Optimal mass transportation and Mather theory

Patrick Bernard and Boris Buffoni

10 decembre 2004

Several observations have recently renewed the interest for the classical topic of optimal mass transportation, whose primary origin is attributed to Monge a few years before French revolution. The framework is as follows. A space M is given, which in the present paper will be a compact manifold, as well as a continuous cost function $c(x,y) : M \times M \longrightarrow \mathbb{R}$. Given two probability measures μ_0 and μ_1 on M, the mappings $\Psi : M \longrightarrow M$ which transport μ_0 into μ_1 and minimize the total cost $\int_M c(x, \Psi(x)) d\mu_0$ are studied. It turns out, and it was the core of the investigations of Monge, that these mappings have very remarkable geometric properties, at least at a formal level.

Only much more recently was the question of the existence of optimal objects rigorously solved by Kantorovitch in a famous paper of 1942. Here we speak of optimal objects, and not of optimal mappings, because the question of existence of an optimal mapping is ill-posed, so that the notion of optimal objects has to be relaxed, in a way that nowadays seems very natural, and that was discovered by Kantorovitch.

Our purpose here is to continue the work initiated by Monge, recently awakened by Brenier and enriched by other authors, on the study of geometric properties of optimal objects. The costs functions we consider are natural generalizations of the cost $c(x,y) = d(x,y)^2$ considered by Brenier and many other authors. The book [34] gives some ideas of the applications expected from this kind of questions. More precisely, we consider a Lagrangian function $L(x,v,t) : TM \times \mathbb{R} \longrightarrow \mathbb{R}$ which is convex in v and satisfies standard hypotheses recalled later, and define our cost by

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

where the minimum is taken on the set of curves $\gamma : [0,1] \longrightarrow M$ satisfying $\gamma(0) = x$ and $\gamma(1) = y$. Note that this class of costs does not contain the very natural cost $c(x,y) = d(x,y)$. We will study such costs in a second paper, but much weaker results are expected.

Our main result is that the optimal transports can be interpolated by measured Lipschitz laminations, or geometric currents in the sense of Ruelle and Sullivan. Interpolations of transport have already been considered by Brenier and McCann for less general cost functions, and with different purposes. Our methods are inspired by the theory of Mather, Mañé and Fathi on Lagrangian dynamics, and we will detail rigorously the relations between these theories. Roughly, they are exactly similar except that mass transportation is a Dirichlet boundary value problem, while Mather theory is a periodic boundary value Problem. We will also prove, extending works of Brenier, Gangbo, Mc Cann, Carlier , and other authors, that the optimal transportation can be performed by a Borel map with the additional assumption that the transported measure is absolutely continuous.

Various connections between Mather-Fathi theory, optimal mass transportation and Hamilton-Jacobi equations have recently been discussed, mainly at a formal level, in the literature, see for exemple [34], or [17], where they are all presented as infinite dimensional linear programming problems. This have motivated a lot of activity around the interface between Aubry-Mather theory and optimal transportation, some of which overlap partly the present work. For exemple, at the moment of submitting the paper, we have been informed of the existence of the recent preprints of De Pascal, Stella and Granieri, [31], and of Granieri, [22]. We had also been aware of a preprint by Wolansky [35] for a few weeks, which, independently, and by somewhat different methods, studies questions very similar to the ones we are interested in.

This paper emanates from the collaboration of the Authors during the end of the stay of the first author in EPFL for the academic year 2002-2003, granted by the Swiss National Fund for Research.

1 Introduction

We present the main context and the main results of the paper.

1.1 Lagrangian, Hamiltonian and cost

In all the present paper, the space M will be a compact manifold without boundary. Some standing notations are gathered in the appendix. Let us fix a positive real number T, and a Lagrangian function

$$L \in C^2(TM \times [0,T], \mathbb{R}).$$

A curve $\gamma \in C^2([0,T], M)$ is called an extremal if it is a critical point of the action

$$\int_0^T L(\gamma(t), \dot{\gamma}(t), t)dt$$

with fixed endpoints. It is called a minimizing extremal if it is minimizing the action. We assume:

convexity For each $(x,t) \in M \times [0,T]$, the function $v \longmapsto L(x,v,t)$ is convex with positive definite Hessian at each point.

superlinearity For each $(x,t) \in M \times [0,T]$, we have $L(x,v,t)/\|v\| \longrightarrow \infty$ as $\|v\| \longrightarrow \infty$.

completeness For each $(x,v,t) \in TM \times [0,T]$, there exists one and only one extremal $\gamma \in C^2([0,T], M)$ such that $(\gamma(t), \dot{\gamma}(t)) = (x,v)$.

Under these hypotheses, there exists a time-dependent vector-field E, on TM, the Euler-Lagrange vector-field, which is such that the extremals are the projection of the integral curves of E. For $0 \leqslant s \leqslant t \leqslant T$ there exists a well defined flow $\psi_s^t : TM \longrightarrow TM$ which is such that $\psi_s^s = Id$ and $\partial_t \psi_s^t = E_t \circ \psi_s^t$, where as usual E_t denotes the vector-field $E(.,t)$ on TM. This flow satisfies the relation

$$\partial_t(\pi \circ \psi_s^t)(x,v,t) = \psi_s^t(x,v). \tag{1}$$

We associate to the Lagrangian L a Hamiltonian function $H \in C^2(T^*M \times [0,T], \mathbb{R})$ given by

$$H(x,p,t) = \max_v p(v) - L(x,v,t).$$

We endow the cotangent bundle T^*M with its canonical symplectic structure, and associate to the Hamiltonian H the time-dependent vector-field Y on T^*M, which is given by

$$Y = (\partial_p H, -\partial_x H)$$

in any canonical local trivialisation of T^*M. The hypotheses on L can be expressed in terms of the function H:

convexity For each $(x,t) \in M \times [0,T]$, the function $p \longmapsto H(x,p,t)$ is convex with positive definite Hessian at each point.

superlinearity For each $(x,t) \in M \times [0,T]$, we have $H(x,p,t)/\|p\| \longrightarrow \infty$ as $\|p\| \longrightarrow \infty$.

completeness Each solution of the equation $(\dot{x}(t), \dot{p}(t)) = Y(x(t), p(t), t)$ can be extended to the interval $[0, T]$. We can then define, for each $0 \leqslant s \leqslant t \leqslant T$, the flow φ_s^t of Y from times s to time t.

In addition, the mapping $\partial_v L : TM \times [0, T] \longrightarrow T^*M \times [0, T]$ is a C^1 diffeomorphism, whose inverse is the mapping $\partial_p H$. These diffeomorphisms conjugate E and Y, and the flows ψ_s^t and φ_s^t.

For each $0 \leqslant s \leqslant t \leqslant T$, we define the cost function

$$c_s^t(x, y) = \min_{\gamma} \int_s^t L(\gamma(\sigma), \dot{\gamma}(\sigma), \sigma) d\sigma$$

where the minimum is taken on the set of curves $\gamma \in C^2([s, t], M)$ satisfying $\gamma(s) = x$ and $\gamma(t) = y$. That this minimum exists is a standard result under our hypotheses, see [28] or [18].

Proposition 1. *Let us fix a subinterval $[s, t] \subset [0, T]$. The set $\mathcal{E} \subset C^2([s, t], M)$ of minimizing extremals is compact for the C^2 topology.*

Let us mention that, for each $(x_0, s) \in M \times [0, T]$, the function $(x, t) \longmapsto c_s^t(x_0, x)$ is a viscosity solution of the Hamilton-Jacobi equation

$$\partial_t c_s^t + H(x, \partial_x c_s^t, t) = 0$$

on $M \times [s, T]$. This remark may help the reader in understanding the key role which will be played by this equation in the sequel.

1.2 Monge-Kantorovitch theory

We recall the basics of Monge-Kantorovitch duality. The proofs are available in many texts on the subjects, for example [1, 32, 34]. We assume that M is a compact manifold and that $c(x, y)$ is a continuous cost function on $M \times M$, which will later be one of the costs c_s^t defined above. Given two Borel probability measures μ_0 and μ_1 on M, a transport plan between μ_0 and μ_1 is a measure on $M \times M$ which satisfies

$$\pi_{0\sharp}(\eta) = \mu_0 \text{ and } \pi_{1\sharp}(\eta) = \mu_1,$$

where $\pi_0 : M \times M \longrightarrow M$ is the projection on the first factor, and π_1 is the projection on the second factor. We denote by $\mathcal{K}(\mu_0, \mu_1)$, after Kantorovitch, the set of transport plans. Kantorovitch proved the existence of a minimum in the expression

$$C(\mu_0, \mu_1) = \min_{\eta \in \mathcal{K}(\mu_0, \mu_1)} \int_{M \times M} c d\eta$$

for each pair (μ_0, μ_1) of probability measures on M. Here we will denote by

$$C_s^t(\mu_0, \mu_1) := \min_{\eta \in \mathcal{K}(\mu_0, \mu_1)} \int_{M \times M} c_s^t(x, y) d\eta(x, y) \tag{2}$$

the optimal value associated to our family of costs c_s^t. The plans which realize this minimum are called optimal transfer plans. A pair (ϕ_0, ϕ_1) of continuous functions is called an admissible Kantorovitch pair if is satisfies the relations

$$\phi_1(x) = \min_{y \in M} \phi_0(y) + c(y, x) \text{ and } \phi_0(x) = \max_{y \in M} \phi_1(y) - c(x, y)$$

for all point $x \in M$. Note that the admissible pairs are composed of Lipschitz functions if the cost c is Lipschitz, which is the case of the costs c_s^t we are going to consider. Another discovery of Kantorovitch is that

$$C(\mu_0, \mu_1) = \max_{\phi_0, \phi_1} \Big(\int_M \phi_1 d\mu_1 - \int_M \phi_0 d\mu_0 \Big) \tag{3}$$

where the maximum is taken on the set of admissible Kantorovitch pairs (ϕ_0, ϕ_1). This maximization problem is called the dual Kantorovitch problem, the admissible pairs which reach this maximum are called optimal Kantorovitch pairs. The direct problem (2) and dual problem (3) are related as follows.

Proposition 2. *If η is an optimal transfer plan, and if (ϕ_0, ϕ_1) is a Kantorovitch optimal pair, then the support of η is contained in the set*

$$\{(x, y) \in M^2 \text{ such that } \phi_1(y) - \phi_0(x) = c(x, y)\} \subset M \times M.$$

Let us remark that the knowledge of the set of Kantorovitch admissible pairs is equivalent to the knowledge of the cost function c.

Lemma 3. *We have*

$$c(x, y) = \max_{(\phi_0, \phi_1)} \phi_1(y) - \phi_0(x)$$

where the maximum is taken on the set of Kantorovitch admissible pairs.

Proof. This maximum is clearly less that $c(x, y)$. For the other inequality, let us fix points x_0 and y_0 in M, and consider the functions $\phi_1(y) = c(x_0, y)$ and $\phi_0(x) = \max_{y \in M} \phi_1(y) - c(x, y)$. We have $\phi_1(y_0) - \phi_0(x_0) = c(x_0, y_0) - 0 = c(x_0, y_0)$. So it is enough to prove that the pair (ϕ_0, ϕ_1) is an admissible Kantorovitch pair, and more precisely that $\phi_1(y) = \min_{x \in M} \phi_0(x) + c(x, y)$. We have

$$\phi_0(x) + c(x, y) \geqslant c(x_0, y) - c(x, y) + c(x, y) \geqslant c(x_0, y) = \phi_1(x)$$

which gives the inequality $\phi_1(y) \leqslant \min_{x \in M} \phi_0(x) + c(x, y)$. On the other hand, we have

$$\min_{x \in M} \phi_0(x) + c(x, y) \leqslant \phi_0(x_0) + c(x_0, y) = c(x_0, y) = \phi_1(y).$$

□

1.3 Interpolations

In this section, the Lagrangian L and time $T > 0$ are fixed. It is not hard to see that, if μ_1, μ_2 and μ_3 are three probability measures on M, and if $t_1 \leqslant t_2 \leqslant t_3 \in [0, T]$ are three times, then we have the inequality

$$C_{t_1}^{t_3}(\mu_1, \mu_3) \leqslant C_{t_1}^{t_2}(\mu_1, \mu_2) + C_{t_2}^{t_3}(\mu_2, \mu_3).$$

The family $\mu_t, t \in [0, T]$ of probability measures on M is called an interpolation between μ_0 and μ_T if it satisfies the equality

$$C_{t_1}^{t_3}(\mu_{t_1}, \mu_{t_3}) = C_{t_1}^{t_2}(\mu_{t_1}, \mu_{t_2}) + C_{t_2}^{t_3}(\mu_{t_2}, \mu_{t_3})$$

for all $0 \leqslant t_1 \leqslant t_2 \leqslant t_3 \leqslant T$. Our main result is the following:

Theorem A. *For each pair μ_0, μ_T of probability measures, there exist interpolations between μ_0 and μ_T. Moreover, each interpolation $(\mu_t), t \in [0, T]$ is given by a Lipschitz measured lamination in the following sense:*

Eulerian description : *There exists a locally Lipschitz vector-field $X(x, t) : M \times]0, T[\longrightarrow TM$ such that, if $\Psi_s^t, (s, t) \in]0, T[^2$ is the flow of X from time s to time t, then $(\Psi_s^t)_\sharp \mu_s = \mu_t$ for each $(s, t) \in]0, T[^2$.*

Lagrangian description : *There exists a family $\mathcal{F} \subset C^2([0, T], M)$ of minimizing extremals γ of L, which is such that the relation $\dot\gamma(t) = X(\gamma(t), t)$ holds for each $t \in]0, T[$ and for each $\gamma \in \mathcal{F}$. The set*

$$\tilde{\mathcal{T}} = \{(\gamma(t), \dot\gamma(t), t), t \in]0, T[, \gamma \in \mathcal{F}\} \in TM \times]0, T[$$

is invariant under the Euler-Lagrange flow ψ. *The measure* μ_t *is supported on* $\mathcal{T}_t = \{\gamma(t), \gamma \in \mathcal{F}\}$. *In addition, there exists a continuous family* $m_t, t \in [0,T]$ *of probability measures on* TM *such that* m_t *is concentrated on* $\tilde{\mathcal{T}}_t = \{(\gamma(t), \dot\gamma(t)), \gamma \in \mathcal{F}\}$ *for each* $t \in]0,T[$, *such that* $\pi_\sharp m_t = \mu_t$ *for each* $t \in [0,T]$, *and such that*

$$m_t = (\psi_s^t)_\sharp m_s$$

for all $(s,t) \in [0,T]^2$.
Hamilton-Jacobi equation : *There exists a* C^1 *function* $v(x,t) : M \times]0,T[\longrightarrow \mathbb{R}$ *which satisfies the inequation*

$$\partial_t v + H(x, \partial_x v, t) \leqslant 0,$$

with equality if and only if $(x,t) \in \mathcal{T} = \{(\gamma(t), t), \gamma \in \mathcal{F}, t \in]0,T[\}$, *and such that* $X(x,t) = \partial_p H(x, \partial_x v(x,t), t)$ *for each* $(x,t) \in \mathcal{T}$.
Uniqueness : *There may exist several different interpolations. However, it is possible to choose the vector-field* X, *the family* $\mathcal{F}$ *and the sub-solution* v *in such a way that the statements above hold for all interpolations* μ_t *with these fixed* X, $\mathcal{F}$ *and* v. *For each* $s < t \in]0,T[$, *the measure* $(Id \times \Psi_s^t)_\sharp \mu_s$ *is the only optimal transport plan in* $\mathcal{K}(\mu_s, \mu_t)$ *for the cost* c_s^t. *This implies that*

$$\int_M c_s^t(x, \Psi_s^t(x)) d\mu_s(x) = C_s^t(\mu_s, \mu_t).$$

Let us comment a bit the preceding statement. The set $\tilde{\mathcal{T}} \subset TM \times]0,T[$ is the image by the Lipschitz vector-field X of the set $\mathcal{T} \subset TM \times]0,T[$. We shall not take $X(x,t) = \partial_p H(x, \partial_x u(x,t), t)$ outside of $\mathcal{T}$ because we do not prove that this vector-field is Lipschitz outside of $\mathcal{T}$. The data of the vector-field X outside of $\mathcal{T}$ is immaterial: any Lipschitz extension of $X_{|\mathcal{T}}$ will fit. Note also that the relation

$$\Psi_s^t = \pi \circ \psi_s^t \circ X_s$$

holds on $\mathcal{T}_s$, where $X_s(.) = X(., s)$.

The vector-field X in the statement depends on the transported measures μ_0 and μ_T. The Lipschitz constant of X, however, can be fixed independently of these measures, as we now state:

Addendum *There exists a decreasing function* $K(\epsilon) :]0, T/2[\longrightarrow]0, \infty[$, *which depends only on the time* T *and on the Lagrangian* L, *and such that, for each pair* μ_0, μ_T *of probability measures, one can choose the vector-field* X *in Theorem A in such a way that* X *is* $K(\epsilon)$*-Lipschitz on* $[\epsilon, T-\epsilon]$ *for each* $\epsilon \in]0, T/2[$.

Proving Theorem A is the main goal of the present paper. We will present in section 2 some direct variational problems which are well-posed and of which the transport interpolations are in some sense the solutions. We believe that these variational problems are interesting in themselves. In order to describe the solutions of the variational problem, we will rely on a dual approach based on the Hamilton-Jacobi equation, inspired from Fathi's approach to Mather theory, as detailed in section 3. The solutions of the problems of section 2, as well as the transport interpolations, are then described in section 4, which ends the proof of Theorem A.

1.4 Case of an absolutely continuous measure μ_0

Additional conclusions concerning optimal transport can usually be obtained when the initial measure μ_0 is absolutely continuous. For example a standard question is whether the optimal transport can be realized by an optimal mapping.

A transport map is a Borel map $\Psi : M \longrightarrow M$ which satisfies $\Psi_\sharp \mu_0 = \mu_1$. To any transport map Ψ is naturally associated the transport plan $(Id \times \Psi)_\sharp \mu_0$, called the induced transport plan. An optimal map is a transport map $\Psi : M \longrightarrow M$ such that

$$\int_M c_T(x, \Psi(x)) d\mu_0 \leqslant \int_M c_T(x, F(x)) d\mu_0$$

for any transport map F. It turns out that, under the assumption that μ_0 has no atoms, a transport map is optimal if and only if the induced transport plan is an optimal transport plan, see [1], Theorem 2.1. In other words, we have

$$\inf_\Psi \int_M c(x, \Psi(x)) d\mu_0(x) = C(\mu_0, \mu_1),$$

where the minimum is taken on the set of transport maps from μ_0 to μ_1. This is a general result which holds for any continuous cost c. It is a standard question, which turns out to be very hard for certain cost functions, whether the infimum above is reached, or in other words whether there exists an optimal transport plan which is induced from a transport map. Part of the result below is that this holds true in the case of the cost c_0^T. The method we use to prove this is an elaboration on ideas due to Brenier, see [10] and developed for instance in [21], (see also [20]) and [13], which is certainly the closest to our needs.

Theorem B. *Assume that μ_0 is absolutely continuous with respect to the Lebesgue class on M. Then for each final measure μ_T, there exists one and only one interpolation $\mu_t, t \in [0, T]$, and each interpolating measure $\mu_t, t < T$ is absolutely continuous. In addition, there exists a family $\Psi_0^t, t \in [0, T]$ of Borel maps such that $(Id \times \Psi_0^t)_\sharp \mu_0$ is the only optimal transfer plan in $\mathcal{K}(\mu_0, \mu_t)$ for the cost function c_0^t. Consequently, we have*

$$\int_M c_0^t(x, \Psi_0^t(x)) d\mu_0(x) = C_0^t(\mu_0, \mu_t).$$

If μ_T, instead of μ_0, is assumed absolutely continuous, then there exists one and only one interpolation, and each interpolating measure $\mu_t, t \in]0, T]$ is absolutely continuous.

This theorem will be proved and commented in section 5.

1.5 Mather theory

Let us now assume that the Lagrangian function is defined for all times, $L \in C^2(TM \times \mathbb{R}, \mathbb{R})$ and, in addition to the standing hypotheses, satisfies the periodicity condition

$$L(x, v, t + 1) = L(x, v, t)$$

for all $(x, v, t) \in TM \times \mathbb{R}$. A Mather measure, see [28], is a compactly supported probability measure m_0 on TM which is invariant in the sense that $(\psi_0^1)_\sharp m_0 = m_0$ and is minimizing the action

$$A_0^1(m_0) = \int_{TM \times [0,1]} L(\psi_0^t(x, v), t) dm_0 dt.$$

The major discovery of [28] is that Mather measures are supported on the graph of a Lipschitz vector-field. Let us call α the action of Mather measures –this number is the value at zero of the α function defined by Mather in [28]. Let us now explain how this theory of Mather is related to, and can be recovered from, the content of our paper.

Theorem C. *We have*

$$\alpha = \min_{\mu} C_0^1(\mu, \mu),$$

where the minimum is taken on the set of probability measures on M. The mapping $m_0 \longmapsto (\pi)_\sharp m_0$ is a bijection between the set of Mather measures m_0 and the set of probability measures μ on M satisfying $C_0^1(\mu, \mu) = \alpha$. There exists a Lipschitz vector-field X_0 on M such that all the Mather measures are supported on the graph of X_0.

This theorem will be proved in section 6, where the bijection between Mather measures and measures minimizing $C_0^1(\mu, \mu)$ will be precised.

2 Direct variational problems

We state two different variational problems whose solutions are the interpolated transports. We believe that these problems are interesting on their own. They will also be used to prove Theorem A.

2.1 Measures

This formulation parallels Mather's theory. Let μ_0 and μ_T be two probability Borel measures on M. Let $m_0 \in \mathcal{B}_1(TM)$ be a Borel probability measure on the tangent bundle TM. We say that m_0 is an initial transport measure if the measure η on $M \times M$ given by

$$\eta = (\pi \times \pi \circ \psi_0^T)_\sharp m_0$$

is a transport plan, where $\pi : TM \longrightarrow M$ is the canonical projection. We call $\mathcal{I}(\mu_0, \mu_1)$ the set of initial transport measures. To an initial transport measure m_0, we associate the continuous family of measures

$$m_t = (\psi_0^t)_\sharp m_0, t \in [0, T]$$

on TM, and the measure m on $TM \times [0, T]$ given by

$$m = \int_0^T m_t dt = \int_0^T (\psi_0^t)_\sharp m_0 \, dt.$$

Note that the linear mapping $m_0 \longmapsto m = \int_0^T (\psi_0^t)_\sharp m_0 \, dt$ is continuous from $\mathcal{B}(TM)$ to $\mathcal{B}(TM \times [0, T])$ endowed with the weak topology, see appendix.

Lemma 4. *The measure m satisfies the relation*

$$\int_{TM \times [0,T]} \partial_t f + \partial_x f(v) dm = \int_M f_T d\mu_T - \int_M f_0 d\mu_0 \tag{4}$$

for each function $f \in C^1(M \times [0, T], \mathbb{R})$, where f_t denotes the function $x \longmapsto f(x, t)$.

Proof. We have

$$\int_{TM \times [0,T]} \partial_t f + \partial_x f(v) dm = \int_0^T \int_{TM} (\partial_t f + \partial_x f) \circ \psi_0^t dm_0 dt.$$

Noticing that, in view of equation (1), we have

$$\partial_t (f \circ \pi \circ \psi_0^t) = \partial_t f \circ \psi_0^t + \partial_x f \circ \psi_0^t$$

we obtain that

$$\int_{TM\times[0,T]} \partial_t f + \partial_x f(v) dm = \int_{TM} (f\circ\pi - f\circ\pi\circ\psi_0^T) dm_0 = \int_M f_T d\mu_T - \int_M f_0 d\mu_0$$

as desired. □

Definition 5. *A finite Borel measure on $TM \times [0,T]$ which satisfies (4) is called a transport measure. We denote by $\mathcal{M}(\mu_0,\mu_1)$ the set of transport measures. A transport measure which is induced from an initial measure m_0 is called an invariant transport measure. The action of the transport measure m is defined by*

$$A(m) = \int_{TM\times[0,T]} L(x,v,t) dm \in \mathbb{R}\cup\{\infty\}$$

The action $A(m_0)$ of an initial transport measure is defined as the action of the associated transport measure m. We will also denote this action by $A_0^T(m_0)$ when we want to insist on the time interval. We have

$$A_0^T(m_0) = \int_{TM\times[0,T]} L(\psi_0^t(x,v),t) dm_0 dt.$$

Although we are going to build minimizers by other means, we believe the following result is worth being mentioned.

Lemma 6. *For each real number a, the set $\mathcal{M}^a(\mu_0,\mu_1)$ of transport measures m which satisfy $A(m)\leqslant a$, as well as the set $\mathcal{I}^a(\mu_0,\mu_1)$ of initial transport measures m_0 which satisfy $A_0^T(m_0)\leqslant a$, are compact. As a consequence, there exist optimal initial transport measures, and optimal transport measures.*

PROOF. This is an easy application of the Prohorov theorem, see [8]. □

Now we have seen that the problem of finding optimal transport measures is well-posed, let us describe its solutions.

Theorem 1. *We have*

$$C_0^T(\mu_0,\mu_T) = \min_{m\in\mathcal{M}(\mu_0,\mu_T)} A(m) = \min_{m_0\in\mathcal{I}(\mu_0,\mu_T)} A(m_0).$$

The mapping

$$m_0 \longmapsto m = \int_0^T (\psi_0^T)_\sharp m_0 \, dt$$

between the set $\mathcal{OI}$ of optimal initial measures and the set $\mathcal{OM}$ of optimal transport measures is a bijection. There exists a locally Lipschitz vector-field $X(x,t): M\times]0,T[\longrightarrow TM$ such that, for each optimal initial measure $m_0\in\mathcal{OI}$, the measure $m_t=(\psi_0^t)_\sharp m_0$ is supported on the graph of X_t for each $t\in]0,T[$.

The proof will be given in section 4.3. Let us just notice now that the inequalities

$$C_0^T(\mu_0,\mu_T) \geqslant \min_{m_0\in\mathcal{I}(\mu_0,\mu_T)} A(m_0) \geqslant \min_{m\in\mathcal{M}(\mu_0,\mu_T)} A(m)$$

hold in view of:

Proposition 7. *The mapping* $(\pi \times \pi \circ \psi_0^T)_\sharp : \mathcal{I}(\mu_0, \mu_1) \longrightarrow \mathcal{K}(\mu_0, \mu_1)$ *is surjective. In addition, for each transport plan* η*, there exists a compactly supported initial transport measure* m_0 *such that* $(\pi \times \pi \circ \psi_0^T)_\sharp m_0 = \eta$ *and such that*

$$A(m_0) = \int_{M \times M} c_0^T(x, y) d\eta.$$

Proof. There exists a compact set $K \in TM$ such that if $\gamma(t) : [0, T] \longrightarrow M$ is a minimizing extremal, then the lifting $(\gamma(t), \dot{\gamma}(t))$ is contained in K for each $t \in [0, T]$. We shall prove that, for each probability measure $\eta \in \mathcal{B}(M \times M)$, there exists a probability measure $m_0 \in \mathcal{B}(K)$ such that $(\pi \times \pi \circ \psi_0^T)_\sharp m_0 = \eta$ and such that

$$A(m_0) = \int_{M \times M} c_0^T(x, y) d\eta.$$

Observing that

- the mappings $m_0 \longrightarrow (\pi \times \pi \circ \psi_0^T)_\sharp m_0$ and $m_0 \longmapsto A(m_0)$ are linear and continuous on the space $\mathcal{B}_1(K)$ of probability measures supported on K,
- the set $\mathcal{B}_1(K)$ is compact for the weak topology, and the action A is continuous on this set,
- the set of probability measures on $M \times M$ is the compact convex closure of the set of Dirac probability measures (probability measures supported in one point), see e. g. [8], p. 73,

it is enough to prove the result when η is a Dirac probability measure (or equivalently when μ_0 and μ_1 are Dirac probability measures). Let η be the Dirac probability measure supported at $(x_0, x_1) \in M \times M$. Let $\gamma(t) : [0, T] \longrightarrow M$ be a minimizing extremal with boundary conditions $\gamma(0) = x_0$ and $\gamma(T) = x_1$. In view of the choice of K, we have $(\gamma(0), \dot{\gamma}(0)) \in K$. Let m_0 be the Dirac probability measure supported at $(\gamma(0), \dot{\gamma}(0))$. It is straightforward that m_t is then the Dirac measure supported at $(\gamma(t), \dot{\gamma}(t))$, so that

$$A(m_0) = \int_0^T L dm_t dt = \int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt = c_0^T(x_0, x_1) = \int_{M \times M} c_0^T d\eta$$

and

$$(\pi \times \pi \circ \psi_0^T)_\sharp m_0 = \eta.$$

□

2.2 Currents

This formulation finds its roots on one hand in the works of Benamou and Brenier, see [4], and then Brenier, see [11], and on the other hand in the work of Bangert [3]. Let $\Omega^0(M \times [0, T])$ be the set of continuous one-forms on $M \times [0, T]$, endowed with the uniform norm. We will often decompose forms $\omega \in \Omega^0(M \times [0, T])$ as

$$\omega = \omega^x + \omega^t dt,$$

where ω^x is a time-dependent form on M and ω^t is a continuous function on $M \times [0, T]$. To each continuous linear form χ on $\Omega^0(M \times [0, T])$, we associate its time component μ_χ, which is the measure on $M \times [0, T]$ defined by

$$\int_{M \times [0,T]} f d\mu_\chi = \chi(f dt)$$

for each continuous function f on $M \times [0,T]$. A Transport current between μ_0 and μ_1 is a continuous linear form χ on $\Omega^0(M \times [0,T])$ which satisfies the two conditions:

1. The measure μ_χ is a non-negative.
2. $d\chi = \mu_1 \otimes \delta_T - \mu_0 \otimes \delta_0$, which means that
$$\chi(df) = \int_M f_T d\mu_1 - \int_M f_0 d\mu_0$$
for each smooth (or equivalently C^1) function $f : M \times [0,T] \longrightarrow \mathbb{R}$.

We call $\mathcal{C}(\mu_0, \mu_1)$ the set of transport currents from μ_0 to μ_1. The set $\mathcal{C}(\mu_0, \mu_1)$ is a closed affine subspace of $\left[\Omega^0(M \times [0,1])\right]^*$. We will endow $\mathcal{C}(\mu_0, \mu_1)$ with the weak topology obtained as the restriction of the weak-$*$ topology of $\left[\Omega^0(M \times [0,1])\right]^*$. Transport currents should be thought of as vector-fields whose components are measures. If Z is a bounded measurable vector-field on $M \times [0,T]$, and if μ is a non-negative measure on $M \times [0,T]$, we define the current $Z \wedge \mu$ by
$$Z \wedge \mu(\omega) := \int_{M \times [0,T]} \omega(Z) d\mu.$$
Every transport current can be written this way, but we shall neither prove nore use this fact. Here are some examples.

Regular transport currents. If X is a bounded measurable section of the projection $TM \times [0,T] \longrightarrow M \times [0,T]$, and if μ is a non-negative measure on $M \times [0,T]$, the current $(X,1) \wedge \mu$ is called a regular transport current. The time component of the current $(X,1) \wedge \mu$ is μ. In addition, if $(X,1) \wedge \mu = (X',1) \wedge \mu$ for two vector-fields X and X', then X and X' agree μ-almost everywhere.

Smooth transport currents. A current is said smooth if it can be written on the form $(X,1) \wedge \lambda$ with a vector-field X which is smooth on $M \times]0,T[$ and a smooth measure λ on $M \times]0,T[$ equivalent to the Lebesgue measure (this means that the measure λ has a positive smooth density in any chart). By standard regularisation arguments, we see that the set of smooth transport currents is dense in $\mathcal{C}(\mu_0, \mu_1)$.

Lipschitz regular transport currents. A transport current is said Lipschitz regular if it can be written on the form $(X,1) \wedge \mu$ with a vector-field X which is locally Lipschitz on $M \times]0,T[$. Smooth currents are Lipschitz regular. Lipschitz regular transport currents have a remarkable structure: If $\chi = (X,1) \wedge \mu_\chi$ is a Lipschitz regular transport current with X locally Lipschitz on $M \times]0,T[$, there exists a unique continuous path $\mu_\chi^t : [0,T] \longmapsto \mathcal{B}(M)$ of measures on M, such that $\mu_\chi = \int_0^T \mu_\chi^t dt$, or in details
$$\int_{M \times [0,T]} f d\mu_\chi = \int_0^T \left(\int_M f_t d\mu_\chi^t \right) dt$$
for each continuous function $f : M \times [0,T] \longrightarrow \mathbb{R}$, where f_t denotes the function $x \longmapsto f(x,t)$. In addition, If $\Psi_s^t, (s,t) \in]0,T[^2$, denotes the flow of the Lipschitz vector-field X from time s to time t, then Ψ_s^t is a bi-Lipschitz homeomorphism of M, and we have
$$(\Psi_s^t)_\sharp \mu_\chi^s = \mu_\chi^t$$
for all $(s,t) \in]0,T[^2$.

Transport current induced from a transport measure. To a transport measure m, we associate the transport current χ_m defined by
$$\chi_m(\omega) = \int_{TM \times [0,T]} \omega^x + \omega^t dm$$

where the form ω is decomposed as $\omega = \omega^x + \omega^t dt$, with ω^x considered as a function on $TM \times [0,T]$ and ω^t as a function on $M \times [0,T]$, and then as a fiberwise constant function on $TM \times [0,T]$. Note that the time component of the current χ_m is $\pi_\sharp m$. We will see that

$$A(\chi_m) \leqslant A(m)$$

with the following definition of the action $A(\chi)$ of a current, with equality if m is concentrated on the graph of a continuous vector-field.

Definition 8. *We define the action of the transport current χ by the formulas*

$$\begin{aligned}
A(\chi) &= \sup_{\omega\in\Omega^0} \Big(\chi(\omega^x,0) - \int_{M\times[0,T]} H(x,\omega^x(x,t),t)d\mu_\chi\Big)\\
&= \sup_{\omega\in\Omega^0} \Big(\chi(\omega) - \int_{M\times[0,T]} \big(H(x,\omega^x(x,t),t)+\omega^t\big)d\mu_\chi\Big)\\
&= \sup_{\omega\in\Omega^0} \Big(\chi(\omega) - T \sup_{(x,t)M\times[0,T]} \big(H(x,\omega^x(x,t),t)+\omega^t\big)\Big)\\
&= \sup_{\omega\in\Omega^0;\omega^t+H(x,\omega^x,t)\leqslant 0} \chi(\omega)\\
&= \sup_{\omega\in\Omega^0;\omega^t+H(x,\omega^x,t)\equiv 0} \chi(\omega).
\end{aligned}$$

PROOF. Let us prove that these expressions are equal. We denote for the moment by A_1, A_2, A_3, A_4 and A_5 these a priori different actions. It is straightforward that $A_1 = A_2$, this just amounts to simplify the term $\int \omega^t d\mu_\chi$. Since μ_χ is a non-negative measure which satisfies $\int_{M\times[0,T]} 1 d\mu_\chi = T$, we have

$$\int_{M\times[0,T]} \big(H(x,\omega^x(x,t),t)+\omega^t\big) \leqslant T \sup_{(x,t)\in M\times[0,T]} \big(H(x,\omega^x(x,t),t)+\omega^t\big)$$

so that $A_3(\chi) \leqslant A_2(\chi)$. In addition, we obviously have $A_5(\chi) \leqslant A_4(\chi) \leqslant A_3(\chi)$. Now notice, in A_1, that the quantity

$$\chi(\omega) - \int_{M\times[0,T]} \big(H(x,\omega^x(x,t),t)+\omega^t\big)d\mu_\chi$$

does not depend on ω^t. Let us consider the form $\tilde\omega = (\omega^x, -H(x,\omega^x,t))$, which satisfies the equality $H(x,\tilde\omega^x,t) + \tilde\omega^t \equiv 0$. We get, for each form ω,

$$\chi(\omega^x,0) - \int_{M\times[0,T]} H(x,\omega^x(x,t),t)d\mu_\chi = \chi(\tilde\omega) \leqslant A_5(\chi)$$

Hence $A_1(\chi) \leqslant A_5(\chi)$. □

Although we are going to provide explicitly a minimum of A, we believe the following Lemma is worth being mentioned.

Lemma 9. *The functional $A : \mathcal{C}(\mu_0,\mu_1) \longrightarrow \mathbb{R}$ is convex, lower semi-continuous and coercive, hence it has a minimum.*

PROOF. Let us define the continuous convex function $\mathbb{H} : \Omega^0(M\times[0,T]) \longrightarrow \mathbb{R}$ by

$$\mathbb{H}(\omega) = T \sup_{(x,t)\in M\times[0,T]} H(x,\omega^x(x,t),t) + \omega^t.$$

Then the action is the restriction to $\mathcal{C}(\mu_0,\mu_T)$ of the Fenchel conjugate $A = \mathbb{H}^* : [\Omega^0(M\times[0,T])]^* \longrightarrow \mathbb{R}$. □

Theorem 2. *We have*

$$C_0^T(\mu_0,\mu_1)=\min_{\chi\in\mathcal{C}(\mu_0,\mu_T)} A(\chi)$$

where the minimum is taken on all transport currents from μ_0 to μ_T. Every optimal transport current is Lipschitz regular. Let $\chi=(X,1)\wedge\mu$ be an optimal transport current, with X locally Lipschitz on $M\times]0,T[$. The measure $m=(X\times\tau)_\sharp\mu\in\mathcal{B}_+(TM\times]0,T[)$ is an optimal transport measure, and χ is the transport current induced from m. Here $\tau:TM\times[0,T]\longrightarrow[0,T]$ is the projection on the second factor, see appendix. We have

$$C_0^T(\mu_0,\mu_1)=A(m)=A(\chi)=\int_{M\times[0,T]} L(x,X(x,t),t)d\mu_\chi.$$

This result will be proved in 4.1 after some essential results on the dual approach have been established. Let us continue here with some useful related remarks.

Lemma 10. *We have*

$$A((X,1)\wedge\mu)=\int_{M\times[0,T]} L(x,X(x,t),t)d\mu$$

for each current $(X,1)\wedge\mu$ with X continuous. If m is a transport measure, and if χ_m is the associated transport current, then $A(\chi_m)\leqslant A(m)$, with equality if m is supported on a continuous graph. As a consequence, we have the inequalities

$$C_0^T(\mu_0,\mu_1)\geqslant \min_{m_0\in\mathcal{I}(\mu_0,\mu_T)} A(m_0)\geqslant \min_{m\in\mathcal{M}(\mu_0,\mu_T)} A(m)\geqslant \min_{\chi\in\mathcal{C}(\mu_0,\mu_T)} A(\chi).$$

Proof. For each continuous form ω, we have

$$\int_{M\times[0,T]} \omega^x(X)-H(x,\omega^x(x,t),t)d\mu\leqslant\int_{M\times[0,T]} L(x,X(x,t),t)d\mu,$$

so that

$$A((X,1)\wedge\mu)\leqslant\int_{M\times[0,T]} L(x,X(x,t),t)d\mu.$$

On the other hand, taking the form $\omega_0^x(x,t)=\partial_v L(x,X(x,t),t)$ –which is continuous because X is continuous– we obtain the pointwise equality

$$L(x,X(x,t),t)=\omega_0^x(X)-H(x,\omega_0^x(x,t),t)$$

and by integration

$$\int_{M\times[0,T]} L(x,X(x,t),t)d\mu=\int_{M\times[0,T]} \omega_0^x(X)-H(x,\omega_0^x(x,t),t)d\mu\leqslant A((X,1)\wedge\mu).$$

This ends the proof of the equality of the two forms of the action "continuous" currents. Now if χ_m is the current associated to a transport measure m, then we have, for each continuous form $\omega\in\Omega^0(M\times[0,T])$,

$$\chi_m(\omega)-\int_{M\times[0,T]} H(x,\omega^x(x,t),t)d\mu=\int_{TM\times[0,T]} \omega^x(v)-H(x,\omega^x(x,t),t)dm$$

by definition of m, so that

$$A(\chi_m)\leqslant\int_{M\times[0,T]} L(x,v,t)dm=A(m)$$

by the Legendre inequality. In addition, if there exists a vector-field $X:M\times[0,T]\longrightarrow TM$ such that the graph of $X\times\tau$ supports m, then we can consider the form ω_0^x associated to X as above, and we get the equality for this form. □

For completeness, we state the following result, which can be proved in the line of [3].

Proposition 11. *Let χ be a transport current of finite action. Then, there exists a Borel measure $\bar{\chi}$ on $C^0([0,T],M)$, which is concentrated on the set of absolutely continuous curves, and such that*

$$\chi(\omega) = \int_{C^0([0,T],M)} \left(\int_0^T \omega^x_{\gamma(t)}(\dot{\gamma}(t)) + \omega^t dt \right) d\bar{\chi}(\gamma)$$

for each $\omega \in \Omega^0$, and

$$A(\chi) = \int_{C^0([0,T],M)} \left(\int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt \right) d\bar{\chi}(\gamma).$$

If, in addition, the current χ is optimal, then the measure $\bar{\chi}$ is concentrated on the set of minimizing extremals, and $(ev_0)_\sharp \bar{\chi}$ is an optimal initial measure, where ev_0 is the evaluation map $\gamma \longmapsto \gamma(0)$.

Proof. We only give a sketch, since this Proposition will not be used. By a regularisation procedure, one can build a sequence χ_n of smooth transport currents such that $\chi_n \longrightarrow \chi$ and $A(\chi_n) \longrightarrow A(\chi)$. For the smooth current χ_n, we can use the structure of Lipschitz currents above to see that there exists a probability measure $\bar{\chi}_n$ on $C^0([0,T],M)$, concentrated on the set of smooth curves, such that the desired conclusion holds. Let $K(a)$ be the subset of $C^0([0,T],M)$ formed by the absolutely continuous curves $\gamma(t)$ which satisfy

$$\int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt \leqslant a.$$

It is standard that the set $K(a)$ is compact in $C^0([0,T],M)$ for all a, see [28] or [18]. Since

$$A(\chi_i) = \int_{C^0([0,T],M)} \int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt d\bar{\chi}_i(\gamma)$$

we conclude that

$$\bar{\chi}_i\big(C^0([0,T],M) - K(a)\big) \leqslant A(\chi_i)/a \leqslant 2A(\chi)/a$$

when i is large enough. Hence the sequence $\bar{\chi}_n$ is a tight sequence of probability measures on $C^0([0,T],M)$, see [8] for the definition of tight sequences of measures. By the Prohorov theorem, the sequence $\bar{\chi}_n$ has an accumulation point $\bar{\chi}$ for the weak topology. It is not hard to see that the measure $\bar{\chi}$ satisfies the desired conclusions. □

3 Hamilton-Jacobi equation

Most of the results stated so far can be proved by direct approaches using Mather's shortening Lemma, which in a sense is an improvement of the initial observation of Monge, see [28] and [3]. We shall however base our proofs on the use of the Hamilton-Jacobi equation, in the spirit of Fathi's approach to Mather theory, see [18], which should be associated to Kantorovitch dual approach of the transportation problem.

3.1 Viscosity solutions and semi-concave functions

It is certainly useful to recall the main properties of viscosity solutions in connection with semi-concave functions. We will not give proofs, and instead refer to [18], [19], [12], as well as the appendix in [6]. We will consider the Hamilton-Jacobi equation

$$\partial_t u + H(x, \partial_x u, t) = 0 \qquad (HJ)$$

as well as the backward Hamilton-Jacobi equation

$$-\partial_t u - H(x, \partial_x u, t) = 0 \tag{$\breve{HJ}$}$$

which are different in the sense of viscosity solutions. The function $u : M \times [0,T] \longrightarrow M$ is called K-semi-concave if, for each chart $\theta \in \Theta$ (see appendix), the function

$$(x,t) \longmapsto u(\theta(x),t) - K(\|x\|^2 + t^2)$$

is concave on $B_2 \times [0,T]$. The function u is called semi-concave if it is K-semi-concave for some K. A function $u : M \times]0,T[\longrightarrow M$ is called locally semi-concave if it is semi-concave on each $M \times [s,t]$, for $0 < s < t < T$. The following regularity result follows from Fathi's work, see [18] and also [6].

Proposition 12. *Let u_1 and u_2 be two K-semi-concave functions. Let A be the set of minima of the function $u_1 + u_2$. Then the functions u_1 and u_2 are differentiable on A, and $du_1(x,t) + du_2(x,t) = 0$ at each point of $(x,t) \in A$. In addition, the mapping $du_1 : M \times [0,T] \longrightarrow T^*M$ is CK-Lipschitz continuous on A, where C is a universal constant.*

Viscosity sub-solutions. Let $u : M \times [0,T] \longrightarrow \mathbb{R}$ be a continuous function. The following properties are equivalent:

1. The function u is a viscosity sub-solution of (HJ), *i.e.* each smooth function $f : M \times [0,T] \longrightarrow \mathbb{R}$ satisfies the inequality $\partial_t f(x,t) + H(x, \partial_x f(x,t), t) \leqslant 0$ at each point of minimum (x,t) of the difference $f - u$.

2. The function u is a viscosity super-solution of $(\breve{HJ})$, *i.e.* each smooth function $f : M \times [0,T] \longrightarrow \mathbb{R}$, satisfies the inequality $-\partial_t f(x,t) - H(x, \partial_x f(x,t), t) \geqslant 0$ at each point of maximum (x,t) of the difference $f - u$.

3. The function u is locally Lipschitz on $M \times]0,T[$ and satisfies the equation (HJ) at almost every point.

4. We have the inequalities $u(x,t) \leqslant u(y,s) + c_s^t(y,x)$ for all $0 \leqslant s \leqslant t \leqslant T$ and all x and y in M.

Viscosity solutions. Let $u : M \times [0,T] \longrightarrow \mathbb{R}$ be a continuous function. The following properties are equivalent:

1. The function u is a viscosity solution of (HJ), *i.e.* it is a viscosity sub-solution of (HJ) and it is a viscosity super-solution, which means that each smooth function $f : M \times [0,T] \longrightarrow \mathbb{R}$ satisfies the inequality $\partial_t f + H(x, \partial_x f(x,t), t) \geqslant 0$ at each point of maximum (x,t) of the difference $f - u$.

2. The function u is the maximum of all viscosity sub-solutions v of (HJ) which satisfy the initial condition $v_0 = u_0$.

3. We have the relation $u(x,t) = \min_{y \in M} u(y,s) + c_s^t(y,x)$ for all $0 \leqslant s \leqslant t \leqslant T$.

4. The function u is locally semi-concave on $M \times]0,T]$ and it solves (HJ) at almost every point.

As a consequence, for each function u_0 on M, there exists one and only one viscosity solution u of (HJ) on $M \times [0,T]$ with u_0 as initial condition, it is locally semi-concave on $M \times]0,T]$ and given explicitly by $u(x,t) = \min_{y \in M} u_0(y) + c_0^t(y,x)$. There exists a decreasing function $K(\epsilon)$, which depends only on the Hamiltonian H (and not on the initial condition u_0), such that the function u is $K(\epsilon)$-semi-concave on $M \times [\epsilon, T]$.

Backward viscosity solutions. Let $\breve{u} : M \times [0,T] \longrightarrow \mathbb{R}$ be a continuous function. The following are equivalent:

1. The function $\breve{u}$ is a viscosity solution of $(\breve{HJ})$, *i.e.* it is a viscosity super-solution of $(\breve{HJ})$ and it is a viscosity sub-solution of $(\breve{HJ})$, which means that each smooth function $f : M \times [0,T] \longrightarrow \mathbb{R}$ satisfies the inequality $-\partial_t f - H(x, \partial_x f(x,t), t) \leqslant 0$ at each point of minimum (x,t) of the difference $f - u$.
2. The function $\breve{u}$ is the minimum of all viscosity sub-solutions v of (HJ) (or equivalently all viscosity super-solutions of $(\breve{HJ})$) which satisfy the final condition $v_T = u_T$.
3. We have the relation $\breve{u}(x,s) = \max_{y \in M} \breve{u}(y,t) - c_s^t(x,y)$ for all $0 \leqslant s \leqslant t \leqslant T$.
4. The function $\breve{u}$ is locally semi-convex on $M \times [0, T[$ and it solves (HJ) (or equivalently $(\breve{HJ})$) at almost every point.

As a consequence, for each continuous function $\breve{u}_T$ on M, there exists one and only one viscosity solution $\breve{u}$ of $(\breve{HJ})$ on $M \times [0,T]$ with $\breve{u}_T$ as final condition, it is locally semi-convex on $M \times [0, T[$ and given explicitly by $\breve{u}(x,t) = \max_{y \in M} u_T(y) - c_t^T(x,y)$. There exists a decreasing function $K(\epsilon)$, which depends only on the Hamiltonian H (and not on the final condition $\breve{u}_T$), such that the function $\breve{u}$ is $K(\epsilon)$-semi-convex on $M \times [0, T-\epsilon]$.

Proximal super-differentials of viscosity solutions. Let $u \in C(M \times [0,T], \mathbb{R})$ be a viscosity solution of (HJ). We have the expression

$$u(x,t) = \min_{\gamma} u_0(\gamma(0)) + \int_0^t L(\gamma(s), \dot{\gamma}(s), s) ds$$

where the minimum is taken on the set of curves $\gamma \in C^2([0,t], M)$ which satisfy the final condition $\gamma(t) = x$. Let us denote by $\Gamma(x,t)$ the set of minimizing curves in this expression, which are obviously minimizing extremals of L.

On the other hand, let $\partial_x^+ u(t,x) \subset T_x^* M$ be the set of differentials $\partial_x f(x,t)$ at the point (x,t) of smooth functions f such that (x,t) is a minimum of the difference $f - u$. The set $\partial_x^+ u(t,x) \subset T_x^* M$ is the proximal super-differential of the function u_t at point x. Let us state from [12], Theorem 6.4.12.

Proposition 13. *Let us fix a point $(x,t) \in M \times]0,T]$. The set $\partial_x^+ u(x,t) \subset T_x^* M$ is a convex and compact, hence it is the closed convex envelop of the set $E(x,t)$ of its extremal points. We have*

$$E(x,t) = \{\partial_v L(x, \dot{\gamma}(t), t), \gamma \in \Gamma(x,t)\}.$$

In particular, the function u_t is differentiable at x if and only $\Gamma(x,t)$ contains a single element γ, and then $\partial_x u(x,t) = \partial_v L(x, \dot{\gamma}(t), t)$.

3.2 Viscosity solutions and Kantorovitch optimal pairs

Given a Kantorovitch optimal pair (ϕ_0, ϕ_1), we define the viscosity solution

$$u(t,x) := \min_{y \in M} \phi_0(x) + c_0^t(y,x)$$

and the backward viscosity solution

$$\breve{u}(t,x) := \max_{y\in M} \phi_1(y) - c_t^T(x,y)$$

which satisfy $u_0 = \breve{u}_0 = \phi_0$, and $u_T = \breve{u}_T = \phi_1$.

Proposition 14. *We have*

$$C_0^T(\mu_0,\mu_1) = \max_u \Big(\int_M u_T d\mu_1 - \int_M u_0 d\mu_0 \Big), \tag{5}$$

where the minimum is taken on the set of viscosity solutions $u : M \times [0,T] \longrightarrow \mathbb{R}$ of the Hamilton-Jacobi equation (HJ). The same conclusion holds if the maximum is taken on the set of backward viscosity solutions. The same conclusion also holds if the maximum is taken on the set of viscosity sub-solutions of (HJ).

Proof. If $u(x,t)$ is a viscosity sub-solution of (HJ), then it satisfies

$$u_T(x) - u_0(y) \leqslant c_0^T(y,x)$$

for each x and $y \in M$, and so, by Kantorovitch duality,

$$\Big(\int_M u_T d\mu_1 - \int_M u_0 d\mu_0 \Big) \leqslant C_0^T(\mu_0,\mu_1).$$

The converse inequality is obtained by using the functions u and $\breve{u}$. □

Definition 15. *If (ϕ_0,ϕ_1) is a Kantorovitch optimal pair, then we denote by $\mathcal{F}(\phi_0,\phi_1) \subset C^2([0,T],M)$ the set of curves $\gamma(t)$ such that*

$$\phi_1(\gamma(T)) = \phi_0(\gamma(0)) + \int_0^T L(\gamma(t),\dot{\gamma}(t),t)dt.$$

We denote by $\mathcal{T}(\phi_0,\phi_1) \subset M\times]0,T[$ the set

$$\mathcal{T}(\phi_0,\phi_1) = \{(\gamma(t),t), t \in]0,T[, \gamma \in \mathcal{F}(\phi_0,\phi_1)\}$$

and by $\tilde{\mathcal{T}}(\phi_0,\phi_1) \subset TM\times]0,T[$ the set

$$\tilde{\mathcal{T}}(\phi_0,\phi_1)) = \{(\gamma(t),\dot{\gamma}(t),t), t \in]0,T[, \gamma \in \mathcal{F}(\phi_0,\phi_1)\},$$

which is obviously invariant under the Euler-Lagrange flow.

Proposition 16. *Let (ϕ_0,ϕ_1) be a Kantorovitch optimal pair, and let u and $\breve{u}$ be the associated viscosity and backward viscosity solutions.*

1. *We have $\breve{u} \leqslant u$, and*

$$\mathcal{T}(\phi_0,\phi_1) = \{(x,t) \in M\times]0,T[\text{ such that } u(x,t) = \breve{u}(x,t)\}.$$

2. *At each point $(x,t) \in \mathcal{T}(\phi_0,\phi_1)$, the functions u and $\breve{u}$ are differentiable, and satisfy $du(x,t) = d\breve{u}(x,t)$. In addition, the mapping $(x,t) \longmapsto du(x,t)$ is locally Lipschitz on $\mathcal{T}(\phi_0,\phi_1)$.*

3. If $\gamma(t) \in \mathcal{F}(\phi_0, \phi_1)$, then $\partial_x u(\gamma(t), t) = \partial_v L(\gamma(t), \dot{\gamma}(t), t)$. As a consequence, the set

$$\mathcal{T}^*(\phi_0, \phi_1) := \{(x, p, t) \in T^*M \times]0, T[\text{ such that } (x, t) \in \mathcal{T} \text{ and } p = \partial_x u(x, t) = \partial_x \breve{u}(x, t)\}$$

is invariant under the Hamiltonian flow, and the restriction to $\tilde{\mathcal{T}}(\phi_0, \phi_1)$ of the projection π is a bi-locally-Lipschitz homeomorphism onto its image $\mathcal{T}(\phi_0, \phi_1)$.

Proof. In order to see that $u \geqslant \breve{u}$, just observe that u is the maximum of all viscosity sub-solutions of (HJ) which satisfy $u_0 = \phi_0$ and $u_T = \phi_T$, while $\breve{u}$ is the minimum of the same family of viscosity sub-solutions. Here is another proof. Let us fix a point $(x, t) \in M \times]0, T[$. There exist points y and z in M such that $u(t, x) = \phi_0(y) + c_0^t(y, x)$ and $\breve{u}(t, x) = \phi_1(z) - c_t^T(x, z)$, so that

$$\begin{aligned} u(t,x) - \breve{u}(t,x) &= \phi_0(y) - \phi_1(z) + c_0^t(y,x) + c_t^T(x,z) \\ &\geqslant c_0^T(y,z) - (\phi_1(z) - \phi_0(y)) \geqslant 0. \end{aligned}$$

In case of equality, we must have $c_0^T(y, z) = c_0^t(y, x) + c_t^T(x, z)$. Let $\gamma_1(s) \in C^2([0, t], M)$ satisfy $\gamma_1(0) = y$, $\gamma_1(t) = x$ and $\int_0^t L(\gamma_1(s), \dot{\gamma}_1(s), s) ds = c_0^t(y, x)$, and let $\gamma_2(s) \in C^2([t, T], M)$ satisfy $\gamma_2(t) = x$, $\gamma_2(T) = z$ and $\int_0^t L(\gamma_2(s), \dot{\gamma}_2(s), s) ds = c_t^T(x, z)$. The curve $\gamma : [0, T] \longrightarrow M$ obtained by pasting γ_1 and γ_2 clearly satisfies $\int_0^T L(\gamma(s), \dot{\gamma}(s), s) ds = c_0^T(y, z)$, it is thus a C^2 minimizer, and belongs to $\mathcal{F}(\phi_0, \phi_1)$. As a consequence, we have $(x, t) \in \mathcal{T}(\phi_0, \phi_1)$.

Conversely, we have:

Lemma 17. *If v is a viscosity sub-solution of (HJ) satisfying $v_0 = \phi_0$ and $v_T = \phi_1$, and if $(x, t) \in \mathcal{T}(\phi_0, \phi_1)$, then $v(x, t) = u(x, t)$.*

Proof. It is enough to prove that $v(\gamma(t), t) = u(\gamma(t), t)$ for each curve $\gamma \in \mathcal{F}(\phi_0, \phi_1)$. If $\gamma(s)$ is such a curve, then we have

$$c_0^t(\gamma(0), \gamma(t)) + c_t^T(\gamma(t), \gamma(T)) = c_0^T(\gamma(0), \gamma(T))$$

Since v is a sub-solution, we have

$$v(\gamma(t), t) \leqslant v(\gamma(0), 0) + c_0^t(\gamma(0), \gamma(t)).$$

Adding with the inequality

$$u(\gamma(T), T) \leqslant u(\gamma(t), t) + c_t^T(\gamma(t), \gamma(T))$$

we obtain

$$u(\gamma(T), T) \leqslant v(\gamma(0), 0) + c_0^T(\gamma(0), \gamma(t))$$

which is an equality because $\gamma \in \mathcal{F}(\phi_0, \phi_1)$. Hence all the inequalities involved are equalities, and we have $v(\gamma(t), t) = v(\gamma(0), 0) + c_0^t(\gamma(0), \gamma(t))$. The same could have been done with u instead of v, so that u has the same value, and we have proved that

$$u(\gamma(t), t) = v(\gamma(t), t) = v(\gamma(0), 0) + c_0^t(\gamma(0), \gamma(t)).$$

□

The end of the proof of the proposition is straightforward. Point 2 follows from Proposition 12 applied to the locally semi-concave functions u and $-\breve{u}$. Point 3 follows from Proposition 13. □

3.3 Optimal C^1 sub-solution

The following result, on which a large part of the present paper is based, is inspired from [19], but seems new in the present context.

Proposition 18. *We have*

$$C_0^T(\mu_0,\mu_1) = \max_v \Big(\int_M v_T d\mu_1 - \int_M v_0 d\mu_0 \Big),$$

where the maximum is taken on the set of continuous functions $v : M \times [0,T] \longrightarrow \mathbb{R}$ which are C^1 on $M\times]0,T[$ and satisfy the inequality

$$\partial_t v(x,t) + H(t,x,\partial_x v(x,t)) \leqslant 0 \tag{6}$$

at each point $(x,t) \in M\times]0,T[$.

Proof. First, let $v(t,x)$ be a continuous function of $M \times [0,T]$ which is differentiable on $M\times]0,T[$, where it satisfies (6). We then have, for each C^1 curve $\gamma(t) : [0,T] \longrightarrow M$, the inequality

$$\int_0^T L(\gamma(t),\dot\gamma(t),t)dt \geqslant \int_0^T \partial_x v_{(\gamma(t),t)}(\dot\gamma(t)) - H(\gamma(t),u_{(\gamma(t),t)},t)dt$$

$$= \int_0^T \partial_x v_{(\gamma(t),t)}(\dot\gamma(t)) + \partial_t v_{(\gamma(t),t)} = u(\gamma(T),T) - v(\gamma(0),0).$$

As a consequence, we get $v(y,T) - v(x,0) \leqslant c^T(x,y)$ for each x and y, so that

$$\int v_T d\mu_1 - \int v_0 d\mu_0 \leqslant C_0^T(\mu_0,\mu_1).$$

The converse follows directly from the next theorem, which is an analog in our context of the main result of [19]. □

Theorem 3. *For each Kantorovitch optimal pair (ϕ_0,ϕ_1), there exists a continuous function $v : M \times [0,T] \longrightarrow \mathbb{R}$ which is C^1 on $M\times]0,T[$, which coincides with u on $M \times \{0,T\} \cup \mathcal{T}(\phi_0,\phi_1)$, and which satisfies the inequality (6) strictly at each point of $M\times]0,T[- \mathcal{T}(\phi_0,\phi_1)$.*

Proof. The proof of [19] can't be translated to our context in a straightforward way. Our proof is different, and, we believe, simpler. It is based on:

Proposition 19. *There exists a function $V \in C^2(M \times [0,T],\mathbb{R})$ which is null on $\mathcal{T}(\phi_0,\phi_1)$ and which is positive on $M\times]0,T[- \mathcal{T}(\phi_0,\phi_1)$, and such that*

$$\phi_1(y) = \min_{\gamma(T)=y} \phi_0(\gamma(0)) + \int_0^T L(\gamma(t),\dot\gamma(t),t) - V(\gamma(t),t)dt. \tag{7}$$

Proof. Let us define the norm

$$\|u\|_2 = \sum_{\theta\in\Theta} \|u\circ\theta\|_{C^2(B_1\times[0,T],\mathbb{R})}$$

of functions $u \in C^2(M \times [0,T],\mathbb{R})$, where Θ is the atlas of M defined in the appendix. If $u_n \in C^2(M\times[0,T],\mathbb{R})$ is a sequence of functions such that $\|u_n - u\|_2 \longrightarrow 0$ then $u \in C^2(M\times[0,T],\mathbb{R})$. Let us denote by U the open set $M\times]0,T[- \mathcal{T}(\phi_0,\phi_1)$. We need a Lemma.

Lemma 20. *Let $U_1 \subset U$ be an open set whose closure $\bar{U}_1$ is compact and contained in U, and let $\epsilon > 0$ be given. There exists a function $V_1 \in C^2(M \times [0,T[, \mathbb{R})$, which is positive on U_1 and null outside of $\bar{U}_1$ which is such that the equality (7) holds with $V = V_1$, and such that $\|V_1\|_2 \leqslant \epsilon$.*

PROOF. Let us fix the open set U_1 and the pair (ϕ_0, ϕ_1). We claim that the conclusion of the Lemma holds for all fuctions V_1 which are supported in $\bar{U}_1$ and are sufficiently small in the C^0 topology. In order to prove the claim, let us denote by $\mathcal{F}(V_1)$ the set of curves $\gamma \in C^2([0,T], M)$ which minimize

$$\phi_0(\gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t), t) - V_1(\gamma(t), t)dt$$

for fixed endpoint $\gamma(T)$. It is not hard to see that the set $\mathcal{F}(V_1)$ depends semi-continuously of the function V_1 in the following sense: For each fixed open set $\mathcal{U}$ in $C^0([0,T], M)$ which contains $\mathcal{F}(0)$, we have $\mathcal{F}(V_1) \subset \mathcal{U}$ when $\|V_1\|_0$ is sufficiently small. Let us choose the open set $\mathcal{U}$ in such a way that, for each $\gamma \in \mathcal{U}$ and for each $t \in [0,T]$, the point $(\gamma(t), t)$ does not belong to $\bar{U}_1$. Now if V_1 is so small that $\mathcal{F}(V_1) \subset \mathcal{U}$, we obtain

$$\min_{\gamma \in C^2([0,T],M), \gamma(T)=x} \phi_0(\gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t), t) - V_1(\gamma(t), t)dt$$

$$= \min_{\gamma \in \mathcal{U} \cap C^2, \gamma(T)=x} \phi_0(\gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t), t) - V_1(\gamma(t), t)dt$$

$$= \min_{\gamma \in \mathcal{U} \cap C^2, \gamma(T)=x} \phi_0(\gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t), t)dt = \phi_1(x)$$

□

Let $U_n \subset U, n \in \mathbb{N}$ be a countable sequence of open sets covering U and whose closures $\bar{U}_n$ are contained in U. There exists a sequence V_n of functions of $C^2(M \times [0,T], \mathbb{R})$ such that, for each $n \in \mathbb{N}$:

- The function V_n is positive in U_n and null outside of $\bar{U}_n$.
- We have $\|V_n\|_2 \leqslant 2^{-n}$.
- The equality (7) holds for the function $V^n = \sum_{i=1}^n V_n$.

Such a sequence can be build inductively by applying the lemma to the Lagrangian $L - V^{n-1}$ with $\epsilon = 2^{-n}$. Since $\|V_n\| \leqslant 2^{-n}$, the sequence V^n is converging in C^2 norm to a limit $V \in C^2(M \times [0,T], \mathbb{R})$. This function V satisfies the desired properties. The proposition is proved.

□

In order to finish the proof of the theorem, we shall consider the new Lagrangian $\tilde{L} = L - V$, and the associated Hamiltonian $\tilde{H} = H + V$, as well as the associated cost functions $\tilde{c}_s^t$. Let

$$\tilde{u}(x,t) := \min_{y \in M} \phi_0(y) + \tilde{c}_0^t(y,x),$$

be the viscosity solution of the Hamilton-Jacobi equation

$$\partial_t \tilde{u} + H(x, \partial_x \tilde{u}, t) = -V(x,t) \qquad (\tilde{H}J)$$

emanating from ϕ_0. The equality (7) says that $\tilde{u}_T = \phi_1 = u_T$.

The function $\tilde{u}$ is semi-concave as a viscosity solution of $(\tilde{H}J)$. It is obviously a viscosity sub-solution of the equation (HJ), which is strict outside of $M \times \{0,T\} \cup \mathcal{T}(\phi_0, \phi_1)$ (where V is positive). We have $\breve{u} \leqslant \tilde{u} \leqslant u$, this relation being satisfied by each viscosity sub-solution of (HJ)

which satisfies $u_0 = \phi_0$ and $u_T = \phi_1$. As a consequence, we have $\breve{u} = \tilde{u} = u$ on $\mathcal{T}(\phi_0, \phi_1)$, and the function $\tilde{u}$ is differentiable at each point of $\mathcal{T}(\phi_0, \phi_1)$. Furthermore, we have $du = d\tilde{u} = d\breve{u}$ on this set.

We then obtain the desired function v of the theorem from the function $\tilde{u}$ by regularisation, applying Theorem 9.2 of [19]. □

4 Optimal objects of the direct problems

We prove Theorem A as well as the results of section 2. The following lemma generalizes a result of Benamou and Brenier, see [4].

Lemma 21. *We have the equality*

$$C_0^T(\mu_0, \mu_1) = \min_{m_0 \in \mathcal{I}(\mu_0, \mu_T)} A(m_0) = \min_{m \in \mathcal{M}(\mu_0, \mu_T)} A(m) = \min_{\chi \in \mathcal{C}(\mu_0, \mu_T)} A(\chi).$$

PROOF. In view of Lemma 10, it is enough to prove that, for each transport current $\chi \in \mathcal{C}(\mu_0, \mu_T)$, we have $A(\chi) \geqslant C_0^T(\mu_0, \mu_1)$. Let v be a sub-solution of (HJ) which is C^1 on $M \times]0, T[$, and such that (v_0, v_T) is a Kantorovitch optimal pair. For each current $\chi \in \mathcal{C}(\mu_0, \mu_T)$, we have $A(\chi) \geqslant \chi(dv) = C_0^T(\mu_0, \mu_1)$, which ends the proof. □

Let us choose an optimal pair (ϕ_0, ϕ_1) of the Kantorovitch dual problem, and fix it for then end of this section. Let us choose a sub-solution $v : M \times [0, T] \longrightarrow \mathbb{R}$ of the Hamilton-Jacobi equation which satisfies $v_0 = \phi_0$ and $v_T = \phi_1$ and which is C^1 on $M \times]0, T[$, and fix it. Let us choose a vector-field $X(x, t) : M \times]0, T[\longrightarrow TM$ which is locally Lipschitz and satisfies $X(x, t) = \partial_p H(x, \partial_x v(x, t), t)$ on $\mathcal{T}(\phi_0, \phi_1)$. The existence of such a vector-field follows from the fact that the vector-field $\partial_p H(x, \partial_x v(x, t), t)$ is locally Lipschitz on $\mathcal{T}(\phi_0, \phi_1)$ and for standard extension results, see appendix. We denote by Ψ_s^t the flow of X. All these objects are fixed once and for all in this section.

4.1 Characterization of optimal currents.

Each optimal transport current χ can be written

$$\chi = (X, 1) \wedge \mu_\chi,$$

with a measure μ_χ concentrated on $\mathcal{T}(\phi_0, \phi_1)$. The current χ is then Lipschitz regular, so that there exists a transport interpolation $\mu_t, t \in [0, T]$ such that $\mu_\chi = \int_0^T \mu_t dt$ and such that $\mu_t = (\Psi_s^t)_\sharp \mu_s$ for each s and t in $]0, T[$.

PROOF. Let χ be an optimal transport current, that is a transport current $\chi \in \mathcal{C}(\mu_0, \mu_T)$ such that $A(\chi) = C_0^T(\mu_0, \mu_T)$. Let us recall the definition of the action $A(\chi)$ that will be used here:

$$A(\chi) = \sup_{\omega \in \Omega^0} \Big(\chi(\omega^x, 0) - \int_{M \times [0,T]} H(x, \omega^x(x, t), t) d\mu_\chi \Big).$$

Since $H(x, \partial_x v, t) + \partial_t v \leqslant 0$, we have

$$A(\chi) = \chi(dv) \leqslant \chi(dv) - \int H(x, \partial_x v(x, t), t) + \partial_t v d\mu_\chi = \chi(\partial_x v, 0) - \int H(x, \partial_x v(x, t), t) d\mu_\chi.$$

The other inequality holds by the definition of A, so that

$$A(\chi) = \chi(dv) - \int H(x, \partial_x v(x, t), t) + \partial_t v d\mu_\chi = \chi(\partial_x v, 0) - \int H(x, \partial_x v(x, t), t) d\mu_\chi,$$

and we conclude that the function $H(x, \partial_x v(x,t), t) + \partial_t v$ vanishes on the support of μ_χ, or in other words that the measure μ_χ is concentrated on the set $\mathcal{T}(\phi_0, \phi_1)$. In addition, for all form $\omega = \omega^x + \omega^t dt$, we have

$$\chi(\partial_x v + \omega^x, 0) - \int H(x, \partial_x v + \omega^x, t) d\mu_\chi \leqslant \chi(\partial_x v, 0) - \int H(x, \partial_x v, t) d\mu_\chi = A(\chi).$$

Hence the equality

$$\chi(\omega^x, 0) = \int \partial_p H(x, \partial_x v, t)(\omega^x) d\mu_\chi$$

holds for each form ω. This equality can be rewritten

$$\chi(\omega) = \int \partial_p H(x, \partial_x v, t)(\omega^x) + \omega^t d\mu_\chi$$

which is precisely saying that

$$\chi = (\partial_p H(x, \partial_x v(x,t), t), 1) \wedge \mu_\chi = (X, 1) \wedge \mu_\chi.$$

The last equality following from the fact that the vector-fields X and $\partial_p H(x, \partial_x v(x,t), t)$ are equal on the support of μ_χ. By the structure of Lipschitz regular transport currents, we obtain the existence of a continuous family $\mu_t, t \in [0,T]$ of probability measures such that $\mu_\chi = \int_0^T \mu_t dt$ and such that $\mu_t = (\Psi_s^t)_\sharp \mu_s$ for each s and t in $]0,T[$. Since the restriction to a subinterval $[s,t] \subset [0,T]$ of an optimal transport current χ is clearly an optimal transport current for the transportation problem between μ_s and μ_t with cost c_s^t, we obtain that the path μ_t is a transport interpolation.

□

4.2 Characterization of transport interpolations.

Each transport interpolation μ_t satisfies

$$\mu_t = (\Psi_s^t)_\sharp \mu_s$$

for each $(s,t) \in]0,T[^2$. The mapping

$$\mu_t \longmapsto (X,1) \wedge \int_0^T \mu_t dt$$

is a bijection between the set of transport interpolations and the set of optimal transport currents.

PROOF. We fix a transport interpolation μ_t and two time $s < s'$ in $]0,T[$. Let χ_1 be a transport current on $M \times [0,s]$ between the measures μ_0 and μ_s which is optimal for the cost c_0^s, let χ_2 be a transport current on $M \times [s,s']$ between the measures μ_s and $\mu_{s'}$ which is optimal for the cost $c_s^{s'}$ and let χ_3 be a transport current on $M \times [s',T]$ between the measures $\mu_{s'}$ and μ_T which is optimal for the cost $c_{s'}^T$. Then the current χ on $M \times [0,T]$ which coincides with χ_1 on $M \times [0,s]$, with χ_2 on $M \times [s,s']$ and with χ_3 on $[s',T]$ belongs to $\mathcal{C}(\mu_0, \mu_T)$. In addition, since μ_t is a transport interpolation, we have

$$A(\chi) = C_0^s(\mu_0, \mu_s) + C_s^{s'}(\mu_s, \mu_{s'}) + C_{s'}^T(\mu_{s'}, \mu_T) = C_0^T(\mu_0, \mu_T).$$

Hence χ is an optimal transport current for the cost c_0^T. In view of the characterisation of optimal currents, this implies that $\chi = (X,1) \wedge \mu_\chi$, and that

$$\mu_\chi = \int_0^T (\Psi_s^t)_\sharp \mu_s dt = \int_0^T (\Psi_{s'}^t)_\sharp \mu_{s'} dt.$$

By uniqueness of the continuous desintegration of μ_χ, we obtain that, for each $t \in]0,T[$, $(\Psi^t_s)_\sharp \mu_s = (\Psi^t_{s'})_\sharp \mu_{s'}$, and since this holds for all s and s', that $(\Psi^t_s)_\sharp \mu_s = \mu_t$ for all $(s,t) \in]0,T[^2$. It follows that $\chi = (X,1) \wedge \int_0^T \mu_t dt$. We have proved that the mapping

$$\mu_t \longmapsto (X,1) \wedge \int_0^T \mu_t dt$$

associates an optimal transport current to each transport interpolation. This mapping is obviously injective, and it is surjective in view of the characterization of optimal currents. □

4.3 Characterization of optimal measures.

The mapping

$$\chi \longmapsto (X \times \tau)_\sharp \mu_\chi$$

is a bijection between the set of optimal transport currents and the set of optimal transport measures. Each optimal transport measure is thus invariant. The mapping

$$m_0 \longmapsto \mu_t = (\pi \circ \psi^t_0)_\sharp m_0$$

is a bijection between the set of optimal initial measures m_0 and the set of interpolations. An invariant measure m is optimal if and only if it is supported on the set $\tilde{\mathcal{T}}(\phi_0,\phi_1)$.

PROOF. If m is an optimal transport measure, then the associated current χ_m is an optimal transport current, and $A(m) = A(\chi_m)$. Let μ_m be the times component of χ_m, which is also the measure $\pi_\sharp m$. In view of the characterization of optimal currents, we have $\chi_m = (X,1) \wedge \mu_m$. We claim that the equality $A(\chi_m) = A(m)$ implies that m is supported on the graph of X. Indeed, we have the pointwise inequality

$$\partial_x v(x,t) \cdot V - H(x, \partial_x v(x,t), t) \leqslant L(x,V,t) \tag{8}$$

for each $(x,V,t) \in TM \times]0,T[$. Integrating with respect to m, we get the equality

$$A(\chi_m) = \int_{TM\times[0,T]} \partial_x v(V) - H(x, \partial_x(x,t), t) dm(x,V,t) = \int_{M\times[0,T]} L(x,V,t) dm(x,V,t) = A(m),$$

which means that m is concentrated on the set where the inequality (8) is an equality, that is on the graph of the vectorfield $\partial_p H(x, \partial_x v(x,t), t)$. Since μ_m is supported on $\mathcal{T}$, the measure m is supported on $\tilde{\mathcal{T}}$ and satisfies $m = (X \times \tau)_\sharp \mu_m$. Let μ_t be the transport interpolation such that $\mu_m = \int_0^T \mu_t dt$. Setting $m_t = (X_t)_\sharp \mu_t$, we have $m = \int_0^T m_t dt$. Observing that the relation

$$X_t \circ \Psi^t_s = \psi^t_s \circ X_s$$

holds on $\mathcal{T}_s$, we conclude, since μ_s is supported on $\mathcal{T}_s$, that

$$(\psi^t_s)_\sharp m_s = m_t,$$

which means that the measure m is invariant.

Conversely, let $m = \int_0^T m_t dt$ be an invariant measure supported on $\tilde{\mathcal{T}}(\phi_0,\phi_1)$. We have

$$A(m) = \int_0^T \int_{TM} L(x,v,t) dm_t(x,v) dt = \int_0^T \int_{TM} L((\psi^t_0(x,v),t) dm_0(x,v) dt,$$

and by Fubini,

$$A(m) = \int_{TM} \int_0^T L((\psi_0^t(x,v),t)dtdm_0(x,v) = \int_{TM} \phi_1(\pi \circ \psi_0^T(x,v)) - \phi_0(x)dm_0(x,v),$$

and since m_0 is an initial transport measure, we get

$$A(m) = \int_{TM} \phi_1 d\mu_T - \int_{TM} \phi_0 d\mu_0 = C_0^T(\mu_0, \mu_T).$$

□

5 Absolute continuity

In this section, we make the additional assumption that the initial measure μ_0 is absolutely continuous, and prove Theorem B. The following lemma answers a question asked to us by Cedric Villani.

Lemma 22. *If μ_0 or μ_T is absolutely continuous with respect to the Lebesgue class, then each interpolating measure $\mu_t, t \in]0,T[$, is absolutely continuous.*

PROOF. If $\mu_t, t \in [0,T]$ is a transport interpolation, we have proved that

$$\mu_t = (\pi \circ \psi_s^t \circ X_s)_\sharp \mu_s$$

for each $s \in]0,T[$, and $t \in [0,T]$. Since the function $\pi \circ \psi_t^s \circ X_t$ is Lipschitz, it maps Lebesgue zero measure sets into Lebesgue zero measure sets, and so it transport singular measures into singular measures. It follows that if, for some $s \in]0,T[$, the measure μ_s is not absolutely continuous, then none of the measures $\mu_t, t \in [0,T]$ are absolutely continuous. □

In order to continue the investigation of the specific properties satisfied when μ_0 is absolutely continuous, we first need some more general results. Let (ϕ_0, ϕ_1) be an optimal Kantorovitch pair for the measures μ_0 and μ_T and for the cost c_0^T. Recall that we have defined $\mathcal{F}(\phi_0,\phi_1) \subset C^2([0,T],M)$ as the the set of curves $\gamma(t)$ such that

$$\phi_1(\gamma(T)) = \phi_0(\gamma(0)) + \int_0^T L(\gamma(t), \dot\gamma(t), t)dt.$$

Let $\mathcal{F}_0(\phi_0,\phi_1)$ be the set of initial velocities $(x,v) \in TM$ such that the curve $t \longmapsto \pi \circ \psi_0^t(x,v)$ belongs to $\mathcal{F}(\phi_0,\phi_1)$. Note that there is a natural bijection between $\mathcal{F}_0(\phi_0,\phi_1)$ and $\mathcal{F}(\phi_0,\phi_1)$.

Lemma 23. *The set $\mathcal{F}_0(\phi_0,\phi_1)$ is compact. The maps π and $\pi \circ \psi_0^T : \mathcal{F}_0(\phi_0,\phi_1) \longrightarrow M$ are surjective. If x is a point of differentiability of ϕ_0, then the set $\pi^{-1}(x) \cap \mathcal{F}_0(\phi_0,\phi_1)$ contains one and only one point. There exists a Borel measurable set $\Sigma \subset M$ of full measure, whose points are points of differentiability of ϕ_0, and such that the map*

$$x \longmapsto S(x) = \pi^{-1}(x) \cap \mathcal{F}_0(\phi_0,\phi_1)$$

is Borel measurable on Σ.

PROOF. The compactness of $\mathcal{F}_0(\phi_0,\phi_1)$ follows from the fact, already mentioned, that the set of minimizing extremals $\gamma : [0,T] \longrightarrow M$ is compact for the C^2- topology.

It is equivalent to say that the projection π restricted to $\mathcal{F}_0(\phi_0, \phi_1)$ is surjective, and to say that, for each point $x \in M$, there exists a curve emanating from x in $\mathcal{F}(\phi_0, \phi_1)$. In order to build such curves, recall that

$$\phi_0(x) = \max_{\gamma} \phi_1(\gamma(T)) - \int_0^T L(\gamma(t, \dot{\gamma}(t), t)dt$$

where the maximum is taken on the set of curves which satisfy $\gamma(0) = x$. Any maximizing curve is then a curve of $\mathcal{F}(\phi_0, \phi_1)$ which satisfies $\gamma(0) = x$. In order to prove that the map $\pi \circ \psi_0^T$ restricted to $\mathcal{F}_0(\phi_0, \phi_1)$ is surjective, it is sufficient to build, for each point $x \in M$, a curve in $\mathcal{F}(\phi_0, \phi_1)$ which ends at x. Such a curve is obtained as a minimizer in the expression

$$\phi_1(x) = \min_{\gamma} \phi_0(\gamma(0)) + \int_0^T L(\gamma(t, \dot{\gamma}(t), t)dt.$$

Now let us consider a point x of differentiability of ϕ_0. Applying the general result on the differentiability of viscosity solutions to the Backward viscosity solution $\breve{u}$, we get thet there exists a unique maximizer to the problem

$$\phi_0(x) = \max_{\gamma} \phi_1(\gamma(T)) - \int_0^T L(\gamma(t, \dot{\gamma}(t), t)dt$$

and that this maximizer is the extremal with initial condition $(x, \partial_p H(x, d\phi_0(x), 0))$. As a consequence, there exists one and only one point $S(x)$ in $\mathcal{F}_0(\phi_0, \phi_1)$ above x, and in addition we have the explicit expression

$$S(x) = \partial_p H(x, d\phi_0(x), 0).$$

Since the set of points of differentiability of ϕ_0 has total Lebesgue measure –because ϕ_0 is Lipschitz– there exists a sequence K_n of compact sets such that ϕ_0 is differentiable at each point of K_n and such that the Lebesgue measure of $M - K_n$ is converging to zero. For each n, the set $\pi^{-1}(K_n) \cap \mathcal{F}_0(\phi_0, \phi_1)$ is compact, and the restriction to this set of the canonical projection π is injective and continuous. It follows that the inverse function S is continuous on K_n. As a consequence, the map S is Borel measurable on $\Sigma := \cup_n K_n$. □

Lemma 24. *The initial transport measure m_0 is optimal if and only if it is an initial transport measure supported on $\mathcal{F}_0(\phi_0, \phi_1)$.*

Proof. This statement is a reformulation of the result in 4.3 stating that the optimal transport measures are the invariant measures supported on $\tilde{\mathcal{T}}(\phi_0, \phi_1)$. □

Proposition 25. *If μ_0 is absolutely continuous, then there exists a unique optimal initial measure m_0. There exists a Borel section $S : M \longrightarrow TM$ of the canonical projection such that $m_0 = S_\sharp \mu_0$, this section is unique μ_0-almost everywhere. For each $t \in [0, T]$, the map $\pi \circ \psi_0^t \circ S : M \longrightarrow M$ is then an optimal transport map between μ_0 and μ_t.*

Proof. Let $S : \Sigma \longrightarrow TM$ be the Borel map constructed in Lemma 23. For convenience, we shall also denote by S the same map extended by zero outside of Σ, which is a Borel section $S : M \longrightarrow TM$. Since the set Σ is of full Lebesgue measure, and since the measure μ_0 is absolutely continuous, we have $\mu_0(\Sigma) = 1$. Let us consider the measure $m_0 = S_\sharp(\mu_{0|\Sigma})$. This is a probability measure on TM, which is concentrated on $\mathcal{F}_0(\phi_0, \phi_1)$, and which satisfies $\pi_\sharp m_0 = \mu_0$. We claim that it is the only measure with these properties. Indeed, if $\tilde{m}_0$ is a measure with these properties, then $\pi_\sharp \tilde{m}_0 = \mu_0$, hence the measure $\tilde{m}_0$ is concentrated on $\pi^{-1}(\Sigma) \cap \mathcal{F}_0(\phi_0, \phi_1)$.

But then, since π induces a Borel isomorphism from $\pi^{-1}(\Sigma) \cap \mathcal{F}_0(\phi_0, \phi_1)$ onto its image Σ, of inverse S, we must have $\tilde{m}_0 = S_\sharp \mu_0$. As a consequence, the measure $m_0 = S_\sharp \mu_0$ is the only candidate to be an optimal initial transport measure. Since we have already proved the existence of an optimal initial transport measure, it implies that m_0 is the only optimal initial transport measure. Of course, we could prove directly that m_0 is an initial transport measure, but as we have seen, it is not necessary. □

5.1 Remark

That there exists an optimal transport map if μ_0 is continuous could be proved directly as a consequence of the following properties of the cost function.

Lemma 26. *The cost function $c_0^T(x, y)$ is semi-concave on $M \times M$. In addition, we have the following injectivity property for each $x \in M$: If the differentials $\partial_x c_0^T(x, y)$ and $\partial_x c_0^T(x, y')$ exist and are equal, then $y = y'$.*

In view of these properties of the cost function, it is not hard to prove the following lemma using a Kantorovitch optimal pair in the spirit of works of Brenier[10] and Carlier [13].

Lemma 27. *There exists a compact subset $K \in M \times M$, such that the fiber $K_x = K \cap \pi_0^{-1}(x)$ contains one and only one point for Lebesgue almost every x, and which contains the support of all optimal plans.*

The proof of the existence of an optimal map for an absolutely continuous measure μ_0 can then be terminated using the following result, see [1], Proposition 2.1.

Proposition 28. *A transport plan η is induced from a transport map if and only if it is concentrated on a η-measurable graph.*

5.2 Remark

Assuming only that μ_0 vanishes on $(d-1)$-rectifiable sets, we can conclude that the same property holds for all interpolating measures $\mu_t, t < T$, and that Proposition 25 hold. This is proved almost identically. The first refinement needed is that the set of singular points of the semi-convex function ϕ_0 is a $(d-1)$-rectifiable, see [12]. The second refinement needed is that there exists a Borel section $S : M \longrightarrow TM$ of the canonical projection such that $S(x) \in \mathcal{F}_0(\phi_0, \phi_1)$ for each $x \in M$. This follows from general statements of set-valued analysis, see for example [14] or the appendix in [12].

6 Aubry-Mather theory

We explain the relations between the results obtained so far and Mather theory, and prove Theorem C. Up to now, we have worked with fixed measures μ_0 and μ_T. Let us study the optimal value $C_0^T(\mu_0, \mu_T)$ as a function of the measures μ_0 and μ_T.

Lemma 29. *The function*

$$(\mu_0, \mu_T) \longmapsto C_0^T(\mu_0, \mu_T)$$

is convex and lower semi-continuous on the set of pairs of probability measures on M.

Proof. It follows directly from the expression

$$C_0^T(\mu_0, \mu_T) = \max_{(\phi_0, \phi_1)} \int_M \phi_1 d\mu_T - \int_M \phi_0 d\mu_0$$

as a maximum of continuous linear functions. □

From now on, we consider that the Lagrangian L is defined for all times, $L \in C^2(TM \times \mathbb{R}, \mathbb{R})$, and satisfies

$$L(x, v, t+1) = L(x, v, t)$$

in addition to the standing hypotheses. Let us restate Theorem C with more details.

Theorem C'. *There exists a Lipschitz vector-field X_0 on M such that all the Mather measures are supported on the graph of X_0. We have*

$$\alpha = \min_{\mu} C_0^1(\mu, \mu),$$

where the minimum is taken on the set of probability measures on M. The mapping $m_0 \longmapsto (\pi)_\sharp m_0$ is a bijection between the set of Mather measures m_0 and the set of probability measures μ on M satisfying $C_0^1(\mu, \mu) = \alpha$. More precisely, if μ is such a probability measure, then there exists one and only one initial transport measure m_0 for the transport problem between $\mu_0 = \mu$ and $\mu_1 = \mu$ with cost c_0^1, this measure is $m_0 = (X_0)_\sharp \mu$, and it is a Mather measure.

The proof, and related digressions, occupy the end of the section.

Lemma 30. *The following minima*

$$\alpha_T := \min_{\mu \in \mathcal{B}_1(M)} \frac{1}{T} C_0^T(\mu, \mu), T \in \mathbb{N}$$

exist and are all equal.

Proof. The existence of the minima follows from the compactness of the set of probability measures and from the semi-continuity of the function C_0^T.

Let μ^1 be a minimizing measure for α_1. We have $C_0^T(\mu^1, \mu^1) \leqslant \sum_{i=1}^T C_{i-1}^i(\mu^1, \mu^1) = T\alpha_1$, which implies the inequality

$$\alpha_T \leqslant \alpha_1.$$

Let us now prove that $\alpha_T \geqslant \alpha_1$. In order to do so, we consider an optimal measure μ^T for α_T, and consider a transport interpolation $\mu_t, t \in [0, T]$ between the measures $\mu_0 = \mu^T$ and $\mu_T = \mu^T$. Let us then consider, for $t \in [0, 1]$, the measure

$$\tilde{\mu}_t^T := \frac{1}{T} \sum_{i=0}^{T-1} \mu_{t+i}^T,$$

and note that $T\tilde{\mu}_0^T = \mu_0^T + \sum_{i=1}^{T-1} \mu_i^T = \mu_T^T + \sum_{i=1}^{T-1} \mu_i^T = T\tilde{\mu}_1^T$. In view of the convexity of the function C_0^1, we have

$$C_0^1(\tilde{\mu}_0^T, \tilde{\mu}_1^T) \leqslant \frac{1}{T} \sum_{i=0}^{T-1} C_i^{i+1}(\mu_i, \mu_{i+1}) = \frac{1}{T} C_0^T(\mu^T, \mu^T) = \alpha_T.$$

Since $\tilde{\mu}_0^T = \tilde{\mu}_1^T$, this implies that $\alpha_1 \leqslant \alpha_T$, as desired. □

Lemma 31. *We have $\alpha_1 \leqslant \alpha$.*

Proof. If m_0 is a Mather measure, then it is an initial measure for the transport problem between $\mu_0 = (\pi)_\sharp m_0$ and $\mu_1 = (\pi)_\sharp m_0$ for the cost c_0^1. As a consequence, we have $\alpha = A_0^1(m_0) \geqslant C_0^1(\mu_0, \mu_0) \geqslant \alpha_1$. □

Lemma 32. *Let μ^1 be a probability measure on M such that $C_0^1(\mu^1, \mu^1) = \alpha_1$. Then there exists a unique initial transport measure m_0 for the transportation problem between $\mu_0 = \mu^1$ and $\mu_1 = \mu^1$ for the cost c_0^1. This measure satisfies $(\psi_0^1)_\sharp m_0 = m_0$. We have $\alpha_1 = A_0^1(m_0) \geqslant \alpha$, so that $\alpha = \alpha_1$ and m_0 is a Mather measure. There exists a constant K, which depends only on L, such that the measure m_0 is supported on the Graph of a K-Lipschitz vector-field.*

Proof. Let μ^1 be a probability measure on M such that $C_0^1(\mu^1, \mu^1) = \alpha_1$, and let m_0 be an optimal initial transport measure for the transportation problem between $\mu_0 = \mu^1$ and $\mu_1 = \mu^1$ for the cost c_0^1. Let $\mu_t = (\pi \circ \psi_0^t)_\sharp m_0, t \in [0, 1]$ be the associated transport interpolation. Let us define the 1-periodic continuous path $\tilde{\mu}_t : \mathbb{R} \longrightarrow \mathcal{B}_1(M)$ by $\tilde{\mu}_t = \mu_{t \bmod 1}$. We claim that the relation

$$C_{s_1}^{s_3}(\tilde{\mu}_{s_1}, \tilde{\mu}_{s_3}) = C_{s_1}^{s_2}(\tilde{\mu}_{s_1}, \tilde{\mu}_{s_2}) + C_{s_2}^{s_3}(\tilde{\mu}_{s_2}, \tilde{\mu}_{s_3})$$

holds for all $s_1 < s_2 < s_3$ in $\mathbb{R}$. Observe that the claim obviously follows from the fact that $\mu_t, t \in [0, 1]$ is a transport interpolation if the three times s_1, s_2 and s_3 lie in a same interval of the form $[n, n + 1]$ with $n \in \mathbb{Z}$. The claim also obviously holds if $s_i \in \mathbb{Z}$, since then

$$C_{s_1}^{s_3}(\tilde{\mu}_{s_1}, \tilde{\mu}_{s_3}) = (s_3 - s_1)\alpha_1 = (s_2 - s_1)\alpha_1 + (s_3 - s_2)\alpha_1 = C_{s_1}^{s_2}(\tilde{\mu}_{s_1}, \tilde{\mu}_{s_2}) + C_{s_2}^{s_3}(\tilde{\mu}_{s_2}, \tilde{\mu}_{s_3}).$$

Let us assume that there exists an integer between s_1 and s_2 and an integer between s_2 and s_3, so that

$$[s_1] \leqslant s_1 \leqslant [s_1] + 1 \leqslant [s_2] \leqslant s_2 \leqslant [s_2] + 1 \leqslant [s_3] \leqslant s_3 \leqslant [s_3] + 1.$$

In order to shorten the expressions, we shall denote by C_s^t the quantity $C_s^t(\tilde{\mu}_s, \tilde{\mu}_t)$. We have the equalities

$$C_{[s_1]}^{[s_3]+1} = C_{[s_1]}^{[s_1]+1} + C_{[s_1]+1}^{[s_2]} + C_{[s_2]}^{[s_2]+1} + C_{[s_2]+1}^{[s_3]} + C_{[s_3]}^{[s_3]+1}.$$

and then

$$C_{[s_1]}^{[s_3]+1} = C_{[s_1]}^{s_1} + C_{s_1}^{[s_1]+1} + C_{[s_1]+1}^{[s_2]} + C_{[s_2]}^{s_2} + C_{s_2}^{[s_2]+1} + C_{[s_2]+1}^{[s_3]} + C_{[s_3]}^{s_3} + C_{s_3}^{[s_3]+1}.$$

This implies, on the one hand that

$$C_{[s_1]}^{[s_3]+1} = C_{[s_1]}^{s_1} + C_{s_1}^{s_3} + C_{s_3}^{[s_3]+1}$$

and on the other hand that

$$C_{[s_1]}^{[s_3]+1} = C_{[s_1]}^{s_1} + C_{s_1}^{s_2} + C_{s_2}^{s_3} + C_{s_3}^{[s_3]+1}.$$

The claim follows. The cases where there does not exist any integer between s_1 and s_2 or between s_2 and s_3 are treated similarly.

As a consequence of the claim, the restriction to the interval $[-2, 3]$ of the path $\tilde{\mu}_t$ is a transport interpolation. It follows from our main results that there exists a constant K which depends only on L, and a K-Lipschitz vector-field $X : M \times [-1, 2] \longrightarrow TM$ whose flow Ψ_s^t satisfies $(\Psi_s^t)_\sharp \mu_s = \mu_t$ for all $(s, t) \in [-1, 2]^2$. In view of the bijection between the optimal initial measures and the interpolations, there exists one and only one optimal initial measure m_0 for the transport problem between μ^1 and itself on the time interval $[0, 1]$, this initial transport measure is given by $m_0 = (X_0)_\sharp \mu^1$. We have, for this measure,

$$A_0^1(m_0) = C_0^1(\mu^1, \mu^1) = \alpha_1 \leqslant \alpha.$$

Let us now consider the measure $m_1 = (\psi_0^1)_\sharp m_0$. We have $(\pi \circ \psi_0^1)_\sharp m_1 = (\pi \circ \psi_0^2)_\sharp m_0 = \mu^1$, so that m_1 is an initial transport measure for the transportation problem between $\mu_0 = \mu^1$ and $\mu_1 = \mu^1$ for the cost c_0^1. Since m_0 is also an optimal initial transport measure for the transportation problem on the time interval $[0,2]$ between $\mu_0 = \mu^1$ and $\mu_2 = \mu^1$ for the cost c_0^2, we get $A_0^2(m_0) = 2\alpha_1$. But then we have

$$A_0^1(m_1) = A_1^2(m_0) = A_0^2(m_0) - A_0^1(m_0) = \alpha_1$$

so that the initial measure m_1 is optimal for the transport problem between $\mu_0 = \mu^1$ and $\mu_1 = \mu^1$ for the cost c_0^1. By uniqueness of such an optimal initial measure, we must have $m_1 = m_0$. As a consequence, $A_0^1(m_0) \geqslant \alpha$. Since we have already obtained the other inequality, we have $A_0^1(m_0) = \alpha$, so that m_0 is a Mather measure. □

Proof of the theorem. Let m_0 be a Mather measure, and let $\mu_0 = \pi_\sharp m_0$. Note that we also have $\mu_0 = (\pi \circ \psi_0^1)_\sharp m_0$. As a consequence, m_0 is an initial transport measure for the transport between μ_0 and μ_0 for the cost c_0^1, and we have

$$\alpha = A_0^1(m_0) \geqslant C_0^1(\mu_0, \mu_0) \geqslant \alpha_1.$$

Since $\alpha_1 = \alpha$, all these inequalities are equalities, so that m_0 is an optimal initial transport, and $C_0^1(\mu_0, \mu_0) = \alpha_1$. It follows from Lemma 32 that m_0 is supported on the graph of a K-Lipschitz vector-field.

Up to now, we have proved that each Mather measure is supported on the graph of a K-Lipschitz vector-field. There remains to prove that all Mather measures are supported on a single K-Lipschitz graph. In order to prove this, let us denote by $\tilde{\mathcal{M}} \subset TM$ the union of the supports of Mather measures. If (x, v) and (x', v') are two points of $\tilde{\mathcal{M}}$, then there exists a Mather measure m_0 whose support contains (x, v) and a measure m_0' whose support contains (x', v'). But then the measure $(m_0 + m_0')/2$ is clearly a Mather measure whose support contains both (x, v) and (x', v'). Since the support of the Mather measure $(m_0 + m_0')/2$ is contained on the graph of a K-Lipschitz vector-field. Assuming that x and x' lie in the image $\theta(B_1)$ of a common chart, see appendix, so that $(x, v) = d\theta(X, V)$ and $(x', v') = d\theta(X', V')$, we obtain

$$\|V - V'\| \leqslant K\|x - x'\|.$$

It follows that the restriction to $\tilde{\mathcal{M}}$ of the canonical projection $TM \longrightarrow M$ is a bi-Lipschitz homeomorphism, or equivalently that the set $\tilde{\mathcal{M}}$ is contained in the graph of a Lipschitz vector-field. □

A Notations and standing conventions

- M is a compact manifold of dimension d, and $\pi : TM \longrightarrow M$ is the canonical projection.
- We denote by $\tau : TM \times [0, T] \longrightarrow [0, T]$ or $M \times [0, T] \longrightarrow [0, T]$ the projection on the second factor.
- If N is any manifold (for exemple M,$M \times [0, T]$, TM or $TM \times [0, T]$)) the sets $\mathcal{B}(N) \subset \mathcal{B}_+(N) \subset \mathcal{B}_1(N)$ are respectively the set of finite Borel signed measures, non-negative Borel finite measures, and Borel probability measures. If $C_c(N)$ is the set of continuous compactly supported functions on N, endowed with the topology of uniform convergence, then the space $\mathcal{B}(N)$ is identified with the set of continuous linear forms on $C_c(M)$ by the Rietz theorem. We will always endow the space $\mathcal{B}(N)$ with the weak-$*$ topology, which will sometimes be called weak topology. Note that the set $\mathcal{B}_1(N)$ is compact if N is.

- Given two manifolds N and N', a Borel application $F : N \longrightarrow N'$, and a measure $\mu \in \mathcal{B}(N)$, we define the push-forward $F_\sharp\mu$ of μ by F as the unique measure on N' which satisfies
$$F_\sharp\mu(B) = \mu(F^{-1}(B))$$
for all Borel set $B \in N$, or equivalently
$$\int_{N'} fd(F_\sharp\mu) = \int_N f \circ F d\mu$$
for all continuous function $f : N' \longrightarrow \mathbb{R}$.

- The set $\mathcal{K}(\mu_0, \mu_T)$ of transport plans is defined in section 1.2.

- The set $\mathcal{I}(\mu_0, \mu_T)$ of initial transport measures is defined in section 2.1.

- The set $\mathcal{M}(\mu_0, \mu_T)$ of transport measures is defined in section 2.1.

- The set $\mathcal{C}(\mu_0, \mu_T)$ of transport currents is defined in section 2.2.

- We fix, once and for all, a finite atlas Θ of M, formed by charts $\theta : B_3 \longrightarrow M$, where B_r is the open ball of radius r centered at zero in $\mathbb{R}^d$. We assume in addition that the sets $\theta(B_1), \theta \in \Theta$ cover M.

- We say that a vector-field $X : M \longrightarrow TM$ is K-Lipschitz if, for each chart $\theta \in \Theta$, the mapping $\Pi \circ (d\theta)^{-1} \circ X \circ \theta : B_3 \longrightarrow \mathbb{R}^d$ is K-Lipschitz on B_1, where Π is the projection $B_3 \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$.

- We mention the following results which are used through the paper : There exists a universal constant C such that, if A is a subset of M, and $X_A : A \longrightarrow TM$ is a K-Lipschitz vector-field, then there exists a CK-Lipschitz vector-field X on M which extends X_A. In addition, if A is a subset of $M \times [0, T]$ and $X_A : A \longrightarrow TM$ is a K-Lipschitz vector-field, then there exists a CK-Lipschitz vector-field X on $M \times [0, T]$ which extends X_A. If A is a compact subset of $M \times [0, T]$ and $X_A : A \cap M \times]0, T[\longrightarrow TM$ is a locallyLipschitz vector-field (which is $K(\epsilon)$-Lipschitz on $A \cap M \times [\epsilon, T - \epsilon]$), then there exists a locally Lipschitz (which is $CK(\epsilon)$-Lipschitz on $M \times [\epsilon, T - \epsilon]$) vector-field X on $M \times]0, T[$ which extends X_A,

References

[1] L. Ambrosio, *Lecture Notes on Optimal Transport Problems.*

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

[3] V. Bangert, *Minimal measures and minimizing closed normal one-currents*, GAFA, **9** (1999), 413-427.

[4] J.-D. Benamou and Y. Brenier, *A computational fluid dynamics solution to the Monge-Kantorovitch mass transfer problem,* Num. Math. **84** (2000), 323-351.

[5] P. Bernard, *Connecting orbits of time dependent Lagrangian systems*, Ann. Inst. Fourier, Grenoble **52** (2002), 1533-1568.

[6] P. Bernard, *The dynamics of pseudographs in convex Hamiltonian systems*, preprint.

[7] P. Bernard and B. Buffoni, *The Monge transport problem in Lagrangian dynamics*, preprint.

[8] P. Billingsley, Convergence of Probability Measures–2nd ed., Wiley-Interscience publication, 1999.

[9] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math. **305** (1987), 805-808.

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

[11] Y. Brenier, *Extended Monge-Kantorovich theory,* Optimal transportation and applications (Martina Franca, 2001), 91–121, Lecture Notes in Math. **1813** , Springer, Berlin, 2003.

[12] P. Cannarsa and C. Sinestrari, Semiconcave functions, Hamilton-Jacobi equations, and optimal control, *Progress in nonlinear differential equations and their applications*, Birkhäuser (2004).

[13] G. Carlier, *Duality and existence for a class of mass transportation problems and economic applications*, preprint found on the WEB.

[14] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics **580**, Springer-Verlag, 1977.

[15] R. M. Dudley, Real Analysis and Probability, Cambridge University Press, 2002.

[16] L. C. Evans and W. Gangbo, *Differential equation methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc. 137, 653 (1999), viii+66.

[17] L. C. Evans and D. Gomes, *Linear programming interpretations of Mather's variational principle*, A tribute to J. L. Lions. ESAIM Control Optim. Calc. Var. **8** (2002), 693–702 (electronic).

[18] A. Fathi, Weak KAM Theorem in Lagrangian Dynamics, preliminary version, Lyon, version 2 February 2001.

[19] A. Fathi and A. Siconolfi, *Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation,* Invent. Math. **155** (2004), no. 2, 363–388.

[20] W. Gangbo, habilitation thesis.

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

[22] L. Granieri, *On action minimizing measures for the Monge-Kantorovitch problem*, preprint.

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

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

[25] M. Knott and C. Smith, *On the optimal mapping of distributions*, J. Optim. Theory Appl. **43** (1984), 39-49.

[26] V. Levin, *Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem*, Set Valued Analysis **7** (1999), 7-32.

[27] R. Mãné, Generic properties and problems of minimizing measures of Lagrangian systems, Nonlinearity **9** (1996), 273-310.

[28] J. N. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Mathematische Zeitschrift, Math. Z. **207** (1991), 169-207.

[29] G. Monge, Mémoire sur la Théorie des Déblais et des Remblais. *Hist. de l'Académie des Sciences de Paris, avec les Mémoires de Mathématiques et de Physique pour la même année*, pages 666-704, 1781.

[30] J. R. Munkres, Topology (second edition), Prentice Hall, 2000.

[31] L. De Pascale, M. Stella and L. Granieri, *Minimal measures, one-dimensional currents and the Monge-Kantorovitch problem* preprint.

[32] S. T. Rachev and L. Rüschendorf, Mass Transportation Problems, Vol. I and II, Springer-Verlag, 1998.

[33] D. Ruelle and D. Sullivan, *Currents, flows and diffeomorphisms*, Topology **14** (1975), 319-327.

[34] C. Villani, Topics in Optimal Transportation, American Mathematical Society, Providence, Rhode Island, 2003.

[35] G. Wolansky, *Optimal transportation in the presence of a prescribed pressure field*, preprint.