ON OPTIMAL STOCHASTIC BALLISTIC TRANSPORTS

ALISTAIR BARTON AND NASSIF GHOUSSOUB

ABSTRACT. For a given Lagrangian $L : [0,T] \times M \times M^* \to \mathbb{R}_+$ and probability measures $\mu \in \mathcal{P}(M^*)$, $\nu \in \mathcal{P}(M)$, we introduce the stochastic ballistic transportation problems

$$(\star) \qquad \underline{B}(\mu,\nu) := \inf \left\{ \mathbb{E}\left[\langle V, X_0 \rangle + \int_0^T L(t, X, \beta(t,X))\, dt \right] \middle| V \sim \mu, X_T \sim \nu \right\}$$

$$(\star\star) \qquad \overline{B}(\nu,\mu) := \sup \left\{ \mathbb{E}\left[\langle V, X_T \rangle - \int_0^T L(t, X, \beta(t,X))\, dt \right] \middle| V \sim \mu, X_0 \sim \nu \right\}$$

where X is a diffusion process with drift β. This cost is based on the stochastic optimal transport problem presented by Mikami and the deterministic ballistic transport introduced by Ghoussoub. We obtain a Kantorovich-style duality result that reformulates this problem in terms of solutions to the Hamilton-Jacobi-Bellman equation

$$\frac{\partial \phi}{\partial t} + \frac{1}{2}\Delta \phi + H(t, x, \nabla \phi) = 0,$$

and show how optimal processes may be thereby attained.

1. INTRODUCTION & RELATED WORKS

In its modern incarnation, the problem of optimal transportation conceived by Monge [11] is formulated as follows: let X, Y be measure spaces with a cost function $c : X \times Y \to \mathbb{R}$. For two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ the minimum cost of transportation between them is then

$$(1.1) \qquad C(\mu,\nu) = \inf \left\{ \int_{X \times Y} c(x,y)\, d\pi(x,y) \middle| \pi \in \mathcal{T}(\mu,\nu) \right\},$$

where $\mathcal{T}(\mu,\nu)$ is the set of admissable *transport plans* for (μ,ν). A transport plan π is admissable if it is a probability measure on $X \times Y$ with first marginal $(\mathrm{proj}_1)_\# \pi = \mu$ and second marginal $(\mathrm{proj}_2)_\# \pi = \nu$.

A well known cost on $\mathbb{R}^n \times \mathbb{R}^n$ is the Wasserstein cost $c(x,y) := \langle x,y \rangle$ studied by Brenier [3], whose associated transportation costs are

$$(\underline{W}) \qquad \underline{W}(\mu,\nu) := \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x,y \rangle\, d\pi(x,y) \middle| \pi \in \mathcal{T}(\mu,\nu) \right\},$$

$$(\overline{W}) \qquad \overline{W}(\mu,\nu) := \sup \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} \langle x,y \rangle\, d\pi(x,y) \middle| \pi \in \mathcal{T}(\mu,\nu) \right\}.$$

These problems are clearly related by noting that $\underline{W}(\mu, \nu) = -\overline{W}(\hat{\mu}, \nu)$ where $\hat{\mu}$ is the reflection of μ (i.e., $\hat{\mu}(A) := \mu(-A)$). Probabilistically, we can rewrite these as:

$$\underline{W}(\mu, \nu) := \inf_{\pi} \left\{ \mathbb{E}\left[\langle X, Y \rangle\right] | X \sim \mu, Y \sim \nu \right\},$$

$$\overline{W}(\mu, \nu) := \sup_{\pi} \left\{ \mathbb{E}\left[\langle X, Y \rangle\right] || X \sim \mu, Y \sim \nu \right\},$$

respectively. Note that, in general, (X, Y) are not independent. Indeed Brenier showed that in the case where μ, ν are absolutely continuous with respect to the Lebesgue measure then, in the optimal case for $\underline{W}$, Y is completely determined by X (in particular, there exists a convex function ϕ such that $Y = \nabla\phi(X)$) [3].

The *dual* formulation, discovered by Kantorovich [7], allows us to equate the *primal* problem considered in eq. (1.1) to a problem on the space of functions:

$$(1.2) \qquad C(\mu, \nu) = \sup \left\{ \int_X f(x)\, d\mu(x) + \int_Y g(y)\, d\nu(y) \middle| (f, g) \in \mathcal{K}(c) \right\}.$$

Here the set of *Kantorovich potentials* $\mathcal{K}(c)$ is the set of functions $(f, g) \in L^1(d\mu) \times L^1(d\nu)$ such that $f(x) + g(y) \leq c(x, y)$. It is clear that the Kantorovich potentials may be assumed to satisfy

$$f(x) = \inf_y \{c(x, y) - g(y)\} \qquad\qquad g(y) = \inf_x \{c(x, y) - f(x)\}.$$

For Wasserstein costs, this amounts to f and g being Legendre duals of each other: $g(y) = \tilde{f}(y)$ and $f(x) = \tilde{g}(x)$ for $\underline{W}$, and $g(y) = f^*(y)$ and $f(x) = g^*(x)$ for $\overline{W}$. Here $\tilde{f}$ (tesp., f^*) denote the concave (resp., convex) Legendre transform of f.

In this paper, we will investigate the ballistic stochastic dynamic transportation problem. This is a variant of the transportation problem developed by combining the ballistic and stochastic dynamic problems that we will briefly summarize for contextual purposes.

Bernard and Buffoni [2] introduced the idea of a dynamic cost function derived from the idea of minimizing the total action associated with a Lagrangian $L : [0, T] \times TM \to \mathbb{R}_+$ in the transportation between two measures ν_0, ν_T on a manifold M (here TM is the tangent bundle on the manifold M). This was motivated to connect Mather theory with optimal transport. Here the cost $c_s^t(x, y)$ associated with the transportation problem denoted $C_s^t(\nu_s, \nu_t)$ is given by

$$(1.3) \qquad c_s^t(x, y) := \inf \left\{ \int_s^t L(t, \gamma(t), \dot{\gamma}(t))\, dt \middle| \gamma \in C^1([s, t], M), \gamma(s) = x, \gamma(t) = y \right\},$$

where $s < t$ and $(s, t) \in [0, T]^2$. They show [2, Proposition 17] that this problem has a dual formulation akin to eq. (1.2), where the Kantorovich potentials may be restricted to the set $\{(u(x, T), -u(x, 0)) | u \in HJ\}$ where HJ is the set of continuous viscosity solutions of the Hamilton-Jacobi equation

$$(\text{HJ}) \qquad\qquad \frac{\partial u}{\partial t} + H(t, x, \nabla u) = 0,$$

and $H(t, x, p) := \sup_v \{\langle p, v \rangle - L(t, x, v)\}$ is the hamiltonian associated with L. They further showed [2, Theorem B] that, in the case where the measures are absolutely continuous, the optimal transportation measure can be described as the Hamiltonian flow of an initial momentum measure.

The ballistic cost was recently introduced by Ghoussoub [6]. This is a cost on $M^* \times M$ (where $M = \mathbb{R}^n$) taking the form:

$$(1.4) \qquad b_s^t(v, x) := \inf \left\{ \langle v, \gamma(s) \rangle + \int_s^t L(t, \gamma(t), \dot{\gamma}(t)) \, dt \bigg| \gamma \in C^1([s, t], M), \gamma(t) = x \right\},$$

This was used to lift the Hopf-Lax formula to Wasserstein space. Ghoussoub also showed that this cost corresponds to the dual problem

$$\underline{B}_c(\mu_0, \nu_T) = \sup \left\{ \int_M V(T, x) \, d\nu_T + \int_{M^*} \tilde{V}_0(v) \, d\mu_0 \bigg| V \in HJ_c \right\}$$

where HJ_c is the set of solutions to the Hamilton-Jacobi equation (HJ) with initial condition $u(x, 0)$ concave. Here $\tilde{V}_0(v) := \inf_{x \in M}\{\langle v, x \rangle + V_0(x)\}$ is the concave legendre transform. This result was obtained through the following interpolation result

$$\underline{B}_c(\mu_0, \nu_T) = \inf \left\{ \underline{W}(\mu_0, \nu_0) + C_c(\nu_0, \nu_T) | \nu_0 \in \mathcal{P}(M) \right\}.$$

This is akin to the Hopf-Lax formula on Wasserstein space for the functional $\mathcal{U}^{\mu_0} : \rho \mapsto \underline{W}(\mu_0, \rho)$ under the Lagrangian

$$\mathcal{L}(t, \rho, v) = \int_M L(t, x, v(x)) \, d\rho(x)$$

where v is a vector field related to ρ via the continuity equation $\frac{d\rho}{dt} + \text{div}(\rho v) = 0$. (See Villani [12, Chapter 8]).

The following stochastic variant of the action based transportation cost was considered by Mikami and Thieullen [10].

$$(1.5) \qquad C_S^\epsilon(\nu_0, \nu_T) := \inf \left\{ \mathbb{E} \left[\int_0^T L(t, X, \beta_X(t, X)) \, dt \right] \bigg| X \in \mathcal{A}_{\nu_0}^{\nu_T}(\epsilon), X_T \sim \nu_T \right\}.$$

The set of *transportation processes* $\mathcal{A}_{\nu_0}^{\nu_T}(\epsilon)$ may be understood as the set of stochastic processes solving the stochastic differential equation $dX = \beta_X(t, X) \, dt + \epsilon dW_t$ for some $\beta_X(t, X)$, where W_t is $\sigma(X_s : 0 \le s \le t)$-Brownian motion, initially distributed according to the measure ν_0 and distributed at time T according to ν_T. An earlier result from Mikami [9] shows that the deterministic case considered by Brenier [3], i.e., when $L(x, v) = \frac{1}{2}|v|^2$, is recovered from the corresponding stochastic transport as one takes $\epsilon \to 0$.

This problem cannot be formulated in the same way as deterministic transportation problems (i.e., eq. (1.1))—indeed, the corresponding cost function could be $c(x, y) = C(\delta_x, \delta_y) = \infty$ everywhere even under reasonable assumptions on the Lagrangian. Therefore, classical Monge-Kantorovich duality need not apply here. Nevertheless, Mikami and Thieullen [10, Theorem 2.1] derived the following dual formulation through Legendre transforms considerations:

$$C(\nu_0, \nu_T) := \sup \left\{ \int_M \phi(T, x) \, d\nu_T(x) - \int_M \phi(0, x) \, d\nu_0(x) \bigg| \phi \in HJB \right\}.$$

Here, HJB is the set of classical solutions to the following Hamilton-Jacobi-Bellman equation

$$(\text{HJB}) \qquad\qquad \frac{\partial \phi}{\partial t} + \frac{\epsilon}{2}\Delta\phi + H(t, x, \nabla\phi) = 0,$$

with final condition smooth and bounded. Furthermore for any optimal process X there is a sequence of solutions ϕ_n to (HJB) such that the SDE

$$dX = \lim_{n \to \infty} \nabla_p H(t, X, \nabla\phi_n(t, x)) dt + \epsilon dW_t$$

is satisfied—this is the counterpart to the Hamiltonian flow for the stochastic cost.

In this paper we aim to combine the results of Ghoussoub, and Mikami and Thieullen, by considering the stochastic ballistic transportation problem:

$$\underline{B}(\mu_0, \nu_T) := \inf\left\{ \mathbb{E}\left[\langle V, X(0)\rangle + \int_0^T L(t, X, \beta_X(t, X))\, dt\right] \middle| V \sim \mu_0, X(\cdot) \in \mathcal{A}^{\nu_T}\right\}.$$

This may be considered in either the framework of stochastic control theory, or a stochastic version of the Hopf-Lax equation in measure space.

2. ASSUMPTIONS & MAIN RESULTS

2.1. **Assumptions.** We will operate on the following assumptions on the Lagrangian, that reduce to those considered by Mikami and Thieullen in [10]:

(A0) $(t, x, v) \mapsto L(t, x, v)$ is $\mathcal{C}^3$, always positive, and has $\nabla_v^2 L(t, x, v) > 0$.

(A1) There exists $\delta > 1$ such that

$$\liminf_{|u| \to \infty} \frac{\inf_{t,x} L(t, x, u)}{|u|^\delta} > 0.$$

   (In section 2.1 we show that this is equivalent $L(t, x, v)$ being bound below by a convex function $\ell(v) \in \Omega(|v|^\delta)$).

(A2) $\Delta L(\epsilon_1, \epsilon_2) := \sup\left\{ \frac{1 + L(t, x, u)}{1 + L(s, y, u)} - 1 \middle| |t - s| < \epsilon_1, |x - y| < \epsilon_2\right\} \xrightarrow{\epsilon_1, \epsilon_2 \to 0} 0.$

(A3)   (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.

(A4) (i) $\Delta L(0, \infty) < \infty$ or (ii) $\delta = 2$ in (A1).

We give a brief summary of the role of the above assumptions:

(A1) is a coercivity result that is necessary for sequential compactness of minimizing transportation processes (proposition 1).

(A2) is used to show that the expected action of a transportation process is lower semicontinuous [9, eqs. (3.17),(3.38)-(3.41)].

(A3) is used with (A0) to derive the Hamilton-Jacobi-Bellman equation in proposition 3 [5, p. 210, Remark 11.2].

(A4,i) allows us to uniformly bound the ratio $\frac{1 + L(t, x, u)}{1 + L(t, y, u)}$, while (A4,ii) ensures the minimizing β_X satisfies $\int_0^T |\beta_X(t, X)|^2\, dt < \infty$ a.s., permitting us to assume the process $X(\cdot)$ is absolutely continuous with respect to $W(\cdot)$ [8, Theorem 7.16]. Either of these results with (A0)-(A3) is sufficient to show convexity of $(\nu_0, \nu_T) \mapsto C(\nu_0, \nu_T)$ [10, Lemma 3.2].

Remark. (A1) is equivalent to $L(t, x, v)$ being bound below by a convex function $\ell(v)$ such that

$$\liminf_{|v| \to \infty} \frac{\ell(v)}{|v|^\delta} > 0.$$

Indeed, if L is bound by such an ℓ, it is simple to see (A1) holds. We prove the reverse direction by construction. By (A1), we may assume that there exists a $U \in \mathbb{R}$ such that for all $|u| > U$

$$\frac{\inf_{t,x} L(t, x, u)}{|u|^\delta} > \alpha, \qquad (\alpha > 0, \delta > 1),$$

defining α, δ. Then the convex function $\ell : M^* \to \mathbb{R}$ defined by

$$\ell(v) := \begin{cases} 0 & |v| < U \\ \alpha \, ||v| - U|^\delta & |v| \geq U \end{cases}$$

is a lower bound on $L(t, x, v)$ and is asymptotically bounded below by $|v|^\delta$.

2.2. **Notation.** The space of probability measures on a space X will be indicated by $\mathcal{P}(X)$, while the subset of measures with finite barycenter will be denoted $\mathcal{P}_1(X) := \{\mu \in \mathcal{P}(X) : \int |x| \, d\mu < \infty\}$. We will work on the space $M := \mathbb{R}^d$ however we will still refer to the dual space M^* explicitly, for clarity. Measures in $\mathcal{P}(M)$ will be designated by ν, while μ will designate measures in $\mathcal{P}(M^*)$.

Given a Lagrangian $L : [0, T] \times M \times M^* \to \mathbb{R}$ satisfying properties (A), we define the (stochastic) dynamic and ballistic variants of the optimal transportation problem mentioned in the introduction:

$$C(\nu_0, \nu_T) := \inf \left\{ \mathbb{E} \left[\int_0^T L(t, X(t), \beta_X(t, X(t))) \, dt \right] \middle| X(\cdot) \in \mathcal{A}_{\nu_0}^{\nu_T} \right\},$$

$$\underline{B}(\mu_0, \nu_T) := \inf \left\{ \mathbb{E} \left[\langle Y, X(0) \rangle + \int_0^T L(t, X(t), \beta_X(t, X(t))) \, dt \right] \middle| Y \sim \mu_0, X(\cdot) \in \mathcal{A}^{\nu_T} \right\}.$$

Here $\mathcal{A}_{\nu_0}^{\nu_T}$ refers to the set of $\mathbb{R}^d$-valued continuous semimartingales $X(\cdot)$—initially distributed according to ν_0 and finally distributed according to ν_T—such that there exists a measurable drift $\beta_X : [0, T] \times C([0, T]) \to M^*$ where

(i) $\omega \mapsto \beta_X(t, \omega)$ is $\mathcal{B}(C([0, t]))_+$-measurable for all t.

(ii) $W(t) := X(t) - X(0) - \int_0^t \beta_X(s, X) \, ds$ is a $\sigma[X(s) : s \in [0, t]]$-Brownian motion.

$\mathcal{A}^{\nu_T}$ refers to the same processes, except with initial distribution unconstrained, likewise $\mathcal{A}_{\nu_0}$ with the final distribution unconstrained and $\mathcal{A}$ with initial and final distribution unconstrained.

For convenience, we define the expected action of stochastic processes $X(\cdot) \in \mathcal{A}$:

$$\mathscr{A}(X) := \mathbb{E} \left[\int_0^T L(t, X(t), \beta_X(t, X(t))) \, dt \right].$$

We will also consider another, analogous, ballistic stochastic cost defined by:

$$\overline{B}(\nu_0, \mu_T) := \sup \left\{ \mathbb{E} \left[\langle V, X(T) \rangle - \int_0^T L(t, X(t), \beta_X(t, X(t))) \, dt \right] \middle| V \sim \mu_T, X(\cdot) \in \mathcal{A}_{\nu_0} \right\},$$

which we will refer to as the *maximizing ballistic cost*—in comparison, $\underline{B}$ may be referred to as the *minimizing ballistic cost*.

We recall the convex and concave Legendre transforms, which we will denote

$$f^*(v) := \sup_{x \in M} \{\langle v, x \rangle - f(x)\},$$

$$\tilde{f}(v) := \inf_{x \in M} \{\langle v, x \rangle - f(x)\} = -(-f)^*(-v),$$

respectively, where v is in the dual space M^*. This notation will also be used for the legendre duals of functions of measures, such as $C_{\nu_0} : \nu \mapsto C(\nu_0, \nu)$ where the dual space is a linear subspace of continuous functions.

A set of functions that is of particular importance in this regard is the set of smooth functions with bound derivatives—notably, this may be realized as the set of mollified Lipschitz functions—which we will denote C_{db}^∞, and the subset of convex such functions cC_{db}^∞.

2.3. **Main Results.** We present an interpolation result that allows us to interpret the ballistic cost in terms of the Wasserstein cost and the stochastic action cost.

Theorem 1 (Interpolation of $\underline{B}$). *If L satisfies the assumptions (A), then our stochastic ballistic cost may be written as:*

$$\underline{B}(\mu_0, \nu_T) = \inf_\nu \{ \underline{W}(\mu_0, \nu) + C(\nu, \nu_T) \}.$$

This minimum is attained in the case where $\mu_0 \in \mathcal{P}_1(M^*)$ and $\nu_T \in \mathcal{P}_1(M)$.

This leads to our main result, the dual of the stochastic ballistic cost:

Theorem 2 (Dual of $\underline{B}$). *If $\nu_T \in \mathcal{P}_1(M)$ and μ_0 with compact support are such that $\underline{B}(\mu_0, \nu_T) < \infty$, and the Lagrangian satisfies (A0)-(A4) then we have*

$$(2.1) \qquad \underline{B}(\mu_0, \nu_T) = \sup_{-f \in cLip(M)} \left\{ \int_M f(x)\, d\nu_T(x) + \int_{M^*} \widetilde{\phi^f}(0, v)\, d\mu_0(v) \right\},$$

where $\widetilde{g}$ denotes the concave legendre transform of g and ϕ^f is the solution to the Hamilton-Jacobi-Bellman equation

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

This formulation of the problem allows us to write optimal processes in terms of a maximizing sequence of ϕ:

Corollary 2.1 (Optimal Processes for $\underline{B}$). *Under the assumptions of Theorem 2, with $d\mu_0 \ll d\lambda$, we have that $(V, X(t))$ is a minimizing process for $\underline{B}(\mu_0, \nu_T)$ if and only if it is a solution to the SDE*

$$dX = \nabla_p \lim_{n \to \infty} H(t, X, \nabla \phi_n(t, X))\, dt + dW_t$$

$$X(0) = \nabla \lim_{n \to \infty} \left(\widetilde{\phi_n^0} \right)(V),$$

where ϕ_n is a sequence of solutions to (HJB) that is maximizing for the dual problem. (Here we use the notation $\phi_n^t(x) := \phi_n(t, x)$).

2.3.1. *Maximizing Cost.* We obtain parallel results for the maximizing cost $\overline{B}$:

Theorem 3 (Interpolation of $\overline{B}$). *If L matches assumptions (A), then*

$$\overline{B}(\nu_0, \mu_T) = \sup \{ \overline{W}(\nu, \mu_T) - C(\nu_0, \nu) \},$$

where, again, the maximum is attained when $\nu_0 \in \mathcal{P}_1(M)$ and $\mu_T \in \mathcal{P}_1(M^*)$.

Theorem 4 (Dual of $\overline{B}$). *Assume the Lagrangian satisfies the assumptions (A). If $\nu_0 \in \mathcal{P}_1(M)$, μ_T has compact support, and $\overline{B}(\nu_0, \mu_T) < \infty$, then*

$$\overline{B}(\nu_0, \mu_T) = \inf_{g \in cC_{\mathrm{db}}^\infty} \left\{ \int_{M^*} g^*\, d\mu_T + \int_M \phi^g\, d\nu_0 \right\},$$

where ϕ^g solves the Hamilton-Jacobi-Bellman equation

$$(\mathrm{HJB2}) \qquad \frac{\partial \phi}{\partial t} + \frac{1}{2} \Delta \phi - H(t, x, \nabla \phi) = 0 \qquad\qquad \phi(x, T) = g(x)$$

Once more we may identify optimal processes by this formulation:

Corollary 4.1 (Optimal Processes for $\overline{B}$). *If the assumptions of theorem 4 are satisfied with $d\mu_T \ll d\lambda$, then $(V, X(t))$ is a maximizing process for $\overline{B}(\nu_0, \mu_T)$ iff it is a solution to the SDE*

$$dX = \lim_{n\to\infty} \nabla_p H(t, x, \nabla\phi_n(t, X)) \, dt + dW_t$$
$$V = \lim_{n\to\infty} \nabla\phi_n^T(X(T)),$$

for a sequence of solutions ϕ_n to (HJB2) that is minimizing for the dual problem.

3. Proofs

We first show a coercivity result that allows us to explore the structure of the problem and some of the consequences of our assumptions. This proposition is useful for showing existence of minimizers to our interpolation formula in theorem 1.

Proposition 1 (Coercivity of C). *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*

$$\tag{3.1} C(\nu, \nu_T) \le N \int |x| \, d\nu(x)$$

is tight. In other words, for every $\epsilon > 0$ there exists an R such that if $\nu \in \mathcal{P}_1(M)$ satsifies eq. (3.1)), then $\nu(B(R, 0)^c) \le \epsilon$.

Proof. First we define the set of measures $\mathcal{T}_{\epsilon, R} := \{\nu \in \mathcal{P}_1(M) : \nu(B(R, 0)^c) > \epsilon\}$. We will assume $\nu \in \mathcal{T}_{\epsilon, R}$ for what follows.

We begin by defining R_1 to be such that $\nu_T(B(R_1, 0)^c) < \epsilon$. Let $X \in \mathcal{A}$ be a stochastic process with $X(0) \sim \nu$ and $X(T) \sim \nu_T$. We leave R to be defined later, but note that if we define the set $\Omega_R := \{|X(0)| > R\}$, then our assumption on ν yields $\mathbb{P}(\Omega_R) > \epsilon$. By positivity of L, this allows us to say that $\mathscr{A}(X) \ge \mathscr{A}(1_{\Omega_R} X)$ (henceforth we define the process $Y(t) := 1_{\Omega_R}(\omega) X(t)$). Furthermore we can assume the probability of a path originating at a distance beyond R from the origin and terminating within distance R_1 from the origin is

$$\tag{3.2} \begin{aligned} \mathbb{P}(\{|X(0)| > R\} \cap \{|X(T)| \le R_1\}) &\ge 1 - \mathbb{P}(|X(0)| \le R) - \mathbb{P}(|X(T)| > R_1) \\ &= \mathbb{P}(|X(0)| > R) - \mathbb{P}(|X(T)| > R_1) \\ &= \nu(B(R, 0)^c) - \nu_T(B(R_1, 0)^c) > \epsilon/2. \end{aligned}$$

Applying (A1) via our remark in section 2.1, we assume that there is a convex function $\ell : M^* \to \mathbb{R}$ such that for all $|u| > U$

$$\tag{3.3} \frac{\ell(u)}{|u|^\delta} > \alpha, \qquad (\alpha > 0, \delta > 1),$$

defining α, δ. Recall that $\ell(|v|)$ is a lower bound on $L(t, x, v)$.

Examining the process Y, we can take advantage of the convexity of ℓ to apply Jensen's inequality:

$$\tag{3.4} \begin{aligned} \mathbb{E}\left[\int_0^T L(t, Y, \beta_Y(t, Y)) \, dt\right] &\ge \mathbb{E}\left[\int_0^T \ell(|\beta_Y(t, Y)|) \, dt\right] \overset{(J)}{\ge} \mathbb{E}[\ell(|V|)T] \\ &\overset{(3.3)}{>} \alpha T \mathbb{E}\left[1_{|V|>U} |V|^\delta\right] \ge \alpha T \left(\mathbb{E}\left[|V|^\delta\right] - U^\delta\right), \end{aligned}$$

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

Putting this all together, we get

$$
\mathscr{A}(X) \geq \mathscr{A}(Y) \overset{(3.4)}{>} \alpha T \left(\mathbb{E}\left[|V|^\delta \right] - U^\delta \right) = \alpha T \left(\mathbb{E}\left[\left| \frac{Y(T) - Y(0)}{T} \right|^\delta \right] - U^\delta \right)
$$

(3.5)

$$
\overset{(\triangle)}{>} \frac{\alpha}{T^{\delta-1}} \left(\mathbb{E}\left[||Y(0)| - |Y(T)||^\delta \right] - (UT)^\delta \right)
$$

Minimizing the expectation over $(Y(0), Y(T))$ is a type of well known optimal transport problem, with the optimal plan given by the monotone Hoeffding-Frechet mapping $x \mapsto G_{|Y(0)|}(G_{|Y(T)|}^{-1}(x))$, where $G_Z(t) := \inf\{x \in \mathbb{R} : t \geq \mathbb{P}(Z \leq x)\}$ is the quantile function associated with the random variable Z [1, Theorem 1.1]. Thus the optimal plan maps each quantile in one measure to the corresponding quantile in the other.

Focussing on the expectation in eq. (3.5), we get

$$
\mathbb{E}\left[||Y(0)| - |Y(T)||^\delta \right] \geq \int |x - y|^\delta \, d\left(\left(G_{|Y(0)|} \times G_{|Y(T)|} \right)_{\#} \lambda_{[0,1]} \right)(x, y)
$$

(3.6)

$$
\overset{(J)}{\geq} \left| \int x \, d\left(G_{|Y(0)|} \right)_{\#} \lambda_{[0,1]}(x) - b \right|^\delta
$$

$$
= \left(\int_{B(R,0)^c} |x| \, d\nu(x) - b \right)^\delta,
$$

where $b := \int |x| \, d\nu_T / \epsilon > \mathbb{E}\left[|Y(T)| \right]$ and we require that $R > b/\epsilon$. We thus want to find R such that

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

Letting $I_\nu(R) := \int_{B(R,0)} |x| \, d\nu(x)$ (which is at least $R\epsilon$ for all $\nu \in \mathcal{T}_{\epsilon,R}$), we find the condition

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

which by using the mentioned bound on $I_\nu(R)$ is satisfied if R satisfies

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

Taking $R \to \infty$ satisfies this inequality, implying this set is non-empty and intersects with our earlier constraints that $R > R_1 + UT$ and $R > b/\epsilon$. $\qquad \square$

Remark. The same argument holds if we add any fixed constant to the right side of eq. (3.1).

Furthermore, eq. (3.6) can be used to show that if $\nu_1 \in \mathcal{P}_1(M)$ and $\nu_0 \in \mathcal{P}(M) \setminus \mathcal{P}_1(M)$, then $C(\nu_0, \nu_1) = C(\nu_1, \nu_0) = \infty$. This leads to our focus on distributions with finite first moment in the following.

We now state some properties of C shown by Mikami:

Proposition 2. *([10, Theorem 2.1]) If the Lagrangian satisfies (A1)-(A4), then*

$$
C(\nu_0, \nu_T) = \sup \left\{ \int_M f(x) \, d\nu_T - \int_M \phi^f(0, x) \, d\nu_0 \,\middle|\, f \in \mathcal{C}_b^\infty \right\},
$$

where ϕ_f solves (HJB). Furthermore, $(\mu, \nu) \mapsto C(\mu, \nu)$ is convex and lsc under the weak topology.

3.1. **Minimizing Ballistic Cost.** Now that we have demonstrated the coercivity and lower semicontinuity of C, we can move on to investigating the minimizing ballistic cost: $\underline{B}$. These previous results are sufficient to show the existence of a minimizer to the interpolation:

Theorem 1 (Interpolation of $\underline{B}$). *If L satisfies the assumptions (A), then our stochastic ballistic cost may be written*

$$\text{(3.7)} \qquad \underline{B}(\mu_0, \nu_T) = \inf_{\nu}\{\underline{W}(\mu_0, \nu) + C(\nu, \nu_T)\}.$$

Furthermore, this minimum is attained in the case where $\mu_0 \in \mathcal{P}_1(M^*)$ and $\nu_T \in \mathcal{P}_1(M)$.

Proof. First, to show that $\underline{B}(\mu_0, \nu_T)$ is greater or equal to its interpolations, consider $(Y_n, X_n(\cdot))$ such that

$$\mathbb{E}\left[\langle Y_n, X_n(0)\rangle + \int_0^T L(t, X_n(t), \beta_{X_n}(X_n, t))\, dt\right] < \underline{B}(\mu_0, \nu_T) + \tfrac{1}{n}.$$

Let $\nu_n := PX_n(0)^{-1}$, then

$$\underline{W}(\mu_0, \nu_n) \leq \mathbb{E}\left[\langle Y_n, X_n(0)\rangle\right],$$

$$C(\nu_n, \nu_T) \leq \mathbb{E}\left[\int_0^T L(t, X_n(t), \beta_{X_n}(X_n, t))\, dt\right],$$

so

$$\inf_{\nu}\{\underline{W}(\mu_0, \nu) + C(\nu, \nu_T)\} \leq \underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_T) < \underline{B}(\mu_0, \nu_T) + \tfrac{1}{n}.$$

Taking $n \to \infty$ gives the inequality.

To get the reverse inequality, take ν_n to be a sequence such that

$$\underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_T) < \inf_{\nu}\{\underline{W}(\mu_0, \nu) + C(\nu, \nu_T)\} + \tfrac{1}{n}.$$

We may then define an admissible stochastic process $Z_n(\cdot) \in \mathcal{A}_{\nu_n}^{\nu_T}$ such that

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

Then, let $W_n(\omega) \sim \gamma_{Z_n(0)(\omega)}^n$, where $d\gamma_x^n(y) \otimes d\mu_0(x) = d\gamma_n(x, y)$ is the disintegration of a measure γ_n such that

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

Thus

$$\underline{B}(\mu_0, \nu_T) \leq \langle W_n, Z_n\rangle + \int_0^T L(t, Z_n(t), \beta_{Z_n}(Z_n, t))\, dt < \underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_T) + \tfrac{2}{n}$$

$$< \inf_{\nu}\{\underline{W}(\mu_0, \nu) + C(\nu, \nu_T)\} + \tfrac{3}{n}.$$

Lastly, to show this minimum is achieved in the case where μ_0 and ν_T have finite mean, we must show that the minimizing sequence ν_n is sequentially compact in the weak topology. To do this, we note that the set of measures ν such that

$$C(\nu, \nu_T) < N \int |y|\, d\nu(y) + \underline{B}(\mu_0, \nu_T) + 1$$

is tight by proposition 1. If we let $N := \int |x| \, d\mu_0(x)$, then the collection of measures such that

$$\underline{B}(\mu_0, \nu_T) + 1 > C(\nu, \nu_T) + \underline{W}(\mu_0, \nu)$$
$$> C(\nu, \nu_T) - \int |x| \, |y| \, d\mu_0(x) \, d\nu(y)$$
$$\overset{(F)}{=} C(\nu, \nu_T) - N \int |y| \, d\nu(y),$$

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

Remark. The attainment of a minimizing ν_0 of eq. (3.7) is sufficient to show that of a minimizer of $\underline{B}$ when it is finite, since the minimum is achieved for both $\underline{W}$ and C [10, Proposition 2.1].

Corollary 1.1 (Properties of $\underline{B}$). *If the Lagrangian satisfies (A), then:*
  a) $\nu \mapsto \underline{B}(\mu_0, \nu)$ *is lower semi-continuous on* $\mathcal{P}_1(M)$ *for all* $\mu_0 \in \mathcal{P}_1(M^*)$.
  b) $(\mu_0, \nu_T) \mapsto \underline{B}(\mu_0, \nu_T)$ *is jointly convex.*

Proof of a). Consider a sequence $\nu_{T,n} \in \mathcal{P}_1(M)$ that converges weakly to $\nu_T \in \mathcal{P}_1(M)$. We have

$$\underline{B}(\mu_0, \nu_{T,n}) = \underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_{T,n}).$$

for a sequence of measures $\nu_n \in \mathcal{P}_1(M)$ by theorem 1. Since C is lsc (proposition 2), we get

$$\liminf_{n \to \infty} \underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_{T,n}) \geq \liminf_{n \to \infty} \underline{W}(\mu_0, \nu_n) + C(\nu_n, \nu_T) \geq \underline{B}(\mu_0, \nu_T).$$

□

Lemma 1.1. *If* $(x, y) \mapsto h(x, y)$ *jointly convex, then* $\bar{h}(x) := \inf_y h(x, y)$ *is convex.*

Proof. Consider $x_1, x_2 \in X$. Then there exist sequences $y_1^n, y_2^n \in Y$ such that

$$\lim_n h(x_i, y_i^n) = \bar{h}(x_i).$$

Now take $x_\lambda := \lambda x_1 + (1 - \lambda)x_2$, and $y_\lambda^n := \lambda y_1^n + (1 - \lambda)y_2^n$, then

$$\bar{h}(x_\lambda) \leq h(x_\lambda, y_\lambda^n) \leq \lambda h(x_1, y_1^n) + (1 - \lambda)h(x_2, y_2^n) \overset{n \to \infty}{\longrightarrow} \lambda \bar{h}(x_1) + (1 - \lambda)\bar{h}(x_2)$$

□

Proof of corollary 1.1b). Define the function

$$(3.8) \qquad\qquad h((\mu_0, \nu_T), \nu) := \underline{W}(\mu_0, \nu) + C(\nu, \nu_T).$$

Then by theorem 1

$$(3.9) \qquad\qquad \underline{B}(\mu_0, \nu_T) = \inf_\nu \{h((\mu_0, \nu_T), \nu)\}.$$

But $h((\mu_0, \nu_T), \nu)$ is jointly convex—as $(\mu_0, \nu) \mapsto \underline{W}(\mu_0, \nu)$ and $(\nu, \nu_T) \mapsto C(\nu, \nu_T)$ are both jointly convex (proposition 2). Hence applying lemma 1.1 gives the joint convexity of eq. (3.9). □

Now that we have shown that $\nu_T \mapsto \underline{B}(\mu_0, \nu_T)$ is convex and lsc, we are able apply convex duality to get our dual formulation.

Theorem 2 (Dual of $\underline{B}$). *If $\nu_T \in \mathcal{P}_1(M^*)$ and μ_0 with compact support are such that $\underline{B}(\mu_0, \nu_T) < \infty$, and the Lagrangian satisfies (A0)-(A4), then we have*

$$(3.10) \qquad \underline{B}(\mu_0, \nu_T) = \sup_{-f \in cLip(M)} \left\{ \int f(x) \, d\nu_T(x) + \int \widetilde{\phi^f}(0, v) \, d\mu_0(v) \right\},$$

where $\widetilde{g}$ is the concave legendre transform of g and ϕ^f is the solution to the Hamilton-Jacobi-Bellman equation

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

Lemma 2.1. *For $\mu \in \mathcal{P}(M)$ with compact support, define $\underline{W}_\mu : \mathcal{M}_1(M^*) \to \mathbb{R}$, (where $\mathcal{M}_1(M^*)$ is the linear space of measures ν on M^* such that $\int_{M^*} (1 + |x|) \, d\mu < \infty$) to be*

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

Then, for $f \in Lip(M)$, the convex Legendre transform of $\underline{W}_\mu$ is given by,

$$\underline{W}_\mu^*(f) = -\int \widetilde{f} \, d\mu.$$

The Lipschitz condition is a consequence of the dual space of $\mathcal{M}_1(M)$ being the space of Lipschitz functions $Lip(M)$ [13].

Proof. It is well known that $\underline{W}_\mu$ is convex and lsc (e.g. via the dual formulation, eq. (1.2)). Thus, by convex analysis, $\underline{W}_\mu = \underline{W}_\mu^{**}$, giving us:

$$\underline{W}_\mu(\nu) = \sup_{f \in Lip(M)} \left\{ \int f \, d\nu - \underline{W}_\mu^*(f) \right\}.$$

But Kantorovich duality (when μ has compact support) also gives us

$$\underline{W}_\mu(\nu) = \sup_{f \in Lip(M)} \left\{ \int f \, d\nu + \int \widetilde{f} \, d\mu \right\} = F_\mu^*(\nu),$$

where $F_\mu : f \mapsto -\int \widetilde{f} \, d\mu = \int (-f)^* \, d\hat{\mu}(v)$ (recall $d\hat{\mu}(v) = d\mu(-v)$). To show that $\underline{W}_\mu^*(f) = F_\mu(f)$ it suffices to show that $f \mapsto F_\mu(f)$ is convex and lower semicontinuous, in which case $\underline{W}_\mu^*(\nu) = F_\mu^{**}(f) = F_\mu(f)$. We will do this by showing $g \mapsto g^*(v)$ is a convex mapping for all $v \in M^*$: Fix $\lambda \in [0, 1]$ and $v \in M^*$ arbitrary.

$$(\lambda g_1 + (1-\lambda)g_2)^*(v) = \sup_x \{\langle x, v \rangle - \lambda g_1(x) - (1-\lambda)g_2(x)\}$$

$$\overset{(1)}{\leq} \sup_x \{\langle x, v \rangle - \lambda(\langle x, v \rangle - g_1^*(v)) - (1-\lambda)(\langle x, v \rangle - g_2^*(v))\}$$

$$= \lambda g_1^*(v) + (1-\lambda)g_2^*(v),$$

where we use $g(x) \geq \langle x, v' \rangle - g^*(v')$ for all $v' \in M^*$ in (1)—taking the liberty to let $v' = v$.

To show F_μ is weakly lsc, consider f_n that converge pointwise to f (a weaker condition than weak convergence). Then

$$\liminf_{n \to \infty} F_\mu(f_n) = \liminf_{n \to \infty} \int (-f_n)^* \, d\hat{\mu} \overset{(\text{F})}{\geq} \int \liminf_{n \to \infty} \sup_x \{\langle v, x \rangle + f_n(x)\} \, d\hat{\mu}$$

$$\geq \int \sup_x \{\langle v, x \rangle + \liminf_{n \to \infty} f_n(x)\} \, d\hat{\mu} = \int (-f)^* \, d\hat{\mu} = F_\mu(f),$$

where we use $\{(-f_n)^*\}$ being bounded below by $\inf_n\{f_n(0)\} > -\infty$ to apply Fatou's lemma (F) and use $\sup g \geq g$ to get to the second line. $\qquad\qquad\qquad\qquad\qquad\square$

Remark. Note that we can limit Kantorovich duality to Lipschitz functions only because μ is compact, hence the optimal f has bounded gradient. We could extend this result to non-compactly supported μ if we consider α-Hölder continuous functions for any $\alpha > 1$ rather than Lipschitz functions. This case is precluded from us here because of the constraint to functions with bounded derivatives in the following proposition.

Since the double dual of $\underline{W}_\mu$ is equivalent to the dual formulation, Brenier's theorem applies. And the maximizing f provides the optimal transport ∇f of the primal problem.

We now state a similar result regarding the dynamic cost:

Proposition 3. *Assume (A) and let $f \in C_{db}^\infty$. Then, the unique solution to eq. (HJB) is given by:*

$$\text{(HJB')} \qquad \phi^f(t,x) = \sup_{X \in \mathcal{A}} \left\{ \mathbb{E}\left[f(X(T)) - \int_t^T L(s, X(s), \beta_X(s, X))\, ds \,\middle|\, X(t) = x \right] \right\}.$$

Moreover, there exists an optimal process X with drift $\beta_X(t, X) = \arg\min_v\{v \cdot \nabla\phi(t,x) + L(t, x, v)\}$.

Remark. This proposition may be shown by removing the boundedness condition in [5, Remark IV.11.2]. In fact it suffices to only have the first three derivatives bounded. Note that integrating eq. (HJB') over $d\nu_0$ yields the legendre transform of $\nu_T \mapsto C(\nu_0, \nu_T)$ for $f \in C_{db}^\infty$. A key step in proving theorem 2 is showing that it is not necessary to consider functions outside of this set.

That this proposition does not necessarily hold for functions with unbounded derivatives confines us to μ with bounded support as we saw in lemma 2.1.

Proof of theorem 2. 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

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

From corollary 1.1, we get that $\underline{B}_{\mu_0}$ is convex and lsc. Thus, by convex analysis, we have

$$\text{(3.11)} \qquad \underline{B}_{\mu_0}(\nu) = \underline{B}_{\mu_0}^{**}(\nu) = \sup_{f \in \text{Lip}(M)} \left\{ \int f\, d\nu - \underline{B}_{\mu_0}^*(f) \right\}.$$

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

$$\begin{aligned} \underline{B}_{\mu_0}^*(f) &:= \sup_{\nu_T \in \mathcal{P}_1(M)} \left\{ \int f\, d\nu_T - \underline{B}(\mu_0, \nu_T) \right\} \\ &\overset{(3.7)}{=} \sup_{\substack{\nu_T \in \mathcal{P}_1(M) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int f\, d\nu_T - C(\nu, \nu_T) - \underline{W}(\mu_0, \nu) \right\} \\ &\overset{(HJB')}{=} \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int \phi^f(0, x)\, d\nu - \underline{W}(\mu_0, \nu) \right\} \\ &= \underline{W}_{\mu_0}^*(\phi^f) = -\int \widetilde{\phi^f}\, d\mu_0. \end{aligned}$$

Thus, plugging this into our dual formula (3.11) and restricting our supremum to C_{db}^∞ gives

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

To show the reverse inequality we will adapt the mollification argument set out in [10, Proof of Theorem 2.1]. We assume our mollifier $\eta_\epsilon(x)$ is such that $\eta_1(x)$ is a smooth function on $[-1,1]^d$ that satisfies $\int \eta_1(x) \, dx = 1$ and $\int x \eta_1(x) \, dx = 0$, then define $\eta_\epsilon(x) = \epsilon^{-d} \eta_1(x/\epsilon)$. Then for Lipschitz f, $f_\epsilon := f * \eta_\epsilon$ is smooth with bounded derivatives. We can derive a bound on $\underline{B}_{\mu * \eta_\epsilon}^*(f)$ by removing the supremum in eq. (3.12) and fixing a process $X \in \mathcal{A}$. This gives us:

$$\mathbb{E}\left[f_\epsilon(X(T)) - \int_0^T L(s, X(s), \beta_X(s, X)) \, ds - \langle X(0), Y \rangle \right] \overset{(A2)}{\leq}$$

$$\mathbb{E}\left[f(X(T) + H_\epsilon) - \int_0^T \frac{L(s, X(s) + H_\epsilon, \beta_X(s, X)) - \Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} \, ds - \langle X(0) + H_\epsilon, Y + H_\epsilon \rangle + |H_\epsilon|^2 \right] \leq$$

$$\frac{\underline{B}_{\mu_0 * \eta_\epsilon}^*(f(1 + \Delta L(0, \epsilon)))}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)} + d\epsilon^2,$$

where H_ϵ is a random variable with distribution $\eta_\epsilon \, dx$ that is independent of $X(\cdot), Y$ (we use this to get the second line), thus $X(T) + H_\epsilon \sim d(\eta_\epsilon * \nu_T)$. The third line arises by maximizing over processes $(X(\cdot) + H_\epsilon, Y + H_\epsilon)$.

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

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

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

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

$\square$

In some sense $\nabla\phi$ is more fundamental than ϕ, since our dual is invariant under $\phi \mapsto \phi + c$. Thus when discussing the convergence of a sequence of ϕ, we refer to the convergence of their gradients. Notably the optimal gradient may not be bounded or smooth, hence may not be achieved within the set C_{db}^∞. In the subsequent corollary, we denote $\mathcal{P}_X$ the measure on $M \times [0, T]$ associated with the process X and refer to solutions to (HJB) by $\phi_n^t(x) := \phi_n(t, x)$ without reference to their final condtion for convenience.

Corollary 2.1 (Optimal Processes for $\underline{B}$). *Suppose the assumptions on theorem 2 are satisfied and $d\mu_0 \ll d\lambda$. Then $(V, X(t))$ minimizes $\underline{B}(\mu_0, \nu_T)$ iff it is a solution to the SDE*

$$(3.13) \qquad\qquad dX = \nabla_p H(t, X, \nabla\phi(t, X)) \, dt + dW_t$$

$$(3.14) \qquad\qquad V = \nabla\bar{\phi}(X(0)),$$

where $\nabla\phi_n(t, x) \to \nabla\phi(t, x)$ $\mathcal{P}_X$-a.s. and $\nabla\phi_n(0, x) \to \nabla\bar{\phi}(x)$ ν_0-a.s. for some sequence of ϕ_n that approach the supremum in eq. (3.10). Furthermore $\bar{\phi}$ is concave.

Proof. First note that there exists optimal (V, X), by theorem 1.

We begin with the forward direction. By theorem 2 (V, X) is optimal iff there exists a sequence of solutions ϕ_n to HJB that is maximizing in eq. (3.19) such that

$$(3.15) \qquad \mathbb{E}\left[\int_0^T L(t, X, \beta_X(t, X))\, dt + \langle X(0), V \rangle\right] = \lim_{n\to\infty} \mathbb{E}\left[\phi_n^T(X(T)) + \widetilde{\phi}_n^0(V)\right]$$

We add 0 to the right hand side to get

$$(3.16) \qquad \lim_{n\to\infty} \mathbb{E}\left[\underbrace{\phi_n^T(X(T)) - \phi_n^0(X(0))}_{(a)} + \underbrace{\phi_n^0(X(0)) - \widetilde{\widetilde{\phi}}_n^0(X(0))}_{(b)} + \underbrace{\widetilde{\widetilde{\phi}}_n^0(X(0)) + \widetilde{\phi}_n^0(V)}_{(c)}\right].$$

where $\widetilde{\widetilde{f}}$ is the concave double dual of f, or its concave hull. Applying Ito's formula to the first two terms, with the knowledge that they satisfy (HJB) we get

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

However, by the definition of the Hamiltonian, we have $\langle v, b \rangle - H(t, x, v) \leq L(t, x, b)$, this along with results from convex analysis provides us with three inequalities on eq. (3.16)

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

(b) $\phi_n^0(X(0)) - \widetilde{\widetilde{\phi}}_n^0(X(0)) \leq 0$

(c) $\widetilde{\widetilde{\phi}}_n^0(X(0)) + \widetilde{\phi}_n^0(V) \leq \langle V, X(0) \rangle.$

These inequalities demonstrate that eq. (3.16) breaks our problem into a stochastic and a Wasserstein transport problem (in the flavour of theorem 1), along with a correction term to account for ϕ_n^0 not being concave. Adding eq. (3.15) to the mix, allows us to obtain L^1 convergence in the (a,b,c) inequalities, hence a.s. convergence of a subsequence ϕ_{n_k}.

By properties on convex analysis, convergence in (b,c) is equivalent to ϕ_n^0 converging ν_0-a.s. to a concave function $\overline{\phi}$ such that $x \mapsto \nabla\overline{\phi}$ is the optimal transport plan for $\underline{W}(\nu_0, \mu_0)$ [3].

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

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

Since ϕ_n^t are deterministic functions, this demonstrates that X_t is a Markov process with drift β_X determined by the inverse transform: $\beta_X(t, X) = \nabla_p H(t, X, \nabla\phi(t, X))$, i.e., eq. (3.13). $\square$

Remark. It is not possible to conclude from the above work that $\overline{\phi}(x) = \phi(0, x)$ since $\overline{\phi}$ is defined on a $\mathcal{P}_X$-null set. To do so, we would require a regularity result on the optimal ϕ in time.

3.2. **Maximizing Ballistic Cost.** We now turn our focus to the related stochastic cost:

$$(3.17)\ \ \overline{B}(\nu_0, \mu_T) := \sup\left\{\mathbb{E}\left[\langle V, X(T) \rangle\right] - \mathbb{E}\left[\int_0^T L(t, X, \beta_X(t, X))\, dt\right]\middle| X \in \mathcal{A}_{\nu_0}, V \sim \mu_T\right\},$$

which we term the *maximizing ballistic cost.*

Remark. The case where μ_T is given by a dirac measure δ_u proves suggestive. Here the maximizing ballistic cost may be interpreted literally in terms of the HJB equation by proposition 3—with $f : x \mapsto \langle u, x \rangle$. Thus, in this particular case we can recover

$$\overline{B}(\nu_0, \delta_u) = \int \phi^f(0, x) \, d\nu_0(x),$$

without the aid of any duality theorem. Notably,

$$f^*(v) = \sup_z \{\langle v, z \rangle - \langle u, z \rangle\} = \begin{cases} 0 & v = u \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 3 further gives us the minimizing process, described by the SDE

$$X(t) = X(0) + \int_0^t \nabla_p H(s, X, \nabla \phi(s, X)) \, ds + W_t.$$

In the case where $L(t, x, v) = \frac{|v|^2}{2} + \ell(x, t)$, the maximizing ballistic cost may be considered as finding the optimal stochastic process with the final drift constrained a.s. to be $\beta_X(T, X) = u$. In this section we will show that these results generalize to all measures in $\mathcal{P}_1(M^*)$.

We begin by obtaining a similar interpolation result:

Theorem 3 (Interpolation of $\overline{B}$)**.** *If the Lagrangian matches assumptions (A), then*

(3.18) $$\overline{B}(\nu_0, \mu_T) = \sup\{\overline{W}(\nu, \mu_T) - C_L(\nu_0, \nu)\},$$

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

Proof. We will proceed as in our proof of the first inequality. First, consider a sequence of rvs $(X_n(\cdot), V_n)$ approximating $\overline{B}(\nu_0, \mu_T)$. Then

$$\mathbb{E}\left[\langle V_n, X_n(T) \rangle\right] - \mathbb{E}\left[\int_0^T L(t, X_n, \beta_{X_n}(t, X_n)) \, dt\right] < \overline{B}(\nu_0, \mu_T) + \frac{1}{n}$$

but if $X_n(T) \sim \mu_n$,

$$\mathbb{E}\left[\langle V_n, X_n(T) \rangle\right] \leq \overline{W}(\mu_n, \mu_T)$$

$$\mathbb{E}\left[\int_0^T L(t, X_n, \beta_{X_n}(t, X_n)) \, dt\right] \geq C_L(\nu_0, \mu_n),$$

showing that $\overline{B}(\nu_0, \mu_T)$ is less than its interpolations. To get the reverse inequality, take μ_n to be a sequence approaching the supremum. Then there exists a random variable $X_n(\cdot) \in \mathcal{A}_{\nu_0}^{\mu_n}$ such that

$$\mathbb{E}\left[\int_0^T L(t, X_n, \beta_{X_n}(t, X_n)) \, dt\right] < C(\nu_0, \mu_n) + \frac{1}{n}.$$

Since $X_n(T) \sim \mu_n$, we may also define $V_n(\omega) \sim \gamma_{X_n(0)(\omega)}^n$ to be distributed according to the disintegration of $d\gamma_x^n(y) \otimes d\mu_T(x) = d\gamma_n$ such that

$$\iint \langle y, x \rangle \, d\gamma_n(x, y) > \overline{W}(\mu_n, \mu_T) - \frac{1}{n}.$$

We may further assume the minimizing sequence μ_n is tight in the case where $\nu_0 \in \mathcal{P}_1(M^*)$ and $\mu_T \in \mathcal{P}_1(M)$ since by proposition 1,

$$\overline{B}(\nu_0, \mu_T) - 1 < \overline{W}(\mu_n, \mu_T) - C_L(\nu_0, \mu_n) < M \int |v| \, d\mu_n - C_L(\nu_0, \mu_n).$$

Again, Prokhorov's theorem concludes the proof. $\qquad\square$

Theorem 4 (Dual of $\overline{B}$). *Assume the Lagrangian satisfies the assumptions (A). If $\nu_0 \in \mathcal{P}_1(M)$, μ_T has compact support, and $\overline{B}(\nu_0, \mu_T) < \infty$, then*

$$(3.19) \qquad \overline{B}(\nu_0, \mu_T) = \inf_{g \in cC_{db}^\infty} \left\{ \int_{M^*} g^* \, d\mu_T + \int_M \phi^g \, d\nu_0 \right\},$$

where ϕ solves the Hamilton-Jacobi-Bellman equation

$$(\text{HJB2}) \qquad \frac{\partial \phi}{\partial t} + \frac{1}{2} \Delta \phi - H(t, x, \nabla \phi) = 0 \qquad\qquad \phi(x, T) = g(x)$$

Lemma 4.1. *Let $\mu \in \mathcal{P}_1(M^*)$ have compact support. If we define $\overline{W}_\mu(\nu) : \nu \mapsto \overline{W}(\mu, \nu)$, then for all $f \in Lip(M^*)$*

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

Proof. This lemma follows from applying lemma 2.1 to $\underline{W}(\hat{\mu}, \nu)$ (recall $\hat{\mu}(A) = \mu(-A)$), and noting that

$$-\int \widetilde{f}(v) \, d\hat{\mu}(v) = -\int \widetilde{f}(-v) \, d\mu(v) = \int (-f)^*(v) \, d\mu(v).$$

$\qquad\square$

Proof of theorem 4. Define $\overline{B}_{\nu_0} : \mu_T \mapsto \overline{B}(\nu_0, \mu_T)$, and note that this is a concave function (which can be seen by applying lemma 1.1 to the negative of eq. (3.18)). Furthermore it is upper semi-continuous on $\mathcal{P}_1(M^*)$ by the same reasoning as corollary 1.1a). Thus we have

$$(3.20) \qquad \overline{B}_{\nu_0}(\mu_T) = -(-\overline{B}_{\nu_0})^{**}(\mu_T) = \inf_{f \in \text{Lip}(M^*)} \left\{ -\int_{M^*} f \, d\mu_T + (-\overline{B}_{\nu_0})^*(f) \right\}.$$

Investigating the dual, we find

$$(-\overline{B}_{\nu_0})^*(f) = \sup_{\mu_T \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} f \, d\mu_T + \overline{B}_{\nu_0}(\mu_T) \right\}$$

$$= \sup_{\substack{\mu_T \in \mathcal{P}_1(M^*) \\ \nu_T \in \mathcal{P}_1(M^*)}} \left\{ \int_{M^*} f \, d\mu_T + \overline{W}(\nu_T, \mu_T) - C_L(\nu_0, \nu_T) \right\}$$

$$(3.21) \qquad = \sup_{\nu_T \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} (-f)^* \, d\nu_T - C_L(\nu_0, \nu_T) \right\}.$$

Note that in the case where $(-f)^* \in C_{\text{db}}^\infty$, this is simply $\int_M \phi^{(-f)^*} \, d\nu_0$, giving us

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

In either case, we can restrict our f to be concave by noting that if we fix $g = (-f)^*$, then the set of corresponding $\{-f | (-f)^* = g\}$ is minimized by the convex function $g^* = (-f)^{**} \leq -f$ [4, Proposition 4.1]. Thus it suffices to consider f convex.

We now show that it is sufficient to consider this infimum over $g \in cC_{\mathrm{db}}^\infty$ by a similar mollification argument to theorem 2 (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:

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

Mikami [10, Proof of Theorem 2.1] further shows that

$$(3.21) = C_{\nu_0}^*(g_\epsilon) \leq \frac{C_{\nu_0 * \eta_\epsilon}^*((1 + \Delta L(0, \epsilon))g)}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)}.$$

Putting these together we get

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

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

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

where $d\mu_{L,\epsilon}(v) := d\mu_T((1 + \Delta L(0, \epsilon)) v)$. Taking $\epsilon \searrow 0$ dominates the right side by $\overline{B}(\nu_0, \mu_T)$ (where we exploit the upper semi-continuity of $\overline{B}$), completing the reverse inequality. $\square$

Corollary 4.1 (Optimal Processes for $\overline{B}$). *Suppose the assumptions on theorem 4 are satisfied, with $d\mu_0 \ll d\lambda$. Then, $(V, X(t))$ is an optimal process iff it is a solution to the SDE*

$$(3.22) \qquad\qquad\qquad dX = \nabla_p H(t, x, \nabla \phi(t, X)) \, dt + dW_t$$

$$(3.23) \qquad\qquad\qquad X(T) = \nabla \bar{\phi}(V, T)$$

where $\lim_{n \to \infty} \phi_n(T, x) \to \bar{\phi}(x)$ ν_T-a.s. and $\lim_{n \to \infty} \phi_n(t, x) = \phi(t, x)$ $\mathcal{P}_X$-a.s. for some sequence of ϕ_n that is minimizing the dual problem.

Proof. If (V, X) is optimal, then theorem 4 means there exists a sequence of solutions ϕ_n to HJB with convex final condition such that

$$(3.24) \qquad \mathbb{E}\left[\langle X(T), V \rangle - \int_0^T L(t, X, \beta_X(t, X)) \, dt \right] = \lim_{n \to \infty} \mathbb{E}\left[\left(\phi_n^T \right)^*(V) + \phi_n^0(X(0)) \right]$$

We add 0 to the right hand side to get

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

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

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

However, by the definition of the Hamiltonian, we have $-\langle v, b \rangle + H(t, x, v) \geq -L(t, x, b)$, similarly $\phi^*(v) + \phi(x) \geq \langle v, x \rangle$. These inequalities allow us to separate the limit in eq. (3.24) into two requirements: (a) $\langle \beta_X, \nabla \phi_n^t(X(t)) \rangle - H(t, X, \nabla \phi_n^t(X(t)))$ must converge to $L(t, X, \beta_X(t, X))$ and (b) $\phi_n^T(X(T)) + \left(\phi_n^T \right)^*(V)$ must converge to $\langle X(T), V \rangle$ in L^1 hence a subsequence ϕ_{n_k} exists such that this convergence is a.e.

The journey from (a) to eq. (3.22) is as in corollary 2.1. The only difference from the earlier corollary is that we know that ϕ_n must converge to a convex function, so (b) implies $V = \nabla \lim_{n\to\infty} \phi_n(X(T))$.

$\square$

Remark. Like for the minimizing cost, it is impossible to conclude $\phi(T, x) = \bar{\phi}(x)$ without a regularity result. **If** one could make this connection, then one could formulate eq. (3.23) as $V = \nabla \phi(T, X(T))$, allowing us to interpret the measure μ_T as a condition on our final momentum distribution—i.e., our final drift $\beta(T)$ would be constrained to be distributed according to the measure ρ where $\mu_T = (\nabla L(t, x, \cdot))_{\#} \rho$. In the particular case where $L(t, x, v) = \frac{|v|^2}{2} + \ell(x, t)$, this would impose a Neumann type boundary condition on our stochastic process: $\beta(T) \sim \mu_T$.

References

[1] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *ArXiv e-prints*, August 2012.

[2] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *Journal of the European Mathematical Society*, 9(1):85–121, 2007.

[3] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[4] I. Ekeland and R. Téman. *Convex Analysis and Variational Problems*, Convex Functions. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[5] W.H. Fleming and H.M. Soner. *Controlled Markov Processes and Viscosity Solutions*. Applications of Mathematics. Springer-Verlag, 1993.

[6] N. Ghoussoub. Optimal Ballistic Transport and Hopf-Lax Formulae on Wasserstein Space. *ArXiv e-prints*, May 2017.

[7] L. Kantorovitch. On the translocation of masses. *Management Science*, 5(1):1–4, 1958.

[8] R.S. Liptser and A.N. Shiryaev. *Statistics of Random Processes: I. General Theory*. Applications of mathematics : stochastic modelling and applied Probability. Springer, 2001.

[9] T. Mikami. Optimal control for absolutely continuous stochastic processes and the mass transportation problem. *Electron. Commun. Probab.*, 7:199–213, 2002.

[10] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Processes and their Applications*, 2006.

[11] G. Monge. *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.

[12] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.

[13] N. Weaver. *Lipschitz Algebras*. World Scientific, 1999.

Mathematics Department, University of British Columbia, Vancouver, BC
E-mail address: `arsbar@math.ubc.ca, nassif@math.ubc.ca`