OPTIMAL TRANSPORT WITH CONTROLLED DYNAMICS AND FREE END TIMES

NASSIF GHOUSSOUB, YOUNG-HEON KIM, AND AARON ZEFF PALMER

ABSTRACT. We consider optimal transport problems where the cost is optimized over controlled dynamics and the end time is free. Unlike the classical setting, the search for optimal transport plans also requires the identification of optimal “stopping plans,” and the corresponding Monge-Kantorovich duality involves the resolution of a Hamilton-Jacobi-Bellman quasi-variational inequality. We discuss both Lagrangian and Eulerian formulations of the problem, and its natural connection to Pontryagin’s maximum principle. We also exhibit a purely dynamic situation, where the optimal stopping plan is a hitting time of a barrier given by the free boundary problem associated to the dual variational inequality. This problem was motivated by its stochastic counterpart, which will be studied in a companion paper.

Contents

1. Introduction 1
2. The primal problem and its dual à la Monge-Kantorovich 6
3. An Eulerian Formulation 8
4. Necessary Optimality Conditions 11
5. When optimal end times are hitting times of a barrier 15
6. One dimensional examples 19
7. Connections with classical optimal transport problems 21
References 23

1. Introduction

With a purpose of connecting mass transport with Mather theory, Bernard and Buffoni [5] considered “generating functions” on a compact manifold $M$ as cost functionals. These are of the following type:

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

where $[0, T]$ is a fixed time interval, and $L : \mathbb{R}^+ \times TM \to \mathbb{R}$ is a given Lagrangian that is convex in the second variable of the tangent bundle $TM$. The corresponding
optimal transport problem consists of minimizing the total cost of “transport plans” between two given probability distributions \( \mu \) and \( \nu \) on \( M \),

\[
V_{ct}(\mu, \nu) := \inf \left\{ \int_{M \times M} c_{T}(x, y) \, d\pi; \pi \in \Pi(\mu, \nu) \right\}.
\]

Here \( \Pi(\mu, \nu) \) is the set of transport plans from \( \mu \) to \( \nu \), that is those probability measures \( \pi \) on \( M \times M \) whose first (resp., second) marginal is \( \mu \) (resp., \( \nu \)).

The consideration of such cost functionals showcased the fact that the theory of optimal transport can be seen as a natural extension of certain aspects of classical mechanics, where one considers an ensemble of identical, non-interacting ‘particles’, which are transported from a given source distribution \( \mu \), to a desired target distribution \( \nu \).

Note that standard cost functionals of the form \( c(x, y) = g(|y - x|) \), where \( g \) is convex, are particular cases of the dynamic formulation, since they correspond to Lagrangians of the form \( L(t, x, v) = g(v) \). The original Monge problem dealt with the cost \( c(x, y) = |y - x| \) (\cite{16, 19, 10, 20, 21}) and was constrained to those probabilities in \( \Pi(\mu, \nu) \) that are supported by graphs of measurable maps from \( X \) to \( Y \) pushing \( \mu \) onto \( \nu \). Brenier \cite{6} considered the important quadratic case \( c(x, y) = |y - x|^2 \). This was followed by a large number of results addressing costs of the form \( g(|y - x|) \), where \( g \) is either a convex or a concave function \cite{13}.

In this paper, we develop mass transport theory in two new directions: For one, we shall formulate it as a natural extension of optimal control theory, the latter being an extension of the classical calculus of variations to ‘nonsmooth’ or ‘semi-discrete’ problems that arise frequently in engineering \cite{18}. Instead of optimizing value functionals when the Lagrangian is a function of velocity, the controlled value might only be optimized over a set of controls, \( A(t) \in A \), that prescribe the velocity. The ‘state’ variable can then obey dynamics of the form

\[
\dot{\gamma}(t) = f(\gamma(t), A(t)).
\]

These type of problems with a fixed end time, that is, when all trajectories stop at a given time \( T \), has been studied by Agrachev and Lee \cite{1}. We shall consider the case when the end time is free, that is, when not given a priori, and is motivated by Skorokhod type embedding problems that we study in \cite{15} and the stochastic control counterpart that we address in \cite{14}. Examples of such control problems arise in engineering problems such as allocation of emergency resources \cite{12}.

The new elements of our approach include:

- The use of Kantorovich type duality to define an obstacle problem associated to a corresponding Hamilton-Jacobi equation. The relevance of this approach will be even more apparent in our companion paper \cite{15}, where it is used to construct Skorokhod embeddings as hitting times of a corresponding free boundary problem. This approach has already been used in particular one-dimensional Skorokhod problems by Cox and collaborators \cite{9, 3}.

- An equivalent Eulerian formulation to the optimal transportation problem, with which we can again employ PDE methods.
The transport associated to control problems with free end times consists of considering cost functionals of the following type:

\[ c(x, y) := \inf_{\tau, A(\cdot)} \left\{ \int_0^\tau K(t, \gamma(t), A(t)) \, dt; \; \dot{\gamma}(t) = f(\gamma(t), A(t)), \; \gamma(0) = x, \; \gamma(\tau) = y \right\}, \]

and the corresponding optimal total cost \( V_c(\mu, \nu) \). To our knowledge (and surprise), this variant of optimal control theory and optimal transportation has not been studied in the mathematical literature, despite the elegant mathematics that it yields, which we present in this paper.

Our approach follows the, by now standard, pattern of identifying the dual of our optimal transportation problem, and use the properties of its extremals to infer properties of the primal problem. Recall that under very general conditions on a cost functional \( c \), the Monge-Kantorovich duality yields that

\[ V_c(\mu, \nu) = D_c(\mu, \nu) := \sup \left\{ \int_M \psi(y) \, d\nu - \int_M \phi(x) \, d\mu; \; \psi, \phi \in K(c) \right\}, \]

where \( K(c) \) is the set of functions \( \psi \in L^1(M, \nu) \) and \( \phi \in L^1(M, \mu) \) such that

\[ \psi(y) - \phi(x) \leq c(x, y) \quad \text{for all} \; (x, y) \in M \times M. \]

The pairs of functions in \( K(c) \) can be assumed to satisfy

\[ \psi(y) = \inf_{x \in M} c(x, y) + \phi(x) \quad \text{and} \quad \phi(x) = \sup_{y \in M} \psi(y) - c(x, y). \]

These will be called optimized Kantorovich potentials, and for reasons that will become clear later, we shall say that \( \phi \) (resp., \( \psi \)) is an initial (resp., final) Kantorovich potential.

A crucial contribution by Bernard-Buffoni was to relate optimized Kantorovich potentials to solutions of corresponding Hamilton-Jacobi equation,

\[ \frac{\partial}{\partial t} J + H(t, x, \nabla J) = 0 \quad \text{on} \; [0, T] \times M, \]

with either the forward or the backward conditions,

\[ J(0, \cdot) = \phi(\cdot), \]

\[ J(T, \cdot) = \psi(\cdot) \quad \text{on} \; M, \]

where the Hamiltonian on \([0, T] \times T^* M\) is defined by

\[ H(t, q, p) := \sup_{v \in T_q M} \{ \langle v, p \rangle - L(t, q, v) \}. \]

This connection, coupled with the fact that the optimal transport map is often given by the Hamiltonian flow, clearly illustrates our claim above on how the theory of mass transport is a natural extension of the calculus of variations view of classical mechanics. The Eulerian formulation of mass transport, as initiated by Brenier-Benamou [4] in the case of a quadratic cost, also leads to its identification as a version of the calculus of variation but on Wasserstein space of probability measures.
Following L. Pontryagin, we define the Hamiltonian in terms of the generalized momentum, or costate, $p$ as follows:

$$H(t, q, p) = \sup_{A \in \mathcal{A}} \left\{ p \cdot f(q, A) - K(t, q, A) \right\}.$$  

(1.1)

Now, under certain natural hypothesis on $K$ and $f$, the cost $c$ becomes lower semi-continuous and Lipschitz in the second variable, and it is then standard (c.f. [20]) to conclude that the Kantorovich duality holds with attainment of both primal and dual problems. In our particular case, we show in Section 2, the following duality

$$V(\mu, \nu) = D(\mu, \nu) := \sup \left\{ \int_{\mathbb{R}^n} \psi \, d\nu - \int_{\mathbb{R}^n} J_\psi(0, \cdot) \, d\mu ; \psi \in C(\mathbb{R}^n) \right\},$$

(1.2)

where $J_\psi(t, \cdot)$ is the ‘value’ or cost-to-go function, which represents maximum payoff when the terminal-payoff is capped at $\psi$, that is

$$J_\psi(t, q) = \sup_{\tau, A(\cdot)} \left\{ \psi(\gamma(\tau)) - \int_0^\tau K(s, \gamma(s), A(s)) \, ds ; \gamma(0) = q \right\}. $$

(1.3)

The dynamic programming principle then equates $J_\psi$ with viscosity solutions of the Hamilton-Jacobi-Bellman quasi-variational inequality [2],

$$\frac{\partial}{\partial t} J(t, q, \nabla J) \leq 0 \text{ on } \mathbb{R}^+ \times \mathbb{R}^n$$

$$\psi - J \leq 0 \text{ on } \mathbb{R}^+ \times \mathbb{R}^n,$$

(1.4)

where one of the two alternatives holds with equality —restated as (2.8) below. This is a key ingredient in our analysis. The function $J_\psi$ represents the backward propagation of the cost such that, if differentiable, $-\nabla J_\psi(0, x)$ is the derivative of the cost of an optimal trajectory with respect to the starting position $x$. While for each $(t, q)$, we have $J_\psi(t, q) \geq \psi(q)$, (1.4) also implies that $J_\psi(0, \gamma(0)) \leq \psi(\gamma(\tau))$ for any maximizing trajectory of (1.3). These competing inequalities determine the end time $\tau$ to lie in the free boundary of the Hamilton-Jacobi inequality, i.e. the boundary of the set where $J_\psi = \psi$.

We also give an Eulerian formulation of the primal problem $V(\mu, \nu)$ as follows. We consider density processes $\rho : \mathbb{R}^+ \rightarrow \mathcal{M}^+ (\mathbb{R}^n \times \mathcal{A})$ and stopping distributions $\tilde{\rho} \in \mathcal{M}^+ (\mathbb{R}^+ \times \mathbb{R}^n)$, where $\mathcal{M}^+$ represent the cone of non-negative Radon measures on the given metric space. We then introduce $\mathcal{C}(\mu, \nu)$ to be the class of all admissible process densities $\rho$, starting at $\mu$, i.e., $d\mu(x) = \int_\mathcal{A} d\rho(0, x, \cdot)$, with a stopping distribution $\tilde{\rho}$, i.e., $d\nu(x) = \int_{\mathbb{R}^+} d\tilde{\rho}(\cdot, x)$, and satisfying (weakly) the following continuity equation:

$$\tilde{\rho} + \frac{\partial}{\partial t} \left( \int_\mathcal{A} \, d\rho \right) + \text{div} \left( \int_\mathcal{A} f \, d\rho \right) = 0.$$

(1.5)

We then show in Theorem 3.4 the following analogue of a celebrated result of Benamou-Brenier [4]:

$$V(\mu, \nu) = W(\mu, \nu) := \inf \left\{ \int_{\mathbb{R}^+} \int_\mathcal{A} \int_{\mathbb{R}^n} K(t, q, A) \rho(t, dq, dA) \, dt ; \rho \in \mathcal{C}(\mu, \nu) \right\}.$$  

(1.6)
We emphasize that unlike the classical case, where mass is conserved while moving in time and then stops at once at the given final time, our admissible process densities $\rho$ drop mass over time before ending with a stopping distribution $\tilde{\rho}$.

We next consider necessary optimality conditions for when the cost $c(x, y)$ of moving state $x$ to $y$ is attained in a –not necessarily unique– time $\tau_{x,y}$, via a control policy $A_{x,y}$. Under suitable conditions, which ensure that the primal problem is attained at a probability measure $\pi$, and the dual problem is attained at some function $\psi$ with $J_\psi$ as the corresponding solution of the Hamilton-Jacobi inequality (1.4), we show in Theorem 4.1 that

$$J_\psi(\tau_{x,y}, y) = \psi(y) \text{ for } \pi\text{-a.e. } (x, y).$$

And just like the fixed-end case, but this time resorting to Pontryagin’s maximum principle (Theorem 4.1, item (2)), if an optimal path exists for $(x, y)$ on the support of $\pi$, then it is given by the projection $\gamma(t)$ of the Hamiltonian flow $(\gamma(t), p(t))$, where $\gamma(t)$ starts at $x$ and ends at $y$, while the costate $p(t)$ starts at $\nabla J_\psi(0, x)$ and ends at $\nabla \psi(y)$. Moreover,

(1) for a.e. $t < \tau_{x,y}$, the control is one where the supremum of the Hamiltonian

$$H(t, \gamma(t), p(t)) = p(t) \cdot f(\gamma(t), A(t)) - K(t, \gamma(t), A(t))$$

is attained,

(2) while at the end time $\tau = \tau_{x,y}$, the transversality condition $H(\tau, \gamma(\tau), p(\tau)) = 0$ holds.

Note that the existence of an optimal transport plan $\pi$ does not necessarily imply that for $\pi$-a.e. $(x, y)$, the cost $c(x, y)$ defined in (2.2) is attained by some optimal pair $(\tau, A)$. Indeed, without the concavity of $p \cdot f(q, A) - K(t, q, A)$ in (1.1), we cannot expect attainment of the optimal control policy by a map $t \mapsto A(t)$ for each initial position $x$ and end position $y$. However, one can go around this problem by relaxing the problem and considering Young probability measures $t \mapsto \alpha_t \in \mathcal{P}(A)$. But these are already subsumed under the admissible process densities introduced for the Eulerian formulation (1.5). With a suitable coercivity condition for $t$ and $A$, we shall show attainment in the Eulerian formulation at an optimal pair $(\rho, \tilde{\rho})$ (see Theorem 4.3). There, we also prove that the support of the optimal stopping distribution $\tilde{\rho}$ is contained in the set where $J_\psi = \psi$.

It is important to note that the end time $\tau$ may not be determined uniquely for each given initial point $x$. In general, it is given by a distribution over $\mathbb{R}^+$, that we shall call stopping plan, by analogy with the randomized stopping times arising in the theory of stochastic processes, and which will be even more apparent in our companion paper [14]. It is a natural question to see when such stopping plan, when optimized, is actually a stopping time, i.e., given by a uniquely determined end time (still depending on the initial point $x$). In Section 5 we shall isolate conditions on the running cost (now considered as a Lagrangian) that allow for the determination of an optimal end time, given by a stopping plan, to be the hitting time $\tau$ of a graph $(y, s(y))$ of a function $s$ determined by an optimal dual function $\psi$ and its associated $J_\psi$. As we show in Theorem 5.5 this turned out to be satisfied under the conditions that $t \mapsto K(t, q, A)$ is either strictly increasing or strictly decreasing in time. In either case, the graph $(y, s(y))$ is, as expected, the free boundary of the variational
inequality (1.4). A similar approach to proving monotonicity of the solution \( J_\psi \) to the variational inequality was taken by one of the authors and A. Vladimirsky for an optimal stopping problem with probabilistic constraints in a discrete setting [17].

In Section 6, we provide a few examples mostly focusing on non-smooth one dimensional cases, while in the final section, we will compare our setting with the classical optimal transport problem, see e.g. [13]. Indeed, if we set the control set to be \( \mathbb{A} = S^{n-1} \subset \mathbb{R}^n \) and \( f(q, A) = A \), we can view the control problem as minimizing over trajectories with velocity bounded by 1. Consider now the cost is given only by a time penalty that is the derivative of a function \( g \) with \( g(0) = 0 \), \( K(t, q, A) = g'(t) \). The optimal trajectory connecting \( x \) and \( y \) is then the straight line that reaches \( y \) at time \( \tau_{x,y} = |y - x| \). Thus the cost associated to travelling from \( x \) to \( y \) in the optimal amount of time is therefore \( c(x, y) = g(|y - x|) \). The cases, considered above, when the running time increases (resp., decreases) correspond to when when \( g \) is convex (resp., concave).

2. The primal problem and its dual à la Monge-Kantorovich

In this section, we prove the no duality gap equation (1.2), namely, that \( V(\mu, \nu) = D(\mu, \nu) \), and that both optimization problems are attained under suitable conditions on the running cost \( K \), the function \( f \), and the source and target measures \( \mu \) and \( \nu \), such as the following.

**H0** The control set \( \mathbb{A} \) is a complete separable metric space, and the running cost \( K(t, q, A) \) is non-negative and continuous in \( t \), \( q \) and \( A \).

**H1** The velocity function \( f \) is Lipschitz with respect to \( q \) and continuous with respect to \( A \), and the span of \( \{ f(q, A); A \in \mathbb{A} \} \) is \( \mathbb{R}^n \) for each \( q \).

**H2** The distributions \( \mu \) and \( \nu \) have compact support in \( \mathbb{R}^n \).

Remark 2.1. These assumptions can be weakened, but doing so incurs technical difficulties and the weakening of some results. In particular, \( f \) could depend also on time for our results through Section 4, but we require that \( f \) is independent of time for the monotonicity properties of Section 5. In many models \( H1 \) will not be satisfied, in which case the feasibility of the primal problem (i.e. \( V(\mu, \nu) < \infty \)) may be difficult to determine, and it is unclear in general if the dual attainment of Section 3 holds. The assumption \( H2 \) is mainly for simplicity, and one may relax it to certain decay conditions of \( \mu \) and \( \nu \) as \( x \to \infty \).

Our primal optimal transport problem is to minimize the total transportation cost of a transportation plan, \( \pi \in \Pi(\mu, \nu) \), with its marginals prescribed as the initial and target distribution:

\[
V(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} c(x, y) \pi(dx, dy) \right\}.
\]

(2.1)

Remark 2.2. We shall use the following definition of the transportation cost:

\[
c(x, y) = \inf_{\tau, A} \left\{ \int_0^\tau K(t, \gamma(t), A(t)) dt; \gamma(t) = \int_0^t f(\gamma(s), A(s)) ds + x, \gamma(\tau) = y \right\},
\]

(2.2)
which is slightly different from the strong form of the dynamics mentioned in the introduction. This allows for non-differentiable paths, in particular if $A(\cdot)$ is not continuous, such as when it is alternating between two discrete values $\pm 1$.

**Lemma 2.3.** Given hypotheses $[\text{H0}]$ and $[\text{H1}]$ the optimal transportation cost $c$ of (2.2) is finite-valued and continuous. Moreover $y \mapsto c(\cdot, y)$ is Lipschitz continuous.

**Proof.** For a compact set $Z \subset \mathbb{R}^n$, $[\text{H1}]$ implies that for any $x$ and $y$ in $Z$ there is a controlled trajectory with $\gamma(0) = x$, $\gamma(\tau) = y$, and $\tau \leq C_1|x - y|$. Hypothesis $[\text{H0}]$ together with the compactness of $Z$, imply that the cost is bounded by $C_2\tau$. Thus $c(x, y) \leq \Lambda_Z|x - y| < \infty$.

We next consider a second pair $x'$ and $y'$ in $Z$, and let $(\tau, A)$ be within $\epsilon$ of attaining the cost $c(x', y)$ and consider a controlled-trajectory connecting $y$ and $y'$ in time $\tau_1 \leq C_1|y' - y|$. Then, the concatenated path gives

$$c(x', y') \leq c(x', y) + \epsilon + \Lambda_Z|y - y'|. \quad (2.3)$$

In the limit as $x' \to x$ and $y' \to y$ the trajectories converge to a trajectory transporting $x$ to $y$ by continuity of the solution to an ODE with respect to the initial condition, thanks to assumption $[\text{H1}]$. Thus

$$c(x', y') \leq c(x, y) + \epsilon + \Lambda_Z|y - y'| + \omega_Z(|x - x'|)$$

where $\omega_Z(r) \to 0$ as $r \to 0$. Since $x'$, $y'$, $\epsilon$ and $Z$ are arbitrary this shows that $c$ is continuous. Moreover, notice that (2.3) also shows, by letting $\epsilon \to 0$, that $c$ is Lipschitz-continuous with respect to $y$. \hfill \Box

Now, let us apply Kantorovich duality $[20]$, which consists of considering the maximization problem:

$$D_1(\mu, \nu) := \sup_{\psi, \phi \in C(\mathbb{R}^n); \psi \equiv (\psi \phi) \leq c} \left\{ \int_{\mathbb{R}^n} \psi(y)\nu(dy) - \int_{\mathbb{R}^n} \phi(x)\mu(dx) \right\}. \quad (2.4)$$

The following result is standard. See for example, Theorem 2.9 in $[20]$. The Lipschitz regularity of $\psi$ is verified by the representation of $\psi$ as the infimum of Lipschitz functions, $\psi(y) = \inf_x \{c(x, y) + \phi(x)\}$.

**Proposition 2.4.** Assume the hypotheses $[\text{H0}]$, $[\text{H1}]$ and $[\text{H2}]$. Then there is $\pi^* \in \Pi(\mu, \nu)$ that attains the optimal value $V(\mu, \nu)$ and there exist potentials $\psi^*$, $\phi^*$ that attain the value $D_1(\mu, \nu)$. Moreover, $\psi^*$ is Lipschitz continuous, and

$$V(\mu, \nu) = D_1(\mu, \nu), \quad \psi^*(y) - \phi^*(x) = c(x, y) \text{ for } \pi^*\text{-a.e. } (x, y). \quad (2.5)$$

Now, we relate the dual potentials with the dynamic programming principle: Given $\psi$ upper semi-continuous, we define the ‘value’ function $J_\psi$ by

$$J_\psi(t, q) = \sup_{\tau, A(\cdot)} \left\{ \psi(\gamma(\tau)) - \int_t^\tau K(s, \gamma(s), A(s))\,ds; \right\}.$$

Thus,
From this definition, we see that for any \( t_1 \leq t_2 \), and \( \gamma(s) = \int_{t_1}^{s} f(\gamma(r), A(r))dr + \gamma(t_1) \), \( A(\cdot) \in \mathcal{A} \),

\[
J_\psi(t_2, \gamma(t_2)) - J_\psi(t_1, \gamma(t_1)) \leq \int_{t_1}^{t_2} K(s, \gamma(s), A(s))ds.
\]

From the dynamic programming principle, \( J_\psi \) is the unique lower semi-continuous viscosity solution to the Hamilton-Jacobi-Bellman type quasi-variational inequality,

\[
\min \left\{ -\frac{\partial}{\partial t} J_\psi(t, q), -\psi(q) + J_\psi(t, q), H(t, q, \nabla J_\psi(t, q)) \right\} = 0,
\]

where the Hamiltonian is \( H(t, q, p) = \sup_{A \in \mathcal{A}} \left\{ p \cdot f(q, A) - K(t, q, A) \right\} \), and the negative sign is introduced to be consistent with the definition of viscosity solutions. The equation (2.6) plays the role of the ‘double convexification’ procedure of optimal transportation, along with the replacement of \( \psi \) by the maximal function satisfying \( \psi \leq J_\psi \), given by \( \psi(q) = \inf J_\psi(t, q) \). Recall that we defined the dual problem in the introduction as

\[
D(\mu, \nu) := \sup_{\psi} \left\{ \int_{\mathbb{R}^n} \psi(y)\nu(dy) - \int_{\mathbb{R}^n} J_\psi(0, x)\mu(dx) ; J_\psi \text{ solves } (2.8) \right\}.
\]

**Proposition 2.5.** Under the same assumptions of Proposition 2.4, we have that \( D(\mu, \nu) = D_1(\mu, \nu) = V(\mu, \nu). \)

Moreover, there is an optimal \( \psi^* \) such that the pair \( \psi^*(\cdot) = \inf_t J_{\psi^*}(t, \cdot) \), \( \phi^*(\cdot) = J_{\psi^*}(0, \cdot) \) is an optimizer for \( D_1(\mu, \nu) \), and hence satisfies (2.5).

**Proof.** Let \((\psi, \phi)\) be an optimized Kantorovich pair for \( D_1(\mu, \nu) \). In other words, we have that

\[
\psi(y) - \phi(x) \leq c(x, y) \quad \forall \, x, y \in \mathbb{R}^n,
\]

\[
\psi(y) = \inf_{x \in \mathbb{R}^n} c(x, y) + \phi(x) \quad \text{and} \quad \phi(x) = \sup_{y \in \mathbb{R}^n} \psi(y) - c(x, y).
\]

Now, from (2.6), the pair \((\psi, J_\psi(0, \cdot))\) obviously satisfies (2.10), thus is a Kantorovich pair for \( D_1(\mu, \nu) \). Since \((\psi, \phi)\) is an optimized Kantorovich pair, then \( J_\psi(0, \cdot) \leq \phi(\cdot) \).

Therefore, the integral in (2.4) for \( D_1(\mu, \nu) \) gets increased when replacing \( \phi \) with \( J_\psi \). Similarly, this integral increases when replacing \( \psi \) with \( \psi^*(\cdot) = \inf_t J_{\psi^*}(t, \cdot) \). This shows that \( D_1(\mu, \nu) = D(\mu, \nu) \) and that \((\psi^*, J_{\psi^*})\) is an optimized Kanotorovich pair. \( \square \)

3. An Eulerian Formulation

We now introduce the Eulerian formulation which has cost \( W(\mu, \nu) \), and its dual formulation with value \( E(\mu, \nu) \). We then prove that

\[
D(\mu, \nu) = E(\mu, \nu) \leq W(\mu, \nu) \leq V(\mu, \nu).
\]

Taking into accounts the duality proven in Section 2, all these values are then equal.

Two key notions are the density process \( \rho : \mathbb{R}^+ \rightarrow \mathcal{M}^+(\mathbb{R}^n \times \mathcal{A}) \) and the stopping distribution \( \hat{\rho} \in \mathcal{M}^+(\mathbb{R}^+ \times \mathbb{R}^n) \), where \( \mathcal{M}^+ \) is the cone of non-negative finite Radon
measures on the given metric space. We assume \( \rho \) is (weak*) measurable in the time variable. The constraint on trajectories becomes the continuity equation (1.5) for \( \rho \) and \( \tilde{\rho} \), which we express in the weak form:

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^+} w(t,q)\tilde{\rho}(dt,dq) = \int_{\mathbb{R}^n} w(0,x)\mu(dx) + \int_{\mathbb{R}_+} \int_{\mathbb{A}} \int_{\mathbb{R}^n} \left[ \frac{\partial}{\partial t} w(t,q) + f(q,A) \cdot \nabla w(t,q) \right] \rho(t,dq,dA)dt
\]

for all \( w \in C^1(\mathbb{R} \times \mathbb{R}^n) \). The target constraint is given by

\[
\int_{\mathbb{R}^n} u(y)\nu(dy) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^+} u(q)\tilde{\rho}(dt,dq) \quad \text{for all } u \in C(\mathbb{R}^n).
\]

If a pair \( (\rho, \tilde{\rho}) \) satisfies (3.1) and (3.2) then we say it is admissible.

We now consider the cost to be

\[
W(\mu, \nu) := \inf \left\{ \int_{\mathbb{R}^+} \int_{\mathbb{A}} \int_{\mathbb{R}^n} K(t,q,A)\rho(t,dq,dA)dt; \ (\rho, \tilde{\rho}) \text{ is admissible} \right\}.
\]

The following proposition associates to any transport plan, an approximating admissible density process.

**Proposition 3.1.** Suppose \( H0, H1, \text{and } H2 \) hold. If \( \pi \) is a transport plan in \( \Pi(\mu, \nu) \), then for any \( \epsilon > 0 \) there is an admissible density process and stopping distribution \( (\rho, \tilde{\rho}) \) such that

\[
\int_{\mathbb{R}^+} \int_{\mathbb{A}} \int_{\mathbb{R}^n} K(t,q,A)\rho(t,dq,dA)dt \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} c(x,y)\pi(dx,dy) + \epsilon
\]

As a consequence, \( W(\mu, \nu) \leq V(\mu, \nu) \).

**Proof.** We must first approximate the transport plan \( \pi \) by nearly optimal trajectories with \( (t, \gamma(\cdot), A(\cdot)) \) remaining in a compact set. Fix \( \epsilon > 0 \). By the definition of \( c(x,y) \) (2.2), for each \( x \) and \( y \), we can find \( A(\cdot; x, y) \) with values in a compact subset of \( \mathbb{A} \) and \( \tau(x,y) \) such that the trajectory

\[
\gamma(t; x, y) := \int_0^t f(\gamma(s; x, y), A(s; x, y))ds + x
\]

satisfies \( \gamma(\tau(x,y), x, y) = y \) and the cost satisfies

\[
\int_0^{\tau(x,y)} K(t, \gamma(t; x, y), A(t; x, y))dt \leq c(x,y) + \frac{\epsilon}{2}.
\]

By the continuity of \( c \) from Lemma 2.3, there is a neighborhood \( O \) of \( (x, y) \) such that for \( (x', y') \in O \) the trajectory beginning at \( x' \) with the same control policy \( A(\cdot; x, y) \) is within \( \epsilon \) of optimality, after concatenating with a trajectory to fix the endpoint at \( y' \). With this construction on a finite open cover of the compact supports of \( \mu \) and \( \nu \), we have that all trajectories and controls remain in a compact set and the times \( \tau(x,y) \) are uniformly bounded.
All of $\gamma(\cdot; x, y)$, $\tau(x, y)$ and $A(\cdot; x, y)$ can be chosen to be measurable in $x$ with respect to $\mu$ and measurable in $y$ with respect to $\nu$ and are thus measurable with respect to $\pi$. Moreover, from $[H1]$ the trajectories $\gamma(\cdot; x, y)$ are Lipschitz continuous in time on a bounded time interval. We now define $\tilde{\rho}$ (by the Riesz representation theorem) such that

$$
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \beta(t, q)\tilde{\rho}(dt, dq) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \beta(\tau(x, y), y)\pi(dx, dy)
$$

for all continuous and compactly supported $\beta$. We also define $\rho$ at time $t$ in such a way that

$$
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \eta(q, A)\rho(t, dq, dA) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} 1\{t < \tau(x, y)\} \eta(\gamma(t; x, y), A(t; x, y))\pi(dx, dy)
$$

for all continuous and compactly supported $\eta$. Since $\pi$ is nonnegative, we see that $\rho, \tilde{\rho}$ are also nonnegative measures, and by construction, $\tilde{\rho}$ satisfies (3.2) and $\rho$ is measurable in time. We consider a smooth test function $w : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$, and (dropping the notational dependence of $\tau, A(\cdot), \gamma(\cdot)$ on $x, y$), we see that

$$
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left[ -\frac{\partial}{\partial t}w(t, q) - f(q, A) \cdot \nabla w(t, q) \right] \rho(t, dq, dA)dt
$$

is our version of the continuity equation (3.1). Note that the first two equations above are possible as $\gamma$ is Lipschitz in $t$ so $t \mapsto w(t, \gamma(t))$ is $W^{1,\infty}$, and $t \mapsto \rho(t, \cdot, \cdot)$ is measurable. Therefore, the pair $(\rho, \tilde{\rho})$ is an admissible density process. Now, a similar calculation shows that for the corresponding costs, we have:

$$
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K(t, q, A)\rho(t, dq, dA)dt = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} 1\{t < \tau\} K(t, \gamma(t), A(t))\pi(dx, dy)dt
$$

$$
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left[ \int_0^\tau K(t, \gamma(t), A(t))dt \right] \pi(dx, dy)
$$

$$
\leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} c(x, y)\pi(dx, dy) + \epsilon.
$$

As $\epsilon$ is arbitrary, we get that $W(\mu, \nu) \leq V(\mu, \nu)$. \hfill \Box

From the duality principle of linear programming, we can formulate the dual problem to the Eulerian formulation as follows:

$$
\text{(3.4)} \quad \text{Maximize} \quad \int_{\mathbb{R}^n} \psi(y)\nu(dy) - \int_{\mathbb{R}^n} J(0, x)\mu(dx),
$$
over $\psi \in C(\mathbb{R}^n)$ and $J \in C^1(\mathbb{R} \times \mathbb{R}^n)$ subject to

\begin{equation}
\frac{\partial}{\partial t} J(t, q) + f(q, A) \cdot \nabla J(t, q) \leq K(t, q, A), \quad \text{for all } (t, q, A) \in \mathbb{R}^+ \times \mathbb{R}^n \times A. \tag{3.5}
\end{equation}

Notice that $\psi$ means that $J$ is a supersolution to (2.8).

We let $E(\mu, \nu)$ denote the optimal value of this dual problem.

**Proposition 3.2.** For all $(J, \psi) \in C^1(\mathbb{R}^+ \times \mathbb{R}^n) \times C(\mathbb{R}^n)$ that satisfy (3.5), and any $(\rho, \tilde{\rho})$ that satisfy (3.1) and (3.2), we have

\begin{equation}
\int_{\mathbb{R}^n} \psi(y)\nu(dy) - \int_{\mathbb{R}^n} J(0, x)\mu(dx) \leq \int_{\mathbb{R}^+} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K(t, q, A)\rho(t, dq, dA)dt. \tag{3.6}
\end{equation}

In particular, $E(\mu, \nu) \leq W(\mu, \nu)$.

**Proof.** This weak duality inequality is immediate from plugging in $J$ and $\psi$ as test functions in (3.1) and (3.2). \hfill \Box

This dual problem (3.4)-(3.5) is indeed equivalent to the duality studied in Section 2, as we can reduce (3.5) to (2.8). To see this, consider the pointwise minimum of $J(t, q)$ amongst all supersolutions and the pointwise maximum of $\psi$ among all functions less than or equal to $J$. These yield a solution to (2.8).

**Proposition 3.3.** Given $\psi \in C(\mathbb{R}^n)$ uniformly bounded, then the infimum over all $J$ that satisfy (3.5) is attained by the function $J_\psi$ defined in (2.6). Equivalently, this is the unique viscosity solution to (2.8), and therefore, the dual value satisfies $E(\mu, \nu) = D(\mu, \nu)$.

**Proof.** Recall the definition of $J_\psi$ from (2.6), which is the unique viscosity solution to (2.8). Therefore, by the comparison principle, if $J$ satisfies (3.5), then $J(t, q) \geq J_\psi(t, q)$ and we have $E(\mu, \nu) \leq D(\mu, \nu)$. Perron's method then yields that $J_\psi$ is the infimum of all supersolutions $J$; see [2]. Comparing with the definition of $D(\mu, \nu)$ in (2.9), this proves that the supremum of the value over $J$ is attained at $J_\psi$ and that $E(\mu, \nu) = D(\mu, \nu)$. \hfill \Box

Now we can verify that the problems are indeed equivalent.

**Theorem 3.4.** Given the hypotheses $\mathbf{H0}$, $\mathbf{H1}$, $\mathbf{H2}$ then the following hold:

\begin{equation}
D(\mu, \nu) = E(\mu, \nu) = W(\mu, \nu) = V(\mu, \nu). \tag{3.7}
\end{equation}

**Proof.** Proposition 3.1 implies that $W(\mu, \nu) \leq V(\mu, \nu)$ and Proposition 3.2 implies $E(\mu, \nu) \leq W(\mu, \nu)$. Proposition 3.3 yields that $E(\mu, \nu) = D(\mu, \nu)$. The rest follows from the fact that $V(\mu, \nu) = D(\mu, \nu)$ shown in Section 2. \hfill \Box

4. **Necessary Optimality Conditions**

The next stage of our analysis is to relate the Hamilton-Jacobi-Bellman quasi-variational inequality (2.8) back into optimality criteria for paths involved in the primal problem.
Theorem 4.1. Suppose $[\text{H0}]$, $[\text{H1}]$, and $[\text{H2}]$ hold and also that $(q,A) \mapsto \nabla f$ is continuous as is $(t,q,A) \mapsto \nabla K$. We let $\pi, \psi$ and $J_\psi$ be optimal for (2.1) and (2.3), resp., and suppose that the optimal control problem (2.2) for the cost $c$ is attained for $\pi$-a.e. $(x,y)$ at $\tau^{x,y}$ and $A^{x,y}(\cdot)$ that are measurable with respect to $x$ and $y$, and for a.e. $(x,y)$, $A^{x,y}(\cdot)$ is piecewise continuous in time. We let $\gamma^{x,y}(\cdot)$ denote the corresponding trajectories. Then the following hold:

1. If $\gamma(\cdot), \tau$ and $A(\cdot)$ attain the optimal cost $c(x,y)$ with $\gamma(0) = x$ and $\gamma(\tau) = y$, then $J_\psi(\tau, y) = \psi(y)$.

2. There is $p \in C([0,\tau];\mathbb{R}^n)$ that solves, at times when $A(\cdot)$ is continuous,

$$p(t) = -p(t)^T \nabla f(t, \gamma(t), A(t)) + \nabla K(t, \gamma(t), A(t)),$$

and such that for a.e. $t \in [0,\tau]$

$$H(t, \gamma(t), p(t)) = p(t) \cdot f(\gamma(t), A(t)) - K(t, \gamma(t), A(t)),$$

while the following transversality condition holds:

$$H(t, \gamma(t), p(t)) = 0.$$

3. If $\psi$ is differentiable at $y$, then $p(\tau) = \nabla \psi(y)$, and if $J_\psi$ is differentiable at $(t, \gamma(t))$, then $p(t) = \nabla J_\psi(t, \gamma(t))$.

Proof. Suppose $\gamma^{x,y}$, $A^{x,y}$, and $\tau^{x,y}$ attain the infimum in the expression of $c(x,y)$. Then, for $\pi$-almost every $(x,y)$ and $t \leq \tau^{x,y}$,

$$J_\psi(t, \gamma^{x,y}(t)) = \psi(y) - \int_{t}^{\tau^{x,y}} K(s, \gamma^{x,y}(s), A^{x,y}(s))ds,$$

from which follows that $J_\psi(\tau^{x,y}, y) = \psi(y)$.

The Pontryagin maximum principle [18] implies that if $\gamma$, $\tau$, $A$ minimize the free end time optimal control problem then there is $p$ that satisfies (4.1), (4.2) and (4.3). For almost every $(x,y)$ in the support of $\pi$, then $\gamma$, $\tau$ and $A$ also minimize

$$J_\psi(t, \gamma(t)) - \psi(\gamma(\tau)) + \int_{t}^{\tau} K(s, \gamma(s), A(s))ds$$

amongst trajectories with free position at times $t$ and $\tau$. It then also follows from the Pontryagin maximum principle and differentiability of $\psi(y)$ and $J_\psi(t, \gamma(t))$ that $p(\tau) = \nabla \psi(y)$ and $p(t) = \nabla J(t, \gamma(t))$. \qed

Remark 4.2. Since the optimal control is generally not attained at a measurable map, we could compactify the problem by using the notion of Young measures, which are weak* measurable maps $t \mapsto \alpha_t \in \mathcal{P}(\mathbb{A})$ (probability measures). As an example consider the case where $\mathbb{A}$ is a finite set, and the optimal direction does not align with $f(q,A)$ for any of these points. A nearly optimal control will rapidly oscillating between points in $\mathbb{A}$ and the optimal control is obtained by a Young measure in $\mathcal{P}(\mathbb{A})$, which may be interpreted as a randomized choice of control at each instant. In this case, we consider controlled trajectories of the form

$$\gamma(t) = \int_{0}^{t} \int_{\mathbb{A}} g(\gamma(s), A) \alpha_s(dA)ds + x.$$
This formulation provides an intermediary step between the primal problem and its Eulerian formulation. However, we choose instead to work with the weak Eulerian formulation of Section 3, similar to what is done in [11], and we later work with the stronger assumptions that the cost is given by a convex Lagrangian.

We now introduce hypotheses that will imply attainment of the Eulerian formulation.

**H3** The set of controls $A \subset \mathbb{R}^m$ is closed, $f(q,A)$ is uniformly bounded in $q$ for each $A$, $|f(q,A)| \leq \Gamma_0(1 + |A|)$, and $K(t,q,A)$ satisfies
\[
\liminf_{A \to \infty} \frac{K(t,q,A)}{|A|} = \infty.
\]

**H4** The following coercivity in time holds:
\[
\liminf_{T \to \infty} \int_0^T \inf_{q,A} K(t,q,A) \, dt = \infty.
\]

We will abbreviate $H_0$ - $H_4$ to include hypotheses $H_0$, $H_1$, $H_2$, $H_3$ and $H_4$.

**Theorem 4.3.** Suppose $H_0$ - $H_4$ hold, then the Eulerian formulation is minimized by some pair $(\rho, \tilde{\rho})$. For optimal $(\psi, J_\psi)$ (cf. Proposition 2.4 and Proposition 3.3):

1. $J_\psi(t,q) = \psi(q)$ for $\tilde{\rho}$ almost every $(t,q)$;
2. If $J_\psi \in C^1(\mathbb{R}^+ \times \mathbb{R}^n)$, then for almost every $t$, and $\rho(t,\cdot,\cdot)$ almost every $(q,A)$, $f(q,A) \in D_pH(t,q,\nabla J_\psi(t,q))$, where $D_pH$ denotes the subdifferential of $H$ with respect to $p$.

**Proof.** The attainment of the minimum at $(\rho, \tilde{\rho})$ is by a compactness argument using coercivity and tightness. We fix a cost $M \in \mathbb{R}$ and show that the set of $(\rho, \tilde{\rho})$ with cost less than $M$ is tight. First, let $\epsilon > 0$ be arbitrary. Given assumption $H_4$ there is $T$ such that
\[
\int_0^T \inf_{q,A} K(t,q,A) \, dt > M/\epsilon.
\]
If $\tilde{\rho}([T, \infty) \times \mathbb{R}^n) \geq \epsilon$ then $\rho(T,\mathbb{R}^n,\mathbb{A}) \geq \epsilon$, because
\[
\int_0^T \int_{\mathbb{R}^n} \tilde{\rho}(dq,dt) + \int_\mathbb{A} \int_{\mathbb{R}^n} \rho(t,dq,dA) = 1,
\]
which implies that the cost of $(\rho, \tilde{\rho})$ is greater than $M$. Similarly, by $H_3$ there is $R_1$ such that $K(t,q,A) > |A|M(1 + \Gamma_0)/\epsilon$ if $|A| > R_1$, which implies
\[
\int_{\mathbb{R}^+} \int_{|A| > R_1} \int_{\mathbb{R}^n} |A|(1 + \Gamma_0)\rho(t,dq,dA) \, dt < \epsilon.
\]
We let $\Gamma_1 = \sup_{q \in \mathbb{R}^n, |A| < R_1} |f(q,A)|$. The amount of mass at time $t$ outside a ball of radius $R_2(t) = 2(t\Gamma_1 + D + 1)$ is less than $\epsilon$, where the support of $\mu$ is contained in a ball of radius $D$. This can be shown from (3.1) by considering a test function, $w(t,q)$, that is zero on a ball of radius $R_2(t)/2$, one outside of a ball of radius $R_2(t)$, nonnegative, and satisfies
\[
\frac{\partial}{\partial t} w(t,q) \leq -|\nabla w(t,q)|\Gamma_1,
\]
and $|\nabla w(t, q)| \leq 1$ for $t \leq T$ (consider a smooth approximation to the piecewise interpolation from the function that is 0 for $|q| \leq R_2(t)/2$ and 1 for $|q| \geq R_2(t)$). Then (3.1) implies

$$
\int_{|q| > R_2(T)} \int_0^T \bar{\rho}(dt, dq) + \int_A \int_{|q| > R_2(T)} \rho(T, dq, dA)
\leq \int_0^T \int_A \int_{\mathbb{R}^n} \left[ \frac{\partial}{\partial t} w(t, q) - f(q, a) \cdot \nabla w(t, q) \right] \rho(t, dq, dA) dt
\leq \int_0^T \int_{|A| > R_1} \int_{\mathbb{R}^n} \left[ |\nabla w| (1 + \Gamma_0 |A|) \right] \rho(t, dq, dA) dt < \epsilon.
$$

These estimates show that the mass outside of $[0, T] \times B_{R_1} \times B_{R_2(T)}$ is less than $\epsilon$. This tightness ensures that a subsequence of a minimizing sequence of $(\rho^i, \bar{\rho}^i)$ converges in the weak* topology to some $(\rho, \bar{\rho})$ that attains the minimum cost.

We now let $\psi$ be optimal and consider a sequence $J^i \in C^1(\mathbb{R}^+ \times \mathbb{R}^n)$ that satisfy (3.5) and converge uniformly to $J_\psi$. Then using $J^i$ and $\psi$ as test functions for (3.1) and (3.2) with optimal $(\rho, \bar{\rho})$, we have that

$$
\int_{\mathbb{R}^n} \psi(y) \nu(dy) - \int_{\mathbb{R}^n} J^i(0, x) \mu(dx)
\leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^+} [\psi(q) - J^i(t, q)] \bar{\rho}(dt, dq) + \int_{\mathbb{R}^+} \int_A \int_{\mathbb{R}^n} K(t, q, A) \rho(t, dq, dA) dt.
$$

In the limit as $i \to 0$ this shows that $0 \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^+} [\psi(q) - J_\psi(t, q)] \bar{\rho}(dt, dq)$, which along with $\psi(q) \leq J_\psi(t, q)$ proves statement (1).

If $J_\psi$ is in fact $C^1$, then using $J_\psi$ as a test function for (3.1) as above also implies that

$$
\frac{\partial}{\partial t} J_\psi(t, q) + f(q, A) \cdot \nabla J_\psi(t, q) - K(t, q, A) = 0
$$

for almost every $t$ and $\rho$ almost every $q$ and $A$. It follows that at these points $A$ maximizes $f(q, A) \cdot \nabla J_\psi(t, q) - K(t, q, A)$, which occurs at $f(q, A) \in D_p H(t, q, \nabla J_\psi(t, q))$.

□

Remark 4.4. While $(\rho, \bar{\rho})$ are convenient for the existence and weak optimality criteria, we can now move to a more familiar Eulerian perspective with the density defined informally by

$$
\eta(t, q) = \int_A \rho(t, q, dA).
$$

The statement $f \in D_p H$ that we have proven in Theorem 4.3 can now be reinterpreted as $\eta$ satisfies a weak continuity inequality, formally equivalent to

$$
\frac{\partial}{\partial t} \eta(t, q) \leq -\text{div} \left( D_p H(t, q, \nabla J_\psi(t, q)) \eta(t, q) \right),
$$

where we have replaced the stopping distribution $\bar{\rho}$ with an inequality.
5. When optimal end times are hitting times of a barrier

In this section, we study the structure of the solutions of the dual problem under different cases of time dependence as well as the implications for the optimal flows. In particular, our analysis will focus on the viscosity solutions to (2.8), and how their properties translate to the primal problem. We consider three cases:

**TS:** The time-stationary case: i.e., when \( K \) is independent of time.

**TC:** The time-compounded case: i.e., when \( K \) is strictly increasing in time.

**TD:** The time-discounted case: i.e., when \( K \) is strictly decreasing in time.

**Proposition 5.1.** We suppose \( \mathbf{H}_0 \) and \( \mathbf{H}_1 \) and that the pair \( (J_\psi, \psi) \) solve (2.8) and \( \psi(q) = \inf_{t \in \mathbb{R}^+} \{J_\psi(t, q)\} \). We consider the three cases of time dependence. We show the following inequalities hold for certain values of \( t \) in each case:

(5.1) 
\[-H(t, q, \nabla J_\psi(t, q)) \leq 0\]

and

(5.2) 
\[-H(t, q, \nabla \psi(q)) \geq 0.\]

1. In the time-stationary case **[TS]**, the dual solution \( J_\psi \) is constant in time, \( J_\psi(t, q) = \psi(q) \), and (5.2) holds for all \( q \).

2. In the time-compounded case **[TC]**, the solution \( J_\psi \) is non-increasing, the free boundary \( s(q) = \inf\{t; J_\psi(t, q) = \psi(q)\} \) is lower semi-continuous, and Equation (5.2) holds for \( t \geq s(q) \) and (5.1) holds for \( t < s(q) \).

3. In the time-discounted case **[TD]**, the solution \( J_\psi \) is non-decreasing, the free boundary \( s(q) = \sup\{t; J_\psi(t, q) = \psi(q)\} \) is upper semi-continuous, and Equation (5.2) holds for \( t \leq s(q) \) and (5.1) holds for \( t > s(q) \).

Moreover, in both cases **[TC]** and **[TD]** it also holds that if \( 0 < s(q) < \infty \) then

(5.3) 
\[-H(s(q), q, \nabla \psi(q)) = 0.\]

**Proof.** The stationary case **[TS]** follows immediately from the uniqueness, given \( \psi \), of viscosity solutions to (2.8) and the symmetry as \( t \mapsto t + \Delta t \). A consequence of (2.8) is that

\[-H(t, q, \nabla J_\psi(t, q)) \leq 0\]

and since \( J_\psi = \psi \), equation (5.2) holds.

For the time-compounded case **[TC]** we let \( \tilde{J}(t, q) = \sup\{J_\psi(s, q); s \geq t\} \), which clearly satisfies \( \tilde{J} \geq J_\psi \) and \( \tilde{J} \) is non-increasing. To prove \( \tilde{J} \leq J_\psi \), we show that \( \tilde{J} \) is a subsolution to (2.8), then use the comparison principle. The subsolution property follows from considering two cases: If \( J_\psi(r, q) = \psi(q) \) for all \( r \geq t \), then \( \tilde{J}(t, q) = J_\psi(t, q) = \psi(q) \) and it is obviously a subsolution; Otherwise the supremum in the definition of \( \tilde{J} \) is attained at a time \( r^* \geq t \) where \( J_\psi(r^*, q) > \psi(q) \). We then consider a test function \( w \) that touches \( \tilde{J} \) from above at \( (t, q) \), and let \( \Delta t = r^* - t \) and \( \tilde{w}(r, q) = w(r + \Delta t, q) \). Then \( \tilde{w} \) touches \( J_\psi \) from above at \( (r^*, q) \). It follows from
the definition of a viscosity subsolution that
\[
0 \geq - \frac{\partial}{\partial s} \hat{w}(r, q) - H(r, q, \nabla \hat{w}(r, q))
\]
\[
\geq - \frac{\partial}{\partial t} w(t, q) - H(t, q, \nabla w(t, q)),
\]
where we have used Assumption [TC] that \( K \) is increasing and therefore the Hamiltonian is decreasing in time. This proves that \( \tilde{J} \) is a viscosity subsolution to (2.8). We conclude that \( \tilde{J} = J_{\psi} \), and that \( J_{\psi} \) is non-increasing, which also proves (5.1) when \( J_{\psi}(t, q) > \psi(q) \). Equation (5.2) follows immediately from (2.8), in the case that \( J_{\psi} = \psi \) and hence \( \psi \) touches \( J_{\psi} \) from below thus \(-H(t, q, \nabla \psi(q)) \geq -H(t, q, \nabla J_{\psi}(t, q)) \geq 0 \), where the last inequality follows from that \( J_{\psi} \) is constant in time for \( t \geq s(q) \) but still satisfies (2.8).

The case of [TD] follows exactly the same argument replacing the supremums by inﬁniums, subsolutions by supersolutions, and touching above by touching below.

That \( s \) is lower (upper) semi-continuous follows from being the inf (sup) over a closed set. For the ﬁnal equation (5.3), that \(-H(s(q), q, \nabla \psi(q)) \geq 0 \) follows from (5.2) with \( t = s(q) \). For the other inequality, suppose to the contrary that there is a smooth test function \( g \) with \( g \geq \psi, g(\overline{q}) = \psi(\overline{q}), \) and \(-H(s(\overline{q}), \overline{q}, \nabla g(\overline{q})) > 0 \). Then we can construct a function \( G(t, q) = g(q) + f(t) \), which satisﬁes \( G \geq \psi, \)
\[-\frac{\partial}{\partial t} G(t, q) - H(t, q, \nabla G(t, q)) \geq 0 \]
and \( G(t, \overline{q}) = \psi(\overline{q}) \) for some \( 0 \leq t < s(\overline{q}) \) (or \( t > s(\overline{q}) \)) in the case of [TD]. In particular, for the case of [TC], we choose some \( \delta > 0 \) and \( f(t) \) that is zero for \( t > s(q) - \delta \) and \( f'(t) \) is sufﬁciently negative for \( t < s(q) - \delta \) to ensure that \(-f'(t) - H(t, q, g(q)) > 0 \). Since \( J_{\psi} \) is the inﬁmum over all such supersolutions, we have that \( J_{\psi} \leq G \), however, this contradicts the deﬁnition of \( s \), which implies that \( J_{\psi}(t, \overline{q}) > \psi(\overline{q}) \) for \( t < s(\overline{q}) \) (or \( t > s(\overline{q}) \)) for [TD].

This next lemma encodes the simple properties that will allow us to prove that the stopping plans are in fact given by a single stopping time. First, we consider that stopping times are given by a function of the Hamiltonian trajectories \((\gamma, p) \mapsto \tau \). In Theorem 5.3 this is used in the setting where the trajectories are uniquely determined by their initial condition, and the initial costate condition is determined by \( \nabla J_{\psi}(0, \cdot) \).

The second setting we consider is where the trajectories are determined by their initial condition, and the initial costate condition is determined by \( \nabla J_{\psi}(0, \cdot) \).

**Lemma 5.2.** Given the same hypotheses as Theorem 4.1 assuming either [TC] or [TD]:

1. Suppose that \( \gamma(\cdot), A(\cdot) \) are optimizers of \( c(x, y) \) with \( x \neq y \), and \( p(\cdot) \) is a costate warranted by the Pontryagin maximum principle. Then along the trajectory the Hamiltonian, \( H(t, \gamma(t), p(t)) \), strictly decreases the case of [TC] or strictly increases in the case of [TD]. In particular, for each Hamiltonian trajectory \((\gamma, p)\) the end time \( \tau \) is determined by the transversality condition (4.3).
Suppose that \( \gamma, A \) are instead optimizers of \( J(0, x) \), i.e. the free end time problem with terminal cost \( -\psi(\cdot) \). If \( \gamma \) ends at \( y \) where \( \psi \) differentiable with end time \( \tau > 0 \), then the end time is \( \tau = s(y) \), i.e. the unique time such that \( H(t, y, \nabla \psi(y)) = 0 \).

Proof. A consequence of the Pontryagin maximum principle is that for almost every \( t < \tau \),

\[
\frac{d}{dt} H(t, \gamma(t), p(t)) = -\frac{\partial}{\partial t} K(t, \gamma(t), p(t)) + [p^T(t)\nabla f(\gamma(t), A(t)) - \nabla k(t, \gamma(t), A(t))]\dot{\gamma}(t) + \dot{p}(t)f(\gamma(t), A(t)),
\]

and the last line is zero from the costate equation \( (4.1) \). Given \( [TC] \), \( -\frac{\partial}{\partial t} K < 0 \) and \( H \) decreases along the trajectory, and given \( [TD] \) the opposite holds. Clearly, this implies that \( H(t, \gamma(t), p(t)) = 0 \) can occur for at most one time.

In the second case, \( p(\tau) = \nabla \psi(\gamma(\tau)) \) from the Pontryagin maximum principle as used in Proposition \( 5.1 \) and if \( \gamma(\tau) = y \), the equation \( H(\tau, y, \nabla \psi(y)) = 0 \) uniquely determines \( \tau = s(y) \). \( \square \)

Remark 5.3. The Pontryagin maximum principle can also be stated for suitable sub/super derivatives of \( \psi \) and \( J_\psi \). The lack of differentiability of \( \psi \) at \( y \) corresponds to the existence of multiple optimal trajectories terminating at \( y \) with possibly different end times satisfying \( H(\tau, y, \beta) = 0 \) for \( \beta \in \partial \psi(y) \). Similarly, lack of differentiability of \( J_\psi \) at \( (t, q) \) corresponds to multiple optimal trajectories emitting from \( \gamma(t) = q \) with different \( p(t) \) values.

Remark 5.4. The time monotonicity of \( J_\psi \) of Proposition \( 5.1 \) and characterization of stopping times of Lemma \( 5.2 \) is closely related to the geometric pathwise monotonicity of \( [3] \). Beiglböck et al. consider stop-go pairs of trajectories \( (\gamma_1(\cdot), \gamma_2(\cdot)) \) where \( \gamma_1(t_1) = \gamma_2(t_2) = y \) for \( t_1 < t_2 \). In the case of \( [TD] \) if \( \gamma_1(\cdot) \) stops at \( t_1 \), then \( \gamma_2(\cdot) \) must also stop at \( t_2 \). Another proof of this principle uses the dynamic programming principle. Consider that \( \gamma_2(\cdot) \) is optimal and continues until time \( \tau_2 > t_2 \), then

\[
J_\psi(t_2, y) = \psi(y) - \int_{t_2}^{\tau_2} L(t, \gamma_2(t)) dt < \psi(y) - \int_{t_1}^{\tau_2-t_2+t_1} L(t, \gamma_2(t_2 - t_1)) dt \leq J_\psi(t_1, y).
\]

This shows that if \( \gamma_2(\cdot) \) continues at time \( t_2 \), then \( \gamma_1(\cdot) \) must also continue at time \( t_1 \) because \( J_\psi(t_1, y) > \psi(y) \), or equivalently if \( \gamma_1(\cdot) \) stops at time \( t_1 \), then \( \gamma_2(\cdot) \) must also stop at time \( t_2 \).

We will now make some assumptions directly on the Hamiltonian that are sufficient for our analysis.
\textbf{H5} We also assume that the Hamiltonian is smooth and satisfies $H(t,q,p) \sim |p|^\beta$ uniformly in $t$ and $q$ for $1 < \beta < \infty$.

\textbf{H6} We assume that $p \cdot f(q,A) - K(t,q,A)$ admits a unique maximizer $A^*$ for all $t$, $q$, and $p$, and $(t,q,p) \mapsto A^*$ is continuous.

We recall the notion of a Hamiltonian flow $\chi_H : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ as $\chi_H(t_1,t_2,q,\beta) = (\gamma(t_2),p(t_2))$ where $\gamma$ and $p$ solve

$$\begin{align*}
\dot{\gamma}(t) &= D_p H(t,\gamma(t),p(t)) \\
\dot{p}(t) &= - D_q H(t,\gamma(t),p(t)),
\end{align*}$$

with $\gamma(t_1) = q$ and $p(t_1) = \beta$. In general this Hamiltonian flow may be set-valued, but under hypothesis H5 the solutions are unique. We let

$$\chi_{H,\psi}(t,x) = \pi^* \chi_H(0,t,x,\nabla J_\psi(0,x))$$

be the spatial flow with the initial momentum determined by $\nabla J_\psi(0,x)$. Here $\pi^* : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the canonical projection. We let $\tau(x)$ denote the unique stopping time determined by the trajectory $(\gamma,p)$ cf. Lemma \cite{lemma5.2}

We also consider the backward flow with terminal end time given by $s(y)$ where for $t < s(y)$ we let

$$\chi_{H,s,s}(t,y) = \pi^* \chi_H(s(y),t,y,\nabla \psi(y))$$

be the backwards flow with terminal condition $\gamma(s(y)) = y$ and $p(s(y)) = \nabla \psi(y)$.

We note the relationship of the end time $\tau(x)$, determined from the initial condition, and the end time $s(y)$ given by the free boundary to \cite{2.8}, which can be expressed in the following two ways by inverting the Hamiltonian flow:

$$s \left( \chi_{H,\psi} (\tau(x),x) \right) = \tau(x),$$

$$s(y) = \tau \left( \chi_{H,\psi,s}(0,y) \right).$$

\textbf{Theorem 5.5.} Suppose hypotheses H6, H4 hold along with H5 and either \cite{TC} or \cite{TD}. Under these assumptions the Hamiltonian flow $\chi_H(t_1,t_2,q,\beta)$ is everywhere single-valued and $J_\psi$ is Lipschitz continuous.

1. If $\mu$ is absolutely continuous with respect to Lebesgue measure and has disjoint support from $\nu$, then the optimal transport plan is unique and is given by $x \mapsto \chi_{H,\psi}(\tau(x),x)$, that is

$$V(\mu,\nu) = \int_{\mathbb{R}^n} c(x,\chi_{H,\psi}(\tau(x),x)) \mu(dx),$$

where the end time $\tau(x)$ is almost everywhere the unique time such that $H(t,\gamma(t),p(t)) = 0$, where $\gamma(0) = x$ and $p(0) = \nabla J_\psi(0,x)$.

If H6, then the pair $(\rho,\hat{p})$ corresponding to the embedding of the optimal transport plan into the Eulerian formulation, cf. Proposition \cite{3.1} is a minimizer of the Eulerian formulation.

2. If $\nu$ is absolutely continuous with respect to Lebesgue measure and has disjoint support from $\mu$, then the optimal transport plan is unique and given by...
Proof. We have verified that $\psi$ is Lipschitz continuous in Proposition 2.4. That $J_\psi$ is Lipschitz continuous follows from hypotheses $H_5$ and the uniform estimates of [8].

The Hamiltonian equations have a unique solution for the initial (or terminal) value problem using the smoothness assumption of $H_5$ thus the Hamiltonian flow is well defined for each $(t_1, t_2, q, \beta)$ and continuous. By Lemma 5.2 and almost everywhere differentiability of $\psi$ and $J_\psi$, the stopping times $\tau(x)$ and $s(y)$ are well defined almost everywhere and measurable, so the maps $x \mapsto \chi_{H,J_\psi}(\tau(x), x)$ and $y \mapsto \chi_{H,J_\psi,s}(0, y)$ exist.

We now invoke Theorem 4.1 to show that, if $\pi$, $\psi$ and $J_\psi$ are optimal, then the optimal trajectories $\gamma_{x,y}^\pi$ solve the Hamiltonian equations with the costate satisfying $p_{x,y}^\pi(0) = \nabla J_\psi(0, x)$ and $p_{x,y}^\pi(x,y) = \nabla \psi(y)$ at points of differentiability of $J_\phi$ and $\psi$. Hence, if $\mu$ is absolutely continuous then $\gamma_{x,y}^\mu, p_{x,y}^\mu$ are determined by $\chi_H(0, \cdot, x, \nabla J_\phi(0, x))$ for almost every $x$ and if $\nu$ is absolutely continuous they are determined by $\chi_H(s(y), \cdot, y, \nabla \psi(y))$ for almost every $y$. It follows in either case that $\pi$ is the unique transport plan supported on the graph $(x, \chi_{H,J_\psi}(\tau(x), x))$ or $(\chi_{H,J_\psi,s}(0, y), y)$.

We now suppose $H_6$ so that for any Hamiltonian trajectory $(\gamma, p)$ there is a unique control $A(t)$. By Proposition 3.1, $\chi_{H,J_\psi}, \tau, A$ embed into a density process and stopping distribution with cost $V(\mu, \nu)$. We have proven in Theorem 3.4 that $V(\mu, \nu) = W(\mu, \nu)$ so $(\rho, \tilde{\rho})$ are optimal. Furthermore, $\tilde{\rho}$ has support in the set $(s(y), y)$ and is given by $\delta_{s(y)}(dt)\nu(dy)$. \qed

Remark 5.6. We also conjecture that under the assumptions of Theorem 5.5 the minimizer of the Eulerian formulation is unique. This holds if $J_\psi \in C^1(\mathbb{R}^+ \times \mathbb{R}^n)$ by uniqueness of the transport equation implied by Theorem 4.3. We expect that an argument can be made to handle the case that $J_\psi$ is Lipschitz by using generalized gradients of $J_\psi$, but we leave this to another work.

6. One dimensional examples

To illustrate our analysis in a very simple setting, we now consider a one-dimensional problem where the control set $A$ consists of two options: travelling to the left or to the right with constant speed. We let the cost be a function of time only, namely as the derivative of a function $g$ with $g(0) = 0$ and $g'(t) \geq 0$. The Hamiltonian is then,

$$H(t, q, p) = \max\{ \pm p - g'(t) \} = |p| - g'(t).$$

The optimal control is to travel left if $p < 0$ and to the right if $p > 0$. Since the Hamiltonian does not depend on $q$, the costate is constant along trajectories, which will therefore be straight lines with slope 1, illustrated in Figure 1. It is then easy to see that the cost $c(x, y) = g(|y - x|)$, hence the corresponding Monge map $Y(x)$ referred to by Theorem 5.5 can also be obtained from classical results; see e.g., [13]. It is however easy to compute in this one dimensional case.
Figure 1. The optimal trajectories in green, which stop upon hitting the free boundary $s$ in blue. The shaded region shows where $\psi(q) = J_\psi(t, q)$.

To illustrate how nonsmooth solutions can occur and to explicitly compute the free boundary $s(y)$, and the optimal dual functions $\psi, J_\psi$, we assume the initial distribution to have uniform density on a connected interval, $\frac{d\mu}{dx} = \frac{1}{2}[\frac{-1}{2}, \frac{1}{2}]$, while the target is $\frac{d\nu}{dy} = \frac{1}{2}[\frac{-2}{2}, \frac{-1}{2}] + \frac{1}{2}[\frac{1}{2}, 2]$.

A. If $g$ is convex, that is if the cost $g'$ is monotonically increasing, then the Monge map $Y(x)$ is monotone and can easily be seen to transport the interval $[-1/2, 0]$ to $[-2, -1]$. To satisfy the target constraint, we require that $Y'(x) = \frac{d\mu}{dx}(x)/\frac{d\nu}{dy}(Y(x)) = 2$. Since $Y(-1/2) = -2$ and $Y(1/2) = 2$, then

$$Y(x) = \begin{cases} -1 + 2x, & x < 0 \\ 1 + 2x, & x > 0. \end{cases}$$

The travel time is given by $\tau(x) = |Y(x) - x| = 1 + |x|$. Inverting $Y$ to express this as the free boundary $s(y)$, we get

$$s(y) = \frac{1}{2} + \frac{1}{2}|y|.$$ 

The costate as a function of the end position, $P(y)$ has the same sign as $y$ and is determined by the transversality condition $[4.3]$, 

$$0 = H(s(y), y, P(y)) = |P(y)| - g'(s(y)).$$

Thus $P(y) = \text{sign}(y)g'(\frac{1}{2} + \frac{1}{2}|y|)$. The dual potential $\psi$ satisfies $\nabla \psi(y) = P(y)$, thus can be integrated to obtain

$$\psi(y) = \int_{\frac{1}{2}}^{\frac{1}{2} + \frac{1}{2}|y|} g'(\frac{1}{2} + \frac{1}{2}q) dq + C = 2g\left(\frac{1}{2} + \frac{1}{2}|y|\right).$$
The function $J_\psi$ is obtained via the formula,
\[
J_\psi(0, x) = \psi(Y(x)) - \int_0^{\tau(x)} g'(s) ds
= \psi(Y(x)) - g(|Y(x) - x|)
= g(1 + |x|).
\]
At later times, with $t < s(q)$, we have
\[
J_\psi(t, q) = J_\psi(0, |q| - t) + \int_0^t g'(s) ds = g(1 + |q| - t) + g(t).
\]
Note that the inequality $\psi(q) \leq J_\psi(t, q)$ is here an expression of the convexity of $g$, since $g \left( \frac{1}{2} + \frac{1}{2} |q| \right) \leq \frac{1}{2} g(1 + |q| - t) + \frac{1}{2} g(t)$.

B. If $g$ is concave, i.e., $g'$ is monotonically decreasing, then the Monge map $Y(x)$ is orientation reversing and satisfies $Y'(x) = -2$. Then $Y(-1/2) = -1$ and $Y(1/2) = 1$ and
\[
Y(x) = \begin{cases} 
-2 - 2x, & x < 0 \\
2 - 2x, & x > 0.
\end{cases}
\]
The travel time is $\tau(x) = 2 - 3|x|$. The free boundary is then
\[
s(y) = 2 - 3|1 - |y||/2 = -1 + 3|y|/2.
\]
The transversality condition (4.3) implies that $\nabla \psi(y) = \text{sign}(y)g'(-1 + 3|y|/2)$, which can be integrated to obtain
\[
\psi(y) = \int_{y}^{y} g' \left( -1 \pm \frac{3}{2} q \right) dq + C = \frac{2}{3} g \left( -1 + \frac{3}{2} |y| \right).
\]
The function $J_\psi$ is then
\[
J_\psi(0, x) = \frac{2}{3} g(2 - 3|x|) - g(2 - 3|x|) = -\frac{1}{3} g(2 - 3|x|).
\]
At later times, with $t > s(q)$, we have
\[
J_\psi(t, q) = J_\psi(0, |q| - t) + g(t) = -\frac{1}{3} g(2 - 3(|q| - t)) + g(t).
\]

Figure 2 illustrates how the function $J_\psi$ evolves in time $t$. In the case A, the graph $J_\psi$ moves downward as time $t$ increases and mass is dropped when it hits the barrier $\psi$ (hitting time). In the case B, the graph $J_\psi$ moves upward as $t$ increases, and mass is dropped when it departs from the barrier $\psi$ (exit time).

7. Connections with classical optimal transport problems

We now generalize the above examples to higher dimensions and show a simple connection with the classical optimal transportation problem of Monge (see e.g. [13]). We let the control set be the unit sphere, $A = S^{n-1} \subset \mathbb{R}^n$ and $f(q, A) = A$, we can
then view the control problem as minimizing over trajectories with velocity bounded by 1. Consider now a cost that is only given by a time penalty,

\[ K(t, q, A) = g'(t), \]

where \( g \) satisfies \( g(0) = 0 \) and \( g'(t) \geq 0 \). The optimal trajectory connecting \( x \) and \( y \) is again the straight line that reaches \( y \) at time \( \tau^{x,y} = |y - x| \). Thus, the cost associated to travelling from \( x \) to \( y \) in the optimal amount of time is

\[ c(x, y) = \int_0^{\tau^{x,y}} g'(t) dt = g(|y - x|). \]

In the case that \( g \) is strictly convex, \( g'(t) \) increases hence satisfies \([TC]\) and if \( g \) is strictly concave then \([TD]\) is satisfied. The case when \( g(t) = t \) is the stationary case \([TS]\) and corresponds to the classical distance function cost of Monge (see e.g., [7]). Proposition 5.1 then identifies the qualitative change in the solution when the cost changes from being induced by a convex or a concave function \( g \). The end time then changes from being a hitting time of a barrier from below to a hitting time from above.

When \( g \) is convex, and as noted in [5], an equivalent dynamic formulation can be posed with fixed endtime \( T = 1 \) and Lagrangian \( L(q, v) = g(|v|) \). The associated Hamilton-Jacobi equation is then

\[
\frac{\partial}{\partial t} I_\psi(t, q) + g^*(|\nabla I_\psi(t, q)|) = 0, \quad 0 \leq t < 1, \\
I_\psi(1, y) = \psi(y).
\]
On the other hand, in our formulation we have the following free boundary problem,

\[ \frac{\partial}{\partial t} J_\psi(t, q) + |\nabla J_\psi(t, q)| = g'(t), \quad 0 \leq t < s(q), \]

\[ J_\psi(t, q) = \psi(q), \quad s(q) \leq t, \]

\[ |\nabla \psi(y)| = g'(s(y)). \]

The equivalence of these two problems implies that for optimal \( \psi \), the solutions satisfy \( I_\psi(0, x) = J_\psi(0, x) \). Both sets of equations have the characteristic curves given by straight lines, along which \( \nabla J_\psi \) or \( \nabla I_\psi \) is constant. The difference is a dynamic rescaling of the length of the curves so that the trajectories in our free end time formulation have constant velocity, while in the classical formulation they have constant end time.

**References**

[1] Andrei Agrachev and Paul Lee. Optimal transportation under nonholonomic constraints. *Transactions of the American Mathematical Society*, 361(11):6019–6047, 2009.

[2] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Modern Birkhäuser Classics. Birkhäuser Boston, 2008.

[3] Mathias Beiglböck, Alexander MG Cox, and Martin Huesmann. Optimal transport and Skorokhod embedding. *Inventiones mathematicae*, 208(2):327–400, 2017.

[4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

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

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

[7] Luis Caffarelli, Mikhail Feldman, and Robert McCann. Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs. *Journal of the American Mathematical Society*, 15(1):1–26, 2002.

[8] Piermarco Cannarsa and Pierre Cardaliaguet. Hölder estimates in space-time for viscosity solutions of Hamilton-Jacobi equations. *Communications on Pure and Applied Mathematics*, 63(5):590–629, 2010.

[9] Alexander MG Cox and Jiajie Wang. Root’s barrier: Construction, optimality and applications to variance options. *The Annals of Applied Probability*, 23(3):859–894, 2013.

[10] Lawrence C Evans and Wilfrid Gangbo. *Differential equations methods for the Monge-Kantorovich mass transfer problem*, volume 653. American Mathematical Soc., 1999.

[11] Lawrence C Evans and D Gomes. Effective Hamiltonians and averaging for Hamiltonian dynamics I. *Archive for rational mechanics and analysis*, 157(1):1–33, 2001.

[12] F Friedrich, F Gehbauer, and U Rickers. Optimized resource allocation for emergency response after earthquake disasters. *Safety Science*, 35(1):41 – 57, 2000.

[13] Wilfrid Gangbo and Robert J McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.

[14] Nassif Ghoussoub, Young-Heon Kim, and Aaron Zeff Palmer. Optimal stopping in stochastic control transportation problems. *In preparation*, 2018.

[15] Nassif Ghoussoub, Young-Heon Kim, and Aaron Zeff Palmer. PDE methods for Skorokhod embeddings. *In preparation*, 2018.

[16] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences de Paris*, 1781.
[17] Aaron Zeff Palmer and Alexander Vladimirsky. Optimal stopping with a probabilistic constraint. *Journal of Optimization Theory and Applications*, 175(3):795–817, Dec 2017.

[18] Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. CRC Press, 1987.

[19] Vladimir N Sudakov. *Geometric problems in the theory of infinite-dimensional probability distributions*. Number 141. American Mathematical Soc., 1979.

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

[21] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

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

E-mail address: nassif@math.ubc.ca, yhkim@math.ubc.ca, azp@math.ubc.ca