# STOCHASTIC OPTIMAL TRANSPORT WITH FREE END TIME

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

ABSTRACT. We consider a stochastic transportation problem between two prescribed probability distributions (a source and a target) over processes with general drift dependence and with free end times. First, and in order to establish a dual principle, we associate two equivalent formulations of the primal problem in order to guarantee its convexity and lower semi-continuity with respect to the source and target distributions. We exhibit an equivalent Eulerian formulation, whose dual variational principle is given by Hamilton-Jacobi-Bellman type variational inequalities. In the case where the drift is bounded, regularity results on the minimizers of the Eulerian problem then enable us to prove attainment in the corresponding dual problem. We also address attainment when the drift component of the cost defining Lagrangian $L$ is superlinear $L \approx |u|^p$ with $1 < p < 2$, in which case the setting is reminiscent of our approach -in a previous work- on deterministic controlled transport problems with free end time. We finally address criteria under which the optimal drift and stopping time are unique, namely strict convexity in the drift component and monotonicity in time of the Lagrangian.

## CONTENTS

## 1. INTRODUCTION

The problem of transporting a distribution from a given source to a prescribed target has been studied since the pioneering work of Monge in 1781 (see [26]) and has many applications in analysis, probability theory, and partial differential equations. In this paper, we consider an optimal transportation problem for stochastic processes with controlled dynamics and

free end time, where the transport cost is given by a general Lagrangian $L$ on $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d$ as follows: if $\mu$ and $\nu$ are two probability measures, then the stochastic transport problem is formally stated as

$$(1.1) \qquad \mathcal{P}_L(\mu, \nu) := \inf_{\beta, \tau} \left\{ \mathbb{E}\left[ \int_0^\tau L(t, X_t, \beta_t) dt \right]; \ dX_t = \beta_t \, dt + dW_t, \ X_0 \sim \mu, \ X_\tau \sim \nu \right\}.$$

More precise definitions will be given later, but for now we mention that the minimization is over all suitable drifts $\beta$, and all stopping times $\tau$. The notation $X_0 \sim \mu$ means that the initial position of the path has $\mu$ as its distribution, and $X_\tau \sim \nu$ means that the processed stopped at the random time $\tau$ has $\nu$ as its distribution.

In [23, 25, 28], the authors consider Problem (1.1) in the case where the end time is fixed ($\tau = 1$), i.e. they minimize

$$(1.2) \qquad \inf_\beta \left\{ \mathbb{E}\left[ \int_0^1 L(t, X_t, \beta_t) dt \right]; \ dX_t = \beta_t \, dt + dW_t, \ X_0 \sim \mu, \ X_1 \sim \nu \right\}.$$

Under some assumptions on the Lagrangian $L$, they establish a weak duality principle for (1.2), namely that the primal value of (1.2) equals the value of the following dual problem:

$$(1.3) \qquad \sup_\psi \left\{ \int_{\mathbb{R}^d} \psi(y) \, \nu(\mathrm{d}y) - \int_{\mathbb{R}^d} J(0, x) \, \mu(\mathrm{d}x); \ J \text{ solves } (1.4) \right\}$$

where

$$(1.4) \qquad \begin{cases} \partial_t J(t, x) + \frac{1}{2} \Delta J(t, x) + H(t, x, \nabla J(t, x)) = 0 & \text{in } (0, 1) \times \mathbb{R}^d, \\ J(1, x) = \psi(x) & \text{on } \mathbb{R}^d, \end{cases}$$

where $H$ is the Hamiltonian associated to $L$. A special case that has received recent attention is when the Lagrangian is of the form $L(t, x, u) = \frac{1}{2}|u|^2 + V(x)$, which has connections with the Schrödinger bridge problem [20, 6, 24]. On the other hand, stopping uncontrolled processes with distribution constraints has a vast literature, in particular pertaining to applications in finance; for some of the approaches related to (1.1) see [1, 2, 3, 17].

Coming back to Problem (1.1), we shall pose the initial problem in a weaker sense so that it involves randomized stopping times and weak solutions to the SDE, analogous to the Kantorovich relaxation of the optimal transport [18] problem. The ultimate goal is to obtain minimizers involving true stopping times and representing strong solutions to the SDE. For that, we shall give two formulations of the stochastic transport problem (see Section 2), which we will ultimately prove equivalent:

- A *weak stochastic* formulation that poses the optimization problem over (weakly) controlled processes and randomized stopping times.
- A *convex stochastic* formulation which poses the optimization over probability measures on a space of randomly stopped paths for both state and drift.

Under appropriate conditions on $L$, the latter equivalent formulation renders the problem l.s.c. and convex in $\nu$. This will allow us to identify a corresponding dual problem, which can be described as follows:

$$(1.5) \qquad \mathcal{D}_L(\mu, \nu) = \sup_\psi \left\{ \int_{\mathbb{R}^d} \psi(y) \, \nu(dy) - \int_{\mathbb{R}^d} J_\psi(0, x) \, \mu(dx) \right\},$$

where $J_\psi$ can be viewed as the viscosity solution (equivalently minimal supersolution) of the following second order Hamilton-Jacobi-Bellman quasivariational inequality,

$$(1.6) \qquad \begin{cases} \partial_t J(t,x) + \frac{1}{2}\Delta J(t,x) + H\big(t,x,\nabla J(t,x)\big) \leq 0 & \text{in } \mathbb{R}^+ \times \mathbb{R}^d, \\ \psi(x) - J(t,x) \leq 0 & \text{on } \mathbb{R}^+ \times \mathbb{R}^d, \end{cases}$$

where $H$ is again the Hamiltonian associated to $L$. We then establish the (weak) duality principle (see Section 3),

$$(1.7) \qquad\qquad \mathcal{D}_L(\mu,\nu) = \mathcal{P}_L(\mu,\nu).$$

The most crucial part of the analysis is to find an optimal end potential $\psi$ (hence $J_\psi$) for the dual problem (1.5). In fact, there is no general result in the literature about the attainment in the dual problem (1.3). However, one of the main advantages of considering (1.1) instead of (1.2) is that the constraints on the potentials $(\psi, J_\psi)$ are now somehow relaxed, i.e. we have $J \geq \psi$ on $\mathbb{R}^+ \times \mathbb{R}^d$ instead of $J(1,\cdot) = \psi$ and so, this will allow us to replace $\psi$ by an end potential $\bar\psi \geq \psi$ that satisfies

$$\frac{1}{2}\Delta\bar\psi(x) + \inf_{t\in\mathbb{R}^+} H\big(t,x,\nabla\bar\psi(x)\big) \leq 0,$$

and we get some Sobolev/Hölder estimates on $\bar\psi$. We note that this attainment, in tandem with the weak duality (1.7), allows - through a verification type theorem - to characterize both the optimal process and the stopping time that resolve the primal problem. In our quest to prove attainment in the dual problem, we will focus on two cases:

(1) When the drift is bounded, the distribution of the process then possesses additional Sobolev regularity due to the diffusion. However, the dual potential may be unbounded with singularities similar to the fundamental solution of the Laplace equation. We find that a sufficient condition to solve the problem is for the target distribution to lie in the dual to an appropriately weighted $L^1$ space. This case builds upon the Sobolev space approach for the Skorokhod problem (without drift) studied in [13].

(2) When the drift is strong (that is if $1 < p < 2$), there is no additional regularity on the density due to the possibility of 'local controllability', i.e. the ability to transport to a Dirac-mass with finite-cost. However, this 'local controllability' allows for uniform bounds and Hölder estimates on the end potential, as in [5], and generalizes the approach for deterministic control problems [11]. In particular, any compactly supported target measure may be reached with finite cost.

In either case, we shall prove attainment in the dual problem (Section 5). But for that, we need to introduce two Eulerian formulations (see Section 4) for the weak stochastic formulation of the primal problem (1.1).

- *The strong Eulerian formulation*, which poses the problem with a velocity field and the solution to a Fokker-Planck equation with stopping. More precisely,

$$(1.8) \qquad \mathcal{P}_L^{\mathcal{E}}(\mu,\nu) = \inf_{(m,v,\rho)\in\mathcal{E}(\mu)} \left\{ \int_{\mathbb{R}^+}\int_{\mathbb{R}^d} L\big(t,x,v(t,x)\big)m(t,x)\,dxdt; \int_{\mathbb{R}^+} \rho(d\tau,\cdot) = \nu \right\},$$

where a triplet $(m,v,\rho)$ belongs to $\mathcal{E}(\mu)$ if the following hold: $\rho$ is a probability measure on $\mathbb{R}^+ \times \mathbb{R}^d$, $m_t$ is a nonnegative density in the Sobolev space $H^1(\mathbb{R}^d)$ for

each time $t \in \mathbb{R}^+$, and $v$ is a measurable velocity field, and for all smooth test functions $\phi$ on $\mathbb{R}^+ \times \mathbb{R}^d$, we have the following:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \left[ \partial_t \phi(t, x) \, m(t, x) + \nabla \phi(t, x) \cdot \left( v(t, x) \, m(t, x) - \frac{1}{2} \nabla m(t, x) \right) \right] dx dt$$

(1.9)
$$= \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \phi(\tau, y) \rho(d\tau, dy) - \int_{\mathbb{R}^d} \phi(0, x) \mu(dx).$$

- *The convex Eulerian formulation*, which poses the problem over phase-space distributions satisfying a convex set of inequalities. More precisely,

(1.10) $$\mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu) = \inf_{(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)} \left\{ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} L(t, x, u) \eta_t(dx, du) dt \; : \; \int_{\mathbb{R}^+} \rho(d\tau, \cdot) = \nu \right\},$$

where a pair of density process and stopping time $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ if $\eta$ is a measurable map from $\mathbb{R}^+$ to nonnegative measures on $\mathbb{R}^d \times \mathbb{R}^d$, $\rho$ is a probability measure on $\mathbb{R}^+ \times \mathbb{R}^d$, and for all smooth test functions $\phi$ on $\mathbb{R}^+ \times \mathbb{R}^d$:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left[ \partial_t \phi(t, x) + \frac{1}{2} \Delta \phi(t, x) + \nabla \phi(t, x) \cdot u \right] \eta_t(dx, du) dt$$

(1.11)
$$= \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \phi(\tau, y) \rho(d\tau, dy) - \int_{\mathbb{R}^d} \phi(0, x) \mu(dx).$$

We shall prove the latter to be equivalent to the primal problem (1.1) first by embedding phase-space distributions into the stochastic formulation and then showing the reverse inequality by using weak duality. We then prove that, when the drift is bounded, the strong Eulerian formulation is also equivalent by using suitable Sobolev estimates.

After proving attainment in the dual problem (see Section 5), we proceed to obtain useful information on the primal problem, and in some special cases (e.g., when $t \mapsto L(t, x, u)$ is monotone), we show (see Section 6) that the unique optimizer is given by the hitting time to a space-time barrier,

$$\tau^* = \inf \left\{ t; \; J_\psi(t, X_t) = \psi(X_t) \right\},$$

which is reminiscent of the graphical structure describing the optimizers in the deterministic mass transports problems studied by Brenier [4], Gangbo-McCann [10] and others.

## 2. Stochastic Formulations

2.1. **Basic assumptions and notations.** We shall assume throughout the paper that the Lagrangian, $(t, x, u) \in \mathbb{R}^+ \times \mathbb{R}^d \times U \mapsto L(t, x, u) \in \mathbb{R}^+$, where $U \subset \mathbb{R}^d$, is a continuous function of time, position, and drift, uniformly continuous in $(t, x)$ (uniformly with respect to $u$) and is convex with respect to the drift, i.e. $u \mapsto L(t, x, u)$ is convex for all $(t, x) \in \mathbb{R}^+ \times \mathbb{R}^d$.

In addition, we assume either that $U = \mathbb{R}^d$ and the Lagrangian $L$ is superlinear with respect to the drift and bounded from below, i.e. there is some $c > 0$ and $p > 1$ such that

(2.1) $$c(|u|^p + 1) \le L(t, x, u), \quad \text{for all } (t, x, u) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d,$$

or that $U$ is a bounded convex subset of $\mathbb{R}^d$. Let $\Omega = C(\mathbb{R}^+, \mathbb{R}^d)$ be the space of continuous paths from $\mathbb{R}^+$ to $\mathbb{R}^d$, $(X_t)_{t \in \mathbb{R}^+}$ be the canonical process, i.e. $X_t(\omega) = \omega(t)$ for every $\omega \in \Omega$, and $\mathbb{F} = \{\mathcal{F}_t\}_{t \in \mathbb{R}^+}$ be the canonical filtration generated by $X$. Let $\mu, \nu$ be two probability measures on $\mathbb{R}^d$. The two (equivalent) formulations of the stochastic transport problem are the following:

2.2. **The weak stochastic formulation.** We say that the triplet $(\mathbb{P}, \beta, \alpha)$ belongs to $\mathcal{A}(\mu)$ if the following conditions hold:

  i) $(\Omega, \mathbb{F}, \mathbb{P})$ is a filtered probability space.
  ii) The initial distribution is given by $\mu$, i.e. $X_0 \sim_{\mathbb{P}} \mu$ which means that $X_{0\#}\mathbb{P} = \mu$.
  iii) The drift, $\beta : \mathbb{R}^+ \times \Omega \to U$, is $\mathbb{F}$-progressively measurable and locally integrable, i.e. for each $\tau$, the map $\beta^\tau : \Omega \to L^1([0, \tau], U)$, which is given by $\beta^\tau(\omega)(t) = \beta_t(\omega)$ for every $t \in [0, \tau]$, satisfies $\beta^\tau$ is $\mathcal{F}_\tau$−measurable and

$$\mathbb{E}^{\mathbb{P}}\big[\|\beta^\tau\|_{L^1([0,\tau],U)}\big] < +\infty.$$

  iv) The process $W^\beta$ given by

$$W_t^\beta := X_t - X_0 - \int_0^t \beta_s \, ds, \text{ for every } t \in \mathbb{R}^+,$$

is the standard Brownian motion, i.e. $W^\beta{}_{\#}\mathbb{P}$ is the Wiener measure on $\Omega$ with $W_0 = 0$. In other words, $X_t$ has the following semimartingale decomposition:

$$(2.2) \qquad X_t = X_0 + \int_0^t \beta_s \, ds + W_t^\beta.$$

  v) $\alpha : \Omega \to \mathcal{M}(\mathbb{R}^+)$, where $\mathcal{M}(\mathbb{R}^+)$ is the space of measures on $\mathbb{R}^+$, is a randomized stopping time or equivalently, $A_t(\omega) := \alpha(\omega)([0, t])$ is increasing, right continuous, adapted to $\mathbb{F}$, with $A_0 \geq 0$ and $\lim_{t \to \infty} A_t(\omega) = 1$.

Let $(\alpha \ltimes \mathbb{P})(d\tau, d\omega) = \alpha(\omega)(d\tau)\mathbb{P}(d\omega)$ denote the measure on $\mathbb{R}^+ \times \Omega$ corresponding to the variable $(\tau, \omega)$. The constraint $X_\tau \sim_{\alpha \ltimes \mathbb{P}} \nu$ is equivalently defined by $X_{\tau\#}(\alpha \ltimes \mathbb{P}) = \nu$, i.e. we have

$$(2.3) \qquad \int_\Omega \int_{\mathbb{R}^+} g\big(\omega(\tau)\big)\alpha(\omega)(d\tau)\mathbb{P}(d\omega) = \int_{\mathbb{R}^d} g(y)\nu(dy), \text{ for all } g \in C_b(\mathbb{R}^d).$$

Now, we let $\mathcal{A}(\mu, \nu) = \Big\{(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu) : X_\tau \sim_{\alpha \ltimes \mathbb{P}} \nu\Big\}$. The cost, defined on $\mathcal{A}(\mu)$ with possible value of $+\infty$, is given by the following:

$$(2.4) \qquad \mathcal{J}_L(\mathbb{P}, \beta, \alpha) = \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\int_0^\tau L\big(t, X_t, \beta_t\big)dt\Big]$$
$$= \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L\big(t, X_t(\omega), \beta_t(\omega)\big)dt \, \alpha(\omega)(d\tau)\mathbb{P}(d\omega).$$

We can state our primal stochastic transportation problem as the minimization of the stochastic transport cost $\mathcal{J}_L(\mathbb{P}, \beta, \alpha)$ among all admissible $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$, i.e.

$$(2.5) \qquad \mathcal{P}_L(\mu, \nu) := \inf\Big\{\mathcal{J}_L(\mathbb{P}, \beta, \alpha) : (\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)\Big\}$$

with the convention that $\mathcal{P}_L(\mu, \nu) = +\infty$ if $\mathcal{A}(\mu, \nu) = \emptyset$ (in fact, we will prove in Section 5 that, under some assumptions on $\mu$ and $\nu$, this set $\mathcal{A}(\mu, \nu)$ is non-empty).

2.3. **The convex stochastic formulation.** With this formulation, we seek to linearize the functional in (2.4) by considering probability measures $\tilde{\mathbb{P}}$ on $\tilde{\Omega} := \Omega \times \Omega$ (this idea is due to Haussmann [14] and, it is used later by Tan & Touzi [28]). In other words, let $(X, B)$ be the canonical process on $\tilde{\Omega}$ (i.e. $(X_t, B_t)(\omega, b) = (\omega(t), b(t))$, for every $(\omega, b) \in \tilde{\Omega}$) and let $\tilde{\mathbb{F}}$ be the corresponding canonical filtration. We now have a process

$$(2.6) \qquad \tilde{W}_t(\omega, b) := X_t(\omega) - X_0(\omega) - B_t(b),$$

which will play the role of $W^\beta$ in the definition of $\mathcal{A}(\mu)$. We denote by $\tilde{\mathbb{P}}_X$ the projection onto the first component of $\Omega \times \Omega$, and $\tilde{\mathbb{P}}_B$ the projection onto the second component. We say that $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$ if the following conditions hold:

  i) $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}})$ is a filtered probability space.
  ii) We have $X_0 \sim_{\tilde{\mathbb{P}}_X} \mu$ which means that $X_{0\#}\tilde{\mathbb{P}}_X = \mu$.
  iii) For $\tilde{\mathbb{P}}_B$ almost every $b$, $B$ is differentiable and $\tilde{\beta} = B'$ is locally integrable, i.e. for each $\tau$, the map $\tilde{\beta}^\tau : \Omega \to L^1([0, \tau], U)$ defined by $\tilde{\beta}^\tau(b)(t) = \tilde{\beta}_t(b)$ satisfies

$$\mathbb{E}^{\tilde{\mathbb{P}}}\big[\|\tilde{\beta}^\tau\|_{L^1([0,\tau],\mathbb{R}^d)}\big] < +\infty.$$

  iv) The process $\tilde{W}$ defined in (2.6) is a standard Brownian motion, i.e. $\tilde{W}_{\#}\tilde{\mathbb{P}}$ is the Wiener measure on $\Omega$ with $\tilde{W}_0 = 0$.
  v) $\tilde{\alpha} : \tilde{\Omega} \to \mathcal{M}(\mathbb{R}^+)$ is a randomized stopping time, equivalently $\tilde{A}_t(\omega, b) := \tilde{\alpha}(\omega, b)([0, t])$ is increasing, right continuous, adapted to $\tilde{\mathbb{F}}$, with $\tilde{A}_0 \geq 0$ and $\lim_{t \to \infty} \tilde{A}_t(\omega, b) = 1$.

We also define $\tilde{\alpha} \ltimes \tilde{\mathbb{P}}$ to be the associated probability measure on $\mathbb{R}^+ \times \Omega \times \Omega$ and $(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}$ its projection onto the first two components. The constraint $X_\tau \sim_{(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}} \nu$ is similarly defined by

$$(2.7) \qquad \int_\Omega \int_\Omega \int_{\mathbb{R}^+} g\big(\omega(\tau)\big)\tilde{\alpha}(\omega, b)(d\tau)\tilde{\mathbb{P}}(d\omega, db) = \int_{\mathbb{R}^d} g(y)\nu(dy), \text{ for all } g \in C_b(\mathbb{R}^d).$$

Now, we say that $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu, \nu)$ if $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$ and $X_\tau \sim_{(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}} \nu$. The cost is similarly given by the following:

$$(2.8) \qquad \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}) = \mathbb{E}^{\tilde{\alpha} \ltimes \tilde{\mathbb{P}}}\Big[\int_0^\tau L\big(t, X_t, \tilde{\beta}_t\big)dt\Big]$$

$$= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L\big(t, X_t(\omega), \tilde{\beta}_t(b)\big)dt \, \tilde{\alpha}(\omega, b)(d\tau)\tilde{\mathbb{P}}(d\omega, db).$$

Then, we consider the following relaxation of (2.5) (again, we let $\tilde{\mathcal{P}}_L(\mu, \nu) = +\infty$ if $\tilde{\mathcal{A}}(\mu, \nu)$ is empty):

$$(2.9) \qquad \tilde{\mathcal{P}}_L(\mu, \nu) := \inf\Big\{\tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}) : (\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu, \nu)\Big\}.$$

We now show that problems (2.9) and (2.5) are equivalent in the sense that the two problems have the same minimal values. First, we show that the convex stochastic problem (2.9) is a relaxation of the weak stochastic formulation (2.5), and then show that there is a projection from the convex stochastic problem back onto the weak stochastic formulation that does not increase the cost. More precisely, we have the following:

**Proposition 2.1.** *With the above notations, the following hold:*
  *(1) For every $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, there exists $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$ with $(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X} = \alpha \ltimes \mathbb{P}$ and*
  *$\mathcal{J}_L(\mathbb{P}, \beta, \alpha) = \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}).$*

(2) *Conversely, for every* $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$, *we can find* $\alpha, \beta$ *such that* $(\tilde{\mathbb{P}}_X, \beta, \alpha) \in \mathcal{A}(\mu)$ *with* $(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X} = \alpha \ltimes \tilde{\mathbb{P}}_X$ *and* $\mathcal{J}_L(\tilde{\mathbb{P}}_X, \beta, \alpha) \leq \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha})$.

(3) *In particular, we have* $\mathcal{P}_L(\mu, \nu) = \tilde{\mathcal{P}}_L(\mu, \nu)$.

*Proof.* Take $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$ and define $B_t^\beta(\omega) := \int_0^t \beta_s(\omega)\, ds$. We set $\tilde{\mathbb{P}} = (X, B^\beta)_{\#}\mathbb{P}$ and $\tilde{\alpha}(\omega, b) = \alpha(\omega)$. We can easily check the five properties defining $\tilde{\mathcal{A}}(\mu)$:

i) $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}})$ is clearly a filtered probability space.

ii) Since $\tilde{\mathbb{P}}_X = \mathbb{P}$, we are also dealing with the same initial distribution $(X_{0\#}\tilde{\mathbb{P}}_X = X_{0\#}\mathbb{P} = \mu)$.

iii) We have that
$$\mathbb{E}^{\tilde{\mathbb{P}}}\big[\|\tilde{\beta}^\tau\|_{L^1([0,\tau],\mathbb{R}^d)}\big] = \mathbb{E}^{\mathbb{P}}\big[\|\beta^\tau\|_{L^1([0,\tau],\mathbb{R}^d)}\big] < +\infty.$$

iv) Note that $\tilde{W}(\omega, B^\beta(\omega)) = W^\beta(\omega)$ and therefore $\tilde{W}_{\#}\tilde{\mathbb{P}} = W^\beta_{\#}\mathbb{P}$ is the Wiener measure on $\Omega$ with $\tilde{W}_0 = 0$.

v) It is easy to check that $\tilde{\alpha}$ is still a randomized stopping time on the extended space.

On the other hand, we also have $(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X} = \alpha \ltimes \mathbb{P}$. From the definitions of $\mathcal{J}_L$ and $\tilde{\mathcal{J}}_L$, we have

$$\tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}) = \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L\big(t, X_t(\omega), \tilde{\beta}_t(b)\big) dt\, \alpha(\omega)(d\tau)(X, B^\beta)_{\#}\mathbb{P}(d\omega, db)$$
$$= \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L\big(t, X_t(\omega), \beta_t(\omega)\big) dt\, \alpha(\omega)(d\tau)\mathbb{P}(d\omega)$$

and so, it follows that $\mathcal{J}_L(\mathbb{P}, \beta, \alpha) = \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha})$.

For the second claim, take $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$ and set $\mathbb{P} = \tilde{\mathbb{P}}_X$. We define $\alpha$ by disintegration in such a way that $(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X} = \alpha \ltimes \mathbb{P}$ and $\beta$ by the conditional expectation $\beta_t(\omega) := \mathbb{E}^{\tilde{\mathbb{P}}}\big[\tilde{\beta}_t | \tilde{\mathcal{F}}_t^X\big](\omega)$ (here, $\tilde{\mathcal{F}}_t^X$ is the filtration generated by process $X$). Again, we have the following five properties:

i) $(\Omega, \mathbb{F}, \mathbb{P})$ is a filtered probability space.

ii) Since $\tilde{\mathbb{P}}_X = \mathbb{P}$, we still have the same initial distribution.

iii) By Jensen's inequality, we have that
$$\mathbb{E}^{\mathbb{P}}\big[\|\mathbb{E}^{\tilde{\mathbb{P}}}[\tilde{\beta}^\tau | \tilde{\mathcal{F}}_t^X]\|_{L^1([0,\tau],\mathbb{R}^d)}\big] \leq \mathbb{E}^{\tilde{\mathbb{P}}}\big[\|\tilde{\beta}^\tau\|_{L^1([0,\tau],\mathbb{R}^d)}\big] < +\infty.$$

iv) From [29, Theorem 4.3], we have that the following process is a standard Brownian motion:
$$W_t^\beta := X_t - X_0 - \int_0^t \mathbb{E}^{\tilde{\mathbb{P}}}\big[\tilde{\beta}_s | \tilde{\mathcal{F}}_s^X\big] ds.$$

v) Finally, it is straightforward to verify that $\alpha$ is a randomized stopping time.

On the other hand, using the definitions of $\mathbb{P}$, $\beta$ and $\alpha$ and Jensen's inequality, we get that
$$\mathcal{J}_L(\mathbb{P}, \beta, \alpha) \leq \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}).$$

Since we have identity of the distributions $\alpha \ltimes \mathbb{P} = (\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}$, it follows that the values of the primal weak stochastic and convex stochastic problems are equal. $\quad\square$

The convex problem (2.9) satisfies a compactness property for a suitable topology on the measures $\tilde{\alpha} \ltimes \tilde{\mathbb{P}}$ on $\mathbb{R}^+ \times \Omega \times \Omega$. We note that a simple truncation of the stopping time allows us to restrict to a compact domain in time and space. We define the truncation for

$(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$ by considering the randomized stopping corresponding to $\tau \wedge \sup\{\tau; \tau \leq T, |X_\tau| \leq R\}$, i.e. $\alpha^{T,R}(\omega) := S^{T,R}_\omega \sharp \alpha$ where $S^{T,R}_\omega(\tau) = \tau \wedge \sup\{t; \ t \leq T, |X_t(\omega)| \leq R\}$. We set $S^{T,R}(\tau, w) := (S^{T,R}_w(\tau), w)$.

**Lemma 2.2.** *For any* $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, *the truncation with* $T, R \in \mathbb{R}^+$ *satisfies* $(\mathbb{P}, \beta, \alpha^{T,R}) \in \mathcal{A}(\mu)$ *with* $\alpha^{T,R}$ *is supported on* $[0, T]$ *and* $|X_t(w)| \leq R$ *almost surely. Furthermore,*

$$\lim_{T,R \to \infty} \mathcal{J}_L(\mathbb{P}, \beta, \alpha^{T,R}) = \mathcal{J}_L(\mathbb{P}, \beta, \alpha),$$

*and for all* $g \in C_b(\mathbb{R}^d)$,

$$\lim_{T,R \to \infty} \mathbb{E}^{\alpha^{T,R} \ltimes \mathbb{P}}\big[g(X_\tau)\big] = \mathbb{E}^{\alpha \ltimes \mathbb{P}}\big[g(X_\tau)\big].$$

*Proof.* Since the range of $S^{T,R}$ is contained in $[0, T] \times \Omega$, it is clear that $\alpha^{T,R}$ is supported on $[0, T]$ and $|X_t(w)| \leq R$ holds $(\alpha^{T,R} \ltimes \mathbb{P})$-almost surely. We then have that $(\mathbb{P}, \beta, \alpha^{T,R}) \in \mathcal{A}(\mu)$ since $S^{T,R}$ is continuous and maps $[0, \tau] \times \Omega$ into $[0, \tau] \times \Omega$, for all $\tau \in \mathbb{R}^+$, and leaves $\mathbb{P}$ and $\beta$ unchanged so that the decomposition (2.2) holds. Furthermore, it follows by the monotone convergence theorem that the limit of $\mathcal{J}_L(\mathbb{P}, \beta, \alpha^{T,R})$ is $\mathcal{J}_L(\mathbb{P}, \beta, \alpha)$ and that the end measure is also recovered.   □

Before we have compactness we also need a truncation of the measure $\tilde{\mathbb{P}}$ so that weak limits do not lead to measures associated with badly behaved drifts after the end-time.

**Proposition 2.3.** *The set* $\{\tilde{\alpha} \ltimes \tilde{\mathbb{P}} : (\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)\}$ *is convex. Moreover, given any sequence* $(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i) \in \tilde{\mathcal{A}}(\mu, \nu^i)$ *with* $\nu^i \rightharpoonup \nu$, *there is a* $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu, \nu)$ *such that* $\tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}) \leq \liminf_{i \to \infty} \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i)$.

*Proof.* The convexity follows immediately from [16, Corollary III.2.8]. Now, for each $\tau \in \mathbb{R}^+$, let us define the map $S^\tau : \Omega \times \Omega \to \Omega \times \Omega$ by

$$(2.10) \qquad\qquad S^\tau(\omega, b) := (\hat{\omega}^\tau, \hat{b}^\tau),$$

where

$$\hat{\omega}^\tau(t) := \omega(t) - b(t) + b(\tau \wedge t) \ \ \text{and} \ \ \hat{b}^\tau(t) := b(\tau \wedge t).$$

Notice that, for every $\tau \in \mathbb{R}^+$, the map $S^\tau$ is invariant for $\tilde{W}$ in the sense that

$$\tilde{W}_t\big(S^\tau(\omega, b)\big) = \hat{\omega}^\tau(t) - \hat{\omega}^\tau(0) - \hat{b}^\tau(t) = \omega(t) - \omega(0) - b(t) = \tilde{W}_t(\omega, b).$$

We define the pair $(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i)$ by duality such that

$$\int_\Omega \int_\Omega \int_{\mathbb{R}^+} H(\tau, \omega, b) \tilde{\gamma}^i(\omega, b)(d\tau) \tilde{\mathbb{Q}}^i(d\omega, db) = \int_\Omega \int_\Omega \int_{\mathbb{R}^+} H(\tau, \hat{\omega}^\tau, \hat{b}^\tau) \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db),$$

for all $H \in C_b(\mathbb{R}^+ \times \Omega \times \Omega)$. In particular, the truncated measure $\tilde{\mathbb{Q}}^i$ satisfies

$$\int_\Omega \int_\Omega F(\omega, b) \tilde{\mathbb{Q}}^i(d\omega, db) = \int_\Omega \int_\Omega \int_{\mathbb{R}^+} F(\hat{\omega}^\tau, \hat{b}^\tau) \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db), \text{ for all } F \in C_b(\Omega \times \Omega).$$

Now, we will show that $(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i) \in \tilde{\mathcal{A}}(\mu, \nu)$ with the same cost. First, it is not difficult to check that the properties (i), (ii), (iii), (iv) and (v) in the definition of $\tilde{\mathcal{A}}(\mu)$ hold. In fact,

for all $G \in C_b(\Omega)$, we have

$$
\int_\Omega G(\omega) \tilde{W}_\# \tilde{\mathbb{Q}}^i(d\omega) = \int_{\tilde{\Omega}} G(\tilde{W}(\omega, b)) \tilde{\mathbb{Q}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} G(\tilde{W}(\hat{\omega}^\tau, \hat{b}^\tau)) \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega G(\tilde{W}(\omega, b)) \tilde{\mathbb{P}}^i(d\omega, db).
$$

Thus, $\tilde{W}_\# \tilde{\mathbb{Q}}^i = \tilde{W}_\# \tilde{\mathbb{P}}^i$. On the other hand, we have

$$
\int_\Omega \int_\Omega \int_{\mathbb{R}^+} f(\omega(\tau)) \tilde{\gamma}^i(\omega, b)(d\tau) \tilde{\mathbb{Q}}^i(d\omega, db) = \int_\Omega \int_\Omega \int_{\mathbb{R}^+} f(\hat{\omega}^\tau(\tau)) \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} f(\omega(\tau)) \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db)
$$

$$
= \int f \, d\nu, \text{ for all } f \in C_b(\mathbb{R}^d).
$$

This implies that $X_\tau \sim_{(\tilde{\gamma}^i \ltimes \tilde{\mathbb{Q}}^i)_{T,X}} \nu$. Similarly, we see that the cost does not change since

$$
\int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L(t, X_t(\omega), \tilde{\beta}_t(b)) dt \, \tilde{\gamma}^i(\omega, b)(d\tau) \tilde{\mathbb{Q}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L(t, X_t(\hat{\omega}^\tau), \tilde{\beta}_t(\hat{b}^\tau)) dt \, \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau L(t, X_t(\omega), \tilde{\beta}_t(b)) dt \, \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db).
$$

The advantage of this truncation is that we now have that

$$
\int_\Omega \int_\Omega \int_{\mathbb{R}^+} |\tilde{\beta}_t(b)|^p dt \, \tilde{\mathbb{Q}}^i(d\omega, db) = \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} |\tilde{\beta}_t(\hat{b}^\tau)|^p dt \, \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db)
$$

$$
= \int_\Omega \int_\Omega \int_{\mathbb{R}^+} \int_0^\tau |\tilde{\beta}_t(b)|^p dt \, \tilde{\alpha}^i(\omega, b)(d\tau) \tilde{\mathbb{P}}^i(d\omega, db),
$$

$$
\leq c^{-1} \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i),
$$

where the last inequality comes from the fact that the Lagrangian $L$ is superlinear, or in the case that $U$ is bounded this holds for any $p$ with $c = \inf_{(t,x,u) \in \mathbb{R}^+ \times \mathbb{R}^d \times U} L(t, x, u) / |u|^p$. Furthermore, we have

$$
\int_\Omega \int_\Omega \tilde{\gamma}^i(\omega, b)([R, +\infty)) \tilde{\mathbb{Q}}^i(d\omega, db)
$$

$$
\leq \frac{1}{cR} \int_{\tilde{\Omega}} \int_R^\infty \int_0^\tau L(t, X_t(\omega), \tilde{\beta}_t(b)) dt \, \tilde{\gamma}^i(\omega, b)(d\tau) \tilde{\mathbb{Q}}^i(d\omega, db)
$$

$$
\leq \frac{1}{cR} \tilde{\mathcal{J}}_L(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i).
$$

Then, by [22, 30], we infer that the sequence $\tilde{\gamma}^i \ltimes \tilde{\mathbb{Q}}^i$ is tight and so, there is some $(\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\mu)$ such that $(\tilde{\gamma}^{i_k} \ltimes \tilde{\mathbb{Q}}^{i_k}) \rightharpoonup (\tilde{\alpha} \ltimes \tilde{\mathbb{P}})$. In particular, this implies that $(\tilde{\gamma}^{i_k} \ltimes \tilde{\mathbb{Q}}^{i_k})_{T,X} \rightharpoonup (\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}$ and $X_\tau \sim_{(\tilde{\alpha} \ltimes \tilde{\mathbb{P}})_{T,X}} \nu$.

Now, we define

$$U_{\varepsilon,\delta,\tau}(\omega) = \begin{cases} 1 & \text{if } \sup_{s\in[0,\tau-\varepsilon],|t-s|\leq\varepsilon} |X_s(w) - X_t(w)| < \delta, \\ 0 & \text{otherwise.} \end{cases}$$

Using the fact that $L$ is uniformly continuous in $(t,x)$ and by Jensen's inequality, it is not difficult to check that, when $U_{\varepsilon,\delta,\tau}(\omega) = 1$, we have

$$\int_0^{\tau-\varepsilon} L\Big(s, X_s(\omega), \frac{1}{\varepsilon}\int_s^{s+\varepsilon} \tilde{\beta}_t(b)\,dt\Big)ds \leq \int_0^\tau L\big(t, X_t(\omega), \tilde{\beta}_t(b)\big)\,dt + C(\varepsilon+\delta)\tau,$$

where $C$ is the modulus of uniform continuity of the Lagrangian $L$ with respect to $t$ and $x$. Then, we get

$$\tilde{\mathcal{J}}_L(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i) \geq \mathbb{E}^{\tilde{\gamma}^i \ltimes \tilde{\mathbb{Q}}^i}\Big[U_{\varepsilon,\delta,\tau}\int_0^{(\tau-\varepsilon)\vee 0} L\Big(s, X_s, \frac{1}{\varepsilon}(B_{s+\varepsilon}(b) - B_s(b)))\Big)ds\Big] - C(\varepsilon+\delta).$$

The map $(\tau, w, b) \mapsto U_{\varepsilon,\delta,\tau}(w)\int_0^{(\tau-\varepsilon)\vee 0} L(s, X_s(w), \frac{1}{\varepsilon}(B_{s+\varepsilon}(b) - B_s(b)))\mathrm{d}s$ is lower semi-continuous and $\tilde{\gamma}^i \ltimes \tilde{\mathbb{Q}}^i \rightharpoonup \tilde{\alpha} \ltimes \tilde{\mathbb{P}}$. So, passing to the limit when $i \to +\infty$, we get

$$\liminf_{i\to\infty} \tilde{\mathcal{J}}_L(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i) \geq \mathbb{E}^{\tilde{\alpha} \ltimes \tilde{\mathbb{P}}}\Big[U_{\varepsilon,\delta,\tau}\int_0^{(\tau-\varepsilon)\vee 0} L\Big(s, X_s, \frac{1}{\varepsilon}(B_{s+\varepsilon}(b) - B_s(b)))\Big)ds\Big] - C(\varepsilon+\delta).$$

Letting $\varepsilon \to 0$ and using Fatou's Lemma as well as the continuity of $C$, we infer that

$$\liminf_{i\to\infty} \tilde{\mathcal{J}}_L(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i) \geq \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha}) - C(\delta),$$

which completes the proof since $\delta > 0$ is arbitrary and, thanks to the fact that $\tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i) = \tilde{\mathcal{J}}_L(\tilde{\mathbb{Q}}^i, \tilde{\gamma}^i)$, for every $i \in \mathbb{N}$.   $\square$

As a consequence, we get

**Corollary 2.4.** *The problem* (2.5) *reaches a minimum. Moreover, the map* $\nu \mapsto \mathcal{P}_L(\mu, \nu)$ *is convex and lower semi-continuous.*

*Proof.* Let $(\mathbb{P}^i, \alpha^i)$ be a minimizing sequence in Problem (2.5) and $(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i)$ be the corresponding measures such that $\tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i) = \mathcal{J}_L(\mathbb{P}^i, \alpha^i)$ (see Proposition 2.1). By Proposition 2.3, there is a $(\tilde{\mathbb{P}}, \tilde{\alpha})$ such that $\liminf_{i\to\infty} \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}^i, \tilde{\alpha}^i) \geq \tilde{\mathcal{J}}_L(\tilde{\mathbb{P}}, \tilde{\alpha})$. This yields that the projection $(\mathbb{P}, \beta, \alpha)$ of $(\tilde{\mathbb{P}}, \tilde{\alpha})$ (see again Proposition 2.1) is a minimizer for (2.5). The second statement can be proved in a similar way using Propositions 2.1 & 2.3.   $\square$

## 3. Duality and dynamic programming

In this section, we verify a duality principle for the stochastic transportation problem (2.5). More precisely, we consider the following maximization principle:

$$(3.1) \qquad \mathcal{D}_L(\mu, \nu) := \sup_{(\phi,\psi)\in\mathcal{Z}_L} \Big\{ \int_{\mathbb{R}^d} \psi(y)\nu(dy) - \int_{\mathbb{R}^d} \phi(x)\mu(dx) \Big\},$$

where $\mathcal{Z}_L$ is the set of functions $(\phi, \psi)$ on $\mathbb{R}^d$ with $\psi \in C_b(\mathbb{R}^d)$, continuous and bounded, and $\phi \in B_\mu(\mathbb{R}^d)$, measurable w.r.t. the Borel $\sigma-$field on $\mathbb{R}^d$ completed by $\mu$, such that the following holds:

$$\phi(x) \geq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt\Big],$$

for all $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_x)$, $\mu-$ a.e. $x \in \mathbb{R}^d$.

We can characterize the dual problem further using the dynamic programming principle. We first define a translation map $T^{t,x} : \Omega \to C([t, +\infty), \mathbb{R}^d)$, where

$$T^{t,x}(\omega)(s) = \omega(s-t) + x.$$

Given $\psi \in C_b(\mathbb{R}^d)$, we introduce the function

$$J_\psi(t,x) := \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}_t(\delta_x)} \mathbb{E}^{\alpha\ltimes\mathbb{P}}\Big[\psi(X_\tau) - \int_t^\tau L(s, X_s, \beta_s)ds\Big]$$

where

$$\mathcal{A}_t(\delta_x) := \{T_\#^{t,0}(\mathbb{P}, \beta, \alpha); (\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_x)\}.$$

Here, the pushforward on $\mathbb{P}$ is as a measure and on $\beta$ and $\alpha$ are given respectively by $\beta(T^{t,0-1}(\omega))$ and $\alpha(T^{t,0-1}(\omega))$ for all $\omega \in C([t, +\infty), \mathbb{R}^d)$. Moreover, we note that

$$J_\psi(t,x) := \sup_{(\tilde{\mathbb{P}},\tilde{\alpha})\in\tilde{\mathcal{A}}_t(\delta_x)} \mathbb{E}^{\tilde{\alpha}\ltimes\tilde{\mathbb{P}}}\Big[\psi(X_\tau) - \int_t^\tau L(s, X_s, \tilde{\beta}_s)ds\Big]$$

where

$$\tilde{\mathcal{A}}_t(\delta_x) := \{\tilde{T}_\#^{t,0}(\tilde{\mathbb{P}}, \tilde{\alpha}); (\tilde{\mathbb{P}}, \tilde{\alpha}) \in \tilde{\mathcal{A}}(\delta_x)\}$$

with

$$\tilde{T}^{t,0}(w, b)(s) = (w(s-t), b(s-t))$$

and

$$\tilde{T}_\#^{t,0}\tilde{\alpha} := \tilde{\alpha}(\tilde{T}^{t,0-1}(\omega, b)) \text{ for all } (\omega, b) \in C([t, +\infty), \mathbb{R}^d \times \mathbb{R}^d).$$

For any $(\phi, \psi) \in \mathcal{Z}_L$, we have that $J_\psi(0, x) \leq \phi(x)$ for $\mu-$a.e. $x \in \mathbb{R}^d$. So, the dual problem (3.1) becomes

$$D_L(\mu, \nu) = \sup_{\psi\in C_b(\mathbb{R}^d)} \Big\{ \int_{\mathbb{R}^d} \psi(y)\,\nu(dy) - \int_{\mathbb{R}^d} J_\psi(0, x)\,\mu(dx)\Big\}.$$

We first prove a Lemma that verifies a basic level of regularity for $J_\psi$ needed in the proof of Theorem 3.2. Such a result is standard, see [28] for a setting similar to ours.

**Lemma 3.1.** *We suppose that $\psi \in C_b(\mathbb{R}^d)$. Then, the map $(t, x) \mapsto J_\psi(t, x)$ is lower semi-continuous and bounded from below. In addition, we have*

$$(3.2) \qquad \int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx) = \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}(\mu)} \Big\{\mathbb{E}^{\alpha\ltimes\mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt\Big]\Big\}.$$

*Proof.* We first note that we can express $J_\psi(t, x)$ as

$$J_\psi(t,x) = \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}(\delta_0)} \Big\{\mathbb{E}^{T_\#^{t,x}(\alpha\ltimes\mathbb{P})}\Big[\psi(X_\tau) - \int_t^\tau L(s, X_s, T_\#^{t,x}\beta_s)ds\Big]\Big\}$$

$$= \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}(\delta_0)} \Big\{\mathbb{E}^{\alpha\ltimes\mathbb{P}}\Big[\psi(X_\tau + x) - \int_0^\tau L(s+t, X_s + x, \beta_s)ds\Big]\Big\}.$$

For each $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_0)$, this defines a continuous function of $t$ and $x$ by the continuity of the translation maps $T^{t,x}$, the uniform continuity of $L$ and the bound on $\psi$, and so $J_\psi(t, x)$ is the supremum over all these functions, making it lower semi-continuous. Clearly, $J_\psi(t, x)$ is bounded below by $\psi(x)$ (we also note that $J_\psi(t, x)$ is bounded above by $\sup_x \psi$).

To prove (3.2), we first note that any $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$ disintegrates w.r.t. $\mu$ by the map $S(\tau, \omega) = \omega(0)$ and yields, using the definition of $J_\psi$, that

$$\int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx) \geq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt\Big].$$

Furthermore, for each $\varepsilon > 0$, the set

$$\Big\{(x, \alpha \ltimes \mathbb{P}); \; (\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_x), \; \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(s, X_s, \beta_s)ds\Big] \geq J_\psi(0, x) - \varepsilon\Big\}$$

is non-empty for $\mu-$a.e. $x$ and closed by Propositions 2.1 & 2.3 and thanks to the lower semi-continuity of $J_\psi$. Thus, by the Kuratowski and Ryll-Nardzewski selection theorem, we may find a $\mu-$Borel measurable map $Q(x) = \alpha^x \ltimes \mathbb{P}^x$ with $(\mathbb{P}^x, \beta^x, \alpha^x) \in \mathcal{A}(\delta_x)$ and $\mathbb{E}^{\alpha^x \ltimes \mathbb{P}^x}\Big[\psi(X_\tau) - \int_0^\tau L(s, X_s, \beta_s^x)ds\Big] \geq J_\psi(0, x) - \varepsilon$, for $\mu-$a.e. $x$. Therefore, we may consider $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$ such that $\alpha \ltimes \mathbb{P}$ disintegrates w.r.t. $\mu$ as $Q(x)$, i.e. $\alpha \ltimes \mathbb{P} = Q \ltimes \mu$. Thus, we have

$$\int_{\mathbb{R}^d} J_\psi(0, x) \, \mu(dx) - \varepsilon \leq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt\Big],$$

which proves the equality (3.2) when $\varepsilon \to 0$.   $\square$

The scheme to prove duality is by now standard in convex analysis (we refer, for instance, to [8]). The idea is that if $\nu \mapsto V(\nu) := \mathcal{P}_L(\mu, \nu)$ is l.s.c. and convex, then $V^{\star\star} = V$, where $V^\star$ denotes the Legendre-Fenchel transform of $V$. The value $V(\nu)$ is the minimal value of our stochastic primal problem (2.5), while $V^{\star\star}(\nu)$ is identified with the maximal value of the dual problem (3.1). This strategy was already used by many authors to establish dual principles for various optimal transport problems that do not fit in the Monge-Kantorovich theory (see [25, 28]).

**Theorem 3.2.** *The following equality holds:*

$$\mathcal{P}_L(\mu, \nu) = D_L(\mu, \nu).$$

*Proof.* First, it is clear that the maximal value of the dual problem is less than or equal to the minimal value of the primal one, since, for all admissible $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$ and $(\phi, \psi) \in \mathcal{Z}_L$, we have

$$\int_{\mathbb{R}^d} \psi(y) \, \nu(dy) - \int_{\mathbb{R}^d} \phi(x) \, \mu(dx) = \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \phi(X_0)\Big] \leq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\int_0^\tau L(t, X_t, \beta_t)dt\Big]$$

and so, it follows that $\mathcal{D}_L(\mu, \nu) \leq \mathcal{P}_L(\mu, \nu)$.

For the other direction, we have that $V(\nu)$ is convex and lower semi-continuous (see Corollary 2.4). Then, we have

$$V(\nu) = V^{\star\star}(\nu) = \sup_{\psi \in C_b(\mathbb{R}^d)} \Big\{\int_{\mathbb{R}^d} \psi(x)\nu(dx) - V^\star(\psi)\Big\}$$

where

$$V^\star(\psi) = \sup_{\nu \in \mathcal{M}(\mathbb{R}^d)} \Big\{\int_{\mathbb{R}^d} \psi(x)\nu(dx) - V(\nu)\Big\}$$

$$= \sup_{(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)} \Big\{\mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt\Big]\Big\}.$$

Note that we used that if $\nu$ is not a probability measure then $\mathcal{A}(\mu, \nu) = \emptyset$ and $V(\nu) = +\infty$, which does not affect the supremum in the definition of $V^\star$. We then have by Lemma 3.1 that

$$V^*(\psi) = \int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx)$$

and

$$\mathcal{P}_L(\mu, \nu) = V^{**}(\nu) = \sup_{\psi \in C_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \psi(y)\nu(dy) - \int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx) \right\}.$$

By setting $\phi(x) = J_\psi(0, x)$, we have $(\phi, \psi) \in \mathcal{Z}_L$ and we have shown the reverse inequality, that is

$$\mathcal{D}_L(\mu, \nu) \geq \mathcal{P}_L(\mu, \nu),$$

which completes the proof. $\square$

On the other hand, by the standard dynamic programming principle, we get that $J_\psi$ is a viscosity solution of the following dynamic programming equation (see, for instance, [28]):

$$\min \left\{ J_\psi(t, x) - \psi(x), -\partial_t J_\psi(t, x) - \frac{1}{2}\Delta J_\psi(t, x) - H\big(t, x, \nabla J_\psi(t, x)\big) \right\} = 0,$$

where the Hamiltonian $H$ is given by the Legendre dual of $L$, i.e. we have

$$H(t, x, z) = \sup_{u \in U} \big[z \cdot u - L(t, x, u)\big], \quad \text{for all } (t, x, z) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d.$$

This viscosity solution $J_\psi$ can be viewed as the minimal supersolution, given by the infimum of the smooth supersolutions of

$$(3.3) \qquad \begin{cases} \partial_t J(t, x) + \frac{1}{2}\Delta J(t, x) + H\big(t, x, \nabla J(t, x)\big) \leq 0 & \text{in } \mathbb{R}^+ \times \mathbb{R}^d, \\ \psi(x) \leq J(t, x) & \text{on } \mathbb{R}^+ \times \mathbb{R}^d. \end{cases}$$

Finally, we get the following:

**Proposition 3.3.** *We have*

$$\mathcal{D}_L(\mu, \nu) = \sup_{\psi \in C_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \psi(y)\, \nu(dy) - \int_{\mathbb{R}^d} J_\psi(0, x)\, \mu(dx) \right\}$$

$$= \sup_{\psi \in C_b(\mathbb{R}^d),\, J \in C_b^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \psi(y)\nu(dy) - \int_{\mathbb{R}^d} J(0, x)\mu(dx);\ (\psi, J)\ \text{satisfies}\ (3.3) \right\}.$$

## 4. Eulerian formulations

In this section, we express two Eulerian formulations for (2.5); the strong Eulerian formulation poses the problem with a velocity field and the solution to a Fokker-Planck equation with stopping, while the convex Eulerian formulation poses the problem over phase-space distributions satisfying a convex set of inequalities. We prove the later to be equivalent to (2.5) by embedding the stochastic formulation and showing a weak duality inequality. We then prove that when the drift is uniformly bounded, the strong Eulerian formulation is also equivalent by Sobolev estimates. We note that these estimates will be used to prove dual attainment in the next section.

4.1. **Strong and convex Eulerian formulations.** First consider strong Eulerian formulation. We say that $(m, v, \rho)$ belongs to $\mathcal{E}(\mu)$ if the following holds:

- $m \in L^1(\mathbb{R}^+ \times \mathbb{R}^d) \cap L^2_{loc}(\mathbb{R}^+, H^1(\mathbb{R}^d))$ with $||m_t||_{L^1(\mathbb{R}^d)} \leq 1$, $v \in L^\infty(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}^d)$
- $\rho$ is a probability measure on $\mathbb{R}^+ \times \mathbb{R}^d$.
- For all smooth $\phi$ with compact support in $\mathbb{R}^+ \times \mathbb{R}^d$, we have

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \left[ \partial_t \phi(t, x) \, m(t, x) + \nabla \phi(t, x) \cdot \left( v(t, x) \, m(t, x) - \frac{1}{2} \nabla m(t, x) \right) \right] dxdt$$

(4.1)
$$= \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \phi(\tau, y) \rho(d\tau, dy) - \int_{\mathbb{R}^d} \phi(0, x) \mu(dx).$$

Then, we consider the following problem:

(4.2)     $$\mathcal{P}^{\mathcal{E}}_L(\mu, \nu) = \inf_{(m,v,\rho) \in \mathcal{E}(\mu)} \left\{ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} L(t, x, v(t, x)) m(t, x) \, dxdt \ : \ \int_{\mathbb{R}^+} \rho(d\tau, \cdot) = \nu \right\}.$$

Now, let us introduce the following convex Eulerian formulation:

**Definition 4.1.** *We say that $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ if the following holds:*

- *$\eta$ is a measurable map from $\mathbb{R}^+$ to sub-probability measures on $\mathbb{R}^d \times U$,*
- *$\rho$ is a probability measure on $\mathbb{R}^+ \times \mathbb{R}^d$,*
- *The following equation holds for all smooth $\phi$ with $\phi, \partial_t \phi, \nabla \phi$ and $\Delta \phi$ uniformly bounded on $\mathbb{R}^+ \times \mathbb{R}^d$:*

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \left[ \partial_t \phi(t, x) + \frac{1}{2} \Delta \phi(t, x) + \nabla \phi(t, x) \cdot u \right] \eta(dt, dx, du)$$

(4.3)
$$= \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \phi(\tau, y) \rho(d\tau, dy) - \int_{\mathbb{R}^d} \phi(0, x) \mu(dx).$$

We let

(4.4)     $$\mathcal{P}^{\tilde{\mathcal{E}}}_L(\mu, \nu) = \inf_{(\eta,\rho) \in \tilde{\mathcal{E}}(\mu)} \left\{ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u) \eta(dt, dx, du) \ : \ \int_{\mathbb{R}^+} \rho(d\tau, \cdot) = \nu \right\}.$$

It is clear that $\mathcal{P}^{\tilde{\mathcal{E}}}_L(\mu, \nu) \leq \mathcal{P}^{\mathcal{E}}_L(\mu, \nu)$ because if $(m, v, \rho) \in \mathcal{E}(\mu)$ then for $\eta(dt, dx, du) = \delta_{v(t,x)}(du) m(t, x) dtdx$, we have $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ with the same cost and target distribution. The following proposition shows that from a probability distribution, drift, and stopping time $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, we can construct an admissible pair $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$.

**Proposition 4.1.** *Given $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, there is a pair $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ such that*

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u) \eta(dt, dx, \mathrm{d}u) = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \left[ \int_0^\tau L(t, X_t, \beta_t) dt \right].$$

*In particular, we have*

$$\mathcal{P}^{\tilde{\mathcal{E}}}_L(\mu, \nu) \leq \mathcal{P}_L(\mu, \nu).$$

*Proof.* Given $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, we find the pair $(\eta, \rho)$ from Riesz representation given by the formula,

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \phi(t, x, u) \eta(dt, dx, du) = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \left[ \int_0^\tau \phi(t, X_t, \beta_t) dt \right], \text{ for all } \phi \in C_b(\mathbb{R}^+ \times \mathbb{R}^d \times U),$$

and

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} h(\tau, y) \rho(d\tau, dy) = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \left[ h(\tau, X_\tau) \right], \text{ for all } h \in C_b(\mathbb{R}^+ \times \mathbb{R}^d).$$

Then, one can check easily that the pair $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$. In fact, if $\eta_t$ is the disintegration of $\eta$ with respect to the Lebesgue measure on $\mathbb{R}^+$ then, for every smooth $\phi$ with compact support in $\mathbb{R}^+ \times \mathbb{R}^d$, we have using Ito's formula the following:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \Big[ \partial_t \phi(t,x) + \frac{1}{2} \Delta \phi(t,x) + u \cdot \nabla \phi(t,x) \Big] \eta_t(dx, du) dt$$

$$= \mathbb{E}^{\alpha \ltimes \mathbb{P}} \Big[ \int_0^\tau \Big( \partial_t \phi(t, X_t) + \beta_t \cdot \nabla \phi(t, X_t) + \frac{1}{2} \Delta \phi(t, X_t) \Big) dt \Big]$$

$$= \mathbb{E}^{\alpha \ltimes \mathbb{P}} \Big[ \phi(\tau, X_\tau) - \phi(0, X_0) \Big] = \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \phi(\tau, y) \rho(d\tau, dy) - \int_{\mathbb{R}^d} \phi(0, x) \mu(dx).$$

Clearly, we have that if $X_\tau \sim_{\alpha \ltimes \mathbb{P}} \nu$ then $\int_{\mathbb{R}^+} \rho(d\tau, \cdot) = \nu$. Moreover, we have the following:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u) \eta(dt, dx, du) = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \Big[ \int_0^\tau L(t, X_t, \beta_t) dt \Big],$$

which completes the proof that $\mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu) \leq \mathcal{P}_L(\mu, \nu)$. $\square$

On the other hand, we have the following duality for the convex Eulerian problem (4.4).

**Theorem 4.2.** *The following equalities hold:*

$$D_L(\mu, \nu) = \mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu) = \mathcal{P}_L(\mu, \nu).$$

*Proof.* Take an admissible pair $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ with $\int_{\mathbb{R}^+} \rho(d\tau, \cdot) = \nu$ and let $(J, \psi)$ satisfy (3.3). Then, we have

$$\int_{\mathbb{R}^d} \psi(y) \, \nu(dy) - \int_{\mathbb{R}^d} J(0, x) \, \mu(dx)$$

$$\leq \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} J(t, x) \, \rho(dt, dx) - \int_{\mathbb{R}^d} J(0, x) \, \mu(dx)$$

$$= \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \Big[ \partial_t J(t, x) + \frac{1}{2} \Delta J(t, x) + u \cdot \nabla J(t, x) \Big] \eta_t(dx, du) dt$$

$$\leq \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u) \eta(dt, dx, du).$$

This shows that $\mathcal{D}_L(\mu, \nu) \leq \mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu)$ in view of Proposition 3.3. We have shown that $\mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu) \leq \mathcal{P}_L(\mu, \nu)$ in Proposition 4.1, and that $\mathcal{P}_L(\mu, \nu) = \mathcal{D}_L(\mu, \nu)$ in Theorem 3.2, which completes the chain of equalities. $\square$

4.2. **Regularity.** We will now partly complete the equivalence (between the strong and the convex Eulerian formulations) by addressing the strong Eulerian formulation in the case where the drift is bounded. We first need a result on the truncation in time and space of pairs $(\eta, \rho) \in \mathcal{E}(\mu)$. We make use of the truncation in time and space of Lemma 2.2.

**Theorem 4.3.** *Suppose that $\mu = \hat{\mu} \, dx$ for $\hat{\mu} \in L^2(\mathbb{R}^d)$. Then, for any $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ with compact support in $\mathbb{R}^+ \times \mathbb{R}^d \times U$ $(0 \leq t < T, |x| < R, |u| < \bar{u})$, there is a $(m, v, \rho) \in \mathcal{E}(\mu)$ such that $\eta_t(dx, \mathbb{R}^d) = m(t, x) dx$, and $(m, v, \rho)$ has a cost less than or equal to that of $(\eta, \rho)$. Moreover, we have the following uniform estimate*

$$(4.5) \qquad \int_0^T \int_{\mathbb{R}^d} |\nabla m(t, x)|^2 dx \, dt \leq 2 \|\mu\|_{L^2(\mathbb{R}^d)}^2 + CT,$$

where the constant $C$ depends only on the bound of the drift $\bar{\mathbf{u}}$ and the dimension $d$. In particular,

$$f \mapsto \int_{\mathbb{R}^d} \int_0^T f(x)\rho(d\tau, dx)$$

is a continuous linear functional of $H^1(\mathbb{R}^d)$. Finally, when the drift is bounded, we have

$$\mathcal{P}_L^{\tilde{\mathcal{E}}}(\mu, \nu) = \mathcal{P}_L^{\mathcal{E}}(\mu, \nu).$$

*Proof.* First, we use convolution to approximate the pair $(\eta, \rho)$ by smooth densities $(\eta^\epsilon, \rho^\epsilon)$ and the measure $\mu$ by $\mu^\varepsilon$ with $(\eta^\epsilon, \rho^\epsilon) \in \tilde{\mathcal{E}}(\mu^\epsilon)$. We define

$$m^\epsilon(t, x) = \int_{\mathbb{R}^d} \eta^\epsilon(t, x, u)du$$

and

$$v^\epsilon(t, x) = \frac{\int_{\mathbb{R}^d} u\,\eta^\epsilon(t, x, u)du}{m^\epsilon(t, x)},$$

with $v^\epsilon(t, x) = 0$ if $m^\epsilon(t, x) = 0$. We note that $(m^\epsilon, v^\epsilon, \rho^\epsilon) \in \mathcal{E}(\mu^\epsilon)$. By Jensen's inequality, the cost of $(m^\epsilon, v^\epsilon, \rho^\epsilon)$ is less or equal to that of $(\eta, \rho)$ within a factor of $\epsilon$. Then, using $m^\epsilon$ as a test function in (4.1), we obtain

$$\int_{\mathbb{R}^d} \int_0^T m^\epsilon(\tau, y)\rho^\epsilon(d\tau, dy) - \int_{\mathbb{R}^d} \left(\hat{\mu}^\epsilon(x)\right)^2 dx$$

$$= \int_0^T \int_{\mathbb{R}^d} \left[\frac{1}{2}\partial_t\left(m^\epsilon(t, x)^2\right) - \frac{1}{2}|\nabla m^\epsilon(t, x)|^2 + v^\epsilon(t, x) \cdot \nabla m^\epsilon(t, x)\, m^\epsilon(t, x)\right] dx\, dt.$$

Then,

$$\frac{1}{2}\int_0^T \int_{\mathbb{R}^d} |\nabla m^\epsilon(t, x)|^2 dx\, dt \leq \frac{1}{2}\|\hat{\mu}^\epsilon\|_{L^2(\mathbb{R}^d)}^2 + \bar{\mathbf{u}}\int_0^T \int_{\mathbb{R}^d} |\nabla m^\epsilon(t, x)|m^\epsilon(t, x)dx\, dt.$$

Yet, we have

$$\int_0^T \int_{\mathbb{R}^d} |\nabla m^\epsilon(t, x)|m^\epsilon(t, x)dx\, dt \leq \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} \left(\delta|\nabla m^\epsilon(t, x)|^2 + \frac{1}{\delta}m^\epsilon(t, x)^2\right)dx\, dt.$$

For the last term using Gagliardo-Nirenberg-Sobolev embedding and an interpolation of the $L^1$ and $L^{2^*}$ norms, we have

$$\|m_t^\epsilon\|_{L^2(\mathbb{R}^d)} \leq \delta\|\nabla m_t^\epsilon\|_{L^2(\mathbb{R}^d)} + C(\delta)\|m_t^\epsilon\|_{L^1(\mathbb{R}^d)},$$

which implies that

$$\int_0^T \int_{\mathbb{R}^d} m^\epsilon(t, x)^2 dx\, dt \leq 2\delta^2 \int_0^T \int_{\mathbb{R}^d} |\nabla m^\epsilon(t, x)|^2 dx\, dt + 2C(\delta)^2 \int_0^T \left(\int_{\mathbb{R}^d} m^\epsilon(t, x)dx\right)^2 dt.$$

Yet, $\|m^\epsilon(t, \cdot)\|_{L^1(\mathbb{R}^d)} \leq 1 + \epsilon$, for every $t$. Then, choosing $\delta > 0$ small enough, we get (4.5) for $m^\epsilon$. In particular, the uniform estimates imply that $m^\epsilon$ converge weakly to $m$ in $L_{loc}^2(\mathbb{R}^+, H^1(\mathbb{R}^d))$ such that

$$m(t, \cdot) = \frac{\int_{\mathbb{R}^d} \eta_t(\cdot, du)}{dx}.$$

We then define the field $v(t, x)$ by the vector-valued Radon-Nikodym derivative

$$v(t, \cdot) = \frac{\int_{\mathbb{R}^d} u\,\eta_t(\cdot, du)}{m_t\, dx}.$$

It is then straightforward to see $(m, v, \rho) \in \mathcal{E}(\mu)$ which by Jensen's inequality, has lesser or equal cost than $(\eta, \rho)$.

For any $f \in H^1(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} f(x) \rho(d\tau, dx)$$

$$= \int_{\mathbb{R}^d} f(x) \hat{\mu}(x) \, dx + \int_0^T \int_{\mathbb{R}^d} \left[ -\frac{1}{2} \nabla f(x) \cdot \nabla m(t, x) + v(t, x) \cdot \nabla f(x) \, m(t, x) \right] dx \, dt,$$

and the similar estimates imply that $f \mapsto \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} f(x) \rho(d\tau, dx)$ is a linear functional of $H^1(\mathbb{R}^d)$.

Finally, to conclude equivalence of the convex and strong formulations, we use Proposition 4.1 and Lemma 2.2 to construct $(\eta^{T,R}, \rho^{T,R})$ with compact support in time and space for any $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu)$, corresponding to the truncation of Lemma 2.2 with $T, R \in \mathbb{R}^+$. We then have corresponding $(m^{T,R}, v^{T,R}, \rho^{T,R}) \in \mathcal{E}(\mu)$. When $T_1 < T_2$ and $R_1 < R_2$, we have $(\eta^{T_2, R_2}, \rho^{T_2, R_2}) = (\eta^{T_1, R_1}, \rho^{T_1, R_1})$ on $[0, T_1] \times B(0, R_1)$, and thus there exists a density $(m, v, \rho) \in \mathcal{E}(\mu, \nu)$ taking the limit as $T, R \to \infty$, with the same cost and end distribution as $(\mathbb{P}, \beta, \alpha)$. Thus, we have shown that

$$\mathcal{P}_L^{\mathcal{E}}(\mu, \nu) = \mathcal{P}_L(\mu, \nu),$$

and applying the result of Theorem 4.2 completes the proof of equivalence. $\qquad\square$

We will also need the following moment bound.

**Proposition 4.4.** *Suppose that* $c(|u|^p + |x|^q + 1) \le L(t, x, u)$ *or* $c(|x| + 1) \le L(t, x, u)$ *and* $u$ *is uniformly bounded, for* $c > 0$. *We assume that* $\mu$ *satisfies* $\int_{\mathbb{R}^d} |x|^2 \mu(dx) < +\infty$. *Then, for any* $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ *with finite cost, we have*

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} |y|^2 \rho(d\tau, dy) \le \int_{\mathbb{R}^d} |x|^2 \mu(dx) + C \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u) \eta_t(dx, du) dt.$$

*Proof.* To prove that, we simply apply (4.3) with the test function $w(y) = |y|^2$:

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} |y|^2 \rho(d\tau, dy) - \int_{\mathbb{R}^d} |x|^2 \mu(dx) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \left[ d + u \cdot 2x \right] \eta_t(dx, du) dt$$

$$\le \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U \left[ d + H(t, x, 2x) + L(t, x, u) \right] \eta_t(dx, du) dt$$

$$\le C \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left[ L(t, x, u) \right] \eta_t(dx, du) dt,$$

where we have used that

$$H(t, x, 2x) \le C |x|^q \le C \, L(t, x, u).$$

In the case that $c|x| \le L(t, x, u)$ and $u$ is uniformly bounded, the proof is simpler, using that

$$u \cdot x \le \frac{\overline{u}}{c} L(t, x, u),$$

and the result follows. $\quad\square$

## 5. Dual attainment

In this section, we prove dual attainment in the cases where either the Lagrangian $L \approx |u|^p$ with $1 < p < 2$ or the drift is uniformly bounded ($|u| \leq \bar{u}$). This relies on a normalization that makes $\psi$ as a supersolution to an HJB equation. First, we define

$$\bar{H}(x, z) = \inf_{t \in \mathbb{R}^+} H(t, x, z), \text{ for all } (x, z) \in \mathbb{R}^d \times \mathbb{R}^d,$$

and

$$\bar{L}(x, u) = \sup_{t \in \mathbb{R}^+} L(t, x, u), \text{ for all } (x, u) \in \mathbb{R}^d \times U.$$

We suppose, strengthening assumption (2.1), that there are constants $c, C > 0$ such that $L$ satisfies

(5.1) $$c\big(|u|^p + |x|^q + 1\big) \leq L(t, x, u) \leq C(|u|^p + |x|^q + 1)$$

for all $(t, x, u) \in \mathbb{R}^+ \times \mathbb{R}^d \times U$, or equivalently, there are constants $\lambda, \Lambda, c, C > 0$ such that $H$ satisfies

$$\lambda |z|^q - C\big(|x|^q + 1\big) \leq H(t, x, z) \leq \Lambda |z|^q - c\big(|x|^q + 1\big), \text{ for all } (t, x, z) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d.$$

In the case where the drift is bounded, this becomes

$$c(|x| + 1) \leq L(t, x, u) \leq C(|x| + 1), \text{ for all } (t, x, u) \in \mathbb{R}^+ \times \mathbb{R}^d \times U,$$

and

$$\lambda |z| - C\big(|x| + 1\big) \leq H(t, x, z) \leq \Lambda |z| - c\big(|x| + 1\big), \text{ for all } (t, x, z) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d.$$

**Proposition 5.1.** *For $\psi \in C_b(\mathbb{R}^d)$, we define*

$$\bar{\psi}(x) := \sup_{(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_x)} \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau \bar{L}(X_t, \beta_t) dt\Big].$$

*Then, $\bar{\psi}$ is lower semi-continuous and bounded with $J_{\bar{\psi}} = J_\psi$ and thus greater or equal dual value. Furthermore, $\bar{\psi}$ satisfies, in the viscosity sense,*

(5.2) $$\frac{1}{2}\Delta\bar{\psi}(x) + \overline{H}\big(x, \nabla\bar{\psi}(x)\big) \leq 0.$$

*Proof.* From the definition of $\bar{\psi}$, we have obviously $\bar{\psi} \geq \psi$, and we get as in Lemma 3.1 that $\bar{\psi}$ is lower semi-continuous and bounded. The standard viscosity solution theory implies that $\bar{\psi}$ is a viscosity supersolution of (5.2).

From the definition of $J_\psi$, we get that $J_{\bar{\psi}} \geq J_\psi$. Let us prove the reverse inequality, that is $J_{\bar{\psi}} \leq J_\psi$, so that we get $J_{\bar{\psi}} = J_\psi$. First, we note that $\bar{\psi}(x) \leq J_\psi(t, x)$, for every $(t, x) \in \mathbb{R}^+ \times \mathbb{R}^d$, which follows from the definitions after noting that $\bar{L} \geq L$.

Now, suppose that $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}_t(\delta_x)$ is within $\epsilon$ of optimality for $J_{\bar{\psi}}$, then we have the following:

$$J_{\bar{\psi}}(t, x) - \epsilon \leq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[\bar{\psi}(X_\tau) - \int_t^\tau L(s, X_s, \beta_s) ds\Big]$$

$$\leq \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[J_\psi(\tau, X_\tau) - \int_t^\tau L(s, X_s, \beta_s) ds\Big] \leq J_\psi(t, x),$$

where the last inequality is a result of dynamic programming principle [19, Theorem 6, Ch. 3] for $J_\psi$. Taking $\epsilon$ to zero proves the desired inequality.    $\square$

The following proposition proves a quadratic lower bound on supersolutions to (5.2) as well as an Hölder continuity in the case that $L \approx |u|^p$ with $1 < p < 2$.

**Proposition 5.2.** *We assume* (5.1) *holds. Suppose $\psi$ is bounded, lower semi-continuous and satisfies* (5.2). *Suppose $d > 1$ and $1 < p < 2$. Fix $0 < \delta \leq 2 - p < 1$. Then, for each $x_0 \in \mathbb{R}^d$, there are two constants $B$ and $E$ (depending only on $\delta$, $p$, $d$, $\lambda$, $C$ and $|x_0|$) such that*

$$(5.3) \qquad \psi(x_0) - \psi(x_1) \leq B|x_1 - x_0|^\delta + E|x_1 - x_0|^2, \quad \text{for all } x_1 \in \mathbb{R}^d.$$

*In particular, $\psi$ is uniformly $\delta$-Hölder continuous on compact sets and, under the assumption that $\psi(0) = 0$, $\psi$ is uniformly globally bounded from below by a quadratic function.*

*In the case $d = 1$, the result holds with $\delta = 1$ for all $p > 1$.*

*Proof.* We will prove that, for each $x_0$, the function

$$w(x) = A - B|x - x_0|^\delta - E|x - x_0|^2$$

with appropriate constants $A$, $B$ and $E$, will touch $\psi$ from below at $x_0$. By computing $\nabla w$ at $x \neq x_0$, we see that

$$\lambda|\nabla w(x)|^q \geq b_1|x - x_0|^{q(\delta-1)} + b_2|x - x_0|^q$$

where

$$b_1 = \delta^q \lambda B^q \text{ and } b_2 = \lambda 2^q E^q.$$

Notice that as $\psi$ is l.s.c. and bounded, we can by adjusting the constant $A$, let the function $w$ touch $\psi$ from below at some point. Suppose $w$ touches $\psi$ from below at $x_1 \neq x_0$. Then, from the equation (5.2) and the assumption (5.1), we have that

$$C\big(|x_1|^q + 1\big) \geq \frac{1}{2}\Delta w(x_1) + \lambda|\nabla w(x_1)|^q$$

$$\geq b_1|x_1 - x_0|^{q(\delta-1)} + b_2|x_1 - x_0|^q - \frac{\delta B}{2}(d + \delta - 2)|x_1 - x_0|^{\delta-2} - dE.$$

To draw a contradiction, we will select two constants $B$ and $E$ such that the following holds

$$b_1|x_1 - x_0|^{q(\delta-1)} + b_2|x_1 - x_0|^q$$

$$> 3\max\left\{\frac{\delta B}{2}(d + \delta - 2)|x_1 - x_0|^{\delta-2}, \ dE + C, \ C|x_1|^q\right\}.$$

Now, it is clear that the inequality with the first and the second item in the maximum is satisfied for large enough $B$ and $E$ such that

$$\min\{b_1, b_2\} > 3\max\left\{\frac{\delta B}{2}(d + \delta - 2), dE + C\right\}.$$

For the third item, note that

$$|x_1|^q \leq 2^q \max\big\{|x_1 - x_0|, |x_0|\big\}^q.$$

Thus, we may choose $B$ and $E$ large enough, depending on $|x_0|$, such that $b_1 > 3C(1+|x_0|)^q$ and $b_2 > 3C2^q \max\{1, |x_0|^q\}$, and these choices of $B$ and $E$ yield a contradiction. Thus, we must have that $w$ touches $\psi$ from below at $x_1 = x_0$, and it follows that (5.3) holds for all $x_0$ and $x_1$. We note that in particular, fixing $x_0$, this establishes a global quadratic lower bound on $\psi$, which becomes uniform (in $\psi$) as soon as $\psi(0) = 0$.

The proof for $d = 1$ is simpler with $\delta = 1$ since the Laplacian of the second term of $w$ vanishes. $\square$

**Theorem 5.3.** *Suppose that $d \geq 2$, $1 < p < 2$, $\int_{\mathbb{R}^d} |x|^2 \mu(dx) < +\infty$, $\nu$ has a compact support, and $L$ satisfies (5.1). Then, the dual problem $\mathcal{D}_L(\mu, \nu)$ is attained at $\psi \in C_{loc}^\delta(\mathbb{R}^d)$, $0 < \delta < 2 - p$, and $\psi$ is globally bounded below by a quadratic function.*

*Furthermore, in this case there exists $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$ with finite cost.*

*Proof.* We take a maximizing sequence $\{\psi^i\}$. We assume that, for each $i \in \mathbb{N}$, $\psi^i$ is l.s.c., bounded, $\psi^i(0) = 0$ and $\psi^i$ satisfies (5.2) (this is possible thanks to Proposition 5.1), and we apply Proposition 5.2. Then, by Arzela-Ascoli, $\{\psi^i\}$ converges uniformly on compact sets to $\psi$ with $\psi(0) = 0$. Furthermore, these $\psi^i$ are uniformly bounded below by a quadratic function. With the target measure $\nu$ compactly supported, for such a limit function $\psi$ to have the maximal dual value, it is enough to show that

$$
(5.4) \qquad \liminf_i \int_{\mathbb{R}^d} J_{\psi^i}(0, x)\mu(dx) \geq \int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx).
$$

But again, we have

$$
\int_{\mathbb{R}^d} J_{\psi^i}(0, x)\mu(dx) = \sup_{(\eta,\rho) \in \tilde{\mathcal{E}}(\mu)} \left[ \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \psi^i(y)\rho(dt, dy) - \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_U L(t, x, u)\eta(dt, dx, du) \right].
$$

For each $(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)$ with finite cost, Proposition 4.4 implies that such $\rho$ has finite second moment. Thus, the uniform quadratic lower bound on $\psi^i$ implies that

$$
\liminf_i \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \psi^i(y)\rho(dt, dy) \geq \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \psi(y)\rho(dt, dy).
$$

This yields (5.4). Finally, the existence of $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$ with finite cost follows from duality (Theorem 3.2).  $\square$

In 1D the result holds for all values of $p > 1$.

**Theorem 5.4.** *Let $d = 1$. Assume that $\int_{\mathbb{R}^d} |x|^2 \mu(dx) < +\infty$, $\nu$ has a compact support and $L$ satisfies (5.1). Then, the dual problem is attained at $\psi$, which is locally Lipschitz continuous and bounded below quadratically.*

*Proof.* The proof follows as in Theorem 5.3 using the Lipschitz estimates in 1D from Proposition 5.2.  $\square$

In order to better handle the behavior as $|x| \to +\infty$ in the case that $p$ is large (for instance, assume that $p = +\infty$ which means that the drift is bounded; $|u| \leq \bar{u}$), we will introduce a weighting measure on $\mathbb{R}^d$. We consider now a smooth convex function $V$ such that $\int_{\mathbb{R}^d} e^{-V(x)} dx = 1$ and $\frac{1}{2}|\nabla V| \leq \lambda$, for instance $V(x) = \gamma|x| + m$ (we assume that $\lambda > \frac{\gamma}{2}$) is easily seen to fit all the criteria we will require. We define the norm

$$
\|f\|_{L_V^q(\mathbb{R}^d)} = \left( \int_{\mathbb{R}^d} |f(x)|^q e^{-V(x)} dx \right)^{\frac{1}{q}}.
$$

Set

$$
W_V^{1;q}(\mathbb{R}^d) := \left\{ f; \ \|f\|_{L_V^q(\mathbb{R}^d)} + \|\nabla f\|_{L_V^q(\mathbb{R}^d)} < +\infty \right\}.
$$

We denote

$$
(f)_V = \int_{\mathbb{R}^d} f(x) e^{-V(x)} dx.
$$

We note that if $V$ is a radially increasing function, then we have, for any $q \geq 1$, the following Poincaré inequality (see, for example, [9, 7]):

$$\|f - (f)_V\|_{L_V^q(\mathbb{R}^d)} \leq C\|\nabla f\|_{L_V^q(\mathbb{R}^d)}.$$

Then, we define

(5.5) $$\mathcal{B} := \left\{ f \in L_V^1(\mathbb{R}^d), \ f \text{ is l.s.c and quadratically bounded from below} \right\}.$$

We will say that $f \in H_V^1(\mathbb{R}^d) := W_V^{1,2}(\mathbb{R}^d)$ satisfies (5.2) in a weak sense if for all compactly supported $h \in H_V^1(\mathbb{R}^d)$ with $h \geq 0$, we have

$$\int_{\mathbb{R}^d} \left[ -\frac{1}{2}\nabla f(x) \cdot \nabla h(x) + \bar{H}(x, \nabla f(x))h(x) \right] dx \leq 0.$$

Then, we have the following:

**Proposition 5.5.** *Suppose that $\bar{H}$ is uniformly continuous in $x$ (uniformly w.r.t. $z$) and satisfies (5.1). Let $\psi$ be a supersolution to (5.2), lower semi-continuous and bounded from below with $(\psi)_V = 0$, then, for any $M \geq 0$, the truncation of $\psi$ satisfies $\psi \wedge M \in H_V^1(\mathbb{R}^d)$ and solves (5.2) in a weak sense. More precisely, there exists a uniform constant $\overline{C}$ (which does not depend on $\psi$) such that*

(5.6) $$\|\psi\|_{W_V^{1,1}(\mathbb{R}^d)} \leq \overline{C},$$

(5.7) $$\psi \geq -\overline{C}(1 + |x|^2),$$

*and, for every $M \geq 0$, there is a constant $C(M)$ (again uniform in $\psi$) such that the following holds:*

(5.8) $$\|\psi \wedge M\|_{H_V^1(\mathbb{R}^d)} \leq C(M).$$

*On the other hand, if the truncation $\psi \wedge M \in H_V^1(\mathbb{R}^d)$ solves (5.2) weakly for all $M \geq 0$, then $\psi$ is l.s.c.*

*Proof.* First, we fix $M \geq 0$ and show there is a sequence $(\psi_\varepsilon^M)_\varepsilon$ such that $\psi_\varepsilon^M$ is $1/\varepsilon-$Lipschitz and semi-concave, $\psi_\varepsilon^M \rightharpoonup \psi \wedge M$ in $H_V^1(\mathbb{R}^d)$ and $\psi_\varepsilon^M$ satisfies (in a strong sense)

(5.9) $$\frac{1}{2}\Delta\psi_\varepsilon^M(x) + \overline{H}(x, \nabla\psi_\varepsilon^M(x)) \leq C(M, \varepsilon).$$

Set $\psi_\varepsilon^M(x) = \inf_{z \in \mathbb{R}^d}\{\psi(z) \wedge M + \frac{1}{2\varepsilon^2}|x - z|^2\}$, for all $x \in \mathbb{R}^d$. We let $z_x$ (depending also on $\varepsilon, M$) be such that $\psi_\varepsilon^M(x) = \psi(z_x) \wedge M + \frac{1}{2\varepsilon^2}|x - z_x|^2$, which is well defined for sufficiently small $\varepsilon$ by lower semi-continuity and the lower bound on $\psi$. Fix $\varepsilon > 0$ and $x \in \mathbb{R}^d$ and let $\phi$ be a smooth function such that $\phi \leq \psi_\varepsilon^M$ with $\phi(x) = \psi_\varepsilon^M(x)$. We define $w(y) := \phi(y + x - z_x) - \frac{1}{2\varepsilon^2}|x - z_x|^2$. We have $w(z_x) = \psi(z_x) \wedge M$ and $w \leq \psi \wedge M$. Hence, by (5.2) and that $\overline{H}(x, 0) \leq 0$,

$$\frac{1}{2}\Delta w(z_x) + \overline{H}(z_x, \nabla w(z_x)) \leq 0.$$

Yet, $\nabla w(z_x) = \nabla\phi(x)$ and $\Delta w(z_x) = \Delta\phi(x)$. Using also the fact that $\bar{H}$ is uniformly continuous, we get

$$\frac{1}{2}\Delta\phi(x) + \overline{H}(x, \nabla\phi(x)) \leq C(|x - z_x|).$$

Moreover, we have
$$\psi(z_x) \wedge M + \frac{1}{2\varepsilon^2}|x - z_x|^2 \leq \psi(x) \wedge M.$$

Hence, $|x - z_x| \leq C(M, \varepsilon)$, and it follows that $\psi_\varepsilon^M$ is a supersolution in the viscosity sense with error $C(M, \varepsilon)$ where $C(M, \varepsilon) \to 0$ as $\varepsilon \to 0$, but also in the sense of distributions thanks to the semi-concavity of $\psi_\varepsilon^M$. Using $e^{-V}$ as a test function, we obtain

$$C(M, \varepsilon) \geq \int_{\mathbb{R}^d} \Big[\frac{1}{2}\Delta\psi_\varepsilon^M(x) + \overline{H}\big(x, \nabla\psi_\varepsilon^M(x)\big)\Big]e^{-V(x)}dx$$

$$\geq -\frac{1}{2}\int_{\mathbb{R}^d} |\nabla\psi_\varepsilon^M(x)|\,|\nabla V(x)|e^{-V(x)}dx + \lambda\|\nabla\psi_\varepsilon^M\|_{L_V^1} - C.$$

Yet, $\gamma < 2\lambda$. Then, we infer that
$$\|\nabla\psi_\varepsilon^M\|_{L_V^1(\mathbb{R}^d)} \leq C,$$

which shows the estimate (5.6) after taking $\varepsilon \to 0$. For the uniform quadratic lower bound: let us consider as in Proposition 5.2 the function $w(y) := A - \frac{E}{2}|y|^2$ that touches $\psi_\varepsilon^0$ from below at $x$. We then have that

$$-\frac{d}{2}E + \lambda E|x| - C(|x| + 1) \leq C(\varepsilon),$$

which implies that, for sufficiently large $E$, $x$ is in a ball of radius $C_1 = C_1(d, \lambda, C, E)$ (we note that the constant $E$ can be taken independent of $x$). Then,

$$\psi_\varepsilon^0(y) \geq \inf_{|x| \leq C_1} \psi_\varepsilon^0(x) + \frac{E}{2}|x|^2 - \frac{E}{2}|y|^2.$$

To bound $\psi_\varepsilon^0$ near the origin we will consider a construction that will be useful for the remainder of the proof. For any $x$, we let $Q_{x,R}$ denote the uniform distribution on the ball of radius $R$ centered at $x$. Fix $0 < R_1 < R_2$ and assume that the drift equals zero, then we can find a pair $(\eta, \rho) = (m, 0, \rho) \in \mathcal{E}(Q_{x,R_1}, Q_{x,R_2})$ such that $m$ has support in $[0, T] \times Q_{x,R_2}$. From (4.3), we have

$$\int_{\mathbb{R}^d} \frac{|y - x|^2}{d} Q_{x,R_2}(dy) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \eta_t(dy, du)dt + \int_{\mathbb{R}^d} \frac{|y - x|^2}{d} Q_{x,R_1}(dy),$$

which implies that

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} m(t, y)dy\, dt = C(R_2^2 - R_1^2).$$

Define
$$\psi_\varepsilon^{M,R}(x) := \int_{\mathbb{R}^d} \psi_\varepsilon^M(y)Q_{x,R}(dy).$$

As $m(t, \cdot) \in H^1(\mathbb{R}^d)$ (see Theorem 4.3), for each $t \in [0, T]$, with compact support, then again by (4.3) and thanks to the fact that $\overline{H}$ is uniformly continuous w.r.t. $x$, we have

$$\psi_\varepsilon^{M,R_2}(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi_\varepsilon^M(y)\rho(d\tau, dy)$$

$$= \int_{\mathbb{R}^d} \psi_\varepsilon^M(y)Q_{x,R_1}(dy) - \frac{1}{2}\int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \nabla\psi_\varepsilon^M(y) \cdot \nabla m(t, y)\, dy\, dt$$

$$\leq \psi_\varepsilon^{M,R_1}(x) - \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \overline{H}\big(y, \nabla\psi_\varepsilon^M(y)\big)m(t, y)\, dy\, dt$$

$$\leq \psi_\varepsilon^{M,R_1}(x) + C(x)(R_2^2 - R_1^2).$$

Letting $R_1 \to 0^+$ and $R_2 = 1$, we find a uniform lower bound for $\psi_\varepsilon$ with $|x| \leq C_1$ using (5.6) and the Poincaré inequality to bound $\psi_\varepsilon$ on balls of radius 1. This concludes the proof of (5.7). We now use $M - \psi_\varepsilon^M$ as a test function in (5.9), so we get

$$\int_{\mathbb{R}^d} C(M,\varepsilon)\big(M - \psi_\varepsilon^M(x)\big)e^{-V(x)}dx$$

$$\geq \int_{\mathbb{R}^d} \big(M - \psi_\varepsilon^M(x)\big)\Big[\frac{1}{2}\Delta\psi_\varepsilon^M(x) + \overline{H}\big(x, \nabla\psi_\varepsilon^M(x)\big)\Big]e^{-V(x)}dx$$

$$= \int_{\mathbb{R}^d} \Big[\frac{1}{2}|\nabla\psi_\varepsilon^M(x)|^2 + \big(M - \psi_\varepsilon^M(x)\big)\Big(\frac{1}{2}\nabla\psi_\varepsilon^M(x)\cdot\nabla V(x) + \overline{H}\big(x, \nabla\psi_\varepsilon^M(x)\big)\Big)\Big]e^{-V(x)}dx$$

$$\geq \int_{\mathbb{R}^d} \Big[\frac{1}{2}|\nabla\psi_\varepsilon^M(x)|^2 - \big(M - \psi_\varepsilon^M(x)\big)\overline{L}\big(x, -\frac{1}{2}\nabla V(x)\big)\Big]e^{-V(x)}dx.$$

Using the uniform quadratic lower bound on $\psi_\varepsilon^M$, the bounds on $L$ and the estimate (5.6), we get

$$\|\nabla\psi_\varepsilon^M\|_{L^2_V(\mathbb{R}^d)}^2 \leq M\big(C + C(M,\varepsilon)\big) + C,$$

which yields the estimate (5.8). To see that $\psi \wedge M$ is a distributional solution to (5.2), we fix a smooth compactly supported test function $h \geq 0$ and consider the limit

$$0 \geq \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} \Big[-\frac{1}{2}\nabla\psi_\varepsilon^M(x)\cdot\nabla h(x) + \overline{H}\big(x, \nabla\psi_\varepsilon^M(x)\big)h(x)\Big]dx$$

$$\geq \int_{\mathbb{R}^d} \Big[-\frac{1}{2}\nabla\psi^M(x)\cdot\nabla h(x) + \overline{H}\big(x, \nabla\psi^M(x)\big)h(x)\Big]dx,$$

where the second inequality follows from convexity of $\overline{H}$ and hence weak lower semi-continuity of the integral.

For the last statement: assume that $\psi \wedge M$ solves (5.2) weakly for all $M \geq 0$. Recalling the previous estimates, we have

$$\psi^{M,R_2}(x) - C(x)R_2^2 \leq \psi^{M,R_1}(x) - C(x)R_1^2.$$

It follows that $R \mapsto \psi^{M,R}(x) - C(x)R^2$ is monotonically decreasing, which implies that $\psi^M$ (and so, $\psi$) is lower semi-continuous (see [27]). $\square$

**Remark 5.6.** *In the case that $\psi \in H^1_V(\mathbb{R}^d)$, bounded above, and $\psi$ satisfies (5.2) in a weak sense, $\psi$ is lower semi-continuous by Proposition 5.5 and $\psi$ will satisfy (5.2) in the viscosity sense. Arguments for this can be found in [21] and [15]. To sketch an argument for this, we note that $\psi^R(x) := \int_{\mathbb{R}^d} \psi(y)Q_{x,R}(dy)$ satisfies*

$$\psi^R(x) \geq \mathbb{E}^{\alpha \times \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau \overline{L}(X_t, \beta_t)dt\Big]$$

*for all $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(Q_{x,R})$. Taking the limit as $R \to 0$ we have $\psi(x) = \lim_{R \to 0} \psi^R(x)$, and any $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\delta_x)$ can be translated to $(\mathbb{P}^R, \beta^R, \alpha^R) \in \mathcal{A}(Q_{x,R})$, showing that*

$$\psi(x) \geq \lim_{R \to 0} \mathbb{E}^{\alpha^R \times \mathbb{P}^R}\Big[\psi(X_\tau) - \int_0^\tau \overline{L}(X_t, \beta_t)dt\Big] \geq \mathbb{E}^{\alpha \times \mathbb{P}}\Big[\psi(X_\tau) - \int_0^\tau \overline{L}(X_t, \beta_t)dt\Big],$$

*which implies that $\psi$ solves (5.2) in the sense of viscosity.*

**Lemma 5.7.** *The map $(t,x) \mapsto J_\psi(t,x)$ is lower semi-continuous for all $\psi \in \mathcal{B}$. Moreover, under the assumptions of Proposition 5.5, we have, for any $\mu \in L^2_{-V}(\mathbb{R}^d)$, $\psi \mapsto \int_{\mathbb{R}^d} J_\psi(0,x)\mu(dx)$ is lower semi-continuous with respect to the strong $L^1_V-$ convergence on $\mathcal{B} \cap \{\psi : \|\psi \wedge M\|_{H^1_V} \leq C(M), \text{ for any } M \geq 0\}$.*

*Proof.* The proof that $(t, x) \mapsto J_\psi(t, x)$ is lower semi-continuous is the same as for Lemma 3.1. For $\mu \in L^2_{-V}(\mathbb{R}^d)$, Lemma 3.1 and Theorem 4.2 imply that

$$\int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx) = \sup_{(\eta, \rho) \in \tilde{\mathcal{E}}(\mu)} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi(x)\rho(d\tau, dx) - \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} L(t, x, u)\eta_t(dx, du)dt \right\}.$$

By Lemma 2.2, using the truncation $\psi^M = \psi \wedge M$, this becomes

$$= \sup_{(m^{T,R}, v^{T,R}, \rho^{T,R}) \in \mathcal{E}(\mu), \, M \geq 0} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi^M(x)\rho^{T,R}(d\tau, dx) - \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} L(t, x, v^{T,R})m^{T,R} \, dxdt \right\},$$

since we have that

$$\liminf_{T,R,M \to \infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi^M(x)\rho^{T,R}(d\tau, dx) \geq \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi(x)\rho(d\tau, dx)$$

by lower semi-continuity of $\psi$, the quadratic lower bound (5.7) on $\psi$ and the quadratic moment bound for $\rho$ (see Proposition 4.4). We will show that for each $(m^{T,R}, v^{T,R}, \rho^{T,R}) \in \mathcal{E}(\mu)$ with finite cost and $M \geq 0$,

$$\psi \mapsto \int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \psi^M(x)\rho^{T,R}(d\tau, dx)$$

is lower semi-continuous, which follows from demonstrating that the map (without the truncation) is a continuous linear functional of $H^1_V(\mathbb{R}^d)$. We now take $(m^{T,R}, v^{T,R}, \rho^{T,R}) \in \mathcal{E}(\mu)$. Then, for every $f \in H^1_V(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} f(x)\rho^{T,R}(d\tau, dx)$$

$$= \int_{\mathbb{R}^d} f(x)\mu(dx) + \int_0^T \int_{B(0,R)} \nabla f(x) \cdot \left( v(t, x)m^{T,R}(t, x) - \frac{1}{2}\nabla m^{T,R}(t, x) \right)$$

and it follows from the estimates on $m^{T,R}$ (see Theorem 4.3) that this is a continuous linear functional, completing the proof. □

Now, we are ready to prove attainment in the dual problem (1.5):

**Theorem 5.8.** *We suppose that the Lagrangian $L$ is bounded, $\mu \in L^2_{-V}(\mathbb{R}^d)$, $e^V \nu \in L^\infty(\mathbb{R}^d)$. Then, the dual problem is attained at $\psi^* \in \mathcal{B}$. Furthermore, the set $\mathcal{A}(\mu, \nu)$ is non-empty and the minimizer $(\mathbb{P}, \beta, \alpha)$ of the stochastic primal problem $\mathcal{P}_L(\mu, \nu)$ satisfies*

$$(5.10) \qquad \mathbb{E}^{\alpha \ltimes \mathbb{P}}\left[ \psi^*(X_\tau) - \int_0^\tau L(t, X_t, \beta_t)dt \right] = \int_{\mathbb{R}^d} J_{\psi^*}(0, x)\mu(dx).$$

*Proof.* Let $(\psi_k)_k \subset C_b(\mathbb{R}^d)$ be a maximizing sequence in the dual problem (1.5) with $(\psi_k)_V = 0$, for all $k$. By Proposition 5.1, we can assume that, for each $k$, $\psi_k$ is lower semi-continuous and satisfies (5.2) in the sense of viscosity. Thanks to Proposition 5.5, we also have

$$\|\psi_k\|_{W^{1,1}_V(\mathbb{R}^d)} \leq \overline{C}, \; \psi_k \geq -\overline{C}(1 + |x|^2), \; \text{and} \; \|\psi_k^M\|_{H^1_V(\mathbb{R}^d)} \leq C(M).$$

In particular, this implies that

$$\mathcal{D}_L(\mu, \nu) \leq \sup_{\psi \in \mathcal{B}} \left\{ \int_{\mathbb{R}^d} \psi(y)\nu(dy) - \int_{\mathbb{R}^d} J_\psi(0, x)\mu(dx) \right\} \leq \mathcal{P}_L(\mu, \nu),$$

which means that

$$\mathcal{D}_L(\mu,\nu) = \sup_{\psi\in\mathcal{B}}\Big\{\int_{\mathbb{R}^d}\psi(y)\nu(dy) - \int_{\mathbb{R}^d}J_\psi(0,x)\mu(dx)\Big\}.$$

Moreover, there is a function $\psi^\star \in L_V^1(\mathbb{R}^d)$ such that, up to a subsequence, $\psi_k \to \psi^\star$ in $L_V^1$. As $e^V\nu \in L^\infty(\mathbb{R}^d)$, then we have

$$\int_{\mathbb{R}^d}\psi_k(x)\nu(dx) \to \int_{\mathbb{R}^d}\psi^\star(x)\nu(dx).$$

And, it is easy to see that for each $M \geq 0$, $\psi^* \wedge M$ is in $H_V^1(\mathbb{R}^d)$ and is a weak supersolution of (5.2). We then have that $\psi^* \in \mathcal{B}$ (thanks again to Proposition 5.5). On the other hand, by Lemma 5.7, we have that

$$\liminf_k \int_{\mathbb{R}^d}J_{\psi_k}(0,x)\,\mu(dx) \geq \int_{\mathbb{R}^d}J_{\psi^\star}(0,x)\,\mu(dx).$$

Then, the existence of a maximizer for the dual problem (1.5) follows. Finally, notice that

$$\int_{\mathbb{R}^d}J_\psi(0,x)\mu(dx) \geq \int_{\mathbb{R}^d}[\psi(x)\wedge 0]\,\mu(dx) \geq -\|\psi\wedge 0\|_{H_V^1(\mathbb{R}^d)}\|\mu\|_{H_{-V}^{-1}(\mathbb{R}^d)}.$$

Then, $\mathcal{D}_L(\mu,\nu) < +\infty$. This implies that $\mathcal{P}_L(\mu,\nu)$ is finite by Theorem 3.2 and $\mathcal{A}(\mu,\nu)$ is non-empty. The equation (5.10) follows directly from the duality $\mathcal{D}_L(\mu,\nu) = \mathcal{P}_L(\mu,\nu)$ and the existence of optimizers for both problems (1.1) & (1.5). $\square$

## 6. HITTING TIMES AND STRONG SOLUTIONS

In this section we address the additional structure of monotonicity of $t \mapsto L(t,x,u)$. In this case the set

$$R = \big\{(t,x);\ J_\psi(t,x) = \psi(x)\big\}$$

has the structure of a barrier. In particular, if $t \mapsto L(t,x,u)$ is increasing then for $(t,x) \in R$ and $s > t$ we have $(s,x) \in R$, and if $t \mapsto L(t,x,u)$ is decreasing then for $(t,x) \in R$ and $s < t$ we have $(s,x) \in R$. In either of these cases strict monotonicity implies uniqueness of the optimizer. The monotonicity of $R$ in $t$ follows the same argument in [11].

**Proposition 6.1.** *If $t \mapsto L(t,x,u)$ is increasing then $t \mapsto J_\psi(t,x)$ is nonincreasing, and if $t \mapsto L(t,x,u)$ is decreasing then $t \mapsto J_\psi(t,x)$ is nondecreasing.*

*Proof.* We suppose $t \mapsto L(t,x,u)$ is increasing and select $0 \leq t < s$. We can express the value function at time $s$ by

$$J_\psi(s,x) = \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}_t(\delta_x)}\Big\{\mathbb{E}^{\alpha\ltimes\mathbb{P}}\Big[\psi(X_\tau) - \int_t^\tau L\big(r-t+s,X_r,\beta_r\big)dr\Big]\Big\}$$

$$\leq \sup_{(\mathbb{P},\beta,\alpha)\in\mathcal{A}_t(\delta_x)}\Big\{\mathbb{E}^{\alpha\ltimes\mathbb{P}}\Big[\psi(X_\tau) - \int_t^\tau L\big(r,X_r,\beta_r\big)dr\Big]\Big\} = J_\psi(t,x).$$

The proof in the case that $t \mapsto L(t,x,u)$ is decreasing is the same with the inequality reversed. $\square$

We require a verification type theorem that will allows us to characterize the optimal process and stopping time by the dual optimizer.

**Theorem 6.2.** *Suppose the dual problem $\mathcal{D}_L(\mu, \nu)$ is attained at $(\psi, J_\psi)$ and that $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$ minimizes the primal problem $\mathcal{P}_L(\mu, \nu)$. Then,*

$$(6.1) \qquad J_\psi(\tau, X_\tau) = \psi(X_\tau) \quad \alpha \ltimes \mathbb{P} \text{ almost surely,}$$

*and*

$$(6.2) \qquad M_t := J_\psi(t, X_t) - \int_0^t L(s, X_s, \beta_s) ds$$

*satisfies for any $t > s$,*

$$(6.3) \qquad \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ M_{t \wedge \tau} \Big| \tilde{\mathcal{F}}_s \Big] = M_{s \wedge \tau} \quad \alpha \ltimes \mathbb{P} \text{ almost surely.}$$

*Proof.* We first note that $J_\psi(t, x) \geq \psi(x)$ for all $(t, x)$ and $M_t$ is a supermartingale. Then by the duality of Theorem 3.2 we have

$$(6.4) \qquad 0 = \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ \psi(X_\tau) - J_\psi(0, X_0) - \int_0^\tau L(t, X_t, \beta_t) dt \Big]$$

$$(6.5) \qquad = \mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ \psi(X_\tau) - J_\psi(\tau, X_\tau) + M_\tau - M_0 \Big],$$

and (6.1), (6.3) follow. □

We can now state a theorem that consolidates our results to show a structure to optimizers under these monotonicity conditions on $L$.

**Theorem 6.3.** *We suppose duality and dual attainment (following from either Theorem 5.8, 5.3, or 5.4), and that $u \mapsto L(t, x, u)$ is strictly convex. In addition, assume $t \mapsto L(t, x, u)$ is strictly increasing and $\psi$, $J_\psi$ are continuous, then the optimizer $(\mathbb{P}, \beta, \alpha) \in \mathcal{A}(\mu, \nu)$ is unique and the optimal stopping time is given by*

$$(6.6) \qquad \tau^* = \inf\{t; \ J_\psi(t, X_t) = \psi(X_t)\}.$$

*Alternatively, if $t \mapsto L(t, x, u)$ is strictly decreasing and $\mu$ and $\nu$ have disjoint support, then the optimizer is unique and also give by (6.6).*

*Proof.* We first show that the optimal $(\mathbb{P}, \beta, \alpha)$ for $\mathcal{P}_L(\mu, \nu)$ satisfies:

(i) the process stops in $R$,

$$\mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ \mathbf{1}\{(\tau, X_\tau) \in R\} \Big] = 1,$$

(ii) and

$$\mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ \int_0^\tau \mathbf{1}\{(t, X_t) \in R\} dt \Big] = 0.$$

We then show that there is a unique $(\mathbb{P}, \beta, \alpha)$ that minimizes $\mathcal{J}_L(\mathbb{P}, \beta, \alpha)$ subject to (i) and (ii), and the unique optimal randomized stopping time is given by $\tau^*$.

By Theorem 6.2, we immediately have (i) from (6.1). We let $(\mathbb{P}^{t,x}, \beta^{t,x}, \alpha^{t,x}) \in \mathcal{A}_t(\delta_x)$ be the conditional expectation, i.e. satisfying for each $t$ and $G \in C_b(\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d)$,

$$\mathbb{E}^{\alpha \ltimes \mathbb{P}}\Big[ \int_t^\tau G(s, X_s, \beta_s) \Big] = \int_{\mathbb{R}^d} \mathbb{E}^{\alpha^{t,x} \ltimes \mathbb{P}^{t,x}}\Big[ \int_t^\tau G(s, X_s, \beta_s^{t,x}) \Big] \eta_t(dx, \mathbb{R}^d).$$

Then,

$$0 = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \Big[ M_\tau - M_{t \wedge \tau} \Big]$$
$$= \int_{\mathbb{R}^d} \mathbb{E}^{\alpha^{t,x} \ltimes \mathbb{P}^{t,x}} \Big[ \psi(X_\tau) - J_\psi(t,x) + \int_t^\tau L\big(s, X_s, \beta_s^{t;x}\big) ds \Big] \eta_t(dx, \mathbb{R}^d),$$

and it follows that

$$J_\psi(t,x) = \mathbb{E}^{\alpha^{t,x} \ltimes \mathbb{P}^{t,x}} \Big[ \psi(X_\tau) + \int_t^\tau L\big(s, X_s, \beta_s^{t,x}\big) ds \Big]$$

for all $t$ and $\eta_t$-a.e. $x$. If $J_\psi(t,x) = \psi(x)$ and either $t \mapsto L(t,x,u)$ is strictly increasing and $s > t + \epsilon$ for $\epsilon > 0$ or $t \mapsto L(t,x,u)$ is strictly decreasing and $0 \leq s < t + \epsilon$ for $\epsilon < 0$, then $T_\#^{t+\epsilon-s,0} \alpha^{s,x} \ltimes \mathbb{P}^{s,x}$ will satisfy

$$J_\psi(t+\epsilon, x) \geq \mathbb{E}^{\alpha^{s,x} \ltimes \mathbb{P}^{s,x}} \Big[ \psi(X_\tau) + \int_s^\tau L\big(r - s + t + \epsilon, X_r, \beta_{r-s+t+\epsilon}^{s,x}\big) dr \Big],$$

which is only satisfied if $\tau = s$ with probability one, and (ii) follows. We now consider $(\mathbb{P}', \beta', \alpha') \in \mathcal{A}(\mu)$ that satisfies (i) and (ii), and we have

$$\mathbb{E}^{\alpha' \ltimes \mathbb{P}'} \Big[ \int_0^\tau L(t, X_t, \beta_t') dt \Big] \geq \mathbb{E}^{\alpha' \ltimes \mathbb{P}'} \Big[ \psi(X_\tau) - J_\psi(0, X_0) \Big] = \mathbb{E}^{\alpha \ltimes \mathbb{P}} \Big[ \int_0^\tau L(t, X_t, \beta_t) dt \Big],$$

by the definition of $J_\psi$, optimality of $(\mathbb{P}, \beta, \alpha)$ and Theorem 3.2, which implies that $(\mathbb{P}', \beta', \alpha')$ optimizes the cost over policies in $\mathcal{A}(\mu)$ that satisfy (i) and (ii).

Given $\mathbb{P}$ and $\beta$, let $\alpha^*$ be the randomized stopping time corresponding to $\tau^*$ defined in (6.6). We then have that (i) and (ii) are satisfied by $(\mathbb{P}, \beta, \alpha^*)$ using continuity of $\psi$ and $J_\psi$. Furthermore the cost of $(\mathbb{P}, \beta, \alpha^*)$ is less than or equal to the cost of $(\mathbb{P}, \beta, \alpha)$ because $L$ is nonnegative. Furthermore, since the cost is equal, this implies that $\alpha^* = \alpha$. Finally, uniqueness of $\mathbb{P}$ and $\beta$ follow from strict convexity of $L$. □

**Remark 6.4.** *The assumption that $\psi$ and $J_\psi$ are continuous in Theorem 6.3 is satisfied if $1 < p < 2$ as in Theorem 5.3, and is probably not needed if the drift is bounded ($p = +\infty$) as in Theorem 5.8 if one pursues further the Sobolev regularity of the stopping distribution $\rho^*$ for $\alpha^*$ as was done in [13, 12].*

*From the control theory point of view, it is a natural question whether the optimal control policy $\beta$ satisfies the Pontryagin maximum principle: namely, $\beta_t = -D_z H(t, X_t, \nabla J_\psi(t, X_t))$ and solves the SDE: $dX_t = \beta(X_t)dt + dW_t$ in a strong sense. This seems to require $J_\psi$ to be $C^{1,1}$. We leave it as an open question.*

## REFERENCES

[1] Erhan Bayraktar and Christopher W Miller. Distribution-constrained optimal stopping. *Mathematical Finance*, 29(1):368–406, 2019.

[2] Mathias Beiglböck, Manu Eder, Christiane Elgert, and Uwe Schmock. Geometry of distribution-constrained optimal stopping problems. *Probability theory and related fields*, 172(1-2):71–101, 2018.

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

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

[5] 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.

[6] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

[7] Bartłomiej Dyda and Moritz Kassmann. On weighted poincaré inequalities. *Ann. Acad. Sci. Fenn. Math. in print, see also http://arxiv. org/abs/1209.3125*, 2013.

[8] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*, volume 28. Siam, 1999.

[9] A. R. Elcrat and H. A. MacLean. Weighted wirtinger and poincar inequalities on unbounded domains. *Indiana University Mathematics Journal*, 29:321–332, 1980.

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

[11] Nassif Ghoussoub, Young-Heon Kim, and Aaron Zeff Palmer. Optimal transport with controlled dynamics and free end times. *SIAM Journal on Control and Optimization*, 56(5):3239–3259, 2019.

[12] Nassif Ghoussoub, Young-Heon Kim, and Aaron Zeff Palmer. A solution to the Monge transport problem for Brownian martingales. *Submitted*, pages 1–23, 2019.

[13] Nassif Ghoussoub, Young-Heon Kim, and Aaron Zeff Palmer. PDE methods for optimal Skorokhod embeddings. *Calc. of Variations and PDEs*, pages 1–33, 2019.

[14] UG Haussmann. Existence of optimal Markovian controls for degenerate diffusions. In *Stochastic differential systems*, pages 171–186. Springer, 1986.

[15] Hitoshi Ishii. On the equivalence of two notions of weak solutions, viscosity solutions and distribution solutions. *Funkcial. Ekvac*, 38(1):101–120, 1995.

[16] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2013.

[17] Sigrid Källblad. A dynamic programming principle for distribution-constrained optimal stopping. *arXiv preprint arXiv:1703.08534*, 2017.

[18] Leonid Vitalievich Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[19] Nikolaj Vladimirovič Krylov. *Controlled diffusion processes*. Springer Verlag, 1980.

[20] Christian Léonard. From the Schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[21] Pierre-Louis Lions. Optimal control of diffusion processes and hamilton–jacobi–bellman equations part 2: viscosity solutions and uniqueness. *Communications in partial differential equations*, 8(11):1229–1276, 1983.

[22] Paul André Meyer and WA Zheng. Tightness criteria for laws of semimartingales. In *Annales de l'IHP Probabilités et statistiques*, volume 20, pages 353–372, 1984.

[23] Toshio Mikami. Optimal control for absolutely continuous stochastic processes and the mass transportation problem. *Electronic Communications in Probability*, 7:199–213, 2002.

[24] Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability theory and related fields*, 129(2):245–260, 2004.

[25] Toshio Mikami and Michele Thieullen. Optimal transportation problem by stochastic optimal control. *SIAM Journal on Control and Optimization*, 47(3):1127–1139, 2008.

[26] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[27] Luis Sylvestre. Viscosity solutions of elliptic equations. *Lecture notes in Second Chicago Summer School In Analysis, http://math.uchicago.edu/ luis/preprints/viscosity-solutions.pdf*, 2015.

[28] Xiaolu Tan, Nizar Touzi, et al. Optimal transportation under controlled stochastic dynamics. *The annals of probability*, 41(5):3201–3240, 2013.

[29] E Wong. Representation of martingales, quadratic variation and applications. *SIAM J. Control*, 9:621–633, 1971.

[30] WA Zheng. Tightness results for laws of diffusion processes application to stochastic mechanics. In *Annales de l'IHP Probabilités et statistiques*, volume 21, pages 103–124, 1985.

* Samer Dweik, Nassif Ghoussoub, Young-Heon Kim and Aaron Zeff Palmer

Department of Mathematics, University of British Columbia, Vancouver, V6T 1Z2 Canada
*E-mail address*: dweik@math.ubc.ca, nassif@math.ubc.ca, yhkim@math.ubc.ca, azp@math.ubc.ca