Advanced Studies in Pure Mathematics 47-2, 2007
Asymptotic Analysis and Singularities
pp. 397–420

Weak KAM pairs and Monge-Kantorovich duality

Patrick Bernard and Boris Buffoni

Abstract.

The dynamics of globally minimizing orbits of Lagrangian systems can be studied using the Barrier function, as Mather first did, or using the pairs of weak KAM solutions introduced by Fathi. The central observation of the present paper is that Fathi weak KAM pairs are precisely the admissible pairs for the Kantorovich problem dual to the Monge transportation problem with the Barrier function as cost. We exploit this observation to recover several relations between the Barrier functions and the set of weak KAM pairs in an axiomatic and elementary way.

§1. Introduction

Let M be a compact connected manifold and consider a C^2 Lagrangian function

$$L : TM \times \mathbb{R} \to \mathbb{R}$$

that satisfies the standard hypotheses of the calculus of variations,

$$L(x, v, t+1) = L(x, v, t) \quad \text{on} \quad TM \times \mathbb{R}, \tag{L1}$$

$$\partial^2_{vv} L(x, v, t) > 0 \quad \text{on} \quad TM \times \mathbb{R}, \tag{L2}$$

$$\lim_{\|v\| \to \infty} L(x, v, t)/\|v\| = \infty \quad \text{on} \quad M \times \mathbb{R}. \tag{L3}$$

It is standard that, under these assumptions, there exists a well-defined time-periodic continuous vectorfield $E(x, v, t)$ on TM such that the integral curves of E satisfy the Euler-Lagrange equations associate to L. We assume in addition that this vectorfield generates a complete flow, and denote by φ the time-one flow, which is a diffeomorphism of TM.

Received November 21, 2005.
Revised March 8, 2006.

In this paper we show that the theory developed by Mather [11], Mañé [14] and Fathi [10] amounts for a large part to the analysis of the function $A : M \times M \to \mathbb{R}$ defined by the expression

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

where the minimum is taken on the set of C^2 curves $\gamma : [0, 1] \to M$ which satisfy $\gamma(0) = x$ and $\gamma(1) = y$.

To emphasize this point of view, we develop an abstract theory based solely on an arbitrary continuous function $A : M \times M \to \mathbb{R}$, where M is a connected compact metric space. We then define $A_1 = A$ and

$$A_n(x, y) = \min_{z_1,\dots,z_{n-1}\in M} A(x, z_1) + A(z_1, z_2) + \dots + A(z_{n-1}, y)$$

for all integers $n \geqslant 2$. It turns out that the family (A_n) is equicontinuous and our only hypothesis on A is that the family (A_n) is uniformly bounded (this can be achieved by adding some constant to A). It then follows that the expression

$$c(x, y) = \liminf_{n\to\infty} A_n(x, y)$$

defines a continuous function $c : M \times M \to \mathbb{R}$.

We call (ϕ_0, ϕ_1) an *admissible Kantorovich pair* for c if

$$\forall y \in M \quad \phi_1(y) = \min_{x\in M} \phi_0(x) + c(x, y)$$

and

$$\forall x \in M \quad \phi_0(x) = \max_{y\in M} \phi_1(y) - c(x, y).$$

The first main result (Theorem 12) states that (ϕ_0, ϕ_1) is an admissible Kantorovich pair for c if and only if

- $\phi_0(x) = \max_{y\in M} \phi_0(y) - A(x, y)$ for all $x \in M$,
- $\phi_1(x) = \min_{y\in M} \phi_1(y) + A(y, x)$ for all $x \in M$,
- and $\phi_0(x) = \phi_1(x)$ whenever $c(x, x) = 0$.

The second main result (Theorem 13) concerns the minimization problem

$$\min_{\eta} \int_{M\times M} A(x, y)d\eta(x, y),$$

where the minimum is taken on the set of Borel probability measures η on $M \times M$ with equal marginal measures, that is, $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$ with π_0 and π_1 denoting the canonical projections on M. Among all admissible measures, the minimizing ones are shown to be exactly those supported on the set

$$D = \{(x, y) \in M \times M \mid A(x, y) + c(y, x) = 0\}.$$

This is also restated in the following way in Theorem 15. Let $X = M^{\mathbb{Z}}$ be endowed with the product topology and denote by $\mathcal{M}_T(X)$ the set of Borelian probability measures on X which are invariant by translation. Consider the minimization problem

$$\min_{\nu \in \mathcal{M}_T(X)} \int_X A(x_0, x_1) d\nu(x),$$

where a generic $x \in X$ is written $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$. Then we show with the help of the Ergodic Decomposition Theorem that ν in $\mathcal{M}_T(X)$ is minimizing exactly when the push-forward of ν by the projection $x \to (x_0, x_1)$ is concentrated on D.

The paper ends with the interpretation of these abstract theorems in the setting of the Aubry-Mather theory, recovering in this way some key results of [14, 11, 10].

§2. Monge-Kantorovich theory

We present some standard facts of Monge-Kantorovich theory, first in the general case, and then when the cost satisfies some given assumptions.

2.1. Generalities

We recall the basics of Monge-Kantorovich duality. The proofs are available in many texts on the subjects, for example [1, 15, 16]. We assume that M and N are compact metric spaces, and that $c(x, y)$ is a continuous cost function on $M \times N$. Given Borel probability measures μ_0 on M and μ_1 on N, a transport plan between μ_0 and μ_1 is a measure on $M \times N$ which satisfies

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

where $\pi_0 : M \times N \to M$ is the projection on the first factor, and $\pi_1 : M \times N \to N$ is the projection on the second factor. We denote by $\mathcal{K}(\mu_0, \mu_1)$, after Kantorovich, the set of transport plans. Kantorovich proved the existence of a minimum in the expression

$$C(\mu_0, \mu_1) = \min_{\eta \in \mathcal{K}(\mu_0, \mu_1)} \int_{M \times N} c d\eta \tag{1}$$

for each pair (μ_0, μ_1) of probability measures. The plans which realize this minimum are called optimal transfer plans. Let ϕ_0 be a real function on M and ϕ_1 a real function on N. The pair (ϕ_0, ϕ_1) is called an admissible Kantorovich pair if it satisfies the relations

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

for all point $x \in M$ and $y \in N$. Another discovery of Kantorovich is that

$$C(\mu_0, \mu_1) = \max_{\phi_0, \phi_1} \Big(\int_N \phi_1 d\mu_1 - \int_M \phi_0 d\mu_0 \Big) \tag{2}$$

where the maximum is taken on the non-empty set of admissible Kantorovich pairs (ϕ_0, ϕ_1). This maximization problem is called the dual Kantorovich problem, the admissible pairs which reach this maximum are called optimal Kantorovich pairs. The direct problem (1) and dual problem (2) are related as follows.

Proposition 1. *If η is an optimal transfer plan, and if (ϕ_0, ϕ_1) is a Kantorovich optimal pair, then the support of η is contained in the set*

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

which is a closed subset of $M \times N$ because ϕ_0 and ϕ_1 are continuous.

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

Lemma 2. *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 Kantorovich admissible pairs.

Proof. This Lemma is elementary and can be proved by easy manipulation of inequalities, see [4]. However, we present a short proof based on the non-elementary Monge-Kantorovich duality. Let us fix points $x \in M$ and $y \in N$, and let μ_0 be the Dirac measure at x and μ_1 be the Dirac measure at y. There exists one and only one transport plan between μ_0 and μ_1, it is the Dirac measure at (x,y). As a consequence, we have $c(x,y) = C(\mu_0, \mu_1)$. Hence the equality above is precisely the conclusion of Kantorovich duality for the transportation problem between μ_0 and μ_1. □

Proposition 3. *Let (ϕ_0, ϕ_1) be an admissible pair, and let μ_0 be a probability measure on M. Then there exists a probability measure μ_1 on N such that the pair (ϕ_0, ϕ_1) is optimal for the transportation problem of the measure μ_0 onto the measure μ_1*

Proof. If μ_0 is the Dirac at x, then take a point y such that $\phi_1(y) = \phi_0(x) + c(x, y)$, and observe that the conclusion obviously holds if μ_1 is the Dirac at y. The set of measures μ_0 for which the conclusion holds (given ϕ_0, ϕ_1) is clearly convex and closed (with respect to the weak topology), it contains the Dirac measures, hence it is the whole set of probability measures. □

2.2. Distance-like costs

Kantorovich stated his duality theorem first in the case where $M = N$ and the cost is a distance. Then, the dual problem takes a simpler form that we now describe. In fact, it is not necessary to assume that the cost is a distance. It is sufficient to assume that, for all x, y and z in M, we have

$$c(x, z) \leqslant c(x, y) + c(y, z), \tag{C1}$$

$$c(x, x) = 0. \tag{C2}$$

A function $\phi : M \to \mathbb{R}$ is called c-Lipschitz if it satisfies the inequality

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

for all x and y in M. Note that, in the above and in what follows, we assume that $M = N$ is a compact and connected metric space, and that $c : M \times M \to \mathbb{R}$ is a continuous cost function.

Theorem 4. *Assume that the cost $c \in C(M^2, \mathbb{R})$ satisfies the assumptions $(C1)$ and $(C2)$. Then for each pair μ_0, μ_1 of probability measures on M, we have*

$$C(\mu_0, \mu_1) = \max_{\phi} \int_M \phi d(\mu_1 - \mu_0)$$

where the maximum is taken on the set of c-Lipschitz functions ϕ.

This is a well-known direct rewriting of Kantorovich duality in view of the following description of admissible pairs.

Lemma 5. *If the cost satisfies $(C1)$ and $(C2)$, then the Kantorovich admissible pairs are precisely the pairs of the form (ϕ, ϕ), with ϕ c-Lipschitz.*

Proof. If ϕ is a c-Lipschitz function, then (ϕ, ϕ) is an admissible pair. Indeed, let us prove for example that $\phi(x) = \min_y \phi(y) + c(y, x)$. On the one hand, we have $\phi(x) \leqslant \phi(y) + c(y, x)$ because ϕ is c-Lipschitz, hence $\phi(x) \leqslant \min_y \phi(y) + c(y, x)$. On the other hand, $\phi(x) = \phi(x) + c(x, x) \geqslant \min_y \phi(y) + c(y, x)$. One can prove similarly that $\phi(x) = \max_y \phi(y) - c(x, y)$. It follows that (ϕ, ϕ) is an admissible pair. Conversely, if (ϕ_0, ϕ_1) is an admissible pair, then $\phi_0 = \phi_1$ is a c-Lipschitz function. This is a special case of Lemma 6 below. □

Let us now study costs which satisfy $(C1)$ but not necessarily $(C2)$. It is then useful to define the set

$$\mathcal{A} := \{x \in M, c(x, x) = 0\} \subset M.$$

Note that the restriction of the cost c to $\mathcal{A} \times \mathcal{A}$ obviously satisfies $(C1)$ and $(C2)$. In this more general case, we have:

Lemma 6. *Let $c \in C(M^2, \mathbb{R})$ satisfy $(C1)$. Let (ϕ_0, ϕ_1) be an admissible pair. Then the functions ϕ_0 and ϕ_1 are c-Lipschitz. In addition, we have $\phi_0 \leqslant \phi_1$ with equality on $\mathcal{A}$.*

Proof. Let us first prove that the function ϕ_1 is c-Lipschitz. Given $x \in M$, there exists y such that $\phi_1(x) = \phi_0(y) + c(y, x)$, and then, for each z,

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

One can prove similarly that ϕ_0 is c-Lipschitz.

We then have

$$\phi_0(x) = \max_y \phi_1(y) - c(x, y) \leqslant \max_y \phi_1(x) = \phi_1(x).$$

because ϕ_1 is c-Lipschitz. If $x \in \mathcal{A}$, we have, in addition,

$$\phi_0(x) = \max_y \phi_1(y) - c(x, y) \geqslant \phi_1(x) - c(x, x) = \phi_1(x).$$

□

We now introduce another hypothesis which is certainly less natural than $(C1)$ and $(C2)$, but is useful for the applications we have in mind. We assume that

(C3) $$\mathcal{A} \neq \emptyset \quad \text{and} \quad c(x, y) = \min_{a \in \mathcal{A}} c(x, a) + c(a, y)$$

for each x and y in M. Note that, under $(C1)$, $(C3)$ is implied by $(C2)$. The hypothesis $(C3)$ implies that each optimal transport can be factored through the set $\mathcal{A}$.

Lemma 7. *If the cost satisfies* $(C1)$ *and* $(C3)$, *then for each pair* (μ_0, μ_1) *of probability measures, there exists a probability measure* μ *supported on* $\mathcal{A}$ *and such that*

$$C(\mu_0, \mu_1) = C(\mu_0, \mu) + C(\mu, \mu_1)$$

Proof. First note that $C(\mu_0, \mu_1) \leqslant C(\mu_0, \mu) + C(\mu, \mu_1)$ is true for all Borelian probability measures μ on M. This can be seen as follows. Let η_0 and η_1 be optimal transport plans for (μ_0, μ) and (μ, μ_1) respectively. Disintegrate η_0 with respect to π_1 and η_1 with respect to π_0: $\eta_0 = \int_M \eta_{0z} d\mu(z)$ and $\eta_1 = \int_M \eta_{1z} d\mu(z)$ (see e.g. Theorem 5.3.1 in [2] for the disintegration theorem; here η_{0z} and η_{1z} are seen as probability measures on M). Following Section 5.3 in [2], define the probability measure η on M^2 by

$$\eta(A \times B) = \int_M \eta_{0z}(A)\eta_{1z}(B)\, d\mu(z)$$

for all Borelian subsets $A, B \subset M$. Then $\eta \in \mathcal{K}(\mu_0, \mu_1)$ and

$$\int_{M^2} c\, d\eta = \int_{M^3} c(x,y) d\eta_{0z}(x) d\eta_{1z}(y) d\mu(z)$$
$$\leqslant \int_{M^3} \{c(x,z) + c(z,y)\} d\eta_{0z}(x) d\eta_{1z}(y) d\mu(z) = \int_{M^2} c\, d\eta_0 + \int_{M^2} c\, d\eta_1.$$

Let us now prove the reverse inequality when μ_0 and μ_1 are Dirac measures supported in x and y. In this case, one can take for μ the Dirac measure supported at a, where a is any point such that $c(x,y) = c(x,a) + c(a,y)$. The general case is then deduced once again using the fact that, on M^2, the set of probability measures is the closed convex envelop of the set of Dirac measures, so that we can approximate any optimal transfer plan in $\mathcal{K}(\mu_0, \mu_1)$ by Dirac measures. □

Proposition 8. *If the cost* $c \in C(M^2, \mathbb{R})$ *satisfies* $(C1)$ *and* $(C3)$, *then for each admissible pair* (ϕ_0, ϕ_1), *there exists a function* $\phi : \mathcal{A} \to \mathbb{R}$, *which is c-Lipschitz, and such that*

$$\phi_1(a) = \phi_0(a) = \phi(a)$$

for all $a \in \mathcal{A}$,

$$\phi_1(x) = \min_{a \in \mathcal{A}} \phi(a) + c(a,x) \tag{3}$$

for all $x \in M$ *and*

$$\phi_0(x) = \max_{a \in \mathcal{A}} \phi(a) - c(x,a). \tag{4}$$

Conversely, given any c-Lipschitz function ϕ on $\mathcal{A}$, the functions ϕ_0 and ϕ_1 defined by (4) and (3) form an admissible pair. In other words, there is a bijection between the set of admissible pairs and the set of c-Lipschitz functions on $\mathcal{A}$.

Proof. The fact that ϕ_0 and ϕ_1 are c-Lipschitz and that, on $\mathcal{A}$, $\phi_0 = \phi_1 := \phi$ results from Lemma 6. Let us prove (4), the proof of (3) being similar:

$$\phi_0(x) = \max_y \phi_1(y) - c(x,y) \overset{(C3)}{=} \max_{y\in M, a\in\mathcal{A}} \phi_1(y) - c(x,a) - c(a,y)$$

$$= \max_{a\in\mathcal{A}} \phi_0(a) - c(x,a) = \max_{a\in\mathcal{A}} \phi(a) - c(x,a).$$

Conversely, let ϕ be a c-Lipschitz function on $\mathcal{A}$, and let ϕ_0 and ϕ_1 be defined by (4) and (3). The reader will easily check that ϕ_0 and ϕ_1 are c-Lipschitz, and that $\phi_1 \leqslant \phi \leqslant \phi_0$ on $\mathcal{A}$. We now prove that $\phi_0 \leqslant \phi_1$ (and then that there is equality on $\mathcal{A}$):

$$\phi_0(x) - \phi_1(x) = \max_{a,b\in\mathcal{A}} \phi(a) - c(x,a) - \phi(b) - c(b,x)$$

$$\leqslant \max_{a,b\in\mathcal{A}} \phi(a) - \phi(b) - c(b,a) \leqslant 0$$

because ϕ is c-Lipschitz on $\mathcal{A}$. In order to check that the pair (ϕ_0, ϕ_1) is an admissible pair, we shall prove that

$$\phi_0(x) = \max_y \phi_1(y) - c(x,y)$$

and leave the other half to the reader. For each x in M, we have

$$\phi_0(x) = \max_{a\in\mathcal{A}} \phi(a) - c(x,a) = \max_{a\in\mathcal{A}} \phi_1(a) - c(x,a) \leqslant \max_{y\in M} \phi_1(y) - c(x,y).$$

In order to obtain the other inequality, let us prove that

$$\phi_1(y) - \phi_0(x) \leqslant c(x,y)$$

for all x an y in M. Indeed, we have

$$\phi_1(y) - \phi_0(x) = \min_{a,b\in\mathcal{A}} \phi(a) + c(a,y) - \phi(b) + c(x,b)$$

$$\leqslant \min_{a,b\in\mathcal{A}} c(b,a) + c(a,y) + c(x,b) = \min_{a\in\mathcal{A}} c(x,a) + c(a,y) = c(x,y)$$

by $(C3)$. □

Since ϕ_1 and ϕ_2 are c-Lipschitz (Lemma 6), equations (3) and (4) imply

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

for all $x \in M$.

§3. Abstract Mather-Fathi Theory

In this section, we consider a continuous function $A(x,y) : M \times M \to \mathbb{R}$. Recall that M is a compact connected metric space. We shall build several functions out of A. First, we define the sequence of functions $A_n(x,y)$ by setting $A_1 = A$ and

$$\begin{aligned} A_n(x,y) &= \min_{z \in M} A(x,z) + A_{n-1}(z,y) \\ &= \min_{z_1,\dots,z_{n-1} \in M} A(x,z_1) + A(z_1,z_2) + \ldots + A(z_{n-1},y). \end{aligned}$$

Lemma 9. *The functions A_n are equicontinuous. In addition, there exists a real number l and a positive constant C such that*

$$|A_n(x,y) - ln| \leqslant C$$

for all $n \in \mathbb{N}$ and all x and y in M.

Proof. The function A is continuous, hence uniformly continuous, hence there exists a modulus of continuity $\delta : [0,\infty) \to [0,\infty)$ such that $\lim_{\epsilon \to 0} \delta(\epsilon) = \delta(0) = 0$ and such that

$$|A(x,y) - A(X,Y)| \leqslant \delta(d(x,X)) + \delta(d(y,Y))$$

for all x, y, X, Y in M. Clearly, for all $n \geqslant 2$ and all $z_1, \dots, z_{n-1} \in M$, the function $(x,y) \to A(x,z_1) + A(z_1,z_2) + \ldots + A(z_{n-1},y)$ is uniformly continuous, with the same modulus of continuity as A. Hence A_n is uniformly continuous with the same modulus of continuity as A because it is the infimum of functions having all the same modulus of continuity.

Let us define the sequences $M_n := \max_{(x,y) \in M^2} A_n(x,y)$ and $m_n := \min_{(x,y) \in M^2} A_n(x,y)$. It is clear that the sequence M_n is subadditive, *i. e.* that $M_{n+k} \leqslant M_n + M_k$ for all n and k in $\mathbb{N}$. In order to check this claim, we take x and y in M such that $A_{n+k}(x,y) = M_{n+k}$. Then there exists a point z in M such that

$$M_{n+k} = A_{n+k}(x,y) = A_n(x,z) + A_k(z,y) \leqslant M_n + M_k.$$

Similarly, the sequence m_n is super-additive, *i. e.* $m_{n+k} \geqslant m_n + m_k$. On the other hand, on view of the equicontinuity of A_n, there exists a constant C such that $M_n - m_n \leqslant C$. Applying a standard result on subadditive sequences (see e.g. Lemma 1.18 in [5]), we obtain that M_n/n converges to its infimum M, and that m_n/n converges to its supremum m. Then for each x and y,

$$nM - C \leqslant M_n - C \leqslant m_n \leqslant A_n(x,y) \leqslant M_n \leqslant m_n + C \leqslant nm + C$$

which implies that $M = m$ and proves the Lemma. □

We make, on the function A, the hypothesis

$$l = 0. \tag{A1}$$

Note that this hypothesis implies that $A(x,x) \geqslant 0$ for all x, and more generally that $A_n(x,x) \geqslant 0$ for all x. Then, we can define a cost function c by the expression

$$c(x,y) = \liminf_{n\to\infty} A_n(x,y). \tag{5}$$

In view of Lemma 9, the function c takes finite values and is continuous. We have $c(x,x) \geqslant 0$ and, by Lemma 11 below, $c(x,y) + c(y,x) \geqslant c(x,x) \geqslant 0$ for all x and y in M.

Lemma 10. *For each $n \in \mathbb{N}$, we have*

$$c(x,y) = \min_{z\in M} c(x,z) + A_n(z,y) = \min_{z\in M} A_n(x,z) + c(z,y).$$

Proof. Let us fix n. Passing at the liminf ($m \to \infty$) in the inequality

$$A_{m+n}(x,y) \leqslant A_m(x,z) + A_n(z,y),$$

we obtain

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

For the opposite inequality, let us notice that, for each m, there exists a point z_m in M such that

$$A_{m+n}(x,y) = A_m(x,z_m) + A_n(z_m,y).$$

Let us consider an increasing sequence of integers m_k such that the subsequence z_{m_k} has a limit z and $\lim_{k\to\infty} A_{m_k+n}(x,y) = c(x,y)$. At the liminf, we get, taking advantage of the equicontinuity of the functions A_n,

$$c(x,y) \geqslant c(x,z) + A_n(z,y).$$

This proves that

$$c(x,y) = \min_z c(x,z) + A_n(z,y).$$

The proof of the second equality of the statement is similar. □

Lemma 11. *The cost function c satisfies $(C1)$ and $(C3)$.*

Proof. The triangle inequality is easily deduced from Lemma 10. Let us now prove $(C3)$. We first prove that, given x and y in M, there exists a point z in M such that $c(x,y) = c(x,z) + c(z,y)$. Indeed, for each n in $\mathbb{N}$, there exists a point z_n such that $c(x,y) = c(x,z_n) + A_n(z_n,y)$. Considering an increasing sequence of integers n_k such that the subsequence z_{n_k} has a limit z, we obtain at the liminf along this subsequence that $c(x,y) \geqslant c(x,z) + c(z,y)$ which is then an equality.

By recurrence, there exists a sequence $Z_n \in M$ such that, for each $k \in \mathbb{N}$, we have

$$c(x,y) = c(x,Z_1) + c(Z_1,Z_2) + \ldots + c(Z_{k-1},Z_k) + c(Z_k,y).$$

Note that $\sum_{i=\ell}^{m} c(Z_i,Z_{i+1}) = c(Z_\ell,Z_{m+1})$ if $0 \leqslant \ell < m \leqslant k$, where $Z_0 = x$ and $Z_{k+1} = y$.

Let Z be an accumulation point of the sequence Z_n. For each $\epsilon > 0$, we can suppose, by taking a subsequence in Z_n, that all the points Z_n belong to the ball of radius ϵ centered at Z. We conclude that, for each $k \in \mathbb{N}$,

$$c(x,y) \geqslant c(x,Z) + (k-1)c(Z,Z) + c(Z,y) - 2(k+1)\delta(\epsilon).$$

This is possible only if $c(Z,Z) \leqslant 2\delta(\epsilon)$, and since this should hold for all ϵ we conclude that $c(Z,Z) \leqslant 0$, hence $c(Z,Z) = 0$. We have proved the existence of a point $Z \in \mathcal{A}$ such that $c(x,y) = c(x,Z) + c(Z,y)$. □

Let us define, the two operators $T^{\pm}$ on the space $C(M,\mathbb{R})$ of continuous functions on M by the expressions

$$T^- u(x) = \min_{y \in M} u(y) + A(y,x)$$

and

$$T^+ u(x) = \max_{y \in M} u(y) - A(x,y).$$

We have the following relation between the fixed points of these operators and the admissible pairs of the Kantorovich dual problem with cost c. Recall the definition $\mathcal{A} := \{x \in M, c(x,x) = 0\} \subset M$.

Theorem 12. *Let A be a function satisfying* (A1)*, and let c be the cost defined by (5). The pair (ϕ_0, ϕ_1) of functions on M is a Kantorovich admissible pair (for c) if and only if*

- *the function ϕ_0 is a fixed point of T^+,*
- *the function ϕ_1 is a fixed point of T^-,*
- *$\phi_0 = \phi_1$ on $\mathcal{A}$.*

Finally, for each fixed point ϕ_1 of T^-, there exists one and only one function ϕ_0 such that (ϕ_0, ϕ_1) is an admissible pair.

Proof. Let (ϕ_0, ϕ_1) be an admissible pair. Then we have the expression

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

We obtain that

$$T^-\phi_1(z) = \min_{x,y \in M} \phi_0(x) + c(x,y) + A(y,z)$$

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

We prove in the same way that the function ϕ_0 is a fixed point of T^+. Lemma 6 implies that $\phi_0 = \phi_1$ on $\mathcal{A}$.

Conversely, let (ϕ_0, ϕ_1) satisfy the three conditions of the statement. We first observe that the functions ϕ_0 and ϕ_1 are c-Lipschitz. Indeed, we have, for each n,

$$\phi_i(y) - \phi_i(x) \leqslant A_n(x,y).$$

When $n = 1$, this is a direct consequence of the fact that ϕ_i is a fixed point of $T^\pm$, and the general case is proved by induction. We get

$$\phi_i(y) - \phi_i(x) \leqslant \liminf_{n \to \infty} A_n(x,y) = c(x,y).$$

The function ϕ_1 being a fixed point of T^-, for each $n \in \mathbb{N}$, there exists a point y_n in M such that $\phi_1(x) = \phi_1(y_n) + A_n(y_n, x)$. Indeed, we can find successively $y_1, y_2, \ldots$ such that

$$\phi_1(x) = \phi_1(y_1) + A(y_1, x) = \phi_1(y_2) + A(y_2, y_1) + A(y_1, x)$$
$$= \ldots = \phi_1(y_n) + A(y_n, y_{n-1}) + \ldots + A(y_1, x).$$

By definition of A_n, we get $\phi_1(x) \geqslant \phi_1(y_n) + A_n(y_n, x)$. The reverse inequality has just been proved above.

Let n_k be a subsequence such that y_{n_k} has a limit y. At the limit, we obtain the inequality

$$\phi_1(x) \geqslant \phi_1(y) + c(y, x),$$

which is then an equality. We have proved that

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

Let us call ϕ the common value of ϕ_0 and ϕ_1 on $\mathcal{A}$. In view of $(C3)$, we have

$$\phi_1(x) = \min_{y \in M, a \in \mathcal{A}} \phi_1(y) + c(y, a) + c(a, x)$$

$$= \min_{a \in \mathcal{A}} \phi_1(a) + c(a, x) = \min_{a \in \mathcal{A}} \phi(a) + c(a, x).$$

One can prove in a similar way that

$$\phi_0(x) = \max_{a \in \mathcal{A}} \phi_0(a) - c(x, a) = \max_{a \in \mathcal{A}} \phi(a) - c(x, a).$$

We conclude that (ϕ_0, ϕ_1) is an admissible pair by Proposition 8. This also proves the uniqueness claim.

In order to prove the last part of the statement, let us consider a fixed point ϕ_1 of T^-. Let us define the function ϕ_0 by

$$\phi_0(x) = \max_{a \in \mathcal{A}} \phi_1(a) - c(x, a).$$

Since the function ϕ_1 is c-Lipschitz (as seen above), we have $\phi_0 \leqslant \phi_1$. On the other hand, it is clear that $\phi_1 \leqslant \phi_0$ on $\mathcal{A}$. As a consequence, we have $\phi_0 = \phi_1$ on $\mathcal{A}$. By Lemma 10, we have for all $z \in M$ that

$$\max_{x \in M} \phi_0(x) - A(z, x) = \max_{x \in M, a \in \mathcal{A}} \phi_1(a) - c(x, a) - A(z, x)$$
$$= \max_{a \in \mathcal{A}} \phi_1(a) - c(z, a) = \phi_0(z).$$

Hence the function ϕ_0 is a fixed point of T^+ and, as a consequence, the pair (ϕ_0, ϕ_1) is an admissible pair. □

§4. Dynamics

Let us define the subset

$$D := \{(x, y) \in M \times M \text{ s. t. } A(x, y) + c(y, x) = 0\} \subset \mathcal{A} \times \mathcal{A}$$

(see Lemma 10). We shall explain in two different ways that the Borel probability measures η on $M \times M$ which are supported on D and satisfy $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$ can be seen in a natural way as the analog of Mather minimizing measures in our setting.

4.1. Construction via Kantorovich pairs

We first expose a construction based on Kantorovich pairs.

Theorem 13. *Under the assumption* $(A1)$*, we have*

$$\min_{\eta} \int_{M \times M} A(x, y) d\eta(x, y) = 0,$$

where the minimum is taken on the set of Borel probability measures η *on* $M \times M$ *such that* $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$*. The minimizing measures are those which are supported on* D*.*

Proof. Let us first prove that there exists a measure η on $M \times M$ which is supported on D and such that $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$. By Lemma 10, for each $x_0 \in \mathcal{A}$, there exists a point x_1 in $\mathcal{A}$ such that $(x_0, x_1) \in D$. Hence there exists a sequence $x_0, x_1, x_2, \dots x_n, \dots$ of points of $\mathcal{A}$ such that $(x_n, x_{n+1}) \in D$ for each n. Let us now consider the sequence

$$\eta_n = \frac{\delta_{(x_0, x_1)} + \delta_{(x_1, x_2)} + \cdots + \delta_{(x_{n-1}, x_n)}}{n}$$

of probability measures on $\mathcal{A} \times \mathcal{A}$. Every accumulation point (for the weak topology) of the sequence η_n satisfies the desired property. Since the set of probability measures on $M \times M$ is compact for the weak topology, such accumulation points exist.

Consider a measure η on $M \times M$ which is supported on D and such that $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$. We have

$$\int A(x, y) d\eta(x, y) = \int -c(y, x) d\eta(y, x) \leqslant \int \phi(y) - \phi(x) d\eta(x, y) = 0,$$

where ϕ is any c-Lipschitz function.

On the other hand, let η be a probability measure on $M \times M$ such that $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$. Consider a function ϕ which is A-Lipschitz. Such functions exist, for example, take $z_2 \to c(z_1, z_2)$ for any $z_1 \in M$ (see Lemma 10) or fixed points of T^- or T^+. We have

$$0 = \int \phi(y) - \phi(x) d\eta(x,y) \leqslant \int A(x,y) d\eta(x,y). \tag{6}$$

We have proved that the minimum in the statement is indeed zero, and that the measures supported on D are minimizing. There remains to prove that every minimizing measure is supported on D.

It is clear that a measure η is minimizing if and only if, for each A-Lipschitz function ϕ, there is equality in (6), which means that the measure η is supported on the set

$$D_1 = \{(x,y) \in M^2 \mid \phi(y) - \phi(x) = A(x,y) \text{ for all } A\text{-Lipschitz functions } \phi\}.$$

Let D_∞ be the set of pairs (x_0, x_1) such that there exists a sequence x_i, $i \in \mathbb{Z}$ satisfying $(x_i, x_{i+1}) \in D_1$ for all $i \in \mathbb{Z}$ (and of course with the given points x_0 and x_1).

We claim that $D_\infty \subset D$. In order to prove this claim, let ϕ be A-Lipschitz. Observe that ϕ is A_n-Lipschitz for all $n \in \mathbb{N}$ and c-Lipschitz. If (x_0, x_1) is a point in D_∞, then there exists a sequence x_i, $i \in \mathbb{Z}$ such that

$$\phi(x_j) - \phi(x_i) = A_{j-i}(x_i, x_j)$$

for each $i < j$ in $\mathbb{Z}$. If α is an accumulation point of the sequence x_i at $-\infty$, we get the equality

$$\phi(x_j) - \phi(\alpha) = c(\alpha, x_j)$$

for each $j \in \mathbb{Z}$ and then, in the same way, $c(\alpha, \alpha) = \phi(\alpha) - \phi(\alpha) = 0$, hence $\alpha \in \mathcal{A}$. Let (ϕ_0, ϕ_1) be a Kantorovich pair for c, so that both ϕ_0 and ϕ_1 are A-Lipschitz (see Theorem 12). We get $\phi_1(\alpha) = \phi_0(\alpha)$ (because $\alpha \in \mathcal{A}$, see Theorem 12) hence $\phi_1(x_j) = \phi_0(x_j)$. Since this holds for all Kantorovich pairs, we get that $x_j \in \mathcal{A}$ (see Lemma 2). In other words, we have proved that $D_\infty \subset \mathcal{A} \times \mathcal{A}$. Now let (x_0, x_1) be a point of D_∞. We have $x_1 \in \mathcal{A}$, and, since the function $c(x_1, .)$ is A-Lipschitz, we have the equality $c(x_1, x_1) - c(x_1, x_0) = A(x_0, x_1)$. Recalling that $c(x_1, x_1) = 0$, we get $c(x_1, x_0) + A(x_0, x_1) = 0$, hence $(x_0, x_1) \in D$. The proof of the Theorem then follows from the next Lemma. □

Lemma 14. *If η is a probability measure on $M \times M$ which is supported on D_1 and such that $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$, then η is concentrated on D_∞.*

Proof. Let us set $\mu = \pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$ and let

$$X_1 = \pi_0(D_1) \cap \pi_1(D_1) \subset M$$

be the set of points $x_0 \in M$ such that a sequence x_{-1}, x_0, x_1 exists, with $(x_{-1}, x_0) \in D_1$ and $(x_0, x_1) \in D_1$. Clearly, we have $\mu(\pi_0(D_1)) = \mu(\pi_1(D_1)) = 1$ hence $\mu(X_1) = 1$. Let

$$D_2 = D_1 \cap (X_1 \times X_1)$$

be the set of pairs $(x_0, x_1) \in M^2$ such that there exist x_{-1}, x_0, x_1, x_2 with $(x_i, x_{i+1}) \in D_1$ for $i = -1, 0, 1$. Let

$$X_2 = \pi_0(D_2) \cap \pi_1(D_2) \subset M$$

be the set of points $x_0 \in M$ such that a sequence $x_{-2}, x_{-1}, x_0, x_1, x_2$ exists, with $(x_i, x_{i+1}) \in D_1$ for all $-2 \leqslant i \leqslant 1$. Since $\mu(X_1) = 1$, we have $\eta(D_2) = 1$, hence $\mu(X_2) = 1$. By recurrence, we build a sequences $D_n \subset M \times M$ and $X_n \subset M$ such that

$$D_n = D_1 \cap (X_{n-1} \times X_{n-1})$$

and

$$X_n = \pi_0(D_n) \cap \pi_1(D_n) \subset M.$$

By recurrence, we see that $\eta(D_n) = 1$ and that $\mu(X_n) = 1$. Now we have

$$D_\infty = \bigcap_{n \in \mathbb{Z}} D_n$$

hence $\eta(D_\infty) = 1$. □

4.2. Ergodic Construction

It is worth explaining that the preceding construction could have been performed in a quite different way, which does not use our theory of Kantorovich pairs, but relies on Ergodic theory, as the first papers of Mather.

Consider $X = M^{\mathbb{Z}}$ endowed with the product topology, so that X is a metrizable compact space. We shall denote by $\mathcal{M}_T(X)$ the set of

Borelian probability measures on X which are invariant by translation. More precisely, we denote by $T : X \to X$ the translation map

$$T(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots) = (\ldots, b_{-2}, b_{-1}, b_0, b_1, b_2, \ldots)$$

with $b_i = a_{i+1}$ for all $i \in \mathbb{Z}$, so that $\mathcal{M}_T(X)$ is the set of probability measures ν on X such that $T_\sharp \nu = \nu$.

Theorem 15. *We have*

$$\min_{\nu \in \mathcal{M}_T(X)} \int_X A(x_0, x_1)\, d\nu(x) = 0.$$

The measure ν is minimizing if and only if its marginal $\eta = (\pi_0 \times \pi_1)_\sharp \nu$ is concentrated on D.

Note that Theorem 15 is equivalent to Theorem 13 in view of the following:

Lemma 16. *Let η be a Borelian probability measure on M^2 such that $\pi_{0\sharp}(\eta) = \pi_{1\sharp}(\eta)$. Then there exists a Borelian measure ν on X that is T-invariant and such that η is its push-forward by the map $X \ni x \to (x_0, x_1) \in M^2$.*

Proof. This follows from the Hahn-Kolmogorov extension theorem (see e.g. Theorem 0.1.5 in [12], Lemma 10.2.4 in [7] and Theorem 12.1.2 in [7]). Let Ω be the algebra of finite unions of subsets G of X of the type $G = \Pi_{i \in \mathbb{Z}} G_i$ where $G_i \neq M$ for at most a finite number of indices i (the number depending on G) and every G_i is a Borelian subset of M. We first define the T-invariant probability measure ν on Ω and then apply the Hahn-Kolmogorov extension theorem, which provides an unique extension to the Borel σ-algebra (by uniqueness, the extension is T-invariant).

Let $\eta = \int_M \eta_{x_1}\, d\mu(x_1)$ be the disintegration of η with respect to the projection $M^2 \ni (x_0, x_1) \to x_1 \in M$. In particular $\mu = \pi_{1\sharp}(\eta)$ (see e.g. Theorem 5.3.1 in [2] for the disintegration theorem). Define for $m < n$

$$\begin{aligned} &\nu(\ldots \times M \times M \times G_m \times \ldots \times G_n \times M \times M \times \ldots) \\ = &\int_{G_m \times \ldots \times G_n} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) d\mu(x_n). \end{aligned}$$

This is well defined because if $G_{m-1} = M$ then

$$\begin{aligned} &\int_{G_{m-1} \times G_m \times \ldots \times G_n} d\eta_{x_m}(x_{m-1}) d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) d\mu(x_n) \\ = &\int_{G_m \times \ldots \times G_n} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) d\mu(x_n) \end{aligned}$$

and if $G_{n+1} = M$ then

$$= \int_{G_m \times \ldots \times G_n \times G_{n+1}} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) d\eta_{x_{n+1}}(x_n) d\mu(x_{n+1})$$

$$= \int_{G_n \times M} \left\{ \int_{G_m \times \ldots \times G_{n-1}} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) \right\} d\eta_{x_{n+1}}(x_n) d\mu(x_{n+1})$$

$$= \int_{G_n \times M} \left\{ \int_{G_m \times \ldots \times G_{n-1}} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) \right\} d\eta(x_n, x_{n+1})$$

$$= \int_{G_n} \left\{ \int_{G_m \times \ldots \times G_{n-1}} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) \right\} d\mu(x_n)$$

$$= \int_{G_m \times \ldots \times G_n} d\eta_{x_{m+1}}(x_m) \ldots d\eta_{x_n}(x_{n-1}) d\mu(x_n)$$

because $\mu = \pi_{1\sharp}(\eta) = \pi_{0\sharp}(\eta)$.

Clearly $\nu(X) = 1$ and ν is T-invariant on Ω. □

Although we have proved the equivalence between Theorem 15 and Theorem 13 we shall, as announced, detail another proof of Theorem 15.

For $x \in X$ and every Borelian subset B, we define

$$\tau_B(x) = \lim_{n \to +\infty} \frac{1}{n} \operatorname{card}\{0 \leqslant j \leqslant n-1 \,|\, T^j(x) \in B\}$$

(when the notation is used, it is understood that the limit exists). A Borelian probability ν on X is ergodic if and only if, for every Borelian subset $B \subset X$, there holds $\tau_B(x) = \nu(B)$ ν-almost surely.

Following Section II.6 in the book by Mañé [12], there exists a Borel set $\Sigma \subset X$ such that $\nu(\Sigma) = 1$ for each $\nu \in \mathcal{M}_T(X)$, and, for each $x \in \Sigma$, the measure

$$\nu_x := \lim_{n \to +\infty} \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j(x)}$$

is well defined and ergodic, where the limit is understood in the sense of the weak topology, that is

$$\forall f \in C(X, \mathbb{R}) \;\; \int_X f \, d\nu_x = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)). \tag{7}$$

Moreover $\nu_x \in \mathcal{M}_T(X)$ and x belongs to the support of ν_x for all $x \in \Sigma$. In addition, still following [12], we have that the function $x \longmapsto \int f d\nu_x$

is ν-integrable and T-invariant, and that

$$\int_X \left(\int_X f d\nu_x \right) d\nu = \int_X f\, d\nu. \tag{8}$$

holds for every $f \in \mathcal{L}^1(X,\nu)$. Note that the measure ν_x is the conditional probability measure of ν with respect to the σ-algebra of T-invariant Borel sets.

We define the continuous function $\Gamma : X \to \mathbb{R}$ by $\Gamma(x) = A(x_0, x_1)$. By standard convexity arguments, the following minimum is reached:

$$\alpha = \min_{\nu \in \mathcal{M}_T(X)} \int_X \Gamma(x) d\nu(x).$$

Let us prove that $\alpha \leqslant 0$. Fix $x_0 \in M$. For all $\epsilon > 0$, we can find $n \geqslant 1$ and $x_1, \ldots, x_n \in M$ such that

$$x_n = x_0 \quad \text{and} \quad \frac{1}{n}\sum_{j=0}^{n-1} A(x_j, x_{j+1}) < \epsilon$$

(thanks to assumption (A1)). Let $x = (\ldots, x_0, \ldots, x_n, \ldots) \in X$ have periodic components with period n and define $\nu \in \mathcal{M}_T(X)$ by

$$\nu = \frac{1}{n}\sum_{j=0}^{n-1} \delta_{T^j(x)}$$

where $\delta_{T^j(x)}$ is the Dirac measure at $T^j(x)$. Then $\int_X \Gamma\, d\nu < \epsilon$, which proves that $\alpha \leqslant 0$ (because ϵ can be chosen arbitrarily small).

Let $\nu \in \mathcal{M}_T(X)$ be any optimal measure. The equality

$$\int_X \left(\int_X \Gamma d\nu_x \right) d\nu = \int_X \Gamma\, d\nu = \alpha$$

shows that $\int_X \Gamma d\nu_x = \alpha$ for ν-almost all $x \in \Sigma$. For such a x, we get

$$0 \geqslant \alpha = \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} \Gamma(T^j(x)) = \lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} A(x_j, x_{j+1}). \tag{9}$$

Assume for a while that $x_0 \notin \mathcal{A}$. Then there exists a neighborhood U of x_0 in M, $\delta > 0$ and $N \geqslant 1$ such that

$$A_n(y_0, z_0) > \delta > 0 \text{ for all } y_0, z_0 \in U \text{ and } n \geqslant N \tag{10}$$

(we use here the equicontinuity of the functions A_n). Setting $\widetilde{U} = \{y \in X \mid y_0 \in U\}$, we get

$$0 < \nu_x(\widetilde{U}) \leqslant \liminf_{n\to+\infty} \frac{1}{n} \operatorname{card}\{0 \leqslant j \leqslant n-1 \mid x_j \in U\}.$$

The first inequality is a consequence of the fact that x is in the support of ν_x and the second one follows from (7) and the fact that the characteristic function of U is the supremum of an increasing sequence of continuous functions. We denote by $(x_{j_k} : k \geqslant 0)$ the sequence of components of x in U (of non-negative index). We obtain (see (10))

$$0 < \nu_x(\widetilde{U}) \leqslant \liminf_{m\to+\infty} \frac{mN}{j_{mN}}$$

and the contradiction

$$\begin{aligned}
&\liminf_{m\to\infty} \frac{1}{j_{mN}} \sum_{j=0}^{j_{mN}-1} A(x_j, x_{j+1}) \\
&= \liminf_{m\to\infty} \frac{1}{j_{mN}} \sum_{k=0}^{m-1} \sum_{j=j_{kN}}^{j_{(k+1)N}-1} A(x_j, x_{j+1}) \\
&\geqslant \liminf_{m\to\infty} \frac{1}{j_{mN}} \sum_{k=0}^{m-1} A_{j_{(k+1)N}-j_{kN}}(x_{j_{kN}}, x_{j_{(k+1)N}}) \\
&\geqslant \liminf_{m\to\infty} \frac{m\delta}{j_{mN}} \geqslant \nu_x(\widetilde{U})\delta/N \;>\; 0
\end{aligned}$$

(compare with (9)). This contradiction shows that $x_0 \in \mathcal{A}$ for ν-almost all x, that is, the marginal $\mu = \pi_{0\sharp}\nu$ is concentrated on $\mathcal{A}$.

Let us now check that $\alpha \geqslant 0$. For contradiction, suppose $\alpha < 0$. Then for $x \in \Sigma$ as above such that $\nu_x \in \mathcal{M}_T(X)$ and $\Gamma(\nu_x) = \alpha$, we get

$$\begin{aligned}
0 \quad &> \quad \alpha = \Gamma(\nu_x) = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=0}^{n-1} A(x_j, x_{j+1}) \\
&= \quad \lim_{n\to+\infty} \frac{1}{n+1} \left(\sum_{j=0}^{n-1} A(x_j, x_{j+1}) + A(x_n, x_0) \right) \\
&\geqslant \quad \limsup_{n\to+\infty} \frac{1}{n+1} A_{n+1}(x_0, x_0).
\end{aligned}$$

This contradicts $l = 0$ (see hypothesis (A1)).

We have proved that $\alpha = 0$, and that every minimizing T-invariant measure ν has its marginal $\mu = \pi_{0\sharp}\nu$ concentrated on $\mathcal{A}$. Let us now prove that every minimizing measure $\nu \in \mathcal{M}_T(X)$ is supported on $\{x \in X \mid A(x_0, x_1) + c(x_1, x_0) = 0\}$. Let x belong to the support of ν and observe that (see Lemma 10) $c(x_1, y_1) \leqslant c(x_1, y_0) + A(y_0, y_1)$ for all $y_0, y_1 \in M$. Therefore

$$0 = \int_X c(x_1, y_1) - c(x_1, y_0)\, d\nu(y) \leqslant \int_X A(y_0, y_1)\, d\nu(y) = \alpha = 0$$

and

$$\int_X A(y_0, y_1) - c(x_1, y_1) + c(x_1, y_0)\, d\nu(y) = 0$$

where the integrand is non negative. Hence $c(x_1, y_1) = c(x_1, y_0) + A(y_0, y_1)$ for ν-almost all y. Since x is in the support of ν, we get $c(x_1, x_1) = c(x_1, x_0) + A(x_0, x_1)$. We have just seen that $y_0 \in \mathcal{A}$ for ν-almost all y. By the T-invariance of ν, we also have $y_1 \in \mathcal{A}$ for ν-almost all y. Since x is in the support of ν, we therefore obtain $x_1 \in \mathcal{A}$ and $0 = c(x_1, x_1) = c(x_1, x_0) + A(x_0, x_1)$.

Finally let $\nu \in \mathcal{M}_T(X)$ be concentrated on

$$\widetilde{D} = \{y \in X \mid A(y_0, y_1) + c(y_1, y_0) = 0\}$$

and let us prove that $\int_X \Gamma d\nu = \alpha$. By (8) applied to the characteristic function of $\widetilde{D}$, we get that $\nu_x(\widetilde{D}) = 1$ for ν-almost all $x \in \Sigma$. By (8) applied to Γ, we see that it suffices to check that $\int_X \Gamma d\nu_x = \alpha$ for all $x \in \Sigma$ such that ν_x is concentrated on $\widetilde{D}$. This follows from (C1):

$$\begin{aligned} 0 = \alpha \leqslant \int_X A(y_0, y_1)\, d\nu_x(y) &= -\int_X c(y_1, y_0)\, d\nu_x(y) \\ &= -\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} c(x_{j+1}, x_j) \leqslant -\liminf_{n\to\infty} \frac{1}{n} c(x_n, x_0) = 0. \end{aligned}$$

□

§5. Aubry-Mather theory

We now briefly explain the relations between our discussions and the literature on Aubry-Mather theory, and especially [11], [14] and [10]. From now on, the space M is a compact connected manifold and we consider a C^2 Lagrangian function $L : TM \times \mathbb{R} \to \mathbb{R}$ as in the Introduction.

In this context, we define $A : M \times M \to \mathbb{R}$ by

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

where the minimum is taken on the set of C^2 curves $\gamma : [0,1] \to M$ which satisfy $\gamma(0) = x$ and $\gamma(1) = y$.

The function c defined by (5) is one of the central objects of Mather's theory of globally minimizing orbits, see [11]. He called it the Peierls barrier. It contains most of the information concerning the globally minimizing orbits, as was explained by Mather, see also [3]. The set $\mathcal{A}$ of points $x \in M$ such that $c(x,x) = 0$ is called the projected Aubry set. It is especially important because Mather proved the existence of a vectorfield $X(x)$ on $\mathcal{A}$ whose graph is invariant under the Lagrangian flow φ. This invariant set is called the Aubry set. The analog of the Aubry set in our general theory is the set D defined in the beginning of section 4.

The operators $T^{\pm}$ have been introduced by Albert Fathi in this context, see [8],[9] and [10]. He called Weak KAM solutions the fixed points of T^-, and we call backward weak KAM solutions the fixed points of T^+. He also noticed that, for each weak KAM solution ϕ_1, there exists one and only one backward weak KAM solution ϕ_0 which is equal to ϕ_1 on the projected Aubry set. This is the main part of our Theorem 12. Albert Fathi also proved Lemma 2 in this context. Our novelty in these matters consists of pointing out and using the equivalence with Kantorovich admissible pairs, which allows, for example, a strikingly simple proof of the important result of Fathi called Lemma 2 in our paper. The representation of weak KAM solutions given in Proposition 8 was obtained by Contreras in [6].

The minimizing measures of Theorem 13 are the famous Mather measures, see [11]. To be more precise, we should say that there is a natural bijection between the set of minimizing measures in Theorem 13 and the set of Mather measures. This bijection is described in [4]. In order to give the reader a clue of this bijection, let us recall that the Mather measures are probability measures on the tangent bundle TM, and that the minimizing measures of Theorem 13 are probability measures on $M \times M$. Denoting by φ the time-one Lagrangian flow, and by $\pi : TM \to M$ the standard projection, we have a well-defined mapping $(\pi, \pi \circ \varphi)_\sharp$ from the set of probability measures on TM to the set of probability measures on $M \times M$. This mapping induces a bijection between the set of Mather measures on TM and the set of minimizing measures of Theorem 13.

The part of Theorem 13 stating that the minimizing measures are precisely the measures supported on D is the analogous in our setting of the theorem of Mañé stating that all invariant measures supported on the Aubry set are minimizing, see [13].

References

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

[2] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows, Lectures in Math. ETH Zürich, Birkhäuser, 2005.

[3] P. Bernard, Symplectic aspects of Aubry-Mather theory, preprint.

[4] P. Bernard and B. Buffoni, Optimal mass transportation and Mather theory, to appear in JEMS.

[5] R. Bowen, Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms, Lecture Notes in Math., **470**, Springer-Verlag, 1975.

[6] G. Contreras, Action potential and weak KAM solutions, Calc. Var. Partial Differential Equations, **13** (2001), 427–458.

[7] R. M. Dudley, Real Analysis and Probability, Cambridge Univ. Press, 2002.

[8] A. Fathi, Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens, (French) [A weak KAM theorem and Mather's theory of Lagrangian systems], C. R. Acad. Sci. Paris Ser. I Math., **324** (1997), 1043–1046.

[9] A. Fathi, Solutions KAM faibles conjuguées et barrières de Peierls, (French) [Weakly conjugate KAM solutions and Peierls's barriers], C. R. Acad. Sci. Paris Ser. I Math., **325** (1997), 649–652.

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

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

[12] R. Mañé, Ergodic Theory and Differentiable Dynamics, Springer-Verlag, 1987.

[13] R. Mañé, On the minimizing measures of Lagrangian dynamical systems, Nonlinearity, **5** (1992), 623–638.

[14] R. Mañé, Lagrangian flows: The dynamics of globally minimizing orbits, Bol. Soc. Bras. Mat, **28** (1997), 141–153.

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

[16] C. Villani, Topics in optimal transportation, Amer. Math. Soc., Providence, RI, 2003.

Patrick Bernard
CEREMADE
Université de Paris Dauphine
Pl. du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
France

Boris Buffoni
School of Mathematics
École Polytechnique Fédérale-Lausanne
SB/IACS/ANA Station 8
1015 Lausanne
Switzerland