*Euro. Jnl of Applied Mathematics* (2019), *vol.* 30, *pp.* 1264–1299  
doi:10.1017/S0956792519000032

# Dynamic and stochastic propagation of the Brenier optimal mass transport†

ALISTAIR BARTON and NASSIF GHOUSSOUB

*Department of Mathematics, University of British Columbia, Vancouver, V6T 1Z2, Canada*
*email: alistair.barton@mail.mcgill.ca; nassif@math.ubc.ca*

(*Received 12 February 2018*; *revised 24 January 2019*; *accepted 1 February 2019*; *first published online 20 March 2019*)

Similar to how Hopf–Lax–Oleinik-type formula yield variational solutions for Hamilton–Jacobi equations on Euclidean space, optimal mass transportations can sometimes provide variational formulations for solutions of certain mean-field games. We investigate here the particular case of transports that maximize and minimize the following 'ballistic' cost functional on phase space $TM$, which propagates Brenier's transport along a Lagrangian $L$,

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

where $M = \mathbb{R}^d$, and $T > 0$. 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. Both cases lead to Lax–Oleinik-type formulas on Wasserstein space that 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. While inf-convolution easily covers 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. We also write Eulerian formulations and point to links with the theory of mean-field games.

**Key words:** Deterministic and stochastic mass transportation, Hamilton-Jacobi equation, stochastic Bolza duality, mean field games.

**2010 Mathematics Subject Classification:** 49J45, 49Q20, 35J60, 49M29, 90C46

## 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 $\mu$ on $X_0$ and $\nu$ on $X_1$, we consider the problem of optimizing the total cost of *transport plans* and its corresponding dual principle as formulated by Kantorovich

† This work is part of a Master's thesis prepared by A. Barton under the supervision of N. Ghoussoub. It was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

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

where $\mathcal{K}(\mu,\nu)$ is the set of *transport plans* between $\mu$ and $\nu$, which is the set of probability measures $\Pi$ on $X_0\times X_1$ whose marginal on $X_0$ (resp. on $X_1$) is $\mu$ (resp., $\nu$). 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 $\varphi_1(x)-\varphi_0(y)\leqslant c(y,x)$ for all $(y,x)\in X_0\times X_1$. The pairs of functions in $\mathcal{K}(c)$ can be assumed to satisfy

$$\varphi_1(x)=\inf_{y\in X_0} c(y,x)+\varphi_0(y)\quad\text{and}\quad \varphi_0(y)=\sup_{x\in X_1}\varphi_1(x)-c(y,x). \tag{1.1}$$

They will be called *admissible Kantorovich potentials*, and for reasons that will become clear later, we shall say that $\varphi_0$ (resp., $\varphi_1$) is an initial (resp., final) Kantorovich potential.

The original Monge problem dealt with the cost $c(y,x)=|x-y|$ [2, 12, 19, 20, 25, 26] and was constrained to those probabilities in $\mathcal{K}(\mu,\nu)$ that are supported by graphs of measurable maps from $X_0$ to $X_1$ pushing $\mu$ onto $\nu$. 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 [14]. 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:

$$C_T(\nu_0,\nu_T):=\inf\left\{\int_{M\times M} c_T(y,x)\,d\Pi;\,\Pi\in\mathcal{K}(\nu_0,\nu_T)\right\}, \tag{1.2}$$

where

$$c_T(y,x):=\inf\left\{\int_0^T L(t,\gamma(t),\dot\gamma(t))\,dt;\,\gamma\in C^1([0,T],M);\,\gamma(0)=y,\gamma(T)=x\right\}, \tag{1.3}$$

$[0,T]$ being a fixed time interval, $\nu_0$, $\nu_T$ two probability measures on $M$, and $L:[0,T]\times TM\to\mathbb{R}\cup\{+\infty\}$ is a given Lagrangian that is convex in the velocity variable of the tangent bundle $TM$. Fathi and Figalli [13] 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

$$b_T(v,x):=\inf\left\{\langle v,\gamma(0)\rangle+\int_0^T L(t,\gamma(t),\dot\gamma(t))\,dt;\,\gamma\in C^1([0,T],M);\,\gamma(T)=x\right\}, \tag{1.4}$$

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

$$\underline{B}_T(\mu_0,\nu_T):=\inf\left\{\int_{M^*\times M} b_T(v,x)\,d\Pi;\ \Pi\in\mathcal{K}(\mu_0,\nu_T)\right\}, \tag{1.5}$$

where $\mu_0$ (resp., $\nu_T$) is a given probability measure on $M^*$ (resp., $M$), and

$$\overline{B}_T(\mu_0,\nu_T):=\sup\left\{\int_{M^*\times M} b_T(v,x)\,d\Pi;\ \Pi\in\mathcal{K}(\mu_0,\nu_T)\right\}. \tag{1.6}$$

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,

$$\underline{W}(\mu_0,\nu_0):=\inf\left\{\int_{M^*\times M}\langle v,x\rangle\,d\Pi;\ \Pi\in\mathcal{K}(\mu_0,\nu_0)\right\},\tag{1.7}$$

and

$$\overline{W}(\mu_0,\nu_0):=\sup\left\{\int_{M^*\times M}\langle v,x\rangle\,d\Pi;\ \Pi\in\mathcal{K}(\mu_0,\nu_0)\right\},\tag{1.8}$$

making (1.6) a suitable dynamic version of the Wasserstein distance.

We shall also consider stochastic versions of the above problems. The cost of transport is then defined between two random variables $Y$ and $Z$ on a fixed complete probability space $(\Omega,\mathcal{F},P)$ as follows:

$$c_T^s(Y,Z)=\inf\left\{\mathbb{E}\left[\int_0^T L(t,X,\beta(t,X))\,dt\right];X\in\mathcal{A},X_0=Y,X_T=Z\text{ a.s}\right\},\tag{1.9}$$

as well as the ballistic cost of using a random input $V$ valued in $M^*$ to get to the random state $Z$ is then

$$b_T^s(V,Z)=\inf\left\{\mathbb{E}\left[\langle V,X_0\rangle+\int_0^T L(t,X,\beta(t,X))\,dt\right];X\in\mathcal{A},X_T=Z\text{ a.s}\right\},\tag{1.10}$$

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 $\sigma(X_s:0\le s\le t)$-Brownian motion. We obtain the corresponding mass transports (from the random variable transports) by letting the random variables vary over all those that have given initial/final probability distributions. Thus, the stochastic version of the dynamic transport (1.2) is

$$C_T^s(\nu_0,\nu_T):=\inf\left\{c_T^s(Y,Z);\,Y\sim\nu_0,Z\sim\nu_T\right\}\tag{1.11}$$

$$=\inf\left\{\mathbb{E}\left[\int_0^T L(t,X,\beta_X(t,X))\,dt\right];X\in\mathcal{A},\ X_0\sim\nu_0,\ X_T\sim\nu_T\right\},\tag{1.12}$$

which was considered by Mikami and Thieullen [18], while the stochastic versions of the ballistic transports are

$$\underline{B}_T^s(\mu_0,\nu_T):=\inf\left\{b_T^s(V,Z);\,V\sim\mu_0,Z\sim\nu_T\right\}\tag{1.13}$$

$$=\inf\left\{\mathbb{E}\left[\langle V,X_0\rangle+\int_0^T L(t,X,\beta(t,X))\,dt\right];X\in\mathcal{A},\,V\sim\mu_0,X_T\sim\nu_T\right\},\tag{1.14}$$

$$\overline{B}_T^s(\mu_0,\nu_T):=\sup\left\{b_T^s(V,Z);\,V\sim\mu_0,Z\sim\nu_T\right\}\tag{1.15}$$

$$=\sup_{V\sim\mu_0,Z\sim\nu_T}\ \inf_{X\in\mathcal{A},X_T=Z}\left\{\mathbb{E}\left[\langle V,X_0\rangle+\int_0^T L(t,X,\beta(t,X))\,dt\right]\right\},\tag{1.16}$$

which we shall consider in the sequel.

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

$$\underline{B}_T(\mu_0, \nu_T) = \inf\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T);\ \nu \in \mathcal{P}(M)\}. \tag{1.17}$$

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

$$\varphi_f(t,x) = \inf\{f(y) + c_t(y,x);\ y \in M\}\,. \tag{1.18}$$

In the case where the Lagrangian $L(t,x,p) = L_0(p)$ is only a convex function of the momentum $p$, then $c_t(y,x) = tL_0(\frac{1}{t}|x-y|)$ and (1.18) is nothing but the Hopf–Lax formula used to generate solutions for corresponding Hamilton–Jacobi equations (see 1.20). If $f$ is the linear functional $f(x) = \langle v, x\rangle$, then $b_t(v,x)$ is itself a solution to the Hamilton–Jacobi equation, since

$$b_t(v,x) = \inf\{\langle v, y\rangle + c_t(y,x);\ y \in M\}. \tag{1.19}$$

In other words, (1.17) can now be seen as extensions of (1.19) to the space of probability measures, where the Wasserstein distance fills the role of the scalar product.

For the duality, we consider the forward Hamilton–Jacobi equations:

$$\begin{cases} \partial_t\varphi + H(t,x,\nabla_x\varphi) = 0 \text{ on } [0,T]\times M, \\ \qquad\qquad \varphi(0,x) = f(x), \end{cases} \tag{1.20}$$

and backward Hamilton–Jacobi equations:

$$\begin{cases} \partial_t\varphi + H(t,x,\nabla_x\varphi) = 0 \text{ on } [0,T]\times M, \\ \qquad\qquad \varphi(T,x) = f(x), \end{cases} \tag{1.21}$$

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 (1.20) and (1.21), which are formally given by the following formulae:

$$\Phi^t_{f,+}(x) := \Phi_{f,+}(t,x) = \inf\left\{f(\gamma(0)) + \int_0^t L(s,\gamma(s),\dot\gamma(s))\,ds;\ \gamma \in C^1([0,T],M);\ \gamma(t) = x\right\}, \tag{1.22}$$

$$\Phi^t_{f,-}(x) := \Phi_{f,-}(t,x) = \sup\left\{f(\gamma(T)) - \int_t^T L(s,\gamma(s),\dot\gamma(s))\,ds;\ \gamma \in C^1([0,T],M);\ \gamma(t) = x\right\}. \tag{1.23}$$

Additional conditions on the Lagrangian are needed in order to guarantee that $\Phi_{f,+}$ and $\Phi_{f,-}$ are at least viscosity solutions (see the thesis [17]). We could have chosen to have a blanket assumption that $L$ is a $C^2$ Tonelli Lagrangian. However, the case where $L$ is jointly convex in position and momentum – a setting we also consider here – requires lesser regularity since convex analysis provides suitable notions of differentiability (see Rockafellar et al. [22, 23]).

We shall then prove the following duality formulae:

$$\underline{B}_T(\mu_0, \nu_T) = \sup\left\{\int_M \Phi_{f_*,+}(T,x)\,d\nu_T(x) + \int_{M^*} f(v)\,d\mu_0(v);\ f \text{ concave in Lip(M}^*)\right\} \tag{1.24}$$

$$= \sup\left\{\int_M g(x)\,d\nu_T(x) + \int_{M^*} (\Phi^0_{g,-})_*(v)\,d\mu_0(v);\ g \text{ in Lip(M)}\right\}, \tag{1.25}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

where $h_*$ is the concave Legendre transform of $h$, that is,

$$h_*(v) = \inf\{\langle v, y\rangle - h(y);\ y \in M\}.$$

We shall also sometimes use the convex Legendre transform $h^*$ defined as follows:

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

As to the question of attainment, we use a result by Fathi–Figalli [13] to show that if $L$ is a Tonelli Lagrangian, and if $\mu_0$ is absolutely continuous with respect to Lebesgue measure, then there exists a probability measure $\Pi_0$ on $M^* \times M$, and a concave function $k : M \to \mathbb{R}$ such that $\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, x)d\Pi_0$, and $\Pi_0$ is supported on the possibly set-valued map $v \to \pi^* \varphi_T^H(\nabla k_*(v), v)$, with $\pi^* : M \times M^* \to M$ being the canonical projection, and $(x, v) \to \varphi_t^H(x, v)$ is the corresponding Hamiltonian flow.

In Section 3, we prove an analogous Hopf–Lax formula on Wasserstein space associated with the stochastic minimization problem:

$$\underline{B}_T^s(\mu_0, \nu_T) = \inf\{\underline{W}(\mu_0, \nu) + C_T^s(\nu, \nu_T);\ \nu \in \mathcal{P}(M)\}. \tag{1.26}$$

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 does not 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:

$$\underline{B}_T^s(\mu_0, \nu_T) = \sup\left\{\int_M g(x)\, d\nu_T(x) + \int_{M^*} (\Psi^0_{g,-})_*(v)\, d\mu_0(v);\ \text{g in Lip(M)}\right\}, \tag{1.27}$$

where this time $\Psi_{g,-}$ is the solution to the backward Hamilton–Jacobi–Bellman equation (1.28).

$$\begin{cases} \partial_t \psi + \frac{1}{2}\Delta\psi + H(x, \nabla_x \psi) = 0 \ \text{ on } [0, T] \times M, \\ \psi(T, x) = g(x), \end{cases} \tag{1.28}$$

whose formal variational solutions are given by the following formula:

$$\Psi_{g,-}(t, x) = \sup_{X \in \mathcal{A}} \left\{\mathbb{E}\left[g(X(T)) - \int_t^T L(s, X(s), \beta_X(s, X))\, ds \middle| X(t) = x\right]\right\}. \tag{1.29}$$

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 purpose, we shall assume that for each time $t \in [0, T]$, the Lagrangian $L(t, \cdot, \cdot)$ is jointly convex in both variables. In Section 4, we then consider the dual Lagrangian $\tilde{L}$ defined on $M^* \times M^*$ by

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

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

$$\tilde{c}_T(u, v) := \inf\left\{\int_0^T \tilde{L}(t, \gamma(t), \dot{\gamma}(t))\, dt; \gamma \in C^1([0, T], M^*); \gamma(0) = u, \gamma(T) = v\right\}, \tag{1.30}$$

and its associated transport

$$\tilde{C}_T(\mu_0, \mu_T) := \inf\left\{\int_{M^* \times M^*} \tilde{c}_T(x, y)\, d\Pi;\ \Pi \in \mathcal{K}(\mu_0, \mu_T)\right\}. \tag{1.31}$$

We then recall the standard deterministic Bolza duality of Rockafellar [22], and the newly established stochastic Bolza duality of Boroushaki–Ghoussoub [9].

We use these results in Section 5 to establish the following results for $\overline{B}_T(\mu_0, \nu_T)$:

$$\overline{B}_T(\mu_0, \nu_T) = \sup\{\overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu);\ \mu \in \mathcal{P}(M^*)\}, \tag{1.32}$$

and

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M g(x)\, d\nu_T(x) + \int_{M^*} \tilde{\Phi}^0_{g^*,-}(v)\, d\mu_0(v);\ g \text{ convex on } M\right\}, \tag{1.33}$$

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:

$$\begin{cases} \partial_t \varphi - H(t, \nabla_v \varphi, v) = 0 \text{ on } [0, T] \times M^*, \\ \varphi(T, v) = k(v), \end{cases} \tag{1.34}$$

whose variational solution is given by

$$\tilde{\Phi}_{k,-}(t, v) = \sup\left\{k(\gamma(T)) - \int_0^t \tilde{L}(s, \gamma(s), \dot{\gamma}(s))\, ds;\ \gamma \in C^1([0, T], M^*);\ \gamma(0) = v\right\}. \tag{1.35}$$

In Section 6, we deal with the stochastic counterpart $\overline{B}^s_T(\mu_0, \nu_T)$ and prove the following:

$$\overline{B}^s_T(\mu_0, \nu_T) := \sup\left\{\mathbb{E}\left[\langle X, V(T)\rangle - \int_0^T \bar{L}(t, V, \beta(t, V))\, dt\right];\ V \in \mathcal{A}, V_0 \sim \mu_0, X \sim \nu_T\right\}, \tag{1.36}$$

where $\bar{L}(x, p) = L^*(-p, -x)$, and therefore

$$\overline{B}^s_T(\mu_0, \nu_T) = \sup\left\{\overline{W}(\nu_T, \mu) - \bar{C}^s_T(\mu_0, \mu);\ \mu \in \mathcal{P}(M^*)\right\}, \tag{1.37}$$

as well as the following duality formula:

$$\overline{B}^s_T(\mu_0, \mu_T) = \inf\left\{\int_{M^*} g(x)\, d\nu_T + \int_M \bar{\Psi}^0_{g^*,-}(v)\, d\mu_0;\ \ g \text{ convex in } C^\infty_{\text{db}}(M^*)\right\}, \tag{1.38}$$

where $\bar{\Psi}^t_{k,-}$ solves the Hamilton–Jacobi–Bellman equation

$$\begin{cases} \partial_t \psi + \frac{1}{2}\Delta\psi - \overline{H}(\nabla_v \psi, v) = 0 \text{ on } [0, T] \times M^*, \\ \psi(T, v) = k(v), \end{cases} \tag{1.39}$$

where $\overline{H}(x, v) := H(-x, -v)$, whose formal variational solutions are given by the formula:

$$\bar{\Psi}^t_{k,-}(v) = \bar{\Psi}_{k,-}(t, v) = \sup_{X \in \mathcal{A}}\left\{\mathbb{E}\left[k(X(T)) - \int_t^T \bar{L}(s, X(s), \beta_X(s, X))\, ds \,\middle|\, X(t) = v\right]\right\}. \tag{1.40}$$

Finally, a few words about our notation: We shall denote by $\partial g$ 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 $\mathrm{Lip}(X)$ of all bounded uniformly Lipschitz functions on $X$. For the stochastic part, we shall also need to work with the space $C^\infty_{\mathrm{db}}(X) := \mathrm{Lip}(X) \cap C^\infty(X)$.

Several of the above results appeared in the posted but non-published manuscripts [16], which dealt with the deterministic case and [5], 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.

We are grateful to the anonymous referee for a careful reading of a first version of this paper, which led to several corrections and improvements.

## 2 Minimizing the ballistic cost: Deterministic case

In this section, we deal with the standard transportation problem associated with 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 continuous, bounded below, and for all $(t, x) \in [0, T] \times M$, the function $v \mapsto L(t, x, v)$ is convex and $\delta$-coercive in the sense that there is a $\delta > 1$ such that

$$\lim_{|v|\to\infty} \frac{L(t, x, v)}{|v|^\delta} = +\infty. \tag{2.1}$$

**Theorem 1** *Assume that $L$ satisfies (A0) and let $\mu_0$ (resp. $\nu_T$) be a probability measure on $M^*$ (resp., $M$) with finite first moment. Then, the following interpolation formula holds:*

$$\underline{B}_T(\mu_0, \nu_T) = \inf\left\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T);\ \nu \in \mathcal{P}_1(M)\right\}. \tag{2.2}$$

*The infimum is attained at some probability measure $\nu_0$ on $M$.*

**Proof** To prove the formula, it suffices to note that

$$\begin{aligned}
&\inf\left\{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T);\ \nu \in \mathcal{P}_1(M)\right\} \\
&= \inf_{\nu\in\mathcal{P}_1(M)} \left\{\int_{M^*\times M} \langle v, x\rangle\, d\Pi_W(v, x) + \int_{M\times M} c_T(x, y)\, d\Pi_C(x, y); \Pi_W \in \mathcal{K}(\mu_0, \nu), \Pi_C \in \mathcal{K}(\nu, \nu_T)\right\} \\
&= \inf_{\Pi\in\mathcal{P}_1(M^*\times M\times M)} \left\{\int_{M^*\times M\times M} \langle v, x\rangle + c_T(x, y)\, d\Pi(v, x, y); \Pi_1 = \mu_0, \Pi_3 = \nu_T\}\right\} \\
&\geq \underline{B}_T(\mu_0, \nu_T).
\end{aligned}$$

For the reverse inequality, use your favourite selection theorem to find a measurable function $y_\epsilon : M^* \times M \to M$ 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 $\Pi_\epsilon := (\mathrm{Id} \times \mathrm{Id} \times y_\epsilon)_\# \Pi \in \mathcal{P}(M^* \times M \times M)$:

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

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 purpose, 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

$$C_T(\nu, \nu_T) \le N \int_M |x| \; d\nu(x) \tag{2.3}$$

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 $B := B(0,R)^c \times M$. Let $\nu \in \mathcal{T}_{\epsilon,R} := \{\nu;\, \nu(B(0,\, R)^c) > \epsilon\}$. We shall assume, without loss of generality, that $L$ and hence $c_T$ is non-negative; hence, for any optimal transport plan $\Pi \in \mathcal{K}(\nu, \nu_T)$

$$C_T(\nu, \nu_T) \ge \frac{N}{T^{\delta}} \int_B ||x| - |y||^{\delta} \; d\Pi(x,y) - K. \tag{2.4}$$

We manipulate this into an optimal transport problem on the half-line by applying the following transformations to $\Pi$, leaving the cost on the right of (2.4) unchanged. We first restrict $\Pi$ to $B$ while adding a dirac measure at $(0,0)$ to compensate for mass lost during the restriction. We then push-forward $\Pi$ onto $\mathbb{R}_+ \times \mathbb{R}_+$ using the absolute value function. This results in the probability measure $\bar{\Pi} := (|\cdot| \times |\cdot|)_{\#}(\Pi|_B) + \nu(B(0,R))\delta_{0,0}$, and transforms

$$\int_B ||x| - |y||^{\delta} \; d\Pi(x,y) \ge \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x-y|^{\delta} \; d\bar{\Pi}(x,y). \tag{2.5}$$

We can obtain a lower estimate for this by minimizing over transportation measures sharing $\bar{\Pi}$'s marginals (i.e. 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 $x \mapsto G_{\bar{\Pi}_1}(G^{-1}_{\bar{\Pi}_2}(x))$, where $G_\nu(t) := \inf\{z \in \mathbb{R} : t \ge \nu(\{x \le z\})\}$ is the quantile function associated with the measure $\nu$ [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 (J):

$$\begin{aligned}\int_{\mathbb{R}_+ \times \mathbb{R}_+} |x-y|^{\delta} \; d\bar{\Pi}(x,y) &\ge \int_{\mathbb{R}_+ \times \mathbb{R}_+} |x-y|^{\delta} \; d\left(\left(G_{\bar{\Pi}_1} \times G_{\bar{\Pi}_2}\right)_{\#} \lambda_{[0,1]}\right)(x,y) \\ &\overset{(J)}{\ge} \left(\int_{B(0,\,R)^c} |x| \; d\nu(x) - m(\nu_T)\right)^{\delta},\end{aligned} \tag{2.6}$$

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

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

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

$$\left(I_\nu(R)^{1-\frac{1}{\delta}} - m(\nu_T) I_\nu(R)^{-\frac{1}{\delta}}\right)^{\delta} - \frac{K}{I_\nu(R)} \overset{?}{>} 1 + \frac{R(1-\epsilon)}{I_\nu(R)}.$$

Observing that the left side is an increasing function of $I_\nu(R)$, while the right side is a decreasing function of $I_\nu(R)$, we may apply the mentioned bound on $I_\nu(R)$ to obtain the following sufficient condition, satisfied for sufficiently large $R$:

$$\left((R\epsilon)^{1-\frac{1}{\delta}} - m(\nu_T)(R\epsilon)^{-\frac{1}{\delta}}\right)^{\delta} - \frac{K}{R\epsilon} \overset{?}{>} \frac{1}{\epsilon}.$$

To show the minimizer is achieved, fix any $\nu_0$ in $\mathcal{P}_1(M)$, and note that by coercivity the set of probability measures $\nu$ such that

$$\underline{W}(\mu_0,\nu)+C_T(\nu,\nu_T)\le m(\mu_0)\int_M |x|\ d\nu + C_T(\nu,\nu_T)\leqslant m(\mu_0)\int_M |x|\ d\nu + C_T(\nu_0,\nu_T)$$

is tight. The fact that $\nu\to\underline{W}(\mu_0,\nu)+C_T(\nu,\nu_T)$ is weak*-lower semi-continuous on $\mathcal{P}_1(M)$ follows from the duality formula that each one of these transports satisfy since both the Brenier cost and $c_T$ are lower semi-continuous (Theorem 1.3 in [25]). □

**Remark 1** *Note that (2.6) 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 the Lagrangian L is* $C^2$ *and satisfies* $(A0)$*, and let* $\mu_0$ *(resp.* $\nu_T$*) be a probability measure on* $M^*$ *(resp., M) with finite first moment.*

1. *If* $\mu_0$ *has compact support, then we have the following duality formula:*

$$\underline{B}_T(\mu_0,\nu_T)=\sup\left\{\int_M g(x)\,d\nu_T(x)+\int_{M^*}(\Phi^0_{g,-})_*(v)\,d\mu_0(v);\ g\ \text{in } Lip(M)\right\}. \tag{2.7}$$

2. *If* $\nu_T$ *has compact support, then we have*

$$\underline{B}_T(\mu_0,\nu_T)=\sup\left\{\int_M \Phi^T_{f_*,+}(x)\,d\nu_T(x)+\int_{M^*} f(v)\,d\mu_0(v);\ f\ \text{concave in } Lip(M^*)\right\}. \tag{2.8}$$

To obtain this result, we shall need the following identifications of Legendre transforms in the Banach space $\mathcal{M}_1(\mathbb{R}^n)$ of measures $\nu$ on $\mathbb{R}^n$ with $\int_{\mathbb{R}^n}(1+|x|)\ d\nu<\infty$ in duality with the space of Lipschitz functions $\mathrm{Lip}(\mathbb{R}^n)$.

**Lemma 1** *Under the hypothesis of Theorem 2, we have the following:*
*(a) For* $\mu_0\in\mathcal{P}(\mathbb{R}^n)$ *with compact support, define* $\underline{W}_{\mu_0}:\mathcal{M}_1(\mathbb{R}^n)\to\mathbb{R}\cup\{\infty\}$ *to be*

$$\underline{W}_{\mu_0}(\nu):=\begin{cases}\underline{W}(\mu_0,\nu) & \nu\in\mathcal{P}_1(\mathbb{R}^n)\\ +\infty & \textit{otherwise.}\end{cases}$$

*Then, the convex Legendre transform of* $\underline{W}_{\mu_0}$ *is given for* $f\in Lip(\mathbb{R}^n)$ *by* $\underline{W}^*_{\mu_0}(f)=-\int_{\mathbb{R}^n}f_*\,d\mu_0$.
*(b) For* $\nu_0\in\mathcal{P}(\mathbb{R}^n)$*, define the function* $C_{\nu_0}:\mathcal{M}_1(\mathbb{R}^n)\to\mathbb{R}\cup\{\infty\}$ *to be*

$$C_{\nu_0}(\nu):=\begin{cases}C_T(\nu_0,\nu) & \nu\in\mathcal{P}_1(\mathbb{R}^n)\\ +\infty & \textit{otherwise.}\end{cases} \tag{2.9}$$

*Then, the convex Legendre transform of* $C_{\nu_0}$ *is given for* $f\in Lip(\mathbb{R}^n)$ *by* $C^*_{\nu_0}(f)=\int_{\mathbb{R}^n}\varphi_{f,-}(0,x)\,d\nu_0(x)$*, where* $\varphi_{f,-}$ *is the solution to the backward HJ-equation (1.21) 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 $\mu_0$ has compact support, Brenier's duality yields

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

We then have

$$\underline{W}^*_{\mu_0}(f) = \sup_{\nu \in \mathcal{M}_1(M)} \inf_{g \in \mathrm{Lip}(M)} \left\{ \int_M f\, d\nu - \int_M g\, d\nu - \int_{M^*} g_*\, d\mu_0 \right\}. \tag{2.10}$$

Note that the functional $g \mapsto -\int_{M^*} g_*\, d\mu = \int_{M^*} (-g)^*\, d\hat{\mu}(v)$ (where $d\hat{\mu}(v) := d\mu(-v)$) is convex and lower semicontinuous, and we may therefore apply a variant of Von Neuman minimax theorem ( [1] Theorem 2.4.1) as the expression is linear in $\nu$ and convex in $g$. We obtain

$$\underline{W}^*_{\mu_0}(f) = \inf_{g \in \mathrm{Lip}(M)} \sup_{\nu \in \mathcal{M}_1(M)} \left\{ \int_M f\, d\nu - \int_M g\, d\nu - \int_{M^*} g_*\, d\mu_0 \right\}. \tag{2.11}$$

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

The same proof applies to $C_{\nu_0}$, since in view of the duality formula of Bernard and Buffoni [7, Proposition 21] or [24] for the case of $\mathbb{R}^n$:

$$C_{\nu_0}(\nu) = \sup_{g \in \mathrm{Lip}(M)} \left\{ \int_M g\, d\nu - \int_M \Phi^0_{g,-}\, d\nu_0 \right\}. \tag{2.12}$$

Note that this holds for all $\nu \in \mathcal{M}_1(M)$, since if $g$ solves the Hamilton–Jacobi equation, then so does $g + c$ for arbitrarily large $c$. We may again apply the minimax theorem as the expression is linear in $\nu$ and convex in $g$. □

**Proof of Theorem 2:** 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

$$\underline{B}_{\mu_0}(\nu) = \underline{B}^{**}_{\mu_0}(\nu) := \sup \left\{ \int_{\mathbb{R}^n} f\, d\nu - \underline{B}^*_{\mu_0}(f);\, f \in \mathrm{Lip}(\mathbb{R}^n) \right\}. \tag{2.13}$$

Now use the Hopf–Lax formula established above to write

$$\begin{aligned} \underline{B}^*_{\mu_0}(f) &:= \sup \left\{ \int_M f\, d\nu - \underline{B}_{\mu_0}(\nu);\, \nu \in \mathcal{P}_1(M) \right\} \\ &= \sup \left\{ \int_M f\, d\nu - \underline{W}(\mu_0, \nu') - C_T(\nu', \nu);\, \nu, \nu' \in \mathcal{P}_1(M) \right\} \\ &= \sup \left\{ \int_M \Phi^T_{f,-}\, d\nu' - \underline{W}(\mu_0, \nu');\, \nu' \in \mathcal{P}_1(M) \right\} \\ &= -\int_M (\Phi^T_{f,-})_*\, d\mu_0. \end{aligned} \tag{2.14}$$

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 expression of the corresponding Kantorovich potentials given in (1.1), and the linearity of $b_T$ in $v$. □

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

We now consider the problem of attainment for $\underline{B}_T(\mu, \nu)$. For that purpose, we shall consider Tonelli Lagrangians studied in the compact case by Bernard–Buffoni [7], and by Fathi–Figalli [13] 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 $v \to L(x, v)$ is strictly convex on $M$.

We also recall the following [4, Definition 5.5.1, page 129]:

**Definition 3** Say that $f : M \to \mathbb{R}$ has an *approximate differential* at $x \in M$ if there exists a function $h : M \to \mathbb{R}$ 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 $H : M \times M^* \to \mathbb{R}$ is then $C^2$, and the Hamiltonian vector field $X_H(x, v) = (\frac{\partial H}{\partial v}(x, v), -\frac{\partial H}{\partial x}(x, v))$ on $M \times M^*$ is then $C^1$, and the associated system of ODEs given by

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial v}(x, v) \\ \dot{v} = -\dfrac{\partial H}{\partial x}(x, v), \end{cases} \tag{2.15}$$

defines a (partial) $C^1$ flow $\varphi_t^H$ (see [13]). The connection between $L$-minimizers $\gamma : [a, b] \to M$ and solutions of (2.15) 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 (2.15).

There is also a (partial) $C^1$ flow $\varphi_t^L$ on $M \times M^*$ such that every unit speed curve of an $L$-minimizer is a part of an orbit of $\varphi_t^L$. This flow is called the Euler–Lagrange flow and is defined by $\varphi_t^L = \mathcal{L}^{-1} \circ \varphi_t^H \circ \mathcal{L}$, where $\mathcal{L} : M \times M \to M \times M^*$ 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* $\mu_0$ *is absolutely continuous with respect to Lebesgue measure. Then there exists a concave function* $k : M \to \mathbb{R}$ *such that*

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \tag{2.16}$$

*where* $S_T(y) = \pi^* \varphi_T^H(y, \nabla k(y))$, $\pi^* : M \times M^* \to M$ *being the canonical projection, and* $\varphi_t^H$ *the Hamiltonian flow associated with L. In other words, an optimal map for* $\underline{B}_T(\mu_0, \nu_T)$ *is given by* $v \to \pi^* \varphi_T^H(\nabla k_*(v), v)$.

**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 $\nu_0$. By the above and Kantorovich duality, there exists a concave function $k : M \to \mathbb{R}$ and another function $h : M \to \mathbb{R}$ 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 [13] 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

$$\underline{B}_T(\mu_0, \nu_T) \leqslant \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \tag{2.17}$$

since $(\nabla k_*)_\# \mu_0 = \nu_0$ and $(S_T)_\# \nu_0 = \nu_T$, and therefore $(\mathrm{Id} \times S_T \circ \nabla k_*)_\# \mu_0$ belongs to $\mathcal{K}(\mu_0, \nu_T)$.

On the other hand, since $b_T(v, x) \leqslant c_T(\nabla k_*(v), x) + \langle \nabla k_*(v), v \rangle$ for every $v \in M^*$, we have

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

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 [18] 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 proper convex function $\underline{L}(v)$ that is 2-coercive in the sense that $\lim_{|v| \to \infty} \frac{\underline{L}(v)}{|v|^2} > 0$.

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

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

(A3) The following are 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) $\sup\{|\nabla_v L(t,x,u)| : |u| \le R\} < \infty$ 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:

$$X(t) = X_0 + \int_0^t \beta_X(s)\,ds + \int_0^t \sigma_X(s)\,dW_s. \tag{3.1}$$

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

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 we have the following:*

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

$$\underline{B}_T^s(\mu_0,\nu_T) = \inf\left\{\underline{W}(\mu_0,\nu) + C_T^s(\nu,\nu_T);\ \nu\in\mathcal{P}_1(M)\right\}. \tag{3.2}$$

*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*

$$\underline{B}_T^s(\mu_0,\nu_T) = \sup\left\{\int_M f(x)\,d\nu_T(x) + \int_{M^*} (\Psi^0_{f,-})_*(v)\,d\mu_0(v);\ f\in C_{db}^\infty\right\}, \tag{3.3}$$

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

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

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

$$\begin{aligned}&\inf\left\{\underline{W}(\mu_0,\nu) + C_T^s(\nu,\nu_T);\ \nu\in\mathcal{P}_1(M)\right\}\\ &\quad= \inf\left\{\mathbb{E}\left[\langle V, X_0\rangle + \int_0^T L(t,X(t),\beta_X(t))\,dt\right]; V\sim\mu_0, X_0\sim\nu, X(t)\in\mathcal{A}_\nu^{\nu_T}; \nu\in\mathcal{P}_1(M)\right\}\\ &\quad\leqslant \underline{B}(\mu_0,\nu_T).\end{aligned}$$

To obtain the reverse inequality, let $\nu_n$ be a sequence of measures approximating the infimum in (3.2). Then for each $\nu_n$, there exists a stochastic process $Z_n \in \mathcal{A}_{\nu_n}^{\nu_T}$ such that

$$\mathbb{E}\left[\int_0^T L(t, Z_n(t), \beta_{Z_n}(Z_n,t))\,dt\right] < C_T^s(\nu_n,\nu_T) + \tfrac{1}{n}. \tag{3.4}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

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

and define $U_n: M\times\Omega\to M^*$ to be a random variable such that $U_n[x]\sim\gamma_x^n$ for $\nu_n$-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 follows:

$$\mathbb{E}\left[\langle U[Z_n(0)],Z_n(0)\rangle+\int_0^T L(t,Z_n(t),\beta_{Z_n}(t))\,dt\right]\le \inf\left\{\underline{W}(\mu_0,\nu)+C_T^s(\nu,\nu_T);\ \nu\in\mathcal{P}_1(M)\right\}+\frac{3}{n}. \tag{3.5}$$

To show that the infimum in $\nu$ 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 $C_T^s(\nu,\nu_T)\le N\int|x|\ d\nu(x)$ is tight.

We will assume $\nu\in\mathcal{T}_{\epsilon,R}:=\{\nu\in\mathcal{P}_1(M):\nu(B(0,\ R)^c)>\epsilon\}$ 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 $\nu$ yields $\mathbb{P}(\Omega_R)>\epsilon$. Note that $\underline{L}$ and hence $L$ is bounded below by (A1), allowing us to assume – without loss of generality – the non-negativity of $L$ and hence of $\underline{L}$. This allows us to focus on the set $\Omega_R$, since

$$\mathbb{E}\left[\int_0^T L(t,X(t),\beta_X(X,t))\,dt\right]\ge\mathbb{E}\left[1_{\Omega_R}\int_0^T L(t,X(t),\beta_X(X,t))\,dt\right].$$

By (A1), there is $C>0$ such that for all $|u|>U$, $\frac{L(u)}{|u|^2}>C$. Recall that $\underline{L}:M^*\to\mathbb{R}$ is a convex lower bound on $L(t,x,v)$. This imposes a lower bound on the expected action of $Y$, which can be refined through Jensen's inequality (J):

$$\begin{aligned}\mathbb{E}\left[\int_0^T 1_{\Omega_R}L(t,X,\beta_X(t,X))\,dt\right]&\ge\mathbb{E}\left[\int_0^T 1_{\Omega_R}\underline{L}(|\beta_X(t,X)|)\,dt\right]\overset{(J)}{\ge}\mathbb{E}\left[\underline{L}(|V|)T\right]\\ &\overset{(A1)}{>}CT\mathbb{E}\left[1_{|V|>U}\,|V|^2\right]\ge CT\left(\mathbb{E}\left[|V|^2\right]-U^2\right),\end{aligned} \tag{3.6}$$

where the $\mathcal{F}_0$-measurable random variable $V:=1_{\Omega_R}(\mathbb{E}\left[X(T)|\mathcal{F}_0\right]-X(0))/T$ is the time-average drift ($\mathcal{F}_t$ being the natural filtration). Omitting the constant term, this may be manipulated into a familiar form by applying reverse triangle and Jensen's inequality:

$$\begin{aligned}CT\mathbb{E}\left[1_{\Omega_R}\left|\frac{\mathbb{E}\left[X(T)|\mathcal{F}_0\right]-X(0)}{T}\right|^2\right]&\ge\frac{C}{T}\mathbb{E}\left[1_{\Omega_R}\,||X(0)|-|\mathbb{E}\left[X(T)|\mathcal{F}_0\right]||^2\right]\\ &\overset{(J)}{\ge}\frac{C}{T}\left|\mathbb{E}\left[1_{\Omega_R}\,|X(0)|\right]-\mathbb{E}\left[1_{\Omega_R}\,|X(T)|\right]\right|^2.\end{aligned} \tag{3.7}$$

Note that we assume $R\epsilon\ge m(\nu_T):=\mathbb{E}\left[|X(T)|\right]$in order to have a lower bound on $\mathbb{E}\left[1_{\Omega_R}\,|X(0)|\right]$ to get the last line. This leaves us with the same formulation as in (2.6) of the deterministic coercivity result, the remainder of the proof is identical, and the claim is proved.

To show that a minimizing sequence $\nu_n$ is sequentially compact in the weak topology, we use the fact that the set of measures $\nu$ 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

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

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.

Now the same reasoning as in Section 2 yields that $C_T(\nu_0,\nu_1)=C_T(\nu_1,\nu_0)=\infty$ 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 (3.2) over $\mathcal{P}_1(M)$.

Also note that the attainment of a minimizing interpolating measure $\nu_0$ is sufficient to show the existence of a minimizing $(V,X)$ for $\underline{B}^s_T(\nu_0,\nu_T)$ whenever the latter is finite. This is a consequence of the existence of minimizers for both $\underline{W}(\mu_0,\nu_0)$ and $C^s_T(\nu_0,\nu_T)$ [18, Proposition 2.1].

(2) To prove the duality formula, we will proceed as in the deterministic case and use the Legendre dual of the optimal cost functional $\nu\to C^s_T(\nu_0,\nu)$, which was derived by Mikami and Thieullen [18]. Indeed, they show that if the Lagrangian satisfies (A1)–(A3), then

$$C^s_T(\nu_0,\nu_T)=\sup\left\{\int_M f\,d\nu_T-\int_M \Psi^0_{f,-}\,d\nu_0; f\in C^\infty_b\right\}, \tag{3.8}$$

where $\Psi_{f,-}$ is the unique solution to the Hamilton–Jacobi–Bellman equation (1.28) that is given by

$$\Psi_{f,-}(t,x)=\sup_{X\in\mathcal{A}}\left\{\mathbb{E}\left[f(X(T))-\int_t^T L(s,X(s),\beta_X(s,X))\,ds\,\middle|\,X(t)=x\right]\right\}. \tag{3.9}$$

Moreover, there exists an optimal process $X$ with drift $\beta_X(t,X)=\mathrm{argmin}_v\{v\cdot\nabla\Psi_{f,-}(t,x)+L(t,x,v)\}$.

Furthermore, $(\mu,\nu)\mapsto C^s_T(\mu,\nu)$ is convex and lower semi-continuous under the weak*-topology. It follows that $\nu\mapsto\underline{B}^s_T(\mu_0,\nu)$ is weak*-lower semi-continuous on $\mathcal{P}_1(M)$ for all $\mu_0\in\mathcal{P}_1(M^*)$, and that $(\mu_0,\nu_T)\mapsto\underline{B}^s_T(\mu_0,\nu_T)$ is jointly convex. Note that integrating $\Psi^0_{f,+}$ over $d\nu_0$ yields the Legendre transform of $\nu\mapsto C_T(\nu_0,\nu)$ for $f\in\mathrm{Lip}(M)$.

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

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

Since $\underline{B}_{\mu_0}$ is convex and weak*-lower semi-continuous, we have

$$\underline{B}_{\mu_0}(\nu)=\underline{B}^{**}_{\mu_0}(\nu)=\sup_{f\in\mathrm{Lip}(M)}\left\{\int f\,d\nu-\underline{B}^*_{\mu_0}(f)\right\}. \tag{3.10}$$

We break this into two steps. First we show that when $f\in C^\infty_{\mathrm{db}}:=\mathrm{Lip}(M)\cap C^\infty$ the dual is appropriate:

$$\begin{aligned}\underline{B}^*_{\mu_0}(f) &:= \sup_{\nu_T\in\mathcal{P}_1(M)}\left\{\int f\,d\nu_T-\underline{B}_T(\mu_0,\nu_T)\right\}\\ &\overset{(3.2)}{=} \sup_{\substack{\nu_T\in\mathcal{P}_1(M)\\ \nu\in\mathcal{P}_1(M)}}\left\{\int f\,d\nu_T-C_T(\nu,\nu_T)-\underline{W}(\mu_0,\nu)\right\}\end{aligned} \tag{3.11}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

$$\stackrel{(3.9)}{=} \sup_{\nu\in\mathcal{P}_1(M)} \left\{ \int \Psi^0_{f,-}(x)\, d\nu(x) - \underline{W}(\mu_0,\nu) \right\}$$
$$= \underline{W}^*_{\mu_0}(\Psi^0_{f,-}) = -\int (\Psi^0_{f,-})_*\, d\mu_0.$$

Thus, plugging this into our dual formula (3.10) and restricting our supremum to $C^\infty_{\text{db}}$ gives

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

To show the reverse inequality, we will adapt the mollification argument used in [18, 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 $\eta_\epsilon(x) = \epsilon^{-d}\eta_1(x/\epsilon)$. We shall use $\epsilon$-subscript to indicate convolution of a measure with $\eta_\epsilon$. Note that 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 (3.11) and fixing a process $X \in \mathcal{A}^{\nu_T}$:

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

where $D_\epsilon(\nu) := \inf\{(1+\Delta L(0,\epsilon))\underline{W}(\mu_\epsilon, \nu_0) + C(\nu_0,\nu); \nu_0\}$, and $H_\epsilon \sim \eta_\epsilon$ is independent of $(X(\cdot), V)$ so that $X(T) + H_\epsilon \sim \eta_\epsilon * \nu_T$. The third line arises by taking $(X(\cdot) + H_\epsilon)$ to be the maximizing process given $V + H_\epsilon \sim \mu_\epsilon$, resulting in the convex dual of $D_\epsilon(\nu)$. Note that $\epsilon \mapsto D_\epsilon(\nu)$ is lower semi-continuous as the infimum of a jointly lower semi-continuous function (the same reason $\nu \mapsto \underline{B}_\mu(\nu)$ is), and converges to $\underline{B}_{\mu_0}(\nu)$ as $\epsilon \to 0$.

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

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

Taking the supremum over $f \in \text{Lip}(M)$, we get the reverse inequality:

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

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

**Corollary 4** *Suppose that the assumptions on Theorem 4(2) are satisfied and that $\mu_0$ 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:*

$$dX = \nabla_p H(t, X, \nabla\psi(t, X))\, dt + dW_t \tag{3.12}$$

$$V = \nabla\bar{\psi}(X(0)), \tag{3.13}$$

*where $\nabla\psi_n(t,x) \to \nabla\psi(t,x)$ $\mathbb{P}_X$-a.s. and $\nabla\psi_n(0,x) \to \nabla\bar{\psi}(x)$ $\nu_0$-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 (3.3). 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 $\psi_n$ to HJB that is maximizing in (3.3) such that

$$\mathbb{E}\left[\int_0^T L(t, X, \beta_X(t, X))\, dt + \langle X(0), V\rangle\right] = \lim_{n\to\infty} \mathbb{E}\left[\psi_n^T(X(T)) + (\psi_n^0)_*(V)\right], \tag{3.14}$$

which we can write as follows:

$$\lim_{n\to\infty} \mathbb{E}\left[\underbrace{\psi_n^T(X(T)) - \psi_n^0(X(0))}_{(a)} + \underbrace{\psi_n^0(X(0)) - (\psi_n^0)_{**}(X(0))}_{(b)} + \underbrace{(\psi_n^0)_{**}(X(0)) + (\psi_n^0)_*(V)}_{(c)}\right], \tag{3.15}$$

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

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

However, by the definition of the Hamiltonian, we have $\langle v, b\rangle - H(t,x,v) \le L(t,x,b)$, which means that (3.15) yield the following three inequalities:

$$\langle \beta_X, \nabla\psi_n^t(X(t))\rangle - H(t, X, \nabla\psi_n^t(X(t))) \le L(t, X, \beta_X(t, X)) \tag{a}$$

$$\psi_n^0(X(0)) - (\psi_n^0)_{**}(X(0)) \le 0 \tag{b}$$

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

In other words, (3.15) 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 $\psi_n^0$ not being necessarily concave. Adding (3.14) to the mix, allows us to obtain $L^1$ convergence in the (a,b,c) inequalities, hence a.s. convergence of a subsequence $\psi_{n_k}$.

Note that the convergence in (b,c) means that $\psi_n^0$ converges $\nu_0$-a.s. to a concave function $\overline{\psi}$ such that $x \mapsto \nabla\overline{\psi}$ 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

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

Since $\psi_n^t$ are deterministic functions, this demonstrates that $X_t$ is a Markov process with drift $\beta_X$ determined by the inverse transform: $\beta_X(t,X) = \nabla_p H(t, X, \nabla\psi(t,X))$, i.e. (3.12) □

**Remark 2** *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 $\psi$, since $\bar{\psi}$ 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 $M^* \times M^*$,

$$\tilde{c}_T(u,v) := \inf\left\{\int_0^T \tilde{L}(\gamma(t),\dot{\gamma}(t))\,dt; \gamma \in C^1([0,T],M^*); \gamma(0)=u, \gamma(T)=v\right\}, \tag{4.1}$$

and its associated optimal transport

$$\tilde{C}_T(\mu_0,\mu_T) := \inf\left\{\int_{M^*\times M^*} \tilde{c}_T(x,y)\,d\Pi;\ \Pi \in \mathcal{K}(\mu_0,\mu_T)\right\}. \tag{4.2}$$

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) $L : M \times M \to \mathbb{R} \cup \{+\infty\}$ 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 $\mathrm{dist}(0, F(x)) \leqslant \varrho(1+|x|)$ for all $x \in M$.
- (B3) For all $(x,p) \in M \times M$, we have $L(x,p) \geqslant \theta(\max\{0, |p| - \alpha|x|\}) - \beta|x|$, where $\alpha, \beta$ are constants, and $\theta$ is a coercive, proper, and 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

$$\psi(x) - (\gamma|x| + \delta)|q| \leqslant H(x,q) \leqslant \varphi(q) + (\alpha|q| + \beta)|x| \ \text{ for all } x, q \text{ in } M \times M^*, \tag{4.3}$$

where $\alpha, \beta, \gamma, \delta$ are constants, $\varphi$ is finite and convex, and $\psi$ is finite and concave (see [23]).

We note that under these conditions, the cost $(x,y) \to c_T(x,y)$ 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 [21] and co-authors [22, 23]. The stochastic counterpart is more recent and has been established by Boroushaki and Ghoussoub [9].

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

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

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

$$(\mathcal{P}) \qquad \inf \left\{ \int_0^T L(\gamma(s), \dot{\gamma}(s))\, ds + \ell(\gamma(0), \gamma(T));\ \gamma \in C^1([0,T], M) \right\}, \tag{4.4}$$

and

$$(\tilde{\mathcal{P}}) \qquad \inf \left\{ \int_0^T \tilde{L}(\gamma(s), \dot{\gamma}(s))\, ds + \ell^*(\gamma(0), -\gamma(T));\ \gamma \in C^1([0,T], M) \right\}. \tag{4.5}$$

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

1. *If there exists $\xi$ such that $\ell(\cdot, \xi)$ is finite, or there exists $\xi'$ 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 $v(t) \in \mathcal{A}^2_M$ for $(\tilde{\mathcal{P}})$.*
2. *A similar statement holds if we replace $\ell$ 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 $\mathcal{A}^2_M$.*

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

$$\dot{x}(t) \in \partial_v H(x(t), v(t)), \tag{4.6}$$

$$-\dot{v}(t) \in \tilde{\partial}_x H(x(t), v(t)), \tag{4.7}$$

coupled with the boundary conditions

$$(v(0), -v(T)) \in \partial \ell(x(0), x(T)). \tag{4.8}$$

See for example [21]. The above duality has several consequences.

**Proposition 5** *The value function $\Phi_{g,+}(t,x) = \inf\{g(y) + c_t(y,x);\ y \in M\}$, which is the variational solution of the Hamilton–Jacobi equation (1.20) 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 3** *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., convex) 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* $\Phi_g$ *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* $x\to\Phi_{g,+}(t,x)=b_t(v,x)$ *is* $w\to\tilde{c}_t(v,w)$*. 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* $t\geqslant 0$*,* $(x,y)\to c_t(x,y)$ *is convex proper and lower semi-continuous on* $M\times M$*.*
2. *For each* $t\geqslant 0$*,* $v\to b_t(v,x)$ *is concave on* $M^*$*, while* $x\to b_t(v,x)$ *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 $w$.

### 4.1 The stochastic Bolza duality and its applications

We now deal with the stochastic case. By denoting $W_s$ one-dimensional Brownian motion, we define the *Itô space* $\mathcal{I}_M^p$ consisting of all $M$-valued processes of the following form:

$$\mathcal{I}_M^p = \Big\{ X : \Omega_T \to M;\ X(t) = X_0 + \int_0^t \beta_X(s)ds + \int_0^t \sigma_X(s)dW_s,$$
$$\text{for } X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; M),\ \beta_X \in L^p(\Omega_T; M),\ \sigma_X \in L^2(\Omega_T; M) \Big\}, \tag{4.9}$$

where $\beta_X$ and $\sigma_X$ 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

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

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 follows:

$$\langle X, q\rangle_{\mathcal{I}_M^2 \times (\mathcal{I}_M^2)^*} = \mathbb{E}\left\{ \langle q_0, X(0)\rangle_M + \int_0^T \langle q_1(t), \beta_X(t)\rangle_M\, dt + \frac{1}{2}\int_0^T \langle Q(t), \sigma_X(t)\rangle_M\, dt \right\}. \tag{4.10}$$

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

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 $N$ be two (possibly $\Omega_T$ dependent) jointly convex Lagrangians on $M \times M$. Assume $\ell$ is a (possibly $\Omega$ dependent) convex lsc function on $M \times M$. Consider the Lagrangian on $\mathcal{I}_M^2 \times (\mathcal{I}_M^2)^*$ defined by*

$$\mathcal{L}(X, p) = \mathbb{E}\left\{ \int_0^T L(X(t) - p_1(t), -\beta_X(t))\, dt + \ell(X(0) - p_0, X(T)) \right.$$
$$\left. + \frac{1}{2}\int_0^T N(\sigma_X(t) - P(t), -\sigma_X(t))\, dt \right\}, \tag{4.11}$$

*where $p = (p_0, p_1(t), P(t)) \in L^2(\Omega; M) \times L^2(\Omega_T; M) \times L^2(\Omega_T; M)$. Its Legendre dual is then given for each $q := (0, q_1, Q)$ by*

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

Note that standard duality theory implies that in general

$$\inf_{X \in \mathcal{I}_M^p} \{\mathcal{L}(X, 0)\} \geq \sup_{Y \in \mathcal{I}_M^p} \{-\mathcal{L}^*(0, Y)\}. \tag{4.12}$$

In our case, we shall restrict ourselves to processes of fixed diffusion. This facilitates the proving of a stochastic analogue to Bolza duality for $p = 1$.

**Proposition 8** *Assume $(\ell, L, N)$ satisfy the assumptions of Theorem 6 and the lagrangian $\tilde{L} : (t, x, v) \mapsto L^*(t, v, x)$ satisfies (A2). If, furthermore, there exist (a.s.-)unique $Y_0 \in L^2(\Omega)$ such that $\ell^*(Y_0, \cdot) < \infty$ and (a.s.-)unique $\sigma_Y \in L^2(\Omega_T)$ such that $N^*(\cdot, \sigma_Y) < \infty$, then there is no duality gap, that is,*

$$\inf_{X \in \mathcal{I}_M^1} \{\mathcal{L}(X, 0)\} = \sup_{Y \in \mathcal{I}_M^1} \{-\mathcal{L}^*(0, Y)\}. \tag{4.13}$$

**Remark 4** *Note that, unlike the deterministic case, there is no backwards condition that works if there is an $Y_T \in L^2(\Omega)$ such that $\ell^*(\cdot, Y_T) < \infty$, this is because stochastic processes, in general, are irreversible.*

**Proof** We proceed by a variational method outlined by Rockafellar [21]. First, we define

$$\varphi(q) := \inf_{Y \in \mathcal{I}_M^1} \{\mathcal{L}^*(q, Y)\}. \tag{4.14}$$

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

$$\varphi^*(X) = \sup\{\langle X, q\rangle - \mathcal{L}^*(q, v); (q, v) \in \mathcal{I}_M^\infty \times \mathcal{I}_M^1\} = \mathcal{L}^{**}(X, 0) = \mathcal{L}(X, 0). \tag{4.15}$$

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

$$X \in \partial\varphi(0) \quad \Longleftrightarrow \quad \varphi(0) + \varphi^*(X) = 0 \quad \Longleftrightarrow \quad \mathcal{L}(X, 0) = -\inf_{Y \in \mathcal{I}_M^1} \{\mathcal{L}^*(0, Y)\}. \tag{4.16}$$

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 $\mathcal{N}$ of the origin in $\mathcal{I}_M^\infty = (\mathcal{I}_M^1)^*$ such that $\mathcal{L}^*(q, Y) < \infty$ for $q \in \mathcal{N}$.

By our assumptions, we may fix $Y_0, \sigma_Y$ to be the unique elements such that $\ell^*(Y_0, \cdot) < \infty$ and $N^*(\cdot, \sigma_Y) < \infty$ (guaranteeing subdifferentiability in these variables), and let $Y = (Y_0, \beta_Y, \sigma_Y)$ be such that $\mathcal{L}^*(0, Y) < \infty$. Thus, for a perturbation $V \in \mathcal{I}_M^\infty$, $V_0 = \sigma_V = 0$ is constrained and we only concern ourselves with $\beta_V \in L^\infty(\Omega_T)$. When our perturbation satisfies $\|\beta_V\|_\infty < \epsilon$, note that (A2) gives for all $(t, u) \in [0, T] \times M^*$,

$$\tilde{L}(Y_t - \beta_V, u) < \left(1 + \Delta\tilde{L}(0, \epsilon)\right) \tilde{L}(Y_t, u) + \Delta\tilde{L}(0, \epsilon), \tag{4.17}$$

and if we let $V = (0, \beta_V, 0)$

$$\begin{aligned} \varphi(V) = &\inf_{Y \in \mathcal{I}_M^1} \mathcal{L}^*(V, Y) \\ \leq &\mathbb{E}\ell^*(-Y_0, Y_T) + \mathbb{E}\int_0^T \tilde{L}(Y_t - \beta_V(t), -\beta_Y(t))\, dt + \frac{1}{2}\mathbb{E}\int_0^T N^*(-\sigma_Y(t), \sigma_Y(t))\, dt \\ \leq &\mathbb{E}\ell^*(-Y_0, Y_T) + (1 + \Delta\tilde{L}(0, \epsilon))\mathbb{E}\int_0^T \tilde{L}(Y_t, -\beta_Y(t))\, dt + T\Delta\tilde{L}(0, \epsilon) \\ &+ \frac{1}{2}\mathbb{E}\int_0^T N^*(-\sigma_Y(t), \sigma_Y(t))\, dt, \end{aligned} \tag{4.18}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

Note that if $\tilde{L}$ satisfies (A1) then this optimum is achieved over $Y \in \mathcal{I}^2_M$. This is because $\mathcal{L}^*(0, Y) = \infty$ for $\beta_Y \in L^1(\Omega_T) \setminus L^2(\Omega_T)$. Indeed, in this case we have $\mathbb{E} \int_0^T \tilde{L}(Y_t, -\beta_Y) \geq \mathbb{E} \int_0^T \underline{L}(\beta_Y)\, dt \geq C\mathbb{E} \int_0^T |\beta_Y|^2 - B\, dt = \infty$ (where $C, B > 0$ 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 $\nu_T$ be a probability measure with compact support on $M$, which is also absolutely continuous with respect to Lebesgue measure. Then,*

1. *The following interpolation formula holds:*

$$\overline{B}_T(\mu_0, \nu_T) = \sup\{\overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu);\ \mu \in \mathcal{P}(M^*)\}. \tag{5.1}$$

*The supremum is attained at some probability measure $\mu_T$ 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:*

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M h(x)\, d\nu_T(x) + \int_{M^*} \tilde{\Phi}^0_{h^*,-}(v)\, d\mu_0(v);\ h \text{ convex in } Lip(M)\right\}. \tag{5.2}$$

*and*

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M (\tilde{\Phi}^T_{g,+})^*(x)\, d\nu_T(x) + \int_{M^*} g(v)\, d\mu_0(v);\ g \text{ in } Lip(M^*)\right\}. \tag{5.3}$$

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

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v), \tag{5.4}$$

*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 with $\tilde{L}$ (i.e. $\tilde{H}(v, x) = -H(x, v)$, and $h_0 = \tilde{\Phi}^0_{h,-}$ is the solution $h(0, v)$ of the backward Hamilton–Jacobi equation (1.34) with $h(T, v) = h(v)$. In other words, an optimal map for $\overline{B}_T(\mu_0, \nu_T)$ is given by the map $x \to \nabla h(\pi^* \varphi_T^{\tilde{H}}(v, \tilde{d}_v h_0))$.*

4. *We also have*

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(S_T^* \circ \nabla h^*(x), x) d\nu_T(x), \tag{5.5}$$

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

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

**Proof** To show (5.1) and (5.2), first note that for any probability measure $\mu$ on $M^*$, we have

$$\overline{B}_T(\mu_0,\nu_T)\geqslant \overline{W}(\nu_T,\mu)-\tilde{C}_T(\mu_0,\mu). \tag{5.6}$$

Indeed, since $\nu_T$ is assumed to be absolutely continuous with respect to Lebesgue measure, Brenier's theorem yields a convex function $h$ that is differentiable $\nu_T$-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 $\Pi_0$ 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 $b_T(v,y)\geqslant\langle\nabla h(x),y\rangle-\tilde{c}_T(v,\nabla h(x))$ for every $(y,x,v)\in M\times M\times M^*$, we have

$$\begin{aligned}\overline{B}_T(\mu_0,\nu_T)&\geqslant\int_{M^*\times M}b_T(v,x)\,d\tilde{\Pi}_0(v,x)\\&\geqslant\int_{M^*\times M}\{\langle\nabla h(x),x\rangle-\tilde{c}_T(v,\nabla h(x))\}d\tilde{\Pi}_0(v,x)\\&=\int_M\langle x,\nabla h(x)\rangle\,d\nu_T(x)-\int_{M^*\times M^*}\tilde{c}_T(v,w)\}d\Pi_0(v,w)\\&=\overline{W}(\nu_T,\mu)-\tilde{C}_T(\mu_0,\mu).\end{aligned}$$

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

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

where the infimum is taken over all admissible Kantorovich pairs $(g,h)$ of functions, that is, 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 $\Pi_0$ 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 $y\to h(y)-g(v)-b_T(v,y)$ 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

$$b_T(v,x)+\tilde{c}_T(v,\nabla h(x))=\langle x,\nabla h(x)\rangle\ \text{ on }\mathcal{O}. \tag{5.7}$$

Integrating with $\Pi_0$, we get since $\Pi_0\in\mathcal{K}(\mu_0,\nu_T)$,

$$\int_{M^*\times M}b_T(v,x)\,d\Pi_0+\int_{M^*\times M}\tilde{c}_T(v,\nabla h(x))d\Pi_0=\int_M\langle x,\nabla h(x)\rangle\,d\nu_T. \tag{5.8}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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

$$\overline{B}_T(\mu_0, \nu_T) + \int_{M^*\times M} \tilde{c}_T(v, \nabla h(x))d\Pi_0 = \overline{W}(\nu_T, \mu_T), \tag{5.9}$$

where $\overline{W}(\nu_T, \mu_T) = \sup\left\{\int_{M\times M^*}\langle x, v\rangle\, d\Pi;\ \Pi \in \mathcal{K}(\nu_T, \mu_T)\right\}$. Note that we have used here that $h$ is convex to deduce that $\overline{W}(\nu_T, \mu_T) = \int_M \langle x, \nabla h(x)\, d\mu_T$ by the uniqueness in Brenier's decomposition. We now prove that

$$\int_{M^*\times M} \tilde{c}_T(v, \nabla h(x))d\Pi_0 = \tilde{C}_T(\mu_0, \mu_T). \tag{5.10}$$

Indeed, we have $\int_{M^*\times M} \tilde{c}_T(v, \nabla h(x))d\Pi_0 \geqslant \tilde{C}_T(\mu_0, \mu_T)$ since the measure $\Pi = S_{\#}\Pi_0$, where $S(v, x) = (v, \nabla h(x))$ has marginals $\mu_0$ and $\mu_T$, respectively. On the other hand, (5.9) yields

$$\begin{aligned}
\int_{M^*\times M} \tilde{c}_T(v, \nabla h(x))d\Pi_0 &= \int_M \langle x, \nabla h(x)\rangle\, d\nu_T(x) - \int_{M^*\times M} b_T(v, x)\, d\Pi_0 \\
&= \int_M h^*(\nabla h(x))d\nu_T(x) + \int_M h(x)\, d\nu_T(x) \\
&\quad + \int_{M^*} g(v)\, d\mu_0(v) - \int_M h(x)\, d\nu_T(x) \\
&= \int_{M^*} h^*(w)d\mu_T(w) + \int_{M^*} g(v)\, d\mu_0(v).
\end{aligned}$$

Moreover, since $h(x) - g(v) \geqslant b(v, x)$, we have $h^*(w) + g(v) \leqslant \tilde{c}_T(v, w)$. 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 $y \in M$,

$$\tilde{c}_T(v, w) \geqslant \langle w, y\rangle - b_T(v, y) \geqslant \langle w, y\rangle + g(v) - h(y),$$

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

$$\begin{aligned}
\tilde{C}_T(\mu_0, \mu_T) &\leqslant \int_{M^*\times M} \tilde{c}_T(v, \nabla h(x))d\Pi_0 \\
&= \int_M h^*(w)d\mu_T(w) + \int_{M^*} g(v)\, d\mu_0(v) \\
&\leqslant \sup\left\{\int_{M^*} \varphi_T(w)\, d\mu_T(w) - \int_{M^*} \varphi_0(v)\, d\mu_0(v);\ \varphi_T(w) - \varphi_0(v) \leqslant \tilde{c}_T(v, w)\right\} \\
&= \tilde{C}_T(\mu_0, \mu_T).
\end{aligned}$$

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 (5.6) is attained by the measure $\mu_T$. Note that the final optimal Kantorovich potential for $\tilde{C}_T(\mu_0, \mu_T)$ is $h^*$, and hence is convex. □

The first duality formula (5.3) 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 (5.2).

To show (5.3), we can – now that the interpolation (5.1) is established – proceed as in Section 2, by identifying the Legendre transform of the functionals $\nu \to \overline{W}(\nu, \nu_T)$ and $\mu \to \tilde{C}_T(\mu, \mu_T)$.

To prove Part (3), use first the interpolation inequality to write

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

for some probability measure $\mu_T$. The proof also shows that there exists a convex function $h: M^* \to \mathbb{R}$ and another function $k: M^* \to \mathbb{R}$ such that $(\nabla h)_\# \mu_T = \nu_T$, $\overline{W}(\nu_T, \mu_T) = \int_M \langle \nabla h(v), v\rangle d\mu_T(v)$, 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

$$\tilde{C}_T(\mu_0, \mu_T) = \int_{M^*} c_T(v, \tilde{S}_T v) d\mu_0(v), \tag{5.11}$$

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

$$\overline{B}_T(\mu_0, \nu_T) \geqslant \textstyle\int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v), \tag{5.12}$$

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

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

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

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

To get Part (4), use the pushforward $\nu_T = (\nabla h \circ \tilde{S}_T)_\# \mu_0$ to write the above in terms of the measure $\nu_T$, 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 with the Hamiltonian $H_*(v, x) := -H(-x, v)$. This gives us

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

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

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

**Remark 5** *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 (2.2) of $\underline{B}_T(\mu_0, \nu_T)$ selects a $\nu_0$ such that $C_T(\nu_0, \nu_T)$ has a concave initial Kantorovich potential, while the interpolation (5.1) of $\overline{B}_T(\mu_0, \nu_T)$ selects a $\mu_T$ such that $\tilde{C}_T(\mu_0, \mu_T)$ has a convex final Kantorovich potential.*

*Furthermore, one wonders whether the formula*

$$c_t(y,x) = \sup\{b_t(v,x) - \langle v,y\rangle;\, v \in M^*\}, \tag{5.13}$$

*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 $\nu_0$ and $\nu_T$ are probability measures on $M$ such that $\nu_0$ 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:*

$$C_T(\nu_0,\nu_T) = \sup\{\underline{B}_T(\mu,\nu_T) - \underline{W}(\nu_0,\mu);\ \mu \in \mathcal{P}(M^*)\}, \tag{5.14}$$

*and the supremum is attained.*

**Proof** Again, it is easy to show that

$$C_T(\nu_0,\nu_T) \geqslant \sup\{\underline{B}_T(\mu,\nu_T) - \underline{W}(\nu_0,\mu);\ \mu \in \mathcal{P}(M^*)\}. \tag{5.15}$$

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

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

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 $\Pi_0$ 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 $z \to h(x) - g(z) - c_T(z,x)$ attains its maxmum at $y$, to deduce that $-\nabla g(y) \in \partial_y c_T(y,x)$.

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

$$c_T(y,x) = b_T(\nabla g(y),x) - \langle \nabla g(y), y\rangle. \tag{5.16}$$

Integrating with $\Pi_0$, we get since $\Pi_0 \in \mathcal{K}(\nu_0,\nu_T)$,

$$\int_{M\times M} c_T(y,x)\,d\Pi_0 = \int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0 - \int_M \langle \nabla g(y), y\rangle\,d\nu_0. \tag{5.17}$$

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

$$C_T(\nu_0,\nu_T) = \int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0 - \underline{W}(\nu_0,\mu_0). \tag{5.18}$$

We now prove that

$$\int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0(y,x) = \underline{B}_T(\mu_0,\nu_T). \tag{5.19}$$

Indeed, we have $\int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0 \geqslant \underline{B}_T(\mu_0,\nu_T)$, since the measure $\Pi = S_\#\Pi_0$ where $S(y,x) = (\nabla g(y),x)$ has $\mu_0$ and $\nu_T$ as marginals. On the other hand, (5.18) yields

$$\begin{aligned}\int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0 &= \int_{M\times M} c_T(y,x)\,d\Pi_0 + \int_M \langle y,\nabla g(y)\rangle\,d\nu_0(y)\\ &= \int_M h(x)\,d\nu_T(x) - \int_M g(y)\,d\nu_0(y)\\ &\quad - \int_M (-g)^*(-\nabla g(y))d\nu_0(y) + \int_M g(y)\,d\nu_0(y)\\ &= \int_M h(x)\,d\nu_T(x) - \int_{M^*} (-g)^*(-v)d\mu_0(v).\end{aligned}$$

Moreover, since $h(x) - g(y) \leqslant c_T(y,x)$, it is easy to see that $h(x) - (-g)^*(-v) \leqslant b_T(v,x)$, that is the couple $((-g)^*(-v), h(x))$ is an admissible Kantorovich pair for the cost $b_T$. It follows that

$$\begin{aligned}\underline{B}_T(\mu_0,\nu_T) &\leqslant \int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0\\ &= \int_M h(x)\,d\nu_T(x) - \int_M (-g)^*(-v)d\mu_0(v)\\ &\leqslant \sup\left\{\int_M \varphi_T(x)\,d\mu_T(x) - \int_{M^*} \varphi_0(v)\,d\mu_0(v);\ \varphi_T(x) - \varphi_0(v) \leqslant b_T(v,x)\right\}\\ &= \underline{B}_T(\mu_0,\nu_T),\end{aligned}$$

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 (5.14) is attained by the measure $\mu_0$.

**Corollary 9** *Assume $M = \mathbb{R}^d$ and that $L$ satisfies hypothesis (B1), (B2), and (B3). Assume $\nu_0$ and $\nu_T$ are probability measures on $M$ such that $\nu_0$ 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 $X_0^T : M^* \to M$ and a concave function $g$ on $M$ such that*

$$C_T(\nu_0,\nu_T) = \int_M c_T(y, X_0^T\circ\nabla g(y))d\nu_0(y). \tag{5.20}$$

**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 $\mu_0$ 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 $X_0^T : M^* \to M$ such that $(X_0^T)_\#\mu_0 = \nu_T$ and

$$\underline{B}_T(\mu_0,\nu_T) = \int_{M^*} b_T(v, X_0^T v)d\mu_0(v). \tag{5.21}$$

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 $\Pi_0$ 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

$$\begin{aligned}\int_{M\times M} b_T(\nabla g(y),x)\,d\Pi_0 = \underline{B}_T(\mu_0,\nu_T) &= \int_{M^*} b_T(v, X_0^T v)d\mu_0(v)\\ &= \int_M b_T(\nabla g(y), X_0^T\circ\nabla g(y))d\nu_0(y).\end{aligned}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

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 $\nu_0,\nu_1$ be probability measures on $M$ such that the initial Kantorovich potential associated with $C_T(\nu_0,\nu_T)$ is concave. Then, there exist concave functions $\varphi : M\to\mathbb{R}$, $\psi : M^*\to\mathbb{R}$ and a probability measure $\mu_0$ on $M^*$ such that*

$$(\nabla\psi\circ\nabla\varphi)_\#\nu_0 = \nu_1, \tag{5.22}$$

*and*

$$\begin{aligned} C_1(\nu_0,\nu_1) + \int_{M^*} c^*(v)\, d\mu_0(v) &= \int_M c(\nabla\psi\circ\nabla\varphi(y) - y)d\nu_0(y) \\ &= \int_M \langle\nabla\psi_*(y) - \nabla\varphi(y), y\rangle\, d\nu_0(y). \end{aligned} \tag{5.23}$$

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

$$c_1(y,x) = \inf\left\{\int_0^1 c(\dot\gamma(t))\,dt;\, \gamma\in C^1([0,1),M);\, \gamma(0) = y, \gamma(1) = x\right\} = c(x-y). \tag{5.24}$$

It follows from (5.14) that there is a probability measure $\mu_0$ 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, $b_1(v,x) = \inf\{\langle v,y\rangle + c(x-y);\ y\in M\} = \langle v,x\rangle - c^*(v)$, and hence we have

$$C_1(\nu_0,\nu_1) = \underline{B}_1(\mu_0,\nu_1) - \underline{W}(\nu_0,\mu_0) = \underline{W}(\mu_0,\nu_1) - \int_{M^*} c^*(v)\, d\mu_0(v) - \underline{W}(\nu_0,\mu_0). \tag{5.25}$$

In other words,

$$C_1(\nu_0,\nu_1) + K = \underline{W}_1(\mu_0,\nu_1) - \underline{W}(\nu_0,\mu_0), \tag{5.26}$$

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

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

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

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

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

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

## 6 Maximizing the ballistic cost: Stochastic case

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

$$b_T^s(V,Y) := \inf\left\{\mathbb{E}\left[\langle V, X(0)\rangle + \int_0^T L(X_t, \beta_X(t))\,dt\right]; X \in \mathcal{A}, X(T) = Y \text{ a.s.}\right\}, \tag{6.1}$$

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

$$\underline{B}_T^s(\mu_0, \nu_T) = \inf\left\{b_T^s(V,Y); V \sim \mu_0, Y \sim \nu_T\right\}, \tag{6.2}$$

while the maximizing cost is defined as follows:

$$\overline{B}_T^s(\mu_0, \nu_T) := \sup\left\{b_T^s(V,Y); V \sim \mu_0, Y \sim \nu_T\right\}. \tag{6.3}$$

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

1. *The following formula then holds:*

$$\overline{B}_T^s(\mu_0, \nu_T) := \sup\left\{\mathbb{E}\left[\langle X, V(T)\rangle - \int_0^T \bar{L}(V, \beta_V(t,V))\,dt\right]; V \in \mathcal{A}, V_0 \sim \mu_0, X \sim \nu_T\right\}. \tag{6.4}$$

   *where $\bar{L}(x,p) := L^*(-p,-x)$.*

2. *The following duality holds:*

$$\overline{B}_T^s(\mu_0, \nu_T) = \sup\left\{\overline{W}(\mu, \nu_T) - \bar{C}_T^s(\mu_0, \mu);\ \mu \in \mathcal{P}_1(M^*)\right\}, \tag{6.5}$$

   *where $\bar{C}_T^s$ is the action corresponding to the Lagrangian $\bar{L}$.*

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

$$\overline{B}_s(\mu_0, \nu_T) = \inf\left\{\int_M g\,d\nu_T + \int_{M^*} \bar{\Psi}^0_{g^*,-}\,d\mu_0;\ g \in C^\infty_{db}(M) \text{ and convex}\right\}, \tag{6.6}$$

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

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

   *where $\overline{H}(x,v) := H(-x,-v)$ is the Hamiltonian associated with $\overline{L}$.*

**Proof** (1) For a fixed pair $(V,Y)$, we consider the Bolza energy $\mathcal{L}_{(V,Y)}$ –defined in (4.11)– associated with $L$ and the two Lagrangians $\ell$ and $M$ defined as follows:

$$\ell_{(V,Y)}(\omega, x_0, x_T) := \begin{cases} \langle x_0, V(\omega)\rangle & x_T = Y(\omega) \\ \infty & \text{else} \end{cases} \qquad N(\sigma,\tau) := \begin{cases} \tau + 1 & \sigma = 1 \\ \infty & \text{else.} \end{cases} \tag{6.7}$$

Note that the minimizing stochastic cost can be written as follows:

$$\underline{B}_s(\mu_0, \nu_T) := \inf\left\{\inf\left\{\mathcal{L}_{(V,Y)}(X(t), 0); X \in \mathcal{I}^1_M\right\}; V \sim \mu_0, Y \sim \nu_T\right\} \tag{6.8}$$

while the maximizing cost is

$$\overline{B}_s(\mu_0, \nu_T) = \sup\left\{\inf\left\{\mathcal{L}_{(V,Y)}(X(t), 0); X \in \mathcal{I}^1_M\right\}; V \sim \mu_0, Y \sim \nu_T\right\}. \tag{6.9}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

Applying Bolza duality turns the infimum to a supremum:

$$\overline{B}_s(\mu_0, \nu_T) = \sup \Big\{ \sup \big\{ -\mathcal{L}^*_{(V,Y)}(0, X(t)); X \in \mathcal{I}^1_M \big\}; V \sim \mu_0, Y \sim \nu_T \Big\}, \tag{6.10}$$

which results in (6.4).

(2) With equation (6.4) in hand, the proof of the interpolation result can now follow closely the proof for the minimization problem. As in that proof, showing

$$\overline{B}(\mu_0, \nu_T) \le \sup \Big\{ \overline{W}(\mu, \nu_T) - \bar{C}^s_T(\mu_0, \mu); \mu \in \mathcal{P}_1(M^*) \Big\} \tag{6.11}$$

is merely a case of expanding out definitions, while the reverse direction can be shown by constructing a sequence random variables distributed according to the optimal transport measures.

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

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

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

Define $\overline{B}_{\mu_0} : \nu \mapsto \overline{B}_T(\mu_0, \nu)$, and note that the interpolation formula (6.5) and a result of Mikima–Thieullen [18] (Lemma 3.2) concerning the convexity of $\bar{C}^s_T$ yields that $\overline{B}_{\mu_0}$ is a concave function. Furthermore it is weak$^\star$-upper semi-continuous on $\mathcal{P}_1(M)$. Thus, we have

$$\overline{B}_{\mu_0}(\nu_T) = -(-\overline{B}_{\mu_0})^{**}(\nu_T) = \inf_{f \in \mathrm{Lip}(M)} \left\{ -\int_M f \, d\nu_T + (-\overline{B}_{\mu_0})^*(f) \right\}. \tag{6.12}$$

Investigating the dual, we find

$$\begin{aligned}
(-\overline{B}_{\mu_0})^*(f) &= \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int_M f \, d\nu + \overline{B}_{\mu_0}(\nu) \right\} \\
&= \sup_{\substack{\mu \in \mathcal{P}_1(M^*) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int_M f \, d\nu + \overline{W}(\mu, \nu) - \bar{C}^s_T(\mu_0, \mu) \right\} \\
&= \sup_{\mu \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} (-f)^* \, d\mu - \bar{C}^s_T(\mu_0, \mu) \right\}.
\end{aligned} \tag{6.13}$$

Note that in the case where $(-f)^* \in C^\infty_{\mathrm{db}}$, this is simply $\int_{M^*} \bar{\Psi}_{(-f)^*,-} \, d\mu_0$, yielding

$$\overline{B}_{\nu_0}(\mu_T) \le \inf_{(-f)^* \in C^\infty_{\mathrm{db}}} \left\{ -\int_{M^*} f \, d\mu_T + \int_{M^*} \bar{\Psi}_{(-f)^*,-} \, d\mu_0 \right\}.$$

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 $g^* = (-f)^{**} \le -f$ [11, Proposition 4.1]. Thus it suffices to consider $f$ convex.

We now show that it is sufficient to consider this infimum over convex $g \in C^\infty_{\text{db}}$ 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:

$$g^*_\epsilon(v) = \sup_x \left\{\langle v, x\rangle - \mathbb{E}\left[g(x+H_\epsilon)\right]\right\} \overset{(\mathrm{J})}{\leq} \sup_x \{\langle v, x\rangle - g(x)\} = g^*(v).$$

Mikami [18, Proof of Theorem 2.1] further shows that

$$(6.13) = C^*_{\nu_0}(g_\epsilon) \leq \frac{C^*_{\nu_0 * \eta_\epsilon}((1+\Delta L(0,\epsilon))g)}{1+\Delta L(0,\epsilon)} + T\frac{\Delta L(0,\epsilon)}{1+\Delta L(0,\epsilon)}.$$

Putting these together, we get

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

Once we take the infimum over convex $g \in \mathrm{Lip}(M)$, we get

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

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 $\mu_0$ absolutely continuous with respect to Lebesgue measure. Then, the pair $(V, X)$ is optimal for (6.3) if and only if there is an Ito process $V(t)$ that satisfies the backward stochastic differential equation:*

$$dV = \nabla_p \overline{H}(\nabla\psi(t,V), V)\, dt + dW_t, \tag{6.14}$$
$$X = \nabla\bar{\psi}(V(T)), \tag{6.15}$$

*where* $\lim_{n\to\infty} \psi_n(T,x) \to \bar{\psi}(x)$ $\nu_T$*-a.s. and* $\lim_{n\to\infty} \psi_n(t,x) = \psi(t,x)$ $\mathbb{P}_V$*-a.s. for some sequence* $\psi_n(t,x)$ *that solves (HJB2) in such a way that* $\psi^0_n = \psi_n(0,\cdot)$ *and* $\psi^T_n = \psi_n(T,\cdot)$ *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 (HJB2) with convex final condition $\psi^T_n$, such that

$$\mathbb{E}\left[\langle X, V(T)\rangle - \int_0^T \bar{L}(V, \beta_V(t,V))\, dt\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\psi^T_n\right)^*(X) + \psi^0_n(V(0))\right], \tag{6.16}$$

which we write as

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

Applying Itô's formula to the last two terms, with the knowledge that $\psi_n$ satisfies (HJB2), we get

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

However, by the definition of the Hamiltonian, we have $-\langle q, x\rangle - \overline{H}(x, v) \geq -\bar{L}(v, q)$, similarly $\psi^*(v) + \psi(x) \geq \langle v, x\rangle$. These inequalities allow us to separate the limit in (6.16) into two requirements:

(a) $\langle \beta_V, \nabla\psi_n^t(V(t))\rangle + \overline{H}(\nabla\psi_n^t(V(t)), V(t))$ must converge to $\bar{L}(V, \beta_V(t, V))$ and
(b) $\psi_n^T(V(T)) + \left(\psi_n^T\right)^*(X)$ must converge to $\langle X, V(T)\rangle$ in $L^1$ hence a subsequence $\psi_{n_k}$ exists such that this convergence is a.e.

The journey from (a) to (6.14) is as in Corollary 4. The only difference from the earlier corollary is that we know that $\psi_n$ must converge to a convex function, so (b) implies $X = \nabla \lim_{n\to\infty} \psi_n(V(T))$.

## 7 Final remarks

The interpolation formula can be seen as a Hopf–Lax formula on Wasserstein space. Indeed, define for a fixed $\mu_0$ on $M^*$ (resp., fixed $\nu_T$ on $M$), the following function $\underline{\mathcal{B}}^{\mu_0}$ (resp., $\overline{\mathcal{B}}^{\nu_T}$) of the terminal (resp., initial) measure,

$$\underline{\mathcal{B}}^{\mu_0}(t, \nu) = \inf\left\{\underline{\mathcal{U}}^{\mu_0}(\varrho) + C_t(\varrho, \nu);\ \varrho \in \mathcal{P}(M)\right\}, \tag{7.1}$$

and

$$\overline{\mathcal{B}}^{\nu_T}(t, \mu) = \inf\left\{\overline{\mathcal{U}}^{\nu_T}(\varrho) - \tilde{C}_t(\varrho, \mu);\ \varrho \in \mathcal{P}(M^*)\right\}, \tag{7.2}$$

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 $\varrho$ is any probability density on $M$ (resp., $M^*$) and $w$ is a vector field on $M$ (resp., $M^*$). By noting that $\underline{\mathcal{B}}^{\mu_0}(T, \nu) = \underline{B}_T(\mu_0, \nu)$, one deduces the following.

**Corollary 12** *Assume $L$ satisfies hypothesis (A0) and (A1), and let $\mu_0$ be a probability measure on $M^*$ with compact support, then*

$$\underline{\mathcal{B}}^{\mu_0}(T, \nu) = \inf\left\{\underline{\mathcal{U}}^{\mu_0}(\varrho_0) + \int_0^T \mathcal{L}(\varrho_t, w_t)dt;\ \partial_t\varrho + \nabla\cdot(\varrho w) = 0,\ \varrho_T = \nu\right\}, \tag{7.3}$$

*where the set of pairs $(\varrho, w)$ considered above are such that $t \to \varrho_t \in \mathcal{P}_1(M)$ (resp., $t \to w_t(x) \in$* 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

$$\begin{cases} \partial_t B + \mathcal{H}(t, \nu, \nabla_\nu B(t, \nu)) = 0, \\ B(0, \nu) = \underline{W}(\mu_0, \nu). \end{cases} \tag{7.4}$$

https://doi.org/10.1017/S0956792519000032 Published online by Cambridge University Press

Here the Hamiltonian is defined as follows:

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

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 (7.4). As importantly, Gangbo–Sweich [15] have shown recently that under certain conditions, value functionals yield solutions to the so-called *Master equations* of mean field games [10]. We refer to their paper for the relevant definitions.

**Theorem 10** *(Gangbo–Swiech) Assume $\mathcal{U}_0 : \mathcal{P}(M) \to \mathbb{R}$, and $U_0 : M \times \mathcal{P}(M) \to \mathbb{R}$ 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 $U : [0, T] \times M \times \mathcal{P}(M) \to \mathbb{R}$ 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 (7.5).*

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 $\mu_0, \nu_T$, of a function $\beta : [0, T] \times M \times \mathcal{P}(M) \to \mathbb{R}$ such that

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

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

$$\begin{cases} \partial_t \beta + \int \langle \nabla_\nu \beta(t, x, \nu) \cdot \nabla H(x, \nabla_x \beta) \rangle \, d\nu + H(x, \nabla_x \beta(t, x, \nu)) = 0, \\ \partial_t \varrho + \nabla(\varrho \nabla H(x, \nabla_x \beta)) = 0, \\ \beta(0, \cdot, \cdot) = \beta_0, \quad \varrho(T, \cdot) = \nu_T, \end{cases} \tag{7.5}$$

where $\beta_0(x, \varrho) = \varphi_\varrho(x)$, where $\varphi_\varrho$ is the convex function such that $\nabla \varphi_\varrho$ pushes $\mu_0$ into $\varrho$.

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*

$$\begin{aligned} \underline{\mathcal{B}}^{\mu_0}(T, \nu) &:= \underline{B}^s_T(\mu_0, \nu) \\ &= \inf\{\underline{\mathcal{U}}^{\mu_0}(\varrho_0) + \int_0^T \mathcal{L}(t, \varrho_t, w_t) \, dt;\ \partial_t \varrho + \nabla \cdot (w\varrho) - \frac{1}{2} \Delta \varrho = 0, \varrho_T = \nu\}. \end{aligned} \tag{7.6}$$

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

$$b_g(T, v, x) := \inf \left\{ g(v, \gamma(0)) + \int_0^T L(\gamma(t), \dot{\gamma}(t))\, dt; \gamma \in C^1([0, T], M) \right\}, \tag{7.7}$$

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

## Conflicts of interest

None.

## References


[1] ADAMS, D. R. & HEDBERG, L. I. (1996) *Function Spaces and Potential Theory,* Vol. 314, Grundlehren der Mathematischen Wissenschaften Springer-Verlag, Berlin.

[2] AMBROSIO, L. (2003) Lecture notes on optimal transport problems. In: *Mathematical Aspects of Evolving Interfaces*, Lecture Notes in Mathematics, Vol. 1812, Springer-Verlag, Berlin/New York, pp. 1–52.

[3] AMBROSIO, L., & FENG, J. (2014) On a class of first order Hamilton Jacobi equations in metric spaces. *J. Differ. Equations* **256**(7), 2194–2245.

[4] AMBROSIO, L., GIGLI, N. & SAVARÉ, G. (2005) *Gradient Flows in Metric Spaces and in the Wasserstein Space of Probability Measures.* Lectures in Mathematics, ETH Zurich, Birkhäuser.

[5] BARTON, A. & GHOUSSOUB, N. (2017) *On Optimal Stochastic Ballistic Transports*, arXiv:1712.00047, 18 pp.

[6] BEIGLBÖCK, M. & JUILLET, N. (2016) On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44**(1), 42–106

[7] BERNARD, P. & BUFFONI, B. (2007) Optimal mass transportation and Mather theory. *J. Eur. Math. Soc.* **9**(1), 85–121.

[8] BRENIER, Y. (1991) Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44**, 375–417.

[9] BOROUSHAKI, S. & GHOUSSOUB, N. (2019) A self-dual variational approach to stochastic partial differential equations. *J. Funct. Anal.* **276**(4), 1201–1243.

[10] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. & LIONS, P.-L. (2019) *The master equations and the convergence problem in Mean Field Games*, Annals of Mathematics Studies (AMS-201), Princeton University Press, ISBN 9780691190709 (232 pp.)

[11] EKELAND, I. & TÉMAN, R. (1987) *Convex Analysis and Variational Problems.* Classics in Applied Mathematics, Vol. 28, Society for Industrial and Applied Mathematics.

[12] EVANS, L. C. & GANGBO, W. (1999) Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.* **137**, 1–66.

[13] FATHI, A, & FIGALLI, A. (2010) Optimal transportation on non-compact manifolds. *Isr. J. Math.* **175**(1), 1–59

[14] GANGBO, W. & MCCANN, R. J. (1996) The geometry of optimal transportation. *Acta Math.* **177**, 113–161.

[15] GANGBO, W. & SWIECH, A. (2015) Existence of a solution to an equation arising from the theory of Mean Field Games. *J. Differ. Equations* **259**(11), 6573–6643.

[16] GHOUSSOUB, N. (2017) *Optimal Ballistic Transport and Hopf-Lax Formulae on Wasserstein Space*, ArXiv e-print: 1705.05951.

[17] JOUKOVSKAIA, T. (1991) *Singularités de Minimax et Solutions Faibles d'Équations aux Dérivées Partielles*. Thèse de Doctorat, Université de Paris VII, Denis Diderot.

[18] MIKAMI, T. & THIEULLEN, M. (2006) Duality theorem for the stochastic optimal control problem. *Stoch. Process. Appl.* **116**(12), 1815–1835

[19] MONGE, G. (1781) Mémoire sur la Théorie des Déblais et des Remblais. Hist. de l'Acad. des Sciences de Paris, Paris, 666–704.

[20] SUDAKOV, V. N. (1979) Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* **141**, 1–178.

[21] ROCKAFELLAR, R. T. (1971) Existence and duality theorems for convex problems of Bolza. *Trans. Amer. Math. Soc.* **159**, 1–40.

[22] ROCKAFELLAR, R. T. & WOLENSKI, P. R. (2001) Convexity in Hamilton-Jacobi theory I: dynamics and duality. *SIAM J. Control Opt.* **39**, 1323–1350.

[23] ROCKAFELLAR, R. T. & WOLENSKI, P. R. (2001) Convexity in Hamilton-Jacobi theory II: envelope representations. *SIAM J. Control Opt.* **39**, 1351–1372.

[24] SCHACHTER, B. (2017) *An Eulerian Approach to Optimal Transport with Applications to the Otto Calculus,* Thesis, U. of Toronto.

[25] VILLANI, C. (2004) Topics in mass transportation. In: *Graduate Studies in Mathematics*, Vol. 58, American Mathematical Society, Providence, RI.

[26] VILLANI, C. (2005) *Optimal Transport, Old and New*, volume 338. Springer Science & Business Media, 2008.