Dynamic and Stochastic Propagation of Brenier's Optimal Mass Transport

by

ALISTAIR BARTON

B.Sc., McGill University, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2018

Abstract

I present analysis of how the mass transports that optimize the inner product cost—considered by Y. Brenier—propagate in time along a given Lagrangian in both deterministic and stochastic settings. While for the minimizing transports one may easily obtain Hopf-Lax formulas on Wasserstein space by inf-convolution, this is not the case for the maximizing transports, which are sup-inf problems. In this case, we assume that the Lagrangian is jointly convex on phase space, which allow us to use Bolza-type duality, a well known phenomenon in the deterministic case but, as far as I know, novel in the stochastic case. Hopf-Lax formulas help relate optimal ballistic transports to those associated with the dynamic fixed-end transports studied by Bernard-Buffoni and Fathi-Figalli in the deterministic case, and by Mikami-Thieullen in the stochastic setting.

Lay Summary

My work is in the mathematical field of optimal transportation, which examines how to efficiently change one distribution into another, given the cost of transporting one unit of material between different locations in the distributions—a common motivation is efficiently transport mining material to construction sites. I analyze the combination of two such transportations in sequence, where the second transportation uses a path-dependent cost. This is a natural way to consider the evolution of the first transportation cost function (seen as a function the initial distribution) along the paths of least resistance given by the second cost. I also consider the case where the second transportation is stochastic (its paths diffuse in some sense), as well as a cost maximization version of each case.

My main contribution is to demonstrate various equivalent reformulations of this problem, including a method of converting certain stochastic minimizing problems to maximizing problems.

Preface

My thesis research is inspired by a paper authored by my supervising professor, Professor N. Ghoussoub that considered the deterministic case [12]. He proposed that I study the stochastic case, and directed me to relevant papers. The research program was designed collaboratively with frequent meetings to discuss possible strategies, techniques, and directions to focus.

While the deterministic case—Sections 2.1, 3.1, 4.1—is discussed in a paper on the ArXiv authored by my supervisor (mentioned above), we also have a joint-authored paper discussing some of the stochastic content—Sections 2.2, and 4.2—of this thesis uploaded to the ArXiv [2]. The two papers are combined in a recent submission to the European Journal of Applied Mathematics for publication later this year (2018) or early next year (2019).

The stochastic sections are entirely my work, excepting some editing and revision by my supervisor. As mentioned, the deterministic results contained in this thesis were previously obtained by my supervisor, however in Section 2.1 most results are reframed using new techniques from the stochastic section. On the other hand, Sections 3.1 and 4.1 contain primarily my supervisor's work.

No ethic approval was required for this research.

Table of Contents

Chapter 1

Introduction

1.1 Results

The problem of optimal transportation, as stated by Kantorovich consists of minimizing the transportation cost between two probability measures ν_0, ν_1 on Polish spaces X_0, X_1 for a given cost function $c : X_0 \times X_1 \to \mathbb{R}$:

$$\inf\{\int_{X_0 \times X_1} c(x,y)\, d\pi(x,y); \pi \in \mathcal{K}(\nu_0, \nu_1)\} \tag{1.1}$$

where $\mathcal{K}(\nu_0, \nu_1)$ is the set of *transport plans*: probability measures on $X_0 \times X_1$ whose marginal on X_0 (resp. X_1) is ν_0 (resp. ν_1). Kantorovich also provides us the so-called *dual formulation* of this cost

$$\sup\{\int_{X_1} \phi_1(y)\, d\nu_1(y) - \int_{X_0} \phi_0(x)\, d\nu_0(x); (\phi_0, \phi_1) \in \mathcal{K}(c)\}. \tag{1.2}$$

Here $\mathcal{K}(c)$ is the set of functions (ϕ_0, ϕ_1) with $\phi_0 \in L^1(X_0, \nu_0)$ and $\phi_1 \in L^1(X_1, \nu_1)$ satisfying the inequality $\phi_1(y) - \phi_0(x) \le c(x,y)$. By maximizing (resp. minimizing) ϕ_1 (resp. ϕ_0) for a given ϕ_0 (resp. ϕ), we may assume that the functions satisfy the below relation

$$\phi_1(y) = \inf_{x \in X_0} \{c(x,y) + \phi_0(x)\} \qquad \phi_0(y) = \sup_{y \in X_1} \{\phi_1(y) - c(x,y)\}. \tag{1.3}$$

Since Monge first considered the problem using distance based costs of the form $c(x,y) = |x-y|$ ([14], [18], [8], [19], [20]), where $X_0 = X_1$ is a normed space, the structure of this problem has been studied under several cost functions. We review a couple cost functions that are relevant for this thesis. Brenier [6] uncovered many connections with convex analysis and transportation when the cost is quadratic $c(x,y) = |x-y|^2$, demonstrating that the optimal transportation plan is given by the gradient of a convex function: $\pi^o = (\mathrm{Id} \times \nabla\phi)_{\#}\nu_0$ (where $f_{\#}\nu$ indicates the push-forward of ν by the function f, ie. the measure μ defined by $\mu(A) = \nu(f^{-1}(A))$ for all measurable sets A). Notably, this is the same optimal plan for the cost function

$c(x,y) = -\langle x, y\rangle = |x-y|^2 - |x|^2 - |y|^2$. 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 [10].

Bernard and Buffoni [4] considered dynamic costs on a manifold M of the form

$$c_T(x,y) := \inf\{\int_0^T L(t,\gamma(t),\dot\gamma(t))\,dt; \gamma \in C^1([0,T],M), \gamma(0)=x, \gamma(T)=y\} \tag{1.4}$$

where T is a fixed time and $L : TM \to \mathbb{R}\cup\{+\infty\}$ is a Lagrangian convex in the second variable of the tangent bundle. This formulation encompasses costs functions of the form $c(x,y) = f(|x-y|)$ for f convex, which correspond to $L(t,x,p) = Tf(|p|/T)$. Fathi and Figalli [9] eventually dealt with the case where M is a non-compact Finsler manifold.

In this thesis I will consider "ballistic cost functions" $b_T : M^* \times M \to \mathbb{R}\cup\{+\infty\}$ (where M^* is dual to the Banach space $M = \mathbb{R}^d$) derived by propogating the inner product cost by a Lagrangian L:

$$b_T(v,y) := \inf\{\langle v, \gamma(0)\rangle + \int_0^T L(t,\gamma(t),\dot\gamma(t))\,dt; \gamma \in C^1([0,T],M), \gamma(T)=y\} \tag{1.5}$$

This leads to the minimzing and maximizing transportation problems that will be discussed in Section 2.1 and 4.1 respectively:

$$\underline{B}_T(\mu_0,\nu_T) := \inf\{\int_{M^*\times M} b_T(v,y)\,d\pi(v,y); \pi \in \mathcal{K}(\mu_0,\nu_T)\} \tag{1.6}$$

$$\overline{B}_T(\mu_0,\nu_T) := \sup\{\int_{M^*\times M} b_T(v,y)\,d\pi(v,y); \pi \in \mathcal{K}(\mu_0,\nu_T)\}. \tag{1.7}$$

The latter cost is a sup-inf problem, however it can be made into a sup-sup problem using Bolza duality theory as reviewed in section 3.1. The ballistic cost may be considered a propagation of the Wasserstein cost considered by Brenier [6], as when $T = 0$ we recover the Wasserstein cost

$$\underline{W}(\mu_0,\nu_0) := \inf\{\int_{M^*\times M} \langle v, y\rangle\,d\pi(v,y); \pi \in \mathcal{K}(\mu_0,\nu_0)\} \tag{1.8}$$

$$\overline{W}(\mu_0,\nu_0) := \sup\{\int_{M^*\times M} \langle v, y\rangle\,d\pi(v,y); \pi \in \mathcal{K}(\mu_0,\nu_0)\}. \tag{1.9}$$

We will also consider a stochastic propagation of these cost functions. To do so, we define a stochastic version of the dynamic transportation problem between two random variables $Y, Z : \Omega \to M$ to be

$$c_T^s(Y,Z) := \inf\{\mathbb{E}\left[\int_0^T L(t,X_t,\beta_t)\,dt\right]; (X,\beta) \in \mathcal{A}, X_0 = Y, X_T = Z\} \tag{1.10}$$

where $\mathcal{A}$ denotes the set of stochastic processes X with previsible drift β such that $X_t = X_0 + \int_0^t \beta_s\,ds + W_t$ where W_t is $\sigma(X_s; 0 \le s \le t)$-Brownian motion. The stochastic transportation between two measures can then be given by

$$\begin{aligned} C_T^s(\nu_0,\nu_T) :=& \inf\{c_T^s(Y,Z); Y \sim \nu_0, Z \sim \nu_T\} \\ =& \inf\{\mathbb{E}\left[\int_0^T L(t,X_t,\beta_t)\,dt\right]; (X,\beta) \in \mathcal{A}, X_0 \sim \nu_0, X_T \sim \nu_T\} \end{aligned} \tag{1.11}$$

as considered by Mikami and Theiullen [13], where $X \sim \nu$ indicates that the random variable X has law ν.

Similarly, we define the ballistic transportation between two random variables as

$$b_T^s(V,Y) := \inf\{\mathbb{E}\left[\langle V, X_0\rangle + \int_0^T L(t,X_t,\beta_t)\,dt\right]; (X,\beta) \in \mathcal{A}, X_T = Y\} \tag{1.12}$$

allowing us to define the stochastic version of the ballistic cost as

$$\begin{aligned} \underline{B}_T^s(\mu_0,\nu_T) :=& \inf\{b_T^s(V,Z); V \sim \mu_0, Z \sim \nu_T\} \\ =& \inf\{\mathbb{E}\left[\langle V, X_0\rangle + \int_0^T L(t,X_t,\beta_t)\,dt\right]; (X,\beta) \in \mathcal{A}, V \sim \mu_0, X_T \sim \nu_T\} \end{aligned} \tag{1.13}$$

and

$$\overline{B}_T^s(\mu_0,\nu_T) := \sup\{b_T^s(V,Z); Y \sim \mu_0, Z \sim \nu_T\} \tag{1.14}$$

which will be analyzed in section 2.2 and 4.2 respectively. Once more the maximizing cost is a sup-inf problem that can be simplified using a stochastic analog of Bolza duality, which is shown in section 3.2.

1.2 Main Results

The chief results of the analysis are duality formulae that allow us to reformulate the ballistic costs in a manner similar to eq. 1.2. We will show that the role of $\mathcal{K}(c)$ will be played by solutions of the forward Hamilton-Jacobi equation

$$\left\{ \begin{array}{rcl} \partial_t \phi + H(t,x,\nabla_x \phi) & = & 0 \text{ on } [0,T]\times M, \\ \phi(0,x) & = & f(x), \end{array} \right. \tag{1.15}$$

and the backward Hamilton-Jacobi equation.

$$\left\{ \begin{array}{rcl} \partial_t \phi + H(t,x,\nabla_x \phi) & = & 0 \text{ on } [0,T]\times M, \\ \phi(T,x) & = & f(x), \end{array} \right. \tag{1.16}$$

where the Hamiltonian on $[0,T]\times M \times M$ is defined by

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

In particular, we concern ourselves with the variational solutions to the above equations:

$$\Phi^t_{f,+}(x) := \Phi_{f,+}(t,x) = \inf\Big\{f(\gamma(0)) + \int_0^t L(s,\gamma(s),\dot\gamma(s))\,ds; \gamma\in C^1([0,T),M);\gamma(t)=x\Big\}, \tag{1.17}$$

and

$$\Phi^t_{f,-}(x) := \Phi_{f,-}(t,x) = \sup\Big\{f(\gamma(T)) - \int_t^T L(s,\gamma(s),\dot\gamma(s))\,ds; \gamma\in C^1([0,T),M);\gamma(t)=x\Big\} \tag{1.18}$$

respectively. We will show that we can describe $\underline{B}_T(\mu_0,\nu_T)$ by the forward and backward duality formulae

$$\begin{aligned} \underline{B}_T(\mu_0,\nu_T) &= \sup\{\int_M \Phi_{f_*,+}(T,x)\,d\nu_T(x) + \int_{M^*} f(v)\,d\mu_0(v);\ f \text{ concave in Lip(M}^*)\} \\ &= \sup\{\int_M g(x)\,d\nu_T(x) + \int_{M^*} (\Phi_{g,-})_*(0,v)\,d\mu_0(v); g\in \mathrm{Lip}(M)\} \end{aligned} \tag{1.19}$$

where h_* is the concave legendre transform of h:

$$h_*(v) := \inf_{x\in M}\{\langle v,x\rangle - h(x)\}. \tag{1.20}$$

As to the question of attainment, we use a result by Fathi-Figalli [9] to show that if L is a Tonelli Lagrangian, and if μ_0 is absolutely continuous with respect to Lebesgue measure, then there exists a probability measure π_0 on $M^* \times M$, and a concave function $k : M \to \mathbb{R}$ such that

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, x) d\pi_0, \tag{1.21}$$

and π_0 is supported on the possibly set-valued map $v \to \pi^* \phi_T^H(\nabla k_*(v), v)$, with $\pi^* : M \times M^* \to M$ being the canonical projection, and $(x, v) \to \phi_t^H(x, v)$ is the corresponding Hamiltonian flow.
These results rely on interpolation formulation of $\underline{B}_T(\mu_0, \nu_T)$:

$$\underline{B}_T(\mu_0, \nu_T) = \inf_{\nu \in \mathcal{P}(M)} \{\underline{W}(\mu_0, \nu) + C_T(\nu, \nu_T)\}. \tag{1.22}$$

The interpolation formula can be seen as extensions of those by Hopf-Lax on state space to Wassertsein space. Indeed, for any (initial) function g, the associated value function $\Phi_{f,+}(t, x)$ can be written as

$$\phi_g(t, x) = \inf\{g(y) + c_t(y, x);\ y \in M\}. \tag{1.23}$$

In the case where the Lagrangian $L(x, p) = L_0(p)$ is only a function of p, and if H_0 is the associated Hamiltonian, then $c(t, y, x) = tL_0(\frac{1}{t}|x - y|)$ and (1.23) is nothing but the Hopf-Lax formula used to generate solutions for corresponding Hamilton-Jacobi equations. When g is the linear functional $g(x) = \langle v, x \rangle$, then $b(t, v, x)$ is itself a solution to the Hamilton-Jacobi equation, since

$$b(t, v, x) = \inf\{\langle v, y \rangle + c(t, y, x);\ y \in M\}. \tag{1.24}$$

In other words, (1.22) can now be seen as extensions of (1.24) to the space of probability measures, where the Wasserstein cost fills the role of the scalar product. This interpolative result is extended to the rest of the costs in their respective sections.

Stochastic Minimizing Cost The stochastic problem in Section 2.2 presents two differences. Firstly, it cannot formulated as a classical transportation problem (1.1) hence there is no Monge-Kantorovich duality, secondly the irreversibility of stochastic processes means we only have the one duality formula

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

where this time $\Psi_{g,-}$ is the solution to the backward Hamilton-Jacobi-Bellman equation (1.26).

$$\begin{cases} \partial_t \psi + \frac{1}{2}\Delta\psi + H(t,x,\nabla_x\psi) &= 0 \text{ on } [0,T]\times M, \\ \psi(T,x) &= g(x), \end{cases} \tag{1.26}$$

whose formal variational solutions are given by the formula:

$$\Psi_{g,-}(t,x) = \sup_{(X,\beta)\in\mathcal{A}} \left\{ \mathbb{E}\left[g(X_T) - \int_t^T L(s,X_s,\beta_s)\,ds \middle| X_t = x \right] \right\}. \tag{1.27}$$

Bolza Duality In order to deal with the maximization problems $\overline{B}_T(\mu_0,\nu_T)$ and $\overline{B}^s_T(\mu_0,\nu_T)$, we need to use Bolza-type duality discussed in chapter 3 to convert the sup-inf problem to a concave maximization problem. For that, we shall assume that the Lagrangian L is jointly convex in both variables. We then consider the dual Lagrangian $\tilde{L}$ defined on $M^*\times M^*$ by

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

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

$$\tilde{c}_T(u,v) := \inf\{\int_0^T \tilde{L}(t,\gamma(t),\dot{\gamma}(t))\,dt; \gamma\in C^1([0,T),M^*); \gamma(0)=u, \gamma(T)=v\}, \tag{1.28}$$

and its associated transport

$$\tilde{C}_T(\mu_0,\mu_T) := \inf\{\int_{M^*\times M^*} \tilde{c}_T(x,y)\,d\pi;\ \pi\in\mathcal{K}(\mu_0,\mu_T)\}. \tag{1.29}$$

We then recall the deterministic Bolza duality, and establish a new stochastic Bolza duality, allowing us to write the maximizing costs as

$$\overline{B}_T(\mu_0,\nu_T) = \sup\{\int_0^T \tilde{b}_T(v,x)\,d\pi(v,x); \pi\in\mathcal{P}(\mu_0,\nu_T)\} \tag{1.30}$$

$$\overline{B}^s_T(\mu_0,\nu_T) = \sup\{\mathbb{E}\left[\langle V_T,X\rangle - \int_0^T \tilde{L}(t,V_t,\beta_t)\,dt\right]; (V,\beta)\in\mathcal{A}, V_0\sim\mu_0, X\sim\nu_T\} \tag{1.31}$$

respectively, where $\tilde{b}_T(v,x) := \sup\{\langle u,x\rangle - \tilde{c}_T(v,u)\}$.

Maximizing Deterministic Cost We use this Bolza theory to establish the following duality result for $\overline{B}_T(\mu_0, \nu_T)$ in section 4.1:

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M g(x)\, d\nu_T(x) + \int_{M^*} (\tilde{\Phi}^0_{g,-})^*(v)\, d\mu_0(v);\ g \text{ convex on } M\right\}, \tag{1.32}$$

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

$$\begin{cases} \partial_t \phi - H(t, \nabla_v \phi, v) &= 0 \text{ on } [0,T] \times M^*, \\ \phi(T, v) &= k(v), \end{cases} \tag{1.33}$$

whose variational solution is given by

$$\tilde{\Phi}_{k,-}(t, v) = \sup\left\{k(\gamma(T)) - \int_0^t \tilde{L}(s, \gamma(s), \dot{\gamma}(s))\, ds; \gamma \in C^1([0,T), M^*); \gamma(0) = v\right\}. \tag{1.34}$$

Maximizing Stochastic Cost We follow this by establishing the duality formula for the stochastic version:

$$\overline{B}^s(\nu_0, \mu_T) = \inf\left\{\int_{M^*} g(x)\, d\nu_T + \int_M (\tilde{\Psi}^0_{g^*,-})^*(v)\, d\mu_0;\ g \text{ in } C^\infty_{\text{db}}(M^*)\right\}, \tag{1.35}$$

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

$$\begin{cases} \partial_t \psi + \frac{1}{2}\Delta\psi - H(t, \nabla_v \psi, v) &= 0 \text{ on } [0,T] \times M^*, \\ \psi(T, v) &= k(v), \end{cases} \tag{1.36}$$

whose formal variational solutions are given by the formula:

$$\tilde{\Psi}_{k,-}(t, v) = \sup_{X \in \mathcal{A}} \left\{\mathbb{E}\left[k(X(T)) - \int_t^T \tilde{L}(s, X(s), \beta_X(s, X))\, ds \middle| X(t) = v\right]\right\}. \tag{1.37}$$

1.3 Notation

We will use the standard notation of $(\Omega, \mathbb{P})$ to denote a probability space Ω equipped with measure $\mathbb{P}$. We will for convenience denote the probability of an event such as $\{X < 3\} := \{\omega \in \Omega; X(\omega) < 3\}$ by $\mathbb{P}(X < 3) := \mathbb{P}(\{X(\omega) < 3\})$.

I will use $\mathcal{P}(M)$ to indicate the space of probability measures on a space M situated within the larger space of finite measures $\mathcal{M}(M)$. When $M = \mathbb{R}^d$, I will use $\mathcal{P}_1(M)$ to be the subset of probability measures with finite first moment(ie. satisfying $\int_M |x| \, d\nu(x) = \mathbb{E}_{X\sim\nu}\left[|X|^\alpha\right] < \infty$ where $\mathbb{E}_{X\sim\nu}\left[\cdot\right]$ denotes expectation with $X \sim \nu$), which is likewise situated within $\mathcal{M}_1(M) := \{\nu; \int_M 1 + |x| \, d\nu(x) < \infty\}$.

The dual space of $\mathcal{M}_1(M)$ can be identified with the space of uniformly Lipschitz functions $\mathrm{Lip}(M)$ under the inner product $\langle f, \nu\rangle = \int f(x)\, d\nu(x)$. However, we will often be forced to work with the smooth subset of Lipschitz functions which I will denote $C^\infty_{\mathrm{db}}(M) := C^\infty(M) \cap \mathrm{Lip}(M)$ ('db' stands for derivative bounded functions).

When operating on a vector space V, I will use ν to represent measures on the primal space (ie. $\nu \in \mathcal{P}(V)$) and μ to represent measures on the dual space (ie. $\mu \in \mathcal{P}(V^*)$) for convenience.

Chapter 2

Minimizing Ballistic Costs

2.1 Deterministic Minimizing Cost

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

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

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

This coercivity condition is inherited by the transportation cost $C_T(\nu,\cdot)$, allowing the attainment of a minimizer in the following Theorem.

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

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

In the case where ν_T has finite first moment, the infimum is attained at some probability measure ν_0 on M, and the initial Kantorovich potential for $C_T(\nu_0,\nu_T)$ is concave.

Proof: To prove the formula it suffices to note that

$$\begin{aligned}
&\inf\{\underline{W}(\mu_0,\nu) + C_T(\nu,\nu_T)\} \\
&\quad = \inf_{\nu\in\mathcal{P}(M)}\{\int \langle v,x\rangle\, d\pi_W(v,x) + \int c(x,y)\, d\pi_C(x,y); \pi_W \in \mathcal{K}(\mu_0,\nu), \pi_C \in \mathcal{K}(\nu,\nu_T)\} \\
&\quad = \inf\{\int \langle v,x\rangle + c(x,y)\, d\pi(v,x,y); \pi_1 = \mu_0, \pi_3 = \nu_T\} \geq \underline{B}(\mu_0,\nu_T).
\end{aligned}$$

For the reverse inequality, we may use a selection theorem to find a measurable function $y_\epsilon : M^* \times M \to M$ that satisfies $\langle v, y_\epsilon(v,x)\rangle + c(y_\epsilon(v,x),x) - \epsilon <$

$b_T(v,x)$ (e.g. [21]). Fixing $\pi \in \mathcal{K}(\mu_0,\nu_T)$ and letting $\pi_\epsilon := (\mathrm{Id}\times\mathrm{Id}\times y_\epsilon)_\#(\pi) \in \mathcal{P}(M^*\times M\times M)$

$$\underline{B}(\mu_0,\nu_T) = \int b_T(v,x)\,d\pi(v,x) \geq \int \langle v,y\rangle + c(y,x)\,d\pi_\epsilon(v,x,y) - \epsilon.$$

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

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

is tight. Indeed, from $(A0)$, there exists a constant K such that $c_T(x,y) > A\left|\frac{x-y}{T}\right|^\delta - K$. Hence for any optimal transport plan $\pi \in \mathcal{K}(\nu,\nu_T)$

$$C(\nu,\nu_T) \geq \frac{A}{T^\delta}\int_{M\times M} ||x|-|y||^\delta\ d\pi(x,y) - K. \tag{2.4}$$

We transfer the problem to $\mathbb{R}_+$ by using the push-forward $\bar{\pi} := (|\cdot|\times|\cdot|)_\#\pi$ to obtain,

$$\int_{M^2} ||x|-|y||^\delta\ d\pi(x,y) = \int_{\mathbb{R}_+^2} |x-y|^\delta\ d\bar{\pi}(x,y). \tag{2.5}$$

We can obtain a lower estimate for this by minimizing over transportation measures sharing $\bar{\pi}$'s marginals (i.e., $\gamma \in \mathcal{K}(\bar{\nu},\bar{\nu}_T)$ where $\bar{\nu} := |\cdot|_\#\,\nu$ and $\bar{\nu}_T := |\cdot|_\#\,\nu_T$). This is a well known optimal transport problem, whose optimal plan given by the monotone Hoeffding-Frechet mapping $x \mapsto G_\nu(G_{\nu_T}^{-1}(x))$, where $G_\nu(t) := \inf\{z\in\mathbb{R} : t \geq \nu(\{x\leq z\})\}$ is the quantile function associated with the measure ν [3]. Thus the optimal plan maps each quantile in one measure to the corresponding quantile in the other. Substituting this into the integral and applying Jensen's inequality:

$$\begin{aligned}\int_{\mathbb{R}_+^2} |x-y|^\delta\ d\bar{\pi}(x,y) &\geq \int_{\mathbb{R}_+^2} |x-y|^\delta\ d\left[[G_{\bar{\nu}}\times G_{\bar{\nu}_T}]_\#\,\lambda_{[0,1]}\right](x,y)\\ &\geq \left[\int_M |x|\ d\nu(x) - b(\nu_T)\right]^\delta,\end{aligned} \tag{2.6}$$

where $b(\nu_T) := \int_M |x|\ d\nu_T$, and we assume $\nu \in \mathcal{T}_{\epsilon,R} := \{\nu;\nu(B(R,0)^c) > \epsilon\}$ for some $R > b(\nu_T)/\epsilon$. It then suffices to find R such that, for all $\nu \in \mathcal{T}_{\epsilon,R}$,

$$\frac{A}{T^\delta}\left[\left[\int_M |x|\ d\nu(x) - b(\nu_T)\right]^\delta - K\right] \geq N\int_M |x|\ d\nu(x).$$

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

$$\frac{A}{T^\delta}\left[(I_\nu(R) - b(\nu_T))^\delta - K\right] \geq N(I_\nu(R) + R(1-\epsilon))$$

for $\nu \in \mathcal{T}_{\epsilon,R}$. Using the fact that $I_\nu(R) \geq R\epsilon$ for $\nu \in \mathcal{T}_{\epsilon,R}$, we can say that this condition is satisfied if R large enough that

$$\frac{A}{T^{\delta-1}}\left[\left[(R\epsilon)^{1-\frac{1}{\delta}} - b(\nu_T)(R\epsilon)^{-\frac{1}{\delta}}\right] - \frac{K}{R\epsilon}\right] > \frac{N}{\epsilon}$$

in addition to our earlier conditions that $R > b(\nu_T)/\epsilon$.

To show the minimizer is achieved, fix ν_0, and note that by coercivity the set of probability measures ν such that

$$-b(\mu_0)\int |x|\, d\nu + C_T(\nu,\nu_T) \leq \underline{W}(\mu_0,\nu_0) + C_T(\nu_0,\nu_T) \tag{2.7}$$

is tight. Since $\underline{W}(\mu_0,\nu)+C_T(\nu,\nu_T) \leq -b(\mu_0)\int |x|\, d\nu + C_T(\nu,\nu_T)$ this implies the set of measures $\nu \in \mathcal{P}_1(M)$ such that

$$\underline{W}(\mu_0,\nu) + C_T(\nu,\nu_T) \leq \underline{W}(\mu_0,\nu_0) + C_T(\nu_0,\nu_T) \tag{2.8}$$

is also tight. Combining this result with the lower semi-continuity of $\nu \mapsto \underline{W}(\mu_0,\nu) + C_T(\nu,\nu_T)$ guarantees the existence of a minimizing measure. □

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

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

$$\underline{B}_T(\mu_0,\nu_T) = \sup\left\{\int_M g(x)\, d\nu_T(x) + \int_{M^*} (\Phi^0_{g,-})_*(v)\, d\mu_0(v);\ g \text{ in } Lip(M)\right\}. \tag{2.9}$$

2. *If the optimal interpolant ν_0 has compact support, then*

$$\underline{B}_T(\mu_0,\nu_T) = \sup\left\{\int_M \Phi^T_{f_*,+}(x)\, d\nu_T(x) + \int_{M^*} f(v)\, d\mu_0(v);\ f \text{ concave in } Lip(M^*)\right\}. \tag{2.10}$$

We shall need the following identifications of the Legendre transforms in the Banach space $\mathcal{M}_1(\mathbb{R}^n)$ of measures ν on $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n} [1+|x|]\, d\nu < \infty$, whose dual space under weak convergence is the space of Lipschitz functions $\mathrm{Lip}(\mathbb{R}^n)$.

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

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

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

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

$$C_{\nu_0}(\nu) := \begin{cases} C_T(\nu_0, \nu) & \nu \in \mathcal{P}_1(\mathbb{R}^n) \\ +\infty & \textit{otherwise.} \end{cases} \tag{2.11}$$

Then, the convex Legendre transform of C_{ν_0} is given for $f \in Lip(\mathbb{R}^n)$ by $C^*_{\nu_0}(f) = \int \phi_{f,-}(0,x)\, d\nu_0(x)$, where $\phi_{f,-}(t,x)$ is the solution to the backward Hamilton-Jacobi equation (1.16) with final condition $\phi_{f,-}(T,x) = f(x)$.

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

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

Note that this holds for all $\nu \in \mathcal{M}_1(M)$ (not merely $\nu \in \mathcal{P}_1(M)$ as the original theorem concerns itself with), since if $\nu(\mathbb{R}^n) > 1$, we can apply the operation $g \mapsto g + c$ resulting in $g_* \mapsto g_* + c$ for arbitrarily large c. Likewise we may choose arbitrarily small c if $\nu(\mathbb{R}^n) < 1$. Compact support is necessary as Brenier's theorem implies that non-compactly supported μ_0 will require $g \notin \mathrm{Lip}(M)$. We then have

$$\underline{W}^*_{\mu_0}(f) = \sup_{\nu \in \mathcal{M}_1(M)} \inf_{g \in \mathrm{Lip}(M)} \{ \int f\, d\nu - \int g\, d\nu - \int g_*\, d\mu_0 \}. \tag{2.12}$$

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

$$\underline{W}^*_{\mu_0}(f) = \inf_{g \in \mathrm{Lip}(M)} \sup_{\nu \in \mathcal{M}_1(M)} \{\int f\, d\nu - \int g\, d\nu - \int g_*\, d\mu_0\}. \tag{2.13}$$

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

The same proof applies to C_{ν_0}, since in view of the duality formula of Bernard and Buffoni [4][Proposition 21]:

$$C_{\nu_0}(\nu) = \sup_{g \in \mathrm{Lip}(M)} \{\int g\, d\nu - \int \phi_{g,-}(0,\cdot)\, d\nu_0\}. \tag{2.14}$$

As for $W_{\nu_0}(\nu)$, this equation holds for all $\nu \in \mathcal{M}_1(M)$. We may again apply the minimax theorem as the expression is linear in ν 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

$$\underline{B}_{\mu_0}(\nu) = \underline{B}^{**}_{\mu_0}(\nu) := \sup\{\int_{\mathbb{R}^n} f\, d\nu - \underline{B}^*_{\mu_0}(f);\ f \in \mathrm{Lip}(\mathbb{R}^n)\}. \tag{2.15}$$

Now use the Hopf-Lax formula established above to write

$$\begin{aligned}
\underline{B}^*_{\mu_0}(f) :=& \sup\{\int_M f\, d\nu - \underline{B}_{\mu_0}(\nu); \nu \in \mathcal{P}_1(M)\} \\
=& \sup\{\int_M f\, d\nu - \underline{W}(\mu_0, \nu') - C_T(\nu', \nu); \nu, \nu' \in \mathcal{P}_1(M)\} \\
=& \sup\{\int_M \Phi_{f,-}(T,\cdot)\, d\nu' - \underline{W}(\mu_0, \nu'); \nu' \in \mathcal{P}_1(M)\} \\
=& -\int_M [\Phi_{f,-}(T,\cdot)]_*\, d\mu_0.
\end{aligned} \tag{2.16}$$

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 $\underline{B}(\mu, \nu)$, and the concavity of f follows from the Kantorovich dual condition (1.3) and the linearity of b_T in v. □

We now turn to the structure of the optimal transport plan, for which we shall consider Tonelli Lagrangians studied in the compact case by Bernard-Buffoni [4], and by Fathi-Figalli [9] 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.

If L is a Tonelli Lagrangian, the Hamiltonian $H : M \times M^* \to \mathbb{R}$ is then C^1, and the Hamiltonian vector field X_H on $M \times M^*$ is then $X_H(x,v) = (\frac{\partial H}{\partial v}(x,v), -\frac{\partial H}{\partial x}(x,v))$, whose flow ϕ_t^H solves the associated system of ODEs

$$\begin{cases} \dot{x} = \dfrac{\partial H}{\partial v}(x,v) \\ \dot{v} = -\dfrac{\partial H}{\partial x}(x,v). \end{cases} \tag{2.17}$$

The connection between minimizers $\gamma : [0,T] \to M$ of $c_T(x,y)$ and solutions of (2.17) 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 (2.17). Note also that since L is a Tonelli Lagrangian, the Hamiltonian H is actually C^2, and the vector field X_H is C^1. It therefore defines a (partial) C^1 flow ϕ_t^H.

There is also a (partial) C^1 flow ϕ_t^L on $M \times M^*$ such that every speed curve of an L-minimizer is a part of an orbit of ϕ_t^L. This flow is called the Euler-Lagrange flow, is defined by $\phi_t^L = \mathcal{L}^{-1} \circ \phi_t^H \circ \mathcal{L}$, 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* μ_0 *is absolutely continuous with respect to Lebesgue measure. Then, there exists a concave function* $k : M \to \mathbb{R}$ *such that*

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \tag{2.18}$$

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

Proof: We start with the interpolation formula, $\underline{B}_T(\mu_0, \nu_T) = C_T(\nu_0, \nu_T) + \underline{W}(\mu_0, \nu_0)$ for some probability measure ν_0. By our duality result and Brenier's theorem, 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

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

(in fact $h(x) = \Phi^T_{k,+}(x)$). Now use a result of Fathi-Figalli [9] to write $C_T(\nu_0, \nu_T) = \int_M c_T(y, S_T y) d\nu_0(y)$, where $S_T(y) = \pi^* \phi^H_T(y, \tilde{d}_y k)$. Note that

$$\underline{B}_T(\mu_0, \nu_T) \leq \int_{M^*} b_T(v, S_T \circ \nabla k_*(v)) d\mu_0(v), \tag{2.19}$$

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

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

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

It follows that

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

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 $\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \pi^* \phi^H_T(\nabla k_*(v), v)) d\mu_0(v)$.
□

2.2 Stochastic Minimizing Problem

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 [13] and therefore use the same assumptions that he imposed on the Lagrangian, namely

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

(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 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 $X = (X_0,\beta,\sigma)$ to refer to an Itô process X_t of the form:

$$X_t = X_0 + \int_0^t \beta_s\, ds + \int_0^t \sigma_s\, dW_s. \tag{2.20}$$

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

Our main result is the stochastic counterpart to Theorem 2:

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

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

$$\underline{B}_T^s(\mu_0,\nu_T) = \inf\{\underline{W}(\mu_0,\nu) + C_T^s(\nu,\nu_T);\ \nu \in \mathcal{P}_1(M)\}. \tag{2.21}$$

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*

$$\underline{B}_T^s(\mu_0,\nu_T) = \sup\left\{\int_M f(x)\, d\nu_T(x) + \int_{M^*} (\Psi^0{}_{f,-})_*(v)\, d\mu_0(v);\ f \text{ convex and in } Lip(M)\right\}, \tag{2.22}$$

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:

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

To obtain the reverse inequality, let ν_n be a sequence of measures approximating the infimum in (2.21). Then for each ν_n, there exists a stochastic process $(Z_n,\beta_n)\in\mathcal{A}_{\nu_n}^{\nu_T}$ such that

$$\mathbb{E}\left[\int_0^T L(t,Z_n(t),\beta_n(t))\,dt\right]<C_T^s(\nu_n,\nu_T)+\tfrac{1}{n}. \tag{2.23}$$

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

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

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

$$\mathbb{E}\left[\langle U[Z_n(0)],Z_n(0)\rangle+\int_0^T L(t,Z_n(t),\beta_n(t))\,dt\right]\leq\inf\{\underline{W}(\mu_0,\nu)+C_T^s(\nu,\nu_T);\ \nu\in\mathcal{P}_1(M)\}+\frac{3}{n}. \tag{2.24}$$

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

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

We will assume $\nu\in\mathcal{T}_{\epsilon,R}:=\{\nu\in\mathcal{P}_1(M):\nu(B(R,0)^c)>\epsilon\}$ for what follows. We leave R to be defined later, but note that if we define the set $\Omega_R:=\{|X_0|>R\}$, then our assumption on ν yields $\mathbb{P}(\Omega_R)>\epsilon$. By positivity of L, this allows us to say that $\mathbb{E}\left[\int_0^T L(t,X_t,\beta_t)\,dt\right]\geq\mathbb{E}\left[\int_0^T 1_{\Omega_R}L(t,X_t,\beta_t)\,dt\right]$ (henceforth we focus only on the event Ω_R).

By (A1), we assume that there is a convex function $\underline{L}:M^*\to\mathbb{R}$ and $C>0$ such that for all $|u|>U$, $\frac{L(u)}{|u|^2}>C$. Recall that $\underline{L}(|v|)$ is a lower bound on

$L(t,x,v)$. This imposes a lower bound on the expected action of X:

$$\begin{aligned}\mathbb{E}\left[\int_0^T L(t,X_t,\beta_t)\,dt\right] &\geq \mathbb{E}\left[\int_0^T \underline{L}(|\beta_t|)\,dt\right] \overset{(J)}{\geq} \mathbb{E}\left[\underline{L}(|V|)T\right] \\ &\overset{(A1)}{>} CT\mathbb{E}\left[1_{|V|>U}\,|V|^2\right] \geq CT\left[\mathbb{E}\left[|V|^2\right]-U^2\right],\end{aligned} \tag{2.25}$$

where $V := (Y(T)-Y(0))/T$ is the time-average of the drift. Hence the expected action of the stochastic process X is bounded:

$$\begin{aligned}\mathbb{E}\left[\int_0^T 1_{\Omega_R}L(t,X_t,\beta_t)\,dt\right] &\geq \epsilon CT\mathbb{E}\left[\left|\frac{Y(T)-Y(0)}{T}\right|^2\right]-CU^2T \\ &> \epsilon\frac{C}{T}\mathbb{E}\left[||X(0)|-|X(T)||^2\right]-CU^2T.\end{aligned} \tag{2.26}$$

This leaves us with the same formulation as in (2.4) of the deterministic coercivity result, the remainder of the proof is identical, and the claim is proven.
The existence of a minimizing $\nu_0 \in \mathcal{P}_1(M)$ follows similarly, with the only distinction being that the lower semi-continuity of $\nu \mapsto C^s(\nu,\nu_T)$ follows from Mikami-Thieullen [13].
Remark: a) The same reasoning as in Section 2 yields that $C(\nu_0,\nu_1) = C(\nu_1,\nu_0) = \infty$ for $\nu_1 \in \mathcal{P}_1(M)$ and $\nu_0 \in \mathcal{P}(M)\setminus\mathcal{P}_1(M)$. This implies that it suffices to take the infimum in (2.21) over $\mathcal{P}_1(M)$.
b) The attainment of a minimizing interpolating measure ν_0 is sufficient to show the existence of a minimizing (V,X) for $\underline{B}_T^s(\nu_0,\nu_T)$ whenever the latter is finite. This is a consequence of the existence of minimizers for both $\underline{W}(\mu_0,\nu_0)$ and $C_T^s(\nu_0,\nu_T)$ [13, Proposition 2.1].

To establish the duality formula in 2), we will proceed as in the the deterministic case and use the Legendre dual of the optimal cost functional $\nu \to C_T^s(\nu_0,\nu)$, which was derived by Mikami and Thieullen [13]

Proposition 3. *if the Lagrangian satisfies (A1)-(A3), then*

$$C_T^s(\nu_0,\nu_T) = \sup\left\{\int_M f(x)\,d\nu_T - \int_M \Psi^0_{f,-}(x)\,d\nu_0; f\in\mathcal{C}_b^\infty\right\}, \tag{2.27}$$

where $\Psi_{f,-}$ is the unique solution to the Hamilton-Jacobi-Bellman equation (1.26) that is given by:

$$\Psi_{f,-}(t,x) = \sup_{(X,\beta)\in\mathcal{A}}\left\{\mathbb{E}\left[f(X_T)-\int_t^T L(s,X_s,\beta_s)\,ds\middle|X_t=x\right]\right\}. \tag{2.28}$$

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

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

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

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

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

$$\underline{B}_{\mu_0}(\nu) = \underline{B}_{\mu_0}^{**}(\nu) = \sup_{f \in \text{Lip}(M)} \left\{ \int f \, d\nu - \underline{B}_{\mu_0}^*(f) \right\}. \tag{2.29}$$

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

$$\begin{aligned} \underline{B}_{\mu_0}^*(f) &:= \sup_{\nu_T \in \mathcal{P}_1(M)} \left\{ \int f \, d\nu_T - \underline{B}(\mu_0, \nu_T) \right\} \\ &\overset{(2.21)}{=} \sup_{\substack{\nu_T \in \mathcal{P}_1(M) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int f \, d\nu_T - C(\nu, \nu_T) - \underline{W}(\mu_0, \nu) \right\} \\ &\overset{(2.28)}{=} \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int \Psi_{f,-}^0(x) \, d\nu(x) - \underline{W}(\mu_0, \nu) \right\} \\ &= \underline{W}_{\mu_0}^*(\Psi_{f,-}^0) = -\int (\Psi_{f,-}^0)_* \, d\mu_0. \end{aligned} \tag{2.30}$$

Thus, plugging this into our dual formula (2.29) and restricting our supremum to C_{db}^∞ gives

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

To show the reverse inequality we will adapt the mollification argument used in [13, 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 $\eta_\epsilon(x) = \epsilon^{-d}\eta_1(x/\epsilon)$. For Lipschitz f, $f_\epsilon := f * \eta_\epsilon$ is then smooth with bounded derivatives. We can derive a bound on $\underline{B}^*_{\mu_\epsilon}(f)$ (where $\mu_\epsilon := \mu_0 * \eta$) by removing the supremum in (2.30) and fixing a process $(X, \beta) \in \mathcal{A}^{\nu_T}$:

$$\mathbb{E}\left[f_\epsilon(X_T) - \int_0^T L(s, X_s, \beta_s)\,ds - \langle X_0, V\rangle\right] \overset{\text{(A2)}}{\leq}$$
$$\mathbb{E}\left[f(X_T + H_\epsilon) - \int_0^T \frac{L(s, X_s + H_\epsilon, \beta_s) - \Delta L(0,\epsilon)}{1 + \Delta L(0,\epsilon)}\,ds - \langle X_0 + H_\epsilon, V + H_\epsilon\rangle + |H_\epsilon|^2\right] \leq$$
$$\frac{D^*_\epsilon(f\,[1 + \Delta L(0,\epsilon)])}{1 + \Delta L(0,\epsilon)} + T\frac{\Delta L(0,\epsilon)}{1 + \Delta L(0,\epsilon)} + d\epsilon^2,$$

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

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

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

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

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

In the following corollary, we will discuss results pertaining to the convergence of 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 ψ, since our dual is invariant under $\psi \mapsto \psi + c$, thus when discussing the convergence of a sequence of ψ, we refer to the convergence of their gradients. In the subsequent corollary, we denote $\mathbb{P}_X$ the measure on $M \times [0, T]$ associated with the process X.

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

$$dX = \nabla_p H(t, X_t, \nabla\psi(t, X_t))\, dt + dW_t \tag{2.31}$$
$$V = \nabla\bar{\psi}(X_0), \tag{2.32}$$

and hence a (time inhomogeneous) Markov process, where $\nabla\psi_n(t,x) \to \nabla\psi(t,x)$ $\mathbb{P}_X$*-a.s. and* $\nabla\psi_n(0,x) \to \nabla\bar{\psi}(x)$ ν_0*-a.s. for some sequence* $\psi_n(t,x)$ *that solves (HJB) in such a way that* $\psi_n^T := \psi_n(T,\cdot)$ *and* $(\psi_n^0)_* := [\psi_n(0,\cdot)]_*$ *are maximizing sequences for the dual problem (2.22). 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 is optimal iff there exists a sequence of solutions ψ_n to HJB that is maximizing in (2.22) such that

$$\mathbb{E}\left[\int_0^T L(t, X_t, \beta_t)\, dt + \langle X_0, V\rangle\right] = \lim_{n\to\infty} \mathbb{E}\left[\psi_n^T(X_T) + (\psi_n^0)_*(V)\right], \tag{2.33}$$

which we can write as

$$\lim_{n\to\infty} \mathbb{E}\left[\underbrace{\psi_n^T(X_T) - \psi_n^0(X_0)}_{\text{(a)}} + \underbrace{\psi_n^0(X_0) - (\psi_n^0)_{**}(X_0)}_{\text{(b)}} + \underbrace{(\psi_n^0)_{**}(X_0) + (\psi_n^0)_*(V)}_{\text{(c)}}\right], \tag{2.34}$$

where f_{**} is the concave hull of f. Applying Itô's formula to a), with the knowledge ψ_n satisfies (HJB), we get

$$\mathbb{E}\left[\psi_n^T(X_T) - \psi_n^0(X_0)\right] = \mathbb{E}\left[\int_0^T \langle \beta_t, \nabla\psi_n^t(X_t)\rangle - H(t, X_t, \nabla\psi_n^t(X_t))\, dt\right].$$

However, by the definition of the Hamiltonian, we have $\langle v, b\rangle - H(t,x,v) \leq L(t,x,b)$, which mean that (2.34) yield the following three inequalities:

$$\langle \beta_t, \nabla\psi_n^t(X_t)\rangle - H(t, X, \nabla\psi_n^t(X_t)) \leq L(t, X, \beta_t) \tag{a}$$
$$\psi_n^0(X_0) - (\psi_n^0)_{**}(X_0) \leq 0 \tag{b}$$
$$(\psi_n^0)_{**}(X_0) + (\psi_n^0)_*(V) \leq \langle V, X_0\rangle. \tag{c}$$

In other words, (2.34) breaks the problem into a stochastic and a Wasserstein transport problem (in the flavour of Theorem 4), along with a correction term to account for ψ_n^0 not being necessarily concave. Adding (2.33) to the mix, allows us to obtain L^1 convergence in the (a,b,c) inequalities, hence a.s. convergence of a subsequence ψ_{n_k}.

Note that the convergence in (b,c) means that ψ_n^0 converges ν_0-a.s. to a concave function $\overline{\psi}$ such that $x \mapsto \nabla\overline{\psi}$ is the optimal transport plan for $\underline{W}(\nu_0, \mu_0)$ [6].
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

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

Since ψ_n^t are deterministic functions, this demonstrates that X_t is a Markov process with drift β_t determined by the inverse transform: $\beta_t(X_t) = \nabla_p H(t, X_t, \nabla\psi(t, X_t))$, i.e., (2.31). □

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

Chapter 3

Bolza Duality

For the rest of this thesis we will assume that the Lagrangian L is convex, proper and lower semi-continuous in both variables, so that we 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 costs on $M^* \times M^*$,

$$\tilde{c}_T(u,v) := \inf\{\int_0^T \tilde{L}(t,\gamma(t),\dot{\gamma}(t))\,dt; \gamma \in C^1([0,T),M^*); \gamma(0)=u, \gamma(T)=v\}, \tag{3.1}$$

and its associated optimal transport

$$\tilde{C}_T(\mu_0,\mu_T) := \inf\{\int_{M^*\times M^*} \tilde{c}_T(x,y)\,d\pi;\ \pi \in \mathcal{K}(\mu_0,\mu_T)\}. \tag{3.2}$$

More specifcally, 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 $\rho > 0$, we have $\operatorname{dist}(0,F(x)) \le \rho(1+|x|)$ for all $x \in M$.

(B3) For all $(x,p) \in M \times M$, we have $L(x,p) \ge \theta(\max\{0,|p|-\alpha|x|\}) - \beta|x|$, where α,β are constants, and θ is a coercive, proper, non-decreasing function on $[0,\infty)$.

These conditions on the Lagrangian make sure that the Hamiltonian H is finite, concave in x and convex in q, hence locally Lipschitz. Moreover, we have

$$\psi(x) - (\gamma|x|+\delta)|q| \le H(x,q) \le \phi(q) + (\alpha|q|+\beta)|x| \text{for all } x,q \text{ in } M \times M^*, \tag{3.3}$$

where $\alpha, \beta, \gamma, \delta$ are constants, ϕ is finite and convex and ψ is finite and concave (see [17]).
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 [15] and co-authors [16, 17]. The stochastic counterpart is more recent and has been established by Boroushaki and Ghoussoub [5].

3.1 Deterministic Bolza Duality

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

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

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

$$(\mathcal{P}) \qquad \inf \left\{ \int_0^T L(\gamma(s), \dot{\gamma}(s))\, ds + \ell(\gamma(0), \gamma(T));\ \gamma \in C^1([0, T), M) \right\}, \tag{3.4}$$

and

$$(\tilde{\mathcal{P}}) \qquad \inf \left\{ \int_0^T \tilde{L}(\gamma(s), \dot{\gamma}(s))\, ds + \ell^*(\gamma(0), -\gamma(T));\ \gamma \in C^1([0, T), M^*) \right\} \tag{3.5}$$

where ℓ^* is the legendre transform of ℓ in both variables.

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

1. *If there exists ξ such that $\ell(\cdot, \xi)$ is finite, or there exists ξ' such that $\ell(\xi', \cdot)$ is finite, then*
$$\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}}).$$
This value is not $+\infty$, and if it is also not $-\infty$, then there is an optimal arc $v(t) \in A^2[0, T]$ for $(\tilde{\mathcal{P}})$.

2. *A similar statement holds if we replace* ℓ *by* ℓ^* *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* $A^2[0,T]$.

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

$$\dot{x}(t) \in \partial_v H(x(t), v(t)) \tag{3.6}$$

$$-\dot{v}(t) \in \tilde{\partial}_x H(x(t), v(t)), \tag{3.7}$$

coupled with the boundary conditions

$$(v(0), -v(T)) \in \partial \ell(x(0), x(T)). \tag{3.8}$$

See for example [15]. The above duality has several consequences.

Proposition 5. *The value function* $\Phi_{g,+}(x) = \inf\{g(y) + c(t,y,x);\, y \in M\}$, *which is the variational solution of the Hamilton-Jacobi equation (**??**) 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 Lipchitzian 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: The above shows that in the case when L is jointly convex, the corresponding forward Hamilton-Jacobi equation has convex solutions whenever the initial state is convex, while the corresponding backward Hamilton-Jacobi equation has concave solutions if the final state is concave. Unfortunately, we shall see that in the mass transport problems we are considering, one mostly propagates concave (resp., concave) functions forward (resp., backward), hence losing their concavity (resp., convexity).

This said, the cost functionals c_T, $\tilde{c}_T$, b_T are all value functions Φ_g starting or ending with affine function g. Indeed, $b(t,v,x) = \Phi_{g,+}(t,x)$, when $g_v(y) = \langle v, y\rangle$. In this case, $g_v^*(u) = 0$ if $u = v$ and $+\infty$ if $u \neq v$, which yields that the Legendre dual of $x \to \Phi_{g,+}(t,x) = b(t,v,x)$ is $w \to \tilde{c}(t,v,w)$. One can also deduce the following.

Proposition 6. *Under assumptions* $(B1), (B2), (B3)$ *on the Lagrangian* L*, the costs* c *and* b *have the following properties:*

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

3.2 Stochastic Bolza duality and its applications

We define the *Itô space* $\mathcal{A}^p_M$ consisting of all M-valued processes of the following form:

$$\mathcal{A}^p_M = \Big\{X : \Omega_T \to M;\ X_t = X_0 + \int_0^t \beta^X_s\, ds + \int_0^t \sigma^X_s dW_s,$$
$$\text{for } X_0 \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}(;)M),\ \beta^X \in L^p(\Omega_T; M),\ \sigma^X \in L^2(\Omega_T; M)\Big\}, \tag{3.9}$$

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

$$\|X\|^2_{\mathcal{A}^2_M} = \mathbb{E}\left(\|X_0\|^2_M + \int_0^T \|\beta^X_t\|^2_M\, dt + \int_0^T \|\sigma^X_t\|^2_M\, dt\right),$$

so that it becomes a Hilbert space. Its dual space $(\mathcal{A}^2_M)^*$ 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 $q \in (\mathcal{A}^2_M)^*$ can be represented by the triplet

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

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

$$\langle X, q\rangle_{\mathcal{A}^2_M \times (\mathcal{A}^2_M)^*} = \mathbb{E}\Big\{\langle q_0, X_0\rangle_M + \int_0^T \langle q_1(t), \beta^X_t\rangle_M\, dt + \frac{1}{2}\int_0^T \langle Q(t), \sigma^X_t\rangle_M\, dt\Big\}. \tag{3.10}$$

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

We shall use the following result recently established in [5].

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

$$\begin{aligned}\mathcal{L}(X, p) = \mathbb{E}\Big\{ &\int_0^T L(t, X_t - p_1(t), -\beta^X_t)\, dt + \ell(X_0 - p_0, X_T)\\ &+ \frac{1}{2}\int_0^T M(\sigma^X_t - P(t), -\sigma^X_t)\, dt\Big\}.\end{aligned} \tag{3.11}$$

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

$$\mathcal{L}^*(q, Y) = \mathbb{E}\Big\{\ell^*(-Y(0), Y(T)) + \int_0^T L^*(t, -\beta_t^Y, Y(t) - q_1(t))\, dt \\ + \frac{1}{2}\int_0^T M^*(-\sigma_t^Y, \sigma_t^Y - Q(t))\, dt\Big\}.$$

Note that standard duality theory implies that in general

$$\inf_{X\in\mathcal{A}^2}\{\mathcal{L}(X, 0)\} \geq \sup_{Y\in(\mathcal{A}^2)^*}\{-\mathcal{L}(0, Y)\}. \tag{3.12}$$

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

Proposition 8. *Assume $\tilde{L}$ satisfies (A1) and (A2) in addition to the assumptions in Theorem 6, and there exists (a.s.-)unique $V_0 \in L^2(\mathbb{P})$ such that $\ell^*(V_0, \cdot) < \infty$ and (a.s.)-unique $\sigma^V \in L^2(\mathbb{P}\times\lambda_{[0,T]})$ such that $M^*(\sigma^V, \cdot) < \infty$, then there is no duality gap, ie.*

$$\inf_{X\in\mathcal{A}^2}\{\mathcal{L}(X, 0)\} = \sup_{Y\in(\mathcal{A}^2)^*}\{-\mathcal{L}^*(0, Y)\} \tag{3.13}$$

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

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

We proceed by a variational method outlined by Rockafellar [15]. First, we define

$$\phi(q) := \inf_{Y\in(\mathcal{A}^1)^*}\{\mathcal{L}^*(q, Y)\}. \tag{3.14}$$

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

$$\phi^*(X) = \sup_{q,v}\{\langle X, q\rangle - \mathcal{L}^*(q, v)\} = \mathcal{L}^{**}(X, 0) = \mathcal{L}(X, 0). \tag{3.15}$$

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

$$X \in \partial\phi(0) \iff \phi(0)+\phi^*(X) = 0 \iff \mathcal{L}(X, 0) = -\inf_{Y\in(\mathcal{B}^2)^*}\{\mathcal{L}^*(0, Y)\}. \tag{3.16}$$

In other words, there is no duality gap if and only if $\partial\phi(0)$ is non-empty. Note that this holds if there is an open (relative to $\{q; \phi(q) < \infty\}$) neighbourhood N of the origin in $\mathcal{A}^\infty$ such that $\mathcal{L}^*(q, Y) < \infty$ for $q \in N$.
By our assumptions, we may fix Y_0, σ^Y to be the unique elements such that $\ell(Y_0, \cdot) < \infty$ and $M^*(\sigma^Y, \cdot) < \infty$ (guaranteeing subdifferentiability in these variables), and let $Y = (Y_0, \beta^Y, \sigma^Y)$ be such that $\mathcal{L}^*(0, Y) < \infty$. For a perturbation $\beta^V \in L^\infty(\mathbb{P} \times \lambda_{[0,T]})$ with $\left\|\beta^V\right\|_\infty < \epsilon$, note that (A2) gives for all $(t, u) \in [0, T] \times M^*$,

$$L(t, Y_t - \beta^V, u) < (1 + \Delta L(0, \epsilon))L(t, Y_t, u) + \Delta L(0, \epsilon), \tag{3.17}$$

and

$$\begin{aligned} \phi(V) = &\inf_{Y \in \mathcal{A}^1_M} \mathcal{L}^*(V, Y) \\ \leq &\mathbb{E}\ell^*(-Y_0, Y_T) + \mathbb{E}\int_0^T \tilde{L}(t, Y_t - \beta^V_t, \beta^Y_t)\, dt \\ \leq &\mathbb{E}\ell^*(-Y_0, Y_T) + (1 + \Delta\tilde{L}(0, \epsilon))\mathbb{E}\int_0^T \tilde{L}(t, Y_t, \beta^Y_t)\, dt + T\Delta\tilde{L}(0, \epsilon), \end{aligned} \tag{3.18}$$

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

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

Chapter 4

Maximizing Ballistic Costs

With Bolza duality in mind, it becomes possible to work with the maximizing costs. We present a non-variational method of achieving duality in the deterministic case, while the stochastic case becomes more natural to work with in this setting, as it becomes a case of propagating the cost backwards in time.

4.1 Deterministic Maximizing Cost

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

1. *The following interpolation formula holds:*

$$\overline{B}_T(\mu_0, \nu_T) = \sup\{\overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu);\ \mu \in \mathcal{P}(M^*)\}. \tag{4.1}$$

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

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

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M h(x)\, d\nu_T(x) + \int_{M^*} \tilde{\Phi}^0_{h^*,-}(v)\, d\mu_0(v);\ h \text{ convex in } Lip(M)\right\}. \tag{4.2}$$

and

$$\overline{B}_T(\mu_0, \nu_T) = \inf\left\{\int_M (\tilde{\Phi}^T_{g,+})^*(x)\, d\nu_T(x) + \int_{M^*} g(v)\, d\mu_0(v);\ g \text{ in } Lip(M^*)\right\}. \tag{4.3}$$

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

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(S^*_T \circ \nabla h^*(x), x) d\nu_T(x), \tag{4.4}$$

where $S_T^*(v) = \pi^* \phi_T^{H_*}(v, \nabla h)$*, and* $\phi_t^{H_*}$ *the flow associated to the Hamiltonian* $H_*(v,x) = -H(-x,v)$*, whose Lagrangian is* $L_*(v,q) = L^*(-q,v)$*. In other words, an optimal map for* $\overline{B}_T(\mu_0, \nu_T)$ *is given by the inverse of the map* $x \to \pi^* \phi_T^{H_*}(\nabla h^*(x), x)$*.*

4. *We also have*

$$\overline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T v) d\mu_0(v), \tag{4.5}$$

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

h_0 *the solution* $h(0,v)$ *of the backward Hamilton-Jacobi equation (1.33) with* $h(T,v) = h(v)$*.*

Proof: To show (4.1) and (4.2), first note that for any probability measure μ on M^*, we have

$$\overline{B}_T(\mu_0, \nu_T) \geq \overline{W}(\nu_T, \mu) - \tilde{C}_T(\mu_0, \mu). \tag{4.6}$$

Indeed, since ν_T is assumed to be absolutely continuous with respect to Lebesgue measure, Brenier's theorem yields a convex function h that is differentiable μ_T-almost everywhere on M such that $(\nabla h)_\# \nu_T = \mu$, and $\overline{W}(\nu_T, \mu) = \int_M \langle x, \nabla h(x) \rangle \, d\nu_T(x)$. Let π_0 be an optimal transport plan for $\tilde{C}_T(\mu_0, \mu)$, that is $\pi_0 \in \mathcal{K}(\mu_0, \mu)$ such that $\tilde{C}_T(\mu_0, \mu) = \int_{M^* \times M^*} \tilde{c}_T(v, w) \, d\pi_0(v, w)$. Let $\tilde{\pi}_0 := S_\# \pi_0$, where $S(v,w) = (v, \nabla h^*(w))$, which is a transport plan in $\mathcal{K}(\mu_0, \nu_T)$. Since $b_T(v,y) \geq \langle \nabla h(x), y \rangle - \tilde{c}_T(v, \nabla h(x))$ for every $(y,x,v) \in M \times M \times M^*$, we have

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

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

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

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

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

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

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

We now exploit the convexity of h, and use the fact that for each $(v,x) \in \mathcal{O}$, the function $y \mapsto h(y) - g(v) - b_T(v,y)$ 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

$$b_T(v,x) + \tilde{c}_T(v,\nabla h(x)) = \langle x, \nabla h(x)\rangle \text{ on } \mathcal{O}. \tag{4.7}$$

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

$$\int_{M^*\times M} b_T(v,x)\, d\pi_0 + \int_{M^*\times M} \tilde{c}_T(v,\nabla h(x)) d\pi_0 = \int_M \langle x, \nabla h(x)\rangle\, d\nu_T. \tag{4.8}$$

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

$$\overline{B}_T(\mu_0,\nu_T) + \int_{M^*\times M} \tilde{c}_T(v,\nabla h(x)) d\pi_0 = \overline{W}(\nu_T,\mu_T), \tag{4.9}$$

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

$$\int_{M^*\times M} \tilde{c}_T(v,\nabla h(x)) d\pi_0 = \tilde{C}_T(\mu_0,\mu_T). \tag{4.10}$$

Indeed, we have $\int_{M^*\times M} \tilde{c}_T(v,\nabla h(x)) d\pi_0 \geq \tilde{C}_T(\mu_0,\mu_T)$ since the measure $\pi = S_\#\pi_0$, where $S(v,x) = (v,\nabla h(x))$ has marginals μ_0 and μ_T respectively.

On the other hand, (4.9) 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 $h(x) - g(v) \geq b(v,x)$, we have $h^*(w) + g(v) \leq \tilde{c}_T(v,w)$. Indeed, since for any $(v,w) \in M^* \times M^*$, we have $\tilde{c}(t,v,w) = \sup\{\langle w,x\rangle - b(t,v,x); x \in M\}$, it follows that for any $y \in M$,

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

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

$$\begin{aligned}\tilde{C}_T(\mu_0,\mu_T) &\leq \int_{M^*\times M}\tilde{c}_T(v,\nabla h(x))d\pi_0\\ &= \int_M h^*(w)d\mu_T(w) + \int_{M^*} g(v)\, d\mu_0(v)\\ &\leq \sup\{\int_{M^*}\phi_T(w)\, d\mu_T(w) - \int_{M^*}\phi_0(v)\, d\mu_0(v);\ \phi_T(w) - \phi_0(v) \leq \tilde{c}_T(v,w)\}\\ &= \tilde{C}_T(\mu_0,\mu_T).\end{aligned}$$

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 (4.6) is attained by the measure μ_T. Note that the final optimal Kantorovich potential for $\tilde{C}_T(\mu_0,\mu_T)$ is h^*, hence is convex.
The first duality formula (4.3) 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 (4.2).
To show (4.3), we can –now that the interpolation (4.1) is established– proceed as in Section 2.1, 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 show part 4), we start with the interpolation inequality and write that

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

for some probability measure μ_T. The proof also shows that there exists a convex function $h : M^* \to \mathbb{R}$ 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

$$\tilde{C}_T(\mu_0, \mu_T) = \int_{M^*} c_T(v, \tilde{S}_T v) d\mu_0(v), \tag{4.11}$$

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

$$\overline{B}_T(\mu_0, \nu_T) \geq \textstyle\int_{M^*} b_T(v, \nabla h \circ \tilde{S}_T(v)) d\mu_0(v), \tag{4.12}$$

since $(\tilde{S}_T)_\# \mu_0 = \mu_T$ and $\nabla h_\# \mu_T = \nu_T$, and therefore $(I \times \nabla h \circ \tilde{S}_T)_\# \mu_0$ belongs to $\mathcal{K}(\mu_0, \nu_T)$.
On the other hand, since $b_T(u, x) \geq \langle \nabla h(v), x \rangle - \tilde{c}_T(u, \nabla h(v))$ for every $v \in M^*$, we have

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

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

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

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

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

Remark: 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 ν_0 such that $C_T(\nu_0, \nu_T)$ has a concave initial Kantorovich potential, while the interpolation (4.1) of $\overline{B}_T(\mu_0, \nu_T)$ selects a μ_T such that $\tilde{C}_T(\mu_0, \mu_T)$ has a convex final Kantorovich potential.

Furthermore, one wonders whether the formula

$$c(t, y, x) = \sup\{b(t, v, x) - \langle v, y\rangle; v \in M^*\}, \tag{4.13}$$

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 ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure. If the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave then the following holds:*

$$C_T(\nu_0, \nu_T) = \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu);\ \mu \in \mathcal{P}(M^*)\}, \tag{4.14}$$

and the supremum is attained.

Proof: Again, it is easy to show that

$$C_T(\nu_0, \nu_T) \geq \sup\{\underline{B}_T(\mu, \nu_T) - \underline{W}(\nu_0, \mu);\ \mu \in \mathcal{P}(M^*)\}. \tag{4.15}$$

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

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

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

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

Since g is concave, use the fact that for each $(y, x) \in \mathcal{O}$, the function $z \to h(x) - g(z) - c_T(z, x)$ 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 $(y,x) \in \mathcal{O}$,

$$c_T(y,x) = b_T(\nabla g(y),x) - \langle \nabla g(y), y\rangle. \tag{4.16}$$

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

$$\int_{M\times M} c_T(y,x)\,d\pi_0 = \int_{M\times M} b_T(\nabla g(y),x)\,d\pi_0 - \int_M \langle \nabla g(y), y\rangle\,d\nu_0. \tag{4.17}$$

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

$$C_T(\nu_0,\nu_T) = \int_{M\times M} b_T(\nabla g(y),x)\,d\pi_0 - \underline{W}(\nu_0,\mu_0). \tag{4.18}$$

We now prove that

$$\int_{M\times M} b_T(\nabla g(y),x)\,d\pi_0(y,x) = \underline{B}_T(\mu_0,\nu_T). \tag{4.19}$$

Indeed, we have $\int_{M\times M} b_T(\nabla g(y),x)\,d\pi_0 \geq \underline{B}_T(\mu_0,\nu_T)$, since the measure $\pi = S_\#\pi_0$ where $S(y,x) = (\nabla g(y),x)$ has μ_0 and ν_T as marginals. On the other hand, (4.18) 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 $h(x)-g(y) \leq c_T(y,x)$, it is easy to see that $h(x)-(-g)^*(-v) \leq b_T(v,x)$, that is the couple $((-g)^*(-v), h(x))$ is an admissible Kantorovich pair for the cost b_T. It follows that

$$\begin{aligned}
\underline{B}_T(\mu_0,\nu_T) &\leq \int_{M\times M} b_T(\nabla g(y),x)\,d\pi_0\\
&= \int_M h(x)\,d\nu_T(x) - \int_M (-g)^*(-v)d\mu_0(v)\\
&\leq \sup\{\int_M \phi_T(x)\,d\mu_T(x) - \int_{M^*} \phi_0(v)\,d\mu_0(v);\, \phi_T(x) - \phi_0(v) \leq b_T(v,x)\}\\
&= \underline{B}_T(\mu_0,\nu_T),
\end{aligned}$$

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 (4.14) is attained by the measure μ_0.

Corollary 9. *Assume $M = \mathbb{R}^d$ and that L satisfies hypothesis (B1), (B2) and (B3). Assume ν_0 and ν_T are probability measures on M such that ν_0 is absolutely continuous with respect to Lebesgue measure, and that the initial Kantorovich potential of $C_T(\nu_0, \nu_T)$ is concave. If b_T satisfies the twist condition, then there exists a map $X_0^T : M^* \to M$ and a concave function g on M such that*

$$C_T(\nu_0, \nu_T) = \int_M c_T(y, X_0^T \circ \nabla g(y)) d\nu_0(y). \tag{4.20}$$

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 μ_0 on M^*. Let g be the concave function on M such that $(\nabla g)_\# \nu_0 = \mu_0$ and $\underline{W}(\nu_0, \mu_0) = \int_M \langle \nabla g(y), y \rangle d\nu_0(y)$. Since b_T satisfies the twist condition, there exists a map $X_0^T : M^* \to M$ such that $(X_0^T)_\# \mu_0 = \nu_T$ and

$$\underline{B}_T(\mu_0, \nu_T) = \int_{M^*} b_T(v, X_0^T v) d\mu_0(v). \tag{4.21}$$

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 π_0 is supported on a subset $\mathcal{O}$ of $M \times M$ such that for $(y, x) \in \mathcal{O}$, $c_T(y, x) = b_T(\nabla g(y), x) - \langle \nabla g(y), y \rangle$. Moreover, $C_T(\nu_0, \nu_T) = \int_{M \times M} b_T(\nabla g(y), x)\, d\pi_0 - \underline{W}(\nu_0, \mu_0)$, and

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

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

Corollary 10. *Consider the cost $c_1(y, x) = c(x - y)$, where c is a convex function on M and let ν_0, ν_1 be probability measures on M such that the initial Kantorovich potential associated to $C_T(\nu_0, \nu_T)$ is concave. Then, there exist concave functions $\phi : M \to \mathbb{R}$ and $\psi : M^* \to \mathbb{R}$ such that*

$$(\nabla \psi \circ \nabla \phi)_\# \nu_0 = \nu_1, \tag{4.22}$$

and

$$C_1(\nu_0, \nu_1) + K = \int_M c(\nabla \psi \circ \nabla \phi(y) - y) d\nu_0(y) = \int_M \langle \nabla \psi_*(y) - \nabla \phi(y), y \rangle \, d\nu_0(y), \tag{4.23}$$

where $K > 0$ is a constant and ψ_* is the concave Legendre transform of ψ.

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

$$c_1(y,x)=\inf\{\int_0^1 c(\dot\gamma(t))\,dt;\gamma\in C^1([0,1),M);\gamma(0)=y,\gamma(1)=x\}=c(x-y). \tag{4.24}$$

It follows from (4.14) that there is a probability measure μ_0 on M^* such that $C_1(\nu_0,\nu_1)=\underline{B}_1(\mu_0,\nu_1)-\underline{W}(\nu_0,\mu_0)$. But in this case, $b_1(v,x)=\inf\{\langle v,y\rangle + c(x-y);\, y\in M\}=\langle v,x\rangle - c^*(v)$, hence

$$C_1(\nu_0,\nu_1)=\underline{B}_1(\mu_0,\nu_1)-\underline{W}(\nu_0,\mu_0)=\underline{W}(\mu_0,\nu_1)-\int_{M^*}c^*(v)\,d\mu_0(v)-\underline{W}(\nu_0,\mu_0). \tag{4.25}$$

In other words,

$$C_1(\nu_0,\nu_1)+K=\underline{W}_1(\mu_0,\nu_1)-\underline{W}(\nu_0,\mu_0), \tag{4.26}$$

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

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

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

It follows from the preceeding corollary that

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

Note also that

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

4.2 Stochastic Maximizing Cost

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

$$b^s(V,Z):=\inf\{\mathbb{E}\left[\langle V,X_0\rangle+\int L(t,X_t,\beta_t)\,dt\right];(X,\beta)\in\mathcal{A},X_T=Z\text{ a.s.}\}, \tag{4.27}$$

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

$$\underline{B}_T^s(\mu_0, \nu_T) = \inf\{b^s(V(\cdot), Y(\cdot)); V \sim \mu_0, Y \sim \nu_T\}, \tag{4.28}$$

while the maximizing cost is defined as:

$$\overline{B}_T^s(\mu_0, \nu_T) := \sup\{b^s(V(\cdot), Y(\cdot)); V \sim \mu_0, Y \sim \nu_T\}. \tag{4.29}$$

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

1. *The following formula holds:*

$$\overline{B}_T^s(\mu_0, \nu_T) := \sup\left\{\mathbb{E}\left[\langle V, X_T\rangle - \int_0^T \tilde{L}(t, X_t, \beta_t(X_t))\, dt\right]; (X, \beta) \in \mathcal{A}, X_0 \sim \mu_0, V \sim \nu_T\right\}. \tag{4.30}$$

2. *The following duality holds:*

$$\overline{B}_T^s(\mu_0, \nu_T) = \sup\{\overline{W}(\mu, \nu_T) - \tilde{C}_T^s(\mu_0, \mu)\}, \tag{4.31}$$

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

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

$$\overline{B}(\mu_0, \nu_T) = \inf\left\{\int_M g^*\, d\nu_T + \int_{M^*} \tilde{\Psi}_{g,-}\, d\mu_0;\ g \in C_{db}^\infty(M^*) \text{ and convex}\right\}, \tag{4.32}$$

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

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

Remark: The case where ν_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 eq. 2.28—with $k : x \mapsto \langle u, x\rangle$. Thus, in this particular case we can recover

$$\overline{B}(\mu_0, \delta_u) = \int \tilde{\Psi}_{k,-}(0, x)\, d\mu_0(x),$$

without the aid of any duality theorem. Notably,

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

Moreover, this method indicates a minimizing process, described by the SDE

$$X_t = X_0 + \int_0^t \nabla_p H(s, X, \nabla\tilde{\Psi}_{k,-}(s, X))\, ds + W_t.$$

Proof: 1) For a fixed pair (V, Y), we consider the Bolza energy $\mathcal{L}_{(V,Y)}$ (3.11) associated to L and the two Lagrangians ℓ and M defined as:

$$\ell_{(Y,U)}(\omega, y, z) := \begin{cases} \langle z, U(\omega)\rangle & y = Y(\omega) \\ \infty & \text{else} \end{cases} \qquad M(\xi, \zeta) := \begin{cases} -\zeta - 1 & \xi = 1 \\ \infty & \text{else} \end{cases} \tag{4.33}$$

Note that the minimizing stochastic cost can be written as,

$$\underline{B}_s(\mu_0, \nu_T) := \inf\{\inf\{\mathcal{L}_{(V,Y)}(X_t, 0); (X, \beta) \in \mathcal{A}^2\}; V \sim \mu_0, Y \sim \nu_T\} \tag{4.34}$$

while the maximizing cost is

$$\overline{B}_s(\mu_0, \nu_T) = \sup\{\inf\{\mathcal{L}_{(V,Y)}(X_t, 0); (X, \beta) \in \mathcal{A}^2\}; V \sim \mu_0, Y \sim \nu_T\}. \tag{4.35}$$

Applying Bolza duality turns the infimum to a supremum:

$$\overline{B}_s(\mu_0, \nu_T) = \sup\{\sup\{-\mathcal{L}^*_{(V,Y)}(0, U(t)); U \in \mathcal{A}^2\}; V \sim \mu_0, Y \sim \nu_T\}, \tag{4.36}$$

which results in (4.30).

2) The proof of the interpolation result can now follow closely the proof for the minimization problem.
3) We again try to identify the Legendre transforms of the functionals $\nu \mapsto \overline{W}(\mu, \nu)$ and $\mu \to \tilde{C}_T^s(\mu_0, \nu)$. We obtain easily that

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

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

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

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

Define $\overline{B}_{\mu_0} : \nu \mapsto \overline{B}(\mu_0, \nu)$, and note that the interpolation formula (4.31) and a result of Mikima-Thieullen [13] concerning $\tilde{C}^s_T$ yields that $\overline{B}_{\mu_0}$ is a concave function. Furthermore it is weak*-upper semi-continuous on $\mathcal{P}_1(M)$. Thus we have

$$\overline{B}_{\mu_0}(\nu_T) = -(-\overline{B}_{\mu_0})^{**}(\nu_T) = \inf_{f \in \mathrm{Lip}(M)} \left\{ -\int_M f \, d\nu_T + (-\overline{B}_{\mu_0})^*(f) \right\}. \tag{4.37}$$

Investigating the dual, we find

$$\begin{aligned}
(-\overline{B}_{\mu_0})^*(f) &= \sup_{\nu \in \mathcal{P}_1(M)} \left\{ \int_M f \, d\nu + \overline{B}_{\mu_0}(\nu) \right\} \\
&= \sup_{\substack{\mu \in \mathcal{P}_1(M^*) \\ \nu \in \mathcal{P}_1(M)}} \left\{ \int_M f \, d\nu + \overline{W}(\mu, \nu) - \tilde{C}^s_T(\mu_0, \mu) \right\} \\
&= \sup_{\mu \in \mathcal{P}_1(M^*)} \left\{ \int_{M^*} (-f)^* \, d\mu - \tilde{C}^s_T(\mu_0, \mu) \right\}. \qquad (4.38)
\end{aligned}$$

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

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

In either case, we can restrict our f to be concave by noting that if we fix $g = (-f)^*$, then the set of corresponding $\{-f; (-f)^* = g\}$ is minimized by the convex function $g^* = (-f)^{**} \leq -f$ [7, Proposition 4.1]. Thus it suffices to consider f convex.

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

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

Mikami [13, Proof of Theorem 2.1] further shows that

$$(4.38) = C^*_{\nu_0}(g_\epsilon) \leq \frac{C^*_{\nu_0 * \eta_\epsilon}((1 + \Delta L(0, \epsilon))g)}{1 + \Delta L(0, \epsilon)} + T \frac{\Delta L(0, \epsilon)}{1 + \Delta L(0, \epsilon)}.$$

Putting these together we get

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

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

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

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 $d\mu_0 \ll d\lambda$. Then, (V_t, X) is an optimal process if and only if it is a solution to the backward Stochastic differential equation,*

$$dV = \nabla_p H(t, \nabla\psi(t,V), V)\, dt + dW_t \tag{4.39}$$

$$X = \nabla\bar{\psi}(V_T), \tag{4.40}$$

where $\lim_{n\to\infty} \psi_n(T,x) \to \bar{\psi}(x)$ *ν_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 (HJB) in such a way that $\psi_n^0 = \psi_n(0,\cdot)$ and $\psi_n^T = \psi_n(T,\cdot)$ are a minimizing pair for the dual problem.*

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

$$\mathbb{E}\left[\langle X, V_T\rangle - \int_0^T \tilde{L}(t, V, \beta_t^V(V))\, dt\right] = \lim_{n\to\infty} \mathbb{E}\left[\left[\psi_n^T\right]^*(X) + \psi_n^0(V_0)\right], \tag{4.41}$$

which we write as

$$\lim_{n\to\infty} \mathbb{E}\left[\left[\psi_n^T\right]^*(X) + \psi_n^T(V_T) - \psi_n^T(V_T) + \psi_n^0(V_0)\right].$$

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

$$\mathbb{E}\left[-\psi_n^T(V_T) + \psi_n^0(V_0)\right] = \mathbb{E}\left[\int_0^T -\langle \beta_t^V, \nabla\psi_n^t(V_t)\rangle - H(t, \nabla\psi_n^t(V_t), V_t)\, dt\right]$$

However, by the definition of the Hamiltonian, we have $-\langle q, v\rangle - H(t,x,v) \ge -\tilde{L}(t,v,q)$, similarly $\psi^*(v) + \psi(x) \ge \langle v, x\rangle$. These inequalities allow us to

separate the limit in (4.41) into two requirements:
(a) $\langle \beta_t^V, \nabla\psi_n^t(V_t)\rangle + H(t, \nabla\psi_n^t(V_t), V_t)$ must converge to $\tilde{L}(t, V, \beta_t^V(t, V))$ 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 ψ_{n_k} exists such that this convergence is a.e.

The journey from (a) to (4.39) is as in Corollary 4. The only difference from the earlier corollary is that we know that ψ_n must converge to a convex function, so (b) implies $X = \nabla \lim_{n\to\infty} \psi_n(V_T)$.

4.3 Final Remarks

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

$$\underline{\mathcal{B}}^{\mu_0}(t,\nu) = \inf\{\underline{\mathcal{U}}^{\mu_0}(\rho) + C_t(\rho,\nu);\ \rho \in \mathcal{P}(M)\} \text{ and } \overline{\mathcal{B}}^{\nu_T}(t,\mu) = \inf\{\overline{\mathcal{U}}^{\nu_T}(\rho) - \tilde{C}_t(\rho,\mu);\ \rho \in \mathcal{P}(M^*)\}, \tag{4.42}$$

where

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

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

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

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

Corollary 12. *Assume L satisfies hypothesis (D0) and (D1), and let μ_0 be a probability measure on M^* with compact support, then*

$$\underline{\mathcal{B}}^{\mu_0}(T,\nu) := \inf\left\{\underline{\mathcal{U}}^{\mu_0}(\rho_0) + \int_0^T \mathcal{L}(\rho_t, w_t)dt;\ \partial_t\varrho + \nabla\cdot(\varrho w) = 0,\ \varrho_T = \nu\right\}, \tag{4.43}$$

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

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

$$\begin{cases} \partial_t B + \mathcal{H}(t, \nu, \nabla_\nu B(t,\nu)) = 0, \\ B(0,\nu) = \underline{W}(\mu_0, \nu). \end{cases} \tag{4.44}$$

Here the Hamiltonian is defined as

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

We note that Ambrosio-Feng [1] 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 (4.44). As importantly, Gangbo-Sweich [11] have shown recently that under certain conditions, value functionals yield solutions to the so-called *Master equations* of mean field games.

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 below (4.45).

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

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

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

$$\begin{cases} \partial_t \beta + \int \langle \nabla_\nu \beta(t,x,\nu) \cdot \nabla H(x, \nabla_x \beta)\rangle\, d\nu + H(x, \nabla_x \beta(t,x,\nu)) = 0, \\ \partial_t \rho + \nabla(\rho \nabla H(x, \nabla_x \beta)) = 0, \\ \beta(0,\cdot,\cdot) = \beta_0, \quad \rho(T,\cdot) = \nu_T, \end{cases} \tag{4.45}$$

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

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

$$b_g(T,v,x) := \inf\left\{ g(v,\gamma(0)) + \int_0^T L(\gamma(t),\dot{\gamma}(t))\,dt; \gamma \in C^1([0,T),M) \right\}, \tag{4.46}$$

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

Bibliography

[1] L. Ambrosio and J. Feng. On a class of first order Hamilton-Jacobi equations in metric spaces. *Journal of Differential Equations*, 256, 4 2014.

[2] A. Barton and N. Ghoussoub. On Optimal Stochastic Ballistic Transports. *ArXiv e-prints*, November 2017.

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

[4] P. Bernard and B. Buffoni. Weak KAM Pairs and Monge-Kantorovich Duality. 2008.

[5] S. Boroushaki and N. Ghoussoub. A Self-dual Variational Approach to Stochastic Partial Differential Equations. *ArXiv e-prints*, October 2017.

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

[7] I. Ekeland and R. Téman. *Convex Analysis and Variational Problems*, chapter 1. Convex Functions. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[8] L.C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem. *Mem. Amer. Math. Soc.*, 137, 1999.

[9] A Fathi and A Figalli. Optimal transportation on non-compact manifolds. *Israel Journal of Mathematics*, 175, 2010.

[10] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177, 1996.

[11] W. Gangbo and A. Swiech. Existence of a solution to an equation arising from the theory of Mean Field Games. *J. Differential Equations*, 259, 2015.

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

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

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

[15] R. T. Rockafellar. Existence and duality theorems for convex problems of Bolza. *Trans. Amer. Math. Soc.*, 159, 1971.

[16] R. T. Rockafellar and P.R. Wolenski. Convexity in Hamilton-Jacobi theory I: dynamics and duality. *SIAM J. Control and Opt.*, 39, 2001.

[17] R. T. Rockafellar and P.R. Wolenski. Convexity in Hamilton-Jacobi theory II: envelope representations. *SIAM J. Control and Opt.*, 39, 2001.

[18] V.N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.*, 141, 1979.

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

[20] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.

[21] Daniel H. Wagner. Survey of measurable selection theorems. *SIAM Journal on Control and Optimization*, 15(5):859–45, 08 1977.