \subset U$ can be covered by a finite numbers of balls $\bar{B}(x_i, r_i)$, $i = 1, \ldots, \ell$, we obtain (i).

To prove (ii), we consider a compact subset $K \subset U$ and apply (i) to obtain the constant A. We denote by D_K the (finite) diameter of the compact set K. For each $x, y \in K$,

$$f(y) - f(x) \leq \langle l_x, y - x \rangle + \|y - x\| \omega(\|y - x\|) \leq (\|l_x\| + \omega(D_K)) \|y - x\|$$
$$\leq (A + \omega(D_K)) \|y - x\|.$$

Exchanging the role of x and y, we conclude that f is Lipschitz on K. ∎

Let us recall that a Lipschitz real valued function defined on an open subset of a Euclidean space is differentiable almost everywhere (with respect to the Lebesgue measure). Therefore, by part (ii) of Lemma A.5 above we obtain the following corollary:

COROLLARY A.6: *A locally semi-concave real valued function defined on an open subset of a Euclidean space is differentiable almost everywhere with respect to the Lebesgue measure.*

In fact, in the case of semi-concave functions there is a better result which is given in Theorem A.8 below, whose proof can be found in [13, Section 4.1]. Let us first give a definition:

Definition A.7: We say that $E \subset \mathbb{R}^n$ is **countably** $(n-1)$**-Lipschitz** if there exists a countable family of compact subsets K_j is a compact subset of $\mathbb{R}^n$ such that:

    (i) E is contained in $\bigcup_j K_j$;

(ii) for each j there exists a hyperplane $H_j \subset \mathbb{R}^n = H_j \oplus H_j^\perp$, where $H_j^\perp$ is the Euclidean orthogonal of H_j, such that K_j is contained in the graph of a Lipschitz function $f_j : A_j \to H_j^\perp$ defined on a compact subset $A_j \subset H_j$.

Note that in the definition above, by the graph property (ii), the compact subset K_j has finite $(n-1)$-dimensional Hausdorff measure. Therefore, any $(n-1)$-Lipschitz set is contained in a Borel (in fact σ-compact) $(n-1)$-Lipschitz set with σ-finite $(n-1)$-dimensional Hausdorff measure.

THEOREM A.8: *If $\varphi : U \to \mathbb{R}$ is a semi-concave function defined on the open subset U of $\mathbb{R}^n$, then φ is differentiable at each point in the complement of a Borel countably $(n-1)$-Lipschitz set.*

In order to extend the definition of locally semi-concave to functions defined on a manifold, it suffices to show that this definition is stable by composition with diffeomorphisms.

LEMMA A.9: *Let $U, V \subset \mathbb{R}^n$ be open subsets. Suppose that $F : V \to U$ is a C^1 map. If $f : U \to \mathbb{R}$ is a locally semi-concave function then $f \circ F : V \to \mathbb{R}$ is also locally semi-concave. Moreover, if F is of class C^2, and $f : U \to \mathbb{R}$ is a locally semi-concave function with a linear modulus, then $f \circ F : V \to \mathbb{R}$ is also locally semi-concave with a linear modulus.*

Proof. Since the nature of the result is local, without loss of generality we can assume that $f : U \to \mathbb{R}$ is semi-concave with modulus ω. We now show that, for every V' convex open subset whose closure $\bar{V}'$ is compact and contained in V, the restriction $f \circ F|_{V'} : V' \to \mathbb{R}$ is a semi-concave function. We set $C_{\bar{V}'} = \max_{z \in \bar{V}'} \|D_z F\|$ and denote by $\hat{\omega}_{\bar{V}'}$ a modulus of continuity for the continuous function $z \mapsto D_z F$ on the compact subset $\bar{V}'$.

For each x, y in the compact convex subset $\bar{V}' \subset V$, we have

$$
\begin{aligned}
f(F(y)) - f(F(x)) \leq &\langle l_{F(x)}, F(y) - F(x) \rangle + \|F(y) - F(x)\| \omega(\|F(y) - F(x)\|) \\
\leq &\langle l_{F(x)}, DF(x)(y - x) \rangle + \|l_{F(x)}\| \hat{\omega}_{\bar{V}'}(\|y - x\|) \|y - x\| \\
&+ C_{\bar{V}'} \|y - x\| \omega(C_{\bar{V}'} \|y - x\|);
\end{aligned}
$$

Since $F(\bar{V}')$ is a compact subset of U we can apply part (i) of Lemma A.5 to obtain that $\tilde{C}_{\bar{V}'} = \sup_{\bar{V}'} \|l_{F(x)}\|$ is finite. This implies that $f \circ F$ on V' is

semi-concave with the modulus

$$\tilde{\omega}(r) = \tilde{C}_{\bar{V}'}\hat{\omega}_{\bar{V}'}(r) + C_{\bar{V}'}\omega(C_{\bar{V}'}r).$$

If F is C^2, then its derivative DF is locally Lipschitz on U, and we can assume that $\hat{\omega}_{\bar{V}'}$ is a linear modulus. Therefore, if ω is a linear modulus, we obtain that $\tilde{\omega}$ is also a linear modulus. $\blacksquare$

Thanks to the previous lemma, we can define a locally semi-concave function (resp., a locally semi-concave function for a linear modulus) on a manifold, as a function whose restrictions to charts is, when computed in coordinates, locally semi-concave (resp., locally semi-concave for a linear modulus). Moreover, it suffices to check this locally semi-concavity in charts for a family of charts whose domains of definition cover the manifold. It is not difficult to see that Theorem A.8 is valid on any (second countable) manifold, since we can cover such a manifold by the domains of definition of a countable family of charts.

Now we want to introduce the notion of uniformly semi-concave family of functions.

Definition A.10: Let $f_i : U \to \mathbb{R}$, $i \in I$, be a family of functions defined on an open subset U of $\mathbb{R}^n$. We will say that the family $(f_i)_{i \in I}$ is **uniformly ω-semi-concave**, where ω is a modulus of continuity, if each f_i is ω-semi-concave. We will say that the family $(f_i)_{i \in I}$ is **uniformly semi-concave** if there exists a modulus of continuity ω such that the family $(f_i)_{i \in I}$ is uniformly ω-semi-concave. We will say that the family $(f_i)_{i \in I}$ is **uniformly semi-concave with a linear modulus**, if it is uniformly ω-semi-concave, with ω of the form $t \mapsto kt$, where k is a fixed constant.

THEOREM A.11: *Suppose that $f_i : U \to \mathbb{R}$, $i \in I$, is a family of functions defined on an open subset U of $\mathbb{R}^n$. Suppose that this family $(f_i)_{i \in I}$ is uniformly ω-semi-concave, where ω is a modulus of continuity. If the function*

$$f(x) = \inf_{i \in I} f_i(x)$$

is finite everywhere on U, then $f : U \to \mathbb{R}$ is also ω-semi-concave.

Proof. Fix $x_0 \in U$. We can find a sequence i_n such that $f_{i_n}(x_0) \searrow f(x_0) > -\infty$. We choose a cube $C \subset U$ with center x_0. Call $y_1, \ldots, y_{2^n}$ the vertices of C. By

the argument in the beginning of the proof of Lemma A.5, we have

$$\min_{1 \leq j \leq 2^n} f_i(y_j) \leq f_i(x_0) + D_C \omega(D_C) \quad \forall x \in C, \ \forall i \in I,$$

where D_C is the diameter of the compact cube C. Using the fact that $f(y_j) = \inf_{i \in I} f_i(y_j)$ is finite, it follows that there exists $A \in \mathbb{R}$ such that

$$f_i(x) \geq A \quad \forall x \in C, \ \forall i \in I.$$

Choose now $\varepsilon > 0$ such that $\bar{B}(x_0, \varepsilon) \subset C$. If $l_i : \mathbb{R}^n \to \mathbb{R}$ is a linear form such that

$$f_i(y) \leq f_i(x_0) + \langle l_i, y - x_0 \rangle + \|y - x_0\| \omega(\|y - x_0\|) \quad \forall y \in U,$$

we obtain that, for every $v \in \mathbb{R}^n$ of norm 1,

$$A \leq f_i(x_0) + \langle l_i, \varepsilon v \rangle + \varepsilon \omega(\varepsilon).$$

Since $f_{i_n}(x_0) \searrow f(x_0)$, we can assume $f_{i_n}(x_0) \leq M < +\infty$ for all n, that implies

$$\|l_{i_n}\| \leq \frac{M - A}{\varepsilon} + \omega(\varepsilon) < \infty.$$

Up to extracting a subsequence, we can assume $l_{i_n} \to l$ in $\mathbb{R}^{n*}$, the dual space of $\mathbb{R}^n$. Then, as for every $y \in U$ we have $f(y) \leq f_{i_n}(y)$, passing to the limit in n in the inequality

$$f(y) \leq f_{i_n}(x_0) + \langle l_{i_n}, y - x_0 \rangle + \|y - x_0\| \omega(\|y - x_0\|),$$

we get

$$f(y) \leq f(x_0) + \langle l, y - x_0 \rangle + \|y - x_0\| \omega(\|y - x_0\|).$$

Since $x_0 \in U$ is arbitrary, this concludes the proof. ∎

Before generalizing the notion of uniformly semi-concave family of functions to manifolds, let us look at the following example.

Example A.12: For $k \in \mathbb{R}$, define $f_k : \mathbb{R} \to \mathbb{R}$ as $f_k(x) = kx$. It is clear that the family $(f_k)_{k \in \mathbb{R}}$ is ω-semi-concave for every modulus of continuity ω. In fact

$$f_k(y) - f_k(x) = k(y - x) \leq k(y - x) + |y - x| \omega(|y - x|),$$

since $\omega \geq 0$. Consider now the diffeomorphism $\varphi : \mathbb{R}_+^* \to \mathbb{R}_+^*$, $\varphi(x) = x^2$. Then there does not exist a non-empty open subset $U \subset \mathbb{R}_+^*$, and a modulus of continuity ω, such that the family $(f_k \circ \varphi|_U)_{k \in \mathbb{R}}$ is (uniformly) ω-semi-concave. Suppose in fact that

$$f_k \circ \varphi(y) - f_k \circ \varphi(x) \leq l_x(y - x) + |y - x| \omega(|y - x|),$$

where l_x depends on k but not ω. Since $f_k \circ \varphi$ is differentiable we must have $l_x(y - x) = (f_k \circ \varphi)'(x)(y - x) = 2kx(y - x)$. Therefore, we should have

$$ky^2 - kx^2 \le 2kx(y - x) + |y - x|\omega(|y - x|).$$

Fix $x, y \in U$, with $y \ne x$ and set $h = y - x$. Then

$$kh^2 \le |h|\omega(|h|) \Rightarrow k \le \frac{\omega(|h|)}{|h|} \quad \forall k,$$

that is obviously a contradiction.

Therefore, the following is the only reasonable definition for the notion of a uniformly locally semi-concave family of functions on a manifold.

Definition A.13: We will say that the family of functions $f_i : M \to \mathbb{R}$, $i \in I$, defined on the manifold M, is **uniformly locally semi-concave** (resp., **with a linear modulus**), if we can find a cover $(U_j)_{j \in J}$ of M by open subsets, with each U_j domain of a chart $\varphi_j : U_j \xrightarrow{\sim} V_j \subset \mathbb{R}^n$ (where n is the dimension of M), such that for every $j \in J$ the family of functions $(f_i \circ \varphi_j^{-1})_{i \in I}$ is a uniformly semi-concave family of functions on the open subset V_j of $\mathbb{R}^n$ (resp., with a linear modulus).

The following corollary is an obvious consequence of Theorem A.11.

COROLLARY A.14: *If the family $f_i : M \to \mathbb{R}$, $i \in I$ is uniformly locally semi-concave (resp., with a linear modulus) and the function*

$$f(x) = \inf_{i \in I} f_i(x)$$

is finite everywhere, then $f : M \to \mathbb{R}$ is locally semi-concave (resp., with a linear modulus).

Definition A.15: Suppose $c : M \times N \to \mathbb{R}$ is a function defined on the product of the manifold M by the topological space N. We will say that the family of functions $(c(\cdot, y))_{y \in N}$ is **locally uniformly locally semi-concave** (resp., **with a linear modulus**), if for each $y_0 \in N$ we can find a neighborhood V_0 of y_0 in N such that the family $(c(\cdot, y))_{y \in V_0}$ is uniformly locally semi-concave on M (resp., with a linear modulus).

PROPOSITION A.16: *Suppose $c : M \times N \to \mathbb{R}$ is a function defined on the product of the manifold M by the topological space N, such that the family*

of functions $(c(\cdot, y))_{y \in N}$ is locally uniformly locally semi-concave (resp., with a linear modulus). If $K \subset N$ is compact, and the function

$$f_K(x) = \inf_{y \in K} c(x, y)$$

is finite everywhere on U, then $f_K : U \to \mathbb{R}$ is locally semi-concave on M (resp., with a linear modulus).

Proof. By compactness of K, we can find a finite family $V_i, i = 1, \ldots, \ell$ of open subsets of N such that $K \subset \bigcup_{i=1}^{\ell} V_i$, and for every $i = 1, \ldots, \ell$, the family $(c(\cdot, y))_{y \in V_i}$ is locally uniformly locally semi-concave (resp., with a linear modulus). The function

$$f_i(x) = \inf_{y \in K \cap V_i} c(x, y)$$

is finite everywhere on U, because $f_i \geq f_K$. It follows from Corollary A.14 that f_i is locally semi-concave on M (resp., with a linear modulus), for $i = 1, \ldots, \ell$. Since $f_K = \min_{i=1}^{\ell} f_i$, we can apply part (ii) of Proposition A.4 to conclude that f_K has the same property.   ■

PROPOSITION A.17: If $c : M \times N \to \mathbb{R}$ is a locally semi-concave function (resp., with a linear modulus) on the product of the manifolds M and N, then the family of functions on M $(c(\cdot, y))_{y \in N}$ is locally uniformly locally semi-concave (resp., with a linear modulus).

Proof. We can cover $M \times N$ by a family $(U_i \times W_j)_{i \in I, j \in J}$ of open sets with U_i open in M, W_j open in N, where U_i is the domain of a chart $\varphi_i : U_i \xrightarrow{\sim} \tilde{U}_i \subset \mathbb{R}^n$ (where n is the dimension of M) and W_j is the domain of a chart $\psi_j : W_j \xrightarrow{\sim} \tilde{W}_j \subset \mathbb{R}^m$ (where m is the dimension of M), and such that

$$(\tilde{x}, \tilde{y}) \mapsto c\left(\varphi_i^{-1}(\tilde{x}), \psi_j^{-1}(\tilde{y})\right)$$

is $\omega_{i,j}$-semi-concave on $\tilde{U}_i \times \tilde{W}_j$, for some modulus $\omega_{i,j}$. It is then clear that the family

$$(c(\varphi_i^{-1}(\tilde{x}), \psi_j^{-1}(\tilde{y})))_{\tilde{y} \in \tilde{W}_j}$$

is uniformly locally $\omega_{i,j}$-semi-concave on $\tilde{U}_i$.   ■

The following corollary is now an obvious consequence of Propositions A.17 and A.16.

COROLLARY A.18: *Suppose $c : M \times N \to \mathbb{R}$ is a locally semi-concave function (resp., with a linear modulus) on the product of the manifolds M and N. Let K be a compact subset of N. If the function*

$$f_K(x) = \inf_{y \in K} c(x, y)$$

is finite everywhere on U, then $f_K : U \to \mathbb{R}$ is locally semi-concave (resp., with a linear modulus).

We end this section with another useful theorem. The proof we give is an adaptation of the proof of [20, Lemma 3.8, page 494].

THEOREM A.19: *Let $\varphi_1, \varphi_2 : M \to \mathbb{R}$ be two functions, with φ_1 locally semi-convex (i.e. $-\varphi_1$ locally semi-concave), and φ_2 locally semi-concave. Assume that $\varphi_1 \leq \varphi_2$. If we define $\mathcal{E} = \{x \in M : \varphi_1(x) = \varphi_2(x)\}$, then both φ_1 and φ_2 are differentiable at each $x \in \mathcal{E}$ with $d_x\varphi_1 = d_x\varphi_2$ at such a point. Moreover, the map $x \mapsto d_x\varphi_1 = d_x\varphi_2$ is continuous on $\mathcal{E}$.*

If φ_1 is locally semi-convex and φ_2 is locally semi-concave, both with a linear modulus, then, in fact, the map $x \mapsto d_x\varphi_1 = d_x\varphi_2$ is locally Lipschitz on $\mathcal{E}$.

Proof. Since the statement is local in nature, we will assume that $M = \overset{\circ}{\mathbb{B}}$ is the Euclidean unit ball of center 0 in $\mathbb{R}^n$, and that $-\varphi_1$ and φ_2 are semi-concave with (common) modulus ω. Suppose now that $x \in \mathcal{E}$. We can find two linear maps $l_{1,x}, l_{2,x} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\varphi_1(y) \geq \varphi_1(x) + l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}})$$
$$\varphi_2(y) \leq \varphi_2(x) + l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}).$$

Using $\varphi_1 \leq \varphi_2$, and $\varphi_1(x) = \varphi_2(x)$, we obtain

$$l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}) \leq \varphi_1(y) - \varphi_1(x)$$
$$\leq \varphi_2(y) - \varphi_2(x)$$
$$(1) \qquad\qquad\qquad\qquad \leq l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}).$$

In particular, we get

$$l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}) \leq l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}),$$

replacing y by $x + v$ with $\|v\|_{\text{euc}}$ small, we conclude

$$l_{1,x}(v) - \|v\|_{\text{euc}}\omega(\|v\|_{\text{euc}}) \leq l_{2,x}(v) + \|v\|_{\text{euc}}\omega(\|v\|_{\text{euc}}).$$

Therefore

$$|[l_{2,x} - l_{1,x}](v)| \le 2\|v\|_{\mathrm{euc}}\omega(\|v\|_{\mathrm{euc}}),$$

for v small enough. Since $l_{2,x} - l_{1,x}$ is linear it must be identically 0. We set $l_x = l_{2,x} = l_{1,x}$. For $i = 1,2$ and $y \in \overset{\circ}{\mathbb{B}}$, we obtain from (1)

(2) $$|\varphi_i(y) - \varphi_i(x) - l_x(y - x)| \le \|y - x\|_{\mathrm{euc}}\omega(\|y - x\|_{\mathrm{euc}}).$$

This implies that φ_i is differentiable at $x \in \mathcal{E}$, with $d_x\varphi_i = l$. It remains to show the continuity of the derivative. Fix $r < 1$. We now find a modulus of continuity of the derivative on the ball $r\,\overset{\circ}{\mathbb{B}}$. If $y_1, y_2 \in \mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$, and $\|k\|_{\mathrm{euc}} \le 1 - r$, we can apply (2) three times to obtain

$$\varphi_1(y_2) - \varphi_1(y_1) - d_{y_1}\varphi_1(y_2 - y_1) \le \|y_2 - y_1\|_{\mathrm{euc}}\omega(\|y_2 - y_1\|_{\mathrm{euc}})$$
$$\varphi_1(y_2 + k) - \varphi_1(y_2) - d_{y_2}\varphi_1(k) \le \|k\|_{\mathrm{euc}}\omega(\|k\|_{\mathrm{euc}})$$

$$-\varphi_1(y_2 + k) + \varphi_1(y_1) + d_{y_1}\varphi_1(y_2 + k - y_1)$$
$$\le \|y_2 + k - y_1\|_{\mathrm{euc}}\omega(\|y_2 + k - y_1\|_{\mathrm{euc}}).$$

If we add the first two inequality to the third one, we obtain

$$[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k) \le \|y_2 - y_1\|_{\mathrm{euc}}\omega(\|y_2 - y_1\|_{\mathrm{euc}}) + \|k\|_{\mathrm{euc}}\omega(\|k\|_{\mathrm{euc}})$$
$$+ [\|y_2 - y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}]\omega(\|y_2 - y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}),$$

which implies, exchanging k with $-k$, and using that the modulus ω is non-decreasing

$$|[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k)| \le 2[\|y_2 - y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}]\omega(\|y_2 - y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}).$$

Since $\|y_2 - y_1\|_{\mathrm{euc}} < 2$, we can apply the inequality (3) above with any k such that $\|k\|_{\mathrm{euc}} = (1 - r)\|y_2 - y_1\|_{\mathrm{euc}}/2$. If we divide the inequality (3) by $\|k\|_{\mathrm{euc}}$, and take the sup over all k such that $\|k\|_{\mathrm{euc}} = (1 - r)\|y_2 - y_1\|_{\mathrm{euc}}/2$, we obtain

$$\|d_{y_1}\varphi_1 - d_{y_2}\varphi_1\|_{\mathrm{euc}} \le 2\Big[\frac{2}{1 - r} + 1\Big]\omega\Big(\Big(1 + \frac{1 - r}{2}\Big)\|y_2 - y_1\|_{\mathrm{euc}}\Big).$$

It follows that a modulus of continuity of $x \mapsto d_x\varphi_1$ on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$ is given by

$$t \mapsto \frac{6 - 2r}{1 - r}\omega\Big(\frac{3 - r}{2}t\Big).$$

This implies the continuity of the map $x \mapsto d_x\varphi_1$ on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$. It also shows that it is Lipschitz on $\mathcal{E} \cap r\,\overset{\circ}{\mathbb{B}}$ when ω is a linear modulus. ∎

Appendix B. Tonelli Lagrangians

B.1. DEFINITION AND BACKGROUND. We recall some of the basic definitions, and some of the results in calculus of variations (in one variable). There are many references on the subject. In [21], one can find an introduction to the subject that is particularly suited for our purpose. Other references are [10] and the first chapters in [35]. A brief and particularly nice description of the main results is contained in [14].

Definition B.1 (Lagrangian): If M is a manifold, a **Lagrangian** on M is a function $L : TM \to \mathbb{R}$. In the following we will assume that L is at least bounded below and continuous.

Definition B.2 (Action): If L is a Lagrangian on M, for an absolutely continuous curve $\gamma : [a, b] \to M, a \leq b$, we can define its **action** $\mathbb{A}_L(\gamma)$ by

$$\mathbb{A}_L(\gamma) = \int_a^b L(\gamma(s), \dot{\gamma}(s)) \, ds.$$

Note that the integral is well-defined with values in $\mathbb{R} \cup \{+\infty\}$, because L is bounded below, and $s \to L(\gamma(s), \dot{\gamma}(s))$ is defined a.e. and measurable. To simplify notation, we set $\mathbb{A}_L(\gamma) = +\infty$ if γ is not absolutely continuous.

Definition B.3 (Minimizer): If L is a Lagrangian on the manifold M, an absolutely continuous curve $\gamma : [a, b] \to M$, with $a \leq b$, is an L-**minimizer**, if $\mathbb{A}_L(\gamma) \leq \mathbb{A}_L(\delta)$ for every absolutely continuous curve $\delta : [a, b] \to M$ with the same endpoints, i.e. such that $\delta(a) = \gamma(a)$ and $\delta(b) = \gamma(b)$.

Definition B.4 (Tonelli Lagrangian): We will say that $L : TM \to \mathbb{R}$ is a **weak Tonelli Lagrangian** on M, if it satisfies the following hypotheses:

(a) L is C^1;

(b) for each $x \in M$, the map $L(x, \cdot) : T_x M \to \mathbb{R}$ is strictly convex;

(c) there exist a complete Riemannian metric g on M and a constant $C > -\infty$ such that

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

where $\|\cdot\|_x$ is the norm on $T_x M$ obtained from the Riemannian metric g;

(d) for every compact subset $K \subset M$ the restriction of L to $T_K M = \bigcup_{x \in K} T_x M$ is superlinear in the fibers of $TM \to M$: this means that

for every $A \geq 0$, there exists a constant $C(A, K) > -\infty$ such that

$$L(x, v) \geq A\|v\|_x + C(A, K) \quad \forall (x, v) \in T_K M.$$

We will say that L is a **Tonelli Lagrangian**, if it is a weak Tonelli Lagrangian, and satisfies the following two strengthening of conditions (a) and (b) above:

(a') L is C^2;

(b') for every $(x, v) \in TM$, the second partial derivative $\frac{\partial^2 L}{\partial v^2}(x, v)$ is positive definite on $T_x M$.

Above a compact subset of a manifold all Riemannian metrics are equivalent. Therefore, if condition (d) in the definition is satisfied for one particular Riemannian metric, then it is satisfied for any other Riemannian metric.

Note that when L is a weak Tonelli Lagrangian on M, and $U : M \to \mathbb{R}$ is a C^1 function *which is bounded below*, then $L + U$, defined by $(L + U)(x, v) = L(x, v) + U(x)$ is a weak Tonelli Lagrangian. If, moreover, L is a Tonelli Lagrangian and U is C^2 and bounded below, then $L + U$ is a Tonelli Lagrangian. Therefore, one can generate a lot of (weak) Tonelli Lagrangians from the following simple example.

Example B.5: Suppose that g is a complete smooth Riemannian metric on M, and $r > 1$. We define the Lagrangian $L_{r,g}$ on M by

$$L_{r,g}(x, v) = \|v\|_x^r = g_x(v, v)^{r/2}.$$

1) $L_{2,g}$ is a Tonelli Lagrangian.
2) For any $r > 1$, the Lagrangian is C^1 and is a weak Tonelli Lagrangian.

In both cases, the Riemannian metric mentioned in condition (c) of Definition B.4 is the same metric g.

Moreover, the vertical derivative of the Lagrangian $L_{r,g}$ is given by

$$\frac{\partial L_{r,g}}{\partial v}(x, v) = r\|v\|_x^{r-2} g_x(v, \cdot),$$

Proof. Since $r > 1$ it is not difficult to check that L has (in coordinates) partial derivatives everywhere with

$$\frac{\partial L_{r,g}}{\partial x}(x, 0) = 0 \quad \text{and} \quad \frac{\partial L_{r,g}}{\partial v}(x, 0) = 0,$$

and that these partial derivatives are continuous. Therefore L is C^1. A simple computation gives

$$\frac{\partial L_{r,g}}{\partial v}(x, v) = r\|v\|_x^{r-2} g_x(v, \cdot).$$

We now prove condition (c) and (d) of Definition B.4 at once. In fact, if A is given, we have

$$L_{r,g}(x, v) = \|v\|_x^r \geq A\|v\|_x - A^{r/r-1},$$

as on can see by considering separately the two cases $\|v\|_x^{r-1} \geq A$ and $\|v\|_x^{r-1} \leq A$. The rest of the proof is easy. ∎

The completeness of the Riemannian metric in condition (c) of Definition B.4 above is crucial to guarantee that a set of the form

$$\mathcal{F} = \{\gamma \in C^0([a, b], M) : \gamma(a) \in K, \ \mathbb{A}_L(\gamma) \leq \kappa\},$$

where K is a compact subset in M, κ is a finite constant and $a \leq b$, is compact in the C^0 topology. In fact, condition (c) implies that the curves in such a set $\mathcal{F}$ have a g-length which is bounded independently of γ. Since K is compact (assuming M connected to simplify things) this implies that there exist $x_0 \in M$ and $R < +\infty$ such that all the curves in $\mathcal{F}$ are contained in the closed ball $\bar{B}(x_0, R) = \{y \in M : d(x, y) \leq R\}$, where d is the distance associated to the Riemannian metric g. But such a ball $\bar{B}(x_0, R)$ is compact since g is complete (Hopf-Rinow Theorem). From there, one obtains that the set $\mathcal{F}$ is compact in the C^0 topology; see [10, Chapters 2 and 3].

The direct method in the Calculus of Variations, see [10, Theorem 3.7, page 114] or [21] for Tonelli Lagrangians, implies:

THEOREM B.6: *Suppose L is a weak Tonelli Lagrangian on the connected manifold M. Then for every $a, b \in \mathbb{R}$, $a < b$ and every $x, y \in M$, there exists an absolutely continuous curve $\gamma : [a, b] \to M$ which is an L-minimizer with $\gamma(a) = x$ and $\gamma(b) = y$.*

In fact, in [10, Theorem 3.7, page 114], the existence of absolutely continuous minimizers is valid under very general hypotheses on the Lagrangian L (the C^1 hypothesis on L is much stronger than necessary). We now come to the problem of regularity of minimizers which uses the C^1 hypothesis on L:

THEOREM B.7: *If L is a weak Tonelli Lagrangian, then every minimizer $\gamma : [a, b] \to M$ is C^1. Moreover, on every interval $[t_0, t_1]$ contained in a domain*

of a chart, it satisfies the following equality written in the coordinate system

$$\text{(IEL)} \qquad \frac{\partial L}{\partial v}(\gamma(t_1), \dot{\gamma}(t_1)) - \frac{\partial L}{\partial v}(\gamma(t_0), \dot{\gamma}(t_0)) = \int_{t_0}^{t_1} \frac{\partial L}{\partial x}(\gamma(s), \dot{\gamma}(s))\, ds,$$

which is an integrated from of the Euler-lagrange equation. This implies that $\partial L/\partial v(\gamma(t), \dot{\gamma}(t))$ *is a* C^1 *function of* t *with*

$$\frac{d}{dt}\left[\frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t))\right] = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t)).$$

Moreover, if L *is a* C^r *Tonelli Lagrangian, with* $r \geq 2$, *then any minimizer is of class* C^r.

Proof. We will only sketch the proof. If L is a Tonelli Lagrangian, this theorem would be a formulation of what is nowadays called Tonelli's existence and regularity theory. In that case its proof can be found in many places, for example [10], [14] and [21]. The fact that the regularity of minimizers holds for C^1 (or even less smooth) Lagrangians is more recent. The fact that a minimizer is Lipschitz has been established by Clarke and Vinter, see [16, Corollary 1, page 77 and Corollary 3.1, page 90] (again, the hypothesis L is C^1 is stronger than the one required in this last work). The same fact under weaker regularity assumptions on L has been proved in [3]. A short and elegant proof of the fact that a minimizer for the class of absolutely continuous curves is necessarily Lipschitz has been given by Clarke, see [15]. The key idea of why $\dot{\gamma}$ should be bounded is that the energy is conserved along absolutely continuous minimizers (this is easy to prove for C^1 minimizers, see Corollary B.15) and the sublevels of the energy are compact (see Proposition B.16).

Once one knows that γ is Lipschitz, when L is C^1 it is possible to differentiate the action, see [10], [14] and [21]; and, using an integration by parts, one can show that γ satisfies the following integrated form (IEL') of the Euler-Lagrange equation for almost every $t \in [t_0, t_1]$, for some fixed linear form c:

$$\text{(IEL')} \qquad \frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t)) = c + \int_{t_0}^{t} \frac{\partial L}{\partial x}(\gamma(s), \dot{\gamma}(s))\, ds.$$

But the continuity of the right hand side in (IEL') implies that $\partial L/\partial v(\gamma(t), \dot{\gamma}(t))$ extends continuously everywhere on $[t_0, t_1]$. Conditions (a) and (b) on L imply

that the global Legendre transform

$$\mathscr{L} : TM \to T^*M,$$

$$(x, v) \mapsto (x, \frac{\partial L}{\partial v}(x, v)),$$

is continuous and injective, therefore a homeomorphism on its image by, for example, Brouwer's Theorem on the Invariance of Domain (see also Proposition B.9 below). We therefore conclude that $\dot{\gamma}(t)$ has a continuous extension to $[t_0, t_1]$. Since γ is Lipschitz this implies that γ is C^1. Equation (IEL) follows from (IEL'), which now holds everywhere by continuity. ∎

In fact, in this paper, we will only use the cases when L is C^2, in which case this regularity of minimizers will follow from the "usual" Tonelli regularity theory, or when L is of the form $L(x, v) = \|v\|_x^p$, $p > 1$, where the norm is obtained from a C^2 Riemannian metric, in which case the minimizers are necessarily geodesics which are of course as smooth as the Riemannian metric, see Proposition B.24 below.

To obtain further properties it is necessary to introduce the global Legendre transform.

Definition B.8 (Global Legendre Transform): If L is a C^1 Lagrangian on the manifold L, its **global Legendre transform** $\mathscr{L} : TM \to T^*M$, where T^*M is the cotangent bundle of M, is defined by

$$\mathscr{L}(x, v) = \left(x, \frac{\partial L}{\partial v}(x, v)\right).$$

PROPOSITION B.9: *If L is a weak Tonelli Lagrangian on the manifold M, then its global Legendre transform $\mathscr{L} : TM \to T^*M$ is a homeomorphism from TM onto T^*M.*

Moreover, if L is a C^r Tonelli Lagrangian with $r \geq 2$, then $\mathscr{L}$ is C^{r-1}.

Proof. We first prove the surjectivity of $\mathscr{L}$. Suppose $p \in T_x^*M$. By condition (d) in Definition B.4, we have

$$p(v) - L(x, v) \leq p(v) - (\|p\|_x + 1)\|v\|_x - C(\|p\|_x + 1, \{x\})$$

$$\leq -\|v\|_x - C(\|p\|_x + 1, \{x\}).$$

But this last quantity tends to $-\infty$, as $\|v\|_x \to +\infty$. Therefore the continuous function $v \mapsto p(v) - L(x, v)$ achieves a maximum at some point $v_p \in T_xM$. Since

this function is C^1, its derivative at v_p must be 0. This yields $p - \partial L/\partial v(x, v_p) = 0$. Hence $(x, p) = \mathscr{L}(x, v_p)$.

To prove injectivity of $\mathscr{L}$, it suffices to show that for $v, v' \in T_x M$, with $v \neq v'$, we have $\partial L/\partial v(x, v) \neq \partial L/\partial v(x, v')$. Consider the function $\varphi : [0, 1] \to \mathbb{R}, t \mapsto L(x, tv + (1-t)v')$, which by condition (b) of Definition B.4 is strictly convex. Since it is C^1, we must have $\varphi'(0) \neq \varphi'(1)$. In fact, if that was not the case, then the non-decreasing function φ' would be constant on $[0, 1]$, and φ would be affine on $[0, 1]$. This contradicts strict convexity. By a simple computation, we therefore get

$$\frac{\partial L}{\partial v}(x, v')(v - v') = \varphi'(0) \neq \varphi'(1) = \frac{\partial L}{\partial v}(x, v)(v - v').$$

This implies $\partial L/\partial v(x, v') \neq \partial L/\partial v(x, v)$. We now show that $\mathscr{L}$ is a homeomorphism. Since this map is continuous, and bijective, we have to check that it is proper, i.e. inverse images under $\mathscr{L}$ of compact subsets of T^*M are (relatively) compact. For this, it suffices to show that for every compact subset $K \subset M$, and every $C < +\infty$, the set

$$\left\{(x, v) \in TM : x \in K, \ \left\|\frac{\partial L}{\partial v}(x, v)\right\|_x \leq C\right\}$$

is compact. By convexity of $v \mapsto L(x, v)$, we obtain

$$\frac{\partial L}{\partial v}(x, v)(v) \geq L(x, v) - L(x, 0).$$

But $\|\partial L/\partial v(x, v)\|_x \geq \partial L/\partial v(x, v)(v/\|v\|_x)$, therefore by condition (d) of Definition B.4, we conclude that

$$\forall A \geq 0, \forall (x, v) \in T_K M, \quad \left\|\frac{\partial L}{\partial v}(x, v)\right\|_x \geq A - [C(K, A)/\|v\|_x].$$

Taking $A = C + 1$, we get the inclusion

$$\left\{(x, v) \in TM : x \in K, \ \|\frac{\partial L}{\partial v}(x, v)\|_x \leq C\right\}$$
$$\subset \{(x, v) \in TM : x \in K, \ \|v\|_x \leq C(K, C + 1)\},$$

and the compactness of the first set follows.

Suppose now that L is a C^r Tonelli Lagrangian with $r \geq 2$. Obviously $\mathscr{L}$ is C^{r-1}. By the inverse function theorem, to show that it is a C^{r-1} diffeomorphism, it suffices to show that the derivative is invertible at each point of TM.

But a simple computation in coordinates show that the derivative of $\mathscr{L}$ at (x,v) is given in matrix form by

$$\begin{pmatrix} \text{Id} & 0 \\ \frac{\partial^2 L}{\partial x \partial v}(x,v) & \frac{\partial^2 L}{\partial v^2}(x,v) \end{pmatrix}.$$

This is clearly invertible by (b') of Definition B.4. ∎

Definition B.10: If L is a Lagrangian on M, we define its **Hamiltonian** H : $T^*M \to \mathbb{R} \cup \{+\infty\}$ by

$$H(x,p) = \sup_{v \in T_x M} p(v) - L(x,v).$$

PROPOSITION B.11: *Let L be a weak Tonelli Lagrangian on the manifold M. Its Hamiltonian H is everywhere finite valued and satisfies the following properties:*

(a*) *H is C^1, and in coordinates*

$$\begin{cases} \frac{\partial H}{\partial p}(\mathscr{L}(x,v)) = v \\ \frac{\partial H}{\partial x}(\mathscr{L}(x,v)) = -\frac{\partial L}{\partial x}(x,v). \end{cases}$$

(b*) *for each $x \in M$, the map $H(x,\cdot) : T_x^* M \to \mathbb{R}$ is strictly convex;*

(d*) *for every compact subset $K \subset M$ the restriction of H to $T_K^* M = \bigcup_{x \in K} T_x^* M$ is superlinear in the fibers of $T^*M \to M$: this means that for every $A \geq 0$, there exists a finite constant $C^*(A,K)$ such that*

$$H(x,p) \geq A\|p\|_x + C^*(A,K), \quad \forall (x,p) \in T_K^* M.$$

In particular, the function H is a proper map, i.e. inverse images under H of compact subsets of $\mathbb{R}$ are compact.

If L is a C^r Tonelli Lagrangian with $r \geq 2$, then

(a'*) *H is C^r;*

(b'*) *for every $(x,v) \in M$, the second partial derivative $\frac{\partial^2 H}{\partial p^2}(x,p)$ is positive definite on $T_x^* M$.*

Proof. To show differentiability, using a chart in M, we can assume that $M = U$ is an open subset in $\mathbb{R}^m$. Moreover, since all Riemannian metrics are equivalent above compact subsets, replacing U by an open subset V with compact closure contained in U, we can assume that the norm used in (c) of Definition B.4 is the

constant standard Euclidean norm $\|\cdot\|_{\mathrm{euc}}$ on the second factor of $TV = V \times \mathbb{R}^m$, that is

$$L(x, v) \geq A\|v\|_{\mathrm{euc}} + C(A), \quad \forall x \in V, \ \forall v \in \mathbb{R}^m,$$

where $C(A)$ is a finite constant, and $\sup_{x \in V} L(x, 0) \leq C < +\infty$.

We have $T^*V = V \times \mathbb{R}^{m*}$, where $\mathbb{R}^{m*}$ is the dual space of $\mathbb{R}^m$. We will denote by $\|\cdot\|_{\mathrm{euc}}$ also the dual norm on $\mathbb{R}^{m*}$ obtained from $\|\cdot\|_{\mathrm{euc}}$ on $\mathbb{R}^m$. We now fix $R > 0$. If $p \in \mathbb{R}^{m*}$ satisfies $\|p\|_{\mathrm{euc}} \leq R$, we have

$$p(v) - L(x, v) \leq \|p\|_{\mathrm{euc}}\|v\|_{\mathrm{euc}} - (R+1)\|v\|_{\mathrm{euc}} - C(R+1)$$

$$\leq -\|v\|_{\mathrm{euc}} - C(R+1).$$

Since $L(x, 0) \leq C$ for $x \in V$, it follows that, for $\|v\|_{\mathrm{euc}} > C - C(R+1)$,

$$p(v) - L(x, v) \leq -C \leq -L(x, 0).$$

This implies

$$H(x, p) = \sup_{v \in \mathbb{R}^m} p(v) - L(x, v) = \sup_{\|v\|_{\mathrm{euc}} \leq C - C(R+1)} p(v) - L(x, v),$$

Therefore, the sup in the definition of $H(x, p)$ is attained at a point $v_{(x,p)}$ with $\|v_{(x,p)}\|_{\mathrm{euc}} \leq C - C(R+1)$. Note that this point $v_{(x,p)}$ is unique (compare with the argument proving that the Legendre transform is surjective). In fact, at its maximum $v_{(x,p)}$, the C^1 function $v \mapsto p(v) - L(x, v)$ must have a derivative equal to 0, and therefore

$$p = \frac{\partial L}{\partial v}(x, v_{(x,p)}).$$

This means $(x, p) = \mathcal{L}(x, v_{(x,p)})$, but the Legendre transform is injective by Proposition B.9.

Note, furthermore, that the map

$$f : \left(V \times \{\|p\|_{\mathrm{euc}} \leq R\}\right) \times \{\|v\|_{\mathrm{euc}} \leq C - C(R+1)\} \to \mathbb{R},$$

$$((x, p), v) \mapsto p(v) - L(x, v),$$

is C^1. Therefore, we obtain that H is C^1 from the following classical lemma whose proof is left to the reader.

LEMMA B.12: *Let $f : N \times K \to \mathbb{R}, (x, k) \mapsto f(x, k)$ be a continuous map, where N is a manifold, and K is a compact space. Define $F : N \to \mathbb{R}$ by $F(x) = \sup_{k \in K} f(x, k)$. Suppose that:*

    (i) *$\frac{\partial f}{\partial x}(x, k)$ exists everywhere and is continuous as a function of both variables (x, k);*

(ii) *for every $x \in N$, the set $\{k \in K : f(x,k) = F(x)\}$ is reduced to a single point, which we will denote by k_x.*

Then F is C^1*, and the derivative D_xF of F at x is given by*

$$D_xF = \frac{\partial f}{\partial x}(x, k_x).$$

Returning to the proof of Proposition B.11, by the last statement of the above lemma we also obtain

$$\frac{\partial H}{\partial p}(x,p) = v_{(x,p)} \quad \text{and} \quad \frac{\partial H}{\partial x}(x,p) = -\frac{\partial L}{\partial x}(x, v_{(x,p)})$$

Since $(x,p) = \mathscr{L}(x, v_{(x,p)})$, this can be rewritten as

$$(5) \qquad \frac{\partial H}{\partial p} \circ \mathscr{L}(x,v) = v \quad \text{and} \quad \frac{\partial H}{\partial x} \circ \mathscr{L}(x,v) = -\frac{\partial L}{\partial x}(x,v),$$

which proves (a*). Note that when L is a C^r Tonelli Lagrangian, by Proposition B.9 the Legendre transform $\mathscr{L}$ is a C^{r-1} global diffeomorphism. From the expression of the partial derivatives above, we conclude that $\partial H/\partial p$ and $\partial H/\partial x$ are both C^{r-1}. This proves (a'*).

We now prove (b'*). Taking the derivative in v of the first equality in (5)

$$\frac{\partial H}{\partial p}\left[x, \frac{\partial L}{\partial v}(x,v)\right] = v,$$

we obtain the matrix equation

$$\frac{\partial^2 H}{\partial p^2}(\mathscr{L}(x,v)) \cdot \frac{\partial^2 L}{\partial v^2}(x,v) = \mathrm{Id}_{\mathbb{R}^m},$$

where the dot $\cdot$ represents the usual product of matrices. This means that the matrix representative of $\partial^2 H/\partial p^2(x,p)$ is the inverse of the matrix of a positive definite quadratic form, therefore $\partial^2 H/\partial p^2(x,p)$ is itself positive definite.

We prove (b*). Suppose $p_1 \neq p_2$ are both in T_x^*M. Fix $t \in]0,1[$, and set $p_3 = tp_1 + (1-t)p_2$. The covectors p_1, p_2, p_3 are all distinct. Call v_1, v_2, v_3 elements in T_xM such that $p_i = \partial L/\partial v(x, v_i)$. By injectivity of the Legendre transform, the tangent vectors v_1, v_2, v_3 are also all distinct. Moreover, for $i = 1, 2$ we have

$$H(x, p_i) = p_i(v_i) - L(x, v_i),$$

$$H(x, p_3) = p_3(v_3) - L(x, v_3) = t[p_1(v_3) - L(x, v_3)] + (1-t)[p_2(v_3) - L(x, v_3)].$$

Since the sup in the definition of $H(x, p)$ is attained at a unique point, and v_1, v_2, v_3 are all distinct, for $i = 1, 2$ we must have

$$p_i(v_3) - L(x, v_3) < p_i(v_i) - L(x, v_i) = H(x, p_i).$$

It follows that

$$H(x, tp_1 + (1 - t)p_2) < tH(x, p_1) + (1 - t)H(x, p_2).$$

It remains to prove (d*). Fix a compact set K in M. Since

$$H(x, p) \geq p(v) - L(x, v),$$

we obtain

$$H(x, p) \geq \sup_{\|v\|_x \leq A} p(v) + \inf_{x \in K, \|v\|_x \leq A} -L(x, v).$$

But $C^*(A, K) = \inf_{x \in K, \|v\|_x \leq A} -L(x, v)$ is finite by compactness, and $\sup_{\|v\|_x \leq A} p(v) = A\|p\|_x$. ∎

Since for a weak Tonelli Lagrangian L, the Hamiltonian $H : T^*M \to \mathbb{R}$ is C^1, we can define the Hamiltonian vector field X_H on T^*M. This is rather standard and uses the fact that the exterior derivative of the Liouville form on M defines a symplectic form on M, see [1] or [28]. The vector field X_H is entirely characterized by the fact that in coordinates obtained from a chart in M, it is given by

$$X_H(x, p) = \Big(\frac{\partial H}{\partial p}(x, p), -\frac{\partial H}{\partial x}(x, p) \Big).$$

So the associated ODE is given by

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p}(x, p) \\ \dot{p} = -\frac{\partial H}{\partial x}(x, p). \end{cases}$$

In this form, it is an easy exercise to check that H is constant on any solution of X_H.

We now come to the simple and important connection between minimizers and solutions of X_H.

THEOREM B.13: *Suppose L is a weak Tonelli Lagrangian on M. If $\gamma{:}[a, b]{\to}M$ is a minimizer for L, then the Legendre transform of its speed curve $t \mapsto \mathscr{L}(\gamma(t), \dot{\gamma}(t))$ is a C^1 solution of the Hamiltonian vector field X_H obtained from the Hamiltonian H associated to L.*

Moreover, if L is a Tonelli Lagrangian, there exists a (partial) C^1 flow ϕ_t^L on TM such that every speed curve of an L-minimizer is a part of an orbit of ϕ_t^L. This flow is called the Euler-Lagrange flow, is defined by

$$\phi_t^L = \mathscr{L}^{-1} \circ \phi_t^H \circ \mathscr{L},$$

where ϕ_t^H is the partial flow of the C^1 vector filed X_H.

Proof. If we write $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot\gamma(t))$ then

$$x(t) = \gamma(t) \quad \text{and} \quad p(t) = \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t)).$$

By Theorem B.7, $x(t) = \gamma(t)$ is C^1 with $\dot x(t) = \dot\gamma(t)$. The fact that $p(t)$ is C^1 follows again from Theorem B.7, which also yields in local coordinates

$$\dot p(t) = \frac{\partial L}{\partial x}(\gamma(t), \dot\gamma(t)).$$

Since $(x(t), p(t)) = \mathscr{L}(\gamma(t), \dot\gamma(t))$, we conclude from Proposition B.11 that $t \mapsto (x(t), p(t))$ satisfies the ODE

$$\begin{cases} \dot x = \frac{\partial H}{\partial p}(x, p) \\ \dot p = -\frac{\partial H}{\partial x}(x, p). \end{cases}$$

Therefore the Legendre transform of the speed curve of a minimizer is a solution of the Hamiltonian vector field X_H.

If L is a Tonelli Lagrangian, by Proposition B.11 the Hamiltonian H is C^2. Therefore the vector field X_H is C^1, and it defines a (partial) C^1 flow ϕ_t^H. The rest follows from what was obtained above and the fact that the Legendre transform is C^1. ∎

We recall the following definition

Definition B.14 (Energy): If L is a C^1 Lagrangian on the manifold M, its **energy** $E : TM \to \mathbb{R}$ is defined by

$$E(x, v) = H \circ \mathscr{L}(x, v) = \frac{\partial L}{\partial v}(x, v)(v) - L(x, v).$$

COROLLARY B.15 (Conservation of Energy): If L is a C^1 Lagrangian on the manifold M, and $\gamma : [a, b] \to M$ is a C^1 minimizer for L, then the energy E is constant on the speed curve

$$s \mapsto (\gamma(s), \dot\gamma(s)).$$

Proof. In fact $E(\gamma(s), \dot{\gamma}(s)) = H \circ \mathscr{L}(\gamma(s), \dot{\gamma}(s))$. But $s \mapsto \mathscr{L}(\gamma(s), \dot{\gamma}(s))$ is a solution of the vector field H, and the Hamiltonian H is constant on orbits of X_H. ∎

PROPOSITION B.16: *If L is a weak Tonelli Lagrangian on the manifold M, then for every compact subset $K \subset M$, and every $C < +\infty$, the set*

$$\{(x, v) \in TM : x \in K, \ E(x, v) \le C\}$$

is compact, i.e., the map $E : TM \to \mathbb{R}$ is proper on every subset of the form $\pi^{-1}(K)$, where K is a compact subset of M.

Proof. Since $E = H \circ \mathscr{L}$, this follows from the fact that H is proper and $\mathscr{L}$ is a homeomorphism. ∎

PROPOSITION B.17: *Let L be a weak Tonelli Lagrangian on M. Suppose K is a compact subset of M, and $t > 0$. Then we can find a compact subset $\tilde{K} \subset M$ and a finite constant A, such that every minimizer $\gamma : [0, t] \to M$ with $\gamma(0), \gamma(t) \in K$ satisfies $\gamma([0, t]) \subset \tilde{K}$ and $\|\dot{\gamma}(s)\|_{\gamma(s)} \le A$ for every $s \in [0, t]$.*

Proof. We will use as a distance d the one coming from the complete Riemannian metric. All finite closed balls in this distance are compact (Hopf-Rinow theorem). We choose $x_0 \in K$, and R such that $K \subset B(x_0, R)$ (we could take $R = \mathrm{diam}(K)$, the diameter of K). We now pick $x, y \in K$. If $\alpha : [0, t] \to M$ is a geodesic with $\alpha(0) = x$, $\alpha(t) = y$ and whose length is $d(x, y)$ (such a geodesic exists by completeness), the inequality

$$d(x, y) \le d(x, x_0) + d(x_0, y) \le 2R$$

implies that $\alpha([0, t]) \subset \bar{B}(x_0, 3R)$. Moreover $\|\dot{\alpha}(s)\|_{\alpha(s)} = d(x, y)/t \le 2R/t$ for every $s \in [0, t]$. By compactness, the Lagrangian L is bounded on the set

$$\mathscr{K} = \{(z, v) \in TM : z \in \bar{B}(x_0, 3R), \ \|v\|_z \le 2R/t\}.$$

We call θ an upper bound of L on $\mathscr{K}$. Obviously the action of α on $[0, t]$ is less than $t\theta$, and therefore if $\gamma : [0, t] \to M$ is a minimizer with $\gamma(0), \gamma(t) \in K$, we get $\int_0^t L(\gamma(s), \dot{\gamma}(s)) \, ds \le t\theta$. Using condition (c) on the Lagrangian L and what we obtained above, we see that

$$Ct + \int_0^t \|\dot{\gamma}(s)\|_{\gamma(s)} \, ds \le t\theta.$$

It follows that we can find $s_0 \in [0, t]$ such that

$$\|\dot{\gamma}(s_0)\|_{\gamma(s_0)} \leq \theta - C.$$

Moreover

$$\gamma([0, t]) \subset \bar{B}(\gamma(0), t(\theta - C)) \subset \bar{B}(x_0, R + t(\theta - C)).$$

We set $\tilde{K} = \bar{B}(x_0, R + t(\theta - C))$. If we define

$$\theta_1 = \sup\{E(z, v) : (z, v) \in TM, \ z \in \tilde{K}, \ \|v\|_z \leq \theta - C\},$$

we see that θ_1 is finite by compactness. Moreover $E(\gamma(s_0), \dot{\gamma}(s_0)) \leq \theta_1$. But, as mentioned earlier, the energy $E(\gamma(s), \dot{\gamma}(s))$ is constant on the curve. This implies that the speed curve

$$s \mapsto (\gamma(s), \dot{\gamma}(s))$$

is contained in the compact set

$$\tilde{\mathscr{K}} = \{(z, v) \in TM : z \in \tilde{K}, \ E(z, v) \leq \theta_1\}.$$

Observing that the set $\tilde{\mathscr{K}}$ does not depend on γ, this finishes the proof.   ∎

B.2. LAGRANGIAN COSTS AND SEMI-CONCAVITY.

Definition B.18 (Costs for a Lagrangian): Suppose $L : TM \to \mathbb{R}$ is a Lagrangian on the connected manifold M, which is bounded from below. For $t > 0$, we define the **cost** $c_{t,L} : M \times M \to \mathbb{R}$ by

$$c_{t,L}(x, y) = \inf_{\gamma(0)=x, \gamma(t)=y} \mathbb{A}_L(\gamma)$$

where the infimum is taken over all the absolutely continuous curves $\gamma : [0, t] \to M$, with $\gamma(0) = x$, and $\gamma(t) = y$, and $\mathbb{A}_L(\gamma)$ is the action $\int_0^t L(\gamma(s), \dot{\gamma}(s)) \, ds$ of γ.

Using a change of variable in the integral defining the action, it is not difficult to see that $c_{t,L} = c_{1,L^t}$ where the Lagrangian L^t on M is defined by $L^t(x, v) = tL(x, t^{-1}v)$. Observe that L^t is a (weak) Tonelli Lagrangian if L is.

THEOREM B.19: *Suppose that $L : TM \to \mathbb{R}$ is a weak Tonelli Lagrangian. Then, for every $t > 0$, the cost $c_{t,L}$ is locally semi-concave on $M \times M$. Moreover, if the derivative of L is locally Lipschitz, then $c_{t,L}$ is locally semi-concave with a linear modulus.*

In particular, if L is a Tonelli Lagrangian for every $t > 0$, the cost $c_{t,L}$ is locally semi-concave on $M \times M$ with a linear modulus.

Proof. By the remark preceding the statement of the theorem, it suffices to prove this for $c = c_{1,L}$. Let n be the dimension of M. Choose two charts $\varphi_i : U_i \xrightarrow{\sim} \mathbb{R}^n$, $i = 0, 1$, on M. We will show that

$$(\tilde{x}_0, \tilde{x}_1) \mapsto c(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1))$$

is semi-concave on $\overset{\circ}{\mathbb{B}} \times \overset{\circ}{\mathbb{B}}$, where $\mathbb{B}$ is the closed Euclidean unit ball of center 0 in $\mathbb{R}^n$. By Proposition B.17, we can find a constant A such that for every minimizer $\gamma : [0,1] \to M$, with $\gamma(i) \in \varphi_i^{-1}(\mathbb{B})$, we have

$$\|\dot{\gamma}(s)\|_{\gamma(s)} \leq A \quad \forall s \in [0,1].$$

We now pick $\delta > 0$ such that for all $z_1, z_2 \in \mathbb{R}^n$, with $\|z_1\|_{\mathrm{euc}} \leq 1$, $\|z_2\|_{\mathrm{euc}} = 2$,

$$d(\varphi_i^{-1}(z_1), \varphi_i^{-1}(z_2)) \geq \delta, \quad i = 0, 1,$$

where $\|\cdot\|_{\mathrm{euc}}$ denote the Euclidean norm. Then we choose $\varepsilon > 0$ such that $A\varepsilon < \delta$. It follows that

$$\gamma([0, \varepsilon]) \subset \varphi_0^{-1}\big(2 \overset{\circ}{\mathbb{B}}\big) \quad \text{and} \quad \gamma([1 - \varepsilon, 1]) \subset \varphi_1^{-1}\big(2 \overset{\circ}{\mathbb{B}}\big).$$

We set $\tilde{x}_i = \varphi_i(\gamma(i))$, $i = 0, 1$. For $h_0, h_1 \in \mathbb{R}^n$ we can define $\tilde{\gamma}_{h_0} : [0, \varepsilon] \to \mathbb{R}^n$ and $\tilde{\gamma}_{h_1} : [1 - \varepsilon, 1] \to \mathbb{R}^n$ as

$$\tilde{\gamma}_{h_0}(s) = \frac{\varepsilon - s}{\varepsilon} h_0 + \varphi_0(\gamma(s)), \quad 0 \leq s \leq \varepsilon,$$

$$\tilde{\gamma}_{h_1}(s) = \frac{s - (1 - \varepsilon)}{\varepsilon} h_1 + \varphi_1(\gamma(s)), \quad 1 - \varepsilon \leq s \leq 1.$$

We observe that when $h_0 = 0$ (or $h_1 = 0$) the curve coincide with γ. Moreover $\tilde{\gamma}_{h_0}(0) = \tilde{x}_0 + h_0$, $\tilde{\gamma}_{h_1}(1) = \tilde{x}_1 + h_1$. We suppose that $\|h_i\|_{\mathrm{euc}} \leq 2$. In that case the images of both $\tilde{\gamma}_{h_0}$ and $\tilde{\gamma}_{h_0}$ are contained in $4 \overset{\circ}{\mathbb{B}}$ and

$$\|\dot{\tilde{\gamma}}_{h_i}(s)\|_{\mathrm{euc}} \leq \|h_i\|_{\mathrm{euc}} + \|(\varphi_i \circ \gamma)'(s)\|_{\mathrm{euc}} \leq 2 + \|(\varphi_i \circ \gamma)'(s)\|_{\mathrm{euc}}.$$

Since we know that the speed of γ is bounded in M, we can find a constant A_1 such that

$$\|\dot{\tilde{\gamma}}_{h_0}(s)\|_{\mathrm{euc}} \leq A_1, \quad \forall s \in [0, \varepsilon],$$

$$\|\dot{\tilde{\gamma}}_{h_1}(s)\|_{\mathrm{euc}} \leq A_1, \quad \forall s \in [1 - \varepsilon, 1].$$

To simplify a little bit the notation, we define the Lagrangian $L_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$L_i(z, v) = L(\varphi_i^{-1}(z), D[\varphi_i^{-1}](v)).$$

If we concatenate the three curves $\varphi_0^{-1} \circ \tilde{\gamma}_{h_0}$, $\gamma|_{[\varepsilon, 1-\varepsilon]}$ and $\varphi_1^{-1} \circ \tilde{\gamma}_{h_1}$, we obtain a curve in M between $\varphi_0^{-1}(\tilde{x}_0 + h_0)$ and $\varphi_1^{-1}(\tilde{x}_1 + h_1)$, and therefore

$$
\begin{aligned}
c\left(\varphi_0^{-1}(\tilde{x}_0 + h_0), \varphi_1^{-1}(\tilde{x}_1 + h_1)\right) \leq & \int_0^\varepsilon L_0(\tilde{\gamma}_{h_0}(t), \dot{\tilde{\gamma}}_{h_0}(t))\, dt \\
& + \int_\varepsilon^{1-\varepsilon} L(\gamma(t), \dot{\gamma}(t))\, dt + \int_{1-\varepsilon}^1 L_1(\tilde{\gamma}_{h_1}(t), \dot{\tilde{\gamma}}_{h_1}(t))\, dt.
\end{aligned}
$$

Hence

$$
\begin{aligned}
& c\left(\varphi_0^{-1}(\tilde{x}_0 + h_0), \varphi_1^{-1}(\tilde{x}_1 + h_1)\right) - c\left(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1)\right) \\
& \quad \leq \int_0^\varepsilon \left[L_0(\tilde{\gamma}_{h_0}(t), \dot{\tilde{\gamma}}_{h_0}(t)) - L_0(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t))\right]\, dt \\
& \qquad + \int_{1-\varepsilon}^1 \left[L_1(\tilde{\gamma}_{h_1}(t), \dot{\tilde{\gamma}}_{h_1}(t)) - L_1(\varphi_1 \circ \gamma(t), (\varphi_1 \circ \gamma)'(t))\right]\, dt.
\end{aligned}
$$

We now call ω a common modulus of continuity for the derivative DL_0 and DL_1 on the compact set $\bar{B}(0, 4) \times \bar{B}(0, A_1)$. Here DL_0 and DL_1 denote the total derivatives of L_0 and L_1, i.e. with respect to all variables. When L has a derivative which is locally Lipschitz, then DL_0 and DL_1 are also locally Lipschitz on $\mathbb{R}^n \times \mathbb{R}^n$, and the modulus ω can be taken linear. Since $\tilde{\gamma}_{h_i}(s) \in \overset{\circ}{B}(0, 4)$ and $\|\dot{\tilde{\gamma}}_{h_i}(s)\| \leq A_1$, we get the estimate

$$
\begin{aligned}
& c\left(\varphi_0^{-1}(\tilde{x}_0 + h_0), \varphi_1^{-1}(\tilde{x}_1 + h_1)\right) - c\left(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1)\right) \\
& \quad \leq \int_0^\varepsilon DL_0\left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right)\left(\frac{\varepsilon - t}{\varepsilon} h_0, -\frac{1}{\varepsilon} h_0\right) dt \\
& \qquad + \int_{1-\varepsilon}^1 DL_1\left(\varphi_1 \circ \gamma(t), (\varphi_1 \circ \gamma)'(t)\right)\left(\frac{t - (1-\varepsilon)}{\varepsilon} h_1, \frac{1}{\varepsilon} h_1\right) dt \\
& \qquad + \omega\left(\frac{1}{\varepsilon}\|h_0\|_{\text{euc}}\right)\frac{1}{\varepsilon}\|h_0\|_{\text{euc}} + \omega\left(\frac{1}{\varepsilon}\|h_1\|_{\text{euc}}\right)\frac{1}{\varepsilon}\|h_1\|_{\text{euc}}.
\end{aligned}
$$

We observe that the sum of the first two terms in the right hand side is linear, while the sum of the last two is bounded by

$$
\frac{1}{\varepsilon}\omega\left(\frac{1}{\varepsilon}\|(h_0, h_1)\|_{\text{euc}}\right)\|(h_0, h_1)\|_{\text{euc}}.
$$

Therefore we obtain that

$$(\tilde{x}_0, \tilde{x}_1) \mapsto c\left(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1)\right)$$

is semi-concave for the modulus $\tilde{\omega}(r) = \frac{1}{\varepsilon}\omega\left(\frac{1}{\varepsilon}r\right)$ on $\overset{\circ}{\mathbb{B}} \times \overset{\circ}{\mathbb{B}}$, as wanted.   ∎

COROLLARY B.20: *If L is a weak Tonelli Lagrangian on the connected manifold M, then, for every $t > 0$, a superdifferential of $c_{t,L}(x,y)$ at (x_0, y_0) is given by*

$$(w_0, w_1) \mapsto \frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t))(w_1) - \frac{\partial L}{\partial v}(\gamma(0), \dot{\gamma}(0))(w_0),$$

where $\gamma : [0,t] \to M$ is a minimizer for L with $\gamma(0) = x_0$, $\gamma(t) = y_0$, and $(w_0, w_1) \in T_x M \times T_y M = T_{(x,y)}(M \times M)$.

Proof. Again, we will do it only for $t = 1$. If we use the notation introduced in the previous proof, we see that a superdifferential of

$$(\tilde{x}_0, \tilde{x}_1) \mapsto c\left(\varphi_0^{-1}(\tilde{x}_0), \varphi_1^{-1}(\tilde{x}_1)\right)$$

is given by

$$(h_0, h_1) \mapsto l_0(h_0) + l_1(h_1),$$

where

$$(6) \quad l_0(h_0) = -\int_0^\varepsilon \left[\frac{t-\varepsilon}{\varepsilon} \frac{\partial L_0}{\partial x}\left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right)(h_0) \right.$$
$$\left. + \frac{1}{\varepsilon} \frac{\partial L_0}{\partial v}\left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right)(h_0) \right] dt,$$

$$l_1(h_1) = \int_{1-\varepsilon}^1 \left[\frac{t-(1-\varepsilon)}{\varepsilon} \frac{\partial L_1}{\partial x}\left(\varphi_1 \circ \gamma(t), (\varphi_1 \circ \gamma)'(t)\right)(h_1) \right.$$
$$\left. + \frac{1}{\varepsilon} \frac{\partial L_1}{\partial v}\left(\varphi_1 \circ \gamma(t), (\varphi_1 \circ \gamma)'(t)\right)(h_1) \right] dt.$$

By Theorem B.7, the curve $t \mapsto \varphi_0 \circ \gamma(t)$ is a C^1 extremal of L_0 and it satisfies the Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial L_0}{\partial v}\left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right) \frac{\partial L_0}{\partial x}\left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right).$$

Using this identity in (6) an integrating by parts, we get

$$l_0(h_0) = -\frac{\partial L_0}{\partial v}\left(\varphi_0 \circ \gamma(0), (\varphi_0 \circ \gamma)'(0)\right)(h_0).$$

This means that l_0, reinterpreted on $T_{x_0} M$ rather than on $\mathbb{R}^n$, is given by $-\frac{\partial L}{\partial v}(\gamma(0), \dot{\gamma}(0))$. The treatment for l_1 is the same.   ∎

We have avoided the first variation formula in the proof of Corollary B.20, because this is usually proved for C^2 variation of curves and C^2 Lagrangians. Of course, our argument to prove this corollary is basically a proof for the first variation formula for C^1 Lagrangians. This is of course already known and the proof is the standard one.

B.3. THE TWIST CONDITION FOR COSTS OBTAINED FROM LAGRANGIANS.

LEMMA B.21: *Let L be a weak Tonelli Lagrangian on the connected manifold M. Suppose that L satisfies the following condition:*

(UC) *If $\gamma_i : [a_i, b_i] \to M, i = 1, 2$ are two L-minimizers such that $\gamma_1(t_0) = \gamma_2(t_0)$ and $\dot\gamma_1(t_0) = \dot\gamma_2(t_0)$, for some $t_0 \in [a_1, b_1] \cap [a_2, b_2]$, then $\gamma_1 = \gamma_2$ on the whole interval $[a_1, b_1] \cap [a_2, b_2]$.*

Then, for every $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ satisfies the left (and the right) twist condition of Definition 2.4.

Moreover, if $(x, y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$, then we have:

(i) *there is a unique L-minimizer $\gamma : [0, t] \to M$ such that $x = \gamma(0)$, and $y = \gamma(t)$;*

(ii) *the speed $\dot\gamma(0)$ is uniquely determined by the equality*

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)).$$

Proof. We first prove part (ii). Pick $\gamma : [0, t] \to M$ an L-minimizer with $x = \gamma(0)$ and $y = \gamma(t)$. From Corollary B.20 we obtain the equality

$$(*)\qquad\qquad \frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)).$$

Since the C^1 map $v \mapsto L(x, v)$ is strictly convex, the Legendre transform $v \in T_x M \mapsto \partial L/\partial v(x, v)$ is injective, and therefore $\dot\gamma(0) \in T_x M$ is indeed uniquely determined by Equation $(*)$ above. This proves (ii).

To prove statement (i), consider another L-minimizer $\gamma_1 : [0, t] \to M$ is $x = \gamma_1(0)$. By what we just said, we also have

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma_1(0)).$$

By the uniqueness already proved in statement (ii), we get $\dot\gamma_1(0) = \dot\gamma(0)$. It now follows from condition (UC) that $\gamma = \gamma_1$ on the whole interval $[0, t]$.

The twist condition follows easily. Consider $(x, y), (x, y_1) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$ such that

$$(**) \qquad \frac{\partial c_{t,L}}{\partial x}(x, y) = \frac{\partial c_{t,L}}{\partial x}(x, y_1).$$

By (i) there is a unique L-minimizer $\gamma : [0, t] \to M$ (resp., $\gamma_1 : [0, t] \to M$) such that $x = \gamma(0), y = \gamma(1)$ (resp., $x = \gamma_1(0), y_1 = \gamma_1(1)$), and

$$\frac{\partial c_{t,L}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot\gamma(0)) \quad \text{and} \quad \frac{\partial c_{t,L}}{\partial x}(x, y_1) = -\frac{\partial L}{\partial v}(x, \dot\gamma_1(0)).$$

From equation $(**)$, and the injectivity of the Legendre transform of L, it follows that $\dot\gamma_1(0) = \dot\gamma(0)$. From condition (UC) we get $\gamma = \gamma_1$ on the whole interval $[0, t]$. In particular, we obtain $y = \gamma(t) = \gamma_1(t) = y_1$. ∎

The next lemma is an easy consequence of Lemma B.21 above.

LEMMA B.22: *Let L be a weak Tonelli Lagrangian on M. If we can find a continuous local flow ϕ_t defined on TM such that:*

(UC') *for every L-minimizer $\gamma : [a, b] \to M$, and every $t_1, t_2 \in [a, b]$, the point $\phi_{t_2-t_1}(\gamma(t_1), \dot\gamma(t_1))$ is defined and $(\gamma(t_2), \dot\gamma(t_2)) = \phi_{t_2-t_1}(\gamma(t_1), \dot\gamma(t_1))$,*

then L satisfies (UC). Therefore, for every $t > 0$, the cost $c_{t,L} : M \times M \to \mathbb{R}$ satisfies the left twist (and the right) condition of Definition 2.4.

Moreover, if $(x, y) \in \mathcal{D}(\Lambda^l_{c_{t,L}})$, then $y = \pi\phi_t(x, v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_xM$ is uniquely determined by the equation

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, v).$$

The curve $s \in [0, t] \mapsto \pi\phi_s(x, v)$ is the unique L-minimizer $\gamma : [0, t] \to M$ with $\gamma(0) = x, \gamma(1) = y$.

Note that the following proposition is contained in Theorem B.13.

PROPOSITION B.23: *If L is a Tonelli Lagrangian, then it satisfies condition (UC') for the Euler Lagrange flow ϕ_t^L.*

PROPOSITION B.24: *Suppose g is a complete Riemannian metric on the connected manifold M, and $r > 1$. For a given $t > 0$, the cost $c_{t,L_{r,g}}$ of the weak Tonelli Lagrangian $L_{r,g}$, defined by*

$$L_{r,g}(x, v) = \|v\|_x^r = g_x(v, v)^{r/2},$$

is given by

$$c_{t,L_{r,g}} = t^{r-1}d_g^r(x,y),$$

where d_g is the distance defined by the Riemannian metric. The Lagrangian $L_{r,g}$ satisfies condition (UC') of Lemma B.22 for the geodesic flow ϕ_t^g of g. Therefore, its cost $c_{t,L_{r,g}}$ satisfies the left (and the right) twist condition. Moreover, if $(x,y) \in \mathcal{D}(\Lambda_{c_{t,L_{r,g}}}^l)$, then $y = \pi\phi_t^g(x,v)$, where $\pi : TM \to M$ is the canonical projection, and $v \in T_xM$ is uniquely determined by the equation

$$\frac{\partial c_{t,L_{r,g}}}{\partial x}(x,y) = -\frac{\partial L_{r,g}}{\partial v}(x,v).$$

Proof. Define s by $1/s + 1/r = 1$. Let $\gamma : [a,b] \to M$ be a piecewise C^1 curve. Denoting by $\ell_g(\gamma)$ the Riemannian length of γ, we can apply Hölder inequality to obtain

$$\int_a^b \|\gamma(s)\|_x \, ds \le (b-a)^{1/s} \left(\int_a^b \|\gamma(s)\|_x^r \, ds \right)^{1/r},$$

with equality if and only if γ is parameterized with $\|\gamma(s)\|_x$ constant, i.e. proportionally to arc-length. This, of course, implies

$$(b-a)^{-r/s}\ell_g(\gamma)^r \le \int_a^b \|\gamma(s)\|_x^r \, ds,$$

with equality if and only if γ is parameterized proportionally to arc-length. Since any curve can be reparametrized proportionally to arc-length and $r/s = r - 1$, we conclude that

$$c_{t,L_{r,g}}(x,y) = t^{1-r}d_g(x,y)^r,$$

and that an $L_{r,g}$-minimizing curve has to minimize the length between its endpoints. Therefore, any $L_{r,g}$-minimizing curve is a geodesic and its speed curve is an orbit of the geodesic flow ϕ_t^g. Therefore, $L_{r,g}$ satisfies condition (UC') of Lemma B.22 for the geodesic flow ϕ_t^g of g. The rest of the proposition follows from Lemma B.22. ∎

References

[1] R. Abraham and J. E. Marsden, *Foundations of mechanics*, second edition, revised and enlarged, Benjamin/Cummings Publishing Co. Inc. Advanced Book Program, Reading, Mass, 1978.

[2] L. Ambrosio, *Lecture notes on optimal transport problems* , in *Mathematical Aspects of Evolving Interfaces*, Lecture Notes in Mathematics **1812**, Springer-Verlag, Berlin/New York, 2003, pp. 1–52.

[3] L. Ambrosio, O. Ascenzi and G. Buttazzo, *Lipschitz regularity for minimizers of integral functionals with highly discontinuous integrands*, Journal of Mathematical Analysis and Applications **142** (1989), 301–316.

[4] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the Wasserstein space of probability measures*, Lectures in Mathematics, ETH Zurich, Birkhaüser Verlag, Basel, 2005.

[5] L. Ambrosio and A. Pratelli, *Existence and stability results in the L^1 theory of optimal transportation, in Optimal Transportation and Applications*, in *Lecture Notes in Mathematics* **1813**, Springer-Verlag, Berlin/New York, 2003, pp. 123–160.

[6] V. Bangert, *Analytische Eigenschaften konvexer Funktionen auf Riemannschen Manigfaltigkeiten*, Journal fuür die Reine und Angewandte Mathematik **307** (1979), 309–324.

[7] P. Bernard and B. Buffoni, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society (JEMS) **9** (2007), 85–121.

[8] P. Bernard and B. Buffoni, *The Monge problem for supercritical Mañé potential on compact manifolds*, Advances in Mathematics **207** (2006), 691–706.

[9] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Communications on Pure and Applied Mathematics **44** (1991), 375–417.

[10] G. Buttazzo, M. Giaquinta and S. Hildebrandt, *One-dimensional variational problems*, in *Oxford Lecture Series in Mathematics and its Applications* **15** , Oxford University Press, New York, 1998.

[11] L. Caffarelli, M. Feldman and R.J. McCann, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs.* Journal of the American Mathematical Society **15** (2002), 1–26.

[12] L. A. Caffarelli and S. Salsa, Editors, *Optimal transportation and applications*, Lecture Notes in Mathematics **1813**, Lectures from the C.I.M.E. Summer School held in Martina Franca, September 2–8, 2001, Springer-Verlag, Berlin, 2003.

[13] P.Cannarsa and C.Sinestrari, *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Progress in Nonlinear Differential Equations and Their Applications **58**, Birkhäuser, Boston, 2004.

[14] F. H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics **57**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[15] F. H. Clarke, *A Lipschitz regularity theorem*, Ergodic Theory and Dynamical Systems **27** (2007), 1713–1718.

[16] F. H. Clarke and R.B. Vinter, *Regularity properties of solutions to the basic problem in the calculus of variations*, Transactions of the American Mathematical Society **289** (1985), 73–98.

[17] D. L. Cohn, *Measure theory*, Birkhäuser Boston, Mass, 1980.

[18] L. C. Evans, *Partial differential equations and Monge-Kantorovich mass transfer,* in *Current Developments in Mathematics* (R. Bott et al., eds.), International Press, Cambridge, 1997, pp. 26–78.

[19] L. C. Evans and W. Gangbo, *Differential Equations Methods for the Monge-Kantorovich Mass Transfer Problem*, Memoirs of the American Mathematical Society **137**, (1999).

[20] A. Fathi, *Regularity of C^1 solutions of the Hamilton-Jacobi equation*, Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6 **12** (2003), 479–516.

[21] A. Fathi, *Weak KAM theorems in Lagrangian Dynamics*, Book to appear.

[22] H. Federer, *Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften* **153**, Springer-Verlag, New York Inc., New York, 1969.

[23] M. Feldman and R. J. McCann, *Monge's transport problem on a Riemannian manifold*, Transactions of the American Mathematical Society **354** (2002), 1667–1697.

[24] A. Figalli, *The Monge problem on non-compact manifolds*, Rendiconti del Seminario Matematico della Università di Padova. The Mathematical Journal of the University of Padova **117** (2007), 147–166.

[25] J. N. Mather and G. Forni, *Action minimizing orbits in Hamiltonian systems, Transition to chaos in classical and quantum mechanics, Montecatini Terme, 1991*, in *Lecture Notes in Mathematics* **1589**, Springer, Berlin, 1994, pp. 92–186.

[26] W. Gangbo, *The Monge mass transfer problem and its applications*, in *Monge Ampère equation: applications to geometry and optimization*, Contemporary Mathematics **226**, American Mathematical Society, Providence, RI, 1999, pp. 79–104.

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

[28] H. Hofer and E. Zehnder, *Symplectic invariants and Hamiltonian dynamics*, Birkhäuser Advanced Texts: Basler Lehrbücher, Birkhäuser Verlag, Basel, 1994.

[29] L. V. Kantorovich, *On the transfer of masses*, Doklady Akademii Nauk SSSR **37** (1942), 227–229.

[30] L. V. Kantorovich, *On a problem of Monge*. Akademiya Nauk SSSR i Moskovskoe Matematicheskoe Obshchestvo. Uspekhi Matematicheskikh Nauk **3** (1948), 225–226.

[31] M. Knott and C. S. Smith, *On the optimal mapping of distributions*, Journal of Optimization Theory and Applications **43** (1984), 39-49.

[32] J. N. Mather, *Existence of quasiperiodic orbits for twist homeomorphisms of the annulus*, Topology **21** (1982), 457–467.

[33] R. McCann, *Polar factorization of maps on Riemannian manifolds*, Geometric and Functional Analysis **11** (2001), no. 3, 589–608.

[34] G. Monge, *Mémoire sur la Théorie des Déblais et des Remblais*, Hist. de l'Acad. des Sciences de Paris (1781), 666–704.

[35] J. Moser, *Selected chapters in the calculus of variations*, in *Lectures in Mathematics ETH Zürich*, Lecture Notes by Oliver Knill, Birkhäuser Verlag, Basel, 2003.

[36] S. T. Rachev and L. Rüschendorf, *A characterization of random variables with minimum L^2-distance*, Journal of Multivariate Analysis **32** (1990), 48–54. Corrigendum in Journal of Multivariate Analysis **34** (1990), 156.

[37] S.T. Rachev and L. Rüschendorf, *Mass transportation problems*, Vol. I: Theory, Vol. II: Applications. in *Probability and its Applications*, Springer-Verlag, New York, 1998.

[38] W. Schachermayer and J. Teichmann, *Characterization of optimal Transport Plans for the Monge-Kantorovich-Problem*, Proceedings of the American Mathematical Society, to appear.

[39] V. N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions*, Proceedings of the Steklov Institute of Mathematics **141** (1979), 1–178.

[40] N. S. Trudinger and X. J. Wang, *On the Monge mass transfer problem*, Calculus of Variations and Partial Differential Equations **13** (2001), 19–31.

[41] C. Villani, *Topics in mass transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence, RI, 2004.

[42] C. Villani, *Optimal Transport, Old and New*, Grundlehern der Mathematischen Wissenschaften, Vol. 338, Springer-Verlag, Berlin, 2009.

[43] R. Vinter, *Optimal control*, Systems & Control: Foundations & Applications, Birkhäuser Boston Inc., Boston, MA, 2000.