

Dynamic and Stochastic Propagation of Brenier's Optimal Mass Transport

Alistair Barton* and Nassif Ghoussoub†

*Department of Mathematics, University of British Columbia
Vancouver BC Canada V6T 1Z2*

February 12, 2018

Abstract

We investigate how mass transports that optimize the inner product cost -considered by Y. Brenier- propagate in time along a given Lagrangian. In the deterministic case, we consider transports that maximize and minimize the following “ballistic” cost functional on phase space $M^* \times M$,

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

where $M = \mathbb{R}^d$, $T > 0$, and $L : M \times M \rightarrow \mathbb{R}$ is a suitable Lagrangian. We also consider the stochastic counterpart:

$$\underline{B}_T^s(\mu, \nu) := \inf \left\{ \mathbb{E} \left[\langle V, X_0 \rangle + \int_0^T L(t, X, \beta(t, X)) dt \right]; X \in \mathcal{A}, V \sim \mu, X_T \sim \nu \right\}$$

where \mathcal{A} is the set of stochastic processes satisfying $dX = \beta_X(t, X) dt + dW_t$, for some drift $\beta_X(t, X)$, and where W_t is $\sigma(X_s : 0 \leq s \leq t)$ -Brownian motion. While inf-convolution allows us to easily obtain Hopf-Lax formulas on Wasserstein space for cost minimizing transports, this is not the case for total cost maximizing transports, which actually are sup-inf problems. However, in the case where the Lagrangian L is jointly convex on phase space, Bolza-type dualities –well known in the deterministic case but novel in the stochastic case–transform sup-inf problems to sup-sup settings. Hopf-Lax formulas relate optimal ballistic transports to those associated with dynamic fixed-end transports studied by Bernard-Buffoni and Fathi-Figalli in the deterministic case, and by Mikami-Thieullen in the stochastic setting. We also write Eulerian formulations and 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

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

*This work is part of a Master's thesis prepared by A. Barton under the supervision of N. Ghoussoub.

†Partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

where $\mathcal{K}(\mu, \nu)$ is the set of *transport plans* between μ and ν , that 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). On the other hand, $\mathcal{K}(c)$ is the set of functions $\varphi_1 \in L^1(X_1, \nu)$ and $\varphi_0 \in L^1(X_0, \mu)$ such that $\varphi_1(x) - \varphi_0(y) \leq c(y, x)$ for all $(y, x) \in X_0 \times X_1$. The pairs of functions in $\mathcal{K}(c)$ can be assumed to satisfy

$$\varphi_1(x) = \inf_{y \in X_0} c(y, x) + \varphi_0(y) \quad \text{and} \quad \varphi_0(y) = \sup_{x \in X_1} \varphi_1(x) - c(y, x). \quad (1)$$

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 $c(y, x) = |x - y|$ ([23], [26], [13], [31], [32]) 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 [8] 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 [18]. With a purpose of connecting mass transport with Mather theory, Bernard and Buffoni [7] considered dynamic cost functions on a given compact manifold M , that deal with fixed end-points problems of the following type:

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

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 [15] 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 $L(t, x, p) = f(p)$.

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

$$b_T(v, x) := \inf \left\{ \langle v, \gamma(0) \rangle + \int_0^T L(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M); \gamma(T) = x \right\}, \quad (3)$$

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

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

where μ_0 (resp., ν_T) is a given probability measure on M^* (resp., M), and

$$\overline{B}_T(\mu_0, \nu_T) := \sup \left\{ \int_{M^* \times M} b_T(v, x) d\pi; \pi \in \mathcal{K}(\mu_0, \nu_T) \right\}. \quad (5)$$

Note that when $T = 0$, we have $b_0(x, v) = \langle v, x \rangle$, which is exactly the case considered by Brenier [8], that is

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

and

$$\overline{W}(\mu_0, \nu_0) := \sup \left\{ \int_{M^* \times M} \langle v, x \rangle d\pi; \pi \in \mathcal{K}(\mu_0, \nu_0) \right\}, \quad (7)$$

making (5) a suitable dynamic version of the Wasserstein distance.

We shall also consider stochastic versions of the above problems, namely the cost of transport between two random variables Y and Z in $L^2(\Omega, M)$ defined as

$$c_T^s(Y, Z) = \inf \left\{ \mathbb{E} \left[\int_0^T L(t, X, \beta(t, X)) dt \right]; X \in \mathcal{A}, X_0 = Y, X_T = Z \text{ a.s.} \right\}, \quad (8)$$

as well as the ballistic cost of using an input V in $L^2(\Omega, M^*)$ to get to the random state Z in $L^2(\Omega, M)$, namely

$$b_T^s(V, Z) = \inf \left\{ \mathbb{E} \left[\langle V, X_0 \rangle + \int_0^T L(t, X, \beta(t, X)) dt \right]; X \in \mathcal{A}, X_T = Z \text{ a.s.} \right\}, \quad (9)$$

where \mathcal{A} is the set of stochastic processes verifying the stochastic differential equation

$$dX = \beta_X(t, X) dt + dW_t,$$

for some drift $\beta_X(t, X)$, where W_t is $\sigma(X_s : 0 \leq s \leq t)$ -Brownian motion. The corresponding mass transports are then

$$C_T^s(\nu_0, \nu_T) := \inf \{c_T^s(Y, Z); Y \sim \nu_0, Z \sim \nu_T\} \quad (10)$$

$$= \inf \left\{ \mathbb{E} \left[\int_0^T L(t, X, \beta_X(t, X)) dt \right]; X \in \mathcal{A}, X_0 \sim \nu_0, X_T \sim \nu_T \right\}, \quad (11)$$

which was considered by Mikami and Thieullen [22], while

$$\underline{B}_T^s(\mu_0, \nu_T) := \inf \{b_T^s(V, Z); V \sim \mu_0, Z \sim \nu_T\} \quad (12)$$

$$= \inf \left\{ \mathbb{E} \left[\langle V, X_0 \rangle + \int_0^T L(t, X, \beta(t, X)) dt \right]; X \in \mathcal{A}, V \sim \mu_0, X_T \sim \nu_T \right\}, \quad (13)$$

$$\overline{B}_T^s(\mu_0, \nu_T) := \sup \{b_T^s(V, Z); V \sim \mu_0, Z \sim \nu_T\} \quad (14)$$

$$= \sup_{V \sim \mu_0, Z \sim \nu_T} \inf_{X \in \mathcal{A}, X_T = Z} \left\{ \mathbb{E} \left[\langle V, X_0 \rangle + \int_0^T L(t, X, \beta(t, X)) dt \right] \right\}, \quad (15)$$

that we shall consider in the sequel.

In Section 2, we shall prove the following interpolation formulae on Wasserstein space associated to the deterministic minimization problem:

$$\underline{B}_T(\mu_0, \nu_T) = \inf \{ \underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}(M) \}. \quad (16)$$

The above formula can be seen as extensions of those by Hopf-Lax on state space to Wasserstein space. Indeed, for any (initial) function g , the associated value function can be written as

$$\varphi_g(t, x) = \inf \{ g(y) + c_t(y, x); y \in M \}. \quad (17)$$

In the case where the Lagrangian $L(t, x, p) = L_0(p)$ is only a function of p , and if H_0 is the associated Hamiltonian, then $c_t(y, x) = tL_0(\frac{1}{t}|x - y|)$ and (17) is nothing but the Hopf-Lax formula used to generate solutions for corresponding Hamilton-Jacobi equations. When g is the linear functional $g(x) = \langle v, x \rangle$, then $b_t(v, x)$ is itself a solution to the Hamilton-Jacobi equation, since

$$b_t(v, x) = \inf \{ \langle v, y \rangle + c_t(y, x); y \in M \}. \quad (18)$$

In other words, (16) can now be seen as extensions of (18) to the space of probability measures, where the Wasserstein distance fill the role of the scalar product.

In order to establish duality formulas, we consider the following forward Hamilton-Jacobi equations:

$$\begin{cases} \partial_t \varphi + H(x, \nabla_x \varphi) &= 0 \text{ on } [0, T] \times M, \\ \varphi(0, x) &= f(x), \end{cases} \quad (19)$$

and backward Hamilton-Jacobi equations:

$$\begin{cases} \partial_t \varphi + H(x, \nabla_x \varphi) &= 0 \text{ on } [0, T] \times M, \\ \varphi(T, x) &= f(x), \end{cases} \quad (20)$$

where the Hamiltonian on $[0, T] \times M \times M^*$ is defined by $H(t, x, q) = \sup_{p \in M} \{\langle p, q \rangle - L(t, x, p)\}$. Unless specified otherwise, we shall consider “variational solutions” for (19) and (20), which are formally given by the formulae

$$\Phi_{f,+}^t(x) := \Phi_{f,+}(t, x) = \inf \left\{ f(\gamma(0)) + \int_0^t L(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M); \gamma(t) = x \right\}, \quad (21)$$

$$\Phi_{f,-}^t(x) := \Phi_{f,-}(t, x) = \sup \left\{ f(\gamma(T)) - \int_t^T L(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M); \gamma(t) = x \right\}. \quad (22)$$

Additional conditions on the Lagrangian are needed in order to verify if $\Phi_{f,+}$ and $\Phi_{f,-}$ are anywhere close to a classical solution. We shall then prove the following duality formulae:

$$\underline{B}_T(\mu_0, \nu_T) = \sup \left\{ \int_M \Phi_{f^*,+}(T, x) d\nu_T(x) + \int_{M^*} f(v) d\mu_0(v); f \text{ concave in } \text{Lip}(M^*) \right\} \quad (23)$$

$$= \sup \left\{ \int_M g(x) d\nu_T(x) + \int_{M^*} (\Phi_{g,-}^0)_*(v) d\mu_0(v); g \text{ in } \text{Lip}(M) \right\}, \quad (24)$$

where h_* is the concave Legendre transform of h , i.e., $h_*(v) = \inf \{\langle v, y \rangle - h(y); y \in M\}$.

As to the question of attainment, we use a result by Fathi-Figalli [15] to show that if L is a Tonelli Lagrangian, and 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 $\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, x) d\pi_0$, and π_0 is supported on the possibly set-valued map $v \rightarrow \pi^* \varphi_T^H(\nabla k_*(v), v)$, with $\pi^* : M \times M^* \rightarrow M$ being the canonical projection, and $(x, v) \rightarrow \varphi_t^H(x, v)$ is the corresponding Hamiltonian flow.

In Section 3, we prove an analogous Hopf-Lax formulae on Wasserstein space associated to the stochastic minimization problem:

$$\underline{B}_T^s(\mu_0, \nu_T) = \inf \{ \underline{W}(\mu_0, \nu) + C_T^s(\nu, \nu_T); \nu \in \mathcal{P}(M) \}. \quad (25)$$

As to the duality, there are two features that distinguish the deterministic case from the stochastic case. For one, there is no Monge-Kantorovich duality for the latter since it doesn't correspond to a cost minimizing transport problem. Moreover, stochastic processes are not reversible as deterministic paths and so we can only prove the following duality formula:

$$\underline{B}_T^s(\mu_0, \nu_T) = \sup \left\{ \int_M g(x) d\nu_T(x) + \int_{M^*} (\Psi_{g,-}^0)_*(v) d\mu_0(v); g \text{ in } \text{Lip}(M) \right\}, \quad (26)$$

where this time $\Psi_{g,-}$ is the solution to the backward Hamilton-Jacobi-Bellman equation (27).

$$\begin{cases} \partial_t \psi + \frac{1}{2} \Delta \psi + H(x, \nabla_x \psi) &= 0 \text{ on } [0, T] \times M, \\ \psi(T, x) &= g(x), \end{cases} \quad (27)$$

whose formal variational solutions are given by the formula:

$$\Psi_{g,-}(t, x) = \sup_{X \in \mathcal{A}} \left\{ \mathbb{E} \left[g(X(T)) - \int_t^T L(s, X(s), \beta_X(s, X)) ds \middle| X(t) = x \right] \right\}. \quad (28)$$

In order to deal with the maximization problems $\overline{B}_T(\mu_0, \nu_T)$ and $\overline{B}_T^s(\mu_0, \nu_T)$, we need to use Bolza-type duality to convert the sup-inf problem to a concave maximization problem. For that, we shall

assume that the Lagrangian L is jointly convex in both variables. In Section 4, we then consider the dual Lagrangian \tilde{L} defined on $M^* \times M^*$ by

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

and the corresponding fixed-end costs on $M^* \times M^*$,

$$\tilde{c}_T(u, v) := \inf\left\{\int_0^T \tilde{L}(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M^*); \gamma(0) = u, \gamma(T) = v\right\}, \quad (29)$$

and its associated transport

$$\tilde{C}_T(\mu_0, \mu_T) := \inf\left\{\int_{M^* \times M^*} \tilde{c}_T(x, y) d\pi; \pi \in \mathcal{K}(\mu_0, \mu_T)\right\}. \quad (30)$$

We then recall the deterministic Bolza duality, and establish a new stochastic Bolza duality. We use these results in Section 5, to establish the following results for $\bar{B}_T(\mu_0, \nu_T)$.

$$\bar{B}_T(\mu_0, \nu_T) = \sup\{\bar{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (31)$$

and

$$\bar{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M g(x) d\nu_T(x) + \int_{M^*} \tilde{\Phi}_{g^*, -}^0(v) d\mu_0(v); g \text{ convex on } M\right\}, \quad (32)$$

where g^* is the convex Legendre transform of g , i.e., $g^*(x) = \sup\{\langle v, x \rangle - g(v); v \in M^*\}$, and $\tilde{\Phi}_{k, -}$ is a solution of the following dual backward Hamilton-Jacobi equation:

$$\begin{cases} \partial_t \varphi - H(t, \nabla_v \varphi, v) &= 0 \text{ on } [0, T] \times M^*, \\ \varphi(T, v) &= k(v), \end{cases} \quad (33)$$

whose variational solution is given by

$$\tilde{\Phi}_{k, -}(t, v) = \sup\left\{k(\gamma(T)) - \int_0^t \tilde{L}(s, \gamma(s), \dot{\gamma}(s)) ds; \gamma \in C^1([0, T], M^*); \gamma(0) = v\right\}. \quad (34)$$

In Section 6, we deal with the stochastic counterpart $\bar{B}_T^s(\mu_0, \nu_T)$ and prove the following

$$\bar{B}_T^s(\mu_0, \nu_T) := \sup\left\{\mathbb{E}\left[\langle X, V(T) \rangle - \int_0^T \tilde{L}(t, V, \beta(t, V)) dt\right]; V \in \mathcal{A}, V_0 \sim \mu_0, X \sim \nu_T\right\} \quad (35)$$

and therefore

$$\bar{B}_T^s(\mu_0, \nu_T) = \sup\{\bar{W}(\nu_T, \mu) - \tilde{C}_T^s(\mu_0, \mu); \mu \in \mathcal{P}(M^*)\}, \quad (36)$$

as well as the following duality formula:

$$\bar{B}_T^s(\mu_0, \mu_T) = \inf\left\{\int_{M^*} g(x) d\nu_T + \int_M \tilde{\Psi}_{g^*, -}^0(v) d\mu_0; g \text{ convex in } C_{\text{db}}^\infty(M^*)\right\}, \quad (37)$$

where $\tilde{\Psi}_k$ solves the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t \psi + \frac{1}{2} \Delta \psi - H(\nabla_v \psi, v) &= 0 \text{ on } [0, T] \times M^*, \\ \psi(T, v) &= k(v), \end{cases} \quad (38)$$

whose formal variational solutions are given by the formula:

$$\tilde{\Psi}_{k, -}(t, v) = \sup_{X \in \mathcal{A}} \left\{ \mathbb{E} \left[k(X(T)) - \int_t^T \tilde{L}(s, X(s), \beta_X(s, X)) ds \middle| X(t) = v \right] \right\}. \quad (39)$$

Finally, a few words about our notation: We shall denote by ∂g the subdifferential of a convex function g , and by $\tilde{\partial} h := -\partial(-h)$ the superdifferential of a concave function h . The set of probability measures on a Banach space X will be denoted $\mathcal{P}(X)$, while the subset of those with finite first moment will be denoted

$$\mathcal{P}_1(X) := \{\nu \in \mathcal{P}(X); \int_X |x| d\nu(x) < \infty\}.$$

$\mathcal{P}_1(X)$ is clearly a subset of the Banach space of all finite measures with finite first moment, denoted similarly $\mathcal{M}_1(X) := \{\nu; \int_X 1 + |x| d\nu(x) < \infty\}$, which is dual to the Banach space $\text{Lip}(X)$ of all bounded uniformly Lipschitz functions on X . For the stochastic part, we shall also need to work with the space $C_{\text{db}}^\infty(X) := \text{Lip}(X) \cap C^\infty(X)$.

Several of the above results appeared in the posted but non-published manuscripts [20], which dealt with the deterministic case and [9], which addressed the stochastic case. We eventually elected to combine them in a single publication so as to illustrate the obvious similarities, but also the subtle differences between the two cases.

2 Minimizing the ballistic cost: Deterministic case

In this section we deal with the standard transportation problem associated to the cost $b_T(v, x)$. We shall assume that the Lagrangian L satisfies the following:

(A0) The Lagrangian $(t, x, v) \mapsto L(t, x, v)$ is bounded below, and for all $(t, x) \in [0, T] \times M$, $v \mapsto L(t, x, v)$ is convex and δ -coercive in the sense that there is a $\delta > 1$ such that

$$\lim_{|v| \rightarrow \infty} \frac{L(t, x, v)}{|v|^\delta} = +\infty. \quad (40)$$

Theorem 1 *Assume that L satisfies (A0) and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M) with finite first moment. Then, the following interpolation formula holds:*

$$\underline{B}_T(\mu_0, \nu_T) = \inf\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}_1(M)\}. \quad (41)$$

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.

Proof: To prove the formula it suffices to note that

$$\begin{aligned} & \inf\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T); \nu \in \mathcal{P}_1(M)\} \\ &= \inf_{\nu \in \mathcal{P}_1(M)} \left\{ \int_{M^* \times M} \langle v, x \rangle d\pi_W(v, x) + \int_{M \times M} c_T(x, y) d\pi_C(x, y); \pi_W \in \mathcal{K}(\mu_0, \nu), \pi_C \in \mathcal{K}(\nu, \nu_T) \right\} \\ &= \inf_{\pi \in \mathcal{P}_1(M^* \times M \times M)} \left\{ \int_{M^* \times M \times M} \langle v, x \rangle + c_T(x, y) d\pi(v, x, y); \pi_1 = \mu_0, \pi_3 = \nu_T \right\} \\ &\geq \underline{B}(\mu_0, \nu_T). \end{aligned}$$

For the reverse inequality, use your favourite selection theorem to find a measurable function $y_\epsilon : M^* \times M \rightarrow M$ that satisfies $\langle v, y_\epsilon(v, x) \rangle + c(y_\epsilon(v, x), x) - \epsilon < b_T(v, x)$. Fixing $\pi \in \mathcal{K}(\mu_0, \nu_T)$ and letting $\pi_\epsilon := (\text{Id} \times \text{Id} \times y_\epsilon)_\# \pi \in \mathcal{P}(M^* \times M \times M)$

$$\underline{B}(\mu_0, \nu_T) = \int_{M^* \times M} b_T(v, x) d\pi(v, x) \geq \int_{M^* \times M \times M} \langle v, y \rangle + c_T(y, x) d\pi_\epsilon(v, x, y) - \epsilon.$$

To show that the minimizer is achieved, we need to prove that C_T satisfies a coercivity condition on the space $\mathcal{P}_1(M)$ of probabilities on M with finite first moments. For that, we show that for any fixed $\nu_T \in \mathcal{P}_1(M)$ and any positive constant $N > 0$, the set of measures $\nu \in \mathcal{P}_1(M)$ satisfying

$$C_T(\nu, \nu_T) \leq N \int_M |x| d\nu(x), \quad (42)$$

is tight. Indeed, from (A0), there exists a constant K such that $c_T(x, y) > N \left| \frac{x-y}{T} \right|^\delta - K$. We concern ourselves with the cylinder set $B := B(R, 0)^c \times M$. Let $\nu \in \mathcal{T}_{\epsilon, R} := \{\nu; \nu(B(R, 0)^c) > \epsilon\}$. We shall assume, without loss of generality, that L and hence c_T is non-negative, hence for any optimal transport plan $\pi \in \mathcal{K}(\nu, \nu_T)$

$$C_T(\nu, \nu_T) \geq \frac{N}{T^\delta} \int_B \left| |x| - |y| \right|^\delta d\pi(x, y) - K\epsilon. \quad (43)$$

We define $\hat{\pi} := \pi|_B$ to be the restriction of π to the set B , and transfer the problem to \mathbb{R}_+ by using the push-forward $\bar{\pi} := (|\cdot| \times |\cdot|)_\# \hat{\pi}$ to obtain,

$$\int_B \left| |x| - |y| \right|^\delta d\hat{\pi}(x, y) = \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y|^\delta d\bar{\pi}(x, y). \quad (44)$$

We can obtain a lower estimate for this by minimizing over transportation measures sharing $\bar{\pi}$'s marginals (i.e., $\gamma \in \mathcal{K}(\bar{\pi}_1, \bar{\pi}_2)$). This is a well known optimal transport problem, whose optimal plan given by the monotone Hoeffding-Frechet mapping $x \mapsto G_\nu(G_{\nu_T}^{-1}(x))$, where $G_\nu(t) := \inf\{z \in \mathbb{R} : t \geq \nu(\{x \leq z\})\}$ is the quantile function associated with the measure ν [6]. Thus the optimal plan maps each quantile in one measure to the corresponding quantile in the other. Substituting this into the integral and applying Jensen's inequality:

$$\begin{aligned} \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y|^\delta d\bar{\pi}(x, y) &\geq \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x - y|^\delta d\left(\left(G_{|Y(0)|} \times G_{|Y(T)|}\right)_\# \lambda_{[0,1]}\right)(x, y) \\ &\geq \left(\int_{B(R, 0)^c} |x| d\nu(x) - b(\nu_T) \right)^\delta, \end{aligned} \quad (45)$$

where $b(\nu_T) := \int |x| d\nu_T$ and $R > b/\epsilon$. We thus want to find R such that

$$N \left(\left(\int_{B(R, 0)^c} |x| d\nu(x) - b(\nu_T) \right)^\delta - K\epsilon \right) > N \left(\int_{B(R, 0)^c} |x| d\nu(x) + R(1 - \epsilon) \right) \geq N \int |x| d\nu(x).$$

Letting $I_\nu(R) := \int_{B(R, 0)^c} |x| d\nu(x)$ ($\geq R\epsilon$ for $\nu \in \mathcal{T}_{\epsilon, R}$), we find the condition

$$\frac{\alpha}{T^{\delta-1}} \left(\left(I_\nu(R)^{1-\frac{1}{\delta}} - b(\nu_T) I_\nu(R)^{-\frac{1}{\delta}} \right)^\delta - \frac{(UT)}{I_\nu(R)} \right) > N \left(1 + \frac{R(1 - \epsilon)}{I_\nu(R)} \right),$$

which by using the mentioned bound on $I_\nu(R)$ is satisfied if R is large enough so that

$$\frac{\alpha}{T^{\delta-1}} \left(\left((R\epsilon)^{1-\frac{1}{\delta}} - b(\nu_T)(R\epsilon)^{-\frac{1}{\delta}} \right)^\delta - \frac{(UT)}{R\epsilon} \right) > \frac{N}{\epsilon}.$$

To show the minimizer is achieved, fix any ν_0 in $\mathcal{P}_1(M)$, and note that by coercivity the set of probability measures ν such that

$$\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T) \leq b(\mu_0) \int_M |x| d\nu + C_T(\nu, \nu_T) \leq b(\mu_0) \int_M |x| d\nu + C_T(\nu_0, \nu_T)$$

is tight.

Remark 1 Note that (45) indicates that when $\nu_1 \in \mathcal{P}_1(M)$ and $\nu_0 \in \mathcal{P}(M) \setminus \mathcal{P}_1(M)$, then $C(\nu_0, \nu_1) = C(\nu_1, \nu_0) = \infty$.

Theorem 2 Assume that L satisfies (A0) and let μ_0 (resp. ν_T) be a probability measure on M^* (resp., M) with finite first moment.

1. If μ_0 has compact support, then we have the following duality formula

$$\underline{B}_T(\mu_0, \nu_T) = \sup \left\{ \int_M g(x) d\nu_T(x) + \int_{M^*} (\Phi_{g,-}^0)_*(v) d\mu_0(v); g \text{ in } Lip(M) \right\}. \quad (46)$$

2. If ν_T has compact support, then

$$\underline{B}_T(\mu_0, \nu_T) = \sup \left\{ \int_M \Phi_{f^*,+}^T(x) d\nu_T(x) + \int_{M^*} f(v) d\mu_0(v); f \text{ concave in } Lip(M^*) \right\}. \quad (47)$$

Proof: We shall need the following identifications of the Legendre transforms in the Banach space $\mathcal{M}_1(\mathbb{R}^n)$ of measures ν on \mathbb{R}^n such that $\int_{\mathbb{R}^n} (1 + |x|) d\nu < \infty$ in duality with the space of Lipschitz functions $Lip(\mathbb{R}^n)$.

Lemma 1 a) For $\mu_0 \in \mathcal{P}(\mathbb{R}^n)$ with compact support, define $\underline{W}_{\mu_0} : \mathcal{M}_1(\mathbb{R}^n) \rightarrow \mathbb{R} \cup \{\infty\}$ to be

$$\underline{W}_{\mu_0}(\nu) := \begin{cases} \underline{W}(\mu_0, \nu) & \nu \in \mathcal{P}_1(M) \\ +\infty & \text{otherwise.} \end{cases}$$

Then, the convex Legendre transform of \underline{W}_{μ_0} is given for $f \in Lip(\mathbb{R}^n)$ by $\underline{W}_{\mu_0}^*(f) = -\int_M f_* d\mu_0$.

b) For $\nu_0 \in \mathcal{P}(\mathbb{R}^n)$, define the function $C_{\nu_0} : \mathcal{M}_1(\mathbb{R}^n) \rightarrow \mathbb{R} \cup \{\infty\}$ to be

$$C_{\nu_0}(\nu) := \begin{cases} C_T(\nu_0, \nu) & \nu \in \mathcal{P}_1(\mathbb{R}^n) \\ +\infty & \text{otherwise.} \end{cases} \quad (48)$$

Then, the convex Legendre transform of C_{ν_0} is given for $f \in Lip(\mathbb{R}^n)$ by $C_{\nu_0}^*(f) = \int_M \varphi_{f,-}(0, x) d\nu_0(x)$, where $\varphi_{f,-}$ is the solution to the backward Hamilton-Jacobi equation (20) with final condition $\varphi_{f,-}(T, x) = f(x)$.

Proof: Both statements follow from Kantorovich duality. Indeed, both functions are convex and weak*-lower semi-continuous on $\mathcal{M}_1(\mathbb{R}^n)$. Since μ_0 has compact support, Brenier's duality yields

$$\underline{W}_{\mu_0}(\nu) = \sup_{g \in Lip(M)} \left\{ \int_M g d\nu + \int_{M^*} g_* d\mu_0 \right\}.$$

We then have

$$\underline{W}_{\mu_0}^*(f) = \sup_{\nu \in \mathcal{M}_1(M)} \inf_{g \in Lip(M)} \left\{ \int_M f d\nu - \int_M g d\nu - \int_{M^*} g_* d\mu_0 \right\}. \quad (49)$$

Note that the functional $g \mapsto -\int_{M^*} g_* d\mu_0 = \int_{M^*} (-g)^* d\hat{\mu}(v)$ (where $d\hat{\mu}(v) := d\mu(-v)$) is convex and lower semicontinuous, and we may therefore apply the Von Neuman minimax theorem as the expression is linear in ν and convex in g . We obtain

$$\underline{W}_{\mu_0}^*(f) = \inf_{g \in Lip(M)} \sup_{\nu \in \mathcal{M}_1(M)} \left\{ \int_M f d\nu - \int_M g d\nu - \int_{M^*} g_* d\mu_0 \right\}. \quad (50)$$

The infimum must occur at $g = f$ since otherwise the sup in ν is $+\infty$, resulting in statement a).

The same proof applies to C_{ν_0} , since in view of the duality formula of Bernard and Buffoni [7][Proposition 21]:

$$C_{\nu_0}(\nu) = \sup_{g \in \text{Lip}(M)} \left\{ \int_M g \, d\nu - \int_M \Phi_{g,-}^0 \, d\nu_0 \right\}. \quad (51)$$

Note that this holds for all $\nu \in \mathcal{M}_1(M)$, since if g solves HJ, then so does $g + c$ for arbitrarily large c . We may again apply the minimax theorem as the expression is linear in ν and convex in g . To complete the proof of the theorem, we first note that Kantorovich duality yields that $\nu \mapsto \underline{B}(\mu_0, \nu)$ is weak*-lower semi-continuous on $\mathcal{P}_1(M)$ for all $\mu_0 \in \mathcal{P}_1(M^*)$ and that $(\mu_0, \nu_T) \mapsto \underline{B}(\mu_0, \nu_T)$ is jointly convex. Let now $\underline{B}_{\mu_0}(\nu) := \underline{B}_T(\mu_0, \nu)$ if $\nu \in \mathcal{P}_1(M)$ and $+\infty$ otherwise. It follows that

$$\underline{B}_{\mu_0}(\nu) = \underline{B}_{\mu_0}^{**}(\nu) := \sup \left\{ \int_{\mathbb{R}^n} f \, d\nu - \underline{B}_{\mu_0}^*(f); f \in \text{Lip}(\mathbb{R}^n) \right\}. \quad (52)$$

Now use the Hopf-Lax formula established above to write

$$\begin{aligned} \underline{B}_{\mu_0}^*(f) &:= \sup \left\{ \int_M f \, d\nu - \underline{B}_{\mu_0}(\nu); \nu \in \mathcal{P}_1(M) \right\} \\ &= \sup \left\{ \int_M f \, d\nu - \underline{W}(\mu_0, \nu') - C_T(\nu', \nu); \nu, \nu' \in \mathcal{P}_1(M) \right\} \\ &= \sup \left\{ \int_M \Phi_{f,-}^T \, d\nu' - \underline{W}(\mu_0, \nu'); \nu' \in \mathcal{P}_1(M) \right\} \\ &= - \int_M (\Phi_{f,-}^T)_* \, d\mu_0. \end{aligned} \quad (53)$$

This completes the proof of the first duality formula.

The second follows in the same way by simply varying the initial measure as opposed to the final measure in $\underline{B}_T(\mu, \nu)$. The concavity of f follows from the Kantorovich dual condition (1) and the linearity of b_T in v .

We now consider the problem of attainment for $\underline{B}_T(\mu, \nu)$. For that, we shall consider Tonelli Lagrangians studied in the compact case by Bernard-Buffoni [7], and by Fathi-Figalli [15] in the case of a Finsler manifold.

Definition 2 We shall say that L is a *Tonelli Lagrangian on $M \times M$* , if it is C^2 and satisfies (A0) with the additional requirement that the function $v \rightarrow L(x, v)$ is strictly convex on M .

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

Definition 3 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 $X_H(x, v) = (\frac{\partial H}{\partial v}(x, v), -\frac{\partial H}{\partial x}(x, v))$, and the associated system of ODEs is given by

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial v}(x, v) \\ \dot{v} = -\frac{\partial H}{\partial x}(x, v). \end{cases} \quad (54)$$

The connection between minimizers $\gamma : [a, b] \rightarrow M$ of I_L and solutions of (54) is as follows. If we write $x(t) = \gamma(t)$ and $v(t) = \frac{\partial L}{\partial p}(\gamma(t), \dot{\gamma}(t))$, 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 (54). 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.

Theorem 3 *In addition to (A0), assume that L is a Tonelli Lagrangian and that μ_0 is absolutely continuous with respect to Lebesgue measure. Then, there exists a concave function $k : M \rightarrow \mathbb{R}$ such that*

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \quad (55)$$

where $S_T(y) = \pi^* \varphi_T^H(y, \nabla k(y))$, $\pi^* : M \times M^* \rightarrow M$ being the canonical projection, 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 $v \rightarrow \pi^* \varphi_T^H(\nabla k_*(v), v)$.

Proof: Start again by the interpolation inequality, $\underline{B}_T(\mu_0, \nu_T) = C_T(\nu_0, \nu_T) + \underline{W}(\mu_0, \nu_0)$ for some probability measure ν_0 . By the above and Kantorovich duality, there exists a concave function $k : M \rightarrow \mathbb{R}$ and another function $h : M \rightarrow \mathbb{R}$ such that $(\nabla k_*)_{\#} \mu_0 = \nu_0$,

$$\underline{W}(\mu_0, \nu_0) = \int_M \langle \nabla k_*(v), v \rangle d\mu_0(v),$$

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 a result of Fathi-Figalli [15] to write $C_T(\nu_0, \nu_T) = \int_M c_T(y, S_T y) d\nu_0(y)$, where $S_T(y) = \pi^* \varphi_T^H(y, \tilde{d}_y k)$. Note that

$$\underline{B}_T(\mu_0, \nu_T) \leq \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \quad (56)$$

since $(\nabla k_*)_{\#} \mu_0 = \nu_0$ and $(S_T)_{\#} \nu_0 = \nu_T$, and therefore $(I \times S_T \circ \nabla k_*)_{\#} \mu_0$ belongs to $\mathcal{K}(\mu_0, \nu_T)$. On the other hand, 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

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

It follows that

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

Since k is concave, we have that $\tilde{d}_x k = \nabla k(x)$, hence $\tilde{d}_{\nabla k_*(v)} k = \nabla k \circ \nabla k_*(v) = v$, which yields our claim that $\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \pi^* \varphi_T^H(\nabla k_*(v), v)) d\mu_0(v)$.

3 Minimizing the ballistic cost: Stochastic case

We now turn to the stochastic version of the minimizing cost. The methods of proof are generally similar to those for the deterministic cost, however there are two complications: The first is that stochastic mass transport does not fit in the framework of cost minimizing transports, hence the Kantorovich duality is not readily available. The second is that stochastic processes are not reversible and therefore there is only one direction to the transport, hence only one duality formula. In order to deal with the first complication, we rely on the results of Mikami-Thieullen [22] and therefore use the same assumptions that they imposed on the Lagrangian, namely

(A1) $L(t, x, v)$ is continuous, convex in v , and uniformly bounded below by a convex function $\underline{L}(v)$ that is 2-coercive in the sense that $\lim_{|v| \rightarrow \infty} \frac{\underline{L}(v)}{|v|^2} > 0$.

(A2) $(t, x) \mapsto \log(1 + L(x, u))$ is uniformly continuous in that

$$\Delta L(\epsilon_1, \epsilon_2) := \sup_{u \in M^*} \left\{ \frac{1 + L(x, u)}{1 + L(y, u)} - 1; |t - s| < \epsilon_1, |x - y| < \epsilon_2 \right\} \xrightarrow{\epsilon_1, \epsilon_2 \rightarrow 0} 0.$$

(A3) The following boundedness conditions:

- (i) $\sup_{t, x} L(t, x, 0) < \infty$.
- (ii) $|\nabla_x L(t, x, v)| / (1 + L(t, x, v))$ is bounded.
- (iii) $\sup \{ |\nabla_v L(t, x, u)| : |u| \leq R \} < \infty$ for all R .

We will use the notation $X = (X_0, \beta_X, \sigma_X)$ to refer to an Itô process $X(t)$ of the form:

$$X(t) = X_0 + \int_0^t \beta_X(s) ds + \int_0^t \sigma_X(s) dW_s. \quad (57)$$

We will use the notation $\mathcal{A}_{\nu_0}^{\nu_T}$ to refer to the set of stochastic processes $X = (X_0, \beta_X, \text{Id})$ with $X(0) \sim \nu_0$ and $X(T) \sim \nu_T$. Notably, (A1) implies that $\mathbb{E}[L(t, X(t), \beta_X)] = \infty$ if $\beta_X(t) \notin L^2(\mathbb{P})$.

Our main result for this section is the stochastic counterpart to Theorem 2:

Theorem 4 *If L satisfies the assumptions (A1), (A2), and (A3), then*

1. *For any given probabilities $\mu_0 \in \mathcal{P}(M^*)$ and $\nu_T \in \mathcal{P}(M)$, we have:*

$$\underline{B}_T^s(\mu_0, \nu_T) = \inf \{ \underline{W}(\mu_0, \nu) + C_T^s(\nu, \nu_T); \nu \in \mathcal{P}_1(M) \}. \quad (58)$$

Furthermore, this infimum is attained whenever $\mu_0 \in \mathcal{P}_1(M^)$ and $\nu_T \in \mathcal{P}_1(M)$.*

2. *If $\nu_T \in \mathcal{P}_1(M)$ and $\mu_0 \in \mathcal{P}_1(M^*)$ are such that $\underline{B}(\mu_0, \nu_T) < \infty$, and if $\mu_0 \in \mathcal{P}_1(M^*)$ has compact support, then*

$$\underline{B}_T^s(\mu_0, \nu_T) = \sup \left\{ \int_M f(x) d\nu_T(x) + \int_{M^*} (\Psi_{f, -}^0)_*(v) d\mu_0(v); f \in \text{Lip}(M) \right\}, \quad (59)$$

where $\Psi_{f, -}$ is the solution to the Hamilton-Jacobi-Bellman equation

$$\frac{\partial \psi}{\partial t} + \frac{1}{2} \Delta \psi(t, x) + H(t, x, \nabla \psi) = 0, \quad \psi(T, x) = f(x). \quad (\text{HJB})$$

Proof: 1) First, expand $\underline{W}(\mu_0, \nu)$ and $C_T(\nu, \nu_T)$ in the interpolation formula to obtain:

$$\begin{aligned} & \inf\{\underline{W}(\mu_0, \nu) + C_T^s(\nu, \nu_T); \nu \in \mathcal{P}_1(M)\} \\ &= \inf\left\{\mathbb{E}\left[\langle V, X \rangle + \int_0^T L(t, X(t), \beta_X(t)) dt\right]; V \sim \mu_0, X \sim \nu, X(t) \in \mathcal{A}_\nu^{\nu_T}; \nu \in \mathcal{P}_1(M)\right\} \\ &\leq \underline{B}(\mu_0, \nu_T). \end{aligned}$$

To obtain the reverse inequality, let ν_n be a sequence of measures approximating the infimum in (58). Then for each ν_n , there exists a stochastic process $Z_n \in \mathcal{A}_{\nu_n}^{\nu_T}$ such that

$$\mathbb{E}\left[\int_0^T L(t, Z_n(t), \beta_{Z_n}(Z_n, t)) dt\right] < C_T^s(\nu_n, \nu_T) + \frac{1}{n}. \quad (60)$$

Similarly, let $d\gamma_x^n(v) \otimes d\nu_n(x) = d\gamma_n(v, x)$ be the disintegration of a measure γ_n such that

$$\int \langle v, x \rangle d\gamma_n(v, x) < \underline{W}(\mu_0, \nu_n) + \frac{1}{n},$$

and define $U_n : M \times \Omega \rightarrow M^*$ to be a random variable such that $U_n[x] \sim \gamma_x^n$ for ν_n -a.a. x . Thus $(U_n[Z_n(0)], Z_n(0)) \sim \gamma_n$ and we have constructed a random variable that approximates the interpolation, as

$$\mathbb{E}\left[\langle U[Z_n(0)], Z_n(0) \rangle + \int_0^T L(t, Z_n(t), \beta_{Z_n}(t)) dt\right] \leq \inf\{\underline{W}(\mu_0, \nu) + C_T^s(\nu, \nu_T); \nu \in \mathcal{P}_1(M)\} + \frac{3}{n}. \quad (61)$$

To show that the infimum in ν is attained in the set $\mathcal{P}_1(M)$, we need again to prove the following coercivity property.

Claim: For any fixed $\nu_T \in \mathcal{P}_1(M)$, $N \in \mathbb{R}$, the set of measures $\nu \in \mathcal{P}_1(M)$ satisfying $C(\nu, \nu_T) \leq N \int |x| d\nu(x)$ is tight.

We will assume $\nu \in \mathcal{T}_{\epsilon, R} := \{\nu \in \mathcal{P}_1(M) : \nu(B(R, 0)^c) > \epsilon\}$ for what follows. We leave R to be defined later, but note that if we define the set $\Omega_R := \{|X(0)| > R\}$, then our assumption on ν yields $\mathbb{P}(\Omega_R) > \epsilon$. By positivity of L , this allows us to say that $\mathcal{A}(X) \geq \mathcal{A}(1_{\Omega_R} X)$ (henceforth we define the process $Y(t) := 1_{\Omega_R} X(t)$).

By (A1), we assume that there is a convex function $\underline{L} : M^* \rightarrow \mathbb{R}$ and $C > 0$ such that for all $|u| > U$, $\frac{\underline{L}(u)}{|u|^2} > C$. Recall that $\underline{L}(|v|)$ is a lower bound on $L(t, x, v)$. This imposes a lower bound on the expected action of Y :

$$\begin{aligned} \mathbb{E}\left[\int_0^T L(t, Y, \beta_Y(t, Y)) dt\right] &\geq \mathbb{E}\left[\int_0^T \underline{L}(|\beta_Y(t, Y)|) dt\right] \stackrel{(J)}{\geq} \mathbb{E}[\underline{L}(|V|)T] \\ &\stackrel{(A1)}{>} CT\mathbb{E}\left[\mathbf{1}_{|V|>U}|V|^2\right] \geq CT\left(\mathbb{E}[|V|^2] - U^2\right), \end{aligned} \quad (62)$$

where $\beta_Y := 1_{\Omega_R}\beta_X$ is the drift associated with the process Y and $V := (Y(T) - Y(0))/T$ is its time-average. Hence the expected action of the stochastic process X is bounded:

$$\mathcal{A}(X) \geq \mathcal{A}(Y) > CT\mathbb{E}\left[\left|\frac{Y(T) - Y(0)}{T}\right|^2\right] - CU^2T > \frac{C}{T}\mathbb{E}\left[||Y(0)| - |Y(T)||^2\right] - CU^2T. \quad (63)$$

This leaves us with the same formulation as in (43) of the deterministic coercivity result, the remainder of the proof is identical, and the claim is proved.

To show that a minimizing sequence ν_n is sequentially compact in the weak topology, we use the fact that the set of measures ν such that $C(\nu, \nu_T) < N \int |y| d\nu(y) + \underline{B}(\mu_0, \nu_T) + 1$ is tight. If we let $N := \int |x| d\mu_0(x)$, then the collection of measures such that

$$\begin{aligned} \underline{B}_T^s(\mu_0, \nu_T) + 1 &> C(\nu, \nu_T) + \underline{W}(\mu_0, \nu) \\ &> C(\nu, \nu_T) - \int |x||y| d\mu_0(x) d\nu(y) \\ &\stackrel{\text{(F)}}{=} C(\nu, \nu_T) - N \int |y| d\nu(y), \end{aligned}$$

is tight, where (F) is an application of Fubini's theorem. Thus, by Prokhorov's theorem the minimizing sequence of interpolating measures necessarily weakly converges to a minimizing measure.

Remark 2 a) The same reasoning as in Section 2 yields that $C(\nu_0, \nu_1) = C(\nu_1, \nu_0) = \infty$ for $\nu_1 \in \mathcal{P}_1(M)$ and $\nu_0 \in \mathcal{P}(M) \setminus \mathcal{P}_1(M)$. This implies that it suffices to take the infimum in (58) over $\mathcal{P}_1(M)$. b) The attainment of a minimizing interpolating measure ν_0 is sufficient to show the existence of a minimizing (V, X) for $\underline{B}_T^s(\nu_0, \nu_T)$ whenever the latter is finite. This is a consequence of the existence of minimizers for both $\underline{W}(\mu_0, \nu_0)$ and $C_T^s(\nu_0, \nu_T)$ [22, Proposition 2.1].

To establish the duality formula, we will proceed as in the deterministic case and use the Legendre dual of the optimal cost functional $\nu \rightarrow C_T^s(\nu_0, \nu)$, which was derived by Mikami and Thieullen [22]. Indeed, they show that if the Lagrangian satisfies (A1)-(A3), then

$$C_T^s(\nu_0, \nu_T) = \sup \left\{ \int_M f d\nu_T - \int_M \Psi_{f,-}^0 d\nu_0; f \in C_b^\infty \right\}, \quad (64)$$

where $\Psi_{f,-}$ is the unique solution to the Hamilton-Jacobi-Bellman equation (27) that is given by:

$$\Psi_{f,-}(t, x) = \sup_{X \in \mathcal{A}} \left\{ \mathbb{E} \left[f(X(T)) - \int_t^T L(s, X(s), \beta_X(s, X)) ds \mid X(t) = x \right] \right\}. \quad (65)$$

Moreover, there exists an optimal process X with drift $\beta_X(t, X) = \operatorname{argmin}_v \{v \cdot \nabla \varphi(t, x) + L(t, x, v)\}$. Furthermore, $(\mu, \nu) \mapsto C_T^s(\mu, \nu)$ is convex and lower semi-continuous under the weak*-topology. It follows that $\nu \mapsto \underline{B}_T^s(\mu_0, \nu)$ is weak*-lower semi-continuous on $\mathcal{P}_1(M)$ for all $\mu_0 \in \mathcal{P}_1(M^*)$, and that $(\mu_0, \nu_T) \mapsto \underline{B}_T^s(\mu_0, \nu_T)$ is jointly convex.

Remark 3 Note that integrating $\Psi_{f,+}^0$ over $d\nu_0$ yields the Legendre transform of $\nu_T \mapsto C(\nu_0, \nu_T)$ for $f \in C_{\text{db}}^\infty$.

For $\mu_0 \in \mathcal{P}_1(M^*)$, define the function $\underline{B}_{\mu_0} : \mathcal{M}_1(M) \rightarrow \mathbb{R} \cup \{\infty\}$ to be

$$\underline{B}_{\mu_0}(\nu) := \begin{cases} \underline{B}(\mu_0, \nu) & \nu \in \mathcal{P}_1(M) \\ \infty & \text{otherwise.} \end{cases}$$

Since \underline{B}_{μ_0} is convex and weak*-lower semi-continuous, we have

$$\underline{B}_{\mu_0}(\nu) = \underline{B}_{\mu_0}^{**}(\nu) = \sup_{f \in \text{Lip}(M)} \left\{ \int f d\nu - \underline{B}_{\mu_0}^*(f) \right\}. \quad (66)$$

We break this into two steps. First we show that when $f \in C_{\text{db}}^\infty$ the dual is appropriate:

$$\begin{aligned}
\underline{B}_{\mu_0}^*(f) &:= \sup_{\nu_T \in \mathcal{P}_1(M)} \left\{ \int f d\nu_T - \underline{B}(\mu_0, \nu_T) \right\} \\
&\stackrel{(58)}{=} \sup_{\substack{\nu_T \in \mathcal{P}_1(M) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int f d\nu_T - C(\nu, \nu_T) - \underline{W}(\mu_0, \nu) \right\} \\
&\stackrel{(65)}{=} \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int \Psi_{f,-}^0(x) d\nu(x) - \underline{W}(\mu_0, \nu) \right\} \\
&= \underline{W}_{\mu_0}^*(\Psi_{f,-}^0) = - \int (\Psi_{f,-}^0)_* d\mu_0.
\end{aligned} \tag{67}$$

Thus, plugging this into our dual formula (66) and restricting our supremum to C_{db}^∞ gives

$$\underline{B}_{\mu_0}(\nu) = \underline{B}_{\mu_0}^{**}(\nu) \geq \sup_{f \in C_{\text{db}}^\infty} \left\{ \int f d\nu + \int (\Psi_{f,-}^0)_* d\mu_0 \right\}.$$

To show the reverse inequality we will adapt the mollification argument used in [22, Proof of Theorem 2.1]. We assume our mollifier $\eta_\epsilon(x)$ is such that $\eta_1(x)$ is a smooth function on $[-1, 1]^d$ that satisfies $\int \eta_1(x) dx = 1$ and $\int x \eta_1(x) dx = 0$, then define $\eta_\epsilon(x) = \epsilon^{-d} \eta_1(x/\epsilon)$. Then for Lipschitz f , $f_\epsilon := f * \eta_\epsilon$ is smooth with bounded derivatives. We can derive a bound on $\underline{B}_{\mu * \eta_\epsilon}^*(f)$ by removing the supremum in (67) and fixing a process $X \in \mathcal{A}^{\nu_T}$:

$$\begin{aligned}
&\mathbb{E} \left[f_\epsilon(X(T)) - \int_0^T L(s, X(s), \beta_X(s, X)) ds - \langle X(0), V \rangle \right] \stackrel{(A2)}{\leq} \\
&\mathbb{E} \left[f(X(T) + H_\epsilon) - \int_0^T \frac{L(s, X(s) + H_\epsilon, \beta_X(s, X)) - \Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} ds - \langle X(0) + H_\epsilon, V + H_\epsilon \rangle + |H_\epsilon|^2 \right] \leq \\
&\qquad \qquad \qquad \frac{D_\epsilon^*(f(1 + \Delta L(0, \epsilon)))}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} + d\epsilon^2,
\end{aligned}$$

where $D_\epsilon(\nu) := \inf\{(1 + \Delta L(0, \epsilon))\underline{W}(\mu_\epsilon, \nu_0) + C(\nu_0, \nu); \nu_0\}$, $H_\epsilon \sim \eta_\epsilon$ is independent of $X(\cdot), V$, thus $X(T) + H_\epsilon \sim d(\eta_\epsilon * \nu_T)$. The third line arises by maximizing over processes $(X(\cdot) + H_\epsilon, V + H_\epsilon)$. Note that $\epsilon \mapsto D_\epsilon(\nu)$ is lower semi-continuous for the same reason that $\nu \mapsto \underline{B}_\mu(\nu)$ is, and converges to $\underline{B}_{\mu_0}(\nu)$ as $\epsilon \rightarrow 0$.

Taking the supremum over $X \in \mathcal{A}_{\mu_0}$ of the left side above, we can retrieve a bound on $\underline{B}_{\mu_0}^*(f_\epsilon)$. This bound allows us to say

$$\int f_\epsilon d\nu - \underline{B}_\mu^*(f_\epsilon) \geq \int f d\nu_\epsilon - \frac{D_\epsilon^*(f(1 + \Delta L(0, \epsilon)))}{1 + \Delta L(0, \epsilon)} - T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} - d\epsilon^2,$$

where we use ϵ -subscript to indicate convolution of a measure with η_ϵ . Taking the supremum over $f \in \text{Lip}(M)$, we get the reverse inequality:

$$\sup_{f \in C_{\text{db}}^\infty} \left\{ \int f d\nu - \underline{B}_\mu^*(f) \right\} \geq \frac{D_\epsilon(\nu_\epsilon)}{1 + \Delta L(0, \epsilon)} - T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} - d\epsilon^2 \stackrel{\epsilon \searrow 0}{\geq} \underline{B}(\mu_0, \nu_T),$$

In the following corollary, we will discuss results pertaining to solutions $\psi_n^t(x) := \psi_n(t, x)$ of the Hamilton-Jacobi-Bellman equation for final conditions $\psi_n^T(x)$. In some sense $\nabla \psi$ is more fundamental than ψ , since our dual is invariant under $\psi \mapsto \psi + c$. Thus when discussing the convergence of a sequence of ψ , we refer to the convergence of their gradients. Notably the optimal gradient may not be bounded or smooth, hence may not be achieved within the set C_{db}^∞ . In the subsequent corollary, we denote \mathbb{P}_X the measure on $M \times [0, T]$ associated with the process X .

Corollary 4 *Suppose the assumptions on Theorem 4.2 are satisfied and that μ_0 is absolutely continuous with respect to Lebesgue measure. Then $(V, X(t))$ minimizes $\underline{B}(\mu_0, \nu_T)$ if and only if it is a solution to the stochastic differential equation*

$$dX = \nabla_p H(t, X, \nabla \psi(t, X)) dt + dW_t \quad (68)$$

$$V = \nabla \bar{\psi}(X(0)), \quad (69)$$

where $\nabla \psi_n(t, x) \rightarrow \nabla \psi(t, x)$ \mathbb{P}_X -a.s. and $\nabla \psi_n(0, x) \rightarrow \nabla \bar{\psi}(x)$ ν_0 -a.s. for some sequence $\psi_n(t, x)$ that solves (HJB) in such a way that $\psi_n^T := \psi_n(T, \cdot)$ and $(\psi_n^0)_* := [\psi_n(0, \cdot)]_*$ are maximizing sequences for the dual problem (59). Furthermore $\bar{\psi}$ is concave.

Proof: First note that there exists such an optimal pair (V, X) , in view of Theorem 4.1. Moreover, the pair is optimal iff there exists a sequence of solutions ψ_n to HJB that is maximizing in (59) such that

$$\mathbb{E} \left[\int_0^T L(t, X, \beta_X(t, X)) dt + \langle X(0), V \rangle \right] = \lim_{n \rightarrow \infty} \mathbb{E} [\psi_n^T(X(T)) + (\psi_n^0)_*(V)], \quad (70)$$

which we can write as

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\underbrace{\psi_n^T(X(T)) - \psi_n^0(X(0))}_{(a)} + \underbrace{\psi_n^0(X(0)) - (\psi_n^0)_{**}(X(0))}_{(b)} + \underbrace{(\psi_n^0)_{**}(X(0)) + (\psi_n^0)_*(V)}_{(c)} \right], \quad (71)$$

where f_{**} is the concave hull of f . Applying Itô's formula to the first two terms, with the knowledge that they satisfy (HJB), we get

$$\mathbb{E} [\psi_n^T(X(T)) - \psi_n^0(X(0))] = \mathbb{E} \left[\int_0^T \langle \beta_X, \nabla \psi_n^t(X(t)) \rangle - H(t, X, \nabla \psi_n^t(X(t))) dt \right].$$

However, by the definition of the Hamiltonian, we have $\langle v, b \rangle - H(t, x, v) \leq L(t, x, b)$, which mean that (71) yield the following three inequalities:

$$\langle \beta_X, \nabla \psi_n^t(X(t)) \rangle - H(t, X, \nabla \psi_n^t(X(t))) \leq L(t, X, \beta_X(t, X)) \quad (a)$$

$$\psi_n^0(X(0)) - (\psi_n^0)_{**}(X(0)) \leq 0 \quad (b)$$

$$(\psi_n^0)_{**}(X(0)) + (\psi_n^0)_*(V) \leq \langle V, X(0) \rangle. \quad (c)$$

In other words, (71) breaks the problem into a stochastic and a Wasserstein transport problem (in the flavour of Theorem 4), along with a correction term to account for ψ_n^0 not being necessarily concave. Adding (70) to the mix, allows us to obtain L^1 convergence in the (a,b,c) inequalities, hence a.s. convergence of a subsequence ψ_{n_k} .

Note that the convergence in (b,c) means that ψ_n^0 converges ν_0 -a.s. to a concave function $\bar{\psi}$ such that $x \mapsto \nabla \bar{\psi}$ is the optimal transport plan for $\underline{W}(\nu_0, \mu_0)$ [8].

To obtain the optimal control for the stochastic process, one needs the uniqueness of the point p achieving equality in (a). This is a consequence of the strict convexity and coercivity of $b \mapsto L(t, x, b)$ for all t, x . The differentiability of L further ensures this value is achieved by $p = \nabla_v L(t, x, b)$. Hence (a) holds iff

$$\nabla \psi_n^t(X_t) \longrightarrow \nabla_v L(t, X, \beta_X(t, X)) \quad \mathbb{P}_X\text{-a.s.}$$

Since ψ_n^t are deterministic functions, this demonstrates that X_t is a Markov process with drift β_X determined by the inverse transform: $\beta_X(t, X) = \nabla_p H(t, X, \nabla \psi(t, X))$, i.e., (68).

Remark 4 *It is not possible to conclude from the above work that $\bar{\psi}(x) = \psi(0, x)$ without a regularity result on $t \mapsto \psi(t, x)$ for the optimal ψ . This is because $\bar{\psi}$ is defined on a \mathbb{P}_X -null set.*

4 Deterministic and stochastic Bolza duality

For the rest of the paper, we shall assume that the Lagrangian L is independent of time, but that it is convex, proper and lower semi-continuous in both variables. We then consider the dual Lagrangian \tilde{L} defined on $M^* \times M^*$ by

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

the corresponding fixed-end costs on $M^* \times M^*$,

$$\tilde{c}_T(u, v) := \inf\left\{\int_0^T \tilde{L}(\gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M^*); \gamma(0) = u, \gamma(T) = v\right\}, \quad (72)$$

and its associated optimal transport

$$\tilde{C}_T(\mu_0, \mu_T) := \inf\left\{\int_{M^* \times M^*} \tilde{c}_T(x, y) d\pi; \pi \in \mathcal{K}(\mu_0, \mu_T)\right\}. \quad (73)$$

More specifically, we shall assume the following conditions on L , which are weaker than (A1), (A2), (A3) but for the crucial condition that L is convex in both variables.

(B1) $L : M \times M \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex, proper and lower semi-continuous in both variables.

(B2) The set $F(x) := \{p; L(x, p) < \infty\}$ is non-empty for all $x \in M$, and for some $\varrho > 0$, we have $\text{dist}(0, F(x)) \leq \varrho(1 + |x|)$ for all $x \in M$.

(B3) 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 $[0, \infty)$.

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

$$\psi(x) - (\gamma|x| + \delta)|q| \leq H(x, q) \leq \varphi(q) + (\alpha|q| + \beta)|x| \text{ for all } x, q \text{ in } M \times M^*, \quad (74)$$

where $\alpha, \beta, \gamma, \delta$ are constants, φ is finite and convex and ψ is finite and concave (see [29]).

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 $[0, \infty) \times M \times M^*$. However, the main addition in the case of joint convexity for L is the following so-called Bolza duality that we briefly describe in the deterministic case since it had been studied in-depth in various articles by T. Rockafellar [27] and co-authors [28, 29]. The stochastic counterpart is more recent and has been established by Boroushaki and Ghoussoub [9].

We consider the path space $\mathcal{A}_M^2 := \mathcal{A}_M^2[0, T] = \{u : [0, T] \rightarrow M; \dot{u} \in L_M^2\}$ equipped with the norm

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

Let L be a convex Lagrangian on $M \times M$ as above, ℓ be a proper convex lower semi-continuous function on $M \times M$ and consider the minimization problems,

$$(\mathcal{P}) \quad \inf \left\{ \int_0^T L(\gamma(s), \dot{\gamma}(s)) ds + \ell(\gamma(0), \gamma(T)); \gamma \in C^1([0, T], M) \right\}, \quad (75)$$

and

$$(\tilde{\mathcal{P}}) \quad \inf \left\{ \int_0^T \tilde{L}(\gamma(s), \dot{\gamma}(s)) ds + \ell^*(\gamma(0), -\gamma(T)); \gamma \in C^1([0, T], M) \right\}. \quad (76)$$

Theorem 5 *Assume L satisfies (B1), (B2) and (B3), and that ℓ is proper, lsc and convex.*

1. *If there exists ξ such that $\ell(\cdot, \xi)$ is finite, or there exists ξ' such that $\ell(\xi', \cdot)$ is finite, then*

$$\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}}).$$

This value is not $+\infty$, and if it is also not $-\infty$, then there is an optimal arc $v(t) \in \mathcal{A}^2[0, T]$ for $(\tilde{\mathcal{P}})$.

2. *A similar statement holds if we replace ℓ by $\tilde{\ell}$ in the above hypothesis and $(\tilde{\mathcal{P}})$ by (\mathcal{P}) in the conclusion.*
3. *If both conditions are satisfied, then both $(\tilde{\mathcal{P}})$ and (\mathcal{P}) are attained respectively by optimal arcs $v(t), x(t)$ in $\mathcal{A}^2[0, T]$.*

In this case, these arcs satisfy $(\dot{v}(t), v(t)) \in \partial L(x(t), \dot{x}(t))$ for a.e. t , which can also be written in a dual form $(\dot{x}(t), x(t)) \in \partial \tilde{L}(v(t), \dot{v}(t))$ for a.e. t , or in a Hamiltonian form as

$$\dot{x}(t) \in \partial_v H(x(t), v(t)) \tag{77}$$

$$-\dot{v}(t) \in \tilde{\partial}_x H(x(t), v(t)), \tag{78}$$

coupled with the boundary conditions

$$(v(0), -v(T)) \in \partial \ell(x(0), x(T)). \tag{79}$$

See for example [27]. The above duality has several consequences.

Proposition 5 *The value function $\Phi_{g,+}(x) = \inf\{g(y) + c_t(y, x); y \in M\}$, which is the variational solution of the Hamilton-Jacobi equation (19) starting at g , can be expressed in terms of the b and \tilde{c} costs as follows:*

1. *If g is convex and lower semi-continuous, then $\Phi_{g,+}(t, x) = \sup\{b_t(v, x) - g^*(v); v \in M^*\}$ is convex lower semi-continuous for every $t \in [0, +\infty)$.*
2. *The convex Legendre transform of $\Phi_{g,+}$ is given by the formula*

$$\tilde{\Phi}_{g^*,+}(t, w) = \inf\{g^*(v) + \tilde{c}_t(v, w); v \in M^*\}.$$

3. *For each t , the graph of the subgradient $\partial \Phi_{g,+}(t, \cdot)$, i.e., $\Gamma_g(t) = \{(x, v); v \in \partial \Phi_{g,+}(t, x)\}$ is a globally Lipschitz manifold of dimension n in $M \times M^*$, which depends continuously on t .*
4. *If a Hamiltonian trajectory $(x(t), v(t))$ over $[0, T]$ starts with $v(0) \in \partial g(x(0))$, then $v(t) \in \partial \Phi_{g,+}(t, x(t))$ for all $t \in [0, T]$. Moreover, this happens if and only if $x(t)$ is optimal in the minimization problem that defines $\Phi_{g,+}(t, x)$ and $v(t)$ is optimal in the minimization problem that defines $\tilde{\Phi}_{g^*,+}(t, w)$.*

Remark 5 The above shows that in the case when L is jointly convex, the corresponding forward Hamilton-Jacobi equation has convex solutions whenever the initial state is convex, while the corresponding backward Hamilton-Jacobi equation has concave solutions if the final state is concave. Unfortunately, we shall see that in the mass transport problems we are considering, one mostly propagates concave (resp., concave) functions forward (resp., backward), hence losing their concavity (resp., convexity).

This said, the cost functionals c_T, \tilde{c}_T, b_T are all value functions Φ_g starting or ending with affine function g . Indeed, $b_t(v, x) = \Phi_{g,+}(t, x)$, 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 $x \rightarrow \Phi_{g,+}(t, x) = b_t(v, x)$ is $w \rightarrow \tilde{c}_t(v, w)$. One can also deduce the following.

Proposition 6 Under assumptions $(B_1), (B_2), (B_3)$ 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 on M^* , while $x \rightarrow b_t(v, x)$ is convex 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 $b_t(v, x) = \sup\{\langle w, x \rangle - \tilde{c}_t(v, w); w \in M^*\}$.
4. 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) .

This leads us to the following standard condition in optimal transport theory.

Definition 7 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 .

The stochastic Bolza duality and its applications

We now deal with the stochastic case. We define the Itô space \mathcal{I}_M^p consisting of all M -valued processes of the following form:

$$\begin{aligned} \mathcal{I}_M^p = \left\{ X : \Omega_T \rightarrow M; X(t) = X_0 + \int_0^t \beta_X(s) ds + \int_0^t \sigma_X(s) dW(s), \right. \\ \left. \text{for } X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; M), \beta_X \in L^p(\Omega_T; M), \sigma_X \in L^2(\Omega_T; M) \right\}, \end{aligned} \quad (80)$$

where β_X and σ_X are both progressively measurable and $\Omega_T := \Omega \times [0, T]$. The cases of $p = 1, 2, \infty$ will be of interest to us. We equip \mathcal{I}_M^2 with the norm

$$\|X\|_{\mathcal{I}_M^2}^2 = \mathbb{E} \left(\|X(0)\|_M^2 + \int_0^T \|\beta_X(t)\|_M^2 dt + \int_0^T \|\sigma_X(t)\|_M^2 dt \right),$$

so that it becomes a Hilbert space. The dual space $(\mathcal{I}_M^2)^*$ can also be identified with $L^2(\Omega; M) \times L^2(\Omega_T; M) \times L^2(\Omega_T; M)$. In other words, each $q \in (\mathcal{I}_M^2)^*$ can be represented by the triplet

$$q = (q_0, q_1(t), Q(t)) \in L^2(\Omega; M) \times L^2(\Omega_T; M) \times L^2(\Omega_T; M),$$

in such a way that the duality can be written as:

$$\langle X, q \rangle_{\mathcal{I}_M^2 \times (\mathcal{I}_M^2)^*} = \mathbb{E} \left\{ \langle q_0, X(0) \rangle_M + \int_0^T \langle q_1(t), \beta_X(t) \rangle_M dt + \frac{1}{2} \int_0^T \langle Q(t), \sigma_X(t) \rangle_M dt \right\}. \quad (81)$$

Similarly, the dual of \mathcal{I}_M^1 can be identified with \mathcal{I}_M^∞ .

We shall use the following result recently established in [9].

Theorem 6 (Boroushaki-Ghoussoub) *Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a complete probability space with normal filtration, and let $L(\cdot, \cdot)$ and M be two jointly convex Lagrangians on $M \times M$, Assume ℓ is a convex lsc function on $M \times M$. Consider the Lagrangian on $\mathcal{I}_M^2 \times (\mathcal{I}_M^2)^*$ defined by*

$$\begin{aligned} \mathcal{L}(X, p) = \mathbb{E} \left\{ \int_0^T L(X(t) - p_1(t), -\beta_X(t)) dt + \ell(X(0) - p_0, X(T)) \right. \\ \left. + \frac{1}{2} \int_0^T M(\sigma_X(t) - P(t), -\sigma_X(t)) dt \right\}. \end{aligned} \quad (82)$$

Its Legendre dual is then given for each $q := (0, q_1, Q)$ by

$$\begin{aligned} \mathcal{L}^*(q, Y) = \mathbb{E} \left\{ \ell^*(-Y(0), Y(T)) + \int_0^T L^*(-\beta_Y(t), Y(t) - q_1(t)) dt \right. \\ \left. + \frac{1}{2} \int_0^T M^*(-\sigma_Y(t), \sigma_Y(t) - Q(t)) dt \right\}. \end{aligned}$$

Note that standard duality theory implies that in general

$$\inf_{X \in \mathcal{I}^2} \{\mathcal{L}(X, 0)\} \geq \sup_{Y \in (\mathcal{I}^2)^*} \{-\mathcal{L}(0, Y)\}. \quad (83)$$

In our case we shall restrict ourselves to processes of fixed diffusion. This facilitates the proving of a stochastic analog to Bolza duality:

Proposition 8 *Assume L satisfies (A1) and (A2), and there exists (a.s.-)unique $V_0 \in L^2(\mathbb{P})$ such that $\ell^*(V_0, \cdot) < \infty$ and (a.s.-)unique $\sigma_V \in L^2(\mathbb{P} \times \lambda_{[0, T]})$ such that $M^*(\sigma_V, \cdot) < \infty$, then there is no duality gap, ie.*

$$\inf_{X \in \mathcal{I}^2} \{\mathcal{L}(X, 0)\} = \sup_{Y \in (\mathcal{I}^2)^*} \{-\mathcal{L}^*(0, Y)\} \quad (84)$$

Note that, unlike the deterministic case, there there is no backwards condition that works if there is an $V_T \in L^2(\mathbb{P})$ such that $\ell^*(\cdot, V_T) < \infty$, this is because stochastic processes, in general, are irreversible.

Proof: We begin with augmenting our space by considering $\beta_V \in L^1(\mathbb{P} \times \lambda_{[0, T]})$ —we call this augmented set \mathcal{I}^1 . If we can show the duality gap is satisfied in \mathcal{I}^1 , by our coercivity condition (A2) we can then show that it must be satisfied in \mathcal{I}^2 .

We proceed by a variational method outlined by Rockafellar [27]. First, we define

$$\varphi(q) := \inf_{Y \in (\mathcal{I}^1)^*} \{\mathcal{L}^*(q, Y)\}. \quad (85)$$

As the infimum of a jointly convex function, φ itself is convex. The benefit of this definition is that

$$\varphi^*(X) = \sup_{q, v} \{\langle X, q \rangle - \mathcal{L}^*(q, v)\} = \mathcal{L}^{**}(X, 0) = \mathcal{L}(X, 0). \quad (86)$$

Hence, X minimizes \mathcal{L} if and only if

$$X \in \partial\varphi(0) \iff \varphi(0) + \varphi^*(X) = 0 \iff \mathcal{L}(X, 0) = - \inf_{Y \in (\mathcal{I}^2)^*} \{\mathcal{L}^*(0, Y)\}. \quad (87)$$

In other words, there is no duality gap if and only if $\partial\varphi(0)$ is non-empty. Note that this holds if there is an open (relative to $\{q; \varphi(q) < \infty\}$) neighbourhood N of the origin in \mathcal{I}^∞ such that $\mathcal{L}^*(q, Y) < \infty$ for $q \in N$.

By our assumptions, we may fix Y_0, σ_Y to be the unique elements such that $\ell(Y_0, \cdot) < \infty$ and $M^*(\sigma_Y, \cdot) < \infty$ (guaranteeing subdifferentiability in these variables), and let $Y = (Y_0, \beta_Y, \sigma_Y)$ be

such that $\mathcal{L}^*(0, Y) < \infty$. For a perturbation $\beta_V \in L^\infty(\mathbb{P} \times \lambda_{[0, T]})$ with $\|\beta_V\|_\infty < \epsilon$, note that (A2) gives for all $(t, u) \in [0, T] \times M^*$,

$$L(t, Y_t - \beta_V, u) < (1 + \Delta L(0, \epsilon)) L(t, Y_t, u) + \Delta L(0, \epsilon), \quad (88)$$

and

$$\begin{aligned} \varphi(V) &= \inf_{Y \in \mathcal{I}_M^1} \mathcal{L}^*(V, Y) \\ &\leq \mathbb{E} \mathcal{L}^*(-Y_0, Y_T) + \mathbb{E} \int_0^T \tilde{L}(t, Y_t - \beta_V(t), \beta_Y(t)) dt \\ &\leq \mathbb{E} \mathcal{L}^*(-Y_0, Y_T) + (1 + \Delta \tilde{L}(0, \epsilon)) \mathbb{E} \int_0^T \tilde{L}(t, Y_t, \beta_Y(t)) dt + T \Delta \tilde{L}(0, \epsilon), \end{aligned} \quad (89)$$

which is finite for $\|\beta_V\|_\infty < \epsilon$ sufficiently small by (A2). Hence φ is finite and continuous in a open set of the origin (all relative to its domain), and duality is achieved on \mathcal{I}^1 .

To show that this duality is achieved in \mathcal{I}^2 , it suffices to remark that $\mathbb{E} \int_0^T L(t, Y_t, \beta_Y) \geq \mathbb{E} \int \underline{L}(\beta_Y) dt \geq C \mathbb{E} \int |\beta_Y|^2 - B dt = \infty$ for $\beta_Y \in L^1(\mathbb{P} \times \lambda_{[0, T]}) \setminus L^2(\mathbb{P} \times \lambda_{[0, T]})$ (where C, B are fixed constants).

5 Maximizing the ballistic cost: Deterministic case

With Bolza duality in mind, we can now turn to the maximizing ballistic cost.

Theorem 7 *Assume that L satisfies hypothesis (B1), (B2) and (B3), and let ν_T be a probability measure with compact support on M , that is also absolutely continuous with respect to Lebesgue measure. Then,*

1. *The following interpolation formula holds:*

$$\overline{B}_T(\mu_0, \nu_T) = \sup\{\overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (90)$$

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.*

2. *We also have the following duality formulae:*

$$\overline{B}_T(\mu_0, \nu_T) = \inf \left\{ \int_M h(x) d\nu_T(x) + \int_{M^*} \tilde{\Phi}_{h^*, -}^0(v) d\mu_0(v); h \text{ convex in } Lip(M) \right\}. \quad (91)$$

and

$$\overline{B}_T(\mu_0, \nu_T) = \inf \left\{ \int_M (\tilde{\Phi}_{g, +}^T)^*(x) d\nu_T(x) + \int_{M^*} g(v) d\mu_0(v); g \text{ in } Lip(M^*) \right\}. \quad (92)$$

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

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(S_T^* \circ \nabla h^*(x), x) d\nu_T(x), \quad (93)$$

where $S_T^*(v) = \pi^* \varphi_T^{H^*}(v, \nabla h)$, 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 $x \rightarrow \pi^* \varphi_T^{H^*}(\nabla h^*(x), x)$.

4. We also have

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v), \quad (94)$$

where $\tilde{S}_T(v) = \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v h_0)$, and $\varphi_t^{\tilde{H}}$ being the Hamiltonian flow associated to \tilde{L} (i.e., $\tilde{H}(v, x) = -H(x, v)$, and $h_0 = \tilde{\Phi}_{h^*, -}^0$).

h_0 the solution $h(0, v)$ of the backward Hamilton-Jacobi equation (33) with $h(T, v) = h(v)$.

Proof: To show (90) and (91), first note that for any probability measure μ on M^* , we have

$$\overline{B}_T(\mu_0, \nu_T) \geq \overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu). \quad (95)$$

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

$$b_T(v, x) + \tilde{c}_T(v, \nabla h(x)) = \langle x, \nabla h(x) \rangle \text{ on } \mathcal{O}. \quad (96)$$

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

$$\int_{M^* \times M} b_T(v, x) d\pi_0 + \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \int_M \langle x, \nabla h(x) \rangle d\nu_T. \quad (97)$$

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

$$\overline{B}_T(\mu_0, \nu_T) + \int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \overline{W}(\nu_T, \mu_T), \quad (98)$$

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

$$\int_{M^* \times M} \tilde{c}_T(v, \nabla h(x)) d\pi_0 = \tilde{C}_T(\mu_0, \mu_T). \quad (99)$$

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, (98) 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 $h^*(w) + g(v) \leq \tilde{c}_T(v, w)$. Indeed, 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 $h^*(w) + g(v) \leq \tilde{c}_T(v, w)$, 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 (95) 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.

The first duality formula (92) follows since we have established that if (g, h) are an optimal pair of Kantorovich functions for $\overline{B}_T(\mu_0, \nu_T)$, then (g, h^*) are an optimal pair of Kantorovich functions for $\tilde{C}_T(\mu_0, \mu_T)$. In other words, the initial Kantorovich function for $\overline{B}_T(\mu_0, \nu_T)$ is $g = \tilde{\Phi}_{h^*, -}(0, \cdot)$. This proves formula (91).

To show (92), we can now that the interpolation (90) is established—proceed as in Section 2, by identifying the Legendre transform of the functionals $\nu \rightarrow \overline{W}(\nu, \nu_T)$ and $\mu \rightarrow \tilde{C}_T(\mu, \mu_T)$.

To show part 3), we start with the interpolation inequality and write that

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

for some probability measure μ_T . The proof also shows that there exists a convex function $h : M^* \rightarrow \mathbb{R}$ and another function $k : M^* \rightarrow \mathbb{R}$ such that $(\nabla h)_\# \mu_T = \nu_T$, $\bar{W}(\nu_T, \mu_T) = \int_M \langle \nabla h(v), v \rangle d\mu_T(v)$. and $\tilde{C}_T(\mu_0, \mu_T) = \int_{M^*} h(u) d\mu_T(u) - \int_{M^*} k(v) d\mu_0(v)$. Now use the theorem of Fathi-Figalli to write

$$\tilde{C}_T(\mu_0, \mu_T) = \int_{M^*} c_T(v, \tilde{S}_T v) d\mu_0(v), \quad (100)$$

where $\tilde{S}_T(v) = \pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v k)$. Note that

$$\bar{B}_T(\mu_0, \nu_T) \geq \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v), \quad (101)$$

since $(\tilde{S}_T)_\# \mu_0 = \mu_T$ and $\nabla h_\# \mu_T = \nu_T$, 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 $b_T(u, x) \geq \langle \nabla h(v), x \rangle - \tilde{c}_T(u, \nabla h(v))$ for every $v \in M^*$, we have

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

It follows that $\bar{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v)$.

To get (3), use the pushforward $\nu_T = (\nabla h \circ \tilde{S}_T)_\# \mu_0$ to write the above in terms of the measure ν_T , using the fact that $(\nabla h)^{-1} = \nabla h^*$ and $\tilde{S}_T^{-1} = S_T^*$ where $S_T^*(v) = \pi^* \varphi_t^{H^*}(v, \tilde{d}_v h)$ and $\varphi_t^{H^*}$ is the Hamiltonian flow associated to the hamiltonian $H_*(v, x) := -H(-x, v)$. This gives us

$$\bar{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(S_T^* \circ \nabla h^*(x), x) d\nu_T(x) = \int_{M^*} b_T(\pi^* \varphi_t^{H^*}(\nabla h^*(x), \tilde{d}_v h), x) d\nu_T(x).$$

Since h is convex, we have that $\tilde{d}_x h = \nabla h(x)$, hence $\tilde{d}_{\nabla h^*(x)} h = \nabla h \circ \nabla h^*(x) = x$, which yields our claim that

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

Remark 6 While the costs c and \tilde{c}_T are themselves jointly convex in both variables, one cannot deduce much in terms of the convexity or concavity of the corresponding Kantorovich potentials. However, we note that the interpolation (41) of $\underline{B}_T(\mu_0, \nu_T)$ selects a ν_0 such that $C_T(\nu_0, \nu_T)$ has a concave initial Kantorovich potential, while the interpolation (90) of $\bar{B}_T(\mu_0, \nu_T)$ selects a μ_T such that $\tilde{C}_T(\mu_0, \mu_T)$ has a convex final Kantorovich potential.

Furthermore, one wonders whether the formula

$$c_t(y, x) = \sup\{b_t(v, x) - \langle v, y \rangle; v \in M^*\}, \quad (102)$$

also extends to Wasserstein space. We show it under the condition that the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave, and conjecture that it is also a necessary condition.

Theorem 8 *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (B1), (B2) and (B3). Assume ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure. If the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave then the following holds:*

$$C_T(\nu_0, \nu_T) = \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu); \mu \in \mathcal{P}(M^*)\}, \quad (103)$$

and the supremum is attained.

Proof: Again, it is easy to show that

$$C_T(\nu_0, \nu_T) \geq \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu); \mu \in \mathcal{P}(M^*)\}. \quad (104)$$

To prove equality, we assume that the initial Kantorovich potential g is concave and 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

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

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

Since g concave and $b_t(v, x) = \inf\{h(x) - g(z) + c_t(z, x); z \in M\}$, this means that for $(y, x) \in \mathcal{O}$,

$$c_T(y, x) = b_T(\nabla g(y), x) - \langle \nabla g(y), y \rangle. \quad (105)$$

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

$$\int_{M \times M} c_T(y, x) d\pi_0 = \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 - \int_M \langle \nabla g(y), y \rangle d\nu_0. \quad (106)$$

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

$$C_T(\nu_0, \nu_T) = \int_{M \times M} b_T(\nabla g(y), x) d\pi_0 - \underline{W}(\nu_0, \mu_0). \quad (107)$$

We now prove that

$$\int_{M \times M} b_T(\nabla g(y), x) d\pi_0 = \underline{B}_T(\mu_0, \nu_T). \quad (108)$$

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, (107) 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, since $h(x) - g(y) \leq c_T(y, x)$, it is easy to see 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 (103) is attained by the measure μ_0 .

Corollary 9 *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (B1), (B2) and (B3). 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*

$$C_T(\nu_0, \nu_T) = \int_M c_T(y, X_0^T \circ \nabla g(y)) d\nu_0(y). \quad (109)$$

Proof: In this case, $C_T(\nu_0, \nu_T) = \underline{B}_T(\mu_0, \nu_T) - \underline{W}(\nu_0, \mu_0)$, 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

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, X_0^T v) d\mu_0(v). \quad (110)$$

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

Corollary 10 *Consider the cost $c_1(y, x) = c(x - y)$, where c 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 $\varphi : M \rightarrow \mathbb{R}$, $\psi : M^* \rightarrow \mathbb{R}$ and a probability measure μ_0 on M^* such that*

$$(\nabla \psi \circ \nabla \varphi)_\# \nu_0 = \nu_1, \quad (111)$$

and

$$C_1(\nu_0, \nu_1) + \int_{M^*} c^*(v) d\mu_0(v) = \int_M c(\nabla \psi \circ \nabla \varphi(y) - y) d\nu_0(y) = \int_M \langle \nabla \psi_*(y) - \nabla \varphi(y), y \rangle d\nu_0(y). \quad (112)$$

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

$$c_1(y, x) = \inf \left\{ \int_0^1 c(\dot{\gamma}(t)) dt; \gamma \in C^1([0, 1], M); \gamma(0) = y, \gamma(1) = x \right\} = c(x - y). \quad (113)$$

It follows from (103) 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

$$C_1(\nu_0, \nu_1) = \underline{B}_1(\mu_0, \nu_1) - \underline{W}(\nu_0, \mu_0) = \underline{W}(\mu_0, \nu_1) - \int_{M^*} c^*(v) d\mu_0(v) - \underline{W}(\nu_0, \mu_0). \quad (114)$$

In other words,

$$C_1(\nu_0, \nu_1) + K = \underline{W}_1(\mu_0, \nu_1) - \underline{W}(\nu_0, \mu_0), \quad (115)$$

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

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

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

It follows from the preceding corollary that

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

Note also that

$$\begin{aligned} C_1(\nu_0, \nu_1) + K &= \int_M \langle v, \nabla\psi(v) \rangle d\mu_0(v) - \int_M \langle y, \nabla\varphi(y) \rangle d\nu_0(y) \\ &= \int_M \langle \nabla\psi_*(y), y \rangle d\nu_0(y) - \int_M \langle y, \nabla\varphi(y) \rangle d\nu_0(y) \\ &= \int_M \langle \nabla\psi_*(y) - \nabla\varphi(y), y \rangle d\nu_0(y). \end{aligned}$$

6 Maximizing the ballistic cost: Stochastic case

Define the transportation cost between two random variables V on M^* and X on M by:

$$b_T^s(V, Y) := \inf \left\{ \mathbb{E} \left[\langle V, X(0) \rangle + \int_0^T L(t, X_t, \beta_X(t)) dt \right]; X \in \mathcal{A}, X(T) = Y \text{ a.s.} \right\}, \quad (116)$$

where \mathcal{A} indicates Itô processes with Brownian diffusion. The minimizing ballistic cost considered earlier is then

$$\bar{B}_T^s(\mu_0, \nu_T) = \inf \{ b_T^s(V, Y); V \sim \mu_0, Y \sim \nu_T \}, \quad (117)$$

while the maximizing cost is defined as:

$$\bar{B}_T^s(\mu_0, \nu_T) := \sup \{ b_T^s(V, Y); V \sim \mu_0, Y \sim \nu_T \}. \quad (118)$$

Theorem 9 *Assume L is a Lagrangian on $M \times M^*$ such that L and its dual \tilde{L} satisfies (A0)-(A3).*

1. *The following formula then holds:*

$$\bar{B}_T^s(\mu_0, \nu_T) := \sup \left\{ \mathbb{E} \left[\langle X, V(T) \rangle - \int_0^T \tilde{L}(t, V, \beta_V(t, V)) dt \right]; V \in \mathcal{A}, V_0 \sim \mu_0, X \sim \nu_T \right\}. \quad (119)$$

2. *The following duality holds:*

$$\bar{B}_T^s(\mu_0, \nu_T) = \sup \{ \bar{W}(\mu, \nu_T) - \tilde{C}_T^s(\mu_0, \mu); \mu \in \mathcal{P}_1(M^*) \}, \quad (120)$$

where \tilde{C}_T^s is the action corresponding to the Lagrangian \tilde{L} . Furthermore, if $\nu_0 \in \mathcal{P}_1(M)$, and $\mu_T \in \mathcal{P}_1(M^*)$ there exist an optimal interpolant μ_T in $\mathcal{P}_1(M^*)$.

3. *If $\mu_0 \in \mathcal{P}_1(M^*)$, ν_T has compact support, and $\bar{B}(\mu_0, \nu_T) < \infty$, then*

$$\bar{B}_s(\mu_0, \nu_T) = \inf \left\{ \int_M g d\nu_T + \int_{M^*} \tilde{\Psi}_{g^*, -}^0 d\mu_0; g \in C_{db}^\infty(M) \text{ and convex} \right\}, \quad (121)$$

where $\tilde{\Psi}_{g^*, -}$ solves the Hamilton-Jacobi-Bellman equation on M^*

$$\frac{\partial\psi}{\partial t} + \frac{1}{2}\Delta\psi - H(t, \nabla_v\psi, v) = 0 \quad \psi(v, T) = g^*(v). \quad (\text{HJB2})$$

Proof: 1) For a fixed pair (V, Y) , we consider the Bolza energy $\mathcal{L}_{(V,Y)}$ –defined in (82)– associated to L and the two Lagrangians ℓ and M defined as:

$$\ell_{(Y,U)}(\omega, y, z) := \begin{cases} \langle z, U(\omega) \rangle & y = Y(\omega) \\ \infty & \text{else} \end{cases} \quad M(\xi, \zeta) := \begin{cases} -\zeta - 1 & \xi = 1 \\ \infty & \text{else.} \end{cases} \quad (122)$$

Note that the minimizing stochastic cost can be written as,

$$\underline{B}_s(\mu_0, \nu_T) := \inf\{\inf\{\mathcal{L}_{(V,Y)}(X(t), 0); X \in \mathcal{I}^2\}; V \sim \mu_0, Y \sim \nu_T\} \quad (123)$$

while the maximizing cost is

$$\overline{B}_s(\mu_0, \nu_T) = \sup\{\inf\{\mathcal{L}_{(V,Y)}(X(t), 0); X \in \mathcal{I}^2\}; V \sim \mu_0, Y \sim \nu_T\}. \quad (124)$$

Applying Bolza duality turns the infimum to a supremum:

$$\overline{B}_s(\mu_0, \nu_T) = \sup\{\sup\{-\mathcal{L}_{(V,Y)}^*(0, U(t)); U \in \mathcal{I}^2\}; V \sim \mu_0, Y \sim \nu_T\}, \quad (125)$$

which results in (119).

2) The proof of the interpolation result can now follow closely the proof for the minimization problem.

3) We again try to identify the Legendre transforms of the functionals $\nu \mapsto \overline{W}(\mu, \nu)$ and $\mu \rightarrow \tilde{C}_T^s(\mu_0, \mu)$. We obtain easily that

- If $\mu \in \mathcal{P}_1(M^*)$ has compact support, then for all $f \in \text{Lip}(M)$, then

$$\sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int_M f d\nu + \overline{W}(\mu, \nu) \right\} = \int_{M^*} (-f)^* d\mu.$$

- If $g \in C_{\text{db}}^\infty(M^*)$, then

$$\sup_{\mu \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} g d\mu - \tilde{C}_T^s(\mu_0, \mu) \right\} = \int_{M^*} \tilde{\Psi}_{g,-} d\mu_0.$$

Define $\overline{B}_{\mu_0} : \nu \mapsto \overline{B}(\mu_0, \nu)$, and note that the interpolation formula (120) and a result of Mikima-Thieullen [22] concerning \tilde{C}_T^s yields that \overline{B}_{μ_0} is a concave function. Furthermore it is weak*-upper semi-continuous on $\mathcal{P}_1(M)$. Thus we have

$$\overline{B}_{\mu_0}(\nu_T) = -(-\overline{B}_{\mu_0})^{**}(\nu_T) = \inf_{f \in \text{Lip}(M)} \left\{ - \int_M f d\nu_T + (-\overline{B}_{\mu_0})^*(f) \right\}. \quad (126)$$

Investigating the dual, we find

$$\begin{aligned} (-\overline{B}_{\mu_0})^*(f) &= \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int_M f d\nu + \overline{B}_{\mu_0}(\nu) \right\} \\ &= \sup_{\substack{\mu \in \mathcal{P}_1(M^*) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int_M f d\nu + \overline{W}(\mu, \nu) - \tilde{C}_T^s(\mu_0, \mu) \right\} \\ &= \sup_{\mu \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} (-f)^* d\mu - \tilde{C}_T^s(\mu_0, \mu) \right\}. \end{aligned} \quad (127)$$

Note that in the case where $(-f)^* \in C_{\text{db}}^\infty$, this is simply $\int_{M^*} \tilde{\Psi}_{(-f)^*, -} d\mu_0$, yielding

$$\overline{B}_{\nu_0}(\mu_T) \leq \inf_{(-f)^* \in C_{\text{db}}^\infty} \left\{ - \int_{M^*} f d\mu_T + \int_{M^*} \tilde{\Psi}_{(-f)^*, -} d\mu_0 \right\}.$$

In either case, we can restrict our f to be concave by noting that if we fix $g = (-f)^*$, then the set of corresponding $\{-f; (-f)^* = g\}$ is minimized by the convex function $g^* = (-f)^{**} \leq -f$ [12, Proposition 4.1]. Thus it suffices to consider f convex.

We now show that it is sufficient to consider this infimum over convex $g \in C_{\text{db}}^\infty$ by a similar mollification argument to that used for \underline{B} (note that the mollifying preserves convexity). Maintaining the same assumptions and notation as in our earlier argument, we first note a useful application of Jensen's inequality to the legendre dual of a mollified function:

$$g_\epsilon^*(v) = \sup_x \{\langle v, x \rangle - \mathbb{E}[g(x + H_\epsilon)]\} \stackrel{(J)}{\leq} \sup_x \{\langle v, x \rangle - g(x)\} = g^*(v).$$

Mikami [22, Proof of Theorem 2.1] further shows that

$$(127) = C_{\nu_0}^*(g_\epsilon) \leq \frac{C_{\nu_0^* \eta_\epsilon}^*((1 + \Delta L(0, \epsilon))g)}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)}.$$

Putting these together we get

$$\int g_\epsilon^* d\mu_T + (-\overline{B}_{\nu_0})^*(g_\epsilon^*) d\nu_0 \leq \int g^* d\mu_T + \frac{C_{\nu_0^* \eta_\epsilon}^*((1 + \Delta L(0, \epsilon))g)}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)}.$$

And once we take the infimum over convex $g \in \text{Lip}(M)$, we get

$$\inf \left\{ \int g^* d\mu_T + (-\overline{B}_\nu)^*(-g^*); g \text{ convex in } C_{\text{db}}^\infty \right\} \leq \frac{-(-\overline{B})_{\nu_0^* \eta_\epsilon}^{**}(\mu_{L, \epsilon})}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)},$$

where $d\mu_{L, \epsilon}(v) := d\mu_T((1 + \Delta L(0, \epsilon))v)$. Taking $\epsilon \searrow 0$ dominates the right side by $\overline{B}(\mu_0, \nu_T)$ (where we exploit the upper semi-continuity of \overline{B}), completing the reverse inequality.

Corollary 11 (Optimal Processes for \overline{B}) *Suppose the assumptions on Theorem 9 are satisfied, with μ_0 absolutely continuous with respect to Lebesgue measure. Then, the pair (V, X) is optimal for (118) if and only there is an Ito process $V(t)$ that satisfy the backward Stochastic differential equation,*

$$dV = \nabla_p H(t, \nabla \psi(t, V), V) dt + dW_t \quad (128)$$

$$X = \nabla \overline{\psi}(V(T)), \quad (129)$$

where $\lim_{n \rightarrow \infty} \psi_n(T, x) \rightarrow \overline{\psi}(x)$ ν_T -a.s. and $\lim_{n \rightarrow \infty} \psi_n(t, x) = \psi(t, x)$ \mathbb{P}_V -a.s. for some sequence $\psi_n(t, x)$ that solves (HJB) in such a way that $\psi_n^0 = \psi_n(0, \cdot)$ and $\psi_n^T = \psi_n(T, \cdot)$ are a minimizing pair for the dual problem.

Proof: If (V, X) is optimal, then Theorem 9 means there exists a sequence of solutions $\psi_n(t, v)$ to (HJB) with convex final condition ψ_n^T , such that

$$\mathbb{E} \left[\langle X, V(T) \rangle - \int_0^T \tilde{L}(t, V, \beta_V(t, V)) dt \right] = \lim_{n \rightarrow \infty} \mathbb{E} \left[(\psi_n^T)^*(X) + \psi_n^0(V(0)) \right], \quad (130)$$

which we write as

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[(\psi_n^T)^*(X) + \psi_n^T(V(T)) - \psi_n^T(V(T)) + \psi_n^0(V(0)) \right].$$

Applying Itô's formula to the last two terms, with the knowledge that ψ_n satisfies (HJB) we get

$$\mathbb{E} \left[-\psi_n^T(V(T)) + \psi_n^0(V(0)) \right] = \mathbb{E} \left[\int_0^T -\langle \beta_V, \nabla \psi_n^t(V(t)) \rangle - H(t, \nabla \psi_n^t(V(t)), V(t)) dt \right]$$

However, by the definition of the Hamiltonian, we have $-\langle q, v \rangle - H(t, x, v) \geq -\tilde{L}(t, v, q)$, similarly $\psi^*(v) + \psi(x) \geq \langle v, x \rangle$. These inequalities allow us to separate the limit in (130) into two requirements: (a) $\langle \beta_V, \nabla \psi_n^t(V(t)) \rangle + H(t, \nabla \psi_n^t(V(t)), V(t))$ must converge to $\tilde{L}(t, V, \beta_V(t, V))$ and (b) $\psi_n^T(V(T)) + (\psi_n^T)^*(X)$ must converge to $\langle X, V(T) \rangle$ in L^1 hence a subsequence ψ_{n_k} exists such that this convergence is a.e.

The journey from (a) to (128) is as in Corollary 4. The only difference from the earlier corollary is that we know that ψ_n must converge to a convex function, so (b) implies $X = \nabla \lim_{n \rightarrow \infty} \psi_n(V(T))$.

7 Final Remarks

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

$$\underline{\mathcal{B}}^{\mu_0}(t, \nu) = \inf\{\underline{\mathcal{U}}^{\mu_0}(\varrho) + C_t(\varrho, \nu); \varrho \in \mathcal{P}(M)\} \quad \text{and} \quad \overline{\mathcal{B}}^{\nu_T}(t, \mu) = \inf\{\overline{\mathcal{U}}^{\nu_T}(\varrho) - \tilde{C}_t(\varrho, \mu); \varrho \in \mathcal{P}(M^*)\}, \quad (131)$$

where

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

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

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

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

Corollary 12 *Assume L satisfies hypothesis (A0) and (A1), and let μ_0 be a probability measure on M^* with compact support, then*

$$\underline{\mathcal{B}}^{\mu_0}(T, \nu) := \underline{\mathcal{B}}_T(\mu_0, \nu) = \inf \left\{ \underline{\mathcal{U}}^{\mu_0}(\varrho_0) + \int_0^T \mathcal{L}(\varrho_t, w_t) dt; \partial_t \varrho + \nabla \cdot (\varrho w) = 0, \varrho_T = \nu \right\} \quad (132)$$

The set of pairs (ϱ, w) considered above are such that $t \rightarrow \varrho_t \in \mathcal{P}_1(M)$ (resp., $t \rightarrow w_t(x) \in \text{Lip}(\mathbb{R}^n)$) are paths of Borel fields.

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

$$\begin{cases} \partial_t B + \mathcal{H}(t, \nu, \nabla_\nu B(t, \nu)) = 0, \\ B(0, \nu) = \underline{W}(\mu_0, \nu). \end{cases} \quad (133)$$

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 (133). As importantly, Gangbo-Swiech [19] have shown recently that under certain conditions, value functionals yield solutions to the so-called *Master equations* of mean field games. We refer to their paper for the relevant definitions.

Theorem 10 (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,*

$$\mathcal{U}(t, \nu) = \inf \left\{ \mathcal{U}_0(\varrho_0) + \int_0^t \mathcal{L}(\varrho, w) dt; \partial_t \varrho + \nabla \cdot (\varrho w) = 0, \varrho_T = \nu \right\}.$$

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 (134).

Applied to the value functional $\underline{\mathcal{B}}^{\mu_0}(t, \nu) := \underline{B}_t(\mu_0, \nu)$, this should then yield the existence, for any probabilities μ_0, ν_T , of 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 $\varrho \in AC^2((0, T) \times \mathcal{P}(M))$ such that

$$\begin{cases} \partial_t \beta + \int \langle \nabla_\nu \beta(t, x, \nu) \cdot \nabla H(x, \nabla_x \beta) \rangle d\nu + H(x, \nabla_x \beta(t, x, \nu)) = 0, \\ \partial_t \varrho + \nabla \cdot (\varrho \nabla H(x, \nabla_x \beta)) = 0, \\ \beta(0, \cdot, \cdot) = \beta_0, \quad \varrho(T, \cdot) = \nu_T, \end{cases} \quad (134)$$

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

We may furthermore derive a Eulerian formulation of the minimizing stochastic problem. Recall that in Corollary 4 we showed that the optimal process for the minimizing stochastic cost is Markovian. Hence its drift may be described by a vector field, allowing an Eulerian formulation of the process:

Corollary 13 *Assume L satisfies the assumptions (A0)-(A3), then*

$$\begin{aligned} \underline{\mathcal{B}}^{\mu_0}(T, \nu) &:= \underline{B}_T^s(\mu_0, \nu) \\ &= \inf \left\{ \underline{\mathcal{U}}^{\mu_0}(\varrho_0) + \int_0^T \mathcal{L}(t, \varrho_t, w_t) dt; \partial_t \varrho + \nabla \cdot (w \varrho) + \frac{1}{2} \Delta \varrho = 0, \varrho_T = \nu \right\}. \end{aligned} \quad (135)$$

Proof: It can be seen by Itô's formula that (ϱ, w) is a solution (in the sense of distributions) to $\partial_t \varrho + \nabla \cdot (w \varrho) + \frac{1}{2} \Delta \varrho = 0$ iff $\varrho_t \in \mathcal{P}(M)$ is the law of X_t where X solves

$$X_t = X_0 + \int_0^t w(s, X_s) ds + W_t. \quad (136)$$

Hence the above Eulerian formulation is equivalent to the stochastic process formulation in the case where the optimal drift is described by a Borel vector field. Corollary 4 shows this is the case for \underline{B}_s . \square

Finally, we mention that 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

$$b_g(T, v, x) := \inf \left\{ g(v, \gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t)) dt; \gamma \in C^1([0, T], M) \right\}, \quad (137)$$

where $g : M^* \times M \rightarrow \mathbb{R}$ is a suitable function.

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] A. BARTON, N. GHOUSSOUB: *On Optimal Stochastic Ballistic Transports*, (November 2017) 18 pp.
- [6] M. BEIGLBÖCK AND N. JUILLET: *On a problem of optimal transport under marginal martingale constraints*, Ann. Probab., Volume 44, Number 1, (2016) 42-106
- [7] P. BERNARD & B. BUFFONI: *Optimal mass transportation and Mather theory*. J. Eur. Math. Soc., **9** (2007), no. 1, 85-121.
- [8] Y. BRENIER: *Polar factorization and monotone rearrangement of vector-valued functions*. Comm. Pure Appl. Math., **44** (1991), 375-417.
- [9] S. BOROUSHAKI, N. GHOUSSOUB: *A Self-dual Variational Approach to Stochastic Partial Differential Equations*, Submitted arXiv:1710.01414v1 (October 2017) 32 pp.
- [10] P. CANNARSA & C. SINISTRARI: *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*. Progress in Nonlinear Differential Equations and Their Applications, **58** (2004), Birkhäuser, Boston.
- [11] P. CARDALIAGUET, F. DELARUE, J-M. LASRY & P-L. LIONS: *The master equations and the convergence problem in Mean Field Games*, Preprint.
- [12] I. EKELAND AND R. TÉMAN: *Convex Analysis and Variational Problems*. Classics in Applied Mathematics, **28** (1987), Society for Industrial and Applied Mathematics.
- [13] L.C. EVANS & W. GANGBO: *Differential equations methods for the Monge-Kantorovich mass transfer problem*. Mem. Amer. Math. Soc. **137** (1999).
- [14] A. FATHI: *Regularity of C^1 solutions of the Hamilton-Jacobi equation*. Ann. Fac. Sci. Toulouse Math. (6), **12** (2003), 479-516.
- [15] A. FATHI, A. FIGALLI: *Optimal transportation on non-compact manifolds*, Israel Journal of Mathematics 175 (1), (2010) 1-59
- [16] H. FEDERER: *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, **153** (1969) Springer-Verlag New York Inc., New York.
- [17] W. H. FLEMING AND H. M. SONER: *Controlled Markov Processes and Viscosity Solutions*. Applications of Mathematics, **25** (1993), Springer-Verlag New York Inc., New York.
- [18] W. GANGBO & R.J. MCCANN: *The geometry of optimal transportation*. Acta Math., **177** (1996), 113-161.
- [19] 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.
- [20] N. GHOUSSOUB: *Optimal Ballistic Transport and Hopf-Lax Formulae on Wasserstein Space*, ArXiv e-print 1705.05951 (2017).

- [21] T. MIKAMI: Optimal Control for Absolutely Continuous Stochastic Processes and the Mass Transportation Problem *Electron. Commun. Probab.* 7 (2002), 199–213
- [22] T. MIKAMI AND M. THIEULLEN: Duality theorem for the stochastic optimal control problem. *Stoch. Process. Appl.* 116 (2006), no. 12, 1815–1835
- [23] G. MONGE: *Mémoire sur la Théorie des Déblais et des Remblais*. Hist. de l'Acad. des Sciences de Paris (1781), 666-704.
- [24] E. PARDOUX, A. RASCANU: *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Stochastic Modelling and Applied Probability, Springer (2014)
- [25] C. PRÉVOT, M. ROCKNER: *A Concise Course on Stochastic Partial Differential Equations*, Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg (2007)
- [26] V.N. SUDAKOV: *Geometric problems in the theory of infinite-dimensional probability distributions*. *Proc. Steklov Inst. Math.* 141 (1979), 1-178.
- [27] R. T. ROCKAFELLAR: *Existence and duality theorems for convex problems of Bolza*, *Trans. Amer. Math. Soc.* 159 (1971), 1-40.
- [28] 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.
- [29] R.T. ROCKAFELLAR AND P.R. WOLENSKI: *Convexity in Hamilton-Jacobi theory II: envelope representations*, *SIAM J. Control Opt.* 39 (2001), 1351-1372.
- [30] B. SCHACHTER *An Eulerian Approach to Optimal Transport with Applications to the Otto Calculus*, Thesis, U. of Toronto (2017)
- [31] C. VILLANI: *Topics in mass transportation*. Graduate Studies in Mathematics, 58 (2004), American Mathematical Society, Providence, RI.
- [32] C. VILLANI: *Optimal transport, old and new*. Lecture notes, 2005 Saint-Flour summer school.