Università degli Studi di Pisa

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Matematica

Tesi di Laurea

23 giugno 2006

Trasporto ottimale su varietà non compatte

Candidato

Alessio Figalli

a.figalli@sns.it

Relatore

Prof. Luigi Ambrosio

Scuola Normale Superiore

Controrelatore

Prof. Giovanni Alberti

Università di Pisa

Anno Accademico 2005/2006

Contents

Introduzione

Il problema di Monge si può formulare nella maniera seguente: date due distribuzioni di probabilità μ e ν su due spazi di misura X e Y, trovare una mappa misurabile $T : X \to Y$ tale che

$$T_\sharp \mu = \nu, \tag{0.1}$$

i.e.

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

e che T minimizzi il costo di trasporto, ossia

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

dove $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ è una funzione di costo data. Quando (0.1) è soddisfatta, si dice che T è una mappa di trasporto (*transport map*), e se inoltre T minimizza il costo la chiameremo mappa di trasporto ottimale (*optimal transport map*).

Anche in spazi Euclidei, il problema dell'esistenza di mappe di trasporto ottimali è tutt'altro che banale. Inoltre è facile costruire esempi in cui il problema di Monge è malposto semplicemente perché non c'è nessuna mappa di trasporto: questo accade per esempio se μ è una massa di Dirac e ν non lo è.

Per superare tali difficoltà, Kantorovich propose in [35], [36] una nozione di soluzione debole del problema di trasporto. L'idea è di cercare piani (*plans*) invece di mappe di trasporto, ossia misure di probabilità γ in $X \times Y$ i cui marginali sono μ e ν, i.e.

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

dove $\pi_X : X \times Y \to X$ e $\pi_Y : X \times Y \to Y$ sono le proiezioni canoniche. Indicando con $\Pi(\mu, \nu)$ la classe dei piani di traporto, il nuovo problema di minimizzazione diventa

$$C(\mu,\nu) := \min_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) \right\}. \tag{0.2}$$

Se γ è un minimo per la formulazione di Kantorovich, si dice che è un piano ottimale (*optimal plan*). Grazie alla linearità del vincolo $\gamma \in \Pi(\mu,\nu)$, l'uso di topologie deboli permette di dimostrare l'esistenza di soluzioni per (0.2): più precisamente, si può dimostrare che esiste un minimo se X e Y sono spazi Polacchi e c è semicontinuo inferiormente (vedi [49], [55, Proposition 2.1]). Il collegamento tra la formulazione di Kantorovich e quella di Monge si può vedere notando che ogni mappa di trasporto T induce un piano γ_T definito da $\gamma_T := (Id\times T)_\sharp\mu$ che è concentrato sul grafico di T, e che ha lo stesso costo:

$$\int_{X\times Y} c(x,y)\, d\gamma_T(x,y) = \int_X c(x,T(x))\, d\mu(x).$$

Il problema di mostrare l'esistenza di mappe ottimali si riduce dunque a dimostrare che un piano ottimale è concentrato su un grafico. È comunque chiaro, per quanto già detto, che senza ipotesi aggiuntive sulle misure e sul costo non si può sperare di ottenere un tale risultato. Il primo risultato di esistenza e unicità di mappe ottimali è dovuto a Brenier [9], che considera il caso $X = Y = \mathbb{R}^n$, $c(x,y) = |x-y|^2$. In tale articolo Brenier mostra come, sotto l'ipotesi che μ sia assolutamente continua rispetto alla misura di Lebesgue, esiste un'unica mappa di trasporto ottimale. Dopo il lavoro di Brenier, in molti hanno lavorato sul problema di Monge mostrando l'esistenza di un trasporto ottimale con costi più generali, sia nel caso euclideo (tra i vari Caffarelli, Evans, Gangbo, Ambrosio e Pratelli, Trudinger e Wang, McCann, Feldman), sia nel caso di varietà compatte (McCann, Bernard et Buffoni), e anche in alcuni casi particolari su varietà non compatte (Feldman e McCann). Nel tempo è diventato chiaro che la scelta del costo modifica profondamente la struttura del problema. In particolare, profondamente diversi sono i casi $c(x,y) = |x-y|^p$ con $p > 1$ rispetto al caso $c(x,y) = |x-y|$, che fu risolto nel caso euclideo nella sua formulazione più generale ben dopo il risultato di Brenier [4]. In particolare, benché anche nel caso $c(x,y) = |x-y|$ l'ipotesi di assoluta continuità di μ rispetto alla misura di Lebesgue assicuri l'esistenza di un trasporto ottimale (ipotesi che nel caso $c(x,y) = |x-y|^p$, $p > 1$, può essere un po' indebolita), non è possibile tuttavia dimostrare l'unicità del trasporto ottimale. Il problema è dovuto al fatto che la norma euclidea non è strettamente convessa, e dunque già nel caso unidimensionale semplici esempi (il cosiddetto *book-shifting*) mostrano che non ci si può aspettare nessun risultato di unicità.

La struttura della tesi è la seguente.

Nel primo capitolo introdurremo alcuni risultati standard della teoria classica del trasporto ottimale, rinviando ad alcuni libri e articoli per le dimostrazioni. Per quanto riguarda alcune definizioni e concetti meno classici, ma necessari per la comprensione dei vari risultati, questi verranno man mano introdotti nei capitoli successivi. In particolare, nel capitolo 2 saranno necessarie alcune proprietà delle funzioni semi-concave e alcuni risultati sui costi indotti da Lagrangiane nel senso di Tonelli. Le definizioni, gli enunciati e le dimostrazioni di quasi tutto ciò che verrà utilizzato sulle funzioni semi-concave e sulle Lagrangiane Tonelli sono riportati in appendice.

Nel capitolo 2, riprendendo un recente lavoro con Albert Fathi [25], si mostrerà come sia possibile dimostrare l'esistenza e l'unicità di un trasporto ottimale per un costo indotto da una Lagrangiana classica su una varietà non compatta. Questo risultato include in particolare i costi della forma $d^p(x, y)$ con $p > 1$, dove $d(x, y)$ denota una distanza Riemanniana completa. Utilizzando poi anche i risultati in [30], dove vengono adattate alcune idee di McCann [42], si generalizzerà il risultato di esistenza e unicità senza fare alcuna ipotesi di integrabilità della funzione di costo. Inoltre nel paragrafo 2.3, mostreremo come, applicando il nostro risultato di esistenza della mappa di trasporto, si possano generalizzare alcuni risultati di [18] al caso di varietà non compatte, e in particolare si riesce a dimostrare che la mappa di trasporto è approssimativamente differenziabile quasi ovunque nel caso in cui il costo è dato da $d^2(x, y)$.

Nel capitolo 3 si riprende [29], in cui vengono estesi al caso di varietà non compatte i risultati di [8], in cui gli autori mostrano l'esistenza di mappe ottimali per una larga classe di costi (che include in particolare il caso $c(x, y) = d(x, y)$) su varietà compatte senza bordo. L'esistenza di una mappa di trasporto ottimale nel caso $c(x, y) = d(x, y)$ su varietà non compatte era stata già dimostrata in [28] sotto l'ipotesi di compattezza dei supporti delle due misure. In [29] si estende tale risultato al caso di misure con supporti non necessariamente compatti e, più in generale, si dimostra l'esistenza di una mappa ottimale per una classe ben più ampia di costi, ossia la classe dei potenziali di Mañé associati a una Lagrangiana supercritica, usando in particolare risultati della "weak KAM theory" su varietà non compatte [26].

Infine, nel capitolo 4, utilizzando il risultato di esistenza e unicità del capitolo 2, riprendendo un lavoro con Cédric Villani [31], si estendono alcuni risultati di [38] in cui gli autori studiano la relazione tra le varie nozioni di tensore di Ricci e la convessità di alcuni funzionali non lineari lungo le geodetiche nello spazio di Wasserstein. In particolare, si mostrerà come le nozioni di *displacement convexity* e di *weak displacement convexity* sono tra loro equivalenti nel caso di varietà Riemanniane non compatte.

Desidero dedicare questa tesi alla mia famiglia, che sempre mi è stata vicina dall'inizio dei miei studi. Ringrazio il mio relatore, prof. Luigi Ambrosio, per la sua continua disponibilità e cordialità con cui mi ha sempre seguito, sia durante la stesura della tesi, sia soprattutto in tutti questi anni universitari. Ringrazio inoltre i professori Albert Fathi e Cédric Villani, con i quali ho collaborato durante il mio soggiorno all'École Normale Supérieure di Lione, dove ho trascorso buona parte di quest'anno e ho iniziato a lavorare sulla tesi. Desidero inoltre ringraziare tutte le persone che hanno contribuito alla mia formazione matematica in questi anni universitari, in particolare il prof. Giovanni Alberti, e tutti coloro che hanno contribuito a farmi trascorrere serenamente questi anni qui a Pisa.

Chapter 1

Preliminaries on the classical theory

Monge transportation problem is more than 200 hundred years old, see [45]. It has generated a huge amount of work. There are now several books and surveys that can help the reader through the literature, see for example among others the books [13, 3, 49, 55, 56], and the surveys [1, 20, 34].

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

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|, \tag{1.1}$$

over all bijections $T : \{x_1, \dots x_n\} \to \{y_1, \dots, y_n\}$, where $|\cdot|$ 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

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|^2. \tag{1.2}$$

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

$$T_\sharp \mu = \nu, \tag{1.3}$$

i.e.

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

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

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

where $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is a given cost function. When condition (1.3) is satisfied, we say that T is a *transport map*, and if T minimizes also the cost we call it *optimal transport map.*

Even in Euclidean spaces, the problem of the existence of optimal transport maps is far from being trivial. Furthermore, it is easy to build examples where the Monge problem is ill-posed simply because the 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 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 (1.2) above is simpler to deal with than case (1.1). 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 [56]. However for the case where the cost is a distance, like in (1.1), one should cite at least the work of Sudakov [53], Evans-Gangbo [21], Feldman-McCann [28], Caffarelli-Feldman-McCann [12], Ambrosio-Pratelli [4], and Bernard-Buffoni [8]. For the case where the cost is the square of the Euclidean or of a Riemannian distance, like in (1.2), one should cite at least the work of Knott-Smith [37], Brenier [9], [10], Rachev-Rüschendorf [48], Gangbo-McCann [33], McCann [44], and Bernard-Buffoni [7].

In order to overcome this difficulties, Kantorovich proposed in [35], [36] 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 class of plans, the new minimization problem becomes then the following:

$$C(\mu, \nu) := \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} c(x, y)\, d\gamma(x, y) \right\}. \tag{1.4}$$

If γ is a minimizer for the Kantorovich formulation, we say that it is an *optimal plan*. Due to the linearity of the constraint $\gamma \in \Pi(\mu, \nu)$, it turns out that weak topologies can be used to provide existence of solutions to (1.4): this happens for instance whenever X and Y are Polish spaces and c is lower semicontinuous (see [49], [55, Proposition 2.1]). The connection between the formulation of Kantorovich and that of Monge can be seen by noticing that any transport map T induces the plan defined by $(Id \times T)_\sharp \mu$ which is concentrated on the graph of T.

It is well-known that a linear minimization problem with convex constraints, like (1.4), admits a dual formulation. Before stating the duality formula, we make some definitions similar to that of the weak KAM theory (see [24]):

Definition 1.0.1 (**c-subsolution).** We say that a pair of Borel functions $\varphi : X \to \mathbb{R} \cup \{-\infty\}$, $\psi : Y \to \mathbb{R} \cup \{+\infty\}$, with

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

is a *c-subsolution* if

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

If a c-subsolution (φ, ψ) exists, then c_- is γ-integrable for any $\gamma \in \Pi(\mu, \nu)$ (where $c_- := \max\{0, -c\}$), and thus, integrating the above inequality on the product $X \times Y$, we get

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

Definition 1.0.2 (**Calibration).** Given an optimal plan γ, we say that a c-subsolution (φ, ψ) is (c, γ)-*calibrated* if

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

We observe that, if $\tilde{\gamma}$ is another optimal plan, then a c-subsolution is (c,γ)-calibrated iff it is $(c,\tilde{\gamma})$-calibrated, and so we will say that a pair is *c-calibrated* if it is (c,γ)-calibrated for some optimal plan γ.

Theorem 1.0.3 (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*

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

Then

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

i.e. there exists a c-subsolution (φ,ψ) which is (c,γ)-calibrated for an optimal plan γ. Moreover the (c,γ)-calibration of (φ,ψ) implies the following equality:

$$\varphi(x) - \psi(y) = c(x,y) \text{ for } \gamma\text{-almost every } (x,y).$$

For a proof of this theorem see [3, Theorem 6.1.5, page 139], [4, Theorems 3.1 and 3.2], [56, Theorem 5.9]. One of the main properties that is used to prove the above duality result, is the property of optimal plans of being c-monotone:

Definition 1.0.4 (c-cyclical monotonicity). A set $S \subset X \times Y$ is said *c-cyclically monotone* if for all $n \in \mathbb{N}$, for all set of pairs $\big((x_i,y_i)\big)_{1\le i\le n}$ and for any permutation $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$, we have

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

A transport plan $\gamma \in \Pi(\mu,\nu)$ is called *c-cyclical monotone* if there exists a c-monotone Borel set $S \subset X \times Y$ where γ is concentrated, that is $\gamma(S) = 1$.

For example, it is not difficult to prove that, if c is continuous and bounded below, then the c-cyclical monotonicity of a measure γ is a necessary condition for the optimality (see [33]). More in general, the c-cyclical monotonicity is a necessary and sufficient condition for optimality under more general assumption on the cost (see [4], [50]).

In order just to see why this fact can be true, observe that, by the above theorem, the optimality a transport plan γ is equivalent to the existence of a c-calibrated pair (φ, ψ) such that γ is concentrated on the set

$$\{(x, y) \in X \times Y \mid \varphi(x) - \psi(y) = c(x, y)\},$$

and it is simple to check that the above set is indeed c-cyclically monotone (this argument is clearly not a proof of the equivalence between optimality and cyclical monotonicity, since we are using Theorem 1.0.3 whose proof relies on the necessity of the cyclical monotonicity).

Chapter 2

Costs induced by Tonelli Lagrangians

We are now interested in studying Monge's problem on manifolds for a large class of cost functions induced by Lagrangians, and in proving the existence of an unique optimal transport map. This has been done already under the assumption that the manifold is compact (see, for instance, [7]). We want to generalize that result to arbitrary non-compact manifolds.
As we will see, 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 2.0.5 (Twist Condition)**.** For a given cost function $c(x, y)$, we define the *skew left Legendre transform* as the map

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

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

whose domain of definition is

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

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

The usefullness of these definitions will be clear in the Section 2.2, in which we will treat the case where $M = N$ and the cost is induced by a Lagrangian. This condition has appeared already in the subject. It has been known (explicitly or not) by several people, among them Gangbo and Villani (see [55, page 90]). It is used in [7], since it is always satisfied for a cost coming from a Lagrangian, as 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 $\dfrac{\partial^2 h}{\partial y \partial x} > \alpha$ everywhere. In that case both maps $\Lambda^l_h : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (x, \partial h/\partial x(x, y))$ and $\Lambda^r_h : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}, (x, y) \mapsto (y, \partial h/\partial y(x, y))$ are C^1 diffeomorphisms. The twist map $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ associated to 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 $f(x_1, v_1) = \Lambda^r_h \circ [\Lambda^l_h]^{-1}(x_1, -v_1)$, see [40] or [32].

2.1 The main theorem

The theorem that we are going to prove is the following:

Theorem 2.1.1. *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 a lower semicontinuous and bounded from below. Suppose that μ and ν are respectively (Borel) probability measures on M and N such that*

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

and let γ_c be an optimal plan for μ, ν with the cost c. If

(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 sets with σ-finite $(n-1)$-dimensional Hausdorff measure,* *then γ_c is concentrated on the graph of a measurable map $T : M \to N$.*

More precisely, if (φ, ψ) is a c-calibrating pair, we will prove that there exists a sequence of Borel subsets $B_n \subset M$, with $B_n \subset B_{n+1}, \mu(B_n) \nearrow 1$, and a sequence locally semi-concave functions $\varphi_n : M \to \mathbb{R}$, with $\varphi \leq \varphi_{n+1} \leq \varphi_n$ everywhere, and $\varphi_n = \varphi$ on B_n, such that φ_n is differentiable on B_n, and γ_c is concentrated on the graph of a map $T : M \to N$, defined uniquely μ-a.e. on B_n, thanks to the twist condition, by

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

Proof. Fix a pair (φ, ψ) c-calibrated for the measures μ and ν. Call $\tilde{N}$ the Borel subset of N where $\psi > -\infty$. Since ψ is ν-integrable, we have $\nu(N \setminus \tilde{N}) = 0$.

Define $G = \mathrm{supp}(\gamma_c)$, and consider the set

$$\tilde{G} := \{(x, y) \in G \mid \varphi(x) - \psi(y) = c(x, y)\}.$$

Since both M and N are Polish and both maps φ and ψ are Borel measurable, $\tilde{G}$ is a Borel subset of $M \times N$ of full γ_c-measure. Since N is a Polish space, by a standard result, for any Borel subset A, the measure $\nu(A)$ is the supremum of the $\nu(K)$ where K is a compact subset of A. Therefore, we can find an increasing sequence of compact subsets $(K_n) \subset \tilde{N}$ such that $\nu(K_n) \nearrow 1$ and $\psi \geq -n$ on K_n: since $\nu(\tilde{N}) = 1$, and $\tilde{N} = \cup_n\{y \in N \mid \psi(y) \geq -n\}$, it suffices to take an increasing sequence of compact sets $K_n \subset \{\psi \geq -n\} \subset \tilde{N}$ such that $\nu(\{\psi \geq -n\} \setminus K_n) \leq \frac{1}{n}$. We set

$$A_n := \pi_M\left(\tilde{G} \cap (M \times K_n)\right) \subset M,$$

where $\pi_M : M \times N \to M$ is the canonical projection. The subset A_n is μ-measurable for each n. In fact, since $\tilde{G} \cap (M \times K_n)$ is Borel and π_M continuous, this A_n is a Souslin set and is therefore μ-measurable, see [19]. We have

$$\mu(A_n) = \gamma_c\left(\pi_M^{-1}(A_n)\right) \geq \gamma_c\left(\tilde{G} \cap (M \times K_n)\right) = \gamma_c(M \times K_n) = \nu(K_n) \nearrow 1.$$

Consider

$$\varphi_n(x) := \inf_{y \in K_n} \psi(y) + c(x, y).$$

Since $\psi \geq -n$ on K_n, and c is bounded from below, we see that φ_n is bounded from below. Using now that K_n is compact, hypothesis (i), and Corollary A.1.13 of the appendix we obtain that φ_n is locally semi-concave. Since $\varphi(x) - \psi(y) \leq c(x, y)$, and $K_n \subset K_{n+1}$, we clearly have

$$\varphi_n \geq \varphi_{n+1} \geq \varphi.$$

A key observation is now the following:

$$\varphi|_{A_n} = \varphi_n|_{A_n}.$$

In fact, if $x \in A_n$, by the definition of A_n, we know that there exists a point $y_x \in K_n$ such that $(x, y_x) \in \tilde{G}$. By the definition of $\tilde{G}$, this implies

$$\varphi(x) = \psi(y) + c(x, y_x) \geq \varphi_n(x) \geq \varphi(x).$$

Since φ_n is locally semi-concave, by Theorem A.1.7 of the appendix, it is differentiable on a Borel subset C_n such that its complement C_n^c is a subset of M with σ-finite $(n-1)$-dimensional Hausdorff measure. Let us then define $C := \cap_n C_n$. The complement $C^c = \cup_n C_n^c$ is a also set with σ-finite $(n-1)$-dimensional Hausdorff measure. Observe that, by hypothesis *(iii)*, the subset C is of full μ-measure. It follows that $\mu(A_n \setminus A_n \cap C) = 0$. We can now find an increasing sequence of Borel subsets $B_n \subset A_n \cap C$ such that $\mu(A_n \setminus B_n) \to 0$. In particular $\mu(B_n) \nearrow 1$.

Since $\gamma_c(G \setminus \tilde{G}) = 0$, $A_n = \pi_M\left(\tilde{G} \cap (M \times K_n)\right)$, and $\varphi_n = \varphi$ on A_n, we conclude that over A_n, the set $\tilde{G} \cap (M \times K_n)$ coincides γ_c-a.e. with the set

$$\{(x,y) \mid x \in A_n,\ y \in K_n,\ \varphi_n(x) - \psi(y) = c(x,y)\}.$$

We now prove that $\tilde{G} \cap (M \times K_n)$ is a graph above $B_n \subset A_n = \pi_M\left(\tilde{G} \cap (M \times K_n)\right)$.

To prove this assertion, fix $x \in B_n \subset A_n$. By the definition of A_n, and what we said above, there exists $y_x \in K_n$ such that

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

Since $x \in B_n$, 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-concave and, by the definition of φ_n, for every $z \in M$, we have $\varphi_n(z) - \psi(y_x) \leq c(z, y_x)$, with equality at when $z = x$. This facts taken together imply that $\dfrac{\partial c}{\partial x}(x, y_x)$ 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 the definition of a locally semi-concave function A.1.3, 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

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

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 $\tilde{G}_x := \{y \in N \mid \varphi(x) - \psi(y) = c(x,y)\}$, we have just shown that $\{x\} \times (\tilde{G}_x \cap K_n) \subset \mathcal{D}(\Lambda_c^l)$ for each $x \in B_n$, and also $\frac{\partial c}{\partial x}(x,y) = d_x\varphi_n$, for every $y \in \tilde{G}_x \cap K_n$. Recalling now that that, by hypothesis (ii), the cost c satisfies the left twist condition, we obtain that $\tilde{G}_x \cap K_n$ is reduced to a single element which is uniquely characterized by the equality

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

and so we have proved that $\tilde{G} \cap (M \times K_n)$ is a graph over B_n. Since $B_{n+1} \supset B_n$, and $K_n \subset K_{n+1}$, we can conclude that $\tilde{G} \cap (M \times (\cup_n K_n))$ is a graph over $\cup_n B_n$. Note that $\cup_n B_n$ is of full μ-measure for μ, since $\mu(B_n) \nearrow 1$. This concludes the proof that γ_c is concentrated on a graph. □

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 [3, Definition 5.5.1, page 129]:

Definition 2.1.2 (Approximate differential). We say that $f : M \to \mathbb{R}^m$ has an approximate differential at $x \in M$ if there exists a function $h : M \to \mathbb{R}^m$ 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 $\{f = h\}$.

We recall that many standard properties of the differential, for example linearity and additivity, still hold for the approximate differential. In particular, it is simple to check that the property of being approximatively differentiable is stable by right composition with smooth maps (say C^1), and, in this case, the standard chain rule formula for the differentials holds. Moreover we remark that it makes sense to speak of approximate differential for maps between manifolds.
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

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

have density 0 at x for each $\varepsilon > 0$ with respect to the Lebesgue measure. This last definition is the one systematically used in [27]. On the other hand, for the purpose of this paper, Definition 2.1.2 is more convenient.

Corollary 2.1.3. *Under the hypothesis of Theorem 2.1.1, if we assume that μ is absolutely continuous for Lebesgue measure (this is stronger than condition (iii) of Theorem 2.1.1), 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.

Proof. We will use the notations and the proof of Theorem 2.1.1. We denote by $D_n \subset B_n$ 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 $\lambda(B_n \setminus D_n) = 0$. Since μ is absolutely continuous with respect to Lebesgue measure, we have $\mu(D_n) = \mu(B_n)$, and therefore $\cup_n D_n$ is of full μ-measure, since $\mu(B_n) \nearrow 1$. 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 D_n with $\tilde{d}_x\varphi = d_x\varphi_n$. □

2.2 Costs obtained from Lagrangians

Now that we have proved the theorem, 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.1.4 of the appendix for the definition of a Tonelli Lagrangian). The cost $c_L : M \times M \to \mathbb{R}$ associated to L is given by

$$c_L(x, y) := \inf_{\gamma(0)=x, \gamma(1)=y} \int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt,$$

where the infimum is taken over all the continuous piecewise C^1 curves $\gamma : [0, 1] \to M$, with $\gamma(0) = x$, and $\gamma(1) = y$ (see Definition B.2.1 of the appendix).

Proposition 2.2.1. *If $L : TM \to \mathbb{R}$ is a Tonelli Lagrangian on the connected manifold M, then the cost $c_L : M \times M \to \mathbb{R}$ associated to the Lagrangian L is continuous, bounded from below, and satisfies condition* (i) *and* (ii) *of Theorem 2.1.1.*

Proof. Observe that L is bounded below by C, hence the cost c_L is also bounded below. By Theorem B.2.2 of the appendix, the cost c_L is locally semi-convex, and therefore continuous. Moreover, we can now apply Proposition A.1.15 of the appendix to conclude that c_L satisfies condition (i) of Theorem 2.1.1.

It remains to verify the twist condition (ii) of Theorem 2.1.1. To simplify notations, we will write $c(x, y)$ instead of $c_L(x, y)$. For $y \in M$ we define $c_y : M \to \mathbb{R}$ by $c_y(x) = c(x, y)$. Given $x, y \in M$, by Theorem B.1.5 of the appendix, can find a C^1 curve $\gamma_{x,y} : [0, 1] \to M$, with $\gamma_{x,y}(0) = x, \gamma_{x,y}(1) = y$ such that

$$c(x, y) = \int_0^1 L(\gamma_{x,y}(t), \dot{\gamma}_{x,y}(t))\, dt.$$

then by Corollary B.2.3 of the appendix

$$-\frac{\partial L}{\partial v}(\gamma_{x,y}(0), \dot{\gamma}_{x,y}(0)) = -\frac{\partial L}{\partial v}(x, \dot{\gamma}_{x,y}(0)) \in D^+_x c_y,$$

therefore, for all $(x, y) \in \mathcal{D}(\Lambda^l_c)$, we have

$$\frac{\partial c}{\partial x}(x, y) = -\frac{\partial L}{\partial v}(x, \dot{\gamma}_{x,y}(0)). \tag{2.1}$$

We know prove the the injectivity of Λ^l_c on $D(\Lambda^l_c)$. Suppose $(x_1, y_1), (x_2, y_2) \in D(\Lambda^l_c)$ are such that $\Lambda^l_c(x_1, y_1) = \Lambda^l_c(x_2, y_2)$, then we have

$$x_1 = x_2 \quad \text{and} \quad \frac{\partial c}{\partial x}(x_1, y_1) = \frac{\partial c}{\partial x}(x_2, y_2).$$

Calling x the common value $x_1 = x_2$, we get from Equation (2.1) that

$$\frac{\partial L}{\partial v}(x, \dot{\gamma}_{x,y_1}(0)) = \frac{\partial L}{\partial v}(x, \dot{\gamma}_{x,y_2}(0)).$$

The strict convexity condition (b) implies that for a given $x \in M$, the map $T_x M \to T^*_x M, v \mapsto \partial L/\partial v(x, v)$, is injective. Therefore we obtain that

$$\dot{\gamma}_{x,y_1}(0) = \dot{\gamma}_{x,y_2}(0).$$

By classical Calculus of Variations, both curves γ_{x,y_1} and γ_{x,y_2} satisfy the same second order differential equation on M. They are therefore equal, because at 0 they have the same value and the same speed. This implies $y_1 = \gamma_{x,y_1}(1) = \gamma_{x,y_2}(1) = y_2$. ☐

We want now to consider the special case in which M is a Riemannian manifold and $L(x, v) = \|v\|_x^p$, with $p > 1$. Obviously this Lagrangian is strictly convex and superlinear in v but not C^2. However, for this particular kind of Lagrangian, it can be proved that the associated cost satisfies the left twist condition (see Proposition B.3.3).

So we know that we can take $L(\gamma(t), \dot\gamma(t), t) = \frac{1}{p}\|\dot\gamma(t)\|_{\gamma(t)}^p$ with $p > 1$ and we get the existence of a unique transport map for the cost function $c(x, y) = \frac{1}{p}d^p(x, y)$. In the particular case when $p = 2$, φ is a $\frac{d^2}{2}$-convex function and the transport map is given by

$$\exp_x[-\tilde\nabla_x\varphi],$$

where $\tilde\nabla_x\varphi$ denotes the approximate gradient of φ at x, which simply corresponds to the element of T_xM obtained from $\tilde d_x\varphi$ using the isomorphism with T_x^*M induced by the Riemannian metric. In fact, by the formula above, if $y = \exp_x(w)$ with $w \in T_xM$, we have

$$\nabla_x c(x, y) = \nabla_x[\frac{1}{2}d^2(x, y)] = -\nabla_v \frac{1}{2}\|\dot\gamma_{x,y}(0)\|_x^2 = -\dot\gamma_{x,y}(0) = -w = -\exp_x^{-1}(y).$$

2.2.1 The interpolation and its absolute continuity

Let μ_0 be an absolutely continuous probability measure, μ_1 a probability measure, and let $T : M \to M$ be the unique optimal transport between μ_0 and μ_1. We recall that T is given by the formula

$$T(x) = [\frac{\partial c}{\partial x}(x, \cdot)]^{-1}(\tilde d_x\varphi),$$

where (φ, ψ) is a c-calibrated pair for the dual problem between μ_0 and μ_1. Moreover we recall that we can assume that φ is given by

$$\varphi(x) = \inf_{y\in M} \psi(y) + c(x, y).$$

In fact, if it is not the case, it suffices to replace φ by

$$\overline\varphi(x) := \inf_{y\in N} \psi(y) + c(x, y). \tag{2.2}$$

Observe that $\overline\varphi$ is measurable as it is upper semicontinuous. In fact, each function

$$x \mapsto c(x, y)$$

is locally semi-concave and it is therefore continuous. It follows that $\overline{\varphi}$ is an infimum of continuous functions, hence it is upper semicontinuous.
We obviously have $\overline{\varphi} \leq \varphi < +\infty$ and $\overline{\varphi}(x) - \psi(y) \leq c(x,y)$. Moreover $\overline{\varphi} \leq \varphi$ implies $\overline{\varphi}_- \leq \varphi_-$ and therefore

$$\int_M \overline{\varphi}_- \, d\mu \leq \int_M \varphi_- \, d\mu < +\infty,$$

that is $\overline{\varphi}_- \in L^1(d\mu)$. As $\overline{\varphi}_+(x) - \overline{\varphi}_-(x) - \psi(y) \leq c(x,y)$,

$$\int_M \overline{\varphi}_+ \, d\mu \leq \int_M \overline{\varphi}_- \, d\mu + \int_N \psi \, d\nu + \int_{M\times N} c \, d\gamma_c < +\infty,$$

that tell us that $\overline{\varphi} \in L^1(d\mu)$ (this fact can be proved also observing that $\overline{\varphi} = \varphi$ μ-a.e. (see below) but we preferred to write this proof as it does not use the fact that (φ, ψ) is optimal). So $(\overline{\varphi}, \psi)$ is still a c-subsolution and, as $\varphi \leq \overline{\varphi}$, it is obviously (c, γ_c)-calibrated.

By what we said in the paragraph above, in the case of costs induced by Lagrangians the formula for $T(x)$ implies that

$$\dot{\gamma}_{x,T(x)}(0) = [\frac{\partial L}{\partial v}(x, \cdot)]^{-1}(-\tilde{d}_x\varphi),$$

where $\gamma_{x,T(x)} : [0,1] \to M$ is a curve from x to $T(x)$ that minimizes the action. We now make the following important remark, that we will need also in the sequel:

Remark 2.2.2. *We observe that, for μ_0-a.e. x, there exists an unique curve from x to $T(x)$ that minimizes the action. In fact, since $\frac{\partial c}{\partial x}(x,y)$ exists at $y = T(x)$ for μ_0-a.e. x, the twist conditions proved in Section 2.2 tells us that its velocity at time 0 is μ_0-a.e. univocally determined.*

The curve γ can be written using the Euler-Lagrange flow $\phi_t : TM \to TM$

$$\gamma_{x,T(x)}(t) = \pi \circ \phi_t(x, \dot{\gamma}_{x,T(x)}(0)),$$

where $\pi : TM \to M$ denotes the canonical projection. We can so define $T_t : M \to M$ as

$$T_t(x) := \gamma_{x,T(x)}(t). \tag{2.3}$$

We observe that $T_0 = \mathrm{id}$ and $T_1 = T$. This allows us to construct an interpolation between μ_0 and μ_1 as

$$\mu_t := (T_t)_\sharp \mu_0.$$

We want to prove that μ_t is absolutely continuous for all $t \in [0,1)$ and that T_t is the unique optimal transport from μ_0 to μ_t for the cost

$$c_t(x,y) := \inf_{\gamma(0)=x,\ \gamma(t)=y} \int_0^t L(\gamma(s),\dot{\gamma}(s))\, ds.$$

In order to prove that μ_t is absolutely continuous for all $t \in (0,1)$, it suffices to prove that the map T_t admits an inverse which is locally Lipschitz. In fact, this would implies that, if μ_t has a singular part with respect to the Lebesgue measure, also $(T_t^{-1})_\sharp \mu_t = \mu_0$ has a singular part, which is absurd. Let us consider an increasing sequence K_n of compact sets such that $\cup_n K_n = M$. We define $A_n = K_n \cap T^{-1}(K_n) \cap B_n$, where B_n was defined in Theorem 2.1.1. In this way we have that A_n is contained in a compact subset, $\mu(A_n) \nearrow 1$ and that, when $x \in A_n$, the image $T(x)$ varies in a compact subset. We will prove that $T_t|_{A_n}$ admits an inverse which is locally Lipschitz. This will implies, as A_n is increasing, that $T_t|_{\cup_n A_n}$ is invertible with a locally Lipschitz inverse, and this will prove the result, as $\mu_0(\cup_n A_n) = 1$.
Let us define the two following semigroups:

$$S_t^- u(x) := \inf_{y \in M} \{u(y) + c_t(y,x)\}, \quad S_t^+ u(x) := \sup_{y \in M} \{u(y) - c_t(x,y)\}.$$

We observe that, with these notations,

$$\varphi(x) = S_1^- \psi(x). \tag{2.4}$$

Moreover we obviously have the property

$$c_{t+s}(x,y) = \min_{z \in M} c_t(x,z) + c_s(z,y) \quad \forall t,s \geq 0. \tag{2.5}$$

So, by (2.4), we have

$$\varphi(x) \leq \psi(y) + c_1(x,y) \leq \psi(y) + c_t(x,z) + c_{1-t}(z,y) \quad \forall x,y,z \in M,$$

that is

$$\varphi(x) - c_t(x,z) \leq \psi(y) + c_{1-t}(z,y) \quad \forall x,y,z \in M. \tag{2.6}$$

Fix now $t \in (0,1)$ and take $x \in A_n$. As $t \mapsto T_t(x)$ is the unique curve from x to $T(x)$ which minimize the action, we have

$$c(x,T(x)) = c_t(x,T_t(x)) + c_{1-t}(T_t(x),T(x)).$$

This tells us that in (2.6) we have equality for $y = T(x)$ and $z = T_t(x)$. So we have, for all $z \in M$,

$$\begin{aligned} f_x(z) := \varphi(x) - c_t(x,z) \le \varphi(x) - c_t(x, T_t(x)) = S_t^+\varphi(T_t(x)) = S_{1-t}^-\psi(T_t(x)) \\ = \psi(T(x)) + c_{1-t}(T_t(x), T(x)) \le \psi(T(x)) - c_{1-t}(z, T(x)) =: F_{T(x)}(z). \end{aligned}$$

So we have that the function $x \mapsto \varphi(x) - c_t(x, T_t(x)) = S_t^-\varphi(T_t(x))$ is below the semi-concave function $F_{T(x)}(z)$ and above the semi-convex function $f_x(z)$. This tells us that $T_t^-\varphi$ is differentiable at the point $T_t(x)$ (this is a simple consequence of the definition of semi-concave and semi-convex functions, see appendix). Moreover, as the Lagrangian is C^2, the modulus of semi-concavity (resp. semi-convexity) of $F_{T(x)}(z)$ (resp. $f_x(z)$) is linear, that is of the form $\omega(\|\cdot\|) = K\|\cdot\|$ (see appendix, Theorem B.2.2). Moreover, as $x \in A_n$, also $T(x)$ varies in a compact. So we can find a bound uniform in x for the modulus of semi-concavity (resp. semi-convexity) of $F_{T(x)}(z)$ (resp. $f_x(z)$), always of the form $K\|\cdot\|$. This fact tells us that $T_t(x) \mapsto d_{T_t(x)}(S_t^-\varphi)$ is a Lipschitz function (see appendix, Theorem A.1.16). Now we recall that

$$d_{T_t(x)}(S_t^-\varphi) = \frac{\partial c_t}{\partial y}(x, T_t(x)) = \frac{\partial L}{\partial v}(T_t(x), \dot{T}_t(x))$$

(see [24] and appendix, Corollary B.2.3). This equation first tells us that all the curves $t \mapsto T_t(x)$ cannot intersect at a certain time $t \in (0,1)$, as their velocity at the time t is univocally determined by $d_{T_t(x)}(S_t^-\varphi)$, and so the map $x \mapsto T_t(x)$ is invertible. Moreover, as the Legendre transform is a C^1 diffeomorphism, we have

$$(T_t(x), \dot{T}_t(x)) = \mathscr{L}^{-1}(T_t(x), d_{T_t(x)}(S_t^-\varphi)).$$

This tells us that T_t^{-1} is given by

$$\begin{aligned} T_t^{-1}(T_t(x)) &= \pi \circ \phi_{-t} \circ \mathscr{L}^{-1}(T_t(x), d_{T_t(x)}[S_t^-\varphi]) \\ &= \pi^* \circ \phi_{-t}^* \circ (T_t(x), d_{T_t(x)}[S_t^-\varphi]), \end{aligned} \tag{2.7}$$

where ϕ_t^* is the Hamiltonian flow and $\pi^* : T^*M \to M$ is the canonical projection. So, as ϕ_t^* is of class C^1, we conclude that T_t^{-1} is a locally Lipschitz function.
We observe that, by (2.7), it follows that also the map $T \circ T_t^{-1} : T_t(x) \mapsto T(x)$ is locally Lipschitz. In fact, it is given by the formula

$$T \circ T_t^{-1}(T_t(x)) = \pi^* \circ \phi_{1-t}^*(T_t(x), d_{T_t(x)}[S_t^-\varphi]).$$

Remark: now that we know that μ_t is absolutely continuous for all $t \in [0,1)$, we know that there exists an optimal transport map $\bar{T}_t$ from μ_t to μ_1 for the cost $c_{1-t}(x,y)$. Now it is simple to prove that T_t is optimal for the transport from μ_0 to μ_t for the cost $c_t(x,y)$, T_t^{-1} is optimal for the transport from μ_t to μ_0 for the cost $c_t^-(x,y) := c_t(y,x)$ and that $\bar{T}_t = T \circ T_t^{-1}$. Indeed, let us define

$$C_t(\mu,\nu) := \min_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{M \times M} c_t(x,y)\, d\gamma(x,y) \right\}$$

(see (1.4)). By (2.5) it follows that

$$\begin{aligned}
\int_M c(x,T(x))\, d\mu_0(x) &= C_1(\mu_0,\mu_1) \le C_t(\mu_0,\mu_t) + C_{1-t}(\mu_t,\mu_1) \\
&\le \int_M c_t(x,T_t(x))\, d\mu_0(x) + \int_M c_{1-t}(x, T \circ T_t^{-1}(x))\, d\mu_t(x) \\
&= \int_M c_t(x,T_t(x))\, d\mu_0(x) + \int_M c_{1-t}(T_t(x), T(x))\, d\mu_0(x) \\
&= \int_M c(x,T(x))\, d\mu_0(x),
\end{aligned}$$

which implies the optimality of T_t and $T \circ T_t^{-1}$, as wanted. Finally, the optimality of T_t^{-1} follows by

$$\begin{aligned}
C_t(\mu_t,\mu_0) = C_t(\mu_0,\mu_t) &= \int_M c_t(x,T_t(x))\, d\mu_0(x) \\
&= \int_M c_t(T_t^{-1}(x), x)\, d\mu_t(x) = \int_M c_t^-(x, T_t^{-1}(x))\, d\mu_t(x).
\end{aligned}$$

2.3 The Wasserstein space W_2

Let (M,g) be a smooth complete Riemannian manifold, equipped with its geodesic distance d and its volume measure vol. We denote with $P_2(M)$ the set of probability measures on M with finite 2-order moment, that is

$$\int_M d^2(x,x_0)\, d\mu(x) < +\infty \quad \text{for a certain } x_0 \in M.$$

We remark that, by the triangle inequality for d, the definition does not depends on the point x_0. The space $P_2(M)$ can be endowed of the so called Wasserstein distance W_2:

$$W_2^2(\mu_0,\mu_1) := \min_{\gamma\in\Pi(\mu_0,\mu_1)} \left\{ \int_{M\times M} d^2(x,y)\, d\gamma(x,y) \right\}.$$

The quantity W_2 will be called the *Wasserstein distance of order 2* between μ_0 and μ_1. It is well-known that it defines a finite metric on $P_2(M)$, and so one can speak about geodesic in the metric space (P_2, W_2). This space turns out, indeed, to be a length space (see for example [55], [56]). We denote with $P_2^{ac}(M)$ the subset of $P_2(M)$ that consists of the Borel probability measures on M that are absolutely continuous with respect to vol.

By all the result proved before, it is simple to prove the following:

Proposition 2.3.1. *$P_2^{ac}(M)$ is a geodesically convex subset of $P_2(M)$. Moreover, if $\mu_0,\mu_1 \in P_2^{ac}(M)$, then there is a unique Wasserstein geodesic $\{\mu_t\}_{t\in[0,1]}$ joining μ_0 to μ_1, which is given by*

$$\mu_t = (T_t)_\sharp \mu_0 := (\exp[-t\tilde{\nabla}\varphi])_\sharp \mu_0,$$

where $T(x) = \exp_x[-\tilde{\nabla}_x\varphi]$ is the unique transport map from μ_0 to μ_1 which is optimal for the cost $\frac{1}{2}d^2(x,y)$ (and so also optimal for the cost $d^2(x,y)$). Moreover:

(i) T_t is the unique optimal transport map from μ_0 to μ_t for all $t \in [0,1]$;

(ii) T_t^{-1} the unique optimal transport map from μ_t to μ_0 for all $t \in [0,1]$ (and, if $t \in [0,1)$, it is locally Lipschitz);

(iii) $T \circ T_t^{-1}$ the unique optimal transport map from μ_t to μ_1 for all $t \in [0,1]$ (and, if $t \in (0,1]$, it is locally Lipschitz).

Since we know that the transport is unique, the proof is quite standard. However, for completeness, we give all the details.

Proof. Let $\{\mu_t\}_{t\in[0,1]}$ be a Wasserstein geodesic joining μ_0 to μ_1. Fix $t \in (0,1)$, and let γ_t (resp. $\hat{\gamma}_t$) be an optimal transport plan between μ_0 and μ_t (resp. μ_t and μ_1) (in effect, we know that γ_t is a graph and it is unique, but we will not use this fact). We now define the probability measure on $M \times M \times M$

$$\lambda_t(dx,dy,dz) := \int_M \gamma_t(dx|y) \times \hat{\gamma}_t(dz|y)\, d\mu_t(y),$$

where $\gamma_t(dx,dy) = \int_M \gamma_t(dx|y)\,d\mu_t(y)$ and $\hat\gamma_t(dy,dz) = \int_M \hat\gamma_t(dz|y)\,d\mu_t(y)$ are the disintegrations of γ_t and $\hat\gamma_t$ with respect to μ_t. Then, if we define

$$\tilde\gamma_t := \pi^{1,3}_\sharp \lambda_t,$$

it is simple to check that $\tilde\gamma_t$ is a transport plan from μ_0 to μ_1. Now, since $\{\mu_t\}_{t\in[0,1]}$ is a geodesic, we have that

$$\begin{aligned} W_2(\mu_0,\mu_1) &\le \|d(x,z)\|_{L^2(\tilde\gamma_t,M\times M)} = \|d(x,z)\|_{L^2(\lambda_t,M\times M\times M)} \\ &\le \|d(x,y)\|_{L^2(\lambda_t,M\times M\times M)} + \|d(y,z)\|_{L^2(\lambda_t,M\times M\times M)} \\ &= \|d(x,y)\|_{L^2(\gamma_t,M\times M)} + \|d(y,z)\|_{L^2(\hat\gamma_t,M\times M)} \\ &= W_2(\mu_0,\mu_t) + W_2(\mu_t,\mu_1) = W_2(\mu_0,\mu_1). \end{aligned} \tag{2.8}$$

This proves that $\tilde\gamma_t$ is an optimal transport plan between μ_0 and μ_1, which implies that $\tilde\gamma_t$ is supported on the graph of T. Moreover, since in (2.8) all the inequalities are indeed equalities, we get that

$$d(x,z) = d(x,y) + d(y,z) \quad \text{for } \lambda_t\text{-a.e. } (x,y,z) \in M\times M\times M$$

that is, y is on a geodesic from x to z. Moreover, since $W_2(\mu_0,\mu_t) = tW_2(\mu_0,\mu_1)$, we also have

$$d(x,y) = td(x,z), \quad d(y,z) = (1-t)d(x,z) \quad \text{for } \lambda_t\text{-a.e. } (x,y,z) \in M\times M\times M.$$

Since, by Remark 2.2.2, the geodesic from x to $T(x)$ is unique for μ_0-a.e. x, we conclude that λ is concentrated on the subset $\{(x,T_t(x),T(x))\}_{x\in\mathrm{supp}(\mu_0)}$, which implies that $\mu_t = (T_t)_\sharp\mu_0$. We now observe that

$$T_t(x) := \exp_x[-t\tilde\nabla_x\varphi]$$

is exactly (2.3) in the particular case $c(x,y) = \frac{t}{2}d^2(x,y)$. Moreover we see that $\mu_t := (T_t)_\sharp\mu_0 \in P_2^{ac}(M)$. In fact,

$$\begin{aligned} \int_M d^2(x,x_0)\,d\mu_t(x) &= \int_M d^2(T_t(x),x_0)\,d\mu_0(x) \\ &\le 2\int_M \left[d^2(x,x_0) + d^2(x,T_t(x))\right]\,d\mu_0(x) \\ &\le 2\int_M \left[d^2(x,x_0) + d^2(x,T(x))\right]\,d\mu_0(x) \\ &\le 4\int_M \left[d^2(x,x_0) + d^2(x_0,T(x))\right]\,d\mu_0(x) \\ &= 4\int_M d^2(x,x_0)\,d\mu_0(x) + 4\int_M d^2(x_0,y)\,d\mu_1(y) < +\infty, \end{aligned}$$

and the result in Section 2.2.1 tells us that μ_t is absolutely continuous. Using the notation of Section 2.2, we have

$$c_t(x,y) = \inf_{\gamma(0)=x,\ \gamma(t)=y} \int_0^t \frac{1}{2}\|\dot\gamma(s)\|^2_{\gamma(s)}\, ds = \frac{1}{2t}d^2(x,y).$$

So, thanks to the remark given at the end of the Section 2.2.1, we know that T_t and T_t^{-1} are optimal for the cost function $\frac{1}{2t}d^2(x,y)$, and $T\circ T_t^{-1}$ is optimal for the cost function $\frac{1}{2(1-t)}d^2(x,y)$. This obviously implies that T_t, T_t^{-1} and $T\circ T_t^{-1}$ are optimal also for the cost $d^2(x,y)$. □

The above result tells us that also $(P_2^{ac}(M), W_2)$ is a length space.

2.3.1 Regularity, concavity estimate and a displacement convexity result

We now consider the cost function $c(x,y) = \frac{1}{2}d^2(x,y)$. Let $\mu,\nu \in P_2^{ac}(M)$ and let us denote with f and g their respective densities. Let

$$T(x) = \exp_x[-\tilde\nabla_x \varphi]$$

be the unique optimal transport map from μ to ν (see Section 2.2).
We recall that a locally semi-concave function with linear modulus admits vol-a.e. a second order Taylor expansion (see [6], [18]). Let us recall the definition of approximate hessian:

Definition 2.3.2 (Approximate hessian). We say that $f : M \to \mathbb{R}^m$ has a *approximate hessian* at $x\in M$ if there exists a function $h : M \to \mathbb{R}$ such that the set $\{f = h\}$ has density 1 at x with respect to the Lebesgue measure and h admits a second order Taylor expansion at x, that is there exists a self-adjoint operator $H : T_xM \to T_xM$ such that

$$h(\exp_x w) = h(x) + \langle \nabla_x h, w\rangle + \frac{1}{2}\langle Hw, w\rangle + o(\|w\|_x^2).$$

In this case the approximate hessian is defined as $\tilde\nabla_x^2 f := H$.

Like in the case of the approximate differential, it is not difficult to show that this definition makes sense.
Observing that $d^2(x,y)$ is locally semi-concave with linear modulus (see Theorem

B.2.2), we get that φ_n is locally semi-concave with linear modulus for each n. Thus we can define μ-a.e. an approximate hessian for φ (see Definition 2.3.2):

$$\tilde{\nabla}_x^2 \varphi := \nabla_x^2 \varphi_n \quad \text{for } x \in D_n \cap E_n,$$

where D_n was defined in the proof of Corollary 2.1.3, E_n denotes the full μ-measure set of points where φ_n admits a second order Taylor expansion, and $\nabla_x^2 \varphi_n$ denotes the self-adjoint operator on T_xM that appears in the Taylor expansion on φ_n at x. Let us now consider, for each set D_n, an increasing sequence of compact sets $K_m^n \subset D_n$ such that $\mu(D_n \setminus \cup_m K_m^n) = 0$. We now define the measures $\mu_m^n := \mu \llcorner K_m^n$ and $\nu_m^n := T_\sharp \mu_m^n = (\exp[-\nabla \varphi_n])_\sharp \mu_m^n$, and we renormalize them in order to obtain two probability measures:

$$\hat{\mu}_m^n := \frac{\mu_m^n}{\mu_m^n(M)} \in P_2^{ac}(M), \quad \hat{\nu}_m^n := \frac{\nu_m^n}{\nu_m^n(M)} = \frac{\nu_m^n}{\mu_m^n(M)} \in P_2^{ac}(M).$$

We now observe observe that T is still optimal. In fact, if is not that case, we would have

$$\int_{M \times M} c(x, S(x))\, d\hat{\mu}_m^n(x) < \int_{M \times M} c(x, T(x))\, d\hat{\mu}_m^n(x)$$

for a certain S transport map from $\hat{\mu}_m^n$ to $\hat{\nu}_m^n$. This would imply that

$$\int_{M \times M} c(x, S(x))\, d\mu_m^n(x) < \int_{M \times M} c(x, T(x))\, d\mu_m^n(x),$$

and so the transport map

$$\tilde{S}(x) := \begin{cases} S(x) & \text{if } x \in K_m^n \\ T(x) & \text{if } x \in M \setminus K_m^n \end{cases}$$

would have a cost strictly less than the cost of T, which would contradict the optimality of T.
We will now apply the results of [18] to the compactly supported measures $\hat{\mu}_m^n$ and $\hat{\nu}_m^n$, in order to get information on the transport problem from μ to ν. In the sequel we will denote by $\nabla_x d_y^2$ and by $\nabla_x^2 d_y^2$ respectively the gradient and the hessian with respect to x of $d^2(x, y)$, and by $d_x \exp$ and $d(\exp_x)_v$ the two components of the differential of the map $TM \ni (x, v) \mapsto \exp_x[v] \in M$ (whenever they exist). By [18, Theorem 4.2], we get the following:

Theorem 2.3.3 (Jacobian identity a.e.). *There exists a subset $E \subset M$ such that $\mu(E) = 1$ and, for each $x \in E$, $Y(x) := d(\exp_x)_{-\tilde\nabla_x\varphi}$ and $H(x) := \frac{1}{2}\nabla_x^2 d^2_{T(x)}$ both exist and we have*

$$f(x) = g(T(x)) \det[Y(x)(H(x) - \tilde\nabla_x^2\varphi)] \neq 0.$$

Proof. It suffices to observe that [18, Theorem 4.2] applied to $\hat\mu_m^n$ and $\hat\nu_m^n$ gives that, for μ-a.e. $x \in K_m^n$,

$$\frac{f(x)}{\mu_m^n(M)} = \frac{g(T(x))}{\mu_m^n(M)} \det[Y(x)(H(x) - \nabla_x^2\varphi_n)] \neq 0,$$

which implies

$$f(x) = g(T(x)) \det[Y(x)(H(x) - \tilde\nabla_x^2\varphi)] \neq 0 \quad \text{for } \mu\text{-a.e. } x \in K_m^n.$$

Passing to the limit as $m, n \to +\infty$ we get the result. □

We can so define μ-a.e. the (weak) differential of the transport map at x as

$$d_xT := Y(x)(H(x) - \tilde\nabla_x^2\varphi).$$

Let us prove now that, indeed, $T(x)$ is approximately differentiable μ-a.e., and that the above differential coincides with the approximate differential of T. In order to prove this fact, let us first make a formal computation. Observe that, since the map $x \mapsto \exp_x[-\frac{1}{2}\nabla_x d_y^2] = y$ is constant, we have

$$0 = d_x(\exp_x[-\frac{1}{2}\nabla_x d_y^2]) = d_x \exp[-\frac{1}{2}\nabla_x d_y^2] - d(\exp_x)_{-\frac{1}{2}\nabla_x d_y^2}(\frac{1}{2}\nabla_x^2 d_y^2) \quad \forall y \in M.$$

By differentiating (in the approximate sense) the equality $T(x) = \exp[-\tilde\nabla_x\varphi]$ and recalling the equality $\tilde\nabla_x\varphi = \frac{1}{2}\nabla_x d^2_{T(x)}$ (proved in Theorem 2.1.1 in the case of general cost functions), we obtain

$$\begin{aligned}
\tilde d_xT &= d(\exp_x)_{-\tilde\nabla_x\varphi}(-\tilde\nabla_x^2\varphi) + d_x \exp[-\tilde\nabla_x\varphi] \\
&= d(\exp_x)_{-\tilde\nabla_x\varphi}(-\tilde\nabla_x^2\varphi) + d(\exp_x)_{-\frac{1}{2}\nabla_x d^2_{T(x)}}(\frac{1}{2}\nabla_x^2 d^2_{T(x)}) \\
&= d(\exp_x)_{-\tilde\nabla_x\varphi}(H(x) - \tilde\nabla_x^2\varphi),
\end{aligned}$$

as wanted. In order to make the above proof rigorous, it suffices to observe that for μ-a.e. x, $T(x) \notin cut(x)$, where $cut(x)$ is defined as the set of points $z \in M$ which cannot be linked to x by an extendable minimizing geodesic. Indeed we recall that the square of the distance fails to be semi-convex at the cut locus, that is, if $x \in cut(y)$, then

$$\inf_{0<\|v\|_x<1} \frac{d_y^2(\exp_x[v]) - 2d_y^2(x) + d_y^2(\exp_x[-v])}{|v|^2} = -\infty$$

(see [18, Proposition 2.5]). Fix now $x \in D_n$. Since we know that $\frac{1}{2}d^2(z, T(x)) \geq \varphi_n(z) - \psi(T(x))$ with equality for $z = x$, we obtain a bound from below of the Hessian of $d^2_{T(x)}$ at x in term of the Hessian of φ_n at x (see the proof of [18, Proposition 4.1(a)]). Thus, since each φ_n admits vol-a.e. a second order Taylor expansion, we obtain that, for μ-a.e. x,

$$x \notin cut(T(x)), \quad \text{or equivalently} \quad T(x) \notin cut(x).$$

This implies that all the computations we made above in order to prove the formula for $\tilde{d}_x T$ are correct. Indeed the exponential map $(x, v) \mapsto \exp_x[v]$ is smooth if $\exp_x[v] \notin cut(x)$, the function d_y^2 is smooth around any $x \notin cut(y)$ (see [18, Paragraph 2]), and $\tilde{\nabla}_x\varphi$ is approximatively differentiable μ-a.e. Thus, recalling that, once we consider the right composition of an approximatively differentiable map with a smooth map, the standard chain rule holds (see the remarks after Definition 2.1.2), we have proved the following regularity result for the transport map:

Proposition 2.3.4 (Approximate differentiability of the transport map)**.** *The transport map is approximatively differentiable for μ-a.e. x and its approximate differential is given by the formula*

$$\tilde{d}_x T = Y(x)(H(x) - \tilde{\nabla}_x^2\varphi),$$

where Y and H are defined in Theorem 2.3.3.

For proving our displacement convexity result, it will be useful the following change of variables formula.

Proposition 2.3.5 (Change of variables for optimal maps)**.** *If $A : [0 + \infty) \to \mathbb{R}$ is a Borel function such that $A(0) = 0$, then*

$$\int_M A(g(y))\, d\,\mathrm{vol}(y) = \int_E A\left(\frac{f(x)}{J(x)}\right) J(x)\, d\,\mathrm{vol}(x),$$

where $J(x) := \det[Y(x)(H(x) - \tilde{\nabla}_x^2\varphi)] = \det[\tilde{d}_x T]$ (either both integrals are undefined or both take the same value in $\overline{\mathbb{R}}$).

The proof follows by the Jacobian identity proved in Theorem 2.3.3 exactly as in [18, Corollary 4.7].

Let us now define for $t \in [0,1]$ the measure $\mu_t := (T_t)_\sharp \mu$, where

$$T_t(x) = \exp_x[-t\tilde{\nabla}_x \varphi].$$

By the results in Section 2.2.1 and Proposition 2.3.1, we know that T_t coincides with the unique optimal map pushing μ forward to μ_t, and that μ_t is absolutely continuous with respect to vol for any $t \in [0,1]$.
Given $x, y \in M$, following [18], we define for $t \in [0,1]$

$$Z_t(x,y) := \{z \in M \mid d(x,z) = td(x,y) \text{ and } d(z,y) = (1-t)d(x,y)\}.$$

If now N is a subset of M, we set

$$Z_t(x,N) := \cup_{y \in N} Z_t(x,y).$$

Letting $B_r(y) \subset M$ denote the open ball of radius $r > 0$ centered at $y \in M$, for $t \in (0,1]$ we define

$$v_t(x,y) := \lim_{r \to 0} \frac{\operatorname{vol}(Z_t(x, B_r(y)))}{\operatorname{vol}(B_{tr}(y))} > 0$$

(the above limit always exists, thought it will be infinite when x and y are conjugate points, see [18]). Arguing as in the proof of Theorem 2.3.3, by [18, Lemma 6.1] we get the following:

Theorem 2.3.6 (Jacobian inequality). *Let E be the set of full μ-measure given by Theorem 2.3.3. Then, for each $x \in E$, $Y_t(x) := d(\exp_x)_{-t\tilde{\nabla}_x\varphi}$ and $H_t(x) := \frac{1}{2}\nabla_x^2 d^2_{T_t(x)}$ both exist for all $t \in [0,1]$ and the Jacobian determinant*

$$J_t(x) := \det[Y_t(x)(H_t(x) - t\tilde{\nabla}_x^2\varphi)] \tag{2.9}$$

satisfies

$$J_t^{\frac{1}{n}}(x) \geq (1-t)\,[v_{1-t}(T(x),x)]^{\frac{1}{n}} + t\,[v_t(x,T(x))]^{\frac{1}{n}}\, J_1^{\frac{1}{n}}(x).$$

We now consider as source measure $\mu_0 = \rho_0\, d\operatorname{vol}(x) \in P_2^{ac}(M)$ and as target measure $\mu_1 = \rho_1\, d\operatorname{vol}(x) \in P_2^{ac}(M)$. By Proposition 2.3.1 we have

$$\mu_t = (T_t)_\sharp[\rho_0\, d\operatorname{vol}] = \rho_t\, d\operatorname{vol} \in P_2^{ac}(M),$$

for a certain $\rho_t \in L^1(M, d\,\mathrm{vol})$.
We now want to consider the behavior of the functional

$$U(\rho) := \int_M A(\rho(x))\, d\,\mathrm{vol}(x)$$

along the path $t \mapsto \rho_t$. In Euclidean spaces, this path is called *displacement interpolation* and the functional U is said to be *displacement convex* if

$$[0,1] \ni t \mapsto U(\rho_t) \quad \text{is convex for every } \rho_0, \rho_1.$$

A sufficient condition for the displacement convexity of U in $\mathbb{R}^n$ is that $A : [0,+\infty) \to \mathbb{R} \cup \{+\infty\}$ satisfy

$$(0,+\infty) \in s \mapsto s^n A(s^{-n}) \text{ is convex and non-increasing, with } A(0) = 0 \qquad (2.10)$$

(see [41], [43]). Typical examples include the entropy $A(\rho) = \rho \log \rho$ and the L^q-norm $A(\rho) = \frac{1}{q-1}\rho^q$ for $q \geq \frac{n-1}{n}$.

By all the results collected above, arguing as in the proof of [18, Theorem 6.2], we can prove that the displacement convexity of U is still true on Ricci non-negative manifolds under the assumption (2.10).

Theorem 2.3.7 (Displacement convexity on Ricci non-negative manifolds)**.** *If* $\mathrm{Ric} \geq 0$ *and A satisfies (2.10), then U is displacement convex.*

Proof. As we remarked above, T_t is the optimal transport map from μ_0 to μ_t. So, by Theorem 2.3.3 and Proposition 2.3.5, we get

$$U(\rho_t) = \int_M A(\rho_t(x))\, d\,\mathrm{vol}(x) = \int_{E_t} A\left(\frac{\rho_0(x)}{\left(J_t^{\frac{1}{n}}(x)\right)^n}\right)\left(J_t^{\frac{1}{n}}(x)\right)^n d\,\mathrm{vol}(x) \qquad (2.11)$$

where E_t is the set of full μ_0-measure given by Theorem 2.3.3 and $J_t(x) \neq 0$ is defined in (2.9). Since $\mathrm{Ric} \geq 0$, we know that $v_t(x,y) \geq 1$ for every $x, y \in M$ (see [18, Corollary 2.2]). Thus, for fixed $x \in E_1$, Theorem 2.3.6 yields the concavity of the map

$$[0,1] \ni t \mapsto J_t^{\frac{1}{n}}(x).$$

Composing this function with the convex non-increasing function $s \mapsto s^n A(s^{-n})$ we get the convexity of the integrand in (2.11). The only problem in order to conclude the displacement convexity of U is that the domain of integration appears to depend on t. But, since by Theorem 2.3.3 E_t is a set of full μ_0-measure for any $t \in [0,1]$, we obtain that, for fixed $t, t', s \in [0,1]$,

$$U(\rho_{(1-s)t+st'}) \leq (1-s)U(\rho_t) + sU(\rho_{t'}),$$

simply by computing each of the three integrals above on the full measure set $E_t \cap E_{t'} \cap E_{(1-s)t+st'}$. □

2.4 A generalization of the existence and uniqueness result

Now we want to generalize this existence and uniqueness result for optimal transport mapping without any integrability assumption on the cost function, adapting the ideas of [42]. We observe that, without the hypothesis

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

in general the minimization problem

$$C(\mu,\nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_{M\times N} c(x,y)\, d\gamma(x,y) \right\}$$

is ill-posed, as it may happen that $C(\mu,\nu) = +\infty$. However, it is known that the optimality of a transport plan γ is equivalent to the c-cyclical monotonicity of the measure-theoretic support of γ whenever $C(\mu,\nu) < +\infty$ (see [4], [50], [56]), and so one may ask whether the fact that the support of γ is c-cyclically monotone implies that γ is supported on a graph. Moreover one can also ask whether this graph is unique, that is is does not depends on γ, which is the case when the cost is $\mu \otimes \nu$ integrable, as Theorem 2.1.1 tells us. The uniqueness in that case, follows by the fact that the functions φ_n are constructed using a pair of function (φ, ψ) which is optimal for the dual problem, and so they are independent of γ. The result we now want to prove is the following:

Theorem 2.4.1. *Assume that $c : M \times N \to \mathbb{R}$ is lower semicontinuous and bounded from below, and let γ be a plan concentrated on a $c-$cyclically monotone set. If*

(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 sets with σ-finite $(n-1)$-dimensional Hausdorff measure,*

then γ is concentrated on a graph of a measurable map $T : M \to N$ (existence). Moreover, if $\tilde{\gamma}$ is another plan concentrated on a $c-$cyclically monotone set, then $\tilde{\gamma}$ is concentrated on the same graph (uniqueness).

Proof. **Existence.** We want to prove that γ is concentrated on a graph. First we recall that, since γ is concentrated on a $c-$cyclically monotone set, there exists a pair of function (φ, ψ), with φ μ-measurable and ψ ν-measurable, such that

$$\varphi(x) = \inf_{y \in N} \psi(y) + c(x, y) \quad \forall x \in M,$$

which implies

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

Moreover we have

$$\varphi(x) - \psi(y) = c(x, y) \quad \gamma - a.e. \tag{2.12}$$

and there exists a point $x_0 \in M$ such that $\varphi(x_0) = 0$ (see [56, Theorem 5.9]). In particular, this implies

$$\psi(y) \geq -c(x_0, y) > -\infty \quad \forall y \in N.$$

So, we can argue as in the proof of Theorem 2.1.1 and Corollary 2.1.3. More precisely, taken a suitable increasing sequence of compact sets $(K_n) \subset N$ such that $\nu(K_n) \nearrow 1$ and $\psi \geq -n$ on K_n (it suffices to take an increasing sequence of compact sets $K_n \subset \{\psi \geq -n\}$ such that $\nu(\{\psi \geq -n\} \setminus K_n) \leq \frac{1}{n}$), we consider the locally semi-concave function

$$\varphi_n(x) := \inf_{y \in K_n} \psi(y) + c(x, y). \tag{2.13}$$

Then, thanks to (2.12), it is possible to find an increasing sequence of Borel sets $D_n \subset \text{supp}(\mu)$, with $\mu(D_n) \nearrow 1$, such that φ_n is differentiable on D_n, $\varphi_n \equiv \varphi$ on D_n and the set $\{\varphi_n = \varphi\}$ has μ-density 1 in all the points of D_n, and γ is concentrated on the graph of the map T determined in D_n by

$$\frac{\partial c}{\partial x}(x, T(x)) = d_x \varphi_n \quad \text{for } x \in D_n.$$

Moreover one has

$$\varphi(x) = \psi(T(x)) + c(x, T(x)) \quad \forall x \in \bigcup_n D_n. \tag{2.14}$$

Uniqueness. As we observed before, the difference here with the case of Theorem 2.1.1 is that the function φ_n depends on the pair (φ, ψ), which in this case depends on γ. Let so $(\tilde{\varphi}, \tilde{\psi})$ be a pair associated to $\tilde{\gamma}$ as above, and let $\tilde{\varphi}_n$ and $\tilde{D}_n$ be such that $\tilde{\gamma}$ is concentrated on the graph of the map $\tilde{T}$ determined in $\tilde{D}_n$ by

$$\frac{\partial c}{\partial x}(x, \tilde{T}(x)) = d_x\tilde{\varphi}_n \quad \text{for } x \in \tilde{D}_n.$$

We need to prove that $T = \tilde{T}$ μ-a.e.
Let us define $C_n := D_n \cap \tilde{D}_n$. Then $\mu(C_n) \nearrow 1$. We want to prove that, if x is a μ-density point of C_n for a certain n, then $T(x) = \tilde{T}(x)$ (we recall that, since $\mu(\cup_n C_n) = 1$, also the union of the μ-density points of C_n is of full μ-measure, see for example [22, Chapter 1.7]).
Let us assume by contradiction that $T(x) \neq \tilde{T}(x)$, that is

$$d_x\varphi_n \neq d_x\tilde{\varphi}_n.$$

Since $x \in \mathrm{supp}(\mu)$, each ball around x must have positive measure under μ. Moreover, the fact that the sets $\{\varphi_n = \varphi\}$ and $\{\tilde{\varphi}_n = \tilde{\varphi}\}$ have μ-density 1 in x implies that the set

$$\{\varphi = \tilde{\varphi}\}$$

has μ-density 0 in x. In fact, as φ_n and $\tilde{\varphi}_n$ are locally semi-concave, up to adding a C^1 function they are concave in a neighborhood of x and their gradients differ at x. So we can apply the non-smooth version of the implicit function theorem proven in [42], which tells us that $\{\varphi_n = \tilde{\varphi}_n\}$ is a set with finite $(n-1)$-dimensional Hausdorff measure in a neighborhood of x (see [42, Theorem 17 and Corollary 19]).
So we have

$$\limsup_{r\to 0} \frac{\mu(\{\varphi = \tilde{\varphi}\} \cap B_r(x))}{\mu(B_r(x))} \leq \limsup_{r\to 0} \left[\frac{\mu(\{\varphi \neq \varphi_n\} \cap B_r(x))}{\mu(B_r(x))} + \frac{\mu(\{\varphi_n = \tilde{\varphi}_n\} \cap B_r(x))}{\mu(B_r(x))} + \frac{\mu(\{\tilde{\varphi}_n \neq \tilde{\varphi}\} \cap B_r(x))}{\mu(B_r(x))} \right] = 0.$$

Now, exchanging φ_n with $\tilde{\varphi}_n$ if necessary, we may assume that

$$\mu(\{\varphi_n < \tilde{\varphi}_n\} \cap B_r(x)) \geq \frac{1}{3}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small,}$$

which implies

$$\mu(\{\varphi < \tilde{\varphi}\} \cap B_r(x)) \geq \frac{1}{4}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small.} \tag{2.15}$$

Let us define $A := \{\varphi < \tilde{\varphi}\}$, $A_n := \{\varphi_n < \tilde{\varphi}_n\}$, $E_n := A \cap A_n \cap C_n$. Since the sets $\{\varphi_n = \varphi\}$ and $\{\tilde{\varphi}_n = \tilde{\varphi}\}$ have μ-density 1 in x, and x is a μ-density point of C_n, we have

$$\lim_{r \to 0} \frac{\mu((A \setminus E_n) \cap B_r(x))}{\mu(B_r(x))} = 0,$$

and so, by (2.15), we get

$$\mu(E_n \cap B_r(x)) \geq \frac{1}{5}\mu(B_r(x)) \quad \text{for } r > 0 \text{ sufficiently small.} \tag{2.16}$$

Now, arguing as in the proof of the Aleksandrov's lemma (see [42, Lemma 13]), we can prove that

$$X := \tilde{T}^{-1}(T(A)) \subset A$$

and $X \cap E_n$ lies a positive distance from x. In fact let us assume, without loss of generality, that

$$\varphi(x) = \varphi_n(x) = \tilde{\varphi}(x) = \tilde{\varphi}_n(x) = 0, \quad d_x\varphi_n \neq d_x\tilde{\varphi}_n = 0.$$

To obtain the inclusion $X \subset A$, let $z \in X$ and $y := \tilde{T}(z)$. Then $y = T(m)$ for a certain $m \in A$. For any $w \in M$, recalling (2.14), we have

$$\varphi(w) \leq c(w, y) - c(m, y) + \varphi(m),$$

$$\tilde{\varphi}(m) \leq c(m, y) - c(z, y) + \tilde{\varphi}(z).$$

Since $\varphi(m) < \tilde{\varphi}(m)$ we get

$$\varphi(w) < c(w, \tilde{T}(z)) - c(z, \tilde{T}(z)) + \tilde{\varphi}(z) \quad \forall w \in M.$$

In particular, taking $w = z$, we obtain $z \in A$, that proves the inclusion $X \subset A$. Let us suppose now, by contradiction, that there exists a sequence $(z_k) \subset X \cap E_n$ such that $z_k \to x$. Again there exists m_k such that $\tilde{T}(z_k) = T(m_k)$. As $d_x\tilde{\varphi}_n = 0$, the closure of the superdifferential of a semi-concave function implies that $d_{z_k}\tilde{\varphi}_n \to 0$. We now observe that, arguing exactly as above with φ_n and $\tilde{\varphi}_n$ instead of φ and $\tilde{\varphi}$, using (2.13), (2.14), and the fact that $\varphi = \varphi_n$ and $\tilde{\varphi} = \tilde{\varphi}_n$ on C_n, one obtains

$$\varphi_n(w) < c(w, \tilde{T}(z_k)) - c(z_k, \tilde{T}(z_k)) + \tilde{\varphi}_n(z_k) \quad \forall w \in M.$$

Taking w sufficiently near to x, we can assume that we are in $\mathbb{R}^n \times N$. We now remark that, since $z_k \in E_n \subset \tilde{D}_n$, $\tilde{T}(z_k)$ vary in a compact subset of N. So, by hypothesis *(i)* on c, we can find a common modulus of continuity ω in a neighborhood of x for the family of uniformly semi-concave functions $z \mapsto c(z, \tilde{T}(z_k))$. Then, we get

$$\begin{aligned}\varphi_n(w) &< \frac{\partial c}{\partial x}(z_k, \tilde{T}(z_k))(w - z_k) + \omega(|w - z_k|)|w - z_k| + \tilde{\varphi}_n(z_k)\\ &= d_{z_k}\tilde{\varphi}_n(w - z_k) + \omega(|w - z_k|)|w - z_k| + \tilde{\varphi}_n(z_k).\end{aligned}$$

Letting $k \to \infty$ and recalling that $d_{z_k}\tilde{\varphi}_n \to 0$ and $\tilde{\varphi}_n(x) = \varphi_n(x) = 0$, we obtain

$$\varphi_n(w) - \varphi_n(x) \le \omega(|w - x|)|w - x| \Rightarrow d_x\varphi_n = 0,$$

which is absurd.
Thus there exists $r > 0$ such that $B_r(x) \cap E_n$ and $X \cap E_n$ are disjoint, and (2.16) holds. Defining now $Y := T(A)$, by (2.16) we obtain

$$\begin{aligned}\nu(Y) &= \mu(T^{-1}(Y)) \ge \mu(A) = \mu(E_n) + \mu(A \setminus E_n) \ge \mu(B_r(x) \cap E_n)\\ &+ \mu(X \cap E_n) + \mu(X \setminus E_n) = \mu(B_r(x) \cap E_n) + \mu(X) \ge \frac{1}{5}\mu(B_r(x)) + \nu(Y),\end{aligned}$$

which is absurd. □

Let now consider the special case $N = M$, with M a complete manifold. As shown in Paragraph 2.2, the above theorem applies in the following cases:

(i) $c : M \times M \to \mathbb{R}$ is defined by

$$c(x, y) := \inf_{\gamma(0)=x,\ \gamma(1)=y} \int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt,$$

where the infimum is taken over all the continuous piecewise C^1 curves, and the Lagrangian $L(x, v) \in C^2(TM, \mathbb{R})$ is C^2-strictly convex and uniform superlinear in v, and verifies an uniform boundeness in the fibers;

(ii) $c(x, y) = d^p(x, y)$ for any $p \in (1, +\infty)$, where $d(x, y)$ denotes a complete Riemannian distance on M.

Chapter 3

Costs induced by Mañé potentials

Regarding the existence-uniqueness of transport maps, the results of the previous paragraph covers all the cases $c(x, y) = d^p(x, y)$ for $p > 1$, but not the limit case $p = 1$.

So, in this chapter, we extend to non-compact manifolds the results of Bernard and Buffoni proved in [8], where the authors showed the existence of optimal transport maps for a large class of costs (that includes in particular the case $c(x, y) = d(x, y)$) on compact manifolds without boundary. The existence of an optimal transport maps in the case $c(x, y) = d(x, y)$ on non-compact manifolds has also been proved in [28] under the assumption of compactness of the supports of the two measures (see the references in [28] for earlier works in the same spirit).

More precisely, we prove the existence of an optimal transport for the class of Mañé potentials associated to a supercritical Lagrangian (that includes the case $c(x, y) = d(x, y)$), using in particular results on weak KAM theory on non compact manifolds (see [26]).

We remark that we do not assume, as usual in the standard theory of optimal transportation, that the cost function is bounded by below. In fact such assumption would be quite nonnatural for a Mañé potential and, also in particular cases, it would not be simple to check its validity. So, in order to apply the standard duality result that gives us an optimal pair for the dual problem, the idea will be to add to our cost a null-Lagrangian, so that the cost becomes non-negative and still satisfies the triangle inequality, and the minimization problem does not change (see Section 3.1.2).

3.0.1 The main result

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

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

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

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

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

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

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

We define the cost function

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

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

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

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

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

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

Theorem 3.0.2. *Assume that c is the cost function associated to a supercritical Lagrangian that satisfies all the assumption above. Suppose that*

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

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

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

among all optimal plans for (1.4), where U is a strict subsolution of the Hamilton Jacobi equation (see Proposition 3.1.2).

We recall that the idea of using a secondary variational problem in order to select a "good" optimal plan was first used in [4] and refined in [5].

Remark 3.0.3. *We observe that, by the triangle inequality for the distance, the condition*

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

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

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

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

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

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

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

We remark that the Lagrangian

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

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

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

3.1 Definitions and preliminary results

3.1.1 Preliminaries in Lagrangian Dynamics

We recall some results of Lagrangian Dynamics that will be useful in the sequel (see [17], [24], [39]) and that shows the naturality of the supercriticality assumption.

Proposition 3.1.1. *Let L be a Lagrangian that satisfies assumption (L1), (L2) and (L3). For $k \in \mathbb{R}$, let us define c_k the Mañé potential associated to the Lagrangian $L + k$. Then there exists a constant k_0 such that*

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

(ii) for $k \geq k_0$, c_k is locally Lipschitz on $M \times M$ and satisfies the triangle inequality
$$c_k(x, z) \leq c_k(x, y) + c_k(y, z) \quad \forall x, y, z \in M;$$
in addition $c_k(x, x) = 0\ \forall x \in M$;

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

The following proposition is a simple corollary of the results proved in [26]:

Proposition 3.1.2. *The Lagrangian L is supercritical if and only if there exist $\delta > 0$ and a C^∞ function U such that*

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

or equivalently
$$L(x, v) - d_x U(v) \geq \delta, \quad \forall (x, v) \in TM,$$
where H is the Hamiltonian associated to the Lagrangian L, that is

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

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

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

(under the assumptions made on the Lagrangian, this value exists and is unique). Then it suffices to apply the following approximation result, also proven in [26]:

Theorem 3.1.3. *If $u : M \to \mathbb{R}$ is locally Lipschitz, with its derivative $d_x u$ satisfying $H(x, d_x u) \leq k$ almost everywhere, then for each $\varepsilon > 0$ there exists a C^∞ function $u_\varepsilon : M \to \mathbb{R}$ such that $H(x, d_x u_\varepsilon) \leq k + \varepsilon$ and $|u(x) - u_\varepsilon(x)| \leq \varepsilon$ for each $x \in M$.*

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

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

3.1.2 Duality and Kantorovich potential

Let us consider a cost function $c(x, y)$ as in Theorem 3.0.2. By hypothesis (L3), c is Lipschitz. In fact, given $x, y \in M$, we consider a geodesic $\gamma_{x,y} : [0, d(x, y)] \to M$ from x to y with $\|\dot\gamma_{x,y}\| = 1$. Then

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

and so we have

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

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

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

(see Proposition 3.1.1). Fix now $z \in M$ and consider the auxiliary cost

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

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

$$0 \leq \overline{c}(x, y) \leq c(x, y) + c(y, x) \leq 2A(1)d(x, y). \tag{3.1}$$

Thus $\overline{c}(x, y)$ is integrable with respect to $\mu \otimes \nu$ if so it is $d(x, y)$, and in this case we can apply Theorem 1.0.3 to prove the following:

Theorem 3.1.4. *Given two probability measures μ and ν on M such that*

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

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

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

and

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

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

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

that is

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

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

Proof. First we remark that, by (3.1),

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

So let (φ, ψ) be a $\overline{c}$-subsolution that realizes the maximum in the dual problem (see Definition 1.0.1 and Theorem 1.0.3). We observe that $\varphi(x) < +\infty$ for all $x \in M$, $\psi(y) > -\infty$ for all $y \in N$ (otherwise, if for example $\psi(y_0) = -\infty$, then $\varphi(x) \le \psi(y_0) + \overline{c}(x, y_0) = -\infty$, i.e. $\varphi \equiv -\infty$ that is not possible as $\varphi \in L^1(d\mu)$). We now remark that we can assume that φ satisfies the formula

$$\psi(y) = \sup_{x\in M} \varphi(x) - \overline{c}(x,y). \tag{3.2}$$

In fact, if it is not the case, it suffices to replace ψ by

$$\tilde{\psi}(y) := \sup_{x\in M} \varphi(x) - \overline{c}(x,y).$$

Observe that $\tilde{\psi}$ is measurable as it is upper semicontinuous. Now we obviously have $\varphi(x) - \tilde{\psi}(y) \le \overline{c}(x,y)$; moreover

$$\tilde{\psi} \le \psi \quad \Rightarrow \quad \tilde{\psi}_+ \le \psi_+ \quad \Rightarrow \quad \int_M \tilde{\psi}_+\, d\nu \le \int_M \psi_+\, d\nu < +\infty$$

that is $\tilde{\psi}_+ \in L^1(d\nu)$, and so, as $\varphi(x) - \tilde{\psi}_+(y) + \tilde{\psi}_-(y) \le \overline{c}(x,y)$,

$$\int_M \tilde{\psi}_- \, d\nu = \int_M \tilde{\psi}_+ \, d\nu - \int_M \varphi \, d\mu + \int_{M\times M} \overline{c} \, d\gamma < +\infty,$$

that tells us that $\tilde{\psi} \in L^1(d\nu)$. So $(\varphi, \tilde{\psi})$ is still a $\overline{c}$-subsolution and, as $\tilde{\psi} \le \psi$, it still realizes the maximum in the dual problem.
We now want to prove that it suffices to take $\overline{u} = -\psi$. Fix $x, y \in M$. By (3.2), the fact that (φ, ψ) is a $\overline{c}$-subsolution and the triangle inequality for $\overline{c}$, we have

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

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

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

Thus, if we define $\overline{u} := -\psi$, by (3.1) we have

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

This inequality tells us that $\overline{u}$ is $2A(1)$-Lipschitz, and so, by Remark 3.0.3, $\overline{u} \in L^1(d\mu) \cap L^1(d\nu)$. In order to conclude the proof, we must show that $(-\overline{u}, -\overline{u})$ realizes the maximum in the dual problem.
This simply follows observing that

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

and thus

$$\int_M \overline{u} \, d(\nu - \mu) \ge \int_M \varphi \, d\mu - \int_M \psi \, d\nu.$$

□

3.1.3 Calibrated curves

Fix a C^∞ function U and a $\delta > 0$ given by Proposition 3.1.2, and a Kantorovich potential u given by Theorem 3.1.4. Following [8], we recall some useful definitions.

Definition 3.1.5 (u-calibrated curve)**.** We say that a continuous piecewise differentiable curve $\gamma : I \to M$ is *u-calibrated* if

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

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

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

Definition 3.1.6 (transport ray)**.** A *transport ray* is the image of a non-trivial u-calibrated curve.

In [24], it is proved that Kantorovich potentials are viscosity subsolutions of the Hamilton-Jacobi equation, that is equivalent to say that u is locally Lipschitz and satisfies

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

or equivalently

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

We recall that, if $\gamma : [a, b] \to \mathbb{R}$ is a u-calibrated curve, then for all $t \in (a, b)$ the function u is differentable at $\gamma(t)$ (see [24]). Then we have the following:

Lemma 3.1.7. *Let $\gamma : [a, b] \to \mathbb{R}$ be a u-calibrated curve. Then for all $t \in (a, b)$ the function u is differentable at $\gamma(t)$ and we have*

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

where U and δ are given by Proposition 3.1.2. This implies that γ is an embedding and transport rays are non-trivial embedded arcs.

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

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

that implies, recalling Proposition 3.1.2,

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

□

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

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

Lemma 3.1.8. *α and β are Borel functions.*

Proof. Let $K^i \subset M$ be a countable increasing sequence of compact set such that $\cup_i K^i = M$. Then we can define the auxiliary functions $\alpha_i(x)$ as the supremum of all times $T \geq 0$ such that there exists a u-calibrated curve $\gamma : [-T, 0] \to M$ such that $\gamma(0) = x$ and $\gamma(-T) \in K^i$. We will prove that α_i is upper semicontinuous for each i, and this will implies the measurability of α as $\alpha(x) = \sup_i \alpha_i(x)$ for all $x \in M$ (the case of β is analogous).
Fix $i \in \mathbb{N}$ and let $(x_j) \subset M$ be a sequence converging to a limit $x \in M$ such that $\alpha_i(x_j) \geq T$ for all j. Then we know that there exists a sequence $\gamma_j : [-T, 0] \to M$ of u-calibrated curves such that $\gamma_j(0) = x_j$ and $\gamma_j(-T) \in K^i$. As $\gamma_j(-T) \in K^i$, we know that there exists a constant A such that $\|\dot{\gamma}_j(0)\|_{\gamma_j(0)} \leq A$ for all j (see appendix, Proposition B.1.8). Then, taking a subsequence if necessary, we can assume that γ_j converges uniformly on $[-T, 0]$ to a curve $\gamma : [-T, 0] \to M$ which is still u-calibrated, as it is easy to see, and satisfies $\gamma(0) = x$, $\gamma(-T) \in K^i$. Then $\alpha_i(x) \geq T$. □

We now can define the following Borel sets:

Definition 3.1.9. We define the set $\mathcal{T}$ given by the union of all the transport rays as

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

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

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

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

We now recall the following:

Theorem 3.1.10. *The function u is differentiable at each point of $\mathcal{T}_0$. For each point $x \in \mathcal{T}_0$, there exists a unique maximal u-calibrated curve*

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

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

$$d_x u = \nabla_v L(x, \dot\gamma_x(0))$$

or equivalently

$$\dot\gamma_x(0) = \nabla_p H(x, d_x u).$$

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

For a proof see [24].

Definition 3.1.11. For $x \in M$, we will denote by R_x the union of the transport rays containing x. We also denote

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

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

In order to conclude this section, we recall two results of [8].

Lemma 3.1.12. *We have*

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

where γ_x is given by Theorem 3.1.10.

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

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

that is

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

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

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

So, if $x \in \mathcal{T}_0$, by Theorem 3.1.10 $\gamma = \gamma_x|_{[0,T]}$ and hence $y = \gamma(T) = \gamma_x(T) \in \gamma_x([0, \beta(x)])$, while, if $x \notin \mathcal{T}$, there is no non-trivial u-calibrated curve and then we must have $y = x$ in the above discussion, that implies $R_x^+ = \{x\}$. □

Proposition 3.1.13. *A transport plan γ is optimal for the cost c if and only if it is supported on the closed set*

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

Proof. By Theorem 3.1.4, γ is optimal if and only if

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

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

3.2 Proof of the main theorem

The line of the proof is essentially the same as in [8], where the authors, using ideas of Lagrangian Dynamics, extend to a Riemannian setting the results obtained in the Euclidean case in [4].
As before, we fix a C^∞ function U and a $\delta > 0$ given by Proposition 3.1.2, and then fix a Kantorovich potential u given by Theorem 3.1.4 (see Sections 3.1.2, 3.1.3). We now define the second cost function

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

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

$$\min_{\gamma \in \mathcal{O}} \int_{M\times M} \tilde{c}(x,y)\, d\gamma(x,y), \tag{3.3}$$

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

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

where $\Gamma_x := \{y \in M \mid (x,y) \in \Gamma\}$ and # denotes the cardinality of the set. In this way, intersecting Λ with a transport ray R, thanks to the monotonicity of $\Lambda \cap R$ it is simple to see that $\Lambda \cap R$ is at most countable. Finally Theorem 3.1.10 allows us to parameterize the transport rays in a locally Lipschitz way. This and the fact that $\Lambda \cap R$ has zero $\mathscr{H}^1$-measure for each transport ray R (where $\mathscr{H}^k$ denote the k-dimensional Hausdorff measure) imply that Λ has null volume measure, and so $\mu(\Lambda) = 0$ as wanted.
We divide the proof in many steps, in order to make the overall strategy more clear.

Step 1: the construction of Γ.

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

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

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

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

So we have

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

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

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

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

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

and, thanks to Proposition 3.1.13, we have that γ_0 is a minimizer for the Kantorovich problem

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

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

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

(for a proof see, for example, [4, Theorem 3.2]). By the definition of ζ this implies the following monotonicity property of $\tilde{\Gamma}$:
if $\big((x_i, y_i)\big)_{1 \le i \le l}$ is a finite family of points of $\tilde{\Gamma}$ and $\sigma(i)$ is a permutation such that $\big((x_i, y_{\sigma(i)})\big)_{1 \le i \le l}$ is still contained in $\tilde{\Gamma}$, then

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

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

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

Now that we have constructed Γ, we define

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

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

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

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

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

is a Borel subset of M.
Let us now show that $\Lambda \subset \mathcal{T}$. If $x \notin \mathcal{T}$, then, by Lemma 3.1.12, $R^+_x = \{x\}$. Hence, as $\Gamma_x \subset R^+_x$ (see Proposition 3.1.13), $\Gamma_x \subset \{x\}$ and $x \notin \Lambda$.

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

We fix a transport ray, that is the image of a non-trivial maximal u-calibrated curve $\gamma : [a,b] \to M$, and we consider the strictly increasing function $f : [a,b] \to \mathbb{R}$ defined by $f = (u - U) \circ \gamma$ (see Lemma 3.1.7). We observe that

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

In view of the monotonicity of Γ we have

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

whenever $(\gamma(s), \gamma(t)) \in \Gamma$, $(\gamma(s'), \gamma(t')) \in \Gamma$, $s \le t'$, $s' \le t$. Now, following [4], we show the implication

$$(\gamma(s), \gamma(t)) \in \Gamma, \quad (\gamma(s'), \gamma(t')) \in \Gamma, \quad s < s' \implies t \le t'. \tag{3.4}$$

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

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

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

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

and therefore we have a contradiction. By (3.4), we obtain that the vertical sections Γ_x of Γ are ordered along a transport ray, i.e.

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

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

Step 4: covering the set Λ.

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

Definition 3.2.1. We call *transport beam* a couple (B, χ) where $B \subset \mathbb{R}^n$ is a Borel subset and $\chi : B \to M$ is a locally Lipschitz map such that:

- there exist a bounded Borel set $\Omega \subset \mathbb{R}^{n-1}$ and two Borel functions $a < b : \Omega \to \mathbb{R}$ such that

$$B = \{(\omega, s) \in \Omega \times \mathbb{R} \mid a(\omega) \le s \le b(\omega)\} \subset \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R};$$

- for each $\omega \in \Omega$, the curve $\chi_\omega : [a(\omega), b(\omega)] \to M$ given by $\chi_\omega(s) = \chi(\omega, s)$ is u-calibrated.

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

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

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

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

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

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

where α and β were defined in Section 3.1.3. We observe that, by Lemma 3.1.8, $a_{j,k}$ and $b_{j,k}$ are Borel functions. We can now define the Borel sets

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

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

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

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

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

is of class C^1. As we have

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

where $\pi_M : TM \to M$ is the canonical projection, in view of Theorem 3.1.10 we deduce that this map is locally Lipschitz. It is clear that, for each transport ray R, there exist $j, k \in \mathbb{N}$ such that R is contained in $\chi_{j,k}(B_{j,k})$.

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

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

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

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

Step 6: uniqueness.

We now prove that the transport plan selected with the secondary variational problem is unique.
Let γ_0, γ_1 be two optimal transport plans, which are optimal also for the secondary variational problem. By what we proved above, we know that they are induced by two transport maps $t_0 : M \to M$ and $t_1 : M \to M$, respectively. Let us now consider $\overline{\gamma} := \frac{\gamma_0+\gamma_1}{2}$. By the linear structure of the two variational problems (1.4) and (3.3), $\overline{\gamma}$ is still optimal for both, and so it is induced by a transport map $\bar{t}$. This implies that both γ_0 and γ_1 are concentrated on the graph of $\bar{t}$, and so $t_0 = t_1$ μ-a.e.

Remark 3.2.2. *We observe that exactly this argument shows also that the set $\mathcal{E}$ of ray ends is negligible with respect to the volume measure.*

Chapter 4

Displacement convexity on Riemannian manifolds

4.1 Introduction and main result

Recent research activity has been devoted to study the geometry of Riemannian manifolds through the geometry of their associated Wasserstein space of probability measures. In particular, it was found that lower bounds on the Ricci curvature tensor can be recast in terms of convexity properties of certain nonlinear functionals defined on spaces of probability measures [47, 18, 52, 38, 51]. In this paper we solve a natural problem in this field by establishing the equivalence of several such formulations.

Before explaining our results in more detail, let us give some notation and background. Let (M, g) be a smooth complete connected n-dimensional Riemannian manifold, equipped with its geodesic distance d and its volume measure vol. Let $P(M)$ be the set of probability measures on M. For any real number $p \geq 1$, we denote by $P_p(M)$ the set of probability measures μ such that

$$\int_M d^p(x, x_0)\, d\mu(x) < \infty \qquad \text{for some } x_0 \in M.$$

The set $P_2(M)$ is equipped with the Wasserstein distance of order 2, denoted by W_2 (see Paragraph 2.3). Then $P_2(M)$ is a metric space, and even a length space; that is, any two probability measures in $P_2(M)$ are joined by at least one geodesic curve $(\mu_t)_{0\leq t\leq 1}$ (here and in the sequel, by convention geodesics are supposed to be globally minimizing and to have constant speed).

A basic representation theorem (see [38, Proposition 2.10] or [56, Corollary 7.20]) states that any Wasserstein geodesic curve necessarily takes the form $\mu_t = (e_t)_\# \Pi$, where Π is a probability measure on the set Γ of minimizing geodesics $[0,1] \to M$, and $e_t : \Gamma \to M$ is the evaluation at time t: $e_t(\gamma) := \gamma(t)$. So the optimal transport problem between two probability measures μ_0 and μ_1 produces three related objects:

- an optimal plan π of μ_0 and μ_1;
- a path $(\mu_t)_{0\le t\le 1}$ in the space of probability measures;
- a probability measure Π on the space of geodesics, such that $(e_t)_\#\Pi = \mu_t$ and $(e_0, e_1)_\#\Pi = \pi$. Such a Π is called a dynamical optimal transference plan.

The core of the studies in [47, 18, 52, 38, 51] lies in the analysis of the convexity properties of certain nonlinear functionals along geodesics in $P_2(M)$, defined below [56, Chapter 29]:

Definition 4.1.1 (Nonlinear functionals of probability measures)**.** Let ν be a reference measure on M, absolutely continuous with respect to the volume measure. Let $U : \mathbb{R}_+ \to \mathbb{R}$ be a continuous convex function with $U(0) = 0$; let $U'(\infty)$ be the limit of $U(r)/r$ as $r \to \infty$. Let μ be a probability measure on M and let $\mu = \rho\nu + \mu_s$ be its Lebesgue decomposition with respect to ν. Then

(i) If $U(\rho)$ is bounded below by a ν-integrable function, then the quantity $U_\nu(\mu)$ is defined by the formula

$$U_\nu(\mu) = \int_M U(\rho(x))\,\nu(dx) + U'(\infty)\,\mu_s[M].$$

(ii) If π is a probability measure on $M \times M$, β is a continuous positive function on $M \times M$, and $U(\rho/\beta)$ is bounded below by a ν-integrable function, then the quantity $U^\beta_{\pi,\nu}(\mu)$ is defined by the formula

$$U^\beta_{\pi,\nu}(\mu) = \int_{M\times M} U\left(\frac{\rho(x)}{\beta(x,y)}\right) \beta(x,y)\pi(dy|x)\,\nu(dx) + U'(\infty)\,\mu_s[M],$$

where $\pi(dy|x)$ is the disintegration of $\pi(dx\,dy)$ with respect to x.

Remark 4.1.2. *If $U'(\infty) = \infty$, then finiteness of $U_\nu(\mu)$ implies that μ is absolutely continuous with respect to ν. This is not true if $U'(\infty) < \infty$.*

The various notions of convexity that are considered in [52, 38, 51] belong to the following ones:

Definition 4.1.3 (Convexity properties)**.** (i) Let U and ν be as in Definition 4.1.1, and let $\lambda \in \mathbb{R}$. We say that the functional U_ν is *λ-displacement convex* if for all Wasserstein geodesics $(\mu_t)_{0\le t\le 1}$ whose image lies in the domain of U_ν,

$$U_\nu(\mu_t) \le (1-t)\, U_\nu(\mu_0) + t\, U_\nu(\mu_1) - \frac{1}{2}\lambda t(1-t) W_2^2(\mu_0,\mu_1), \qquad \forall t \in [0,1]. \quad (4.1)$$

(i') We say that the functional U_ν is *displacement convex with distortion β* if for all Wasserstein geodesics $(\mu_t)_{0\le t\le 1}$ whose image lies in the domain of U_ν, if $\pi(dx\,dy)$ stands for the associated optimal coupling between μ_0 and μ_1, and $\check{\pi}$ is obtained from π by exchanging the two variables, then

$$U_\nu(\mu_t) \le (1-t)\, U^\beta_{\pi,\nu}(\mu_0) + t\, U^\beta_{\check{\pi},\nu}(\mu_1), \qquad \forall t \in [0,1]. \quad (4.2)$$

(ii) We say that U_ν is *weakly λ-displacement convex* (resp. *weakly displacement convex with distortion β*) if for all μ_0, μ_1 in the domain of U_ν, there is *some* Wasserstein geodesic from μ_0 to μ_1 along which (4.1) (resp. (4.2)) is satisfied.

(iii) We say that U_ν is *weakly λ-a.c.c.s. displacement convex* (resp. *weakly a.c.c.s. displacement convex with distortion β*) if condition (4.1) (resp. (4.2)) is satisfied along some Wasserstein geodesic when we further assume that μ_0, μ_1 are absolutely continuous and compactly supported.

Remark 4.1.4. *If U_ν is a λ-displacement convex functional, then the function $t \mapsto U_\nu(\mu_t)$ is λ-convex on $[0,1]$, i.e. for all $0 \le s \le s' \le 1$ and $t \in [0,1]$,*

$$U_\nu(\mu_{ts'} + (1-t)s) \le tU_\nu(\mu_{s'}) + (1-t)U_\nu(\mu_s) - \frac{1}{2}\lambda t(1-t)(s'-s)^2 W_2^2(\mu_0,\mu_1). \quad (4.3)$$

This is not a priori *the case if we only assume that U_ν is weakly λ-displacement convex.*

We may sometimes write "displacement convex" instead of 0-displacement convex. In short, *weakly* means that we require a condition to hold only for *some* geodesic between two measures, as opposed to all geodesics, and *a.c.c.s.* means that we only require the condition to hold when the two measures are absolutely continuous and compactly supported.

There are obvious implications

$$\begin{array}{c}\lambda\text{-displacement convex}\\ \Downarrow \\ \text{weakly } \lambda\text{-displacement convex}\\ \Downarrow \\ \text{weakly } \lambda\text{-a.c.c.s. displacement convex.}\end{array}$$

Although the natural convexity condition is arguably the one appearing in (i) and (i'), that is, holding true along all Wasserstein geodesics, this condition is quite more delicate to study than the weaker conditions appearing in (ii) and (iii), in particular for stability issues (see [38, 51]). In the same references the equivalence between (ii) and (iii) was established, at least for compact spaces [38, Proposition 3.21]. But the implication (ii) $\Rightarrow$ (i) remained open. In the present paper we more or less fill this gap, thus answering a quite natural question in the theory of convexity in the Wasserstein space. Here is our main result:

Theorem 4.1.5. *Let U, ν and β be as in Theorem 4.1.1. For each $a > 0$, define $U_a(r) = U(ar)/a$. Then*

(i) If $(U_a)_\nu$ is weakly λ-a.c.c.s. displacement convex for all $a \in (0, 1]$, then U_ν is λ-displacement convex;

(ii) If $(U_a)_\nu$ is weakly displacement convex with distortion β for all $a \in (0, 1]$, then U_ν is displacement convex with distortion β.

Among the consequences of Theorem 4.1.5 is the following corollary, solving what was presented as an open problem in a preliminary version of [56]:

Corollary 4.1.6. *Let M be a smooth complete Riemannian manifold with non-negative Ricci curvature and dimension n. Let $U(r) = -r^{1-1/n}$, and let ν be the volume measure on M. Then U_ν is displacement convex on $P_p(M)$, where $p = 2$ if $n \geq 3$, and p is any real number greater than 2 if $n = 2$.*

More generally, Theorem 4.1.5 makes it possible to drop the "weakly" in all displacement convexity characterizations of Ricci curvature bounds.

Before turning to the proof of Theorem 4.1.5, let us explain a bit more about the difficulties and the strategy of proof. Obviously, there are two problems to tackle: first, the possibility that μ_0 and/or μ_1 do not have compact support; and secondly, the possibility that μ_0 and/or μ_1 are singular with respect to the volume measure.

It was shown in [38, 51] that inequalities such as (4.1) or (4.2) are *stable* under (weak) convergence. Then it is natural to approximate μ_0, μ_1 by compactly supported, absolutely continuous measures, and pass to the limit. This scheme of proof is enough to show the implication (iii) $\Rightarrow$ (ii) in Definition 4.1.3, but does not guarantee that we can attain *all* Wasserstein geodesics in this way — unless of course we know that there is a unique Wasserstein geodesic between μ_0 and μ_1.

To treat the difficulty arising from the possible non-compactness, we use recent results (see [25], [30]), showing that the Wasserstein geodesic between any two absolutely continuous probability measures on M is unique, even if they are not compactly supported (see Paragraph 2.3).

The difficulty arising from the possible singularity of μ_0, μ_1 is less simple. If μ_0 and μ_1 are both singular, then there are in general several Wasserstein geodesics joining them. A most simple example is constructed by taking $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$, where δ_x stands for the Dirac mass at x, and x_0, x_1 are joined by multiple geodesics. So it is part of the problem to regularize μ_0, μ_1 into absolutely continuous measures $\mu_{0,k}$, $\mu_{1,k}$ so that, as $k \to \infty$, the optimal transport between $\mu_{0,k}$ and $\mu_{1,k}$ converges to a *given* optimal transport between μ_0 and μ_1.

We handle this by a rather nonstandard regularization procedure, which roughly goes as follows. We start from a given dynamical optimal transference plan Π between μ_0 and μ_1, leave intact that part $\Pi^{(a)}$ of Π which corresponds to the absolutely continuous part of μ_0. Then we let displacement occur for a very short time at the level of that part $\Pi^{(s)}$ of Π corresponding to the singular part of μ_0. Next we regularize the resulting contribution of $\Pi^{(s)}$.

Let us illustrate this in the most basic case when $\mu_0 = \delta_{x_0}$ and $\mu_1 = \delta_{x_1}$. Let $\gamma = (\gamma_t)_{0\le t\le 1}$ be a given geodesic between x_0 and x_1; we wish to approximate the Wasserstein geodesic $(\delta_{\gamma_t})_{0\le t\le 1}$. Instead of directly regularizing μ_0 and μ_1, we shall first replace μ_0 by γ_τ, where τ is positive but very small, and then regularize δ_{γ_τ} and δ_{x_1} into probability measures $\mu_{\tau,\varphi}$ and $\mu_{1,\varphi}$. What we have gained is that the geodesic joining γ_τ to $x_1 = \gamma_1$ is unique, so we may let $\tau \to 0$ and $\varphi \to 0$ in such a way that the Wasserstein geodesic joining $\mu_{\tau,\varphi}$ to $\mu_{1,\varphi}$ does converge to $(\delta_{\gamma_t})_{0\le t\le 1}$.

In a more general context, the procedure will be more tricky, and what will make it work is the following important property [56, Theorem 7.26]: *geodesics in dynamical optimal transport plans do not cross at intermediate times.* In fact, if Π is a given dynamical optimal transport plan, then for each $t \in (0,1)$ one can define a measurable map $F_t : M \to \Gamma$ by the requirement that $F_t \circ e_t = \mathrm{Id}$, Π-almost surely. In understandable words, if γ is a geodesic along which there is optimal transport, then the position of γ at time t determines the whole geodesic γ. This property will ensure that $\Pi^{(a)}$ and $\Pi^{(s)}$ "do not overlap at intermediate times".

Finally, we note that the results in this paper can be extended to more general situations outside the category of Riemannian manifolds: It is sufficient that the optimal transport between any two absolutely continuous probability measures be unique. In fact, there is a more general framework where these results still hold true, namely the case of *nonbranching* locally compact, complete length spaces. This extension will be established, by a slightly different approach, in [56, Chapter 30].

4.2 Proofs

In the sequel, we shall use the notation $U_{a,\nu}$ for $(U_a)_\nu$. An important ingredient in the proof of Theorem 4.1.5 will be the following lemma.

Lemma 4.2.1. *Let U be a Lipschitz convex function with $U(0) = 0$. Let μ_1, μ_2 be any two probability measures on M, and let Z_1, Z_2 be two positive numbers with $Z_1 + Z_2 = 1$. Then*

(i) $U_\nu(Z_1\mu_1 + Z_2\mu_2) \geq Z_1\, U_{Z_1,\nu}(\mu_1) + Z_2\, U_{Z_2,\nu}(\mu_2)$, with equality if μ_1 and μ_2 are singular to each other;

(ii) Let π_1, π_2 be two probability measures on $M \times M$, and let β be a positive measurable function on $M \times M$. Then

$$U^\beta_{Z_1\pi_1+Z_2\pi_2,\nu}(Z_1\mu_1 + Z_2\mu_2) \geq Z_1\, U^\beta_{Z_1,\pi_1,\nu}(\mu_1) + Z_2\, U^\beta_{Z_2,\pi_2,\nu}(\mu_2),$$

with equality if μ_1 and μ_2 are singular to each other.

Proof of Lemma 4.2.1. We start by the following remark: If x, y are nonnegative numbers, then

$$U(x + y) \geq U(x) + U(y). \tag{4.4}$$

Inequality (4.4) follows at once from the fact that $U(t)/t$ is a nondecreasing function of t, and thus

$$\frac{U(x)}{x} \leq \frac{U(x+y)}{x+y}, \quad \frac{U(y)}{y} \leq \frac{U(x+y)}{x+y}$$
$$\Rightarrow xU(x + y) + yU(x + y) \geq (x + y)(U(x) + U(y)).$$

Now we turn to the proof of the lemma. With obvious notation,

$$U_\nu(Z_1\mu_1 + Z_2\mu_2) = \int U(Z_1\rho_1 + Z_2\rho_2)\, d\nu + U'(\infty)\big(Z_1\, \mu_{1,s}[M] + Z_2\, \mu_{2,s}[M]\big);$$

$$U_{Z_1,\nu}(\mu_1) = \frac{1}{Z_1}\int U(Z_1\rho_1)\, d\nu + U'(\infty)\mu_{1,s}[M];$$

$$U_{Z_2,\nu}(\mu_2) = \frac{1}{Z_2}\int U(Z_2\rho_2)\, d\nu + U'(\infty)\mu_{2,s}[M];$$

so part (i) of the lemma follows immediately from (4.4). The claim about equality is obvious since it amounts to say that $U(x + y) = U(x) + U(y)$ as soon as either x or y is zero.

To prove part (ii), we observe that, with obvious notation,

$$
\begin{aligned}
&U^{\beta}_{Z_1\pi_1+Z_2\pi_2,\nu}(Z_1\mu_1+Z_2\mu_2)\\
&\qquad\qquad = \int_{M\times M} U\left(\frac{Z_1\rho_1(x)+Z_2\rho_2(x)}{\beta(x,y)}\right)\beta(x,y)(Z_1\pi_1+Z_2\pi_2)(dy|x)\,\nu(dx)\\
&\qquad\qquad + U'(\infty)\big(Z_1\,\mu_{1,s}[M]+Z_2\,\mu_{2,s}[M]\big);
\end{aligned}
$$

$$
U^{\beta}_{Z_1,\pi_1,\nu}(\mu_1) = \int U\left(\frac{Z_1\rho_1(x)}{\beta(x,y)}\right)\beta(x,y)\pi_1(dy|x)\,d\nu + U'(\infty)\mu_{1,s}[M];
$$

$$
U^{\beta}_{Z_2,\pi_2,\nu}(\mu_2) = \int U\left(\frac{Z_2\rho_2(x)}{\beta(x,y)}\right)\beta(x,y)\pi_2(dy|x)\,d\nu + U'(\infty)\mu_{2,s}[M].
$$

Thus the thesis follows again by (4.4). □

Proof of Theorem 4.1.5. First we observe that U_ν is well-defined on $P_2(M)$ since, if $\mu=\rho\nu+\mu_s$ is the Lebesgue decomposition of a probability measure $\mu\in P(M)$, then

$$U(\rho)\ge -\|U\|_{\mathrm{Lip}}\,\rho\in L^1(M,\nu).$$

In fact, there is also an upper bound, so U_ν is well-defined on the whole of $P_2(M)$ with values in $\mathbb{R}$.

Let μ_0, μ_1 be any two measures in $P_2(M)$, and let Π be an optimal dynamical transference plan between μ_0 and μ_1. Let further

$$\mu_0=\rho_0\,\nu+\mu_{0,s}$$

be the Lebesgue decomposition of μ_0 with respect to ν. Let $E^{(a)}$ and $E^{(s)}$ be two disjoint Borel subsets of M such that $\rho_0\,\nu$ is concentrated on E_a and $\mu_{0,s}$ is concentrated on E_s. We decompose Π as

$$\Pi=\Pi^{(a)}+\Pi^{(s)}, \tag{4.5}$$

where

$$\Pi^{(a)}:=\Pi\llcorner\{\gamma\in\Gamma\mid\gamma(0)\in E^{(a)}\},\qquad \Pi^{(s)}:=\Pi\llcorner\{\gamma\in\Gamma\mid\gamma(0)\in E^{(s)}\}.$$

Taking the marginals at time t in (4.5) we get

$$\mu_t=\mu_t^{(a)}+\mu_t^{(s)}.$$

In the end, we renormalize $\mu_t^{(a)}$ and $\mu_t^{(s)}$ into probability measures: we define

$$Z^{(a)} = \Pi^{(a)}[\Gamma] = \mu_0^{(a)}[M] = \mu_t^{(a)}[M]; \qquad Z^{(s)} = \Pi^{(s)}[\Gamma],$$

and

$$\hat{\mu}_t^{(a)} := \frac{\mu_t^{(a)}}{Z^{(a)}}; \qquad \hat{\mu}_t^{(s)} := \frac{\mu_t^{(s)}}{Z^{(s)}}.$$

So

$$\mu_t = Z^{(a)} \hat{\mu}_t^{(a)} + Z^{(s)} \hat{\mu}_t^{(s)}. \tag{4.6}$$

We remark that by the results in [25] $\mu_t^{(a)}$ is absolutely continuous for any $t \in [0,1)$, but $\mu_t^{(s)}$ is not necessarily completely singular.

It follows from [56, Theorem 7.26 (iv)], for any $t \in (0,1)$ there is a Borel map F_t such that $F_t(\gamma_t) = \gamma_0$, $\Pi(d\gamma)$-almost surely. Then $\mu_t^{(s)}$ is concentrated on $F_t^{-1}(E^{(s)})$, while $\mu_t^{(a)}$ is concentrated on $F_t^{-1}(E^{(a)})$; so these measures are singular to each other. Then by Lemma 4.2.1 and (4.6), for any $t \in (0,1)$,

$$U_\nu(\mu_t) = U_{Z^{(a)},\nu}(\hat{\mu}_t^{(a)}) + U_{Z^{(s)},\nu}(\hat{\mu}_t^{(s)}). \tag{4.7}$$

In the sequel, we focus on part (i) of Theorem 4.1.5, since the reasoning is quite the same for part (ii). By construction and the restriction property of optimal transport [56, Theorem 7.26], $\Pi^{(a)}$ is an optimal dynamical transference plan between $\mu_0^{(a)}$ and $\mu_1^{(a)}$, and the associated Wasserstein geodesic is $(\mu_t^{(a)})_{0\le t\le 1}$. Since by construction $\mu_0^{(a)}$ is absolutely continuous, by the results in Paragraph 2.3 $(\mu_t^{(a)})$ is the *unique* Wasserstein geodesic joining $\mu_0^{(a)}$ to $\mu_1^{(a)}$. Then we can apply the displacement convexity inequality of the functional $U_{Z^{(a)},\nu}$ along that geodesic:

$$U_{Z^{(a)},\nu}(\hat{\mu}_t^{(a)}) \le (1-t)\, U_{Z^{(a)},\nu}(\hat{\mu}_0^{(a)}) + t\, U_{Z^{(s)},\nu}(\hat{\mu}_1^{(a)}) - \frac{\lambda}{2}\, t\,(1-t)\, W_2^2(\hat{\mu}_0^{(a)}, \hat{\mu}_1^{(a)}). \tag{4.8}$$

Next, let $\varphi_k \to 0$ be a sequence of positive numbers. From the nonbranching property of $P_2(M)$ [56, Corollary 7.28], there is only one Wasserstein geodesic joining $\mu_{\varphi_k}^{(s)}$ to $\mu_1^{(s)}$, and it is obtained by reparameterizing $(\mu_t^{(s)})_{\varphi_k\le t\le 1}$. So we can also apply the displacement convexity inequality of the functional $U_{Z^{(s)},\nu}$ along that geodesic, and get

$$U_{Z^{(s)},\nu}(\hat{\mu}_t^{(s)}) \le \left(\frac{1-t}{1-\varphi_k}\right) U_{Z^{(s)},\nu}(\hat{\mu}_{\varphi_k}^{(s)}) + \left(\frac{t-\varphi_k}{1-\varphi_k}\right) U_{Z^{(s)},\nu}(\hat{\mu}_1^{(s)}) - \frac{\lambda}{2}\, \frac{(t-\varphi_k)\,(1-t)}{(1-\varphi_k)^2}\, W_2^2(\hat{\mu}_0^{(s)}, \hat{\mu}_1^{(s)}) \tag{4.9}$$

(for the latter term we have used the fact that if $(\mu_t)_{0\le t\le 1}$ is any Wasserstein geodesic, then $W_2(\mu_s,\mu_t) = |t-s|\, W_2(\mu_0,\mu_1)$).

The first term in the right-hand side of (4.9) can be trivially bounded by $U'(\infty)$, which coincides with $U_{Z^{(s)},\nu}(\hat\mu_0^{(s)})$ since $\hat\mu_0^{(s)}$ is totally singular. Indeed, since $\frac{U(r)}{r} \le U'(\infty)$, we have

$$\begin{aligned}
U_{Z^{(s)},\nu}(\hat\mu_{\varphi_k}^{(s)}) &= \frac{1}{Z^{(s)}}\int_M U\left(Z^{(s)}\hat\rho_{\varphi_k}^{(s)}\right)\, d\nu + U'(\infty)\,\hat\mu_{\varphi_k,s}^{(s)}(M)\\
&= \frac{1}{Z^{(s)}}\int_{\{\hat\rho_{\varphi_k}^{(s)}>0\}} U\left(Z^{(s)}\hat\rho_{\varphi_k}^{(s)}\right)\, d\nu + U'(\infty)\,\hat\mu_{\varphi_k,s}^{(s)}(M)\\
&= \int_{\{\hat\rho_{\varphi_k}^{(s)}>0\}} \frac{U\left(Z^{(s)}\hat\rho_{\varphi_k}^{(s)}\right)}{Z^{(s)}\hat\rho_{\varphi_k}^{s}}\hat\rho_{\varphi_k}^{(s)}\, d\nu + U'(\infty)\,\hat\mu_{\varphi_k,s}^{(s)}(M)\\
&\le \int_{\{\hat\rho_{\varphi_k}^{(s)}>0\}} U'(\infty)\hat\rho_{\varphi_k}^{(s)}\, d\nu + U'(\infty)\,\hat\mu_{\varphi_k,s}^{(s)}(M)\\
&= U'(\infty)\,\hat\mu_{\varphi_k}^{(s)}(M) = U'(\infty).
\end{aligned}$$

Then by passing to the lim inf as $k\to\infty$ in (4.9), we recover

$$U_{Z^{(s)},\nu}(\hat\mu_t^{(s)}) \le (1-t)\,U_{Z^{(s)},\nu}(\hat\mu_0^{(s)}) + t\,U_{Z^{(s)},\nu}(\hat\mu_1^{(s)}) - \frac{\lambda}{2}\,t(1-t)\,W_2^2(\hat\mu_0^{(s)},\hat\mu_1^{(s)}). \quad (4.10)$$

By combining together (4.7), (4.8) and (4.9), we obtain

$$\begin{aligned}
U_\nu(\mu_t) \le (1-t)&\left[Z^{(a)}U_{Z^{(a)},\nu}(\hat\mu_0^{(a)}) + Z^{(s)}U_{Z^{(s)},\nu}(\hat\mu_0^{(s)})\right]\\
&+ t\left[Z^{(a)}U_{Z^{(a)},\nu}(\hat\mu_1^{(a)}) + Z^{(s)}U_{Z^{(s)},\nu}(\hat\mu_1^{(s)})\right]\\
&- \frac{\lambda}{2}\,t(1-t)\left[Z^{(a)}W_2^2(\hat\mu_0^{(a)},\hat\mu_1^{(a)}) + Z^{(a)}W_2^2(\hat\mu_0^{(s)},\hat\mu_1^{(s)})\right].
\end{aligned} \quad (4.11)$$

The last term inside square brackets can be rewritten as

$$\int d^2(\gamma_0,\gamma_1)\Pi^{(a)}(d\gamma) + \int d^2(\gamma_0,\gamma_1)\Pi^{(s)}(d\gamma) = \int d^2(\gamma_0,\gamma_1)\Pi(d\gamma) = W_2^2(\mu_0,\mu_1).$$

Plugging this back into (4.11) and using Lemma 4.2.1, we conclude that

$$U_\nu(\mu_t) \le (1-t)\,U_\nu(\mu_0) + t\,U_\nu(\mu_1) - \frac{\lambda}{2}\,t(1-t)\,W_2^2(\mu_0,\mu_1).$$

This concludes the proof of Theorem 4.1.5. □

Proof of Corollary 4.1.6. Let $U := r \to -r^{1-1/N}$. By the estimates derived in [38, Proposition E.17], U_ν is well-defined on $P_p(M)$ (this is made more explicit in [56, Chapter 17]).

Let $\mathcal{DC}_N$ be the displacement convex class of order N, that is the class of functions $U \in C^2(0,\infty) \cap C([0,+\infty))$ such that $U(0) = 0$ and $\delta^N U(\delta^{-N})$ is a convex function of δ. (See Definition 4.3.1, [56, Definition 17.1]). Obviously, $U \in \mathcal{DC}_N$. By [56, Proposition 17.4], there is a sequence $(U^{(\ell)})_{\ell\in\mathbb{N}}$ of Lipschitz functions, all belonging to $\mathcal{DC}_N$, such that $U^{(\ell)}$ converges monotonically to U as $\ell \to \infty$.

Since $U^{(\ell)}$ lies in $\mathcal{DC}_N$, it is by now classical [56, Chapter 17] that $U_\nu^{(\ell)}$ it is a.c.c.s-displacement convex. By Theorem 4.1.5, this functional is also displacement convex. Then it follows by an easy limiting argument that U_ν itself is displacement convex. □

4.3 Application to the study of Ricci curvature bounds

4.3.1 Functionals with distorsion

In the definition of the functionals $U^\beta_{\pi,\nu}$ (see Definition 4.1.1) the factor β is usually called distorsion.

The following particular distorsion coefficients are particularly useful: given $K \in \mathbb{R}$, define for $N \in (1,\infty)$

$$\beta_t^{(K,N)}(x,y) := \begin{cases} \infty & \text{if } K>0 \text{ and } \alpha>\pi, \\ \left(\frac{\sin(t\alpha)}{t\sin\alpha}\right)^{N-1} & \text{if } K>0 \text{ and } \alpha\in[0,\pi], \\ 1 & \text{if } K=0, \\ \left(\frac{\sinh(t\alpha)}{t\sinh\alpha}\right)^{N-1} & \text{if } K<0, \end{cases}$$

with $\alpha := \sqrt{\frac{|K|}{N-1}}d(x,y)$, while, in the limit cases $N=1$ and $N=\infty$,

$$\beta_t^{(K,1)}(x,y) := \begin{cases} \infty & \text{if } K>0, \\ t & \text{if } K\le 0, \end{cases}$$

$$\beta_t^{(K,\infty)}(x,y) := e^{\frac{K}{6}(1-t^2)d^2(x,y)}.$$

4.3.2 The class $\mathcal{DC}_N$

We now define a suitable class of convex functions, introduced by McCann [43]. Consider a continuous convex function $U : [0, \infty) \to \mathbb{R}$ with $U(0) = 0$. We define the nonnegative function

$$\mathrm{p}(r) := rU'_+(r) - U(r), \tag{4.12}$$

with $\mathrm{p}(0) = 0$. If one thinks of U as defining an internal energy for a continuous medium then p can be thought of as a pressure.

Definition 4.3.1. For $N \in [1, \infty)$, we define $\mathcal{DC}_N$ to be the set of all continuous convex functions U on $[0, \infty)$, with $U(0) = 0$, such that the function

$$\psi(\lambda) = \lambda^N U(\lambda^{-N}) \tag{4.13}$$

is convex on $(0, \infty)$.

We further define $\mathcal{DC}_\infty$ to be the set of all continuous convex functions U on $[0, \infty)$, with $U(0) = 0$, such that the function

$$\psi(\lambda) = e^\lambda U(e^{-\lambda}) \tag{4.14}$$

is convex on $(-\infty, \infty)$.

We note that the convexity of U implies that ψ is nonincreasing in λ, as $\frac{U(x)}{x}$ is nondecreasing in x. For some properties of the classes $\mathcal{DC}_N$, see [38], [56].

Important example of functions in $\mathcal{DC}_N$

The following example of functions in $\mathcal{DC}_N$ is the most important:

$$U_N(r) = \begin{cases} Nr(1 - r^{-1/N}) & \text{if } 1 < N < \infty, \\ r \log r & \text{if } N = \infty. \end{cases} \tag{4.15}$$

It is simple to verify that $U_N \in \mathcal{DC}_N$. Now, given $\mu = \rho\nu + \mu_s \in P_2(M)$, the functional $H_{N,\nu} : P_2(M) \to [0, \infty]$ associated to U_N via Definition 4.1.3 is given by:

- For $N \in (1, \infty)$,

$$H_{N,\nu}(\mu) = N - N \int_X \rho^{1-\frac{1}{N}} \, d\nu.$$

- For $N = \infty$,

$$H_{\infty,\nu}(\mu) = \begin{cases} \int_X \rho \log \rho \, d\nu & \text{if } \mu_s = 0, \\ \infty & \text{otherwise.} \end{cases}$$

4.3.3 Displacement convexity (with distorsion) and Ric_N

We can now give the definition of weak curvature dimension condition and of N-Ricci tensor (see [38], [56]).

Definition 4.3.2. Fix $K \in \mathbb{R}$ and let $N \in [1, \infty]$. Then M is said to satisfies a weak $CD(K, N)$ condition if, whenever μ_0, μ_1 are two compactly supported absolutely continuous probability measures, there exists a Wasserstein geodesics $\{\mu_t\}_{t\in[0,1]}$ from μ_0 to μ_1 such that, for any $U \in \mathcal{DC}_N$, we have

$$U_\nu(\mu_t) \le tU_{\pi,\nu}^{\beta_{1-t}^{(K,N)}}(\mu_1) + (1-t)U_{\hat{\pi},\nu}^{\beta_t^{(K,N)}}(\mu_0) \quad \forall t \in [0,1], \tag{4.16}$$

where π is the optimal coupling from μ_0 to μ_1, $\hat{\pi}$ is the optimal coupling from μ_1 to μ^0.

Let Ric denote the Ricci tensor of M.

Definition 4.3.3. For $N \in [1, \infty]$, the N-Ricci tensor of (M, g, ν) is

$$\mathrm{Ric}_N = \begin{cases} \mathrm{Ric} + \mathrm{Hess}(V) & \text{if } N = \infty, \\ \mathrm{Ric} + \mathrm{Hess}(V) - \frac{1}{N-n} dV \otimes dV & \text{if } n < N < \infty, \\ \mathrm{Ric} + \mathrm{Hess}(V) - \infty(dV \otimes dV) & \text{if } N = n, \\ -\infty & \text{if } N < n, \end{cases} \tag{4.17}$$

where by convention $\infty \cdot 0 = 0$ and we recall that $V \in C^2(M)$ is such that $\nu = e^{-V}\,\mathrm{vol}$.

We recall the following result (for a proof, see for example [56, Theorem 29.6]):

Theorem 4.3.4. *For $N \in [1, \infty]$, the following are equivalent :*
(1) $Ric_N \ge Kg$.
(2) M satisfies the $CD(K, N)$ weak condition.

Finally, the following result is proved in [38, Theorem 7.3]:

Theorem 4.3.5. *Suppose M compact.*
a) For $N \in (1, \infty)$, the following are equivalent :
(1) $\mathrm{Ric}_N \ge 0$.
(2) For all $U \in \mathcal{DC}_N$, U_ν is weakly displacement convex on $P_2^{ac}(M)$.
(3) For all $U \in \mathcal{DC}_N$, U_ν is weakly a.c.(c.s.) displacement convex on $P_2^{ac}(M)$.
(4) $H_{N,\nu}$ is weakly a.c.(c.s.) displacement convex on $P_2^{ac}(M)$.

Let us define $\lambda : \mathcal{DC}_\infty \to \mathbb{R} \cup \{-\infty\}$ *as*

$$\lambda(U) = \inf_{r>0} K\frac{p(r)}{r} = \begin{cases} K \lim_{r\to 0^+} \frac{p(r)}{r} & \text{if } K > 0, \\ 0 & \text{if } K = 0, \\ K \lim_{r\to\infty} \frac{p(r)}{r} & \text{if } K < 0, \end{cases} \tag{4.18}$$

where p *is given by (4.12).*
b) For any $K \in \mathbb{R}$*, the following are equivalent :*
(1) $\mathrm{Ric}_\infty \geq Kg$.
(2) For all $U \in \mathcal{DC}_\infty$*,* U_ν *is weakly* $\lambda(U)$*-a.c.(c.s.) displacement convex on* $P_2^{ac}(M)$.
(3) For all $U \in \mathcal{DC}_\infty$*,* U_ν *is weakly* $\lambda(U)$*-a.c.(c.s.) displacement convex on* $P_2^{ac}(M)$.
(4) $H_{\infty,\nu}$ *is weakly* K*-a.c.(c.s.) displacement convex on* $P_2^{ac}(M)$.

By the Theorem 4.1.5, we get as corollaries the following theorems.

Theorem 4.3.6. *Suppose that* M *satisfies the weak* $CD(K,N)$ *condition and that* $U \in \mathcal{DC}_N$ *for a certain* $N \in [1,\infty]$*. Then, if* M *is compact,* U_ν *is* λ*-displacement convex on* $P_2(M)$*. More in general, if* M *is non-compact, if there exists* $p \geq 2$ *such that*

$$\int_M \frac{d\nu(x)}{(1+d(x,x_0))^{p(N-1)}} < \infty \quad \text{for a certain } x_0 \in M \quad (1 \leq N < \infty),$$

$$\int_M e^{cd^p(x,x_0)}\, d\nu(x) < \infty \quad \text{for a certain } x_0 \in M,\ c > 0 \quad (N = \infty),$$

then U_ν *is well-defined and* λ*-displacement convex on* $P_p(M)$.

Theorem 4.3.7. *Suppose that* M *satisfies the weak* $CD(K,N)$ *condition and that* $U \in \mathcal{DC}_N$ *for a certain* $N \in [1,\infty]$*. Then, if* M *is compact, (4.16) is true for all* $\mu_0, \mu_1 \in P_2(M)$ *and for all Wasserstein geodesics. More in general, if* M *is non-compact, the same is true if there exists* $p \geq 2$ *such that*

$$\int_M \frac{d\nu(x)}{(1+d(x,x_0))^{p(N-1)}} < \infty \quad \text{for a certain } x_0 \in M \quad (1 \leq N < \infty),$$

$$\int_M e^{cd^p(x,x_0)}\, d\nu(x) < \infty \quad \text{for a certain } x_0 \in M,\ c > 0 \quad (N = \infty).$$

We now observe that, by Theorems 4.3.6 and 4.3.5, we have:

Corollary 4.3.8. *Assume that there exists* $p \geq 2$ *such that*

$$\int_M \frac{d\nu(x)}{(1+d(x,x_0))^{p(N-1)}} < \infty \quad \textit{for a certain } x_0 \in M \quad (1 < N < \infty),$$

$$\int_M e^{cd^p(x,x_0)}\, d\nu(x) < \infty \quad \textit{for a certain } x_0 \in M,\ c > 0 \quad (N = \infty).$$

a. For $N \in (1,\infty)$*, the following are equivalent :*
(1) $\mathrm{Ric}_N \geq 0$.
(2) For all $U \in \mathcal{DC}_N$, U_ν *is displacement convex on* $P_p(M)$.
(3) For all $U \in \mathcal{DC}_N$, U_ν *is weakly a.c.c.s. displacement convex on* $P_p(M)$.
(4) $H_{N,\nu}$ *is weakly a.c.c.s. displacement convex on* $P_p(M)$.

b. For any $K \in \mathbb{R}$*, the following are equivalent :*
(1) $\mathrm{Ric}_\infty \geq Kg$.
(2) For all $U \in \mathcal{DC}_\infty$, U_ν *is* $\lambda(U)$*-displacement convex on* $P_2^{ac}(M)$.
(3) For all $U \in \mathcal{DC}_\infty$, U_ν *is weakly* $\lambda(U)$*-a.c.c.s. displacement convex on* $P_p(M)$.
(4) $H_{\infty,\nu}$ *is weakly* K*-a.c.c.s. displacement convex on* $P_p(M)$.

Finally, by Theorem 4.3.7 and Theorem 4.3.4, we also get the following:

Corollary 4.3.9. *Assume that there exists* $p \geq 2$ *such that*

$$\int_M \frac{d\nu(x)}{(1+d(x,x_0))^{p(N-1)}} < \infty \quad \textit{for a certain } x_0 \in M \quad (1 \leq N < \infty),$$

$$\int_M e^{cd^p(x,x_0)}\, d\nu(x) < \infty \quad \textit{for a certain } x_0 \in M,\ c > 0 \quad (N = \infty).$$

For $N \in [1,\infty]$*, the following are equivalent :*
(1) $Ric_N \geq Kg$.
(2) For any $U \in \mathcal{DC}_N$*, (4.16) it true for all* $\mu_0, \mu_1 \in P_p(M)$ *and for all Wasserstein geodesics.*
(3) M *satisfies the* $CD(K,N)$ *weak condition.*

Proof of Theorems 4.3.6 and 4.3.7. First we observe that, in the case of M non-compact, the integrability hypothesis on ν ensures that $U_\nu(M)$ is well-defined on $P_p(M)$ (see [56, Theorem 17.5]). As the class $\mathcal{DC}_N$ is invariant by reparameterization, $U_{a,\nu}$ is weakly λ-a.c.c.s. displacement convex for all $a \in (0,1]$ if and only if U_ν is weakly λ-a.c.c.s. displacement convex. So, in order to conclude, it suffices to apply Theorem 4.1.5. □

Appendix A

Semi-concave functions

We give the definition of semi-concave function and we recall their principal properties. The main reference on semi-concave functions is the book [14].

A.1 Definition and principal properties

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

Definition A.1.1 (Modulus). A modulus ω is a non-decreasing function $\omega : [0,+\infty[\to [0,+\infty]$, such that $\omega(0) = 0$, and ω is continuous at 0. We will say that a modulus is *linear* if it is of the form $\omega_K(t) = Kt$, where $K \geq 0$ is some fixed constant.

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

Definition A.1.2 (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 $g \geq f$, $g(x) = f(x)$, 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 an Euclidean space.

Definition A.1.3 (Semi-concavity). Let $U \subset \mathbb{R}^n$ open and let $\omega : \mathbb{R}^+ \to \mathbb{R}^+$ be a continuous nondecreasing function such that $\omega(r) = o(1)$ as $r \to 0$. 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 $l_x \in D^+_x f$. 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.1.4. *1) Suppose $f_i : U \to \mathbb{R}, i = 1, \ldots, k$ is ω_i-semi-concave, where U is an open subset of $\mathbb{R}^n$, and 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 $\inf_{i=1}^k f_i$ 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 we find $i_0 \in \{1, \ldots, k\}$ such that $\inf_{i=1}^k f_i(x) = f_{i_0}(x)$. 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

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

It clearly follows that

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

To prove 2), consider an open convex subset C with $\bar{C}$ compact and contained on 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

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

Therefore f is ω-semi-concave on the open subset C. □

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

Lemma A.1.5. *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 $\{y_1, \dots, y_{2^n}\}$ 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 $i = 1, \dots, 2^n$,

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

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

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

We now know that f is locally bounded below. Using that, 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, this implies

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

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 $\bar{B}(x_i, r_i), i = 1\dots, \ell$, we obtain (i).

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

$$\begin{aligned} f(y) - f(x) &\leq \langle l_x, y - x\rangle + \|y - x\|\omega(\|y - x\|) \\ &\leq (\|l_x\| + \omega(D_K))\,\|y - x\| \\ &\leq (A + \omega(D_K))\|y - x\|. \end{aligned}$$

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 an Euclidean space is differentiable almost everywhere. Therefore by part (ii) of Lemma A.1.5 above we obtain the following corollary:

Corollary A.1.6. *A locally semi-concave real valued function defined on an open subset of an Euclidean space is differentiable almost everywhere.*

In fact, in the case of semi-concave functions there is a better result which is given in the next theorem, whose proof can be found in [14, Section 4.1]:

Theorem A.1.7. *If $\varphi : U \to \mathbb{R}$ is a semi-concave function defined on the open subset U of $\mathbb{R}^n$, then we can find a countable family of Lipschitz functions $f_j : K_j \to U$, where K_j is a compact subset of $\mathbb{R}^{n-1}$, such that φ is differentiable at each point in the complement C of $\cup_j f_j(K_j)$. Moreover, the subset $\cup_j f_j(K_j)$ is σ-compact, and it has σ-finite $(n-1)$-dimensional Hausdorff measure.*

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 for by composition with diffeomorphisms.

Lemma A.1.8. *Let $U, V \subset \mathbb{R}^n$ be open subsets. Suppose that $F : V \to U$ a C^1 diffeomorphism. 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.*

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

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

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

Since $F(\bar{V}')$ is a compact subset of U we can apply part (i) of Lemma A.1.5 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

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

□

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

Definition A.1.9. 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.

Theorem A.1.10. *Suppose $f_i : U \to \mathbb{R}$, $i \in I$, 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 $y_1, \dots, y_{2^n}$ the vertices of C. By the argument in the beginning of the proof of Lemma A.1.5, we have

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

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

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

Choose now $\varepsilon_0 > 0$ such that $\bar{B}(x_0, \varepsilon) \subset C$. If $l_i : \mathbb{R}^n \to \mathbb{R}$ is a linear form such that

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

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

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

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

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

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

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

we get

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

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.1.11. 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

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

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, by absurd, 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

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

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

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

that is obviously absurd.

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.1.12. 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*, 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 an uniformly semi-concave family of functions on the open subset V_j of $\mathbb{R}^n$.

Corollary A.1.13. *If the family $f_i : M \to \mathbb{R}$, $i \in I$ is uniformly locally semi-concave 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.

Definition A.1.14. 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*, 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.

Proposition A.1.15. *If $c : M \times N \to \mathbb{R}$ is a locally semi-concave function on the product of the manifolds M and N, then the family $(c(\cdot, y))_{y\in N}$ of functions on M is locally uniformly locally semi-concave.*

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 semi-concave on $\tilde{U}_i \times \tilde{W}_j$. 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 semi-concave on $\tilde{U}_i$. □

We end this section with another useful theorem. The proof we give is an adaptation of the proof of [23, Lemma 3.8, page 494].

Theorem A.1.16. *Let $\varphi_1, \varphi_2 : M \to \mathbb{R}$ be two functions, with φ_1 locally semi-convex, and φ_2 locally semi-concave. Assume that $\varphi_1 \leq \varphi_2$. If we define $\mathcal{E} = \{x \in M \mid \varphi_1(x) = \varphi_2(x)\}$, 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, with both 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 $M = \mathring{\mathbb{B}}$ 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

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

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

$$\begin{aligned}
l_{1,x}(y - x) - \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}) \leq \varphi_1(y) - \varphi_1(x) \leq \\
\leq \varphi_2(y) - \varphi_1(x) \leq l_{2,x}(y - x) + \|y - x\|_{\text{euc}}\omega(\|y - x\|_{\text{euc}}). \quad \text{(A.1)}
\end{aligned}$$

In particular, we get

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

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

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

Therefore

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

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 (A.1)

$$|\varphi_i(y) - \varphi_i(x) - l(y-x)| \leq \|y-x\|_{\mathrm{euc}}\omega(\|y-x\|_{\mathrm{euc}}). \tag{A.2}$$

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 $\|k\|_{\mathrm{euc}} \leq 1-r$, we can apply three times (A.2) to obtain

$$\begin{aligned}
\varphi_1(y_2) - \varphi_1(y_1) - d_{y_1}\varphi_1(y_2-y_1) &\leq \|y_2-y_1\|_{\mathrm{euc}}\omega(\|y_2-y_1\|_{\mathrm{euc}})\\
\varphi_1(y_2+k) - \varphi_1(y_2) - d_{y_2}\varphi_1(k) &\leq \|k\|_{\mathrm{euc}}\omega(\|k\|_{\mathrm{euc}})\\
-\varphi_1(y_2+k) + \varphi_1(y_1) + d_{y_1}\varphi_1(y_2+k-y_1) &\leq \|y_2+k-y_1\|_{\mathrm{euc}}\omega(\|y_2+k-y_1\|_{\mathrm{euc}}).
\end{aligned}$$

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

$$\begin{aligned}
[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k) \leq{}& \|y_2-y_1\|_{\mathrm{euc}}\omega(\|y_2-y_1\|_{\mathrm{euc}}) + \|k\|_{\mathrm{euc}}\omega(\|k\|_{\mathrm{euc}})\\
&+ [\|y_2-y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}]\omega(\|y_2-y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}),
\end{aligned}$$

which imples, exchanging k with $-k$,

$$|[d_{y_1}\varphi_1 - d_{y_2}\varphi_1](k)| \leq 2[\|y_2-y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}]\omega(\|y_2-y_1\|_{\mathrm{euc}} + \|k\|_{\mathrm{euc}}). \tag{A.3}$$

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

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

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

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

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

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 a lot of references on the subject. In [24], one can find an introduction to the subject that is particularly suited for our purpose. Another reference is [11]. A brief and particularly nice description of the main results is contained in [16].

Definition B.1.1 (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.1.2 (Action)**.** If L is a Lagrangian on M, for a continuous piecewise C^1 (or even 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 $s \mapsto L(\gamma(s), \dot\gamma(s))$ is piecewise continuous (or defined a.e., and measurable if γ is absolutely continuous).

Definition B.1.3 (Minimizer)**.** *If L is a Lagrangian on the manifold M, an absolutely continuous curve $\gamma : [a, b] \to M$, with $a \leq b$, is a minimizer, if $A_L(\gamma) \leq A_L(\delta)$, 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.1.4 (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 $\times \in M$, the map $L(x, \cdot) : T_xM \to M$ is strictly convex;

(c) there exist a complete Riemannian metric g on M and a finite constant C such that
$$\forall (x, v) \in TM, \quad L(x, v) \geq \|v\|_x + C$$
where $\|\cdot\|_x$ is the norm on T_xM obtained from the Riemannian metric g;

(d) for every compact subset K of M the restriction of L to $TM_K := \cup_{x \in K} T_xM$ is superlinear in the fibers of $TM \to M$: this means that for every $A \geq 0$, there exists a finite constant $C(A, K)$ such that
$$\forall (x, v) \in TM, \text{ with } x \in K, \quad L(x, v) \geq A\|v\|_x + C(A, K).$$

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 $(x, v) \in M$, the second partial derivative $\dfrac{\partial^2 L}{\partial v^2}$ is positive definite on T_xM.

Since, all Riemannian metrics are locally equivalent, if condition (d) in the definition is satisfied for one particular Riemannian metric, then it is satisfied for any other Riemannian metric.

On the other hand the completeness of the Riemannian metric in condition (c) above is crucial to guarantee that a set of the form
$$\mathcal{F} = \{\gamma : [a, b] \to M \in C^0([a, b]) \mid \gamma(a) \in K, \mathbb{A}_L \leq C\},$$
where K is a compact subset in M, C 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 $\bar{B}(x_0, R) = \{y \in M \mid d(x, y) \leq R\}$, 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 like for example in [11].

The direct method in the Calculus of Variations, see for example [11, Theorem 3.7, page 114] or [24] for Tonelli Lagrangians, implies:

Theorem B.1.5. *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 a minimizer.*

In fact in [11, Theorem 3.7, page 114], the existence of absolutely continuous minimizers is valid under very general hypotheses on the Lagrangian, and the C^1 hypothesis is too strong. We now come to the problem of regularity of a Lagrangian, which uses the C^1 hypothesis on L:

Theorem B.1.6. *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*

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

which is an integrated from of the Euler-lagrange equation.

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 [11], [16], or [24]. The fact that regularity 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 [15, Corollary 1, page 77, and Corollary 3.1, page 90] (the same fact under weaker regularity assumptions on L has been proved in [2]). The fact that L is C^1 is again too strong. Once one knows that γ is Lipschitz, when L is C^1, one can differentiate the action, see [11], [16], or [24], and 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

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

But the continuity of right hand side in (IEL') implies that $\partial L/\partial v(\gamma(t), \dot\gamma(t))$ extend 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. 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. Now equation (IEL) follows from (IEL'), which now holds everywhere by continuity. □

In fact, in this work, we will only use the case where L is C^2, or 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.

We define the energy $E : TM \to \mathbb{R}$ by

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

If L is a weak Tonelli Lagrangian and γ is a (C^1) minimizer, it is a well-known fact that E is constant on the speed curve

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

and this comes from the fact that L does not depend on time.

Proposition B.1.7. *For every compact set $K \subset M$, for every $A < +\infty$, the set*

$$V(K, A) := \{(x, v) \in TM \mid x \in K,\ E(x, v) \le A\}$$

is compact.

Proof. By condition (b), the global Legendre transform

$$\mathscr{L} : TM \to T^*M,$$
$$(x, v) \mapsto (x, \frac{\partial L}{\partial v}(x, v)),$$

is proper (i.e. inverse images of compact subsets are compact). By condition (a) it is injective, and therefore $\mathscr{L}$ is a surjective homeomorphism. Since

$$V(K, A) = \mathscr{L}\left(\{(x, p) \in T^*M \mid x \in K,\ H(x, p) \le A\}\right),$$

the compactness of $V(K, A)$ follows from the superlinearity of $H(x, p)$ above K, which is a consequence of the hypotheses made on L. □

We now define the cost $c(x, y)$ associated to L by

$$c(x, y) := \inf_{\gamma(0)=x,\ \gamma(1)=y} \int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt.$$

Proposition B.1.8. *Suppose K is a compact subset of M. Then we can find $\tilde{K} \subset M$ a compact subset and A a finite constant such that every curve $\gamma : [0, 1] \to M$ with $\gamma(0), \gamma(1) \in K$ and $\int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt = c(\gamma(0), \gamma(1))$ satisfies $\gamma([0, 1]) \subset \tilde{K}$ and $\|\dot{\gamma}(t)\|_{\gamma(t)} \leq A$ for every $t \in [0, 1]$.*

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 $R = \operatorname{diam}(K)$, the diameter of K). We now pick $x, y \in K$. If $\alpha : [0, 1] \to M$ is a geodesic with $\alpha(0) = x$, $\alpha(1) = y$ and whose length is $d(x, y)$ (such a geodesic exists by completeness), the inequality

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

implies that $\alpha([0, 1]) \subset \bar{B}(x_0, 3R)$. Moreover $\|\dot{\alpha}(s)\|_{\alpha(s)} = d(x, y) \leq 2R$ for every $s \in [0, 1]$. By compactness, the Lagrangian L is bounded on the set

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

We call θ an upper bound of L on $\mathscr{K}$. Obviously the action of α is less then θ, and therefore $c(x, y) \leq \theta$ for every $x, y \in K$.
Suppose now that $\gamma : [0, 1] \to M$ satisfies $\gamma(0), \gamma(1) \in K$ and $\int_0^1 L(\gamma(t), \dot{\gamma}(t))\, dt = c(\gamma(0), \gamma(1))$. Using condition (c) on the Lagrangian L and what we obtained above, we see that

$$C + \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)}\, dt \leq \theta.$$

It follows that we can find $s_0 \in [0, 1]$ such that

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

Moreover

$$\gamma([0, 1]) \subset \bar{B}(\gamma(0), \theta - C) \subset \bar{B}(x_0, R + \theta - C) =: \tilde{K}.$$

If we define

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

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

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

Observing that the set $\tilde{\mathscr{K}}$ does not depend on γ, we have the thesis. □

B.2 The least action or cost of a Lagrangian and its semi-concavity

Definition B.2.1 (Cost 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

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

where the infimum is taken over all the continuous piecewise C^1 curves (or equivalently over all the absolutely continuous curves) $\gamma : [0, t] \to M$, with $\gamma(0) = x$, and $\gamma(t) = y$, and $\mathbb{A}_{t,L}(\gamma)$ 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.2.2. *Suppose that $L : TM \to \mathbb{R}$ is a weak Tonelli Lagrangian. Then, for every $t > 0$, the cost c_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 for 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.1.8, we can find a constant A such that for every $\gamma : [0,1] \to M$, with $\gamma(i) \in \varphi_i^{-1}(\mathbb{B})$, whose action is $c(\gamma(0), \gamma(1))$, we have

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

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

$$d(\varphi_i^{-1}(\tilde{z}_1), \varphi_i^{-1}(\tilde{z}_2)) \geq \delta, \; i = 0, 1,$$

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

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

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 $\|h_i\|_{\text{euc}} \leq 2$. In that case the image of $\tilde{\gamma}_{h_i}$ is contained in $\overset{\circ}{\mathbb{B}}\,(0,4)$ and

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

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

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

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

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

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

Concatening the 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

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

Hence

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

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 $\|\dot{\tilde{\gamma}}_{h_1}(s)\| \leq A_1$, we get the estimate

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

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

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

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

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

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

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

But, the curve $t \mapsto \varphi_0 \circ \gamma(t)$ is a C^1 extremal of L_0, therefore, see Theorem B.1.6 it must must satisfy the following integrated form of the Euler-Lagrange equation

$$\frac{\partial L_0}{\partial v} \left(\varphi_0 \circ \gamma(t), (\varphi_0 \circ \gamma)'(t)\right) - \frac{\partial L_0}{\partial v} \left(\varphi_0 \circ \gamma(0), (\varphi_0 \circ \gamma)'(0)\right) = \int_0^t \frac{\partial L_0}{\partial x} \left(\varphi_0 \circ \gamma(s), (\varphi_0 \circ \gamma)'(s)\right) ds.$$

This gives us

$$l_0(h_0) = -\frac{\partial L_0}{\partial v} \left(\varphi_0 \circ \gamma(0), (\varphi_0 \circ \gamma)'(0)\right) - \frac{1}{\varepsilon} \int_0^\varepsilon \frac{d}{ds} \left[(t-\varepsilon) \int_0^t \frac{\partial L_0}{\partial x} \left(\varphi_0 \circ \gamma(s), (\varphi_0 \circ \gamma)'(s)\right) ds\right] dt.$$

Obviously the second term in the right hand side is 0 and so l_0 reinterpreted on $T_{x_0}M$ rather than on $\mathbb{R}^n$ gives $-\frac{\partial L}{\partial v}(\gamma(0), \dot\gamma(0))$. The treatment for l_1 is the same. □

We have avoided the first variation formula in the proof of Corollary B.2.3, because this is usually proven for C^2 variation of curves and C^2 Lagrangians. Of course, our argument is a proof for the first variation formula for C^1 Lagrangians. This is a standard argument.

B.3 The twist condition for the cost of a Lagrangian

Lemma B.3.1. *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 twist condition of Definition 2.0.5.

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

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

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

(ii) *we have $y = \gamma(t)$, where γ is the minimizer obtained in* (i).

Proof. We first prove the second part. Pick $\gamma : [0, t] \to M$ a minimizer with $x = \gamma(0)$ and $y = \gamma(t)$. Form Corollary B.2.3 we obtain the equality

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

Since by construction $y = \gamma(t)$, it remains to prove the uniqueness part in statement (i). In fact, if $\gamma_1 : [0, t] \to M$ is another minimizer $x = \gamma_1(0)$ and

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

then we get that

$$\frac{\partial L}{\partial v}(x,\dot\gamma(0)) = \frac{\partial L}{\partial v}(x,\dot\gamma_1(0)).$$

Since the C^1 map $v \mapsto L(x,v)$ is strictly convex this implies $\dot\gamma_1(0) = \dot\gamma(0)$. It now follows that $\gamma = \gamma_1$ on the whole interval $[0,t]$.

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

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

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

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

But by (ii), applied twice, we must have $y = \gamma(t) = y_1$. □

The next lemma is an easy consequence of Lemma B.3.1 above.

Lemma B.3.2. *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 condition of Definition 2.0.5.*

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

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

Proposition B.3.3. *Suppose g is a complete Riemannian metric on the connected manifold M, and $r > 1$, then the Lagrangian $g^{r/2}$ on TM defined by*

$$g^{r/2}(x,v) := (g_x(v,v))^{r/2},$$

satisfies condition (UC') *for the geodesic flow ϕ^g_t of g.*

By this proposition, together with Lemma B.3.2, we obtain that the Lagrangian $g^{r/2}$ satisfies condition (UC), and thus we obtain that the cost $c_{t,g^{r/2}}(x,y) = t^{1-r}d^r(x,y)$ satisfies the left twist condition for any $r > 1$.

Bibliography

[1] L. Ambrosio: *Lecture notes on optimal transport problems,* in Mathematical Aspects of Evolving Interfaces, Lecture Notes in Math. **1812**, Springer-Verlag, Berlin/New York (2003), 1-52.

[2] L.Ambrosio, O.Ascenzi & G.Buttazzo: *Lipschitz regularity for minimizers of integral functionals with highly discontinuous integrands.* J. Math. Anal. Appl., **142** (1989), no. 2, 301-316.

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

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

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

[6] V.Bangert: *Analytische Eigenschaften konvexer Funktionen auf Riemannschen Manigfaltigkeiten.* J. Reine Angew. Math., **307** (1979), 309-324.

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

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

[9] Y.Brenier: *Polar decomposition and increasing rearrangement of vector fields.* C. R. Acad. Sci. Paris Sér. I Math., **305** (1987), no. 19, 805-808.

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

[11] G. BUTTAZZO, M. GIAQUINTA &S. HILDEBRANDT: *One-dimensional variational problems.* Oxford Lecture Series in Mathematics and its Applications, **15** (1998), Oxford Univ. Press, New York.

[12] L. CAFFARELLI, M. FELDMAN & R.J. MCCANN: *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs.* J. Amer. Math. Soc., **15** (2002), 1-26.

[13] L. A. CAFFARELLI & 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).

[14] P.CANNARSA & C.SINESTRARI: *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control.* Progress in Nonlinear Differential Equations and Their Applications, **58**, Birkhauser Boston (2004).

[15] F.H. CLARKE & R.B. VINTER: *Regularity properties of solutions to the basic problem in the calculus of variations.* Trans. Amer. Math. Soc., **289** (1985), 73-98.

[16] F.H. CLARKE: *Methods of dynamic and nonsmooth optimization.* CBMS-NSF Regional Conference Series in Applied Mathematics, **57** (1989), Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

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

[18] D.CORDERO-ERASQUIN, R.J.MCCANN & M.SCHMUCKENSCHLAGER: *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb.* Inventiones mathematicae, **146** (2001), 219-257.

[19] D.L. COHN: *Measure theory.* Birkhäuser Boston, (1980), Mass.

[20] L.C. EVANS: *Partial differential equations and Monge-Kantorovich mass transfer*, in R. Bott et al., editors,*Current Developments in Mathematics.* (1997), International Press, Cambridge, 26-78.

[21] L.C. EVANS & W. GANGBO: *Differential equations methods for the Monge-Kantorovich mass transfer problem.* Mem. Amer. Math. Soc., **137** (1999).

[22] L.C.EVANS & R.F.GARIEPY: *Measure Theory and Fine Properties of Functions.* Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[23] A. FATHI: *Regularity of C^1 solutions of the Hamilton-Jacobi equation.* Ann. Fac. Sci. Toulouse Math. (6), **12** (2003), 479-516.

[24] A.FATHI: *Weak KAM theorems in Lagrangian Dynamics.* Book to appear.

[25] A.FATHI & A.FIGALLI: *Optimal transportation on manifolds.* Preprint, 2006.

[26] A.FATHI & E.MADERNA: *Weak KAM theorem on non compact manifolds.* Nonlin. Diff. Eq. and Appl., to appear.

[27] H. FEDERER: *Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften,* **153**, Springer-Verlag New York Inc., New York (1969).

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

[29] A.FIGALLI: *The Monge problem on non-compact manifolds.* Rend. Sem. Mat. Univ. Padova, to appear.

[30] A.FIGALLI: *Existence, uniqueness and regularity of optimal transport maps.* Preprint, 2006.

[31] A.FIGALLI & C.VILLANI: *Displacement convexity on Riemannian manifolds.* Preprint, 2006.

[32] G. FORNI & J.N. MATHER: *Action minimizing orbits in Hamiltonian systems, Transition to chaos in classical and quantum mechanics, Montecatini Terme,1991.* Lecture Notes in Math., **1589** (1994), Springer,Berlin, 92-186.

[33] W. GANGBO & R.J. MCCANN: *The geometry of optimal transportation.* Acta Math., **177** (1996), 113-161.

[34] W. GANGBO: *The Monge mass transfer problem and its applications, in Monge Ampère equation: applications to geometry and optimization.* Contemp. Math., **226** (1999), Amer. Math. Soc., Providence, RI, 79-104.

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

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

[37] M.KNOTT & C.S.SMITH: *On the optimal mapping of distributions.* J. Optim. Theory Appl., **43** (1984), 39-49.

[38] J.LOTT & C.VILLANI: *Ricci curvature via optimal transport.* Preprint.

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

[40] J.N. MATHER: *Existence of quasiperiodic orbits for twist homeomorphisms of the annulus.* Topology, **21** (1982), 457-467.

[41] MCCANN: *A convexity principle for interacting gases and equilibrium crystals.* PhD thesis, Princeton University, 1994.

[42] R.J.MCCANN: *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J. 80 (1995), no. 2, 309-323.

[43] R.J.MCCANN: *A convexity principle for interacting gases.* Adv. Math., **128** (1997), 153-179.

[44] R.J.MCCANN: *Polar factorization of maps on Riemannian manifolds.* Geom. Funct. Anal., **11** (2001), no. 3, 589-608.

[45] G.MONGE: *Mémoire sur la Théorie des Déblais et des Remblais.* Hist. de l'Acad. des Sciences de Paris (1781), 666-704.

[46] J.MOSER: *Selected chapters in the calculus of variations*, Lectures in Mathematics ETH Zürich, Lecture notes by Oliver Knill, Birkhäuser Verlag, Basel (2003).

[47] F.OTTO & C.VILLANI: *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality.* J. Funct. Anal., **173** (2000), no. 2, 361-400.

[48] S.T.RACHEV & L. RÜSCHENDORF: *A characterization of random variables with minimum L^2-distance.* J. Multivariate Anal., **32** (1990), 48-54. Corrigendum in J. Multivariate Anal., **34** (1990), 156.

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

[50] W.SCHACHERMAYER & J.TEICHMANN: *Characterization of optimal transport plans for the Monge-Kantorovich-problem.* Preprint.

[51] K.T.STURM: *On the geometry of metric measure spaces.* Acta mathematica., to appear.

[52] K.T.STURM & M.K.VON RENESSE: *Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature.* Comm. Pure Appl. Math., to appear.

[53] V.N. SUDAKOV: *Geometric problems in the theory of infinite-dimensional probability distributions.* Proc. Steklov Inst. Math., **141** (1979), 1-178.

[54] N.S. TRUDINGER & X.J. WANG: *On the Monge mass transfer problem.* Calc. Var. Partial Differential Equations, **13** (2001), 19-31.

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

[56] C.VILLANI: *Optimal transport, old and new.* Lecture notes, 2005 Saint-Flour summer school, available online at `www.umpa.ens-lyon.fr/~cvillani`.

[57] R. VINTER: *Optimal control.* Systems & Control: Foundations & Applications, (2000), Birkhäuser Boston Inc., Boston, MA.