

Optimal Ballistic Transport and Hopf-Lax Formulae on Wasserstein Space

Nassif Ghoussoub*

*Department of Mathematics, University of British Columbia
Vancouver BC Canada V6T 1Z2
nassif@math.ubc.ca*

June 9, 2017

Abstract

We investigate the optimal mass transport problem associated to the following “ballistic” cost functional on phase space $M \times M^*$,

$$b_T(v, x) := \inf\{\langle v, \gamma(0) \rangle + \int_0^T L(\gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M); \gamma(T) = x\},$$

where $M = \mathbb{R}^d$, $T > 0$, and $L : M \times M \rightarrow \mathbb{R}$ is a Lagrangian that is jointly convex in both variables. Under suitable conditions on the initial and final probability measures, we use convex duality à la Bolza and Monge-Kantorovich theory to lift classical Hopf-Lax formulae from state space to Wasserstein space. This allows us to relate optimal transport maps for the ballistic cost to those associated with the fixed-end cost defined on $M \times M$ by

$$c_T(x, y) := \inf\left\{\int_0^T L(\gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M); \gamma(0) = x, \gamma(T) = y\right\}.$$

We also point to links with the theory of mean field games.

1 Introduction and main results

Given a cost functional $c(y, x)$ on some product measure space $X_0 \times X_1$, and two probability measures μ on X_0 and ν on X_1 , we consider the problem of optimizing the total cost of transport plans and its corresponding dual principle as formulated by Kantorovich

$$\inf\left\{\int_{X_0 \times X_1} c(y, x) d\pi; \pi \in \mathcal{K}(\mu, \nu)\right\} = \sup\left\{\int_{X_1} \varphi_1(x) d\nu(x) - \int_{X_0} \varphi_0(y) d\mu(y); \varphi_1, \varphi_0 \in \mathcal{K}(c)\right\},$$

where $\mathcal{K}(\mu, \nu)$ is the set of probability measures π on $X_0 \times X_1$ whose marginal on X_0 (resp. on X_1) is μ (resp., ν) (*the transport plans*), and where $\mathcal{K}(c)$ is the set of functions $\varphi_1 \in L^1(X_1, \nu)$ and $\varphi_0 \in L^1(X_0, \mu)$ such that

$$\varphi_1(x) - \varphi_0(y) \leq c(y, x) \quad \text{for all } (y, x) \in X_0 \times X_1.$$

*Partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. This work was initiated during a visit of the author to the Schrödinger Institute in Vienna during the month of June 2016.

The pairs of functions in $\mathcal{K}(c)$ can be assumed to satisfy

$$\varphi_1(x) = \inf_{y \in X_0} c(y, x) + \varphi_0(y) \quad \text{and} \quad \varphi_0(y) = \sup_{x \in X_1} \varphi_1(x) - c(y, x).$$

They will be called *admissible Kantorovich potentials*, and for reasons that will become clear later, we shall say that φ_0 (resp., φ_1) is an initial (resp., final) Kantorovich potential.

The original Monge problem dealt with the cost $c(y, x) = |x - y|$ ([15], [16], [9], [21], [22]) and was constrained to those probabilities in $\mathcal{K}(\mu, \nu)$ that are supported by graphs of measurable maps from X to Y pushing μ onto ν . Brenier [6] considered the important quadratic case $c(x, y) = |x - y|^2$. This was followed by a large number of results addressing costs of the form $f(x - y)$, where f is either a convex or a concave function [13]. With a purpose of connecting mass transport with Mather theory, Bernard and Buffoni [5] considered dynamic cost functions on a given compact manifold M , that deal with fixed end-points problems of the following type:

$$c_T(y, x) := \inf \left\{ \int_0^T L(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M); \gamma(0) = y, \gamma(T) = x \right\}, \quad (1)$$

where $[0, T]$ is a fixed time interval, and $L : TM \rightarrow \mathbb{R} \cup \{+\infty\}$ is a given Lagrangian that is convex in the second variable of the tangent bundle TM . Fathi and Figalli [11] eventually dealt with the case where M is a non-compact Finsler manifold. Note that standard cost functionals of the form $f(|x - y|)$, where f is convex, are particular cases of the dynamic formulation, since they correspond to Lagrangians of the form $L(t, x, p) = f(p)$.

In this paper, we shall consider the “*ballistic cost function*,” which is defined on phase space $M^* \times M$ by,

$$b_T(v, x) := \inf \left\{ \langle v, \gamma(0) \rangle + \int_0^T L(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M); \gamma(T) = x \right\}, \quad (2)$$

where M is a Banach space and M^* is its dual. The associated transport problems will be

$$\overline{B}_T(\mu_0, \nu_T) := \sup \left\{ \int_{M^* \times M} b_T(v, x) d\pi; \pi \in \mathcal{K}(\mu_0, \nu_T) \right\}, \quad (3)$$

and

$$\underline{B}_T(\mu_0, \nu_T) := \inf \left\{ \int_{M^* \times M} b_T(v, x) d\pi; \pi \in \mathcal{K}(\mu_0, \nu_T) \right\}, \quad (4)$$

where μ_0 (resp., ν_T) is a given probability measure on M^* (resp., M). Note that when $T = 0$, we have $b_0(x, v) = \langle v, x \rangle$, which is exactly the case considered by Brenier [6], that is

$$\overline{W}(\mu_0, \nu_0) := \sup \left\{ \int_{M^* \times M} \langle v, x \rangle d\pi; \pi \in \mathcal{K}(\mu_0, \nu_0) \right\}, \quad (5)$$

and

$$\underline{W}(\mu_0, \nu_0) := \inf \left\{ \int_{M^* \times M} \langle v, x \rangle d\pi; \pi \in \mathcal{K}(\mu_0, \nu_0) \right\}, \quad (6)$$

making (3) a suitable dynamic version of the Wasserstein distance.

The assumptions on the Lagrangian that we use are as follows:

(A1) $L : M \times M \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, proper and lower semi-continuous in both variables.

(A2) The set $F(x) := \{p; L(x, p) < \infty\}$ is non-empty for all $x \in M$, and for some $\rho > 0$, we have $\text{dist}(0, F(x)) \leq \rho(1 + |x|)$ for all $x \in M$.

(A3) For all $(x, p) \in M \times M$, we have $L(x, p) \geq \theta(\max\{0, |p| - \alpha|x|\}) - \beta|x|$, where α, β are constants, and θ is a coercive, proper, non-decreasing function on $[0, \infty)$.

The associated Hamiltonian on $[0, T] \times M \times M^*$ is defined by

$$H(t, x, q) = \sup_{p \in M} \{ \langle p, q \rangle - L(t, x, p) \}.$$

We shall assume throughout that $M = M^* = \mathbb{R}^d$, while preserving –for pedagogical reasons– the notational distinction between the state space and its dual. These conditions on the Lagrangian make sure that the Hamiltonian H is finite, concave in x and convex in q , hence locally Lipschitz. Moreover, we have

$$\psi(x) - (\gamma|x| + \delta)|q| \leq H(x, q) \leq \varphi(q) + (\alpha|q| + \beta)|x| \text{ for all } x, q \text{ in } M \times M^*, \quad (7)$$

where $\alpha, \beta, \gamma, \delta$ are constants, φ is finite and convex and ψ is finite and concave (see [19]).

We note that under these conditions, the cost $(x, y) \rightarrow c(t, x, y)$ is convex proper and lower semi-continuous on $M \times M$. But the cost b_T is nicer in many ways. For one, it is everywhere finite and locally Lipschitz continuous on $[0, \infty) \times M \times M^*$. Moreover, we shall consider suitable solutions of the Hamilton-Jacobi equation,

$$\begin{cases} \partial_t V + H(t, x, \nabla_x V) &= 0 \text{ on } [0, T] \times M, \\ V(0, x) &= V_0(x), \end{cases} \quad (8)$$

which are formally given by the formula

$$V(t, x) = \inf \left\{ V_0(\gamma(0)) + \int_0^t L(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M); \gamma(t) = x \right\}, \quad (9)$$

as well as the following dual Hamilton-Jacobi equation:

$$\begin{cases} \partial_t W - H(\nabla_v W, v) &= 0 \text{ on } [0, T] \times M^*, \\ W(T, v) &= W_T(v), \end{cases} \quad (10)$$

whose variational solution is given by

$$W(t, v) = \sup \left\{ W_T(\gamma(T)) - \int_0^t \tilde{L}(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M^*); \gamma(0) = v \right\}, \quad (11)$$

where the Lagrangian \tilde{L} is defined on $M^* \times M^*$ by

$$\tilde{L}(t, v, q) := L^*(t, q, v) = \sup \{ \langle v, y \rangle + \langle p, q \rangle - L(t, y, p); (y, p) \in M \times M \}.$$

While the above formula for the “solutions” V (resp., W) of (8) (resp., (10)) is natural, it doesn’t often generate solutions with suitable regularity properties, unless more conditions are imposed on the Lagrangian. Bernard and Buffoni [5] imposed conditions that allowed for the consideration of so-called continuous *viscosity solutions*. There are however instances where our convex setting is completely satisfactory under the above assumptions on L . Indeed, if V_0 (resp., W_T) is convex (resp., concave), then the solution $V(t, \cdot)$ (resp., $W(t, \cdot)$) is convex (resp., concave) for each t , and convex subdifferentiability then provides a good alternative to regularity. This is the case for $b(t, v, x)$, which remarkably satisfies both equations (in the sense of convex analysis), that is

$$\begin{aligned} \partial_t b + H(t, x, \partial_x b) &= 0 \quad \text{on } [0, T] \times M, & b(0, v, x) &= \langle v, x \rangle. \\ \partial_t b - H(t, \partial_v b, v) &= 0 \quad \text{on } [0, T] \times M^*, & b(T, v, x) &= \langle v, x \rangle. \end{aligned}$$

Unfortunately, as we shall see below, the mass transport problems that we consider lead to an opposite situation, where the initial function V_0 is concave, a property that is not propagated forward by L , and the terminal function W_T is convex, which is not propagated backward by \tilde{L} . We shall therefore only consider *variational solutions* for (8) (resp., 10) meaning those that has the form (9) and (11) respectively.

The following duality will be proved in Section 5.

Theorem 1 Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (A1), (A2) and (A3), and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M). The following duality holds:

1. Assume μ_0 is absolutely continuous with respect to Lebesgue measure, then

$$\underline{B}_T(\mu_0, \nu_T) = \sup \left\{ \int_M V(T, x) d\nu_T(x) + \int_{M^*} \tilde{V}_0(v) d\mu_0(v); V_0 \text{ concave on } M \text{ \& } V \text{ solution of (8)} \right\}, \quad (12)$$

where \tilde{V}_0 is the concave Legendre transform of V_0 , i.e.,

$$\tilde{V}_0(v) = \inf \{ \langle v, y \rangle - V_0(y); y \in M \}.$$

2. Assume ν_T is absolutely continuous with respect to Lebesgue measure, then

$$\overline{B}_T(\mu_0, \nu_T) = \inf \left\{ \int_M W_T^*(x) d\nu_T(x) + \int_{M^*} W(0, v) d\mu_0(v); W_T \text{ convex on } M^* \text{ \& } W \text{ solution of (10)} \right\}, \quad (13)$$

where W_T^* is the convex Legendre transform of W_T , i.e.,

$$W_T^*(x) = \sup \{ \langle v, x \rangle - W_T(v); v \in M^* \}.$$

In order to investigate the support of optimal transport plans, we shall need a subgradient form of Hamiltonian dynamics, a solution of which over a time interval $[0, T]$ is any pair of C^1 arcs $(x(t), v(t))$ such that for almost every $t \in [0, T]$,

$$\begin{cases} \dot{x}(t) & \in \partial_v H(x(t), v(t)) \\ -\dot{v}(t) & \in \partial_x H(x(t), v(t)). \end{cases} \quad (14)$$

The associated Hamiltonian flow is the one-parameter family of, possibly set-valued, mappings $(\varphi_t^H)_t$,

$$\varphi_t^H(x_0, v_0) = \{(x, v); \exists \text{ a Hamiltonian trajectory starting at } (x_0, v_0) \text{ with } (x(t), v(t)) = (x, v)\}. \quad (15)$$

We shall prove the following:

Theorem 2 Assume L satisfies hypothesis (A1), (A2) and (A3), and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M).

1. If μ_0 is absolutely continuous with respect to Lebesgue measure, then there exists a probability measure π_0 on $M^* \times M$, and a concave function $k : M \rightarrow \mathbb{R}$ such that

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, x) d\pi_0, \quad (16)$$

and π_0 is supported on the possibly set-valued map $v \rightarrow \pi^* \varphi_T^H(\nabla \tilde{k}(v), v)$, with $\pi^* : M \times M^* \rightarrow M$ being the canonical projection, and \tilde{k} is the concave Legendre transform of k .

2. Assume ν_T is absolutely continuous with respect to Lebesgue measure, then there exists a probability measure $\tilde{\pi}_0$ on $M^* \times M$, and a convex function $h : M^* \rightarrow \mathbb{R}$ such that

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, x) d\tilde{\pi}_0, \quad (17)$$

and $\tilde{\pi}_0$ is supported on the possibly set-valued map $x \rightarrow \pi^* \varphi_T^{-H^*}(\nabla h(x), x)$, where $\varphi_t^{-H^*}$ is the Hamiltonian flow associated to the Lagrangian $L_*(v, q) = L^*(-q, v)$, namely $H_*(x, q) = -H(-x, q)$.

If we assume further that L is a Tonelli Lagrangian on $M \times M^*$, then the above maps are single-valued and completely solve the Monge version of the mass transport problems, that is

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \pi^* \varphi_T^H(\nabla \tilde{k}(v), v)) d\mu_0(v), \quad (18)$$

and

$$\overline{B}_T(\mu_0, \nu_T) = \int_M b_T(\pi^* \varphi_T^{-H}(\nabla h^*(x), x), x) d\nu_T(x). \quad (19)$$

The above two theorems will follow from the following interpolation result, and what is known about the optimal mass transport

$$C_T(\nu_0, \nu_T) := \inf \left\{ \int_{M \times M} c_T(x, y) d\pi; \pi \in \mathcal{K}(\nu_0, \nu_T) \right\}, \quad (20)$$

where ν_0 and ν_T are two given probability measures on M . We shall also need another cost function on $M^* \times M^*$,

$$\tilde{c}_T(u, v) := \inf \left\{ \int_0^T \tilde{L}(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M^*); \gamma(0) = u, \gamma(T) = v \right\}, \quad (21)$$

and its associated transport

$$\tilde{C}_T(\mu_0, \mu_T) := \inf \left\{ \int_{M^* \times M^*} \tilde{c}_T(x, y) d\pi; \pi \in \mathcal{K}(\mu_0, \mu_T) \right\}. \quad (22)$$

Theorem 3 *Assume that L satisfies hypothesis (A1), (A2) and (A3), and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M).*

1. *If μ_0 is absolutely continuous with respect to Lebesgue measure, then*

$$\underline{B}_T(\mu_0, \nu_T) = \inf \{ \underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}(M) \}. \quad (23)$$

The infimum is attained at some probability measure ν_0 on M , and the initial Kantorovich potential for $C_T(\nu_0, \nu_T)$ is concave.

2. *If ν_T is absolutely continuous with respect to Lebesgue measure, then*

$$\overline{B}_T(\mu_0, \nu_T) = \sup \{ \overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu); \mu \in \mathcal{P}(M^*) \}. \quad (24)$$

The supremum is attained at some probability measure μ_T on M^ , and the final Kantorovich potential for $\tilde{C}_T(\mu_0, \mu_T)$ is convex.*

The above interpolation formulas can be seen as extensions of those by Hopf-Lax on state space to Wasserstein space. Indeed, for any convex and lower semi-continuous function g , the associated value function

$$\begin{cases} V_g(t, x) &= \inf \left\{ g(\gamma(0)) + \int_0^t L(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M); \gamma(t) = x \right\}, \\ V(0, x) &= g(x), \end{cases}$$

can be written as

$$V_g(t, x) = \inf \{ g(y) + c(t, y, x); y \in M \} \quad \text{as well as} \quad V_g(t, x) = \sup \{ b(t, v, x) - g^*(v); v \in M^* \}. \quad (25)$$

In the case where the Lagrangian $L(x, p) = L_0(p)$ is only a function of p , and if H_0 is the associated Hamiltonian, then

$$c_t(y, x) = tL_0\left(\frac{1}{t}|x - y|\right) \quad \text{and} \quad b_t(v, x) = \langle v, x \rangle - tH_0(v), \quad (26)$$

and (25) is nothing but the Hopf-Lax formula used to generate solutions for corresponding Hamilton-Jacobi equations. Moreover, when g is the linear functional $g(x) = \langle v, x \rangle$, then $b(t, v, x)$ is itself a solution to two Hamilton-Jacobi equations, since

$$b(t, v, x) = \inf\{\langle v, y \rangle + c(t, y, x); y \in M\} = \sup\{\langle w, x \rangle - \tilde{c}(t, v, w); w \in M^*\}. \quad (27)$$

Note that formulas (23) and (24) can now be seen as extensions of (27) to the space of probability measures, where the Wasserstein distance fill the role of the scalar product.

Surprisingly, the extension of the dual formula

$$c(t, y, x) = \sup\{b(t, v, x) - \langle v, y \rangle; v \in M^*\}. \quad (28)$$

is not always possible. Indeed, we shall see that the key to proving (23) and (24) is that the initial Kantorovich potential for $\underline{B}_T(\mu_0, \nu_T)$ can be taken to be convex, while for $\overline{B}_T(\mu_0, \nu_T)$, the final Kantorovich potential can be assumed to be concave. This contrasts the case of the fixed endpoints cost c_T , which even though it is jointly convex in both variables, one cannot deduce much in terms of the convexity or concavity of the Kantorovich potentials corresponding to $C_T(\nu_0, \nu_T)$. Indeed, if $L(x, v) = \frac{1}{2}|v|^2$, which corresponds to the cost $c(y, x) = \frac{1}{2}|x - y|^2$, the initial Kantorovich potential is then of the form $\varphi_0(y) = g(y) - \frac{1}{2}|y|^2$, where g is a convex function.

Note however, that the optimal transport $C_T(\nu_0, \nu_T)$ corresponding to the initial measure ν_0 obtained via the factorization of ballistic optimal transport problems, has a concave initial Kantorovich potential. This turned out to be a necessary and sufficient condition. In the case where $L(x, v) = \frac{1}{2}|v|^2$, it says that for such a measure the initial Kantorovich potential φ_0 can still be concave, that is $0 \leq D^2 g \leq I$.

Theorem 4 *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (A1), (A2) and (A3). Assume ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure. Then, the following are equivalent:*

1. *The initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave.*
2. *The following formula holds*

$$C_T(\nu_0, \nu_T) = \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (29)$$

Here is an application.

Corollary 1 *Consider the cost $c(y, x) = c_0(x - y)$, where c_0 is a convex function on M and let ν_0, ν_1 be probability measures on M such that the initial Kantorovich potential associated to $C_T(\nu_0, \nu_T)$ is concave. Then, there exist concave functions $\varphi : M \rightarrow \mathbb{R}$ and $\psi : M^* \rightarrow \mathbb{R}$ such that*

$$C_1(\nu_0, \nu_1) - K = \int_M c_0(\nabla\psi \circ \nabla\varphi(y) - y) d\nu_0(y) = \int_M \langle \nabla\tilde{\psi}(y) - \nabla\varphi(y), y \rangle d\nu_0(y), \quad (30)$$

where K is a constant and $\tilde{\psi}$ is the concave Legendre transform of ψ .

The interpolation formula can be seen as a Hopf-Lax formula on Wasserstein space, since for a fixed μ_0 on M^* (resp., fixed ν_T on M), then as a function of the terminal (resp., initial) measure, we have

$$\underline{\mathcal{B}}^{\mu_0}(t, \nu) = \inf\{\underline{\mathcal{U}}^{\mu_0}(\rho) + C_t(\rho, \nu); \rho \in \mathcal{P}(M)\} \text{ and } \overline{\mathcal{B}}^{\nu_T}(t, \mu) = \inf\{\overline{\mathcal{U}}^{\nu_T}(\rho) - \tilde{C}_t(\rho, \mu); \rho \in \mathcal{P}(M^*)\}, \quad (31)$$

where

$$\underline{\mathcal{U}}^{\mu_0}(\rho) = \underline{W}(\mu_0, \rho) \quad \text{and} \quad \overline{\mathcal{U}}^{\nu_T}(\rho) = \overline{W}(\nu_T, \rho).$$

The following Eulerian formulation illustrates best how $\underline{\mathcal{B}}^{\mu_0}(t, \nu)$ and $\overline{\mathcal{B}}^{\nu_T}(t, \mu)$ can be represented as value functionals on Wasserstein space. Indeed, lift the Lagrangian L to the tangent bundle of Wasserstein space via the formula

$$\underline{\mathcal{L}}(\rho, w); = \int_M L(x, w(x)) d\rho(x) \quad \text{and} \quad \tilde{\mathcal{L}}(\rho, w); = \int_{M^*} \tilde{L}(x, w(x)) d\rho(x),$$

where ρ is any probability density on M (resp., M^*) and w is a vector field on M (resp., M^*).

Theorem 5 Assume L satisfies hypothesis (A1), (A2) and (A3), and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M).

1. If μ_0 is absolutely continuous with respect to Lebesgue measure, then

$$\underline{\mathcal{B}}^{\mu_0}(T, \nu) := \underline{B}_T(\mu_0, \nu) = \inf \left\{ \underline{\mathcal{U}}^{\mu_0}(\rho_0) + \int_0^T \mathcal{L}(\rho_t, w_t) dt; \partial_t \rho + \nabla \cdot (\rho w) = 0, \rho_T = \nu \right\}, \quad (32)$$

2. If ν_T is absolutely continuous with respect to Lebesgue measure, then

$$\overline{\mathcal{B}}^{\nu_T}(T, \mu) := \overline{B}_T(\mu, \nu_T) = \sup \left\{ \overline{\mathcal{U}}^{\nu_T}(\rho_T) - \int_0^T \tilde{\mathcal{L}}(\rho_t, w_t) dt; \partial_t \rho + \nabla \cdot (\rho w) = 0, \rho_0 = \mu \right\}. \quad (33)$$

The set of pairs (ρ, w) considered above are such that $t \rightarrow \rho_t \in \mathcal{P}(M)$, and $t \rightarrow w_t(x) \in \text{Lip}(\mathbb{R}^n)$ are paths of Borel vector fields.

One can then ask whether these value functionals also satisfy a Hamilton-Jacobi equation on Wasserstein space such as

$$\begin{cases} \partial_t B + \mathcal{H}(t, \nu, \nabla_\nu B(t, \nu)) = 0, \\ B(0, \nu) = \underline{W}(\mu_0, \nu). \end{cases} \quad (34)$$

Here the Hamiltonian is defined as

$$\mathcal{H}(\nu, \zeta) = \sup \left\{ \int \langle \zeta, \xi \rangle d\nu - \mathcal{L}(\nu, \xi); \xi \in T_\nu^*(\mathcal{P}(M)) \right\}.$$

We note that Ambrosio-Feng [3] have shown recently that –at least in the case where the Hamiltonian is the square– value functionals on Wasserstein space yield a unique *metric viscosity solution* for (34). As importantly, Gangbo-Swiech [14] have shown recently that under certain conditions, value functionals yield solutions to the so-called *Master equations* of mean field games.

Theorem 6 (Gangbo-Swiech) Assume $\mathcal{U}_0 : \mathcal{P}(M) \rightarrow \mathbb{R}$, and $U_0 : M \times \mathcal{P}(M) \rightarrow \mathbb{R}$ are functionals such that $\nabla_x U_0(x, \mu) \equiv \nabla_\mu \mathcal{U}_0(\mu)(x)$ for all $x \in M$, $\mu \in \mathcal{P}(M)$, and consider the value functional,

$$\mathcal{U}(t, \nu) = \inf \left\{ \mathcal{U}_0(\rho_0) + \int_0^t \mathcal{L}(\rho, w) dt; \partial_t \rho + \nabla \cdot (\rho w) = 0, \rho_T = \nu \right\}.$$

Then, there exists $U : [0, T] \times M \times \mathcal{P}(M) \rightarrow \mathbb{R}$ such that

$$\nabla_x U_t(x, \nu) \equiv \nabla_\nu \mathcal{U}_t(\nu)(x) \quad \text{for all } x \in M, \nu \in \mathcal{P}(M),$$

and U satisfies the Master equation below (35).

Applied to the value functional $\underline{\mathcal{B}}^{\mu_0}(t, \nu) := \underline{B}_t(\mu_0, \nu)$, this should then yields the existence for any probabilities μ_0, ν_T , a function $\beta : [0, T] \times M \times \mathcal{P}(M) \rightarrow \mathbb{R}$ such that

$$\nabla_x \beta(t, x, \nu) \equiv \nabla_\nu \underline{\mathcal{B}}^{\mu_0}(t, \nu)(x) \quad \text{for all } x \in M, \nu \in \mathcal{P}(M),$$

and $\rho \in AC^2((0, T) \times \mathcal{P}(M))$ such that

$$\begin{cases} \partial_t \beta + \int \langle \nabla_\nu \beta(t, x, \nu) \cdot \nabla H(x, \nabla_x \beta) \rangle d\nu + H(x, \nabla_x \beta(t, x, \nu)) = 0, \\ \partial_t \rho + \nabla(\rho \nabla H(x, \nabla_x \beta)) = 0, \\ \beta(0, \cdot, \cdot) = \beta_0, \quad \rho(T, \cdot) = \nu_T, \end{cases} \quad (35)$$

where $\beta_0(x, \rho) = \varphi_\rho(x)$, where φ_ρ is the convex function such that $\nabla\varphi_\rho$ pushes μ_0 into ρ .

Needless to say, one would like to consider value functionals on Wasserstein space that are more general than those starting with the Wasserstein distance. One can still obtain such functionals via mass transport by considering more general ballistic costs of the form

$$b_g(T, v, x) := \inf \left\{ g(v, \gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M) \right\}, \quad (36)$$

where $g : M^* \times M \rightarrow \mathbb{R}$ is a suitable function. This, as well as the stochastic counterpart of the ballistic problem and the corresponding ‘‘diffusive’’ master equation, will be considered in a subsequent paper.

I am very grateful to Yann Brenier and Wilfrid Gangbo for several fruitful discussions, and in particular for pointing me towards the connection between value functionals on Wasserstein space and mean field games. I am also indebted to the Schrödinger Institute in Vienna, where most of this work was done during my visit there in June 2016.

2 The Bolza duality and its consequences

We first review properties of the various cost functions that will be needed in the sequel. They are standard and had been studied in detail in various articles by T. Rockafellar [17] and his co-authors [18, 19].

Proposition 2 *Under assumptions (A) on the Lagrangian L , the costs c and b have the following properties:*

1. For each $t \geq 0$, $(x, y) \rightarrow c(t, x, y)$ is convex proper and lower semi-continuous on $M \times M$.
2. For each $t \geq 0$, $v \rightarrow b(t, v, x)$ is concave upper semi-continuous on M^* , while $x \rightarrow b(t, v, x)$ is convex lower semi-continuous on M . Moreover, b is locally Lipschitz continuous on $[0, \infty) \times M \times M^*$.
3. The costs b , c and \tilde{c} are dual to each other in the following sense:
 - For any $(v, x) \in M^* \times M$, we have $b(t, v, x) = \inf\{ \langle v, y \rangle + c(t, y, x); y \in M \}$.
 - For any $(y, x) \in M \times M$, we have $c(t, y, x) = \sup\{ b(t, v, x) - \langle v, y \rangle; v \in M^* \}$.
 - For any $(v, x) \in M^* \times M$, we have $b(t, v, x) = \sup\{ \langle w, x \rangle - \tilde{c}(t, v, w); w \in M^* \}$.

The above statements follow readily from properties of value functions that are consequences of the duality in Bolza’s problem, which is a particular case of the following set up. Consider the path space $A_M^2 = \{u : [0, T] \rightarrow M; \dot{u} \in L_M^2\}$ equipped with the norm

$$\|u\|_{A_M^2} = \left(\|u(0)\|_M^2 + \int_0^T \|\dot{u}\|^2 dt \right)^{\frac{1}{2}}.$$

One way to represent the space A_M^2 is to identify it with the product space $M \times L_M^2$, in such a way that its dual $(A_M^2)^*$ can also be identified with $M \times L_M^2$ via the formula:

$$\langle u, (p_1, p_0) \rangle_{A_M^2, M \times L_M^2} = \langle u(0), p_1 \rangle + \int_0^T \langle \dot{u}(t), p_0(t) \rangle dt, \quad (37)$$

where $u \in A_M^2$ and $(p_1, p_0) \in M \times L_M^2$. We have the following duality formula.

Proposition 3 *Let L be a time-dependent convex Lagrangian on $M \times M$ and let ℓ be a proper convex lower semi-continuous function on $M \times M$. Consider the Lagrangian on $A_M^2 \times (A_M^2)^* = A_M^2 \times (M \times L_M^2)$ defined by*

$$\mathcal{N}(u, p) = \int_0^T L(t, u(t) - p_0(t), -\dot{u}(t)) dt + \ell(u(0) - a, u(T)) \quad (38)$$

where $u \in A_M^2$ and $(p_0(t), a) \in L_M^2 \times M$ represents an element p in the dual of A_M^2 . Then, for any $(v, q) \in A_M^2 \times (A_M^2)^*$ with q of the form $(q_0(t), 0)$, we have

$$\mathcal{N}^*(q, v) = \int_0^T L^*(t, -\dot{v}(t), v(t) - q_0(t)) dt + \ell^*(-v(0), v(T)). \quad (39)$$

Proof: For $(v, q) \in A_M^2 \times (A_M^2)^*$ with q represented by $(q_0(t), 0)$ write:

$$\begin{aligned} \mathcal{N}^*(q, v) = & \sup_{p_1 \in M} \sup_{p_0 \in L_M^2} \sup_{u \in A_M^2} \left\{ \langle p_1, v(0) \rangle + \int_0^T \langle p_0(t), \dot{v}(t) \rangle + \langle q_0(t), \dot{u}(t) \rangle dt \right. \\ & \left. - \int_0^T L(t, u(t) - p_0(t), -\dot{u}(t)) dt - \ell(u(0) - p_1, u(T)) \right\}. \end{aligned}$$

Making a substitution $u(0) - p_1 = a \in M$ and $u(t) - p_0(t) = y(t) \in L_M^2$, we obtain

$$\begin{aligned} \mathcal{N}^*(q, v) = & \sup_{a \in M} \sup_{y \in L_M^2} \sup_{u \in A_M^2} \left\{ \langle u(0) - a, v(0) \rangle - \ell(a, u(T)) \right. \\ & \left. + \int_0^T \{ \langle u(t) - y(t), \dot{v}(t) \rangle + \langle q_0(t), \dot{u}(t) \rangle - L(t, y(t), -\dot{u}(t)) \} dt \right\} \end{aligned}$$

Since \dot{u} and $\dot{v} \in L_M^2$, we have

$$\int_0^T \langle u, \dot{v} \rangle = - \int_0^T \langle \dot{u}, v \rangle + \langle v(T), u(T) \rangle - \langle v(0), u(0) \rangle$$

which implies

$$\begin{aligned} \mathcal{N}^*(q, v) = & \sup_{a \in M} \sup_{y \in L_M^2} \sup_{u \in A_M^2} \left\{ \langle -a, v(0) \rangle + \langle v(T), u(T) \rangle - \ell(a, u(T)) \right. \\ & \left. + \int_0^T [-\langle y(t), \dot{v}(t) \rangle + \langle v(t) - q_0(t), -\dot{u}(t) \rangle - L(t, y(t), -\dot{u}(t))] dt \right\}. \end{aligned}$$

Identify now A_M^2 with $M \times L_M^2$ via the correspondence:

$$\begin{aligned} (b, r) \in M \times L_M^2 & \mapsto b + \int_t^T r(s) ds \in A_M^2 \\ u \in A_M^2 & \mapsto (u(T), -\dot{u}(t)) \in M \times L_M^2. \end{aligned}$$

We finally obtain

$$\begin{aligned} \mathcal{N}^*(q, v) = & \sup_{a \in M} \sup_{b \in M} \{ \langle a, -v(0) \rangle + \langle v(T), b \rangle - \ell(a, b) \} \\ & + \sup_{y \in L_M^2} \sup_{r \in L_M^2} \int_0^T -\langle y(t), \dot{v}(t) \rangle + \langle v(t) - q_0(t), r(t) \rangle - L(t, y(t), r(t)) dt \\ = & \int_0^T L^*(t, -\dot{v}(t), v(t) - q_0(t)) dt + \ell^*(-v(0), v(T)). \end{aligned}$$

Now consider the minimization problems,

$$(\mathcal{P}) \quad \inf \left\{ \int_0^T L(\gamma(s), \dot{\gamma}(s)) ds + \ell(\gamma(0), \gamma(T)); \gamma \in C^1([0, T], M) \right\}, \quad (40)$$

and

$$(\tilde{\mathcal{P}}) \quad \inf \left\{ \int_0^T \tilde{L}(\gamma(s), \dot{\gamma}(s)) ds + \ell^*(\gamma(0), -\gamma(T)); \gamma \in C^1([0, T], M) \right\}. \quad (41)$$

It is clear that $\inf(\mathcal{P}) \geq -\inf(\tilde{\mathcal{P}})$, and that we have equality under very mild conditions (see [17]). In this case, there exists two paths $x(t), v(t)$ in A_M^2 satisfying

$$(\dot{v}(t), v(t)) \in \partial L(x(t), \dot{x}(t)) \quad \text{for a.e. } t, \quad (42)$$

which can also be written in a dual form

$$(\dot{x}(t), x(t)) \in \partial \tilde{L}(v(t), \dot{v}(t)) \quad \text{for a.e. } t, \quad (43)$$

or in a Hamiltonian form as

$$\dot{x}(t) \in \partial_v H(x(t), v(t)) \quad (44)$$

$$-\dot{v}(t) \in \tilde{\partial}_x H(x(t), v(t)), \quad (45)$$

coupled with the boundary conditions

$$(v(0), -v(T)) \in \partial \ell(x(0), x(T)). \quad (46)$$

Note that c_T, \tilde{c}_T, b_T and all value functions V_g for g convex, can be written as in (40), and therefore inherit whatever the duality $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$ provides.

Proposition 4 *The value function V_g is expressed in terms of the costs c and b by the following formulae:*

1. $V_g(t, x) = \inf\{g(y) + c(t, y, x); y \in M\}$.
2. *If g is convex and lower semi-continuous, then $V_g(t, x) = \sup\{b(t, v, x) - g^*(v); v \in M^*\}$.*
3. *For each $t \in [0, +\infty)$, the Legendre transform of the convex function $x \rightarrow V_g(t, x)$ on M is the convex function $w \rightarrow \tilde{V}_{g^*}(t, w)$ given by*

$$\begin{cases} \tilde{V}_{g^*}(t, w) &= \inf\{g^*(\gamma(0)) + \int_0^t \tilde{L}(s, \gamma(s), \dot{\gamma}(s)) ds; u \in C^1([0, T], M); \gamma(t) = w\}, \\ \tilde{V}_{g^*}(0, w) &= g^*(w). \end{cases}$$

4. *For each t , the graph of the subgradient $\partial V_g(t, \cdot)$, i.e.,*

$$\Gamma_g(t) = \{(x, v); v \in \partial V_g(t, x)\}$$

is a globally Lipschitzian manifold of dimension n in $M \times M^$, which depends continuously on t .*

Note that $b(t, v, x) = V_g(t, x)$, when $g_v(y) = \langle v, y \rangle$. In this case, $g_v^*(u) = 0$ if $u = v$ and $+\infty$ if $u \neq v$, which yields that the Legendre dual of $x \rightarrow V_g(t, x) = b(t, v, x)$ is $w \rightarrow \tilde{c}(t, v, w)$. One can also deduce the following.

Proposition 5 *Let g be a proper convex lower semi-continuous function on M .*

1. *If a Hamiltonian trajectory $(x(t), v(t))$ over $[0, T]$ starts with $v(0) \in \partial g(x(0))$, then $v(t) \in \partial V_g(t, x(t))$ for all $t \in [0, T]$.*
2. *Moreover, a pair of arcs $x(t)$ and $v(t)$ gives a Hamiltonian trajectory over $[0, T]$ that starts in the graph of ∂g and ends at a point x, w in the graph of $\partial V_g(T, \cdot)$ if and only if $x(t)$ is optimal in the minimization problem that defines $V_g(t, x)$ and $v(t)$ is optimal in the minimization problem that defines $\tilde{V}_g(t, w)$.*
3. *In particular, the following properties are equivalent:*

- (a) $(-v, w) \in \partial_{y,x} c_T(y, x)$;
- (b) $w \in \partial_x b_T(v, x)$ and $y \in \tilde{\partial}_v b_T(v, x)$.
- (c) There is a Hamiltonian trajectory $(\gamma(t), \eta(t))$ over $[0, T]$ starting at (y, v) and ending at (x, w) .

We finally recall the twist condition.

Definition 6 A cost function c satisfies the twist condition if for each $y \in M$, we have $x = x'$ whenever the differentials $\partial_y c(y, x)$ and $\partial_y c(y, x')$ exist and are equal.

In view of the above proposition, c_T satisfies the twist condition if there is at most one Hamiltonian trajectory starting at a given initial state (v, y) , while the cost b_T satisfies the twist condition if for any given states (v, w) , there is at most one Hamiltonian trajectory starting at v and ending at w .

3 The Hopf-Lax formulas on Wasserstein space

We shall frequently use the following results of Brenier [6] that we summarize in the following proposition.

Proposition 7 (Brenier) If ν_0 (resp., μ_0) is a probability measure on M (resp., M^*) such that ν_0 is absolutely continuous measure with respect to Lebesgue measure, then there exists a unique (up to a constant) convex function h (resp., concave function g) on M such that $(\nabla h)_{\#}\nu_0 = \mu_0$ (resp., $(\nabla g)_{\#}\nu_0 = \mu_0$) and

$$\overline{W}(\nu_0, \mu_0) = \int_M \langle \nabla h(x), x \rangle d\nu_0(x), \quad \text{resp.,} \quad \underline{W}(\nu_0, \mu_0) = \int_M \langle \nabla g(x), x \rangle d\nu_0(x).$$

Moreover, $\underline{W}(\nu_0, \mu_0) = -\overline{W}(\tilde{\nu}_0, \mu_0)$, where $\tilde{\nu}_0(A) = \nu_0(-A)$.

Note that if the convex function h solves $\overline{W}(\tilde{\nu}_0, \mu_0)$, then the concave function $g(x) = -h(-x)$ solves $\underline{W}(\nu_0, \mu_0)$ and vice-versa.

We shall need the following simple lemma regarding Kantorovich potentials.

Lemma 8 Let g be a function on M^* , and let h, k be functions on M .

1. If $h(x) - g(v) \geq b_T(v, x)$ on $M^* \times M$, then $h^*(w) + g(v) \leq \tilde{c}_T(v, w)$ on $M^* \times M^*$.
2. If $h(x) - g(v) \leq b_T(v, x)$ on $M^* \times M$, then $h(x) + g^*(-y) \leq c(y, x)$ on $M \times M$.
3. If $h(x) - g(y) \leq c(y, x)$ on $M \times M$, then $h(x) - (-g)^*(-v) \leq b_T(v, x)$ on $M^* \times M$.

Proof: 1) Since for any $(v, w) \in M^* \times M^*$, we have

$$\tilde{c}(t, v, w) = \sup\{\langle w, x \rangle - b(t, v, x); x \in M\}.$$

It follows that for any $y \in M$,

$$\tilde{c}_T(v, w) \geq \langle w, y \rangle - b(t, v, y) \geq \langle w, y \rangle + g(v) - h(y)$$

hence

$$h^*(w) + g(v) \leq \tilde{c}_T(v, w).$$

2) Since for any $(y, x) \in M \times M$, we have $c(t, y, x) = \sup\{b(t, v, x) - \langle v, y \rangle; v \in M^*\}$, it follows that for any $y \in M$, we have $-g(v) + h(x) - \langle v, y \rangle \leq c(x, y)$, that is

$$h(x) + g^*(-y) \leq c(y, x).$$

3) Since $b(t, v, x) = \inf\{\langle v, y \rangle + c(t, y, x); y \in M\}$, it follows that for any $(v, x) \in M^* \times M$, and any $\epsilon > 0$, there exists $y_\epsilon \in M$ such that

$$c_T(y_0, x) + \langle v, y_0 \rangle \leq b_T(v, x) + \epsilon,$$

hence $h(x) - g(y_0) + \langle v, y_0 \rangle \leq b_T(v, x) + \epsilon$, which means that

$$h(x) + \inf\{-g(y) + \langle v, y \rangle\} \leq b_T(v, x),$$

that is $h(x) - (-g)^*(-v) \leq b_T(v, x)$.

Proof of Theorem 3: First, we prove (23), that is if μ_0 is absolutely continuous with respect to Lebesgue measure, then

$$\underline{B}_T(\mu_0, \nu_T) = \inf\{C_T(\nu, \nu_T) + \underline{W}(\mu_0, \nu); \nu \in \mathcal{P}(M)\}. \quad (47)$$

We note that for any probability measure ν on M , we have

$$\underline{B}_T(\mu_0, \nu_T) \leq C_T(\nu, \nu_T) + \underline{W}(\mu_0, \nu). \quad (48)$$

Indeed, since μ_0 is assumed to be absolutely continuous with respect to Lebesgue measure, Brenier's theorem yields a concave function k on M^* that is differentiable μ_0 -almost everywhere such that $(\nabla k)_\# \mu_0 = \nu$, and

$$\underline{W}(\mu_0, \nu) = \int_{M^*} \langle v, \nabla k(v) \rangle d\mu_0(v).$$

Also let π_0 be an optimal transport plan for $C_T(\nu, \nu_T)$, that is $\pi_0 \in \mathcal{K}(\nu, \nu_T)$ such that

$$C_T(\nu, \nu_T) = \int_{M \times M} c_T(y, x) d\pi_0(y, x).$$

Let $\tilde{\pi}_0 := S_\# \pi_0$, where $S(y, x) = (\nabla \tilde{k}(y), x)$, where \tilde{k} is the concave Legendre transform of k . It is a transport plan in $\mathcal{K}(\mu_0, \nu_T)$. Since $b_T(v, x) \leq c_T(\nabla k(v), x) + \langle \nabla k(v), v \rangle$ for every $v \in M^*$, we have

$$\begin{aligned} \underline{B}_T(\mu_0, \nu_T) &\leq \int_{M^* \times M} b_T(v, x) d\tilde{\pi}_0(v, x) \\ &\leq \int_{M^* \times M} \{c_T(\nabla k(v), x) + \langle \nabla k(v), v \rangle\} d\tilde{\pi}_0(v, x) \\ &= \int_{M \times M} c_T(y, x) d\pi_0(y, x) + \int_M \langle \nabla k(v), v \rangle d\mu_0(v) \\ &= C_T(\nu, \nu_T) + \underline{W}(\mu_0, \nu). \end{aligned}$$

To prove the reverse inequality, use standard Monge-Kantorovich theory to write

$$\begin{aligned} \underline{B}_T(\mu_0, \nu_T) &= \inf\left\{ \int_{M^* \times M} b_T(v, x) d\pi(v, x); \pi \in \mathcal{K}(\mu_0, \nu_T) \right\} \\ &= \sup\left\{ \int_M h(x) d\nu_T(x) - \int_{M^*} g(v) d\mu_0(v); h(x) - g(v) \leq b_T(v, x) \right\}. \end{aligned}$$

Since the supremum can be taken over all admissible Kantorovich pairs (g, h) of functions, i.e. those satisfying the relations

$$g(v) = \sup_{x \in M} h(x) - b_T(v, x) \quad \text{and} \quad h(x) = \inf_{v \in M^*} b_T(v, x) + g(v),$$

we can assume that the initial Kantorovich potential g is convex. Since the cost function b_T is continuous, the infimum $\underline{B}_T(\mu_0, \nu_T)$ is attained at some probability measure $\pi_0 \in \mathcal{K}(\mu_0, \nu_T)$. Moreover, the supremum

in the dual problem is attained at some pair (g, h) of admissible Kantorovich functions. It follows that π_0 is supported on the set

$$\mathcal{O} := \{(v, x) \in M^* \times M; b_T(v, x) = h(x) - g(v)\}.$$

We now exploit the convexity of g , and use the fact that for each $(v, x) \in \mathcal{O}$, the function $w \rightarrow -g(w) + h(x) - b_T(w, x)$ attains its maximum at v , which means that

$$-\nabla g(v) \in \tilde{\partial}_v b_T(v, x),$$

where for a concave function f , we write $\tilde{\partial}f := -\partial(-f)$, the latter being negative the subdifferential of the convex function $-f$. But since $y \rightarrow -c_T(y, x)$ is the concave Legendre transform of $v \rightarrow b_T(v, x)$ with respect to the v -variable, we then have

$$b_T(v, x) - c_T(-\nabla g(v), x) = \langle v, -\nabla g(v) \rangle. \quad (49)$$

Integrating with π_0 , we get since $\pi_0 \in \mathcal{K}(\mu_0, \nu_T)$,

$$\int_{M^* \times M} b_T(v, x) d\pi_0 - \int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 = \int_{M^*} \langle v, -\nabla g(v) \rangle d\mu_0. \quad (50)$$

Letting $\nu_0 = (-\nabla g)_\# \mu_0$, we obtain that

$$\underline{B}_T(\mu_0, \nu_T) - \int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 = \underline{W}(\mu_0, \nu_0). \quad (51)$$

We now prove that

$$\int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 = C_T(\nu_0, \nu_T). \quad (52)$$

Indeed, we have

$$\int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 \geq C_T(\nu_0, \nu_T),$$

since the measure $\pi = S_\# \pi_0$, where $S(v, x) = (-\nabla g(v), x)$ has marginals ν_0 and ν_T respectively. On the other hand, (51) yields

$$\begin{aligned} \int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 &= \underline{B}_T(\mu_0, \nu_T) + \int_{M^*} \langle v, \nabla g(v) \rangle d\mu_0(v) \\ &= \int_M h(x) d\nu_T(x) - \int_{M^*} g(v) d\mu_0(v) + \int_{M^*} g^*(\nabla g(v)) d\mu_0(v) + \int_{M^*} g(v) d\mu_0(v) \\ &= \int_M h(x) d\nu_T(x) + \int_{M^*} g^*(-y) d\nu_0(y). \end{aligned}$$

Moreover, since $h(x) - g(v) \leq b(v, x)$ for all $(x, v) \in M \times M^*$, it follows from Lemma 8 that

$$h(x) + g^*(-y) \leq c_T(y, x), \quad (53)$$

which means that the couple $(-g^*(-y), h(x))$ is an admissible Kantorovich pair for the cost c_T . Hence,

$$\begin{aligned} C_T(\mu_0, \mu_T) &\leq \int_{M^* \times M} c_T(-\nabla g(v), x) d\pi_0 \\ &= \int_M h(x) d\nu_T(x) + \int_{M^*} g^*(-y) d\nu_0(y) \\ &\leq \sup \left\{ \int_M \varphi_1(x) d\nu_T(x) - \int_M \varphi_0(y) d\nu_0(y); \varphi_1(x) - \varphi_0(y) \leq c_T(y, x) \right\} \\ &= C_T(\mu_0, \mu_T). \end{aligned}$$

It follows that

$$\underline{B}_T(\mu_0, \nu_T) = \inf\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}(M)\}. \quad (54)$$

and the infimum is attained by the measure ν_0 . Note that the final optimal Kantorovich potential for $C_T(\nu_0, \nu_T)$ is $y \rightarrow -g^*(-y)$, hence is concave.

Similarly, we can show (24). First, for any probability measure μ on M^* , we have

$$\overline{B}_T(\mu_0, \nu_T) \geq \overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu). \quad (55)$$

Indeed, since ν_T is assumed to be absolutely continuous with respect to Lebesgue measure, Brenier's theorem yields a convex function h that is differentiable μ_T -almost everywhere on M such that $(\nabla h)_\# \nu_T = \mu$, and $\overline{W}(\nu_T, \mu) = \int_M \langle x, \nabla h(x) \rangle d\nu_T(x)$. Let π_0 be an optimal transport plan for $\tilde{C}_T(\mu_0, \mu)$, that is $\pi_0 \in \mathcal{K}(\mu_0, \mu)$ such that

$$\tilde{C}_T(\mu_0, \mu) = \int_{M^* \times M^*} \tilde{c}_T(v, w) d\pi_0(v, w).$$

Let $\tilde{\pi}_0 := S_\# \pi_0$, where $S(v, w) = (v, \nabla h^*(w))$, which is a transport plan in $\mathcal{K}(\mu_0, \nu_T)$. Since $b_T(v, y) \geq \langle \nabla h(x), y \rangle - \tilde{c}_T(v, \nabla h(x))$ for every $(y, x, v) \in M \times M \times M^*$, we have

$$\begin{aligned} \overline{B}_T(\mu_0, \nu_T) &\geq \int_{M^* \times M} b_T(v, x) d\tilde{\pi}_0(v, x) \\ &\geq \int_{M^* \times M} \{\langle \nabla h(x), x \rangle - \tilde{c}_T(v, \nabla h(x))\} d\tilde{\pi}_0(v, x) \\ &= \int_M \langle x, \nabla h(x) \rangle d\nu_T(x) - \int_{M^* \times M^*} \tilde{c}_T(v, w) d\pi_0(v, w) \\ &= \overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu). \end{aligned}$$

To prove the reverse inequality, we use standard Monge-Kantorovich theory to write

$$\begin{aligned} \overline{B}_T(\mu_0, \nu_T) &= \sup \left\{ \int_{M^* \times M} b_T(v, x) d\pi(v, x); \pi \in \mathcal{K}(\mu_0, \nu_T) \right\} \\ &= \inf \left\{ \int_M h(x) d\nu_T(x) - \int_{M^*} g(v) d\mu_0(v); h(x) - g(v) \geq b_T(v, x) \right\}, \end{aligned}$$

where the infimum is taken over all admissible Kantorovich pairs (g, h) of functions, i.e. those satisfying the relations

$$g(v) = \inf_{x \in M} h(x) - b_T(v, x) \quad \text{and} \quad h(x) = \sup_{v \in M^*} b_T(v, x) + g(v)$$

Note that h is convex. Since the cost function b_T is continuous, the supremum $\overline{B}_T(\mu_0, \nu_T)$ is attained at some probability measure $\pi_0 \in \mathcal{K}(\mu_0, \nu_T)$. Moreover, the infimum in the dual problem is attained at some pair (g, h) of admissible Kantorovich functions. It follows that π_0 is supported on the set

$$\mathcal{O} := \{(v, x) \in M^* \times M; b_T(v, x) = h(x) - g(v)\}.$$

We now exploit the convexity of h , and use the fact that for each $(v, x) \in \mathcal{O}$, the function $y \rightarrow h(y) - g(v) - b_T(v, y)$ attains its minimum at x , which means that

$$\nabla h(x) \in \partial_x b_T(v, x).$$

But since \tilde{c}_T is the Legendre transform of b_T with respect to the x -variable, we then have

$$b_T(v, x) + \tilde{c}_T(v, \nabla h(x)) = \langle x, \nabla h(x) \rangle \text{ on } \mathcal{O}. \quad (56)$$

Integrating with π_0 , we get since $\pi_0 \in \mathcal{K}(\mu_0, \nu_T)$,

$$\int_{M^* \times M} b_T(v, x) d\pi_0 + \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \int_M \langle x, \nabla h(x) \rangle d\nu_T. \quad (57)$$

Letting $\mu_T = \nabla h_{\#} \nu_T$, we obtain that

$$\overline{B}_T(\mu_0, \nu_T) + \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \overline{W}(\nu_T, \mu_T), \quad (58)$$

where

$$\overline{W}(\nu_T, \mu_T) = \sup \left\{ \int_{M \times M^*} \langle x, v \rangle d\pi; \pi \in \mathcal{K}(\nu_T, \mu_T) \right\}.$$

Note that we have used here that h is convex to deduce that $\overline{W}(\nu_T, \mu_T) = \int_M \langle x, \nabla h(x) \rangle d\mu_T$ by the uniqueness in Brenier's decomposition. We now prove that

$$\int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \tilde{C}_T(\mu_0, \mu_T), \quad (59)$$

where

$$\tilde{C}_T(\mu_0, \mu_T) := \inf \left\{ \int_{M^* \times M^*} \tilde{c}_T(v, w) d\pi; \pi \in \mathcal{K}(\mu_0, \mu_T) \right\}.$$

Indeed, we have

$$\int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 \geq \tilde{C}_T(\mu_0, \mu_T),$$

since the measure $\pi = S_{\#} \pi_0$, where $S(v, x) = (v, \nabla h(x))$ has marginals μ_0 and μ_T respectively. On the other hand, (58) yields

$$\begin{aligned} \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 &= \int_M \langle x, \nabla h(x) \rangle d\nu_T(x) - \int_{M^* \times M} b_T(v, x) d\pi_0 \\ &= \int_M h^*(\nabla h(x)) d\nu_T(x) + \int_M h(x) d\nu_T(x) + \int_{M^*} g(v) d\mu_0(v) - \int_M h(x) d\nu_T(x) \\ &= \int_{M^*} h^*(w) d\mu_T(w) + \int_{M^*} g(v) d\mu_0(v). \end{aligned}$$

Moreover, since $h(x) - g(v) \geq b(v, x)$, we have by Lemma 8, that

$$h^*(w) + g(v) \leq \tilde{c}_T(v, w),$$

which means that the couple $(-g, h^*)$ is an admissible Kantorovich pair for the cost \tilde{c}_T . Hence,

$$\begin{aligned} \tilde{C}_T(\mu_0, \mu_T) &\leq \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 \\ &= \int_M h^*(w) d\mu_T(w) + \int_{M^*} g(v) d\mu_0(v) \\ &\leq \sup \left\{ \int_{M^*} \varphi_T(w) d\mu_T(w) - \int_{M^*} \varphi_0(v) d\mu_0(v); \varphi_T(w) - \varphi_0(v) \leq \tilde{c}_T(v, w) \right\} \\ &= \tilde{C}_T(\mu_0, \mu_T). \end{aligned}$$

It follows that $\overline{B}_T(\mu_0, \nu_T) = \overline{W}(\nu_T, \mu_T) - \tilde{C}_T(\mu_0, \mu_T)$. In other words, the supremum in (55) is attained by the measure μ_T . Note that the final optimal Kantorovich potential for $\tilde{C}_T(\mu_0, \mu_T)$ is h^* , hence is convex.

4 The reverse interpolation

While the cost c_T is itself jointly convex in both variables, one cannot deduce much in terms of the convexity or concavity of the corresponding Kantorovich potentials. We investigate here what happens in such situations. Note that we get a concave initial Kantorovich potential for $C_T(\nu_0, \nu_T)$, when ν_0 is obtained via the factorization of a ballistic optimal transport problem. Actually, the following somewhat converse statement holds true.

Theorem 7 *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (A1), (A2) and (A3). Assume ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure. Then, the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave if and only if*

$$C_T(\nu_0, \nu_T) = \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (60)$$

Proof: Again, we show first that

$$C_T(\nu_0, \nu_T) \geq \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (61)$$

Indeed, if ν is any probability measure on M , use again Brenier's theorem to find a concave function h on M such that $(\nabla h)_\# \nu_0 = \mu$ and $\underline{W}(\nu_0, \mu) = \int_M \langle x, \nabla h(x) \rangle d\nu_0(x)$. Also let π_0 be an optimal transport plan for $C_T(\nu_0, \nu_T)$. Since $c(t, y, x) \geq b(t, \nabla h(y), x) - \langle \nabla h(y), y \rangle$ for every $x, y \in M$, and since $\tilde{\pi}_0 := S_{\#} \pi_0$ where $S(y, x) = (\nabla h(y), x)$, is in $\mathcal{K}(\mu, \nu_T)$, we have

$$\begin{aligned} C_T(\nu_0, \nu_T) &= \int_{M \times M} c_T(y, x) d\pi_0 \\ &\geq \int_{M \times M} \{b_T(\nabla h(y), x) - \langle \nabla h(y), y \rangle\} d\pi_0(y, x) \\ &= \int_{M \times M} b_T(\nabla h(y), x) d\pi_0(y, x) - \underline{W}(\nu_0, \mu) \\ &= \int_{M^* \times M} b_T(v, x) d\tilde{\pi}_0(v, x) - \underline{W}(\nu_0, \mu) \\ &\geq \underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu). \end{aligned}$$

To prove Theorem 7, we start by proving the reverse inequality and attainment in the case where the initial Kantorovich potential is concave. Indeed, write

$$\begin{aligned} C_T(\nu_0, \nu_T) &= \inf\left\{ \int_{M \times M} c(y, x) d\pi(y, x); \pi \in \mathcal{K}(\nu_0, \nu_T) \right\} \\ &= \sup\left\{ \int_M h(x) d\nu_T(x) - \int_M g(y) d\nu_0(y); h(x) - g(y) \leq c_T(y, x) \right\}. \end{aligned}$$

Since the cost function c_T is continuous, the infimum $C_T(\nu_0, \nu_T)$ is attained at some probability measure $\pi_0 \in \mathcal{K}(\nu_0, \nu_T)$. Moreover, the infimum in the dual problem is attained at some pair (g, h) of admissible Kantorovich functions. It follows that π_0 is supported on the set

$$\mathcal{O} := \{(y, x) \in M \times M; c_T(y, x) = h(x) - g(y)\}$$

Assuming g concave, use the fact that for each $(y, x) \in \mathcal{O}$, the function $z \rightarrow h(x) - g(z) - c_T(z, x)$ attains its maximum at y , to deduce that

$$-\nabla g(y) \in \partial_y c_T(y, x).$$

Assuming g concave and since $b(t, v, x) = \inf\{ \langle v, z \rangle + c(t, z, x); z \in M \}$, this means that for $(y, x) \in \mathcal{O}$,

$$c_T(y, x) = b_T(\nabla g(y), x) - \langle \nabla g(y), y \rangle. \quad (62)$$

Integrating with π_0 , we get since $\pi_0 \in \mathcal{K}(\nu_0, \nu_T)$,

$$\int_{M \times M} c_T(y, x) d\pi_0 = \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 - \int_M \langle \nabla g(y), y \rangle d\nu_0. \quad (63)$$

Letting $\mu_0 = (\nabla g)_\# \nu_0$, and since g is concave, we obtain that

$$C_T(\nu_0, \nu_T) = \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 - \underline{W}(\nu_0, \mu_0). \quad (64)$$

We now prove that

$$\int_{M \times M} b_T(\nabla g(y), x) d\pi_0(y, x) = \underline{B}_T(\mu_0, \nu_T). \quad (65)$$

Indeed, we have

$$\int_{M \times M} b_T(\nabla g(y), x) d\pi_0 \geq \underline{B}_T(\mu_0, \nu_T),$$

since the measure $\pi = S_\# \pi_0$ where $S(y, x) = (\nabla g(y), x)$ has μ_0 and ν_T as marginals. On the other hand, (64) yields

$$\begin{aligned} \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 &= \int_{M \times M} c_T(y, x) d\pi_0 + \int_M \langle y, \nabla g(y) \rangle d\nu_0(y) \\ &= \int_M h(x) d\nu_T(x) - \int_M g(y) d\nu_0(y) - \int_M (-g)^*(-\nabla g(y)) d\nu_0(y) + \int_M g(y) d\nu_0(y) \\ &= \int_M h(x) d\nu_T(x) - \int_{M^*} (-g)^*(-v) d\mu_0(v). \end{aligned}$$

Moreover, by Lemma 8, we have that $h(x) - (-g)^*(-v) \leq b_T(v, x)$, that is the couple $((-g)^*(-v), h(x))$ is an admissible Kantorovich pair for the cost b_T . It follows that

$$\begin{aligned} \underline{B}_T(\mu_0, \nu_T) &\leq \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 \\ &= \int_M h(x) d\nu_T(x) - \int_M (-g)^*(-v) d\mu_0(v) \\ &\leq \sup \left\{ \int_M \varphi_T(x) d\mu_T(x) - \int_{M^*} \varphi_0(v) d\mu_0(v); \varphi_T(x) - \varphi_0(v) \leq b_T(v, x) \right\} \\ &= \underline{B}_T(\mu_0, \nu_T), \end{aligned}$$

and $C_T(\nu_0, \nu_T) = \underline{B}_T(\mu_0, \nu_T) - \underline{W}(\nu_0, \mu_0)$. In other words, the supremum in (60) is attained by the measure μ_0 .

Corollary 9 *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (A1), (A2) and (A3). Assume ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure, and that the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave. If b_T satisfies the twist condition, then there exists a map $X_0^T : M^* \rightarrow M$ and a concave function g on M such that*

$$C_T(\nu_0, \nu_T) = \int_M c_T(y, X_0^T \circ \nabla g(y)) d\nu_0(y). \quad (66)$$

Proof: In this case,

$$C_T(\nu_0, \nu_T) = \underline{B}_T(\mu_0, \nu_T) - \underline{W}(\nu_0, \mu_0), \quad (67)$$

for some probability measure μ_0 on M^* . Let g be the concave function on M such that $(\nabla g)_\# \nu_0 = \mu_0$ and

$$\underline{W}(\nu_0, \mu_0) = \int_M \langle \nabla g(y), y \rangle d\nu_0(y).$$

Since b_T satisfies the twist condition, there exists a map $X_0^T : M^* \rightarrow M$ such that $(X_0^T)_\# \mu_0 = \nu_T$ and

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, X_0^T v) d\mu_0(v). \quad (68)$$

Note that the infimum $C_T(\nu_0, \nu_T)$ is attained at some probability measure $\pi_0 \in \mathcal{K}(\nu_0, \nu_T)$ and that π_0 is supported on a subset \mathcal{O} of $M \times M$ such that for $(y, x) \in \mathcal{O}$,

$$c_T(y, x) = b_T(\nabla g(y), x) - \langle \nabla g(y), y \rangle.$$

Moreover, $C_T(\nu_0, \nu_T) = \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 - \underline{W}(\nu_0, \mu_0)$, and

$$\int_{M \times M} b_T(\nabla g(y), x) d\pi_0 = \underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, X_0^T v) d\mu_0(v) = \int_M b_T(\nabla g(y), X_0^T \circ \nabla g(y)) d\nu_0(y).$$

Since b_T satisfies the twist condition, it follows that for any $(y, x) \in \mathcal{O}$, we have that $x = X_0^T \circ \nabla g(y)$ from which follows that $C_T(\nu_0, \nu_T) = \int_M c_T(y, X_0^T \circ \nabla g(y)) d\nu_0(y)$.

Proof of Corollary 1: The cost $c(x - y)$ corresponds to $c_1(y, x)$, where the Lagrangian is $L(x, v) = c(v)$, that is

$$c_1(y, x) = \inf \left\{ \int_0^1 c(\dot{\gamma}(t)) dt; \gamma \in C^1([0, 1], M); \gamma(0) = y, \gamma(1) = x \right\} = c(x - y). \quad (69)$$

It follows from (60) that there is a probability measure μ_0 on M^* such that $C_1(\nu_0, \nu_1) = \underline{B}_1(\mu_0, \nu_1) - \underline{W}(\nu_0, \mu_0)$. But in this case, $b_1(v, x) = \inf \{ \langle v, y \rangle + c(x - y); y \in M \} = \langle v, x \rangle - c^*(v)$, hence

$$C_1(\nu_0, \nu_1) = \underline{B}_1(\mu_0, \nu_1) - \underline{W}(\nu_0, \mu_0) = \underline{W}(\mu_0, \nu_1) - \int_{M^*} c^*(v) d\mu_0(v) - \underline{W}(\nu_0, \mu_0). \quad (70)$$

In other words,

$$C_1(\nu_0, \nu_1) + K = \underline{W}_1(\mu_0, \nu_1) - \underline{W}(\nu_0, \mu_0), \quad (71)$$

where K is the constant $\int_{M^*} c^*(v) d\mu_0(v)$.

Apply Brenier's theorem [6] twice to find concave functions $\varphi_0 : M \rightarrow \mathbb{R}$ and $\varphi_1 : M^* \rightarrow \mathbb{R}$ such that $(\nabla \varphi_0)_\# \nu_0 = \mu_0$, $(\nabla \varphi_1)_\# \mu_0 = \nu_1$ and

$$\underline{W}(\nu_0, \mu_0) = \int_M \langle y, \nabla \varphi_0(y) \rangle d\nu_0(y) \quad \text{and} \quad \underline{W}(\mu_0, \nu_1) = \int_{M^*} \langle v, \nabla \varphi_1(v) \rangle d\mu_0(v).$$

It follows from the preceding corollary that

$$C_1(\nu_0, \nu_1) + K = \int_M c_1(y, \nabla \varphi_1 \circ \nabla \varphi_0(y)) d\nu_0(y) = \int_M c(\nabla \varphi_1 \circ \nabla \varphi_0(y) - y) d\nu_0(y).$$

Note also that

$$\begin{aligned} C_1(\nu_0, \nu_1) + K &= \int_M \langle v, \nabla \varphi_1(v) \rangle d\mu_0(v) - \int_M \langle y, \nabla \varphi_0(y) \rangle d\nu_0(y) \\ &= \int_M \langle \nabla \tilde{\varphi}_1(y), y \rangle d\nu_0(y) - \int_M \langle y, \nabla \varphi_0(y) \rangle d\nu_0(y) \\ &= \int_M \langle \nabla \tilde{\varphi}_1(y) - \nabla \varphi_0(y), y \rangle d\nu_0(y), \end{aligned}$$

where $\tilde{\varphi}_1$ is the concave Legendre transform of φ_1 .

5 Duality for Ballistic Transports

We now use the Hopf-Lax formulae established in the previous section to prove Theorem 1. The other main ingredient is the duality formula exhibited by Bernard and Buffoni [5] for the optimal mass transport $C_T(\nu_0, \nu_T)$, where ν_0 and ν_T are two given probability measures on M .

Proof of Theorem 1: To prove the duality formula (12), first note that if V_0 is any initial Kantorovich potential for $C_T(\nu_0, \nu_T)$, then the final one can be taken to be

$$\begin{aligned} V_T(x) &= \inf_{y \in M} c_T(y, x) + V_0(y) \\ &= \inf \left\{ V_0(\gamma(0)) + \int_0^T L(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M); \gamma(T) = x \right\}. \end{aligned}$$

In other words, $V(T, x)$ is the final state (at time T) of a variational solution of (8) starting at V_0 .

Now use the Hopf-Lax formula to write $\underline{B}_T(\mu_0, \nu_T) = C_T(\nu_0, \nu_T) + \underline{W}(\mu_0, \nu_0)$, for some probability measure ν_0 on M . The proof of Theorem 3 also yields that

$$\underline{B}_T(\mu_0, \nu_T) = \int_M h(x) d\nu_T(x) - \int_{M^*} g(v) d\mu_0(v), \quad (72)$$

where g (resp h) is a convex initial (resp., final) Kantorovich potential for $\underline{B}_T(\mu_0, \nu_T)$ if and only if

$$\underline{C}_T(\nu_0, \nu_T) = \int_M h(x) d\nu_T(x) - \int_M \tilde{g}(v) d\mu_0(v),$$

where $\tilde{g}(y) = -g^*(-y)$ (resp h) is a concave initial (resp., final) Kantorovich potential for $\underline{C}_T(\nu_0, \nu_T)$. Note that \tilde{g} is the concave Legendre transform of $-g$. In other words, the duality formula (12) can then be obtained by taking as initial Kantorovich potential for $\underline{C}_T(\nu_0, \nu_T)$ any concave functional V_0 . The corresponding final Kantorovich potential for $\underline{C}_T(\nu_0, \nu_T)$ is then the final state V_T of a variational solution of (8) starting at V_0 . As for $\underline{B}_T(\mu_0, \nu_T)$, we then have

$$\underline{B}_T(\mu_0, \nu_T) = \sup \left\{ \int_M V_T(x) d\nu_T(x) + \int_{M^*} \tilde{V}_0(v) d\mu_0(v); V_0 \text{ concave \& } V_t \text{ solution of (8)} \right\}.$$

The proof is similar for the duality formula (13), provided one uses the Hopf-Lax formula (24), replace L by \tilde{L} and note that the corresponding Hamiltonian is now $H_{\tilde{L}}(q, x) = -H_L(x, q)$.

6 Optimal maps for the ballistic cost

We have seen in Section 4, that since the cost function b_T is continuous, the infimum $\underline{B}_T(\mu_0, \nu_T)$ is attained at some probability measure $\pi_0 \in \mathcal{K}(\mu_0, \nu_T)$, and that the supremum in the dual problem is attained at some pair (g, h) of admissible Kantorovich functions, where g is convex. In other words, the optimal transport plan π_0 is supported on the set

$$\mathcal{O} := \{(v, x) \in M^* \times M; b_T(v, x) = h(x) - g(v)\}$$

Moreover, for each $(v, x) \in \mathcal{O}$, the function $w \rightarrow -g(w) + h(x) - b_T(w, x)$ attains its maximum at v , which means that $-\nabla g(v) \in \partial_v b_T(v, x)$. By Proposition 5, there exists a Hamiltonian trajectory $(\gamma(t), \eta(t))$ over $[0, T]$ starting at $(-\nabla g(v), v)$ and ending at x . The rest of the proof of Theorem 2 is clear. It is however instructive to make the connection with the known results regarding c_T .

Indeed, if c_T satisfies *the twist condition*, that is if there is at most one Hamiltonian trajectory starting at a given initial state (y, v) , then x will be determined by v , and \mathcal{O} will be supported by a graph. This is indeed the case for Tonelli Lagrangians, which were considered in the compact case by Bernard-Buffoni [5], and by Fathi-Figalli [11] in the case of a Finsler manifold.

Definition 10 L is said to be a *Tonelli Lagrangian on $M \times M$* , if it satisfies the following properties:

- (a) L is C^2 ;
- (b) for every $x \in M$, the function $v \rightarrow L(x, v)$ is strictly convex on M ;
- (c) there exist a constant $C > -\infty$ and a non-negative function $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ with superlinear growth, i.e.,

$$\lim_{|v| \rightarrow +\infty} \frac{\theta(v)}{|v|} = +\infty$$
, such that $L(x, v) \geq C + \theta(v)$.

We also recall the following [2, Definition 5.5.1, page 129]:

Definition 11 Say that $f : M \rightarrow \mathbb{R}$ has an *approximate differential* at $x \in M$ if there exists a function $h : M \rightarrow \mathbb{R}$ differentiable at x such that the set $\{f = h\}$ has density 1 at x with respect to the Lebesgue measure. In this case, the approximate value of f at x is defined as $\tilde{f}(x) = h(x)$, and the approximate differential of f at x is defined as $\tilde{d}_x f = d_x h$. It is not difficult to show that this definition makes sense. In fact, both $h(x)$, and $d_x h$ do not depend on the choice of h , provided x is a density point of the set $\{f = h\}$.

If L is a Tonelli Lagrangian, the Hamiltonian $H : M \times M^* \rightarrow \mathbb{R}$ is then C^1 , and the Hamiltonian vector field X_H on $M \times M^*$ is then

$$X_H(x, v) = \left(\frac{\partial H}{\partial v}(x, v), -\frac{\partial H}{\partial x}(x, v) \right),$$

and the associated system of ODEs is given by

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial v}(x, v) \\ \dot{v} = -\frac{\partial H}{\partial x}(x, v). \end{cases} \quad (73)$$

The connection between minimizers $\gamma : [a, b] \rightarrow M$ of I_L and solutions of (73) is as follows. If we write

$$x(t) = \gamma(t) \quad \text{and} \quad v(t) = \frac{\partial L}{\partial p}(\gamma(t), \dot{\gamma}(t)),$$

then $x(t) = \gamma(t)$ and $v(t)$ are C^1 with $\dot{x}(t) = \dot{\gamma}(t)$, and the Euler-Lagrange equation yields $\dot{v}(t) = \frac{\partial L}{\partial x}(\gamma(t), \dot{\gamma}(t))$, from which follows that $t \mapsto (x(t), v(t))$ satisfies (73). Note also that since L is a Tonelli Lagrangian, the Hamiltonian H is actually C^2 , and the vector field X_H is C^1 . It therefore defines a (partial) C^1 flow φ_t^H .

There is also a (partial) C^1 flow φ_t^L on $M \times M^*$ such that every speed curve of an L -minimizer is a part of an orbit of φ_t^L . This flow is called the Euler-Lagrange flow, is defined by

$$\varphi_t^L = \mathcal{L}^{-1} \circ \varphi_t^H \circ \mathcal{L},$$

where $\mathcal{L} : M \times M \rightarrow M \times M^*$, is the global Legendre transform $(x, p) \mapsto (x, \frac{\partial L}{\partial p}(x, p))$. Note that \mathcal{L} is a homeomorphism on its image whenever L is a Tonelli Lagrangian. We now recall the following result.

Theorem 8 (*Fathi-Figalli [11]*) *Let L be a Tonelli Lagrangian on M . Fix $T > 0$, ν_0, ν_T a pair of probability measure on M , with ν_0 absolutely continuous with respect to Lebesgue measure. Then there exists a uniquely ν_0 -almost everywhere defined transport map $S : M \rightarrow M$ from ν_0 to ν_T which is optimal for the cost c_T . Moreover, any plan $\gamma \in \mathcal{K}(\nu_0, \nu_T)$, which is optimal for the cost c_T , verifies $\gamma(\text{Graph}(S)) = 1$.*

1. If (h, k) is an optimal Kantorovich pair, that is if

$$h(x) - k(y) = c_T(y, x) \quad \text{for } \gamma\text{-a.e. } (y, x) \text{ in } M \times M,$$

then there is a Borel set B of full ν_0 -measure, such that the approximate differential $\tilde{d}_y k$ of k at y is defined for $y \in B$, the map $y \mapsto \tilde{d}_y k$ is Borel measurable on B , and the transport map S is defined on B (hence μ -almost everywhere) by

$$S(y) = \pi^* \varphi_T^H(y, \tilde{d}_y k),$$

where $\pi^* : M \times M^* \rightarrow M$ is the canonical projection, and φ_t^H is the Hamiltonian flow of the Hamiltonian H associated to L .

2. A Lagrangian description for S valid on B (hence ν_0 -almost everywhere) is given by

$$S(y) = \pi \varphi_t^L(y, \widetilde{\text{grad}}_y^L k),$$

where $\pi = \pi^* \circ \mathcal{L}$, φ_t^L is the Euler-Lagrange flow of L , and $y \rightarrow \widetilde{\text{grad}}_y^L k$ is the measurable vector field on M defined on B by

$$\frac{\partial L}{\partial v}(y, \widetilde{\text{grad}}_y^L k) = \tilde{d}_y k.$$

Moreover, for every $y \in B$, there is a unique L -minimizer $\gamma : [0, T] \rightarrow M$, with $\gamma(0) = y, \gamma(T) = S(y)$, and this curve γ is given by $\gamma(s) = \pi \varphi_s^L(y, \widetilde{\text{grad}}_y^L k)$, for $0 \leq s \leq T$.

We now give the following corresponding result for b_T .

Theorem 9 *In addition to (A_1) , assume that L is a Tonelli Lagrangian, then*

1. *There exists a concave function $k : M \rightarrow \mathbb{R}$ such that*

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, S_T \circ \nabla \tilde{k}(v)) d\mu_0(v). \quad (74)$$

where

$$S_T(y) = \pi^* \varphi_T^H(y, \tilde{d}_y k),$$

$\pi^* : M \times M^* \rightarrow M$ being the canonical projection, \tilde{k} is the concave Legendre transform of k , and φ_t^H the Hamiltonian flow associated to L .

In other words, an optimal map for $\underline{B}_T(\mu_0, \nu_T)$ is given by $v \rightarrow \pi^* \varphi_T^H(\nabla \tilde{k}(v), v)$.

2. *There exists a convex function $h : M^* \rightarrow \mathbb{R}$ such that*

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(S_T^* \circ \nabla h^*(x), x) d\nu_T(x), \quad (75)$$

where $S_T^*(v) = \pi^* \varphi_T^{H^*}(v, \tilde{d}_v h)$, and $\varphi_t^{H^*}$ the flow associated to the Hamiltonian $H_*(v, x) = -H(-x, v)$, whose Lagrangian is $L_*(v, q) = L^*(-q, v)$.

In other words, an optimal map for $\overline{B}_T(\mu_0, \nu_T)$ is given by the inverse of the map $x \rightarrow \pi^* \varphi_T^{H^*}(\nabla h^*(x), x)$.

3. *We also have*

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v), \quad (76)$$

where

$$\tilde{S}_T(v) = \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v h_0)$$

$\varphi_t^{\tilde{H}}$ being the Hamiltonian flow associated to \tilde{L} (i.e., $\tilde{H}(v, x) = -H(x, v)$, and h_0 the solution $h(0, v)$ of the reverse Hamilton-Jacobi equation (10) with $h(T, v) = h(v)$).

Proof: Start again by the interpolation inequality and write that

$$\underline{B}_T(\mu_0, \nu_T) = C_T(\nu_0, \nu_T) + \underline{W}(\mu_0, \nu_0),$$

for some probability measure ν_0 . The proof also shows that there exists a concave function $k : M \rightarrow \mathbb{R}$ and another function $h : M \rightarrow \mathbb{R}$ such that $(\nabla \tilde{k})_{\#} \mu_0 = \nu_0$,

$$\underline{W}(\mu_0, \nu_0) = \int_M \langle \nabla \tilde{k}(v), v \rangle d\mu_0(v).$$

and

$$C_T(\nu_0, \nu_T) = \int_M h(x) d\nu_T(x) - \int_M k(y) d\nu_0(y).$$

Now use the theorem of Fathi-Figalli to write

$$C_T(\nu_0, \nu_T) = \int_M c_T(y, S_T y) d\nu_0(y), \quad (77)$$

where $S_T(y) = \pi^* \varphi_T^H(y, \tilde{d}_y k)$. Note that

$$\underline{B}_T(\mu_0, \nu_T) \leq \int_{M^*} b_T(v, S_T \circ \nabla \tilde{k}(v)) d\mu_0(v), \quad (78)$$

since $\nabla \tilde{k} \# \mu_0 = \nu_0$ and $(S_T) \# \nu_0 = \nu_T$, and therefore $(I \times S_T \circ \nabla \tilde{k}) \# \mu_0$ belongs to $\mathcal{K}(\mu_0, \nu_T)$.

On the other hand, since $b_T(v, x) \leq c_T(\nabla \tilde{k}(v), x) + \langle \nabla \tilde{k}(v), v \rangle$ for every $v \in M^*$, we have

$$\begin{aligned} \underline{B}_T(\mu_0, \nu_T) &\leq \int_{M^*} b_T(v, S_T \circ \nabla \tilde{k}(v)) d\mu_0(v) \\ &\leq \int_{M^*} \{c_T(\nabla \tilde{k}(v), S_T \circ \nabla \tilde{k}(v)) + \langle \nabla \tilde{k}(v), v \rangle\} d\mu_0(v) \\ &= \int_M c_T(y, S_T y) d\nu_0(y) + \int_{M^*} \langle \nabla \tilde{k}(v), v \rangle d\mu_0(v) \\ &= C_T(\nu_0, \nu_T) + \underline{W}(\mu_0, \nu_0) \\ &= \underline{B}_T(\mu_0, \nu_T). \end{aligned}$$

It follows that

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, S_T \circ \nabla \tilde{k}(v)) d\mu_0(v) = \int_{M^*} b_T(v, \pi^* \varphi_T^H(\nabla \tilde{k}(v), \tilde{d}_{\nabla \tilde{k}(v)} k)) d\mu_0(v).$$

Since k is concave, we have that $\tilde{d}_x k = \nabla k(x)$, hence $\tilde{d}_{\nabla \tilde{k}(v)} k = \nabla k \circ \nabla \tilde{k}(v) = v$, which yields our claim that

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \pi^* \varphi_T^H(\nabla \tilde{k}(v), v)) d\mu_0(v).$$

2) The proof for $\overline{B}_T(\mu_0, \nu_T)$ is similar and can be done by just reversing the order and proceeding from ν_T to μ_0 . For 3), there exists a convex function $h : M^* \rightarrow \mathbb{R}$ such that

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v). \quad (79)$$

where

$$\tilde{S}_T(v) = \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v h_0)$$

$\varphi_t^{\tilde{H}}$ being the Hamiltonian flow associated to \tilde{L} , and h_0 the solution $h(0, v)$ of the reverse Hamilton-Jacobi equation (10) with $h(T, v) = h(v)$.

As above, we start with the interpolation formula

$$\overline{B}_T(\mu_0, \nu_T) = \overline{W}(\nu_T, \mu_T) - \tilde{C}_T(\mu_0, \mu_T),$$

where μ_T is a probability measure on M^* . The proof yields a convex function $h : M^* \rightarrow \mathbb{R}$ and another function $h_0 : M^* \rightarrow \mathbb{R}$ as above such that $(\nabla h^*) \# \nu_T = \mu_T$

$$\overline{W}(\nu_T, \mu_T) = \int_{M^*} \langle \nabla^* h(x), x \rangle d\nu_T(x),$$

and

$$\tilde{C}_T(\mu_0, \mu_T) = \int_M h(w) d\mu_T(w) - \int_{M^*} h_0(v) d\mu_0(v).$$

Now use the theorem of Fathi-Figalli to write

$$\tilde{C}_T(\mu_0, \mu_T) = \int_M \tilde{c}_T(v, \tilde{S}_T v) d\mu_0(v), \quad (80)$$

where $\tilde{S}_T(v) = \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v g)$, and $\tilde{H}(v, x) := H_{\tilde{L}}(v, x) = -H_L(x, v)$.

Note that

$$\bar{B}_T(\mu_0, \nu_T) \geq \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v), \quad (81)$$

since $(\nabla h)_{\#} \mu_T = \nu_T$ and $(\tilde{S}_T)_{\#} \mu_0 = \mu_T$, and therefore $(I \times \nabla h \circ \tilde{S}_T)_{\#} \mu_0$ belongs to $\mathcal{K}(\mu_0, \nu_T)$.

On the other hand, since $b_T(v, y) \geq \langle \nabla h(x), y \rangle - \tilde{c}_T(v, \nabla h(x))$ for every $(x, y, v) \in M \times M \times M^*$, we have for every $(x, v) \in M \times M^*$,

$$b_T(v, \nabla h \circ \tilde{S}_T(v)) \geq \langle \nabla h(x), \nabla h \circ \tilde{S}_T(v) \rangle - \tilde{c}_T(v, \nabla h(x)),$$

and by taking $x = \nabla h^* \circ \tilde{S}_T v$, we have

$$b_T(v, \nabla h \circ \tilde{S}_T(v)) \geq \langle \tilde{S}_T v, \nabla h \circ \tilde{S}_T(v) \rangle - \tilde{c}_T(v, \tilde{S}_T(v)).$$

It follows that

$$\begin{aligned} \bar{B}_T(\mu_0, \nu_T) &\geq \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v) \\ &\geq \int_{M^*} \{ \langle \tilde{S}_T v, \nabla h \circ \tilde{S}_T v \rangle - \tilde{c}_T(v, \tilde{S}_T v) \} d\mu_0(v) \\ &= \int_{M^*} \langle v, \nabla h(v) \rangle d\mu_T(v) - \int_{M^*} \tilde{c}_T(v, \tilde{S}_T v) d\mu_0(v) \\ &= \bar{W}(\nu_T, \mu_T) - \tilde{C}_T(\mu_0, \mu_T) \\ &= \bar{B}_T(\mu_0, \nu_T), \end{aligned}$$

and therefore,

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v) = \int_{M^*} b_T(v, \nabla h \circ \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v h_0)) d\mu_0(v).$$

7 The Eulerian formulation

For the sake of brevity, we shall assume that L is a Tonelli Lagrangian and use the following characterization of $C_T(\nu_0, \nu_T)$ given in [5] in the compact setting. See the thesis of Schachter [20] for the case $M = \mathbb{R}^d$.

$$C_T(\nu_0, \nu_T) = \inf \left\{ \int_0^T \int_M L(x, w_t(x)) d\rho_t(x) dt; (\rho, w) \in P(0, T; \nu_0, \nu_T) \right\},$$

where $P(0, T; \nu_0, \nu_T)$ is the set of pairs (ρ, w) such that $t \rightarrow \rho_t \in \mathcal{P}(M)$, $t \rightarrow w_t \in \mathbb{R}^n$ are paths of Borel vector fields such that

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho w) &= 0 \quad \text{in } \mathcal{D}'((0, T) \times M) \\ \rho_0 &= \nu_0 \quad , \quad \rho_T = \nu_T. \end{aligned} \quad (82)$$

Combining this with (24), we get that

$$\begin{aligned} \underline{B}_T(\mu_0, \nu_T) &= \inf\{W(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}(M)\} \\ &= \inf\left\{W(\mu_0, \nu) + \int_0^T \int_M L(x, w_t(x)) d\rho_t(x) dt; \nu \in \mathcal{P}(M), (\rho, w) \in P(0, T; \nu, \nu_T)\right\} \\ &= \inf\left\{W(\mu_0, \rho_0) + \int_0^T \int_M L(x, w_t(x)) d\rho_t(x) dt; (\rho, w) \in P_{ter}(0, T; \nu_T)\right\}, \end{aligned}$$

where $P_{ter}(0, T; \nu_T)$ is the set of pairs (ρ, w) such that

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho w) &= 0 \quad \text{in } \mathcal{D}'((0, T) \times M) \\ \rho_T &= \nu_T. \end{aligned} \tag{83}$$

The same reasoning holds for $\overline{B}_T(\mu_0, \nu_T)$.

References

- [1] L. AMBROSIO: *Lecture notes on optimal transport problems*, in Mathematical Aspects of Evolving Interfaces, Lecture Notes in Math., **1812**, (2003) Springer-Verlag, Berlin/New York, 1-52.
- [2] L. AMBROSIO, N. GIGLI & G. SAVARÉ: *Gradient flows in metric spaces and in the Wasserstein space of probability measures*. Lectures in Mathematics, ETH Zurich, (2005) Birkhäuser.
- [3] L. AMBROSIO, J. FENG: *On a class of first order Hamilton-Jacobi equations in metric spaces*, Journal of Differential Equations, Volume 256, Issue 7, 1 (April 2014) 2194-2245
- [4] E. N. BARRON AND R. JENSEN: *Semicontinuous viscosity solutions for Hamilton-Jacobi equations with convex Hamiltonians*, Commun. PDE 15 (1990), 1713-1742.
- [5] P. BERNARD & B. BUFFONI: *Optimal mass transportation and Mather theory*. J. Eur. Math. Soc., **9** (2007), no. 1, 85-121.
- [6] Y. BRENIER: *Polar factorization and monotone rearrangement of vector-valued functions*. Comm. Pure Appl. Math., **44** (1991), 375-417.
- [7] P. CANNARSA & C. SINISTRARI: *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*. Progress in Nonlinear Differential Equations and Their Applications, **58** (2004), Birkhäuser, Boston.
- [8] P. CARDALIAGUET, F. DELARUE, J-M. LASRY & P-L. LIONS: *The master equations and the convergence problem in Mean Field Games*, Preprint.
- [9] L.C. EVANS & W. GANGBO: *Differential equations methods for the Monge-Ampère mass transfer problem*. Mem. Amer. Math. Soc. **137** (1999).
- [10] A. FATHI: *Regularity of C^1 solutions of the Hamilton-Jacobi equation*. Ann. Fac. Sci. Toulouse Math. (6), **12** (2003), 479-516.
- [11] A. FATHI, A. FIGALLI: *Optimal transportation on non-compact manifolds*, Israel Journal of Mathematics 175 (1), (2010) 1-59
- [12] H. FEDERER: *Geometric measure theory. Die Grundlehren der mathematischen Wissenschaften*, **153** (1969) Springer-Verlag New York Inc., New York.

- [13] W. GANGBO & R.J. MCCANN: *The geometry of optimal transportation*. Acta Math., **177** (1996), 113-161.
- [14] W. GANGBO AND A. SWIECH: Existence of a solution to an equation arising from the theory of Mean Field Games, J. Differential Equations 259 (2015), no. 11, 6573–6643.
- [15] G. MONGE: *Mémoire sur la Théorie des Déblais et des Remblais*. Hist. de l'Acad. des Sciences de Paris (1781), 666-704.
- [16] V.N. SUDAKOV: *Geometric problems in the theory of infinite-dimensional probability distributions*. Proc. Steklov Inst. Math. **141** (1979), 1-178.
- [17] R. T. ROCKAFELLAR: *Existence and duality theorems for convex problems of Bolza*, Trans. Amer. Math. Soc. 159 (1971), 1-40.
- [18] R.T. ROCKAFELLAR AND P.R. WOLENSKI: *Convexity in Hamilton-Jacobi theory I: dynamics and duality*, SIAM J. Control and Opt. 39 (2001), 1323-1350.
- [19] R.T. ROCKAFELLAR AND P.R. WOLENSKI: *Convexity in Hamilton-Jacobi theory II: envelope representations*, SIAM J. Control Opt. 39 (2001), 1351-1372.
- [20] B. SCHACHTER *An Eulerian Approach to Optimal Transport with Applications to the Otto Calculus*, Thesis, U. of Toronto (2017)
- [21] C. VILLANI: *Topics in mass transportation*. Graduate Studies in Mathematics, **58** (2004), American Mathematical Society, Providence, RI.
- [22] C. VILLANI: *Optimal transport, old and new*. Lecture notes, 2005 Saint-Flour summer school.