Minimax Theory and its Applications
Volume 8 (2023), No. 1, 213–234

Minimal Action on the Space of Measures and Hamilton-Jacobi Equations

Antonio Siconolfi
Department of Mathematics, Università degli Studi di Roma "La Sapienza", Italy
siconolf@mat.uniroma1.it

Received: December 31, 2021
Accepted: February 20, 2022

We study the minimal action induced by a convex Lagrangian on the torus in the space of probability measures, in connection with the viscosity solutions of a pair of corresponding conjugate time-dependent Hamilton Jacobi equations. To this scope, we prove a version of Kantorovich duality theorem in the aforementioned framework through an argument that, as far as we can judge, is simple and new. It is based on the construction of approximate linear optimization problems posed in a space of matrices and a passage at the limit exploiting the density property of the convex hull of Dirac measures in the space of the probability measures on the torus with the narrow topology. In the proof it is solely used the finite dimensional version of Hahn-Banach theorem.

Keywords: Wasserstein distances, Hamilton-Jacobi equations, Kantorovich duality theorem, probability measures.

2020 Mathematics Subject Classification: 35D40, 49L25, 49N15, 60B10, 90C05.

1. Introduction

Given a convex Hamiltonian/Lagrangian defined in the cotangent/tangent bundle of the flat torus $\mathbb{T}^N$, this note aims at studying the related minimal action on the space of probability measures in connection with the viscosity solutions of a pair of corresponding conjugate time-dependent Hamilton Jacobi equations.

It is well established and first, as far as we know, highlighted in [2] for first order Hamilton-Jacobi and in [7] for the second order case, namely in presence of a diffusion (see also [6]), that this link is given by the Kantorovich duality theorem, see [1, 10]. The novelty of our contribution is that we provide a different proof of this result in our context, that we believe quite simple and direct. It is based on the formulation of a sequence of linear optimization problems in a finite-dimensional space of matrices plus a passage to the limit exploiting the density property of the convex hull of Dirac measures in the space of the probability measures on $\mathbb{T}^N$ with the narrow topology.

As in the other known proofs, we use as crucial duality tool the Hahn-Banach theorem, however, in contrast with what usually done, we just need it in the finite dimensional case. It is definitely more elementary than the general version since its proof relies on a easy induction argument and does not require Zorn lemma. See Appendix B for further remarks on this issue.

The work of the author was partially supported by Fondo Ateneo 2019, Università di Roma "La Sapienza"00.

ISSN 2199-1413 (printed), ISSN 2199-1421 (electronic) / $ 2.50

We consider a Hamiltonian $H : \mathbb{T}^N \times \mathbb{R}^N \to \mathbb{R}$ which is just assumed to be continuous in both arguments, convex in the momentum variable and superlinear at infinity. We emphasize that no Tonelli type regularity or strict convexity is required. We denote by L the corresponding Lagrangian obtained via Fenchel transform. Given x, y, $t_2 > t_1$, we define the minimal action

$$S(x, t_1, y, t_2) = \inf \int_0^{t_2 - t_1} L(\xi, \dot{\xi})\, dt, \tag{1}$$

where the minimum is over the absolutely continuous curves defined in $[0, t_2 - t_1]$ and linking x to y. We fix $T > 0$ and take for simplicity from now on $t_1 = 0$, $t_2 = T$, but of course any pair of times t_1, t_2 with $t_2 > t_1$ will lead to the same results. The minimal action is related via Lax-Oleinik formula to the pair of conjugate time dependent Hamilton Jacobi equations

$$u_t + H(x, Du) = 0 \qquad \text{in } \mathbb{T}^N \times (0, T) \tag{HJ+}$$

$$-u_t + H(x, -Du) = 0 \qquad \text{in } \mathbb{T}^N \times (0, T) \tag{HJ−}$$

In Section 3 we provide a metric interpretation of (1) which adds some geometrical intuition to the matter. We stress that in our setting the (viscosity) solutions to the above equations could be just continuous, without enjoying any Lipschitz continuity or seminconcavity/convexity property.

We lift the minimal action to the space of probability measures on $\mathbb{T}^N$ setting for any probability measures μ, ν on $\mathbb{T}^N$

$$S(\mu \times \delta_0, \nu \times \delta_T) = \inf_\gamma \int S(x, 0, y, T)\, d\gamma$$

where the infimum is over the probability measures γ on $\mathbb{T}^N \times \mathbb{T}^N$ with first marginal μ and second marginal ν. Our main result is:

Theorem 1.1. *Given probability measures μ, ν on $\mathbb{T}^N$, there exist continuous functions g, $g^\checkmark$ such that the solutions v, $v^\checkmark$ of* (HJ+), (HJ−), *with datum g at $t = 0$ and datum $g^\checkmark$ at $t = T$, respectively, satisfy*

$$\begin{aligned} &S(\mu \times \delta_0, \nu \times \delta_T) = \\ &= \int v(x, T)\, d\nu - \int g(x)\, d\mu = \int v^\checkmark(x, 0)\, d\mu - \int g^\checkmark(x)\, d\nu \\ &= \max_u \left[\int u(x, T)\, d\nu - \int u(x, 0)\, d\mu \right] = \max_{u^\checkmark} \left[\int u^\checkmark(x, 0)\, d\mu - \int u^\checkmark(x, T)\, d\nu \right] \end{aligned} \tag{2}$$

where the maxima are over the (viscosity) subsolutions u, $u^\checkmark$ of (HJ+), (HJ−), *respectively.*

Even if we fully prove the above statement, it is theoretically significant to point out that it is actually redundant, in the sense that the duality formula for (HJ−) follows from that of (HJ+) and vice versa, see Proposition 4.3.

In [2] the duality result is equivalently stated in terms of Kantorovich pairs, namely functions ϕ_0, $\phi_1 : \mathbb{T}^N \to \mathbb{R}$ satisfying

$$\phi_1(y) = \min_{x \in \mathbb{T}^N} \phi_0(x) + S(x, 0, y, T) \qquad \text{and} \qquad \phi_0(x) = \max_{y \in \mathbb{T}^N} \phi_1(y) - S(x, 0, y, T).$$

We would just like to point out that a Kantorovich pair is not so far from being a solution to (HJ+).

It is in fact clear by Lax-Oleinik formula that if ϕ_0, ϕ_1 is a Kantorovich pair then $\phi_1 = v(\cdot, T)$, where v is the solution of (HJ+) with initial datum ϕ_0 at $t = 0$. A converse result says that any solution v of (HJ+), gives rise to a Kantorovich pair, up to modifying v at $t = 0$.

Proposition 1.2. *Let v be a solution to* (HJ+)*, then the functions*

$$\phi(y) = \max_x v(x, T) - S(y, 0, x, T)$$

and $v(x, T)$ makes up a Kantorovich pair.

Proof. Given an arbitrary $x \in \mathbb{T}^N$, we have

$$\min_y \phi(y) + S(y, 0, x, T) \geq v(x, T) - S(y, 0, x, T) + S(y, 0, x, T) = v(x, T),$$

on the other hand

$$\phi(y) \leq v(y, 0) + S(y, 0, x, T) - S(y, 0, x, T) = v(y, 0)$$

which implies

$$\min_y \phi(y) + S(y, 0, x, T) \leq \min_y v(y, 0) + S(y, 0, x, T) = v(y, T),$$

and in the end $\min_y \phi(y) + S(y, 0, x, T) = v(y, T)$, which gives the assertion. □

The paper is organized as follows: in Section 2 we fix the notations and give some preliminaries, in section 3 we make precise the assumptions on the Hamiltonian, provide a metric interpretation of the minimal action and write down Lax-Oleinik formulas.

Section 4 is devoted to the definition of the minimal action in the space of probability measures plus the proof of related result that we use in what follows, in the final part it is discussed the redundancy in the statement of Theorem 1.1. In Section 5 we formulate the optimization problems in a space of matrices we use in the proof of the main result, which is proved along with some related material, in Section 6.

Finally, in Appendix A it is recalled some classical results on the minimization of the action functional over absolutely continuous curves, and in Appendix B it is discussed Hahn-Banach theorem in the finite dimensional case and it is derived a corollary.

Acknowledgments

The author would like to thank an anonymous referee for helpful comments and suggestions which allow to improve the text.

2. Notations and preliminaries

We denote by $\mathbb{T}^N$ the N- dimensional *torus* $\mathbb{R}^N/\mathbb{Z}^N$. Any *probability measure* on $\mathbb{T}^N$ is understood to be *Borel*. We write δ_x (resp δ_t) for the *Dirac* measure

centered at the point $x \in \mathbb{T}^N$ (resp. at a time t). We denote by $\langle \cdot, \cdot \rangle$ the *pairing* between continuous functions defined on $\mathbb{T}^N$ and probability measures. The space of probability measures on $\mathbb{T}^N$ is endowed with the *narrow* topology which makes it a compact space. A sequence μ_n of probability measures narrow converges to a measure μ if

$$\langle \mu_n, f \rangle \to \langle \mu, f \rangle \quad \text{for any continuous } f : \mathbb{T}^N \to \mathbb{R}.$$

We will crucially use the following result, see for instance [8]:

Theorem 2.1. *The convex hull of Dirac measures is dense in the space of probability measures on $\mathbb{T}^N$ with respect to the narrow topology.*

Given an open interval $I \subset (0, +\infty)$ and a C^1 function $u : \mathbb{T}^N \times I \to \mathbb{R}$, we denote by u_t, Du the *time* and *spatial derivative* of u, respectively. The notion of (sub)solution for the Hamilton-Jacobi equations considered throughout the paper must be understood in the *viscosity sense*.

We fix throughout the paper a positive time T. We denote by $AC(0,T)$ the space of *absolutely continuous curves* ξ defined in $[0,T]$ with support in $\mathbb{T}^N$ endowed with the norm

$$\int_0^T |\xi| \, dt + \int_0^T |\dot{\xi}| \, dt$$

where $\dot{\xi}$ stands for the time derivative. We also consider it endowed with the corresponding weak topology.

Given the Euclidean space $\mathbb{R}^M$, we identify it with its dual and denote by $\cdot$ the *scalar product*. Given $x \in \mathbb{R}^M$, $r > 0$, Ww further denote by $B(x,r)$ the ball eith radius r centeres at x. Sums or differences of subsets of $\mathbb{R}^M$ are intended in the *Minkowski sense*. For a convex function $F : \mathbb{R}^M \to \mathbb{R}$ we denote by $\partial F(x)$ its *subdifferential* at a point x. If C is a closed convex subset of $\mathbb{R}^M$, we define, for any $p \in \mathbb{R}^M$, the (possibly infinite) *support function* as

$$\sigma_C(p) = \sup\{p \cdot x \mid x \in C\}.$$

if $x \in C$, the *normal cone* of C at x is defined as

$$N_C(x) = \{p \in \mathbb{R}^M \mid p \cdot x = \sigma_C(p)\}.$$

If 0 is an interior point of C, we define the *gauge function*

$$\gamma_C(x) = \inf\{\lambda > 0 \mid x/\lambda \in C\},$$

which is continuous.

If f is a continuous function from $\mathbb{T}^N$ (or $\mathbb{T}^N \times \mathbb{T}^N$) to $\mathbb{R}$, we say that the function $\omega : [0, +\infty) \to [0, +\infty)$ is a *uniform continuity modulus* for f if $\omega(r) > 0$ for $r > 0$, $\lim_{r \to 0} \omega(r) = 0$ and $f(x) - f(y) \leq \omega(|x - y|)$ for any x, y.

A uniform continuity modulus can be taken, without loosing generality, with the additional properties of being concave, and so continuous, and nondecreasing. More

generally, we call *modulus* any function from $[0,+\infty)$ to $[0,+\infty)$ which is continuous and nondecreasing.

The following result will be useful.

Lemma 2.2. *Let X be a metric space and γ a (positive) measure on $X \times X$ with marginals μ and ν, then*

$$\text{spt } \gamma \subset \text{spt } \mu \times \text{spt } \nu.$$

Proof. Let $(x,y) \in \text{spt } \gamma$. Given any open set E containing x, we have by the very definition of marginals

$$\mu(E) = \pi_1 \# \gamma(E) = \gamma(E \times \mathbb{T}^N),$$

where $\#$ indicates the push-forward. Since $E \times \mathbb{T}^N$ is an open set containing (x,y), we deduce $\mu(E) = \gamma(E \times \mathbb{T}^N) > 0$, which shows that $x \in \text{spt } \mu$. We prove in the same way that $y \in \text{spt } \nu$. □

3. Assumptions

We are given a Hamiltonian $H : \mathbb{T}^N \times \mathbb{R}^N \to \mathbb{R}$ with the following properties:

(A1) H is continuous in $\mathbb{T}^N \times \mathbb{R}^N$;

(A2) $p \mapsto H(x,p)$ is convex for any $x \in \mathbb{T}^N$;

(A3) H is superlinear in p uniformly in x.

We indicate by L the Lagrangian obtained from H via Fenchel transform. Note that(A3) implies that there exists a function $\theta : [0,+\infty) \to \mathbb{R}$ superlinear such that

$$L(x,q) \geq \theta(|q|) \qquad \text{for all } x. \tag{3}$$

We further consider the time dependent Hamilton Jacobi equations (HJ+), (HJ−), written in the Introduction, coupled with the conditions

$$u(\cdot,0) = g \quad \text{in } \mathbb{T}^N \tag{D+}$$

$$u(\cdot,T) = g^{\checkmark} \quad \text{in } \mathbb{T}^N \tag{D−}$$

respectively, for given continuous functions g, $g^{\checkmark}$. These equations correspond to the conjugate Hamiltonians from $\mathbb{T}^N \times \mathbb{R}^N \times \mathbb{R}$ to $\mathbb{R}$

$$F(x,p,s) = s + H(x,p) \quad \text{and} \quad F(x,p,s)^{\checkmark} = -s + H(x,-p)$$

As it is well known, the solutions to the above problems can be represented through Lax-Oleinik formula, based on minimization of the action. We interpret it in metric terms, mostly for cosmetic reasons. This angle highlights that the passage from F to $F^{\checkmark}$ corresponds to a time inversion and adds a modicum of geometric intuition to the matter.

Following the procedure introduced in [9], we associate a semidistance in $\mathbb{T}^N \times [0,T]$, possibly infinite, to (HJ+), (HJ−), looking at the 0-sublevels of F, $F^{\checkmark}$. We define

$$\begin{aligned} Z(x) &= \{(p,s) \mid F(x,p,s) \leq 0\} \\ Z^{\checkmark}(x) &= -Z(x) = \{(p,s) \mid F^{\checkmark}(x,p,s) \leq 0\} \end{aligned}$$

and consider the related support functions

$$\begin{aligned}\sigma(x,q,r) &= \sigma_{Z(x)}(q,r) = \sup\{p\cdot q + s\,r \mid (p,s)\in Z(x)\}\\ \sigma^{\checkmark}(x,q,r) &= \sigma_{Z^{\checkmark}(x)}(q,r) = \sup\{p\cdot q + s\,r \mid (p,s)\in Z^{\checkmark}(x)\}.\end{aligned}$$

We see that $\sigma(x,q,r)$ (resp. $\sigma^{\checkmark}(x,q,r)$) is finite if and only if $r>0$ (resp. $r<0$), by positive homogeneity we can focus on $(q,1)$ (resp. $(q,-1)$). Taking into account that to determine the support functions it is enough to compute the supremum for (p,s) with $F(x,t,p,s)=0$ (resp. $F^{\checkmark}(x,t,p,s)=0$) or in other terms with

$$s=-H(x,p)\qquad (\text{resp. } s=H(x,t,-p)),$$

we have
$$\sigma(x,q,1)=\sup_p -H(x,p)+p\cdot q = L(x,q)$$
$$\sigma^{\checkmark}(x,q,-1)=\sup_p -H(x,-p)+p\cdot q = L(x,-q)$$

Given (x,t_1), (y,t_2) with $t_2>t_1$, we finally define the intrinsic (semi) distances $S(x,t_1,y,t_2)$, $S^{\checkmark}(y,t_2,x,t_1)$ as follows

$$S(x,t_1,y,t_2)=\inf\int_0^{t_2-t_1}\sigma(\xi(t),\dot\xi(t),1)\,dt=\inf\int_0^{t_2-t_2}L(\xi(t),\dot\xi(t))\,dt \tag{4}$$
$$S^{\checkmark}(y,t_2,x,t_1)=\inf\int_0^{t_1-t_2}\sigma^{\checkmark}(\zeta(t),\dot\zeta(t),-1)\,dt=\inf\int_0^{t_1-t_2}L(\zeta(t),-\dot\zeta(t))\,dt$$

where the infimum is taken over the absolutely continuous curves ξ with $\xi(0)=x$, $\xi(t_2-t_1)=y$ (resp. ζ with $\zeta(0)=y$, $\zeta(t_1-t_2)=x$). If ζ is such a curve, by setting

$$\xi(t)=\zeta(-t)$$

we get
$$\int_0^{t_1-t_2}L(\zeta(t),-\dot\zeta(t))\,dt=\int_0^{t_1-t_2}L(\xi(-t),\dot\xi(-t))\,dt$$

and changing integration variable from t to $r=-t$ we further obtain

$$\int_0^{t_1-t_2}L(\zeta(t),-\dot\zeta(t))\,dt=\int_0^{t_2-t_1}L(\xi(r),\dot\xi(r))\,dr$$

which shows in the end the relation

$$S^{\checkmark}(y,t_2,x,t_1)=S(x,t_1,y,t_2)\qquad \text{for any } x,\ y \text{ in } \mathbb{T}^N,\ t_2>t_1.$$

Proposition 3.1. *Given $T>0$, the function*

$$(x,t,y,r)\mapsto S(x,t,y,r) \tag{5}$$

is continuous for x, y varying in $\mathbb{T}^N$, t, r varying in $[0,T]$ with $r>t$.

The proof is given in Appendix A.

The following well known result is about Lax-Oleinik representation formulas, see [5].

Proposition 3.2. *The problems* (HJ+), (D+) *and* (HJ−), (D−) *admit unique solution denoted by* v, $v^\checkmark$, *respectively, given by the representation formulas*

$$\begin{aligned} v(x,t) &= \min_{z\in\mathbb{T}^N}\big(g(z)+S(z,0,x,t)\big) \qquad \text{for } (x,t)\in\mathbb{T}^N\times(0,T) \\ v^\checkmark(x,t) &= \min_{z\in\mathbb{T}^N}\big(g^\checkmark(z)+S(x,t,z,T)\big) \qquad \text{for } (x,t)\in\mathbb{T}^N\times(0,T) \end{aligned}$$

We record for later use:

Corollary 3.3. *Any continuous subsolution* u *of* (HJ+) *(resp. of* (HJ−)*) satisfies*

$$u(y,t_2)-u(x,t_1)\le S(x,t_1,y,t_2) \quad (\text{resp. } u(x,t_1)-u(y,t_2)\le S(x,t_1,y,t_2))$$

for any x, y *in* $\mathbb{T}^N$, $t_2>t_1$.

Proof. We prove the statement for subsolutions of (HJ−), the argument for (HJ+) goes along the same lines. According to Proposition 3.2, the solution v of (HJ−) with final datum $u(x,t_2)$ at $t=t_2$ is

$$v(x,t)=\inf_z u(z,t_2)+S(x,t,z,t_2) \qquad \text{in } \mathbb{T}^N\times(0,t_2).$$

By the comparison principle we get $u(x,t_1)\le v(x,t_1)\le u(y,t_2)+S(x,t_1,y,t_2)$. □

4. Lift of the minimal action

We lift the semidistance defined in (4) to the space of probability measures on $\mathbb{T}^N\times[0,T]$ adapting the usual formula used for Wasserstein distances. We focus on measures of the form $\mu\times\delta_0$, $\nu\times\delta_T$, with μ, ν probability measures on $\mathbb{T}^N$. Taking into account Lemma 2.2, we get

$$S(\mu\times\delta_0,\nu\times\delta_T)=\inf_\gamma\int S(x,0,y,T)\,d\gamma$$

where the infimum is over the probability measures γ on $\mathbb{T}^N\times\mathbb{T}^N$ with first marginal μ and second marginal ν.

We record for later use:

Lemma 4.1. *Let* γ_n *be a sequence of probability measures in* $\mathbb{T}^N\times\mathbb{T}^N$ *narrowly converging to some measure* γ. *Then the first and the second marginals of* γ_n *converge to the first and the second marginal of* γ, *respectively.*

Proof. This is just an application of Portmanteau Theorem. Let μ_n be the first marginals of γ_n. We denote by C a closed set of $\mathbb{T}^N$. We have

$$\limsup\mu_n(C)=\limsup\gamma_n(C\times\mathbb{T}^N)\le\gamma(C\times\mathbb{T}^N)=\mu(C).$$

Same argument applies to the second marginals. This concludes the proof. □

Lemma 4.2. *Let* μ_n, ν_n *be sequences of measures in* $\mathbb{T}^N$ *narrowly converging to* μ, ν, *respectively. Then*

$$\liminf S(\mu_n\times\delta_0,\nu_n\times\delta_T)\ge S(\mu\times\delta_0,\nu\times\delta_T).$$

Proof. Assume that

$$S(\mu_n \times \delta_0, \nu_n \times \delta_T) > \int S(x,0,y,T)\, d\gamma_n - \frac{1}{n}$$

for a sequence of measures γ_n in $\mathbb{T}^N \times \mathbb{T}^N$ with marginals μ_n, ν_n. The γ_n narrowly converge, up to subsequences, to a measure γ with marginals μ, ν by Lemma 4.1, and consequently

$$\begin{aligned}\liminf_n S(\mu_n \times \delta_0, \nu_n \times \delta_T) &\geq \lim_n \int S(x,0,y,T)\, d\gamma_n - \frac{1}{n} \\ &= \int S(x,0,y,T)\, d\gamma \geq S(\mu \times \delta_0, \nu \times \delta_T),\end{aligned}$$

as was asserted. □

In the final part of the section we discuss the redundancy in the statement of Theorem 1.1 already pointed out in the Introduction. We prove:

Proposition 4.3. *Given an Hamiltonian H, assume that for any pair of probability measures μ, ν on $\mathbb{T}^N$, we have*

$$S(\mu \times \delta_0, \nu \times \delta_T) = \sup_u \big[\langle \nu, u(\cdot,T)\rangle - \langle \mu, u(\cdot,0)\rangle\big] \tag{6}$$

where the infimum is over the subsolutions of (HJ+)*, and in addition that there exists a continuous function g, such that the solution v, of* (HJ+)*, with datum g at $t = 0$ satisfies*

$$S(\mu \times \delta_0, \nu \times \delta_0) = \langle \nu, v(\cdot,T)\rangle - \langle \mu, g\rangle. \tag{7}$$

Then we derive that

$$S(\mu \times \delta_0, \nu \times \delta_0) = \sup_{u^\checkmark} \big[\langle \mu, u^\checkmark(\cdot,0)\rangle - \langle \nu, u^\checkmark(\cdot,T)\rangle\big] \tag{8}$$

where the infimum is over the subsolutions of (HJ−)*, and in addition that there exists a continuous function f such that the solution w of* (HJ−) *with datum f at $t = T$ satisfies*

$$S(\mu \times \delta_0, \nu \times \delta_0) = \langle \mu, w(\cdot,0)\rangle - \langle \nu, f\rangle. \tag{9}$$

The converse implication is true as well.

We preliminarily introduce the Hamiltonian $H(x,-p)$ and the equation

$$u_t + H(x,-Du) = 0 \qquad \text{in } \mathbb{T}^N \times (0,T) \tag{$\overline{\text{HJ}}$}$$

possibly coupled with initial condition at $t = 0$. The corresponding minimal action functional $\overline{S}$ is defined, for any x, y in $\mathbb{T}^N$, $t_2 > t_1$, by

$$\overline{S}(x,t_1,y,t_2) = \inf_\xi \int_{t_1}^{t_2} L(\xi,-\xi)\, dt,$$

where ξ varies among the absolute continuous curves defined in $[t_1,t_2]$, with x and y as initial and final point, respectively.

Given one of these curves ξ, we set

$$\zeta(r) = \xi(t_1 + t_2 - r),$$

so that ζ is an absolutely continuous curve defined in $[t_1, t_2]$ with $\zeta(t_1) = y$ and $\zeta(t_2) = x$. Putting $t = t_1 + t_2 - r$, we get

$$\int_{t_1}^{t_2} L(\zeta, \dot{\zeta})\, dr = \int_{t_1}^{t_2} L(\xi, -\dot{\xi})\, dt.$$

Taking into account that a converse construction is also possible, we deduce that

$$\overline{S}(x, t_1, y, t_2) = S(y, t_1, x, t_2). \tag{10}$$

Summing up: we pass from S to $S^{\checkmark}$ inverting the direction of time and interchanging initial and final position, while comparing S to $\overline{S}$ we find that initial and final positions are interchanged as well, but time is not affected.

The minimal action functional $\overline{S}$ can be lifted to the space of probability measures on $\mathbb{T}^N \times [0, T]$ as we did for S in the first part of the section. We derive from (10) that

$$S(\mu \times \delta_0, \nu \times \delta_T) = \bar{S}(\nu \times \delta_0, \mu \times \delta_T) \tag{11}$$

for any pair μ, ν of probability measures on $\mathbb{T}^N$.

The following statement can be easily checked through direct calculation :

Proposition 4.4. *If $w(x,t)$ is subsolution to* (HJ$-$) *then $u(x,t) = w(x, T-t)$ is subsolution of* ($\overline{\text{HJ}}$)*, and conversely if $u(x,t)$ is subsolution of* ($\overline{\text{HJ}}$) *then $w(x,t) = u(x, T-t)$ is subsolution to* (HJ$-$)*.*

Moreover, if $w(x,t)$ is solution to (HJ$-$) *agreeing with a continuous datum f at $t = T$, then $u(x,t) = w(x, T-t)$ is solution of* ($\overline{\text{HJ}}$) *and is equal to f at $t = 0$, and conversely if $u(x,t)$ is solution of* ($\overline{\text{HJ}}$) *agreeing with f at $t = 0$, then $w(x,t) = u(x, T-t)$ is solution to* (HJ$-$) *and is equal to f at $t = T$.*

Proof of Proposition 4.3. We apply the assumption to the Hamiltonian $H(x, -p)$. Given μ, ν we see that (8) can be deduced from (6) by Proposition 4.4. By assumption we can find an initial datum f at $t = 0$ such that the solution $u(x,t)$ of ($\overline{\text{HJ}}$) agreeing with f at $t = 0$ satisfies

$$\overline{S}(\nu \times \delta_0, \mu \times \delta_T) = \langle \mu, u(\cdot, T) \rangle - \langle \nu, f \rangle. \tag{12}$$

We deduce from Proposition 4.4 that $w(x,t) = u(x, T-t)$ solves ($\overline{\text{HJ}}$) plus $w(\cdot, T) = f$, taking into account (11), we derive from (12)

$$S(\mu \times \delta_0, \nu \times \delta_T) = \langle \mu, w(\cdot, T) \rangle - \langle \nu, f \rangle.$$

The converse implication can be proved adapting the above argument. □

5. Linear optimization problems in a space of matrices

We consider two finite collections of points in $\mathbb{T}^N$ denoted by $x_1, \cdots, x_n$; $y_1, \cdots, y_m$ and positive coefficients $\rho_1, \cdots, \rho_n$; $\lambda_1, \cdots, \lambda_m$ with $\sum_i \rho_i = \sum_j \lambda_j = 1$. We further consider the vector space $\mathbf{M}$ of $n \times m$ matrices identified with its the dual space, with duality pairing given by the Frobenius inner product and denoted by $\cdot$, we call positive a linear functional on $\mathbf{M}$ if it represented by a matrix with nonnegative entries. We fix the matrix $S = (S_{ij}) \in \mathbf{M}$ with

$$S_{ij} = S(x_i, 0, y_j, T) \tag{13}$$

and denote by $\mathbf{C}$ the cone $\mathbf{C} = \{A \in \mathbf{M} \mid A \leq S\}$ where $\leq$ indicates the componentwise order in $\mathbf{M}$. We introduce the vector subspace $\mathbf{W} \subset \mathbf{M}$ made up by matrices $A = (a_{ij})$ of the form

$$a_{ij} = v_i - w_j \tag{14}$$

where $v = (v_1, \cdots, v_n)$ and $w = (w_1, \cdots, w_m)$ are vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and define a linear functional P on $\mathbf{W}$ it via

$$P(A) = \sum_i \rho_i \, v_i - \sum_j \lambda_j \, w_j. \tag{15}$$

The above definition is actually well posed. Assume in fact that there are vectors $v' = (v'_1, \cdots, v'_n)$ and $w' = (w'_1, \cdots, w'_m)$ with

$$a_{ij} = v_i - w_j = v'_i - w'_j \qquad \text{for any } i,\, j,$$

this implies

$$\begin{aligned} v_1 - v'_1 &= w_1 - w'_1 \\ v_i - w_1 &= v'_i - w'_1 \qquad \text{for any } i \\ w_j - v_1 &= w'_j - v'_1 \qquad \text{for any } j. \end{aligned} \tag{16}$$

Hence we have

$$\begin{aligned} \sum_i \rho_i \, v_i &= \sum_i \rho_i \, v'_i + (w_1 - w'_1) \\ \sum_j \lambda_j \, w_j &= \sum_j \lambda_j \, w'_j + (v_1 - v'_1) \end{aligned}$$

and, taking into account (16)

$$\begin{aligned} \sum_i \rho_i \, v_i - \sum_j \lambda_j \, w_j &= \sum_i \rho_i \, v'_i - \sum_j \lambda_j \, w'_j + (w_1 - w'_1) - (v_1 - v'_1) \\ &= \sum_i \rho_i \, v'_i - \sum_j \lambda_j \, w'_j. \end{aligned}$$

It can be easily checked by direct computation that the matrix

$$B = (\rho_i \lambda_j)$$

provides an example of a linear functional on $\mathbf{M}$ extending P, namely such that

$$B \cdot A = P(A) \qquad \text{for any } A \in \mathbf{W}.$$

Since B is in addition positive, and any positive functional is bounded from above in C, we immediately deduce

Lemma 5.1. *The functional P is bounded from above in $\mathbf{W} \cap \mathbf{C}$.*

We have the following characterization of linear functional extending P:

Lemma 5.2. *A functional $B = (b_{ij})$ on $\mathbf{M}$ is an extension of P if and only if*

$$\sum_j b_{ij} = \rho_i \quad \text{for any } i, \text{ and} \quad \sum_i b_{ij} = \lambda_j \qquad \text{for any } j. \tag{17}$$

Proof. Let $B = (b_{ij})$ be an extension of P, and take $A = (a_{ij}) \in \mathbf{W}$ of the form (14). The assertion is consequence of the following equalities

$$\begin{aligned} B \cdot A &= \sum_{i,j} b_{ij}\,(v_i - w_j) = \sum_i \Big(\sum_j b_{ij}\Big) v_i - \sum_j \Big(\sum_i b_{ij}\Big) w_i \\ &= \sum_i \rho_i\, v_i - \sum_j \lambda_j\, w_j. \end{aligned}$$

In addition we have: □

Lemma 5.3. *The functional P admits maximizers $\widetilde{A} = (\widetilde{a}_{ij})$, $\bar{A} = (\bar{a}_{ij})$, in $\mathbf{W} \cap \mathbf{C}$ of the form*

$$\widetilde{a}_{ij} = -\widetilde{c}_i + (\min_h \widetilde{c}_h + S_{hj}) \tag{18}$$

$$\bar{a}_{ij} = (\min_k \bar{c}_k + S_{ik}) - \bar{c}_j \tag{19}$$

for suitable $(\widetilde{c}_1, \cdots, \widetilde{c}_n) \in \mathbb{R}^n$, $(\bar{c}_1, \cdots, \bar{c}_m) \in \mathbb{R}^m$.

Proof. Let $A_n = (v_i^n - w_j^n)$ be a maximizing sequence in $\mathbf{W} \cap \mathbf{C}$, since

$$v_i^n - w_j^n \le S_{ij} \qquad \text{for all } i,\, j,$$

we deduce that

$$\begin{aligned} -w_j^n &\le \min_h -v_h^n - S_{hj} \qquad \text{for any } j \\ v_i^n &\le \min_k w_k^n + S_{ik} \qquad \text{for any } i. \end{aligned}$$

If we set

$$\widetilde{A}_n = (v_i^n + \min_h(-v_h^n + S_{hj})) \tag{20}$$

$$\bar{A}_n = (\min_k(w_k^n + S_{ik}) - w_j^n) \tag{21}$$

we find that $\widetilde{A}_n$, $\bar{A}_n \in \mathbf{W} \cap \mathbf{C}$ with

$$P(\widetilde{A}_n) \ge P(A_n) \qquad \text{and} \qquad P(\bar{A}_n) \ge P(A_n)$$

which shows that there are maximizing sequences of the form (20), (21). If we replace in (20) v_i^n by $v_i^n + r_n$ (resp. w_n^j by $w_j^n + r_n$ in (21)), where r_n is an arbitrary sequence of real numbers, the corresponding value of P is not affected, so that we can assume, without loosing generality, that

$$v_1^n = 0 \qquad \text{in (20)} \quad \text{and} \quad w_1^n = 0 \quad \text{in (21)} \quad \text{for any } n. \tag{22}$$

We can further assume, up to extracting subsequences, that the sequences v_i^n, w_j^n admit limit, as n goes to infinity, for all i, j.

We set $\widetilde{d}_n = \max_i v_i^n \geq 0$, $\bar{d}_n = \min_j w_j^n \leq 0$, $s = \max_{i,j} S_{ij}$, taking into account (22), we have

$$\begin{aligned} P(\widetilde{A}_n) &\leq (1-\rho_1)\widetilde{d}_n + \sum_j \lambda_j(-\widetilde{d}_n + s) = -\rho_1\,\widetilde{d}_n + s \\ P(\bar{A}_n) &\leq \sum_i \rho_i(\bar{d}_n + s) - (1-\lambda_1)\,\bar{d}_n = \lambda_1\,\bar{d}_n + s \end{aligned}$$

which implies that the v_i^n's stay bounded from above and the w_j^n's from below as $n \to +\infty$, otherwise $\widetilde{A}_n$, $\bar{A}_n$ could not be maximizing sequences for P.

On the other hand, if there is an index h with v_h^n negatively diverging (resp. w_h^n positively diverging) then we have

$$\begin{aligned} P(\widetilde{A}_n) &\leq \sum_i \rho_i\, v_i^n + \sum_j \lambda_j\, s = \sum_i \rho_i\, v_i^n + s \to -\infty \\ P(\bar{A}_n) &\leq \sum_i \rho_i\, s - \sum_j \lambda_j\, w_j^n = s - \sum_j \lambda_j w_j^n \to -\infty \end{aligned}$$

which is again in contradiction with $\widetilde{A}_n$, $\bar{A}_n$ being maximizing sequences. All in all, we have that $\widetilde{A}_n$, $\bar{A}_n$ are bounded and consequently converging to maximizers $\widetilde{A}$, $\bar{A}$ of P in $\mathbf{W} \cap \mathbf{C}$ of the form (18), (19) respectively. □

Proposition 5.4. *There is a positive functional B on $\mathbf{M}$ extending P, and vectors $(\widetilde{c}_1, \cdots, \widetilde{c}_n) \in \mathbb{R}^n$ and $(\bar{c}_1, \cdots \bar{c}_m) \in \mathbb{R}^m$ such that*

$$B_{ij} = 0 \quad \textit{if} \quad -\widetilde{c}_i + \min_h(\widetilde{c}_h + S_{hj}) < S_{ij} \tag{23}$$

$$B \cdot S = -\sum_i \rho_i\, \widetilde{c}_i + \sum_i \lambda_j(\min_h \widetilde{c}_h + S_{hj}) \tag{24}$$

$$B_{ij} = 0 \quad \textit{if} \quad \min_k (c_k + S_{ik}) - \bar{c}_j < S_{ij} \tag{25}$$

$$B \cdot S = \sum_i \rho_i(\min_k \bar{c}_k + S_{ik}) - \sum_j \lambda_j\, \bar{c}_j. \tag{26}$$

Proof. We note preliminarily that if $m = \min_{\mathbb{T}^N \times \mathbb{T}^N} S(\cdot, 0, \cdot, T)$, then $A = (v_i - w_j)$, where v, w are with all the components equal to $m-1$ and to 0, respectively, belongs to $\mathbf{W}$ intersected with the interior of $\mathbf{C}$.

We can therefore apply the finite dimensional version of Hahn-Banach theorem, see Appendix B and Corollary B.3, which guarantees the existence of a functional B extending P with

$$\max\{P(A) \mid A \in \mathbf{W} \cap \mathbf{C}\} = \sigma_{\mathbf{C}}(B).$$

Clearly B must be positive otherwise it should be unbounded in $\mathbf{C}$, since B is positive then, by the very definition of $\mathbf{C}$, $\sigma_{\mathbf{C}}(B) = B \cdot S$. By Lemma 5.3 there are a maximizers $\widetilde{A}$, $\bar{A}$ of P in $\mathbf{W} \cap \mathbf{C}$ of the form (18), (19) respectively. and consequently

$$P(\widetilde{A}) = B \cdot \widetilde{A} = P(\bar{A}) = B \cdot \bar{A} = \sigma_{\mathbf{C}}(B) = B \cdot S$$

which gives (24), (26).

We further derive that $\widetilde{A}$, $\bar{A} \in \partial\mathbf{C}$ and $B \in N_{\mathbf{C}}(\widetilde{A}) \cap N_{\mathbf{C}}(\bar{A})$. Taking into account that C is a cone we therefore have that

$$\begin{aligned} -\widetilde{c}_i + \min_h(\widetilde{c}_h + S_{hj}) &= S_{ij} \qquad \text{for some } i,\ j \\ \min_k(\bar{c}_k + S_{ik}) - \bar{c}_j &= S_{ij} \qquad \text{for some } i,\ j \end{aligned}$$

and (23), (25) hold true. □

6. Main result

In this section we prove Theorem 1.1. Note that by Corollary 3.3 and the very definition of $S(\mu \times \delta_0, \nu \times \delta_T)$ we get the inequality

$$\begin{aligned} S(\mu \times \delta_0, \nu \times \delta_T) &\leq \sup_u \left[\langle \nu, u(\cdot, T)\rangle - \langle \mu, u(\cdot, 0)\rangle\right] \\ S(\mu \times \delta_0, \nu \times \delta_T) &\leq \sup_{u^\checkmark} \left[\langle \mu, u^\checkmark(\cdot, 0)\rangle - \langle \nu, u^\checkmark(\cdot, T)\rangle\right] \end{aligned}$$

where the suprema are over the subsolutions u, $u^\checkmark$ of (HJ+), (HJ−), respectively. It is therefore enough to prove the first two equalities in (39).

To exploit the finite dimensional analysis set up in the previous section, we first assume

$$\mu = \sum_{i=1}^{n} \rho_i\, \delta_{x_i} \qquad \text{and} \qquad \nu = \sum_{j=1}^{m} \lambda_j\, \delta_{y_j} \tag{27}$$

for some $x_1, \cdots, x_n$; $y_1, \cdots y_m$ in $\mathbb{T}^N$, positive constants $\rho_1, \cdots, \rho_n$; $\lambda_1, \cdots, \lambda_n$ with $\sum_i \rho_i = \sum_j \lambda_j = 1$. We define moreover

$$g(x) = \min_i \big(b_i + \omega(|x_i - x|)\big) \tag{28}$$

$$g^\checkmark(x) = \min_j \big(a_j + \omega(|y_j - x|)\big) \tag{29}$$

where b_i, a_j are arbitrary constants and ω is a modulus strictly greater than the continuity modulus of $(x, y) \mapsto S(x, 0, y, T)$ in $\mathbb{T}^N \times \mathbb{T}^N$, see Proposition 3.1.

Proposition 6.1. *Assume μ, ν to be of the form* (27)*, there is a suitable choice of constants $b_1, \cdots, b_n$, $a_1, \cdots a_m$ such that the assertion of Theorem 1.1 is true with g, $g^\checkmark$ as in* (28), (29)*, respectively.*

A preliminary result is:

Lemma 6.2. *If g, $g^\checkmark$ are taken as in* (28), (29) *then the corresponding solution v, $v^\checkmark$ of* (HJ+), (D+), (HJ−), (D−) *satisfy for any $x \in \mathbb{T}^N$*

$$v(x, T) = \min_i \big(b_i + S(x_i, 0, x, T)\big) \tag{30}$$

$$v^\checkmark(x, 0) = \min_j \big(a_j + S(x, 0, y_j, T)\big) \tag{31}$$

Proof. We just show (31), the proof for (30) is the same with obvious modification. Taking into account the representation formula for solutions to (HJ−), we have that

$$v^{\checkmark}(x,0) \le \min_j g^{\checkmark}(y_j) + S(x,0,y_j,T) \le \min_j a_j + S(x,0,y_j,T)$$

for $x \in \mathbb{T}^N$. Assume for purposes of contradiction that there are points x_0, y_0 with

$$v^{\checkmark}(x_0,0) = g^{\checkmark}(y_0) + S(x_0,0,y_0,T) < \min_j a_j + S(x_0,0,y_j,T). \tag{32}$$

We can write $g^{\checkmark}(y_0) = a_{j_0} + \omega(|y_{j_0} - y_0|)$ for a suitable choice of the index j_0, and we derive from (32)

$$g^{\checkmark}(y_0) + S(x_0,0,y_0,T) < a_{j_0} + S(x_0,0,y_{j_0},T)$$

and consequently

$$S(x_0,0,y_{j_0},T) - S(x_0,0,y_0,T) > g^{\checkmark}(y_0) - a_{j_0} = \omega(|y_{j_0} - y_0|)$$

which is against the choice of ω. Therefore

$$v^{\checkmark}(x,0) = \min_j a_j + S(x,0,y_j,T) \qquad \text{for any } x \in \mathbb{T}^N,$$

as was asserted. □

Corollary 6.3. *We use the same notations as in the previous lemma. Assume that one has for some $x \in \mathbb{T}^N$ and some index k that*

$$\min_i \big(b_i + S(x_i,0,x,T)\big) - b_k = S(x_k,0,x,T)$$
$$\big(\textit{resp. } \min_j \big(a_j + S(x,0,y_j,T)\big) - a_k = S(x,0,y_k,T)\big)$$

then $b_k = g(x_k)$ (resp. $a_k = g^{\checkmark}(y_k)$).

Proof. We show the assertion for g. Since

$$b_k + S(x_k,0,x,T) \le b_i + S(x_i,0,x,T)$$

we have $b_k \le b_i + S(x_i,0,x,T) - S(x_k,0,x,T) \le b_i + \omega(|x_k - x_i|)$ for all i, and consequently $b_k \le g(x_k)$.

The reverse inequality follows from the very definition of g. □

Proof of Proposition 6.1. The probability measures γ in $\mathbb{T}^N \times \mathbb{T}^N$ with marginals μ and ν correspond to positive functionals $B = (b_{ij})$ on $\mathbf{M}$ with the property (17), via the relation $b_{ij} = b_{ij}\,\delta_{(x_i,y_j)}$. They satisfy

$$\int S(x,0,y,T)\,d\gamma = B \cdot S \tag{33}$$

where $S \in \mathbf{M}$ is defined as in (13), and in addition

$$B \in N_{\mathbf{C}}(S)$$

which implies that

$$\int S(x,0,y,T)\,d\gamma = B\cdot S \geq P(A) = \sum_i \rho_i\, v_i - \sum_j \lambda_j\, w_j \tag{34}$$

for any $A = (v_i - w_j) \in \mathbf{W}\cap\mathbf{C}$. Keeping in mind the equality (33), we can find by Proposition 5.4 a measure γ with marginals μ and ν and $(\widetilde{c}_1,\cdots,\widetilde{c}_n)\in\mathbb{R}^N$, $(\bar{c}_1,\cdots,\bar{c}_m)\in\mathbb{R}^m$ such that

$$\int S(x,0,y,T)\,d\gamma = -\sum_i \rho_i\,\widetilde{c}_i + \sum_j \lambda_j\,\min_h(\widetilde{c}_h + S_{hj}) = \sum_i \rho_i\,\min_k(\bar{c}_k + S_{ik}) - \sum_j \lambda_j\,\bar{c}_j$$

and consequently, taking into account (34),

$$\begin{aligned} &S(\mu\times\delta_0,\nu\times\delta_T) = \\ &= -\sum_i \rho_i\,\widetilde{c}_i + \sum_j \lambda_j\,\min_h(\widetilde{c}_h + S_{hj}) = \sum_i \rho_i\,\min_k(\bar{c}_k + S_{ik}) - \sum_j \lambda_j\,\bar{c}_j. \end{aligned} \tag{35}$$

If we take the solution v of (HJ+), (D+) (resp. $v^{\checkmark}$ of (HJ−), (D−)) with datum g at $t=0$ (resp. $g^{\checkmark}$ at $t=T$) as in (28) with $\widetilde{c}_i$ in place of b_i in (28) (resp. with $\bar{c}_j$ in place of a_j in (29)), we get by Lemma 6.2

$$\langle \nu, v(\cdot,T)\rangle - \langle \mu, g\rangle \;=\; \sum_j \lambda_j\,\min_h(\widetilde{c}_h + S_{hj}) - \sum_i \rho_i\, g(x_i). \tag{36}$$

$$\langle \mu, v^{\checkmark}(\cdot,0)\rangle - \langle \nu, g^{\checkmark}\rangle \;=\; \sum_i \rho_i\,\min_k(\bar{c}_k + S_{ik}) - \sum_j \lambda_j\, g^{\checkmark}(y_j). \tag{37}$$

To conclude the proof, it is then enough to show, in view of (35), (36), (37) that

$$\widetilde{c}_i = g(x_i) \qquad \text{and} \qquad \bar{c}_j = g^{\checkmark}(y_j) \qquad \text{for any } i,\, j. \tag{38}$$

We just know, by the very definition of $g^{\checkmark}$, that $\widetilde{c}_i \geq g(x_i)$, $\bar{c}_j \geq g^{\checkmark}(y_j)$. If B is the functional on $\mathbf{M}$ corresponding to the measure γ satisfying (35), we have, according to (23), (25)

$$\begin{aligned} -\widetilde{c}_i + \min_h(\widetilde{c}_h + S_{hj}) < S_{ij} \;&\Rightarrow\; B_{ij} = 0 \\ \min_k(\bar{c}_k + S_{ik}) - \bar{c}_j < S_{ij} \;&\Rightarrow\; B_{ij} = 0. \end{aligned}$$

Taking into account that the matrix B satisfies the condition (17), we deduce that for any i there is a j with

$$-\widetilde{c}_i + \min_h(\widetilde{c}_h + S_{hj}) = S_{ij}$$

and for any j there is an i with

$$\min_k(\bar{c}_k + S_{ik}) - \bar{c}_j = S_{ij}.$$

Formula (38) then comes from Corollary 6.3. □

Proof of Theorem 1.1. Let

$$\mu_k = \sum_{i=1}^{n_k} \rho_i^k \delta_{x_k^n}, \qquad \nu_k = \sum_{j=1}^{m_k} \lambda_j^k \, \delta_{y_j^k}$$

be sequences narrowly converging to μ and ν, respectively. We set

$$g_k(x) = \min_i \left\{ \widetilde{c}_j^k + 2\,\omega(|x_j^k - x|) \right\} \text{ and } g_k^{\checkmark}(x) = \min_j \left\{ \bar{c}_j^k + 2\,\omega(|y_j^k - x|) \right\}$$

with $\widetilde{c}_1^k, \cdots, \widetilde{c}_{n_k}^k$ such that the solution v_k to (HJ+) with datum g_k at $t = 0$ (resp. the solution $v_k^{\checkmark}$ to (HJ−) with datum $g_k^{\checkmark}$ at $t = T$) satisfy

$$S(\mu_k \times \delta_0, \nu_k \times \delta_T) = -\langle \mu_k, g_k \rangle + \langle \nu_k, v_k(\cdot, T) \rangle \qquad \text{for any } k \in \mathbb{N}.$$

$$(\text{resp. } S(\mu_k \times \delta_0, \nu_k \times \delta_T) = \langle \mu_k, u_k^{\checkmark}(\cdot, 0) \rangle - \langle \nu_k, g^{\checkmark} \rangle \qquad \text{for any } k \in \mathbb{N}.)$$

The sequences g_k, $\widetilde{g}_k$ being equibounded and equicontinuous, uniformly converges in $\mathbb{T}^N \times [0, T]$, up to subsequences, to functions g, $\widetilde{g}$, respectively. Accordingly, by basic stability properties of viscosity solutions, v_k uniformly converges to the solution v of (HJ+) with datum g at $t = 0$, and $v_k^{\checkmark}$ to the solution $v^{\checkmark}$ with datum $g^{\checkmark}$ at $t = T$. By Lemma 4.2

$$\liminf S(\mu_k \times \delta_0, \nu_k \times \delta_T) \geq S(\mu \times \delta_0, \nu \times \delta_T),$$

by the uniform convergence of v_k to v, and the narrow convergence of μ_k to μ and of ν_k to ν, we have

$$\langle \mu_k, v_k(\cdot, 0) \rangle \to \langle \mu, v(\cdot, 0) \rangle \quad \text{and} \quad \langle \nu_k, g_k \rangle \to \langle \nu, g \rangle$$

Summing up, we can write

$$\begin{aligned} S(\mu \times \delta_0, \nu \times \delta_T) &\leq \liminf S(\mu_k \times \delta_0, \nu_k \times \delta_T) = \lim -\langle \mu_k, g_k \rangle + \langle \nu_k, v_k(\cdot, T) \rangle \\ &= -\langle \mu, g \rangle + \langle \nu, v(\cdot, T) \rangle, \\ S(\mu \times \delta_0, \nu \times \delta_T) &\leq \liminf S(\mu_k \times \delta_0, \nu_k \times \delta_T) = \lim \langle \mu_k, v_k^{\checkmark}(\cdot, 0) \rangle - \langle \nu_k, g_k^{\checkmark} \rangle \\ &= \langle \mu, v^{\checkmark}(\cdot, 0) \rangle - \langle \nu, g^{\checkmark} \rangle. \end{aligned}$$

This concludes the proof. □

The next result can be seen as a partial converse of Theorem 1.1, see also [3] Proposition 3. Some connections can be found with initial conditions in Mean Field systems.

Proposition 6.4.

(i) *Given a continuous function g and a probability measure ν, there exists a probability measure μ such that the solution v of* (HJ+), (D+) *satisfies*

$$S(\mu \times \delta_0, \nu \times \delta_T) = \langle \nu, v(\cdot, T) \rangle - \langle \mu, g \rangle. \tag{39}$$

(ii) *Given a continuous function $g^{\checkmark}$ and a probability measure μ, there exists a probability measure ν such that the solution $v^{\checkmark}$ of* (HJ−), (D−) *satisfies*

$$S(\mu \times \delta_0, \nu \times \delta_T) = \langle \mu, v^{\checkmark}(\cdot, 0) \rangle - \langle \nu, g^{\checkmark} \rangle. \tag{40}$$

Proof. We prove (i): We assume that ν is approximated in the narrow sense by a sequence ν_k with

$$\nu_k = \sum_{j=1}^{m_k} \lambda_j^k \, \delta_{y_j^k},$$

where λ_j^k are positive coefficients with $\sum_j \lambda_j^k = 1$ for any k and y_j^k are points in $\mathbb{T}^N$. For any y_j^k we select x_j^k in $\mathbb{T}^N$ with

$$g(x_j^k) + S(x_j^k, 0, y_j^k, T) = v(y_j^k, T). \tag{41}$$

We define $\mu_k = \sum_{j=1}^{m_k} \lambda_j^k \, \delta_{x_j^k}$ and by (41) we get $S(\mu_k, \nu_k) = \langle \nu_k, v(\cdot, T) \rangle - \langle \mu_k, g \rangle$.

We get the assertion exploiting Lemma 4.2. The proof of (ii) goes along the same lines with obvious adaptations. □

A. Minimal action

For completeness of the presentation we give some established results on the minimization of the action functional over absolutely continuous curves.

We start recalling the following compactness criterion with respect to the weak topology in $AC(0,T)$:

Proposition A.1. *Let ξ_n be a sequence in $AC(0,T)$ satisfying*

- *ξ_n is equibounded in L^1;*
- *there exists a function $\theta : [0, +\infty) \to \mathbb{R}$ superlinear such that*

$$\int_0^T \theta(|\dot{\xi}_n|) \, dt < M \qquad \text{for any } n.$$

then ξ_n is weakly convergent in $AC(0,T)$, up to subsequences.

We derive:

Proposition A.2. *Let x, y be points of $\mathbb{T}^N$. The optimization problem*

$$\inf \int_0^T L(\xi, \dot{\xi}) \, dt \tag{42}$$

over the absolute continuous curves in $\mathbb{T}^N$ with $\xi(0) = x$, $\xi(T) = y$ admits a minimizer.

Proof. The convexity of $L(x,q)$ in q is equivalent to the integral functional in the statement being lower semicontinuous with respect to the weak topology in $AC(0,T)$. Let ξ_n be a minimizing sequence then

$$M \geq \int_0^T L(\xi_n, \dot{\xi}_n) \, dt \geq \int_0^T \theta(|\dot{\xi}|) \, dt.$$

This first imply that the derivatives are equibounded in L^1 and, since we have fixed the initial point x, the ξ_n's are bounded and so bounded in L^1. So we can apply the compactness criterion in Proposition A.1 to get that the ξ_n converges weakly in $AC(0,T)$. We can therefore conclude the proof exploiting the lower semicontinuity property of the functional. □

We proceed establishing a property of equiLipschitz continuity for the minimizers of (42).

Lemma A.3. *Given $h > 0$, there exists $\ell_h > 0$ such that for any x, y in $\mathbb{T}^N$, t_1, t_2 in $[0, +\infty)$ with $t_2 - t_1 \geq h$, any curve ξ realizing $S(x, t_1, y, t_2)$ satisfies*

$$|\dot{\xi}(t)| \leq \ell_h \qquad \text{for a.e. } t \in [t_1, t_2].$$

Proof. We follow the proof of [4, Theorem 16.18], the unique new point being that the Lipschitz constant of the minimizing curves can be taken uniformly with respect to x, y, t_1, t_2, as indicated in the statement.

We fix such a minimizer ξ and claim that an $\ell_0 > 0$ can be found with

$$|\dot{\xi}(t)| \leq \ell_0 \qquad \text{in a subset of } [t_1, t_2] \text{ with positive measure.} \tag{43}$$

We set $M_h = \max\left\{L(x,q) \mid (x,q) \in \mathbb{T}^N \times B\left(0, \dfrac{\mathrm{diam}\mathbb{T}^N}{h}\right)\right\}$, so that

$$\begin{aligned} S(x, t_1, y, t_2) &\leq \int_{t_1}^{t_2} L(x + (y-x)t/(t_2 - t_1), (y-x)/(t_2 - t_1))\, dt \qquad (44)\\ &\leq M_h\,(t_2 - t_1). \end{aligned}$$

Bearing in mind the coercive character of L, we can select ℓ_0 with

$$|q| > \ell_0 \;\Rightarrow\; L(z,q) > M_h \quad \text{for any } z \in \mathbb{T}^N.$$

If $|\dot{\xi}(t)| > \ell_0$ a.e. then

$$S(x, t_1, y, t_2) = \int_{t_1}^{t_2} L(\xi, \dot{\xi})\, dt > M_h\,(t_2 - t_1)$$

in contradiction with (44). This proves the claim.

We take for granted the validity in our setting, as proved in [4], of the first order condition for optimality named after Erdmann, which reads

$$L(\xi(t), \dot{\xi}(t)) - p(t) \cdot \dot{\xi}(t) = \lambda \tag{45}$$

for a suitable constant λ and a.e. $t \in [0,T]$ and a suitable $p(t) \in \partial_q L(\xi(t), \dot{\xi}(t))$. We deduce by (43) that

$$\lambda \leq \max\{|L(x,q)| \mid (x,q) \in \mathbb{T}^N \times B(0, \ell_0)\} + \ell_0\, R, \tag{46}$$

where R is a Lipschitz constant of $L(x, \cdot)$ in $\mathbb{T}^N \times B(0, \ell_0)$ for any $x \in \mathbb{T}^N$.

We derive from (45) and the fact that $p(t) \in \partial_q L(\xi(t), \dot{\xi}(t))$ that

$$H(\xi(t), p(t)) = \lambda \qquad \text{for a.e. } t.$$

This gives an a.e. bound C_1 of $p(t)$, where C_1 depends only on H, L and ℓ_0, by (46). Taking into account that L has superlinear growth at infinity by (A3), it follows that

$$|\dot{\xi}(t)| \leq \ell_h \qquad \text{for a suitable } \ell_h \text{ and a.e. } t \in [0,T]$$

uniformly with respect to x, y, as was asserted. □

We are now in position to prove Proposition 3.1.

Proof of Proposition 3.1. We assume for the sake of simplicity that $T \geq 1$. Given $h \in (0,T)$, we consider (x_1, y_1), (x_2, y_2) in $\mathbb{T}^N \times \mathbb{T}^N$ and (t_1, r_1), (t_2, r_2) in $[0,T] \times [0,T]$ with $(r_1 - t_1)$ and $(r_2 - t_2)$ greater than or equal to h.
We estimate $|S(x_1, t_1, y_1, r_1) - S(x_2, t_2, y_2, r_2)|$ via the decomposition

$$\begin{aligned} |S(x_1,t_1,y_1,r_1) - S(x_2,t_2,y_2,r_2)| &\leq |S(x_1,t_1,y_1,r_1) - S(x_1,t_1,y_2,r_1)| \\ + \quad |S(x_1,t_1,y_2,r_1) - S(x_2,t_1,y_2,r_1)| &+ |S(x_2,t_1,y_2,r_1) - S(x_2,t_2,y_2,r_2)|. \end{aligned}$$

Assume that $S(x_1, t_1, y_1, r_1) \geq S(x_1, t_1, y_2, r_1)$, and let ξ be a curve realizing the minimal action $S(x_1, t_1, y_2, r_1)$, we have

$$\int_{t_1}^{r_1} L\left(\xi + \frac{1}{r_1 - t_1}(t - t_1)(y_1 - y_2), \dot{\xi}(t) + \frac{1}{r_1 - t_1}(y_1 - y_2)\right) dt \geq S(x_1, t_1, y_1, r_1)$$

By the previous lemma $|\dot{\xi}| \leq \ell_h$ for a.e. $t \in [0,T]$ and a suitable constant ℓ_h, so that we get for $t \in [t_1, r_1]$

$$\begin{aligned} L\left(\xi + \frac{1}{r_1 - t_1}(t - t_1)(y_1 - y_2), \dot{\xi}(t) + \frac{1}{r_1 - t_1}(y_1 - y_2)\right) - L(\xi(t), \dot{\xi}(t)) \\ \leq \tau\left(\frac{2T|y_1 - y_2|}{h}\right), \end{aligned}$$

where τ is an uniform continuity modulus of L in $\mathbb{T}^N \times B\left(0, \ell_h + \frac{\operatorname{diam} \mathbb{T}^N}{h}\right)$, and consequently

$$|S(x_1, t_1, y_1, r_1) - S(x_1, t_1, y_2, r_1)| \leq T\,\tau\left(\frac{2T|y_1 - y_2|}{h}\right). \tag{47}$$

Arguing as above we further obtain

$$|S(x_1, t_1, y_2, r_1) - S(x_2, t_1, y_2, r_1)| \leq T\,\tau\left(\frac{2T|x_1 - x_2|}{h}\right). \tag{48}$$

Finally, assume $S(x_2, t_1, y_2, r_1) \geq S(x_2, t_2, y_2, r_2)$ and let ξ be a curve realizing the minimal action $S(x_2, t_2, y_2, r_2)$.

Then we have

$$\int_{t_1}^{r_1} L\left(\xi\left(t_2 - a\,t_1 + a\,t\right), a\,\dot{\xi}\left(t_2 - a\,t_1 + a\,t\right)\right)\,dt \geq S(x_1, t_1, y_1, r_1)$$

where $a = \frac{r_2 - t_2}{r_1 - t_1}$, and via change of integration variable

$$\int_{t_1}^{r_1} L\left(\xi\left(t_2 - a\,t_1 + a\,t\right), a\,\dot{\xi}\left(t_2 - a\,t_1 + a\,t\right)\right)\,dt = \frac{1}{a}\,S(x_2, t_2, y_2, r_2),$$

so that

$$\begin{aligned} |S(x_1, t_1, y_1, r_1) - S(x_2, t_2, y_2, r_2)| &\leq \left(\frac{1}{a} - 1\right)\,S(x_2, t_2, y_2, r_2) \\ &\leq \frac{(r_1 - t_1) + (r_2 - t_2)}{h}\,S(x_2, t_2, y_2, r_2) \end{aligned} \tag{49}$$

From (47), (48), (49), and the fact that S is bounded in $\mathbb{T}^N \times [0, T]$, we derive the assertion. □

B. Finite dimensional Hahn-Banach theorem

We recall the Hahn-Banach theorem in a finite dimensional space $\mathbb{R}^M$. The characters are:

- a sublinear function $f : \mathbb{R}^M \to \mathbb{R}$, namely positive homogeneous and subadditive;
- a linear functional $P : W \to \mathbb{R}$, with W vector subspace of $\mathbb{R}^M$.

Theorem B.1. *Assume that*

$$P(x) \leq f(x) \quad \textit{for all } x \in W, \tag{50}$$

then there exists an extension $\bar{p}$ of P in $\mathbb{R}^M$ such that

$$\bar{p} \cdot x \leq f(x) \quad \textit{for all } x \in \mathbb{R}^M.$$

The proof of the theorem is based on an induction argument plus the following lemma:

Lemma B.2. *Let P be a linear functional on W satisfying* (50)*, given any nonzero vector x_0 in $\mathbb{R}^M \setminus W$, there exists a constant c such that*

$$P(x) + \lambda\,c \leq f(x + \lambda\,x_0) \qquad \textit{for all } x \in W,\ \lambda \in \mathbb{R}$$

Proof. Given x, y in W, we have

$$P(x - y) \leq f(x - y) \leq f(x - x_0) + f(x_0 - y)$$

and accordingly

$$\begin{aligned} P(x) - f(x - x_0) &\leq P(y) + f(x_0 - y) \\ \sup_{x \in W} P(x) - f(x - x_0) &\leq \inf_{y \in W} P(y) + f(x_0 - y). \end{aligned}$$

If we take a constant c with

$$\sup_{x\in W} P(x) - f(x - x_0) \leq c \leq \inf_{y\in W} P(y) + f(x_0 - y)$$

and $\lambda \geq 0$ we get $\qquad \lambda\, c \leq P(\lambda\, y) + f(\lambda\, x_0 - \lambda\, y)$

and $$P(-\lambda\, y) + \lambda\, c \leq f(\lambda\, x_0 - \lambda\, y) \tag{51}$$

for any y in W, if instead $\lambda < 0$, we have

$$\lambda\, c \leq \lambda\, P(x) + f(-\lambda\, x + \lambda\, x_0)$$

and $$P(-\lambda\, x) + \lambda\, c \leq f(-\lambda\, x + \lambda\, x_0) \tag{52}$$

for any $x \in W$. Taking into account that W is a vector subspace, (51) and (52) give

$$P(x) + \lambda\, c \leq f(x + \lambda\, x_0) \qquad \text{for all } x \in W,\ \lambda \in \mathbb{R},$$

as was asserted. □

The lemma allows constructing the extension $\bar{p}$ dimension by dimension, if the dimension of the ambient space is finite then this can be formalized through an induction argument, otherwise Zorn lemma is needed. A corollary that we use in Section 5 is:

Corollary B.3. *Let $C \subset \mathbb{R}^M$ be a closed convex set with nonempty interior. Let W be a vector subspace possessing a nonempty intersection with the interior of C, and P a linear functional on W, which is in addition bounded from above in $C \cap W$. There is an extension of P, denoted by $\bar{p}$, on $\mathbb{R}^M$ with*

$$\sigma_{C\cap W}(P) = \sigma_C(\bar{p}).$$

Proof. Let $x_0 \in \operatorname{int} C \cap W$, we have

$$\begin{aligned} \sigma_{W\cap(C-x_0)}(P) &= \sigma_{W\cap C} - P(x_0) \\ \sigma_{C-x_0}(\bar{p}) &= \sigma_C(\bar{p}) - \bar{p}\cdot x_0 \end{aligned}$$

so that we can assume without loosing generality that 0 is an interior point of C. The gauge function γ_C can be therefore continuously defined. Notice that in this setting $\sigma_{C\cap W}(P) =: s_0 > 0$. We get

$$\frac{x}{\gamma_C(x)} \in W \cap C \quad \text{whenever } x \in W$$

and consequently $\qquad P(x) \leq s_0\, \gamma_C(x) \qquad \text{for } x \in W.$

We are therefore in position to apply Theorem B.1 with W as linear subspace, P as linear functional on W and $x \mapsto s_0\, \gamma_C(x)$ as sublinear function dominating P. We deduce that there exists a linear functional $\bar{p}$ extending P with

$$\bar{p}\cdot x \leq s_0\, \gamma_C(x) \qquad \text{for any } x \in \mathbb{R}^M.$$

This implies that $\sigma_C(\bar{p}) \leq s_0$ and consequently, the opposite inequality being immediate, $\sigma_C(\bar{p}) = s_0$. □

References

[1] L. Ambrosio, N. Gigli, G. Savaré: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel (2005).

[2] P. Bernard, B. Buffoni: *Optimal mass transportation and Mather theory*, J. Eor. Math. Soc. 9 (2007) 85–121.

[3] P. Bernard, B. Buffoni: *Weak KAM pairs and Monge-Kantorovich duality*, in: *Asymptotic Analysis and Singularities – Elliptic and Parabolic PDEs and Related Problems*, H. Kozono et al. (eds.), Advanced Studies in Pure Mathematics 47/2, Mathematical Society of Japan, Tokyo (2007) 397–420.

[4] F. Clarke: *Functional Analysis, Calculus of Variations and Optimal Control*, Graduate Texts in Mathematics 264, Springer, New York (2013).

[5] A. Fathi: *Weak Kam Theorem in Lagrangian Dynamics*, Cambridge Studies in Advanced Mathematics 88, Cambridge University Press, Cambridge (2004).

[6] T. Mikami: *A simple proof of duality theorem for Monge-Kantorovich problem*, Kodai Math. J. 29 (2006) 1–4.

[7] T. Mikami, M. Thieullen: *Duality theorem for the stochastic optimal control problem*, Stochastic Processes Appl 116 (2006) 1815–1835.

[8] K. R. Parthasarathy: *Probability Measures on Metric Spaces*, Probability and Mathematical Statistics, Academic Press, New York (1967).

[9] A. Siconolfi: *Metric character of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. 355 (2003) 1987–2009.

[10] C. Villani: *Topics in Optimal Transportation*, Graduate Studies in Mathematics 58, American Mathematical Society, Providence (2004).