Available online at www.sciencedirect.com
ScienceDirect
Advances in Mathematics 207 (2006) 691–706
ADVANCES IN Mathematics
www.elsevier.com/locate/aim

The Monge problem for supercritical Mañé potentials on compact manifolds ☆

Patrick Bernard [a,*], Boris Buffoni [b]

[a] *Institut Fourier, Grenoble, France*
[b] *School of Mathematics, École Polytechnique Fédérale-Lausanne, SB/IACS/ANA Station 8, 1015 Lausanne, Switzerland*

Received 7 February 2005; accepted 17 January 2006

Available online 3 March 2006

Communicated by Luis Caffarelli

Abstract

We prove the existence of optimal transport maps for the Monge problem when the cost is a Finsler distance on a compact manifold. Our point of view consists in considering the distance as a Mañé potential, and to rely on recent developments in the theory of viscosity solutions of the Hamilton–Jacobi equation.

Résumé

On montre l'existence d'une application de transport optimale pour le problème de Monge lorsque le cout est une distance Finslerienne sur une variété compacte. Le nouveau point de vue consiste à considérer la distance comme un potentiel de Mañé, et à exploiter des développements récents sur les solutions de viscostité de l'équation de Hamilton–Jacobi.

Keywords: Hamilton–Jacobi equation; Transport map; Monge–Kantorovich problem; Lagrangian dynamics

The Monge transportation problem is to move one distribution of mass into another in an optimal way. Before we discuss this problem, let us describe a precise setting. We fix a space M,

☆ This paper was born during the visit of the first author to the Bernoulli center in EPFL, Lausanne, in Summer 2003. We thank this institution for its support.
* Corresponding author. On move to CEREMADE, Université de Paris Dauphine, Pl. du Maréchal de Lattre de Tassigny, 75775 Paris cedex 16, France.
E-mail addresses: patrick.bernard@ceremade.dauphine.fr (P. Bernard), boris.buffoni@epfl.ch (B. Buffoni).

0001-8708/$ – see front matter
doi:10.1016/j.aim.2006.01.003

which in the present paper will be a manifold, and a cost function $c \in C(M \times M, \mathbb{R})$. Given two probability measures μ_0 and μ_1, we call transport map a Borel map $F: M \to M$ that transports μ_0 onto μ_1. An optimal transport map is a transport map F that minimizes the total cost

$$\int_M c\big(x, F(x)\big)\, d\mu_0$$

among all transport maps. In many situations, optimal transport maps have remarkable geometric properties, at least at a formal level. Some of these properties were investigated by Monge at the end of the eighteenth century.

The question of existence of optimal transport maps was discussed much later in the literature. Some major steps were made by Kantorovich in 1942. He introduced both a relaxed problem and a dual problem that opened new approaches to the existence problem. When the cost is the square of the distance on an Euclidean vector space, Brenier proved the existence of an optimal transport map in [6] and also provided an interesting geometric description on the optimal maps, which have to be the gradient of a convex function. The argument was simplified, taking advantage of the Kantorovich dual problem, by Gangbo [13], and extended in many directions by Gangbo and McCann [14] and other authors, see our paper [4] for more details.

The case where the cost function is the distance on an Euclidean vector space is very natural, but more difficult. Sudakov announced a proof of the existence of an optimal map in 1979, but a gap was recently found in this proof. The strategy was to decompose the space into pieces of smaller dimension on which transport maps can be more easily built, and to glue these maps together. An essential hypothesis of the result is that the measure μ_0 is absolutely continuous. In the construction, it is necessary to control how this hypothesis behaves under decomposition to the subsets. Sudakov was not aware of these difficulties, and made wrong statements at that point, as was discovered only much later. It is interesting to notice for comparison that similar kind of difficulties had been faced and solved ten years earlier by Anosov in his ergodic theory of hyperbolic diffeomorphisms.

Correct proofs in the spirit of the work of Sudakov were written simultaneously by Caffarelli, Feldman and McCann in [7], Trudinger and Wang in [20] and slightly afterwards by Ambrosio [1]. These authors manage to build a decomposition of the space into line segments that have to be preserved by transport maps, called transport rays. They prove that the direction of these rays vary Lipschitz continuously. This regularity implies that the absolutely continuous measure μ_0 has absolutely continuous decompositions on these rays. See [1] for a remarkably written discussion on these works and of Sudakov's mistake. Before the proof of Sudakov were completed in these papers, Evans and Gangbo had provided a different proof under more stringent hypotheses in [10]. This proof is long and complicated, but it now appears as the first proof of the existence of a transport map in the case where the cost is a distance.

The methods inspired from Sudakov seem to allow many kinds of generalizations. The paper [7] treats all norms whose unit ball is smooth and strictly convex. It is worth mentioning that flat part in the unit ball represent a major difficulty. An important progress have been recently made in [3], which studies norms whose unit ball is a polyhedron. In another direction, Feldman and McCann [12] have treated the case where the cost is the distance on a Riemannian manifold. In this generalized setting, transport rays are not any more line segments, but pieces of geodesics. It is in this direction we will pursue in the present text.

Our goal is to prove the existence of transport maps for Finsler distances on manifolds (possibly non symmetric distances). In order to avoid superficial (and less superficial) additional technicalities, we shall work on a compact manifold M. Our first novelty is a new approach of the geometric part of the proof, that is the decomposition into transport rays. We believe that this new approach is interesting because, beyond being more general, it enlightens new links between the Monge problem and the general theory of Hamilton–Jacobi equations as presented by Fathi in [11]. In fact, all the relevant properties of the decomposition into transport rays are obtained by straightforward applications of results of [11]. In order to finish the proof, we rely on a secondary variational principle, in the lines of [2] and [3]. Our treatment of this secondary principle is quite different from these papers, and it is, we believe, shorter and clearer than the methods previously used in the literature.

1. Introduction

We state two versions of our main result, and prove the equivalence between these two statements.

1.1. Optimal transport maps for Finsler distances

In the present paper, the space M is a smooth compact connected manifold without boundary. We equip the manifold M with a C^2 Finsler metric, that is the data, for each $x \in M$, of a non-negative convex function $v \mapsto \|v\|_x$ on T_xM such that

- $\|\lambda v\|_x = \lambda\|v\|_x$ for all $\lambda > 0$ and all $v \in T_xM$ (positive 1-homogeneity),
- the function $(x, v) \mapsto \|v\|_x$ is C^2 outside of the zero section,
- for each $x \in M$, the function $v \mapsto \|v\|_x^2$ has positive definite Hessian at all vectors $v \neq 0$.

As a consequence, $\|v\|_x = 0$ exactly when $v = 0$ (positivity). Note that $\|\cdot\|_x$ is not assumed symmetric (that is, $\|-v\|_x \neq \|v\|_x$ is allowed). In standard terminology, the function $v \mapsto \|v\|_x$ is a Minkowsky metric on the vector space T_xM. See [5] for more one Finsler metrics (in particular, Theorem 1.2.2 and paragraph 6.2). We define the length of each smooth curve $\gamma : [0, T] \to M$ by the expression

$$l(\gamma) = \int_0^T \|\dot{\gamma}(t)\|_{\gamma(t)}\, dt.$$

The Finsler distance c is then given by the expression

$$c(x, y) = \inf_{\gamma} l(\gamma)$$

where the infimum is taken on the set of smooth curves $\gamma : [0, T] \to M$ (where T is any positive number) which satisfy $\gamma(0) = x$ and $\gamma(T) = y$. Note that the value of the infimum would not be changed by imposing the additional requirement that $\|\dot{\gamma}(t)\|_{\gamma(t)} \equiv 1$. The Finsler distance $c(x, y)$ is not necessarily symmetric, and thus is not properly speaking a distance. It does satisfy the triangle inequality, and $c(x, y) = 0$ if and only if $x = y$.

We shall consider the Monge transportation problem for the cost c. Given a Borel measure μ_0 on M, and a Borel map $F : M \to M$, we define the image measure $F_\sharp\mu_0$ by

$$F_\sharp\mu_0(A) := \mu_0\big(F^{-1}(A)\big)$$

for each Borel set $A \subset M$. The map F is said to transport μ_0 onto μ_1 if $F_\sharp\mu_0 = \mu_1$. We will present a short proof of

Theorem 1. *Let $c(x, y)$ be a Finsler distance on M. Let μ_0 and μ_1 be two probability measures on M, such that μ_0 is absolutely continuous with respect to the Lebesgue class. Then there exists a Borel map $F : M \to M$ such that $F_\sharp\mu_0 = \mu_1$, and such that the inequality*

$$\int_M c\big(x, F(x)\big)\,d\mu_0 \leqslant \int_M c\big(x, G(x)\big)\,d\mu_0$$

holds for each Borel map $G : M \to M$ satisfying $G_\sharp\mu_0 = \mu_1$. In other words, there exists an optimal map for the Monge transportation problem.

1.2. Optimal transport maps for supercritical Mañé potentials

We shall now present a generalization of Theorem 1, which is the natural setting for our proof. A function $L : TM \to \mathbb{R}$ is called a Tonelli Lagrangian if it is C^2 and satisfies:

Convexity For each $x \in M$, the function $v \mapsto L(x, v)$ is convex with positive definite Hessian at each point.

Superlinearity For each $x \in M$, we have $L(x, v)/\|v\|_x \to \infty$ as $\|v\|_x \to \infty$.

Given a Tonelli Lagrangian L and a time $T \in]0, \infty)$, we define the cost function

$$c_T^L(x, y) = \min_\gamma \int_0^T L\big(\gamma(t), \dot\gamma(t)\big)\,dt$$

where the minimum is taken on the set of curves $\gamma \in C^2([0, T], M)$ satisfying $\gamma(0) = x$ and $\gamma(T) = y$. That this minimum exists is standard, see [17] or [11]. The function

$$c^L(x, y) := \inf_{T \in]0,\infty)} c_T(x, y)$$

is called the Mañé potential of L. It was introduced and studied by Ricardo Mañé and then his students in [8,16]. Without additional hypothesis, the Mañé potential may be identically $-\infty$. So we assume in addition:

Supercriticality For each $x \neq y \in M^2$, we have $c(x, y) + c(y, x) > 0$.

The following result of Mañé [8,16] makes this hypothesis natural.

Proposition 1. *Let $L \in C^2(TM, \mathbb{R})$ be a Tonelli Lagrangian. For $k \in \mathbb{R}$, let c^{L+k} be the Mañé potential associated to the Lagrangian $L + k$. There exists a constant k_0 such that*:

- *For $k < k_0$, then $c^{L+k} \equiv -\infty$ and the Lagrangian $L + k$ is called subcritical.*

- *For $k \geqslant k_0$, the Mañé potential c^{L+k} is a Lipschitz function on $M \times M$ that satisfies the triangle inequality*

$$c^{L+k}(x,z) \leqslant c^{L+k}(x,y) + c^{L+k}(y,z)$$

for all x, y and z in M. In addition, we have $c^{L+k}(x,x) = 0$ for all $x \in M$.
- *For $k > k_0$, the Lagrangian L is supercritical, which means that $c^{L+k}(x,y) + c^{L+k}(y,x) > 0$ for $x \neq y$ in M.*

We will explain that the following theorem is equivalent to Theorem 1.

Theorem 2. *Let $c^L(x,y)$ be the Mañé potential associated to a supercritical Tonelli Lagrangian L. Let μ_0 and μ_1 be two probability measures on M, such that μ_0 is absolutely continuous with respect to the Lebesgue class. Then there exists an optimal transport map for the Monge transportation problem with cost $c^L(x,y)$.*

1.3. Supercritical Mañé potentials and Finsler distances

We prove the equivalence between Theorems 1 and 2. To each Tonelli Lagrangian L, we associate the Hamiltonian $H \in C^2(T^*M, \mathbb{R})$ defined by

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

and the energy function $E \in C^2(TM, \mathbb{R})$ defined by

$$E(x,v) = \partial_v L(x,v).v - L(x,v).$$

The function H is also convex and superlinear. The mapping $\partial_v L : TM \to T^*M$ is a C^1 diffeomorphism, whose inverse is the mapping $\partial_p H$. We have $E = H \circ \partial_v L$.

Lemma 2. *Let L be a supercritical Tonelli Lagrangian. There exists a constant K such that, for each $x \neq y$ in M, there exists a time $T \in]0, K]$ and a minimizing extremal $\gamma \in C^2([0,T], M)$ such that $\gamma(0) = x$, $\gamma(T) = y$, and $\int_0^T L(\gamma(t), \dot\gamma(t))\,dt = c_T^L(x,y) = c^L(x,y)$. Moreover, if γ is such a curve, then*

$$E\big(\gamma(t), \dot\gamma(t)\big) \equiv 0.$$

Proof. We shall prove, and use, this lemma only in the case where L is positive. Note first that the function $(x,y,T) \mapsto c_T^L(x,y)$ is continuous on $M \times M \times]0, \infty)$. It is not hard to see, in view of the superlinearity of L, that the function $T \mapsto c_T^L(x,y)$ goes to infinity as T goes to zero if $x \neq y$. On the other hand, setting $\delta = \inf L > 0$, we obviously have the minoration $c_T^L \geqslant \delta T$. Since c_1^L is bounded, this implies the existence of a constant K such that $c_T^L > c_1^L$ for $T \geqslant K$. As a consequence, the function $T \mapsto c_T(x,y)$ reaches its minimum on $]0, K]$ for each $x \neq y$.

Let $x \neq y$ be two points on M. There exists a $T \in]0, K]$ such that $c_T^L(x,y) = c^L(x,y)$. Now by standard results on the calculus of variations, there exists a C^2 curve $\gamma : [0,T] \to M$ satisfying $\gamma(0) = x$, $\gamma(T) = y$ and $\int_0^T L(\gamma(t), \dot\gamma(t))\,dt = c^L(x,y)$. In addition, this curve satisfies the Euler–Lagrange equations, and in particular the energy $E(\gamma(t), \dot{\gamma}(t))$ is constant on $[0, T]$.

Let $\gamma_\lambda : [0, \lambda T] \to M$ be defined by $\gamma_\lambda(t) = \gamma(t/\lambda)$. The function

$$f(\lambda) := \int_0^{T\lambda} L(\gamma_\lambda, \dot{\gamma}_\lambda)\, dt = \lambda \int_0^T L\big(\gamma, \lambda^{-1}\dot{\gamma}\big)\, dt$$

clearly has to reach its minimum at $\lambda = 1$. On the other hand, a classical computation shows that the function f is differentiable, and that $f'(1) = -\int_0^T E(\gamma(t), \dot{\gamma}(t))\, dt$. This proves that $E(\gamma(t), \dot{\gamma}(t)) \equiv 0$. □

The following proposition implies that the transportation problem for Finsler distances, the transportation problem for supercritical Mañé potentials, and the transportation problem for the Mañé potentials of positive Tonelli Lagrangians are equivalent problems.

Proposition 3. *If L is a supercritical Tonelli Lagrangian, then there exists a Finsler distance c (associated to a C^2 Finsler metric) and a smooth function $f : M \to \mathbb{R}$ such that*

$$c(x, y) = c^L(x, y) + f(y) - f(x).$$

Conversely, given a Finsler distance c (associated to a C^2 Finsler metric) there exists a positive Tonelli Lagrangian L such that

$$c(x, y) = c^L(x, y).$$

Proof. The first part of this proposition is the content of [15]. For the converse, we consider the Lagrangian

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

Note that the associated energy function is

$$\tilde{E}(x, v) = \frac{\|v\|_x^2 - 1}{2}.$$

Let us now consider a positive Tonelli Lagrangian L such that $0 < L \leqslant \tilde{L}$ and such that $L = \tilde{L}$ on the set $\{\|v\|_x \geqslant 1/2\} \subset TM$. In order to see that such a Lagrangian exists, consider a smooth convex function $f : [0, \infty) \to [0, \infty)$ that vanishes in a small neighborhood of 0 and such that $f(s) \leqslant (1 + s^2)/2$ with equality for $s \geqslant 1/2$; and consider a smooth function $g(x, v) : TM \to \mathbb{R}$ such that g is positive and $\partial_2 g(x, v)$ is positive definite where $f(\|v\|_x)$ vanishes, and such that g is zero on $\{\|v\|_x \geqslant 1/2\} \subset TM$. It is easy to check that the Lagrangian $L(x, v) = f(\|v\|_x) + \epsilon g(x, v)$ satisfies the desired requirements when $\epsilon > 0$ is small enough.

Since $L \leqslant \tilde{L}$, we have $c^L \leqslant c^{\tilde{L}}$. Moreover, for $x \neq y$, let $\gamma : [0, T] \to M$ be the optimal (for c^L) trajectory obtained by Lemma 2. We have $E(\gamma(t), \dot{\gamma}(t)) = \tilde{E}(\gamma(t), \dot{\gamma}(t)) \equiv 0$ hence $\|\dot{\gamma}\| \equiv 1$. As a consequence,

$$c^L(x, y) = \int_0^T L\big(\gamma(t), \dot{\gamma}(t)\big)\, dt = \int_0^T \tilde{L}\big(\gamma(t), \dot{\gamma}(t)\big)\, dt = l(\gamma) \geqslant c(x, y).$$

We have proved that $c \leqslant c^L \leqslant c^{\tilde{L}}$. These are equalities because $c^{\tilde{L}} \leqslant c$. Indeed, for all smooth curves $\gamma : [0, T] \to M$ which satisfy $\gamma(0) = x$, $\gamma(T) = y$ and $\|\dot{\gamma}(t)\|_{\gamma(t)} \equiv 1$ with $T = l(\gamma) > 0$, we get

$$c^{\tilde{L}}(x, y) \leqslant \int_0^T \tilde{L}\big(\gamma(t), \dot{\gamma}(t)\big)\, dt = l(\gamma).$$

Since $c(x, y)$ is the infimum of the lengths of such curves, we have $c^{\tilde{L}}(x, y) \leqslant c(x, y)$. □

1.4. General convention

In the sequel, we will prove Theorem 2 for a positive Lagrangian L and denote c^L simply by c. In view of Proposition 3, this implies the general form of Theorem 2, as well as Theorem 1. We fix, once and for all, a positive Tonelli Lagrangian L, and a positive number $\delta > 0$ such that

$$L(x, v) \geqslant \delta$$

for each $(x, v) \in TM$.

The general scheme of our proof is somewhat similar to the one introduced by Sudakov, and followed, in [1,3,7,12,20] and other papers. Like these papers, our proof involves a decomposition of the space M into distinguished curves, called transport rays. We introduce these rays in Section 3 and describe their geometric properties. In this geometric part of the proof, our point of view is quite different from the literature as we emphasize the link with the theory of viscosity solutions as developed in [11], and manage to obtain all the relevant properties of transport rays as a straightforward application of general results of [11]. For the second part of the proof, all the papers mentioned above involve subtle decompositions of measures on these transport rays. It is at this step that the paper of Sudakov contains a gap. We simplify this step by introducing a secondary variational principle in Section 2. Note that secondary variational principles have already been introduced by Ambrosio, Kirchheim and Pratelli in [3] for related problems. This secondary problem is studied in Section 5 by a quite simple method, which, surprisingly, seems new. This methods allows a neat clarification of the end of the proof compared to the existing literature. All the difficulties involving measurability issues and absolute continuity of disintegrated measures are reduced to a single and simple Fubini-like theorem, exposed in Section 4.

2. Transport plans

We introduce our secondary variational principle and recall the necessary generalities on the Monge problem. Beyond the references provided below, the pedagogical texts [1,19,21] may help the reader who wants more details. We define the quantity

$$C(\mu_0, \mu_1) := \inf_F \int_M c\big(x, F(x)\big)\, d\mu_0,$$

where the infimum is taken on the set of Borel maps $F : M \to M$ that transport μ_0 onto μ_1. It is useful, following Kantorovich, to relax this infimum to a nicer minimization problem. A Borel measure μ on $M \times M$ is called a transport plan if it satisfies the equalities $\pi_{i\sharp}\mu = \mu_i$, where $\pi_0 : M \times M \to M$ is the projection on the first factor, and $\pi_1 : M \times M \to M$ is the projection on the second factor. Clearly, any transport map F can be considered as the transport plan

$$(\mathrm{Id} \times F)_\sharp \mu_0.$$

Following Kantorovich, we consider the minimum

$$K(\mu_0, \mu_1) = \min_\mu \int_{M\times M} c\, d\mu$$

taken on the set of transport plans. It is well known and easy to prove that this minimum exists. The equality

$$K(\mu_0, \mu_1) = C(\mu_0, \mu_1)$$

holds if μ_0 has no atom, see [1, Theorem 2.1]. Note that this equality is very general, see [18] for a discussion. As a consequence, if μ_0 has no atom, it is equivalent to prove the existence of an optimal transport map and to prove that there exists an optimal transport plan concentrated on the graph of a Borel function.

Let us define the second cost function

$$\sigma(x, y) = \big(c(x, y)\big)^2.$$

This cost is chosen in order that the following refined form of Theorem 1 holds.

Theorem 3. *Let $\mathcal{O}$ be the set of optimal transport plans for $K(\mu_0, \mu_1)$ with the cost c. The minimum*

$$\min_{\mu \in \mathcal{O}} \int \sigma\, d\mu$$

exists. In addition, if μ_0 is absolutely continuous, then there is one and only one transport plan μ realizing this optimum, and this transport plan is concentrated on the graph of a Borel function that is an optimal transport map for the cost c.

This result will be proved in Section 5. The idea of introducing secondary variational problem as in this statement has already been used by Ambrosio, Kirchheim and Pratelli, see [3] and also [2]. Our treatment in Section 5 is inspired from these references, although it is somewhat different. It allows substantial simplifications compared to the literature.

3. Kantorovich potential and calibrated curves

We present the decomposition in transport ray, which is the standard initial step in the construction of optimal maps. This construction is based on well-understood general results on viscosity subsolutions of the Hamilton–Jacobi equation, as presented in [11]. Making this connection is one of the novelties of the present paper.

Since the cost function we consider satisfies the triangle inequality

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

for all x, y and z in M, as well as the identity $c(x,x)=0$ for all $x \in M$, we can take advantage of the following general duality result, inspired from Kantorovich, see, for example, [9,12,21].

Proposition 4. *Given two measures μ_0 and μ_1, there exists a function $u \in C(M,\mathbb{R})$ that satisfies*

$$u(y) - u(x) \leqslant c(x,y)$$

for all x and y in M, and

$$K(\mu_0,\mu_1) = \int_M u \, d(\mu_1 - \mu_0).$$

In addition, for each optimal transport plan μ, the equality $u(y) - u(x) = c(x,y)$ holds for μ-almost every $(x,y) \in M^2$. The function u is called a Kantorovich potential.

The present paper is born from the observation that the Kantorovich potentials are viscosity subsolutions of the Hamilton–Jacobi equation as studied in [11]. In order to explain this connection, it is necessary to define the Hamiltonian function $H \in C^2(T^*M,\mathbb{R})$ by

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

Note that the mapping $\partial_v L : TM \to T^*M$ is a C^1 diffeomorphism, whose inverse is the mapping $\partial_p H$. It is proved in [11] that the following properties are equivalent for a function $w \in C(M,\mathbb{R})$.

(1) The function w satisfies the inequality $w(y) - w(x) \leqslant c(x,y)$ for all x and y in M.
(2) The function w is a viscosity subsolution of the Hamilton–Jacobi equation $H(x,dw)=0$, i.e., each smooth function $f : M \to \mathbb{R}$ satisfies the inequality $H(x,df(x)) \leqslant 0$ at each point of minimum x of the difference $f - w$.
(3) The function w is Lipschitz and satisfies the inequation $H(x,dw_x) \leqslant 0$ at almost every point. This inequality then holds at all point of differentiability x of w.

(4) The function w is Lipschitz and, for almost every $x \in M$, it satisfies the inequation $\forall v \in T_x M$, $L(x, v) \geqslant dw_x(v)$. This inequality then holds at all point of differentiability x of w.

Although there may exist several Kantorovich potentials, we shall fix one of them, u, for the sequel.

Definition 5. Following Fathi [11], we call *calibrated curve* a continuous and piecewise C^1 curve $\gamma : I \to M$ that satisfies

$$u\big(\gamma(t)\big) - u\big(\gamma(s)\big) = \int_s^t L\big(\gamma(\tau), \dot\gamma(\tau)\big)\, d\tau = c\big(\gamma(s), \gamma(t)\big) \tag{1}$$

whenever $s \leqslant t$ in I, where I is a non-empty interval of $\mathbb{R}$ (possibly a point). A calibrated curve $\gamma : I \to M$ is called *non-trivial* if the interval I has non-empty interior.

Note that the first of the equalities in (1) implies the second, since the inequalities

$$u\big(\gamma(t)\big) - u\big(\gamma(s)\big) \leqslant c\big(\gamma(s), \gamma(t)\big) \leqslant \int_s^t L\big(\gamma(\tau), \dot\gamma(\tau)\big)\, d\tau$$

hold for any curve γ. It is obvious that non-trivial calibrated curves are minimizing extremals of L, and as a consequence they are C^2 curves. In addition, the concatenation of two calibrated curves is calibrated, so that each calibrated curve can be extended to a maximal calibrated curve, that is, its domain I cannot be further extended without loosing calibration. Note that I is closed when γ is a maximal calibrated curve.

Definition 6. We call *transport ray* the image of a non-trivial maximal calibrated curve.

It will be useful to consider, following [7] and [12], the functions α and $\beta : M \to [0, \infty)$ defined as follows:

- $\alpha(x)$ is the supremum of all times $T \geqslant 0$ such that there exists a calibrated curve $\gamma : [-T, 0] \to M$ that satisfies $\gamma(0) = x$.
- $\beta(x)$ is the supremum of all times $T \geqslant 0$ such that there exists a calibrated curve $\gamma : [0, T] \to M$ that satisfies $\gamma(0) = x$ (the fact that α and β are finite is a consequence of Lemma 9 below).

Definition 7. Let us denote by $\mathcal{T}$ the subset of M obtained as the union of all transport rays, or equivalently the set of points $x \in M$ such that $\alpha(x) + \beta(x) > 0$. For $\epsilon \geqslant 0$, we denote by $\mathcal{T}_\epsilon$ the set of points $x \in M$ that satisfy $\alpha(x) > \epsilon$ and $\beta(x) > \epsilon$. Clearly $\mathcal{T}_\epsilon \subset \mathcal{T}$ for all $\epsilon \geqslant 0$. The set $\mathcal{E} := \mathcal{T} - \mathcal{T}_0$ is the set of ray ends.

Proposition 8. *The function u is differentiable at each point of $\mathcal{T}_0$. For each point $x \in \mathcal{T}_0$, there exists a single maximal calibrated curve*

$$\gamma_x : \big[-\alpha(x), \beta(x)\big] \to M \quad \textit{such that} \quad \gamma_x(0) = x. \tag{2}$$

This curve satisfies the relations

$$du_x = \partial_v L\big(x, \dot\gamma_x(0)\big) \quad \textit{or equivalently} \quad \dot\gamma_x(0) = \partial_p H(x, du_x).$$

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

Proof. This proposition is Theorem 4.5.5 of Fathi's book [11]. □

Lemma 9. *Let $\gamma : [a, b] \to M$ be a non-trivial calibrated curve. Then, for all $t \in \,]a, b[$, the function u is differentiable at $\gamma(t)$ and*

$$du_{\gamma(t)}\big(\dot\gamma(t)\big) \geqslant \delta$$

(*see* 1.4 *for the definition of* δ). *As a consequence, the map $\gamma : [a, b] \to M$ is an embedding* (*it is one-to-one and has non-zero derivative on* $]a, b[$) *and transport rays are non-trivial embedded arcs.*

Proof. Since u is a viscosity subsolution of the Hamilton–Jacobi equation, see the equivalence below Proposition 4, we have $L(x, v) \geqslant du_x(v)$ for all $v \in T_x M$ at each point of differentiability of u. As a consequence, the inequality

$$L\big(\gamma(t), \dot\gamma(t)\big) \geqslant du_{\gamma(t)}\big(\dot\gamma(t)\big)$$

holds for each $t \in \,]a, b[$. Integrating the above inequality gives

$$\int_a^b L\big(\gamma(t), \dot\gamma(t)\big)\, dt \geqslant u\big(\gamma(b)\big) - u\big(\gamma(a)\big),$$

which is an equality because the curve γ is calibrated. As a consequence, we have

$$du_{\gamma(t)}\big(\dot\gamma(t)\big) = L\big(\gamma(t), \dot\gamma(t)\big) \geqslant \delta$$

for all $t \in \,]a, b[$. □

Lemma 10. *The functions α and β are bounded and upper semi-continuous, hence Borel measurable. As a consequence, the sets $\mathcal{T}$ and $\mathcal{T}_\epsilon$, $\epsilon \geqslant 0$, are Borel.*

Proof. We shall consider only the function α. We have just seen that, for each non-trivial calibrated curve $\gamma : I \to M$, the function $f(t) = u \circ \gamma(t)$ is differentiable and satisfies $f'(t) \geqslant \delta$. Since the continuous function u is bounded on the compact manifold M, we conclude that the functions α and β are bounded. In order to prove that the function α is semi-continuous, let us consider a sequence $x_n \in M$ that is converging to a limit x and is such that $\alpha(x_n) \geqslant T$. We have to prove that $\alpha(x) \geqslant T$. There exists a sequence $\gamma_n : [-T, 0] \to M$ of calibrated curves such that $\gamma_n(0) = x_n$. There exists a subsequence of γ_n that is converging uniformly on $[-T, 0]$ to a curve $\gamma : [-T, 0] \to M$. It is easy to see that the curve γ is calibrated and satisfies $\gamma(0) = x$. As a consequence, we have $\alpha(x) \geqslant T$. □

Definition 11. For $x \in M$, let us denote by R_x the union of the transport rays containing x. We also denote by R_x^+ the set of points $y \in M$ such that $u(y) - u(x) = c(x, y)$.

Note that $R_x = \gamma_x([-\alpha(x), \beta(x)])$ when $x \in \mathcal{T}_0$, where γ_x is given in (2).

Lemma 12. *We have $R_x^+ = \gamma_x([0, \beta(x)])$ when $x \in \mathcal{T}_0$, and $R_x^+ = \{x\}$ when $x \in M - \mathcal{T}$.*

Proof. Let x be a point of $\mathcal{T}_0$. By the calibration property of γ_x, we have, for $t \in [0, \beta(x)]$, $\gamma_x(t) - \gamma_x(0) = c(\gamma_x(0), \gamma_x(t))$, which is precisely saying that $\gamma_x(t) \in R_x^+$. Conversely, let us fix a point $x \in M$ and let y be a point of R_x^+. There exists a time $T \geqslant 0$ and a curve $\gamma : [0, T] \to M$ such that $\int_0^T L(\gamma(t), \dot\gamma(t))\,dt = c(x, y)$, $\gamma(0) = x$ and $\gamma(T) = y$. Since $c(x, y) = u(y) - u(x)$, the curve γ is calibrated. If $x \in \mathcal{T}_0$, then $\gamma = \gamma_x|_{[0,T]}$ hence $y = \gamma(T) = \gamma_x(T) \in \gamma_x([0, \beta(x)])$. If $x \notin \mathcal{T}$, then there is no non-trivial calibrated curve starting at x, so we must have $y = x$ in the above discussion, and $R_x^+ = \{x\}$. □

Proposition 13. *The transport plan μ is optimal for the cost c if and only if it is concentrated on the closed set*

$$\bigcup_{x \in M} \{x\} \times R_x^+ = \big\{(x, y) \in M^2 \colon\ c(x, y) = u(y) - u(x)\big\}.$$

Proof. Recall from Proposition 4 that any optimal transport plan is concentrated on this set. Reciprocally, if μ is a transport plan concentrated on this set, then

$$K(\mu_0, \mu_1) = \int_M u\,d(\mu_1 - \mu_0) = \int_{M^2} \big(u(y) - u(x)\big)\,d\mu = \int_{M^2} c\,d\mu$$

and μ is thus optimal. □

4. Fubini theorem

The geometric informations on transport rays that have been obtained in the preceding section imply the following crucial Fubini-like result, whose proof is the goal of the present section:

Proposition 14. *Let Λ be a Borel subset of $\mathcal{T}$ such that the intersection $\Lambda \cap R$ has zero 1-Hausdorff measure for each transport ray R. Then the set Λ has zero Lebesgue measure.*

For comparison with the literature, see [7, Lemma 25], [20, Section 4] or [12, Lemma 24], we mention:

Corollary 15. *The set $\mathcal{E} = \mathcal{T} - \mathcal{T}_0$ of ray ends has zero Lebesgue measure.*

We now turn to the proof of Proposition 14, which occupies the end of this section. The method is standard. Let k be the dimension of M.

Definition 16. We call *transport beam* any given pair (B, χ), where B is a bounded Borel subset of $\mathbb{R}^k$ and $\chi : B \to M$ is a Lipschitz map (not necessarily one-to-one) such that:

- There exists a bounded Borel set $\Omega \in \mathbb{R}^{k-1}$ and two bounded Borel functions $a < b : \Omega \to \mathbb{R}$ such that

$$B = \{(\omega, s) \in \Omega \times \mathbb{R} \text{ s.t. } a(\omega) \leqslant s \leqslant b(\omega)\} \subset \mathbb{R}^k = \mathbb{R}^{k-1} \times \mathbb{R}.$$

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

Lemma 17. *If (B, χ) is a transport beam, then the set $\Lambda \cap \chi(B)$ has zero Lebesgue measure.*

Proof. For each $\omega \in \Omega$, the curve χ_ω is a bi-Lipschitz homeomorphism onto its image. Since in addition, the set $\Lambda \cap \chi(\{\omega\} \times [a(\omega), b(\omega)])$ has zero 1-Hausdorff measure, the set $\chi^{-1}(\Lambda)$ intersects each vertical line $\{\omega\} \times \mathbb{R}$ along a set of zero 1-Hausdorff measure. In view of the classical Fubini theorem, the set $\chi^{-1}(\Lambda)$ has zero Lebesgue measure in $\mathbb{R}^k$. Since the k-Hausdorff measure on B (associated to the restricted Euclidean metric) is the restriction to B of the Lebesgue measure of $\mathbb{R}^k$, the set $\chi^{-1}(\Lambda)$ has zero k-Hausdorff measure in B, for the Hausdorff measure associated to the induced metric. Since Lipschitz maps send sets of zero k-Hausdorff measure onto sets of zero k-Hausdorff measure, we conclude that the set $\Lambda \cap \chi(B) \subset \chi(\chi^{-1}(\Lambda))$ has zero Lebesgue measure in M. □

We can now conclude the proof of Proposition 14 by the following lemma.

Lemma 18. *There exists a countable family $(B_{i,j}, \chi_{i,j})$, $(i, j) \in \mathbb{N}^2$ of transport beams such that the images $\chi_{i,j}(B_{i,j})$ cover the set $\mathcal{T}$.*

Proof. Let D be the closed unit ball in $\mathbb{R}^{k-1}$. Let $\psi_i : D \to M$, $i \in \mathbb{N}$, be a countable family of smooth embeddings such that, for each maximal calibrated curve $\gamma : [a, b] \to M$, the curve $\gamma(]a, b[)$ intersects the image of ψ_i for some $i \in \mathbb{N}$. In order to build such a family of embeddings, let us consider a finite atlas Θ of M composed of charts $\theta : B_3 \to M$, where B_r is the open ball of radius r centered at zero in $\mathbb{R}^k$. We assume that the finite family of open sets $\theta(B_1), \theta \in \Theta$, covers M. For $n = 1, \dots, k$ and $q \in \mathbb{Q} \cap [-1, 1]$, we consider the embedded disk $\mathcal{D}_{n,q} \subset B_3$ formed by points $x = (x_1, \dots, x_k) \in \bar{B}_2$ which satisfy $x_n = q$. The countable family

$$\theta(\mathcal{D}_{n,q}), \quad \theta \in \Theta,\ n = 1, \dots, k,\ q \in \mathbb{Q} \cap [-1, 1],$$

of embedded disks of M forms a web which intersects all non-trivial curves of M, hence all transport rays. We have constructed a countable family of embedded disks which intersects all transport rays.

For each $(i, j) \in \mathbb{N}^2$ let us consider the set $\Omega_{i,j} = D \cap \psi_i^{-1}(\mathcal{T}_{1/j})$. Let $a_{i,j}(\omega)$ and $b_{i,j}(\omega)$: $\Omega_{i,j} \to \mathbb{R}$ be the functions $-\alpha \circ \psi_i$ and $\beta \circ \psi_i$. Let $B_{i,j}$ be the set of points $(\omega, s) \subset \Omega_{i,j} \times \mathbb{R}$ such that $a_{i,j}(\omega) \leqslant s \leqslant b_{i,j}(\omega)$. To finish, we define the map $\chi_{i,j} : B_{i,j} \to M$ by

$$\chi_{i,j}(\omega, s) = \gamma_{\psi_i(\omega)}(s).$$

We claim that, for each $(i, j) \in \mathbb{N}^2$, the map $\chi_{i,j}$ is Lipschitz, so that the pair $(B_{i,j}, \chi_{i,j})$ is a transport beam. In order to prove this claim, remember that there exists a vector field E on TM, the Euler–Lagrange vector field, such that the extremals are the projections of the integral curves of E. Because of energy conservation, this vector field generates a complete flow, denoted by $f_s : TM \to TM$ for $s \in \mathbb{R}$. From the fact that the Hamiltonian H is C^2 and f_s is in Legendre duality with the Hamiltonian flow, we deduce that $(s, x, v) \to f_s(x, v)$ is C^1. We have

$$\chi_{i,j}(x, s) = P_M \circ f_s\big(x, \dot\gamma_{\psi_i(x)}(0)\big) \quad \forall (x, s) \in B_{i,j},$$

where $P_M : TM \to M$ is the canonical projection on M. This map is Lipschitz in view of Proposition 8. If R is a transport ray, it is clear that R is contained in one of the images $\chi_{i,j}(B_{i,j})$. □

5. The distinguished transport plan

We shall now prove Theorem 2, and hence Theorem 1. Our approach is based on remarks in [2] and [3], however it seems new, and is surprisingly simple. Let μ be a transport plan that is optimal for the cost c, and, among these optimal transport plans, minimizes the functional $\int \sigma \, d\mu$. The existence of such a plan is straightforward.

Proposition 19. *There exists a set*

$$\Gamma \subset \bigcup_{x\in M} \{x\} \times R_x^+, \tag{3}$$

which is a countable union of compact sets, such that $\mu(\Gamma) = 1$ and which is monotone in the following sense: If (x_i, y_i), $i \in \{1, \dots, k\}$, is a finite family of points of Γ and if $j(i)$ is a permutation such that $y_{j(i)} \in R_{x_i}^+$ then

$$\sum_{i=1}^k \sigma(x_i, y_{j(i)}) \geqslant \sum_{i=1}^k \sigma(x_i, y_i).$$

Proof. Let us consider the cost function ζ, where $\zeta(x, y) : M \times M \to [0, \infty]$ is the lower semi-continuous function defined by $\zeta(x, y) = \sigma(x, y)$ if $u(y) - u(x) = c(x, y)$ and $\zeta(x, y) = \infty$ if not. Note that $\int \zeta \, d\mu = \int \sigma \, d\mu$ is finite. Theorem 3.2 of [2] implies the existence of a Borel set $\tilde\Gamma$ on which μ is concentrated, and which is monotone. By interior regularity of the Borel measure μ, there exists a set $\Gamma \subset \tilde\Gamma$ which is a countable union of compact sets and on which μ is concentrated. Being a subset of the monotone set $\tilde\Gamma$, the set Γ is itself monotone. □

Definition 20. Let Λ be the set of points $x \in M$ such that the set

$$\Gamma_x := \big\{y \in M \colon (x, y) \in \Gamma\big\}$$

contains more than one point, where Γ is defined in (3).

Lemma 21. *The set Λ is Borel measurable.*

Proof. Let K^n, $n \in \mathbb{N}$ be an increasing sequence of compact sets such that $\Gamma = \bigcup K^n$. For each $x \in M$, let $\delta_n(x)$ be the diameter of the compact set K_x^n of points $y \in M$ such that $(x, y) \in K^n$. It is not hard to see that the function $\delta_n(x)$ is upper semi-continuous, hence Borel measurable. Since

$\Gamma_x = \bigcup_{n\in\mathbb{N}} K_x^n$, we have $\delta(x) = \sup_n \delta_n(x)$, where $\delta(x)$ is the diameter of Γ_x. As a consequence, the function δ is Borel measurable, and the set $\Lambda = \{x \in M,\ \delta(x) > 0\}$ is Borel. □

Proposition 22. *We have $\Lambda \subset \mathcal{T}$, and the intersection $\Lambda \cap R$ is at most countable for each transport ray R.*

Proof. If $x \notin \mathcal{T}$, then $R_x^+ = \{x\}$ hence $\Gamma_x \subset \{x\}$, and $x \notin \Lambda$. Let us now consider a transport ray R that is the image of a maximal calibrated curve $\gamma : [\alpha, \beta] \to M$. Let us denote by $h : [\alpha, \beta] \to \mathbb{R}$ the function $u \circ \gamma$, which is strictly increasing (Lemma 9). Note that $\sigma(\gamma(s), \gamma(t)) = (h(t) - h(s))^2$ for $s \leqslant t$ in $[\alpha, \beta]$. In view of the monotonicity of Γ, we have

$$\big(h(t) - h(s)\big)^2 + \big(h(t') - h(s')\big)^2 \leqslant \big(h(t) - h(s')\big)^2 + \big(h(t') - h(s)\big)^2$$

or equivalently

$$\big(h(t) - h(t')\big)\big(h(s) - h(s')\big) \geqslant 0 \tag{4}$$

whenever $(\gamma(s), \gamma(t)) \in \Gamma$, $(\gamma(s'), \gamma(t')) \in \Gamma$, $s' \leqslant t$, and $s \leqslant t'$. Following [14] or [2], we observe that this implies the property:

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

This property implies that the set of values of s in $[\alpha, \beta]$ such that $\Gamma_{\gamma(s)}$ contains more than one element is at most countable. Indeed, for all integers $n \geqslant 1$, let S_n be the set of values of $s \in [\alpha, \beta]$ such that there exist $t_1, t_2 \in [\alpha, \beta]$ with $(\gamma(s), \gamma(t_1)) \in \Gamma$, $(\gamma(s), \gamma(t_2)) \in \Gamma$ and $t_2 - t_1 \geqslant 1/n$. If $s < s'$ are in S_n and if t_1, t_2 and t_1', t_2' are as above with respect to s and s', then $\alpha \leqslant t_1 \leqslant t_2 - 1/n < t_2 \leqslant t_1' \leqslant t_2' - 1/n < t_2' \leqslant \beta$ and thus $\beta - \alpha \geqslant 2/n$. More generally, if S_n contains at least j points, then $\beta - \alpha \geqslant j/n$. As the interval $[\alpha, \beta]$ is bounded, the set S_n is finite for all n, which leads to the conclusion. □

Theorem 2 can now be proved in a very standard way. In view of Section 4, the set Λ has zero Lebesgue measure in M. The set $Z = M - \Lambda$ is a Borel set of full Lebesgue measure, $\mu_0(Z) = 1$. Denoting by $\pi_0 : M \times M \to M$ the projection on the first factor, we observe that the set $\Gamma_Z = \Gamma \cap \pi_0^{-1}(Z)$ is a Borel graph on which μ is concentrated (because $\mu(\Gamma) = 1$). By the easy Proposition 2.1 of [1], we conclude that the plan μ is induced from a transport map F. We then have

$$\int_{M\times M} c\, d\mu = \int_M c\big(x, F(x)\big)\, d\mu_0(x) = K(\mu_0, \mu_1) = C(\mu_0, \mu_1),$$

so that the map F is optimal for the cost c. This ends the proof of Theorems 2 and 1.

References

[1] L. Ambrosio, Lecture Notes on Optimal Transport Problems.
[2] L. Ambrosio, A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in: Lecture Notes in Math., vol. 1813, Springer, 2003, pp. 123–160.

[3] L. Ambrosio, B. Kirchheim, A. Pratelli, Existence of optimal transport maps for crystalline norms, Duke Math. J. 125 (2) (2004) 207–241.
[4] P. Bernard, B. Buffoni, Optimal mass transportation and Mather theory, J. Eur. Math. Soc., in press.
[5] D. Bao, S.S. Chern, Z. Shen, An Introduction to Riemann–Finsler Geometry, Grad. Texts in Math., vol. 200, Springer, New York, 2000.
[6] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. 44 (1991) 375–417.
[7] L. Caffarelli, M. Feldman, R.J. McCann, Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs, J. Amer. Math. Soc. 15 (2002) 1–26.
[8] G. Contreras, J. Delgado, R. Iturriaga, Lagrangian flows: The dynamics of globally minimizing orbits II, Bol. Soc. Brasil. Mat. (N.S.) 28 (2) (1997) 155–196.
[9] L.C. Evans, Partial differential equations and Monge–Kantorovich mass transfer, in: Current Developments in Mathematics, Cambridge, MA, 1997, Int. Press, Boston, MA, 1999, pp. 65–126.
[10] L.C. Evans, W. Gangbo, Differential equation methods for the Monge–Kantorovich mass transfer problem, Mem. Amer. Math. Soc. 137 (1999) 653, viii+66.
[11] A. Fathi, Weak KAM Theorem in Lagrangian Dynamics, in press.
[12] M. Feldman, R.J. McCann, Monge's transport problem on a Riemannian manifold, Trans. Amer. Math. Soc. 354 (2002) 1667–1697.
[13] W. Gangbo, An elementary proof of the polar factorization of vector-valued functions, Arch. Ration. Mech. Anal. 128 (4) (1994) 381–399.
[14] W. Gangbo, R.J. McCann, The geometry of optimal transportation, Acta Math. 177 (1996) 113–161.
[15] R. Iturriaga, H. Sánchez-Morgado, Finsler metrics and action potentials, Proc. Amer. Math. Soc. 128 (11) (2000) 3311–3316.
[16] R. Mañé, Lagrangian flows: The dynamics of globally minimizing orbits, Bol. Soc. Brasil. Mat. (N.S.) 28 (2) (1997) 141–153.
[17] J.N. Mather, Action minimizing invariant measures for positive definite Lagrangian systems, Math. Z. 207 (1991) 169–207.
[18] A. Pratelli, On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation, preprint, 2004.
[19] S.T. Rachev, L. Rüschendorf, Mass Transportation Problems, vols. I and II, Springer, 1998.
[20] N.S. Trudinger, X.J. Wang, On the Monge mass transfer problem, Calc. Var. Partial Differential Equations 13 (2001) 19–31.
[21] C. Villani, Topics in Optimal Transportation, Amer. Math. Soc., Providence, RI, 2003.

Further reading

[22] L.V. Kantorovich, On the transfer of masses, Dokl. Akad. Nauk USSR 37 (1942) 227–229.
[23] L.V. Kantorovich, On a problem of Monge, Uspekhi Mat. Nauk 3 (1948) 225–226.
[24] 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, 1781, pp. 666–704.
[25] V.N. Sudakov, Geometric problems in the theory of infinite-dimensional probability distributions, Proc. Steklov Inst. Math. 141 (1979) 1–178.