

A SOLUTION TO THE MONGE TRANSPORT PROBLEM FOR BROWNIAN MARTINGALES

BY NASSIF GHOUSOUB^{*}, YOUNG-HEON KIM[†] AND AARON ZEFF PALMER[‡]

Department of Mathematics, University of British Columbia, ^{*}nassif@math.ubc.ca; [†]yhkim@math.ubc.ca; [‡]azp@math.ubc.ca

We provide a solution to the problem of optimal transport by Brownian martingales in general dimensions whenever the transport cost satisfies certain subharmonic properties in the target variable as well as a stochastic version of the standard “twist condition” frequently used in deterministic Monge transport theory. This setting includes, in particular, the case of the distance cost $c(x, y) = |x - y|$. We prove existence and uniqueness of the solution and characterize it as the first time Brownian motion hits a barrier that is determined by solutions to a corresponding dual problem.

1. Introduction. Given a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, the Monge optimal transport problem seeks a map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ that pushes forward a given probability measure μ to ν (written as $T_{\#}\mu = \nu$) while minimizing the total transport cost,

$$(1) \quad \mathcal{T}_c(\mu, \nu) = \inf_{T_{\#}\mu = \nu} \int_{\mathbb{R}^d} c(x, T(x))\mu(dx).$$

Kantorovich [35] proposed the following linear relaxation of the Monge problem, which was eventually shown in [42] to have the same infimum provided c is continuous and μ is nonatomic:

$$(2) \quad \mathcal{T}_c(\mu, \nu) = \inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\pi(dx, dy),$$

where $\Gamma(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$, whose marginals are μ and ν . In turn, the last expression can also be written in terms of an optimal statistical correlation problem between probability measures given by

$$(3) \quad \mathcal{T}_c(\mu, \nu) = \inf\{\mathbb{E}[c(X, Y)]; X \sim \mu, Y \sim \nu\},$$

where $X \sim \mu$ means that the random variable X has μ as its distribution.

Monge [40] originally formulated problem (1) with the distance cost $c(x, y) = |x - y|$, and a solution was provided more than 200 years later [1, 13, 19, 48]. Brenier [11], on the other hand, dealt with the important case of the squared distance $c(x, y) = |x - y|^2$, which was extended by Caffarelli [12], Gangbo and McCann [23], Gangbo [22], Levin [37] and Ma-Trudinger-Wang [39] to more general cost functions c that satisfy the so-called *twist condition*, namely,

$$(4) \quad \text{the map } y \mapsto \nabla_x c(x, y) \text{ is one-to-one for all } x.$$

This condition was shown to be sufficient to ensure that the minimizer of the relaxed problem (2) is indeed supported on the graph of a single-valued map T , hence solving the original problem (1). In the statistical terms of (3), the result means that the optimal random variables are perfectly dependent, that is, $Y = T \circ X$ a.e.

Received July 2019; revised April 2020.

MSC2020 subject classifications. Primary 49, 60; secondary 52.

Key words and phrases. Optimal transport, Skorokhod embedding, variational inequality.

We also note that optimal transport problems associated to cost functions of the form $g(|x - y|)$, where g is convex can be represented by ones given by the generating function of an associated Lagrangian L , that is,

$$(5) \quad c(x, y) = \inf \left\{ \int_0^1 L(t, \gamma(t), \dot{\gamma}(t)) dt; \gamma(0) = x, \gamma(1) = y \right\}.$$

In this case, optimizing maps are closely related to the corresponding Hamiltonian dynamics (e.g., Benamou–Brenier [9], Bernard–Buffoni [10]).

The martingale transport problem was motivated by questions of option pricing in mathematical finance; see, for example, [16, 30, 32]. It consists of a constrained version of (3), when the pair of random variables (X, Y) forms a one-step martingale, that is,

$$(6) \quad \mathcal{M}_c(\mu, \nu) = \inf \{ \mathbb{E}[c(X, Y)]; X \sim \mu, Y \sim \nu \text{ and } \mathbb{E}[Y|X] = X \}.$$

The martingale property on (X, Y) entails that $(w(X), w(Y))$ is a real-valued submartingale for any convex function w on \mathbb{R}^d . This implies a necessary condition on the probability measures μ and ν for such a martingale pairing to exist, namely, that they are in convex order, that

$$\mu \prec_C \nu \quad \text{which means } \int w(x)\mu(dx) \leq \int w(y)\nu(dy) \text{ for all convex functions } w.$$

A remarkable theorem of Strassen [46] states that the convex order between μ and ν is also sufficient for the existence of a one-step martingale starting at μ and ending at ν , hence providing a martingale transport for problem (6). The martingale transport problem is by now well understood in dimension one, including the formulation of corresponding dual principles, cases where the latter are attained, as well as questions of uniqueness and structure of the optimal martingales in the primal problem (see, e.g., [5, 6, 8, 16, 30, 32]). The higher dimensional case, however, is less understood, as could be seen in [25] where a solution for dimension $d = 2$ is given.

In this paper we shall address the *Brownian martingale transport problem*, which involves a more particular class of martingales and can be formulated as

$$(7) \quad \mathcal{P}_c(\mu, \nu) = \inf_{\tau} \{ \mathbb{E}[c(B_0, B_{\tau})]; B_0 \sim \mu \ \& \ B_{\tau} \sim \nu \},$$

where $(B_t)_t$ is Brownian motion starting with distribution μ and ending at a stopping time τ such that B_{τ} realizes the distribution ν . Note that here again, this imposes (in dimension $d \geq 2$) an even more stringent condition on the pair (μ, ν) , namely, that they are in *subharmonic order*, that is,

$$\mu \prec_{SH} \nu \quad \text{which means } \int w(x)\mu(dx) \leq \int w(y)\nu(dy) \text{ for all subharmonic functions } w.$$

This problem was originally proposed by Skorokhod in one dimension [45] with $\mu = \delta_0$, where he proved the existence of a randomized stopping time τ such that $B_{\tau} \sim \nu$, where $(B_t)_t$ is a Brownian motion starting at *zero*. The existence of a nonrandomized stopping time in one dimension was established in [17] and [43]. An alternative construction of a randomized stopping was given by [44] for more general processes. The existence of a nonrandomized stopping time in higher dimensions was established by Baxter and Chacon [2] and extended in [20]. A thorough review of solutions to the Skorokhod problem and its connections with optimal transport has been given in [41].

Another formulation of this problem, similar to Kantorovich’s relaxation but which restricts the *transport plans* to follow Brownian paths, is the following:

$$(8) \quad \mathcal{P}_c(\mu, \nu) = \inf_{\pi \in \mathcal{BM}(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y)\pi(dx, dy),$$

where each $\pi \in \mathcal{BM}(\mu, \nu)$ is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν , satisfying $\delta_x \prec_{\text{SH}} \pi_x$ for μ -a.e. x , where π_x is the disintegration of $\pi(dx, dy) = \pi_x(dy)\mu(dx)$. Again, such transport plans π can be seen as joint distributions of $(B_0, B_\tau) \sim \pi$, where $B_0 \sim \mu$, $B_\tau \sim \nu$ and τ is a possibly randomized stopping time for the Brownian filtration; see, for example, [24]. Under the assumption that μ and ν have compact support, the set $\mathcal{BM}(\mu, \nu)$ is a weak*-compact set of measures in the Banach space dual to the bounded continuous functions, and problem (8) has an optimal solution $\pi^* \in \mathcal{BM}(\mu, \nu)$. However, similar to the issue triggered by Kantorovich’s relaxation of (1), such a solution may very well correspond to a randomized stopping time. A main objective of this paper is to figure out when there exists a true and unique optimal stopping time that solves the problem.

This problem is equivalent to the martingale problem in the one dimensional case, a context addressed by Henry–Labordere–Touzi [31] and others. The duality theory and applications to mathematical finance were developed in [34] and [5] which also establish the existence and uniqueness of an optimal nonrandomized optimal stopping time under certain conditions. Motivated by the methods of standard mass transports, in [3] Beiglböck, Cox, and Huesmann further developed the duality theory and a monotonicity principle for problems with stochastic cost processes of the form

$$\inf_{\tau} \{ \mathbb{E}[G_\tau]; B_0 \sim \mu, B_\tau \sim \nu \}.$$

In particular, they used this theory to solve the problem in the case $G_\tau = g(\tau)$ where g is strictly convex (resp., concave), by showing that the unique optimal stopping time is given by the classical Root (resp. Rost) embedding which is the hitting time of a space-time barrier. The problem of the attainment of the dual problem remained a challenge. In [8] the authors prove that a relaxed dual problem is attained when $d = 1$ or costs of the form (7), and then the attainment of the original dual problem was established for $d = 1$ in [6]. By using PDE methods, dual attainment was extended for $d > 1$ in [27] to the case where the cost is an integral along the path, that is, for problems of the form

$$(9) \quad \inf_{\tau} \left\{ \mathbb{E} \left[\int_0^\tau L(t, B_t) dt \right]; B_0 \sim \mu, B_\tau \sim \nu \right\}.$$

For these costs it is shown in [27] that the unique optimal stopping time is the hitting time of a space-time barrier when L is strictly increasing (resp., decreasing).

Our goal in this paper is to solve problem (7) in general dimensions for costs that only depend on the initial and final states and not on the path traveled between them, and we make special note of the case where the cost is given by the Euclidean distance $c(x, y) = |x - y|$. It is important to note that, unlike standard mass transports, problem (7) cannot be reduced to (9) nor to

$$(10) \quad \inf_{\tau, A} \left\{ \mathbb{E} \left[\int_0^\tau L(t, X_t, A_t) dt \right]; X_0 \sim \mu, X_\tau \sim \nu \right\},$$

where $X_t \in \mathbb{R}^n$ solves the SDE: $dX_t = f(X_t, A_t) dt + \sigma dW_t$. We mention that because, at least in (9), the time monotonicity of the Lagrangian cost enables one to construct an appropriate barrier set in space-time, as was done in [3] and [27], hence establishing both the uniqueness and the hitting-time characterization of the optimal stopping. A similar result was established in [24] in the case where the measures μ, ν and the cost $c(x, y)$ are radially symmetric.

We also note that another advantage of the Lagrangian formulation of the cost is the associated Eulerian (mass flow) formulation to the optimal transport problem [9, 10] which was also extended to the free-end case in both the deterministic [26] and stochastic cases [27] as long as the cost is of Lagrangian type. However, this seminal approach cannot be used for the costs we are considering in this paper, where we prove—among other things—the following result.

THEOREM (See Theorem 7.2). *Let μ and ν be two probability measures that are supported on an open bounded convex set $O \subset \mathbb{R}^d$. Assuming $\mu \prec_{\text{SH}} \nu$ and that they have continuous densities with respect to Lebesgue measure, then there is a unique optimal stopping time τ^* that minimizes*

$$(11) \quad \inf_{\tau} \{ \mathbb{E} |B_0 - B_{\tau}|; B_0 \sim \mu \ \& \ B_{\tau} \sim \nu \}.$$

Moreover, τ^* is only randomized at 0, while otherwise it is given by the hitting time

$$(12) \quad \eta = \inf \{ t > 0; (B_0, B_t) \in R \}$$

for some measurable $R \subset \overline{O} \times \overline{O}$.

We shall first establish the following weak duality formula for the primal problem (7) under the mere assumption that $c \in C(\mathbb{R}^d \times \mathbb{R}^d)$:

$$(13) \quad \mathcal{P}_c(\mu, \nu) = D_c(\mu, \nu) := \sup_{\psi \in \text{LSC}(\overline{O})} \left\{ \int_O \psi(y) \nu(dy) - \int_O J_{\psi}(x, x) \mu(dx) \right\},$$

where $\text{LSC}(\overline{O})$ is the cone of all bounded lower semicontinuous functions on \overline{O} . Throughout the paper we will denote B_t^x as the Brownian motion beginning with $B_0^x = x$, and we will use the restriction of stopping times to this set of paths without changing notation so that, for continuous functions $h \in C(\mathbb{R}^d \times \mathbb{R}^d)$ and stopping times τ ,

$$\mathbb{E}[h(B_0, B_{\tau})] = \int_{\mathbb{R}^d} \mathbb{E}[h(x, B_{\tau}^x)] \mu(dx).$$

The so-called value function J_{ψ} is defined as

$$(14) \quad J_{\psi}(x, y) = \sup_{\tau \leq \tau_O} \mathbb{E}[\psi(B_{\tau}^y) - c(x, B_{\tau}^y)],$$

where τ_O is the first exit time of the set O . We point out two other useful characterizations of the value function J_{ψ} . Under some regularity assumptions on ψ and c and for each fixed $x \in \overline{O}$, the function $y \mapsto J_{\psi}(x, y)$ is the unique viscosity solution to the obstacle problem for $u \in C(\overline{O})$:

$$\begin{cases} u(y) \geq \psi(y) - c(x, y) & \text{for } y \in O, \\ u(y) = \psi(y) - c(x, y) & \text{for } y \in \partial O, \\ \Delta u(y) \leq 0 & \text{for } y \in O, \\ \Delta u(y) = 0 & \text{whenever } u(y) > \psi(y) - c(x, y) \end{cases}$$

as well as the unique minimizer of the variational problem

$$\min_{u \in H^1(O)} \left\{ \int_O |\nabla u(y)|^2 dy; u \geq \psi - c(x, \cdot), u - \psi + c(x, \cdot) \in H_0^1(O) \right\}.$$

We then assume the following condition on the cost:

$$(15) \quad y \mapsto c(x, y) \text{ is subharmonic and } D\text{-superharmonic, that is, } 0 \leq \Delta_y c(x, y) \leq D,$$

which will yield that the dual problem $D_c(\mu, \nu)$ can actually be restricted to a weakly compact set of functions $\mathcal{B}_D \subset H_0^1(O)$ that consists of bounded and nonpositive D -superharmonic functions that vanish on ∂O . These functions are lower semicontinuous and satisfy a remarkable property of continuity along Brownian paths (see Lemma 3.5). These properties of the maximizing set \mathcal{B}_D then guarantee that the dual problem (13) is attained by a sufficiently regular function ψ^* (see Theorem 4.1).

Condition (15) on the cost can be weakened in several cases. Indeed, while the distance cost, $c(x, y) = |x - y|$ does not satisfy D -superharmonicity at the singular points $x = y$, we can avoid this issue under the assumption that the supports of μ and ν are strictly disjoint. We then use a localization argument to handle the general case (without the disjointness assumption) for this important cost. Moreover, the subharmonicity of $y \rightarrow c(x, y)$ can also be replaced by the condition that $M = \inf_{x, y \in O} \Delta_y c(x, y) > -\infty$, since in this case one can replace c with the subharmonic cost $\tilde{c}(x, y) = c(x, y) - h(y)$ where h is any function in $C^\infty(O)$ such that $\Delta h(y) \leq M$. On the other hand, a one-dimensional counterexample to the dual attainment problem in the case where (15) is not satisfied is given in [4]; see [5] and [25] for related results.

For the issue of uniqueness, we isolate a second condition on the cost which can be seen as a stochastic counterpart of the twist condition (4). It can be stated as follows: For each pair of states (x, y) and any stopping time τ for Brownian motion starting at y ,

$$(16) \quad \mathbb{E}[\nabla_x c(x, B_\tau^y)] = \nabla_x c(x, y) \implies \tau = 0.$$

We therefore call it the *stochastic twist condition*. We note that a related condition on cost functions was introduced in the one-dimensional case in [33] as *increasing (resp., decreasing) variation swap kernel* and in [31] as *the martingale counterpart of the Spence–Mirrlees condition*, and it reads as

$$(17) \quad c_{yyx}(x, y) > (<)0.$$

The stochastic twist condition enables us to prove that the optimal stopping time τ^* is unique and is characterized as the first hitting time of a barrier set determined by the dual optimizer ψ^* ; see Theorem 6.1. The delicate proof also uses the differentiability of the value function, which we obtain from a Lipschitz assumption on $x \mapsto c(x, y)$, as well as the dynamic programming principle defining J_ψ . In particular, the distance cost $|x - y|$ satisfies the stochastic twist condition when $x \neq y$; a property that was used in [25] for the general martingale optimal transport problem and was realized to apply to the case of Brownian martingales by Tongseok Lim. We extend our results to this particular and important cost function in Theorems 7.1 and 7.2.

REMARK 1.1. The results in this paper, including Theorems 7.1 and 7.2 and the related results on the dual attainment (Theorem 4.1), hold in more general settings such as for Brownian motion valued in a geodesically convex bounded domain O of a complete nonpositively curved Riemannian manifold, when the cost c is given by the Riemannian distance $d(x, y)$. Here, the Laplace operator is replaced with the Laplace–Beltrami operator [28]. Note that the stochastic twist condition then holds for any differentiable Riemannian distance, while non-positive curvature implies subharmonicity (in fact, convexity) of $y \mapsto d(x, y)$. The proofs are adaptable from the Euclidean case with additional technical complications, which we do not address further.

The paper is organized as follows: In Section 2 we prove the weak duality principle (13) and reduce the problem to optimizing over the end potential by using a dynamic programming principle. This gives a novel point of view for problem (7). In Section 3 we introduce a remarkable set of D -superharmonic functions \mathcal{B}_D and represent it as a weakly compact convex subset of $H_0^1(O)$. In Section 4 we prove our first main result, namely, the attainment in the dual problem (Theorem 4.1), by showing that one can restrict the maximization of (13) to the set \mathcal{B}_D . Some of the key lemmas there use condition (15) in a crucial way. This then yields a barrier set for the verification theorem (Theorem 4.7). Section 5 discusses the key Stochastic Twist condition (16) and includes a few important examples. Section 6 contains the proof for

uniqueness and the characterization of the optimal solution to (7) as a first hitting time (Theorem 6.1). The case of the distance cost is finally addressed in Section 7 (Theorems 7.1, 7.2). The Appendix contains a couple of technical results used in Sections 2 and 3. Additionally, we shall use our result on dual attainment to provide there a quick proof of a version of the monotonicity principle of Beiglböck, Cox, and Huesmann [3] that is adapted to our setting.

2. A dual variational principle. Duality is a key aspect of many optimization problems. In our case we shall see that the dual problem to $\mathcal{P}_c(\mu, \nu)$ arises directly as the linear maximization problem

$$(18) \quad \mathcal{D}'_c(\mu, \nu) = \sup_{(\psi, J) \in \mathcal{A}_c} \left[\int_{\mathbb{R}^d} \psi(y) \nu(dy) - \int_{\mathbb{R}^d} J(x, x) \mu(dx) \right]$$

with linear constraints

$$(19) \quad \begin{aligned} \mathcal{A}_c &= \{(\psi, J) \in C(\mathbb{R}^d) \times C(\mathbb{R}^d \times \mathbb{R}^d); \\ &\quad J(x, y) \geq \psi(y) - c(x, y) \text{ and } \Delta_y J(x, y) \leq 0\}, \end{aligned}$$

where we understand the inequality $\Delta_y J(x, y) \leq 0$ in the sense of viscosity. The equivalence a probabilistic notions of superharmonic functions will be recalled in Lemma 2.3 and later with a weak and variational notions in Proposition 3.2 and Lemma 3.3.

Define the operator $V_{c,x}^\psi$ as

$$V_{c,x}^\psi [J](y) := \min\{J(x, y) - \psi(y) + c(x, y), -\Delta_y J(x, y)\};$$

we can then understand \mathcal{A}_c as the set of pairs (ψ, J) such that for each $x \in \mathbb{R}^d$, $y \mapsto J(x, y)$ is a supersolution of $V_{c,x}^\psi [J(x, \cdot)](y) \geq 0$ in the sense of viscosity.

We assume that the measures μ and ν have support on a bounded open convex domain O . In this case we can determine J in terms of ψ as the minimal viscosity supersolution. We call this the value function, and it is given by

$$(20) \quad J_\psi(x, y) := \sup_{\tau \leq \tau_O} \mathbb{E}[\psi(B_\tau^y) - c(x, B_\tau^y)],$$

where B_t^y is the Brownian motion with $B_0^y = y$, τ_O is the exit time from O and the supremum is over all stopping times prior to τ_O . In the following we denote by $\text{LSC}(\overline{O})$ the set of bounded, lower semicontinuous functions on \overline{O} . Recall the definition of Brownian martingale plans between μ and ν as the finite measures π on $O \times O$ with marginals μ and ν that disintegrate as $\pi(dx, dy) = \pi_x(dy)\mu(dx)$ in such a way that $\delta_x \prec_{\text{SH}} \pi_x$ for μ -a.e. $x \in O$. We set $\mathcal{BM}(\mu, \nu)$ to be the set of such transport plans, that is,

$$(21) \quad \mathcal{BM}(\mu, \nu) = \{\pi \geq 0, \pi(\cdot, O) = \mu, \pi(O, \cdot) = \nu, \delta_x \prec_{\text{SH}} \pi_x \text{ for } \mu\text{-a.e. } x \in O\}.$$

THEOREM 2.1. *Suppose that $c \in C(\mathbb{R}^d \times \mathbb{R}^d)$ and μ and ν are probability measures with support in an open, bounded and convex subset O of \mathbb{R}^d . Then,*

$$(22) \quad \begin{aligned} \mathcal{P}_c(\mu, \nu) &= \mathcal{D}'_c(\mu, \nu) \\ &= \mathcal{D}_c(\mu, \nu) := \sup_{\psi \in \text{LSC}(\overline{O})} \left\{ \int_O \psi(y) \nu(dy) - \int_O J_\psi(x, x) \mu(dx) \right\}, \end{aligned}$$

and there is $\pi^* \in \mathcal{BM}(\mu, \nu)$ such that

$$\mathcal{P}_c(\mu, \nu) = \int_O \int_O c(x, y) \pi^*(dx, dy).$$

Before giving the proof of Theorem 2.1, we start by noting the relationships between the various formulations of our problem. Recall first the following correspondence between (possibly randomized) stopping times and subharmonic martingale measures. While randomized stopping times have many equivalent representations, it suffices to consider their law as probability measures on $C(\mathbb{R}^+, \mathbb{R}^d) \times \mathbb{R}^+$ that disintegrate with respect to Wiener measure as a measure η on \mathbb{R}^+ such that $A_t = \int_0^t \eta(ds)$ is adapted to the filtration of the Brownian motion. The only topology we consider is the weak* convergence of the probability law on $C(\mathbb{R}^+, \mathbb{R}^d) \times \mathbb{R}^+$.

LEMMA 2.2 (See, e.g., Theorem 2.1 in [24]). *Let σ and ρ be probability measures on \overline{O} , and let $(B_t)_t$ denote Brownian motion starting according to $B_0 \sim \sigma$. Then:*

1. *For each (possibly randomized) stopping time $\tau \leq \tau_O$ satisfying $B_\tau \sim \rho$, there is $\pi \in \mathcal{BM}(\sigma, \rho)$ such that $(B_0, B_\tau) \sim \pi$.*
2. *Conversely, for each $\pi \in \mathcal{BM}(\sigma, \rho)$, there exists a (randomized) stopping time $\tau \leq \tau_O$, such that $(B_0, B_\tau) \sim \pi$.*

We shall need the following characterization for superharmonic lower semicontinuous functions. We include a proof for completeness.

LEMMA 2.3. *For $\phi \in \text{LSC}(\overline{O})$ the following are equivalent:*

1. $\Delta\phi(x) \leq 0$ in the sense of viscosity for all $x \in O$;
2. For any stopping time $\tau \leq \tau_O$ and all $y \in \overline{O}$, we have

$$(23) \quad \phi(y) \geq \mathbb{E}[\phi(B_\tau^y)].$$

PROOF. First, we suppose that $\phi \in \text{LSC}(\overline{O})$ satisfies $\Delta\phi(x) \leq 0$ in the sense of viscosity for all $x \in O$. There is then a sequence of smooth functions $\phi^i \in C^\infty(\overline{O})$ with $\phi^i \leq \phi$ in O , constants δ^i such that $\lim_{i \rightarrow \infty} \phi^i(x) = \phi(x)$ for all $x \in \overline{O}$, $\lim_{i \rightarrow \infty} \delta^i = 0$ and $\Delta\phi^i(x) \leq 0$ for $x \in O^{\delta^i}$, where O^{δ^i} is the set of points in O with distance from ∂O greater than δ^i (see Lemma A.1 or similar results in [14, 21, 47].)

For $y \in \partial O$, $B_\tau^y = B_{\tau_0}^y = y$ and (23) holds. For $y \in O$ and any stopping time $\tau \leq \tau_O$, we fix $\delta > 0$ such that $y \in O^\delta$ and set $\tau^\delta = \tau \wedge \tau_{O^\delta}$. Then, by Itô's Lemma

$$\begin{aligned} \phi(y) &= \lim_{i \rightarrow \infty} \phi^i(y) \\ &= \lim_{i \rightarrow \infty} \mathbb{E} \left[\phi^i(B_{\tau^\delta}^y) - \int_0^{\tau^\delta} \Delta\phi^i(B_t^y) dt \right] \\ &\geq \lim_{i \rightarrow \infty} \mathbb{E}[\phi^i(B_{\tau^\delta}^y)] = \mathbb{E}[\phi(B_{\tau^\delta}^y)]. \end{aligned}$$

Then, taking $\delta \rightarrow 0$, we have

$$\liminf_{\delta \rightarrow 0} \mathbb{E}[\phi(B_{\tau^\delta}^y)] \geq \mathbb{E} \left[\liminf_{\delta \rightarrow 0} \phi(B_{\tau^\delta}^y) \right] \geq \mathbb{E}[\phi(B_\tau^y)],$$

by lower semicontinuity of ϕ , Fatou's lemma and the continuity of Brownian paths.

For the other direction (2) \Rightarrow (1), we consider $w \in C^2(\overline{O})$ that touches ϕ from below at $y \in O$, that is, $w(y) = \phi(y)$ and $w(x) \leq \phi(x)$ for all $x \in \overline{O}$. Then, for all stopping times $\tau \leq \tau_O$,

$$\mathbb{E}[w(B_\tau^y)] - w(y) \leq \mathbb{E}[\phi(B_\tau^y)] - \phi(y) \leq 0$$

which implies that $\Delta w(y) \leq 0$ for $w \in C^2$. This completes the proof. \square

PROOF OF THEOREM 2.1. We sketch the proof of $\mathcal{P}_c(\mu, \nu) = \mathcal{D}'_c(\mu, \nu)$ as very similar results have appeared in [3] and [27]. We assume ψ and J are continuous unless stated otherwise.

The constraint that $\pi \in \mathcal{BM}(\mu, \nu)$, defined in (21), is equivalent to:

- The measure π on $O \times O$ is a finite and nonnegative;
- The second marginal of π is ν , or

$$0 = \sup_{\psi} \left\{ \int_{\mathbb{R}^d} \psi(y) \nu(dy) - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(y) \pi(dx, dy) \right\},$$

- For each x , π_x is in subharmonic order with δ_x , and the first marginal is μ , that is,

$$0 = \sup_{J; \Delta_y J \leq 0} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} J(x, y) \pi(dx, dy) - \int_{\mathbb{R}^d} J(x, x) \mu(dx) \right\}.$$

Thus, we have

$$\begin{aligned} & \inf_{\pi \in \mathcal{BM}(\mu, \nu)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c(x, y) \pi(dx, dy) \\ &= \inf_{\pi \geq 0} \sup_{\psi} \sup_{J; \Delta_y J \leq 0} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} c(x, y) \pi(dx, dy) + \int_{\mathbb{R}^d} \psi(y) \nu(dy) \right. \\ & \quad \left. - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(y) \pi(dx, dy) + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} J(x, y) \pi(dx, dy) - \int_{\mathbb{R}^d} J(x, x) \mu(dx) \right\}, \end{aligned}$$

and the Fenchel–Rockafellar duality theorem allows us to interchange the infimum and supremum. This expression becomes

$$\begin{aligned} \mathcal{P}_c(\mu, \nu) &= \sup_{\psi} \sup_{J; \Delta_y J \leq 0} \inf_{\pi \geq 0} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (c(x, y) - \psi(y) + J(x, y)) \pi(dx, dy) \right. \\ & \quad \left. + \int_{\mathbb{R}^d} \psi(y) \nu(dy) - \int_{\mathbb{R}^d} J(x, x) \mu(dx) \right\} \\ &= \mathcal{D}'_c(\mu, \nu). \end{aligned}$$

The attainment of a minimizer π^* for $\mathcal{P}_c(\mu, \nu)$ is immediate from the compactness of probability measures in the weak* topology and the definition of $\mathcal{BM}(\mu, \nu)$ which makes it a weak* closed convex set in the space of Radon measures on $\overline{O} \times \overline{O}$.

To prove that

$$\mathcal{D}'_c(\mu, \nu) = \mathcal{D}_c(\mu, \nu) := \sup_{\psi \in \text{LSC}(\overline{O})} \left\{ \int_O \psi(y) \nu(dy) - \int_O J_{\psi}(x, x) \mu(dx) \right\},$$

we first show the inequality $\mathcal{D}'_c(\mu, \nu) \geq \mathcal{D}_c(\mu, \nu)$. For that, let $\pi^* \in \mathcal{BM}(\mu, \nu)$ be where the infimum of $\mathcal{P}_c(\mu, \nu)$ is attained and its representation by a (randomized) stopping time τ^* ; cf. Lemma 2.2. Then, for any $\psi \in \text{LSC}(\overline{O})$,

$$\begin{aligned} & \int_O \psi(y) \nu(dy) - \int_O J_{\psi}(x, x) \mu(dx) \\ &= \mathbb{E}[\psi(B_{\tau^*}) - J_{\psi}(B_0, B_0)] \\ &\leq \mathbb{E}[\psi(B_{\tau^*}) - J_{\psi}(B_0, B_{\tau^*})] \\ &\leq \mathbb{E}[c(B_0, B_{\tau^*})] = \mathcal{P}_c(\mu, \nu) = \mathcal{D}'_c(\mu, \nu), \end{aligned}$$

where we have used the definition (20) of J_{ψ} . Hence, $\mathcal{D}_c(\mu, \nu) \leq \mathcal{D}'_c(\mu, \nu)$.

For the other direction we consider $(\psi, J) \in \mathcal{A}_c$, then $J_\psi(x, y) \leq J(x, y)$ for $(x, y) \in O \times O$ since by Lemma 2.3 and the definitions (19) and (20), we have

$$J_\psi(x, y) = \sup_{\tau \leq \tau_O} \mathbb{E}[\psi(B_\tau^y) - c(x, B_\tau^y)] \leq \sup_{\tau \leq \tau_O} \mathbb{E}[J(x, B_\tau^y)] \leq J(x, y).$$

It follows that

$$\int_O \psi(y) \nu(dy) - \int_O J_\psi(x, x) \mu(dx) \geq \int_O \psi(y) \nu(dy) - \int_O J(x, x) \mu(dx),$$

and $D'_c(\mu, \nu) \leq D_c(\mu, \nu)$. This completes the proof. \square

3. A weakly compact set of D -superharmonic functions in Sobolev class. In this section we introduce the following convex set of functions \mathcal{B}_D which plays a key role in the sequel. These are the lower semicontinuous D -superharmonic functions that are nonpositive and zero on the boundary of O . A key property will be that these functions can be equivalently defined as members of the Sobolev class $H_0^1(O)$ that are nonpositive and have their Laplacian bounded above by D , weakly.

DEFINITION 3.1. Let O be a bounded convex open subset of \mathbb{R}^d . We say that a function $\psi \in \text{LSC}(\overline{O})$ is in the set \mathcal{B}_D , if the following properties hold:

1. $\psi(y) = 0$ for all $y \in \partial O$;
2. $\psi(y) \leq 0$ for all $y \in O$;
3. ψ is D -superharmonic, in the sense that for all stopping times $\tau \leq \tau_O$,

$$\psi(y) \geq \mathbb{E}[\psi(B_\tau^y) - D\tau].$$

(Equivalently, $\Delta\psi(y) \leq D$ for all $y \in O$ in the sense of viscosity by Lemma 2.3.)

The functions in this class \mathcal{B}_D have a uniform lower bound following from the maximum principle. Indeed, let u_O be the solution to

$$(24) \quad \begin{cases} \Delta u_O(x) = 1 & x \in O, \\ u_O(x) = 0 & x \in \partial O. \end{cases}$$

Equivalently, $u_O(y) = -\mathbb{E}[\tau_O]$, where τ_O is the exit time for the Brownian motion beginning at $B_0^y = y$. Since $\Delta\psi(y) \leq D$, we see that

$$(25) \quad \psi(y) \geq \mathbb{E}[\psi(B_{\tau_O}^y) - D\tau_O] = Du_O(y) \geq -D \sup_{z \in O} |u_O(z)|$$

which provides a lower bound depending only on O and D .

We now prove a key property of \mathcal{B}_D , namely, that it can be identified with a closed convex bounded (hence weakly compact) subset of the Sobolev class $H_0^1(O)$, equipped with the norm $\|f\|_{H_0^1(O)}^2 = \int_O |\nabla f(x)|^2 dx$.

PROPOSITION 3.2. A function ψ is in \mathcal{B}_D if and only if—up to a set of Lebesgue measure zero—it satisfies:

1. $\psi \in H_0^1(O)$;
2. $\psi(y) \leq 0$ for a.e. $y \in O$;
3. $\Delta\psi(y) \leq D$ weakly, in the sense that $\int_O [\nabla\psi(y) \cdot \nabla\phi(x) + D\phi(x)] dx \leq 0$ for all $\phi \in H_0^1(O)$ such that $\phi \leq 0$ a.e. on O .

Furthermore, there is a constant M dependent only on D and O such that

$$(26) \quad \|\psi\|_{H_0^1(O)} \leq M \quad \text{for all } \psi \in \mathcal{B}_D.$$

In other words, \mathcal{B}_D can be identified with a closed bounded convex subset of $H_0^1(O)$.

PROOF. First, use Lemma A.2 to fix a bounded convex domain \tilde{O} containing O such that we can approximate $\psi \in \mathcal{B}_D$ by smooth functions on \tilde{O} that satisfy: $\lim_{i \rightarrow \infty} \psi^i(x) = \psi(x)$, $\psi^i(x) \leq 0$ and $\Delta \psi^i(x) \leq D$ for $x \in \tilde{O}$ and $\psi^i(y) = 0$ for $y \in \partial \tilde{O}$. Notice that $\psi^i(x) \geq Du_{\tilde{O}}(x)$ for $x \in \tilde{O}$ for $u_{\tilde{O}}$ defined similarly as (24).

From the weak Laplacian bound with $\phi = \psi^i$, we get

$$\int_{\tilde{O}} |\nabla \psi^i(x)|^2 dx \leq - \int_{\tilde{O}} D \psi^i(x) dx \leq D^2 \|u_{\tilde{O}}\|_{L^1(O)}.$$

Hence, there is a subsequence of ψ^i that converges weakly in $H^1(O)$ to an equivalence class of ψ in $H_0^1(O)$ with $\|\psi\|_{H_0^1(O)}^2 \leq D^2 \|u_{\tilde{O}}\|_{L^1(\tilde{O})}$.

The properties in (2) and (3) follow as they are stable under weak convergence in $H^1(O)$.

Conversely, if $\psi \in H_0^1(O)$ satisfies $\Delta \psi \leq D$ weakly and $\psi \leq 0$, we can easily check that the extension of the function $\psi - Du_O$ to \mathbb{R}^d by zero is superharmonic in the sense that the average integral

$$r \mapsto \frac{1}{|B_r(y)|} \int_{B_r(y)} (\psi(z) - Du_O(z)) dz$$

is monotonically decreasing in r . It follows that ψ has a representative $\tilde{\psi}$ that is lower semicontinuous and D -superharmonic in the sense of viscosity; see, for instance, the notes [47]. This representative $\tilde{\psi}$ is everywhere nonpositive and $\tilde{\psi}(y) \geq Du_O(y)$ for all $y \in \bar{O}$; thus, $\tilde{\psi}$ is zero on the boundary. We have shown the lower semicontinuous representative $\tilde{\psi}$ of ψ lies in \mathcal{B}_D , completing the proof. \square

We now define the superharmonic envelope of a function $h \in \text{LSC}(\bar{O})$ to be

$$(27) \quad h^{\text{SH}}(y) = \sup_{\tau \leq \tau_O} \mathbb{E}[h(B_\tau^y)].$$

We note that the definition of J_ψ in (20) means that $y \mapsto J_\psi(x, y)$ is the superharmonic envelope of $y \mapsto \psi(y) - c(x, y)$. We will require in the sequel a few results on superharmonic envelopes.

LEMMA 3.3. Given $\phi \in H_0^1(O) \cap C^2(\bar{O})$, then its superharmonic envelope ϕ^{SH} , as defined in (27), is the unique minimizer of the variational problem, that is,

$$\phi^{\text{SH}} = \operatorname{argmin} \left\{ \int_O |\nabla u(y)|^2 dy; u \in H_0^1(O), u \geq \phi \right\}.$$

PROOF. Take $\phi \in H_0^1(O) \cap C^2(\bar{O})$, and let $\tilde{\phi} \in H_0^1(O)$ be the unique minimizer of the variational problem; uniqueness follows from the strict convexity of $u \mapsto \int |\nabla u|^2$. The optimality of $\tilde{\phi}$ implies that $\tilde{\phi}$ satisfies $\Delta \tilde{\phi} \leq 0$ weakly (see [36]). As in the proof of Proposition 3.2, we have that $\tilde{\phi}$ has a lower semicontinuous representative that satisfies $\Delta \tilde{\phi}(x) \leq 0$ for $x \in O$ in the sense of viscosity and $\tilde{\phi}(x) \geq \phi(x)$ for $x \in \bar{O}$. This implies that $\phi^{\text{SH}}(y) \leq \tilde{\phi}(y)$ for all $y \in \bar{O}$ since from Lemma 2.3,

$$\phi^{\text{SH}}(y) = \sup_{\tau \leq \tau_O} \mathbb{E}[\phi(B_\tau^y)] \leq \sup_{\tau \leq \tau_O} \mathbb{E}[\tilde{\phi}(B_\tau^y)] \leq \tilde{\phi}(y).$$

For any smooth superharmonic function $\hat{\phi} \in H_0^1(O) \cap C^2(\bar{O})$ greater than ϕ , we have by the same argument as above, $\hat{\phi} \geq \phi^{\text{SH}}$. It follows that $\phi^{\text{SH}} - \hat{\phi} \in \mathcal{B}_{D'}$ for $D' = \sup_{x \in O} -\Delta \hat{\phi}(x)$. Now use Proposition 3.2 to show that $\phi^{\text{SH}} \in H_0^1(O)$ and satisfies $\Delta \phi^{\text{SH}} \leq 0$ weakly. It follows that

$$\begin{aligned} \|\phi^{\text{SH}}\|_{H_0^1(O)}^2 &= \int_O \phi^{\text{SH}}(y)(-\Delta \phi^{\text{SH}}(y)) dy \\ &\leq \int_O \tilde{\phi}(y)(-\Delta \phi^{\text{SH}}(y)) dy \leq \|\tilde{\phi}\|_{H_0^1(O)} \|\phi^{\text{SH}}\|_{H_0^1(O)}. \end{aligned}$$

Thus, ϕ^{SH} is a minimizer of the variational problem; hence, from uniqueness we have $\phi^{\text{SH}} = \tilde{\phi}$. \square

We shall need the following property of the norm of $H^{-1}(O)$ which is dual to $H_0^1(O)$.

LEMMA 3.4. *Let τ and σ be two stopping times such that $\tau \leq \sigma \leq \tau_O$, and suppose that $B_\tau \sim \rho \in H^{-1}(O)$. Then, the distribution γ of B_σ belongs to $H^{-1}(O)$ and satisfies*

$$\|\gamma\|_{H^{-1}(O)} \leq \|\rho\|_{H^{-1}(O)}.$$

PROOF. For each $\phi \in H_0^1(O) \cap C^2(\bar{O})$, the superharmonic envelope ϕ^{SH} satisfies $\|\phi^{\text{SH}}\|_{H_0^1(O)} \leq \|\phi\|_{H_0^1(O)}$ by Lemma 3.3. Therefore,

$$\int_O \phi(y)\gamma(dy) \leq \int_O \phi^{\text{SH}}(y)\gamma(dy) \leq \int_O \phi^{\text{SH}}(y)\rho(dy) \leq \|\rho\|_{H^{-1}(O)} \|\phi\|_{H_0^1(O)}.$$

Here, the second inequality is due to the subharmonic order $\rho \prec_{\text{SH}} \gamma$ and subharmonicity of $-\phi^{\text{SH}}$. This proves the lemma as smooth functions are dense in $H_0^1(O)$. \square

The next lemma shows continuity of functions in class \mathcal{B}_D with respect to stopping times.

LEMMA 3.5. *We consider a sequence of randomized stopping times $\xi_i \leq \tau_O$, that converge weakly (in law) to a randomized stopping time ξ_∞ . Then, if $h \in \mathcal{B}_D$, we have*

$$(28) \quad \lim_{i \rightarrow \infty} \mathbb{E}[h(B_{\xi_i})] = \mathbb{E}[h(B_{\xi_\infty})].$$

PROOF. We first note that

$$(29) \quad \liminf_{i \rightarrow \infty} \mathbb{E}[h(B_{\xi_i})] \geq \mathbb{E}[h(B_{\xi_\infty})]$$

follows as a consequence of the Portmanteau theorem using the lower semicontinuity of h composed with the continuous paths of Brownian motion.

To prove the opposite inequality,

$$(30) \quad \limsup_{i \rightarrow \infty} \mathbb{E}[h(B_{\xi_i})] \leq \mathbb{E}[h(B_{\xi_\infty})],$$

we use that $h \in H_0^1(O)$ to control larger times and that h is D -superharmonic to control small times. We fix $\epsilon > 0$ and select $\delta > 0$ such that $\delta \leq \frac{\epsilon}{4D}$ and

$$\mathbb{E}[h(B_0)] \leq \mathbb{E}[h(B_{\xi_\infty \wedge \delta})] + \frac{\epsilon}{4}.$$

Note that the latter is possible by (29) since $\lim_{\delta \rightarrow 0} \xi_\infty \wedge \delta = 0$.

We decompose the expectation into two pieces,

$$(31) \quad \mathbb{E}[h(B_{\xi_i})] = \mathbb{E}[h(B_{\xi_i}) - h(B_{\xi_i \wedge \delta})] + \mathbb{E}[h(B_{\xi_i \wedge \delta})].$$

The second term on the right-hand side of (31) satisfies (because of D -superharmonicity of h) that

$$\mathbb{E}[h(B_{\xi_i \wedge \delta})] \leq \mathbb{E}[h(B_0)] + D\delta \leq \mathbb{E}[h(B_{\xi_\infty \wedge \delta})] + \frac{\epsilon}{2}.$$

We define $\tau_\delta = \tau_O \wedge \delta$. For the first term on the right-hand side of (31), because ξ_i and $\xi_i \wedge \delta$ coincide on the set where $\xi_i \leq \tau_\delta$, we have

$$\mathbb{E}[h(B_{\xi_i}) - h(B_{\xi_i \wedge \delta})] = \mathbb{E}[h(B_{\xi_i \vee \tau_\delta}) - h(B_{\tau_\delta})].$$

For $\rho_\delta \sim B_{\tau_\delta}$, the distribution is in $H^{-1}(O)$, by comparison with the heat kernel g_δ on \mathbb{R}^d since (by restricting to nonnegative test functions w and extending by zero outside of O)

$$\begin{aligned} \|\rho_\delta\|_{H^{-1}(O)} &= \sup_{w \geq 0, \|w\|_{H_0^1(O)} \leq 1} \mathbb{E}[w(B_{\tau_\delta})] \\ &\leq \sup_{w \geq 0, \|w\|_{H_0^1(O)} \leq 1} \mathbb{E}[w(B_\delta)] = \|g_\delta\|_{H^{-1}(O)} = C(\delta). \end{aligned}$$

We have that $\xi_i \vee \tau_\delta \geq \tau_\delta$, and thus, by Lemma 3.4, for $\rho_i \sim B_{\xi_i \vee \tau_\delta}$ and $\rho_\infty \sim B_{\xi_\infty \vee \tau_\delta}$,

$$\|\rho_i\|_{H^{-1}(O)} \leq C(\delta), \quad \|\rho_\infty\|_{H^{-1}(O)} \leq C(\delta).$$

Then, weak convergence of ξ_i implies $\rho_i \rightharpoonup \rho_\infty$ in $H^{-1}(O)$, and there is I such that, for $i \geq I$,

$$\mathbb{E}[h(B_{\xi_i \vee \tau_\delta})] \leq \mathbb{E}[h(B_{\xi_\infty \vee \tau_\delta})] + \frac{\epsilon}{2}.$$

Putting everything together, we have that

$$\begin{aligned} \mathbb{E}[h(B_{\xi_i}) - h(B_{\xi_i \wedge \delta})] &= \mathbb{E}[h(B_{\xi_i \vee \tau_\delta}) - h(B_{\tau_\delta})] \\ &\leq \mathbb{E}[h(B_{\xi_\infty \vee \tau_\delta}) - h(B_{\tau_\delta})] + \frac{\epsilon}{2} \\ &= \mathbb{E}[h(B_{\xi_\infty}) - h(B_{\xi_\infty \wedge \delta})] + \frac{\epsilon}{2}. \end{aligned}$$

Thus, we have shown that, for $i \geq I$, $\mathbb{E}[h(B_{\xi_i})] \leq \mathbb{E}[h(B_{\xi_\infty})] + \epsilon$ which implies (30) and completes the proof of (28). \square

The above property of \mathcal{B}_D enables us to prove the following lemma regarding verification of hitting times. We also include the case for $C(\overline{O})$.

LEMMA 3.6. *Assume that either $h \in C(\overline{O})$ or that $h \in \mathcal{B}_D$, and let h^{SH} be its superharmonic envelope defined in (27). For fixed $y \in \overline{O}$, we let*

$$\eta = \inf\{t; h^{\text{SH}}(B_t^y) = h(B_t^y)\}.$$

The stopping time η attains the supremum in the definition of h^{SH} . It also satisfies:

1. $h(B_\eta^y) = h^{\text{SH}}(B_\eta^y)$,
2. For any stopping time $\sigma \leq \tau_O$, we have $h^{\text{SH}}(y) = \mathbb{E}[h^{\text{SH}}(B_{\sigma \wedge \eta}^y)]$.

PROOF. *Case $h \in C(\overline{O})$:* Let us first show that there exist stopping times $\eta^i \leq \tau_O$ such that $h(B_{\eta^i}^y) = h^{\text{SH}}(B_{\eta^i}^y)$ and $\eta^i \rightarrow \eta$ weakly. Define for each $\epsilon > 0$ the set

$$H_\epsilon = \{(t, \omega); h^{\text{SH}}(B_t(\omega)) - h(B_t(\omega)) = 0, t < \eta + \epsilon\},$$

where each ω is a point in the probability space Ω . Note that $h^{SH} = h = 0$ along ∂O . Therefore, for almost every ω , there exists t with $(t, \omega) \in H_\epsilon$ by the definition of η ; thus, the projection of H_ϵ on Ω has full measure. By the “section theorem” in [15], there exists a stopping time η_ϵ such that $(\eta_\epsilon(\omega), \omega) \in H_\epsilon$ whenever $\eta_\epsilon(\omega) < \infty$, and $\mathbb{P}[\eta_\epsilon < \infty] \geq 1 - \epsilon$. Given a sequence ϵ^i converging to zero, we define $\eta^i := \eta_{\epsilon^i} \wedge \tau_O$ which has a subsequence converging weakly as desired.

Using the continuity of h and h^{SH} and the continuity of Brownian paths, we have

$$\mathbb{E}[h^{SH}(B_\eta^y)] = \lim_{i \rightarrow \infty} \mathbb{E}[h^{SH}(B_{\eta^i}^y)] = \lim_{i \rightarrow \infty} \mathbb{E}[h(B_{\eta^i}^y)] = \mathbb{E}[h(B_\eta^y)].$$

The supremum of definition (27) is attained at a randomized stopping time τ by compactness [18]. Optimality of τ implies that

$$h^{SH}(y) = \mathbb{E}[h(B_\tau^y)] = \sup_{\tau \leq \sigma \leq \tau_O} \mathbb{E}[h(B_\sigma^y)] = \mathbb{E}[h^{SH}(B_\tau^y)].$$

Since $h^{SH}(B_\tau^y) \geq h(B_\tau^y)$, we have $h(B_\tau^y) = h^{SH}(B_\tau^y)$ almost surely so $\tau \geq \eta$. Furthermore, h^{SH} is superharmonic, so we have

$$\mathbb{E}[h^{SH}(B_\tau^y)] \leq \mathbb{E}[h^{SH}(B_\eta^y)]$$

which implies $h^{SH}(y) = \mathbb{E}[h(B_\eta^y)]$. Thus, η also attains the supremum of (27), which again implies that

$$h(B_\eta^y) = \sup_{\eta \leq \sigma \leq \tau_O} \mathbb{E}[h(B_\sigma^y) | B_\eta^y] = h^{SH}(B_\eta^y),$$

proving (1).

Finally, for any stopping time $\sigma \leq \tau_O$, we have from the superharmonic property $h^{SH}(y) \geq \mathbb{E}[h^{SH}(B_{\sigma \wedge \eta}^y)] \geq \mathbb{E}[h^{SH}(B_\eta^y)] = h^{SH}(y)$ which implies (2) and completes the proof in the case where h is continuous.

Case $h \in \mathcal{B}_D$: If now h is in \mathcal{B}_D , the above proof is still valid thanks to Lemma 3.5. This can be used to obtain an optimal stopping time τ that maximizes

$$\sup_\tau \mathbb{E}[h(B_\tau^y)],$$

since the expectation of h is then a weakly continuous function of the stopping times. The same lemma can also be used to carry on the limit of $(\eta^i)_i$ to η in the above proof. \square

4. Dual attainment in Sobolev class. In this section we prove one of the main results of the paper, namely, the attainment of the supremum in the dual problem. This has been an elusive problem which has previously only been fully resolved for $d = 1$ in [7]. Recall that we assume $\text{supp } \mu$ and $\text{supp } \nu$ are contained in a bounded convex open set $O \subset \mathbb{R}^d$, and, by Theorem 2.1, we have the dual problem in the form

$$\mathcal{D}_c(\mu, \nu) = \sup_{\psi \in \text{LSC}(\overline{O})} \left\{ \int \psi(y) \nu(dy) - \int J_\psi(x, x) \mu(dx) \right\}.$$

We now state our main result on dual attainment.

THEOREM 4.1. *Assume that $c \in C(\overline{O} \times \overline{O})$ and $y \mapsto c(x, y)$ is subharmonic and D -superharmonic, that is, $0 \leq \Delta_y c(x, y) \leq D$ in the sense of viscosity, and that $\mu \prec_{SH} \nu$ as well as $\mu \in H^{-1}(O)$. There exists then $\psi^* \in \mathcal{B}_D$ that attains the maximum value of the dual problem, that is,*

$$(32) \quad \mathcal{D}_c(\mu, \nu) = \int_O \psi^*(y) \nu(dy) - \int_O J_{\psi^*}(x, x) \mu(dx).$$

REMARK 4.2. As long as $\Delta_y c(x, y) \geq -M$, we can form a problem with subharmonic cost $\tilde{c}(x, y) = c(x, y) + h(y)$, with h solving $\Delta h = M$. The solutions to these two problems are equivalent in the sense that if $\tilde{\psi}$ is an optimizer for the cost \tilde{c} , then $\tilde{\psi} - h$ is an optimizer for the cost c .

Furthermore, for costs of the form $c(x, y) = |x - y|^\alpha$ where $\alpha > 0$ and $d \geq 2$, a version of Theorem 4.1 applies with the additional assumption that the support of μ and ν are disjoint. The argument is given in Theorem 7.1 for the case of the distance function, that is, $\alpha = 1$.

The subharmonicity condition $\Delta_y c(x, y) \geq 0$ is, however, more essential than the D -superharmonic property. For example, we have a counterexample to dual attainment in [4] (see also [25]) for the cost $c(x, y) = -|x - y|$.

Before we prove the Theorem 4.1, we prove a few lemmata that will allow us to utilize the weak compactness of the set \mathcal{B}_D in H_0^1 (Proposition 3.2) for the proof of dual attainment.

LEMMA 4.3. *If μ is a probability measure on O and $\mu \in H^{-1}(O)$, then the map $\psi \in H_0^1(O) \mapsto \int_O J_\psi(x, x)\mu(dx)$ is convex and lower semicontinuous on \mathcal{B}_D .*

PROOF. We first prove the technical result that we may interchange the supremum and the expectation to obtain

$$\int_O J_\psi(x, x)\mu(dx) = \sup_{\tau \leq \tau_O} \mathbb{E}[\psi(B_\tau) - c(B_0, B_\tau)]$$

for the Brownian motion where $B_0 \sim \mu$. To see this equality, fix $\epsilon > 0$, and consider a measurable selection τ_x of stopping times for $B_0^x = x$ such that

$$J_\psi(x, x) \leq \mathbb{E}[\psi(B_{\tau_x}^x) - c(x, B_{\tau_x}^x)] + \epsilon.$$

Integrate with respect to μ to get a stopping time τ for $B_0 \sim \mu$ such that

$$\int_O J_\psi(x, x)\mu(dx) \leq \mathbb{E}[\psi(B_\tau) - c(B_0, B_\tau)] + \epsilon.$$

For the other direction, note that any stopping time τ disintegrates as τ_x , for $B_0^x = x$, and

$$\mathbb{E}[\psi(B_\tau) - c(B_0, B_\tau)] = \int_O \mathbb{E}[\psi(B_{\tau_x}^x) - c(x, B_{\tau_x}^x)]\mu(dx) \leq \int_O J_\psi(x, x)\mu(dx).$$

Now consider $B_0 \sim \mu \in H^{-1}(O)$. From Lemma 3.4, if $\tau \leq \tau_O$ and $B_\tau \sim \rho$, then $\rho \in H^{-1}(O)$. Therefore, for each $\tau \leq \tau_O$, $\psi \in H_0^1(O) \mapsto \mathbb{E}[\psi(B_\tau) - c(B_0, B_\tau)]$ is linear and continuous. This shows the desired convexity and lower semicontinuity, as the map is given by the supremum of continuous linear functionals. \square

The next step is to show that the maximization of the dual problem $\mathcal{D}_c(\mu, \nu)$ can be restricted to the smaller set \mathcal{B}_D .

PROPOSITION 4.4. *We assume that $y \mapsto c(x, y)$ is subharmonic and D -superharmonic, that is, $0 \leq \Delta_y c(x, y) \leq D$ as well as $\mu \prec_{SH} \nu$. Then, it holds that*

$$\mathcal{D}_c(\mu, \nu) = \sup_{\psi \in \mathcal{B}_D} \left\{ \int_O \psi(y)\nu(dy) - \int_O J_\psi(x, x)\mu(dx) \right\}.$$

The proof of this proposition is done as two improvements to ψ . To maximize the dual value, we wish to choose ψ as large as possible while maintaining $\psi(x) \leq J_\psi(x, y) + c(x, y)$ for every $x \in \bar{O}$ which motivates the following lemma.

LEMMA 4.5. *Suppose $y \mapsto h(x, y)$ is D -superharmonic for each $x \in O$:*

1. *The function $\bar{\psi}(y) := \inf_{z \in \bar{O}} h(z, y)$ is D -superharmonic.*
2. *For $h(x, y) = J_\psi(x, y) + c(x, y)$ and if $y \mapsto c(x, y)$ is D -superharmonic, then*
 - *$\bar{\psi}$ is D -superharmonic,*
 - *$\bar{\psi}(y) \geq \psi(y)$ for all $y \in \bar{O}$, and*
 - *$J_{\bar{\psi}}(x, y) = J_\psi(x, y)$ for all $(x, y) \in \bar{O} \times \bar{O}$.*

PROOF. For any $y \in \bar{O}$ and $\epsilon > 0$, there is $z \in \bar{O}$ such that $\bar{\psi}(y) + \epsilon \geq h(z, y)$; hence, for any stopping time $\tau \leq \tau_O$,

$$\bar{\psi}(y) + \epsilon \geq h(z, y) \geq \mathbb{E}[h(z, B_\tau^y) - D\tau] \geq \mathbb{E}[\bar{\psi}(B_\tau^y) - D\tau],$$

where we have used that $y \mapsto h(x, y)$ is D -superharmonic and $\bar{\psi}(y') \leq h(z, y')$ for all $y' \in \bar{O}$. The first item (1) follows by letting $\epsilon \rightarrow 0$.

For the second part of the lemma, we have that $y \mapsto c(x, y) + J_\psi(x, y)$ is D -superharmonic since c is D -superharmonic and J_ψ is superharmonic. That $\bar{\psi}(y) \geq \psi(y)$ is obvious from the fact that $\psi(y) \leq J_\psi(x, y) + c(x, y)$. For the last item in (2), observe that $J_\psi \leq J_{\bar{\psi}}$ as

$$J_\psi(x, y) = \sup_{\tau \leq \tau_O} \mathbb{E}[\psi(B_\tau^y) - c(x, B_\tau^y)] \leq \sup_{\tau \leq \tau_O} \mathbb{E}[\bar{\psi}(B_\tau^y) - c(x, B_\tau^y)] = J_{\bar{\psi}}(x, y).$$

By definition of $\bar{\psi}$, we have that $\bar{\psi}(y) \leq J_\psi(x, y) + c(x, y)$, thus we have that, for all $\tau \leq \tau_O$,

$$\mathbb{E}[\bar{\psi}(B_\tau^y) - c(x, B_\tau^y)] \leq \mathbb{E}[J_\psi(x, B_\tau^y)] \leq J_\psi(x, y),$$

and it follows that $J_\psi \geq J_{\bar{\psi}}$. \square

The second improvement of ψ is slightly more subtle. We make use of the assumption that $\mu \prec_{SH} \nu$ so that subtracting a superharmonic function of y from both ψ and J_ψ will not decrease the dual value. In general, subtracting a superharmonic function would violate the constraint that $y \mapsto J_\psi(x, y)$ is superharmonic, but subtracting the superharmonic envelope of ψ from J_ψ miraculously does not violate this constraint given that $y \mapsto c(x, y)$ is subharmonic.

PROPOSITION 4.6. *Assume that $c \in C(\bar{O} \times \bar{O})$ and for all $x \in \bar{O}$, $y \mapsto c(x, y)$ is subharmonic, that is, $\Delta_y c(x, y) \geq 0$. Then, for each $\psi \in C(\bar{O})$, we have*

$$J_{\psi - \psi^{SH}}(x, y) = J_\psi(x, y) - \psi^{SH}(y), \quad \text{for all } (x, y) \in \bar{O} \times \bar{O}.$$

PROOF. We fix (x, y) in $\bar{O} \times \bar{O}$. The proof that $J_{\psi - \psi^{SH}}(x, y) \geq J_\psi(x, y) - \psi^{SH}(y)$ is easy and does not need the assumption $\Delta_y c(x, y) \geq 0$. Indeed, we let $\bar{\tau}$ be a (randomized) stopping time that attains the supremum for J_ψ , so that

$$J_\psi(x, y) = \mathbb{E}[\psi(B_{\bar{\tau}}^y) - c(x, B_{\bar{\tau}}^y)].$$

Using $\mathbb{E}[\psi^{SH}(B_{\bar{\tau}}^y)] \leq \psi^{SH}(y)$, we have

$$J_\psi(x, y) - \psi^{SH}(y) \leq \mathbb{E}[\psi(B_{\bar{\tau}}^y) - c(x, B_{\bar{\tau}}^y) - \psi^{SH}(B_{\bar{\tau}}^y)] \leq J_{\psi - \psi^{SH}}(x, y).$$

The last inequality is due to the definition of $J_{\psi - \psi^{SH}}$.

The reverse inequality is important for the proof of Proposition 4.4 and requires the assumption $\Delta_y c(x, y) \geq 0$. We let τ be a (randomized) stopping time attaining the supremum for $J_{\psi - \psi^{SH}}$, so that

$$J_{\psi - \psi^{SH}}(x, y) = \mathbb{E}[\psi(B_\tau^y) - \psi^{SH}(B_\tau^y) - c(x, B_\tau^y)].$$

We consider the first hitting time: $\eta = \inf\{t; \psi(B_t^y) = \psi^{\text{SH}}(B_t^y)\}$. We can write

$$\begin{aligned} & \mathbb{E}[\psi(B_\tau^y) - \psi^{\text{SH}}(B_\tau^y) - c(x, B_\tau^y)] \\ &= \mathbb{E}[\psi(B_{\tau \wedge \eta}^y) - \psi^{\text{SH}}(B_{\tau \wedge \eta}^y) - c(x, B_{\tau \wedge \eta}^y)] \\ & \quad + \mathbb{E}[\psi(B_\tau^y) - \psi(B_{\tau \wedge \eta}^y) - \psi^{\text{SH}}(B_\tau^y) + \psi^{\text{SH}}(B_{\tau \wedge \eta}^y)] \\ & \quad - \mathbb{E}[c(x, B_\tau^y) - c(x, B_{\tau \wedge \eta}^y)] \\ &= I + II + III. \end{aligned}$$

For I , use item (2) of Lemma 3.6 and the definition of J_ψ to see

$$I \leq J_\psi(x, y) - \psi^{\text{SH}}(y).$$

For III , by the assumption $\Delta_y c(x, y) \geq 0$, we see that $III \leq 0$. For the term II , notice that

$$\text{if } \tau = \tau \wedge \eta \quad \text{then } \psi(B_\tau^y) = \psi(B_{\tau \wedge \eta}^y) \quad \text{and} \quad \psi^{\text{SH}}(B_\tau^y) = \psi^{\text{SH}}(B_{\tau \wedge \eta}^y).$$

Moreover, by item (1) of Lemma 3.6

$$\text{if } \eta = \tau \wedge \eta, \quad \text{then } \psi(B_{\tau \wedge \eta}^y) = \psi^{\text{SH}}(B_{\tau \wedge \eta}^y).$$

Recall that $\psi \leq \psi^{\text{SH}}$ always. Therefore, we can conclude that $II \leq 0$. All of these together imply that $J_{\psi - \psi^{\text{SH}}}(x, y) \leq J_\psi(x, y) - \psi^{\text{SH}}(y)$, as desired, completing the proof. \square

We now have the necessary ingredients to prove Proposition 4.4, which is the last component of the proof of Theorem 4.1.

PROOF OF PROPOSITION 4.4. The inequality

$$\mathcal{D}_c(\mu, \nu) \geq \sup_{\psi \in \mathcal{B}_D} \left\{ \int_O \psi(y) \nu(dy) - \int_O J_\psi(x, x) \mu(dx) \right\}$$

follows directly from the definitions of \mathcal{D}_c and \mathcal{B}_D as $\mathcal{B}_D \subset \text{LSC}(\overline{O})$.

For the reverse inequality, first notice that, for each $\phi \in \text{LSC}(\overline{O})$, there exists a sequence of continuous functions ϕ^i such that $\lim_{i \rightarrow \infty} \phi^i(x) = \phi(x)$, $\forall x \in \overline{O}$; for example, one can consider the inf-convolution as in the proof of Lemma A.1. Therefore, it suffices to prove that, for any $\psi \in C(\overline{O})$, we can modify it to $\bar{\psi} \in \mathcal{B}_D$ such that

$$\int_O \psi(y) \nu(dy) - \int_O J_\psi(x, x) \mu(dx) \leq \int_O \bar{\psi}(y) \nu(dy) - \int_O J_{\bar{\psi}}(x, x) \mu(dx).$$

(We can then apply the modification to a maximizing sequence of \mathcal{D}_c to get another maximizing sequence but now from the class \mathcal{B}_D .)

Improvement 1 ($\psi = 0$ on ∂O and $\psi \leq 0$ in O). We first modify ψ to $\psi - \psi^{\text{SH}}$, using the superharmonic envelope ψ^{SH} given in (27). We see that

$$\begin{aligned} & \int_O \psi(y) \nu(dy) - \int_O J_\psi(x, x) \mu(dx) \\ & \leq \int (\psi(y) - \psi^{\text{SH}}(y)) \nu(dy) - \int (J_\psi(x, x) - \psi^{\text{SH}}(x)) \mu(dx) \end{aligned}$$

because ψ^{SH} is superharmonic and $\mu \prec_{\text{SH}} \nu$. It is important to notice that, from Proposition 4.6, $J_{\psi - \psi^{\text{SH}}} = J_\psi - \psi^{\text{SH}}$. Also, notice that, because \overline{O} is compact, the continuous function ψ is uniformly continuous and so are $\psi - \psi^{\text{SH}}$ and $J_{\psi - \psi^{\text{SH}}}$.

Improvement 2 ($\Delta\psi \leq D$ in O). Now modify the function $\psi - \psi^{\text{SH}}$ further to

$$\bar{\psi}(y) = \inf_{z \in \bar{O}} \{J_{\psi - \psi^{\text{SH}}}(z, y) + c(z, y)\},$$

as in Lemma 4.5 with $h = J_{\psi - \psi^{\text{SH}}} + c$. Here, $\bar{\psi}$ is continuous on \bar{O} following from uniform continuity of $J_{\psi - \psi^{\text{SH}}}$ and c . From the lemma we have that $\bar{\psi}$ is D -superharmonic, $\psi - \psi^{\text{SH}} \leq \bar{\psi}$ and $J_{\bar{\psi}} = J_{\psi - \psi^{\text{SH}}}$. Then, the last line in the above inequality is less than or equal to

$$\int_O \bar{\psi}(y)\nu(dy) - \int_O J_{\bar{\psi}}(x, x)\mu(dx),$$

getting the desired inequality.

We also have that $\bar{\psi}(y) \leq 0$ for all y , since

$$J_{\psi - \psi^{\text{SH}}}(x, y) + c(x, y) = \sup_{\tau} \mathbb{E}[\psi(B_{\tau}^y) - \psi^{\text{SH}}(B_{\tau}^y) - c(x, B_{\tau}^y) + c(x, y)] \leq 0,$$

and for $y \in \partial O$, $\bar{\psi}(y) = 0$ since $J_{\psi - \psi^{\text{SH}}}(x, y) = -c(x, y)$. Thus, by definition this implies that $\bar{\psi} \in \mathcal{B}_D$. This completes the proof. \square

PROOF OF THEOREM 4.1. We use Proposition 4.4 to find a sequence $\psi_i \in \mathcal{B}_D$ such that

$$\mathcal{D}_c(\mu, \nu) = \lim_{i \rightarrow \infty} \left\{ \int_O \psi_i(y)\nu(dy) - \int_O J_{\psi_i}(x, x)\mu(dx) \right\}.$$

Uniform boundedness of $\|\psi_i\|_{H_0^1(O)}$, given by (26) of Proposition 3.2, implies there is a weak limit $\psi^* \in H_0^1(O)$ of ψ_i . Note also that such a weak limit preserves the property $\psi^* \leq 0$ as well as $\Delta\psi^* \leq D$ in the weak sense, so by the equivalence of Proposition 3.2 we have $\psi^* \in \mathcal{B}_D$. Since $\mu \in H^{-1}(O)$ and $\mu \prec_{\text{SH}} \nu$, we have $\nu \in H^{-1}(O)$ from Lemma 3.4, and

$$\lim_{i \rightarrow \infty} \int_O \psi_i(y)\nu(dy) = \int_O \psi^*(y)\nu(dy).$$

On the other hand, from the lower semicontinuity shown in Lemma 4.3, we have

$$\lim_{i \rightarrow \infty} \int_O J_{\psi_i}(x, x)\mu(dx) \geq \int_O J_{\psi^*}(x, x)\mu(dx)$$

which then implies

$$\int_O \psi^*(y)\nu(dy) - \int_O J_{\psi^*}(x, x)\mu(dx) \geq \mathcal{D}_c(\mu, \nu).$$

Since $\psi^* \in \text{LSC}(\bar{O})$, the above two inequalities are, in fact, equalities, showing the identity (32). This completes the proof. \square

We now demonstrate the verification properties for how the dual optimizers pair with the primal minima. We will use these general results to prove uniqueness of the optimal stopping times in Section 6 after introducing the crucial assumption on the cost, the stochastic twist condition in Section 5.

Let π^* be an optimizer of $\mathcal{P}_c(\mu, \nu)$, and let τ^* be the corresponding optimal (randomized) stopping time. Notice that, for the disintegration $\pi^*(dx, dy) = \pi_x^*(dy)\mu(dx)$, the measure π_x^* describes the distribution of the stopped Brownian paths $B_{\tau^*}^x$. The dual optimizers (ψ^*, J_{ψ^*}) in Theorem 4.1 are useful tools for us to characterize this measure. For example, we have the following important property.

THEOREM 4.7. *Under the same assumptions as in Theorem 4.1, the optimal Brownian path stops at the contact set, namely, for π^* -a.e. $(x, y) \in O \times O$,*

$$J_{\psi^*}(x, y) = \psi^*(y) - c(x, y),$$

and, in particular, $\tau^* \geq \eta := \inf\{t; J_{\psi^*}(B_0, B_t) = \psi^*(B_t) - c(B_0, B_t)\}$.

Furthermore, if $\sigma(dx, dy) = \sigma_x(dy)\mu(dx)$ satisfies $0 \prec_{SH} \sigma_x \prec_{SH} \pi_x^*$ for each μ -a.e. x , then

$$(33) \quad \int_{O \times O} J_{\psi^*}(x, y)\pi^*(dx, dy) = \int_{O \times O} J_{\psi^*}(x, y)\sigma(dx, dy) = \int_O J_{\psi^*}(x, x)\mu(dx).$$

If ξ is a (nonrandomized) stopping time corresponding to σ , then $\xi \leq \tau^*$ and for μ a.e. x and ξ restricted to the paths with $B_0 = x$,

$$(34) \quad \mathbb{E}[J_{\psi^*}(x, B_\xi^x)] = J_{\psi^*}(x, x).$$

Also, for σ a.e. (x, y) and the stopping time $\tau^* - \xi$ restricted to paths satisfying $B_\xi = y$,

$$(35) \quad \mathbb{E}[J_{\psi^*}(x, B_{\tau^*-\xi}^y)] = J_{\psi^*}(x, y).$$

In particular, these hold for $\xi = \eta$.

PROOF. For σ , satisfying $\sigma(dx, dy) = \sigma_x(dy)\mu(dx)$ with $\delta_x \prec_{SH} \sigma_x \prec_{SH} \pi_x$ for each μ -a.e. x , from the superharmonic property of J_{ψ^*} we have the inequalities

$$\int_{O \times O} J_{\psi^*}(x, y)\pi^*(dx, dy) \leq \int_{O \times O} J_{\psi^*}(x, y)\sigma(dx, dy) \leq \int_O J_{\psi^*}(x, x)\mu(dx).$$

Therefore, from the obvious inequality $J_{\psi^*}(x, y) \geq \psi^*(y) - c(x, y)$, we see that

$$\begin{aligned} \int_{O \times O} c(x, y)\pi^*(dx, dy) &\geq \int_{O \times O} [\psi^*(y) - J_{\psi^*}(x, y)]\pi^*(dx, dy) \\ &\geq \int_O \psi^*(y)v(dy) - \int_{O \times O} J_{\psi^*}(x, y)\sigma(dx, dy) \\ &\geq \int_O \psi^*(y)v(dy) - \int_O J_{\psi^*}(x, x)\mu(dx). \end{aligned}$$

Strong duality (Theorem 4.1) implies that the first and the last end of these inequalities are the same, making all the inequalities equalities. In particular, this implies that $J_{\psi^*}(x, y) = \psi^*(y) - c(x, y)$. for π^* -a.e. (x, y) . The equalities (33) also follow. Then, the equalities (34) and (35) follow from (33) and the disintegration of π^* , σ , with respect to σ , μ , respectively. In other words,

$$\int_O \mathbb{E}[J_{\psi^*}(x, B_\xi^x)]\mu(dx) = \int_O J_{\psi^*}(x, x)\mu(dx),$$

(34) follows since $J_{\psi^*}(x, B_\xi^x) \leq J_{\psi^*}(x, x)$ and

$$\begin{aligned} \int_{O \times O} J_{\psi^*}(x, y)\pi^*(dx, dy) &= \int_{O \times O} \mathbb{E}[J_{\psi^*}(x, B_{\tau^*-\xi}^y)]\sigma(dx, dy) \\ &= \int_{O \times O} \mathbb{E}[J_{\psi^*}(x, y)]\sigma(dx, dy); \end{aligned}$$

(35) follows similarly. \square

5. A stochastic twist condition. We now start the discussion of our results on uniqueness and characterization of the optimal Brownian stopping time. In this section we give a stochastic version of the twist condition (4) on the cost in the deterministic optimal transport theory. This condition will allow us to prove later that the optimal stopping time is uniquely given by the first hitting time to the contact set.

DEFINITION 5.1. We say that c satisfies the stochastic twist (ST) condition at (x, y) if for each stopping time $\xi \leq \tau_O$,

$$(36) \quad \mathbb{E}[\nabla_x c(x, B_\xi^y)] = \nabla_x c(x, y) \implies \xi = 0.$$

Notice that this stochastic twist condition **ST** is not a direct generalization of the usual twist condition in optimal transport theory. In particular, the quadratic cost $c(x, y) = |x - y|^2$ does not satisfy **ST**, because $\nabla_x |x - y|^2 = 2(x - y)$; therefore, the equality in (36) holds for any $\tau \geq 0$ due to the martingale property.

We now collect some examples of costs that satisfy **ST**:

- Our most important example of a cost satisfying **ST** is the distance function $c(x, y) = |x - y|$ for dimensions $d \geq 2$ and points $x \neq y$. The gradient is $\nabla_x c(x, y) = \frac{x-y}{|x-y|} \in S^{d-1}$, valued in the unit sphere. In particular, we have $\nabla_x c(x, y) \cdot \nabla_x c(x, y) = 1$ and $\mathbb{E}[\nabla_x c(x, y) \cdot \nabla_x c(x, B_\sigma^y)] < 1$ for any stopping time $\sigma > 0$ which implies **ST**. Note that the same argument shows that any (differentiable) Riemannian distance function $d(x, y)$ satisfies **ST**, since $\nabla_x d(x, y)$ is always a unit tangent vector at x .
- A more general class of costs with **ST** can be described by the local condition: for each x , there exist convex functions f_x such that

$$(37) \quad y \mapsto f_x(\nabla_x c(x, y)) \text{ is strictly superharmonic.}$$

In this case we have for $\sigma \neq 0$

$$f_x(\mathbb{E}[\nabla_x c(x, B_\sigma^y)]) \leq \mathbb{E}[f_x(\nabla_x c(x, B_\sigma^y))] < f_x(\nabla_x c(x, y)).$$

- A simple subclass of the previous example are the separable costs

$$c(x, y) = g(x)h(y)$$

that satisfy $\nabla g(x) \neq 0$ and $y \mapsto h(y)$ is either strictly superharmonic or strictly subharmonic. In either case, $\nabla_x c(x, y) = \nabla g(x)h(y)$, and we select $f_x(z) = \pm \nabla g(x) \cdot z$ where the sign is positive if h is superharmonic and negative if h is subharmonic.

- For cost functions that only satisfy

$$\Delta_y \nabla_x c(x, y) \neq \mathbf{0} \text{ for all } (x, y) \in O \times O,$$

a localized version of **ST** holds, that is, there is $\delta > 0$ such that if $0 \leq \tau \leq \delta$ and (36), then $\tau = 0$. Note, however, that in dimension one this condition is sufficient to imply **ST** (i.e., that $\Delta_y \nabla_x c(x, y) = c_{yyx}(x, y) > 0$ which appears in [33] and [31]).

REMARK 5.2. One can also consider a version of the stochastic twist condition for martingale transport.

DEFINITION 5.3. We say that c satisfies the martingale twist (MT) condition at (x, y) if for each probability measure σ such that $\delta_y \prec_C \sigma$ (for the convex order), we have

$$(38) \quad \int_O \nabla_x c(x, z) \sigma(dz) = \nabla_x c(x, y) \implies \sigma = \delta_y.$$

The above examples for **(ST)** hold for **(MT)** if we replace the superharmonicity, subharmonicity with concavity, convexity, respectively. In particular, the distance cost $c(x, y) = |x - y|$ is **(MT)**, and this property was crucially used in [25] to prove uniqueness and structure of optimal martingale transport under various conditions.

6. Uniqueness of the Monge solution: Optimal stopping as a hitting time. The stochastic twist condition **(ST)** allows us to prove uniqueness of the optimal Brownian martingale and characterize it as the first hitting time to a barrier set. We will assume the technical assumption $\mu(\partial \text{supp } \mu) = 0$ and the simplifying structural assumptions that $\mu \wedge \nu = 0$ which, in most cases, can be omitted by applying Lemma 7.3.

THEOREM 6.1. *Suppose, in addition to the assumptions of Theorem 4.1, that $x \mapsto c(x, y)$ is C^1 and c satisfies the stochastic twist condition **(ST)** for all $(x, y) \in O \times O$. Assume further that $\mu \ll \text{Leb}$, $\mu(\partial \text{supp } \mu) = 0$, and $\mu \wedge \nu = 0$. Then, there exists a unique optimal stopping time that is given by*

$$(39) \quad \eta := \inf\{t; J_{\psi^*}(B_0, B_t) = \psi^*(B_t) - c(B_0, B_t)\},$$

where ψ^* is the dual optimizer of Theorem 4.1, and J_{ψ^*} is the value function satisfying (14).

We will need several technical lemmata that address differentiability issues for $J_{\psi^*}(x, y)$ in our proof of Theorem 6.1. The dynamic programming principle for J_{ψ} allows us easily verify the following remarkable Lipschitz continuity.

LEMMA 6.2. *Assume that, for each $y \in \overline{O}$, we have that $\|x \mapsto c(x, y)\|_{\text{Lip}} \leq K$ for some constant $K > 0$ (independent on y). Then, for each $y \in \overline{O}$, $\|x \mapsto J_{\psi}(x, y)\|_{\text{Lip}} \leq K$.*

PROOF. This is an easy conclusion from the definition of J_{ψ} by stopping times, (14), because

$$\|x \mapsto \mathbb{E}[\psi(B_{\tau}^y) - c(x, B_{\tau}^y)]\|_{\text{Lip}} \leq K$$

for each τ . The supremum over those Lipschitz functions is again Lipschitz with the same Lipschitz constant K . \square

The following two lemmas deal with the differentiability of two relevant integrals (expected values). We first verify harmonicity of $y \mapsto J_{\psi^*}(x, y)$ in a small neighborhood.

LEMMA 6.3. *Use the same assumptions and notation as in Theorem 6.1. Then, for each $x \in \text{int}(\text{supp } \mu)$, $y \mapsto J_{\psi^*}(x, y)$ is harmonic in an open neighborhood around x .*

PROOF. For μ -a.e. $x \in \text{int}(\text{supp } \mu)$, choose an open ball $V_{\epsilon}(x) \subset \text{int}(\text{supp } \mu)$ centered at x , with radius $\epsilon > 0$ such that $\nu(V_{\epsilon}(x)) = 0$. For each $r \leq \epsilon$, let ξ^r be the first hitting time to $\partial V_r(x)$. Define $u : V_{\epsilon}(x) \rightarrow \mathbb{R}$ as

$$u(y) = \mathbb{E}[J_{\psi^*}(x, B_{\xi^{\epsilon}}^y)].$$

Because of Markov property of the Brownian motion, u satisfies the mean value property, so, is harmonic. Recall that $y \mapsto J_{\psi^*}(x, y)$ is superharmonic. Therefore, $J(x, y) \geq u(y)$ for all $y \in V_{\epsilon}(x)$. Moreover, because of our assumption $\mu \wedge \nu = 0$, we have $\xi^r \leq \tau^*$, $0 \leq r \leq \epsilon$, for the optimal stopping τ^* of $\mathcal{P}_c(\mu, \nu)$. Therefore, from the verification theorem (Theorem 4.7), we see that for μ -a.e. x ,

$$(40) \quad J_{\psi^*}(x, x) = \mathbb{E}[J_{\psi^*}(x, B_{\xi^r}^x)] = \mathbb{E}[J_{\psi^*}(x, B_{\tau^*}^x)].$$

In the case $r = \epsilon$, we have $J_{\psi^*}(x, x) = u(x)$. Then, for other r , (40) and the inequality $J_{\psi^*} \geq u$ imply that $J_{\psi^*}(x, y) = u(y)$ for a.e. $y \in V_\epsilon(x)$. Now lower semicontinuity of J_{ψ^*} and the inequality $J_{\psi^*} \geq u$ imply $J_{\psi^*}(x, y) = u(y)$ for all $y \in V_\epsilon(x)$. To have the harmonicity for all $x \in \text{int}(\text{supp } \mu)$, not just μ -a.e., use the (Lipschitz) continuity of $x \mapsto J_{\psi^*}(x, y)$, to extend this harmonicity to all $x \in \text{int}(\text{supp } \mu)$. This completes the proof. \square

LEMMA 6.4. *Use the same assumptions and notation as in Theorem 6.1. Let τ^* be an optimal stopping time of our problem $\mathcal{P}_c(\mu, \nu)$. Let ζ be any stopping time with $\zeta \leq \tau^*$ satisfying*

$$(41) \quad \mathbb{E}[J_{\psi^*}(x, B_\zeta^x)] = \mathbb{E}[\psi^*(B_\zeta^x) - c(x, B_\zeta^x)] \quad \text{for } \mu\text{-a.e. } x.$$

(In particular, $\zeta = \eta$, the hitting time defined in (39), satisfies these from Lemma 3.6(1).) Then, for μ -a.e. x the functions

$$h \mapsto J_{\psi^*}(x + h, x), \quad h \mapsto \mathbb{E}[J_{\psi^*}(x + h, B_\zeta^x)], \quad \text{and} \quad h \mapsto \mathbb{E}[J_{\psi^*}(x + h, B_{\tau^*}^x)]$$

are differentiable at $h = 0$ and

$$(42) \quad \begin{aligned} & \left. \frac{d}{dh} \right|_{h=0} J_{\psi^*}(x + h, x) \\ &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x + h, B_\zeta^x)] = \mathbb{E}[-\nabla_x c(x, B_\zeta^x)] \\ &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x + h, B_{\tau^*}^x)] = \mathbb{E}[-\nabla_x c(x, B_{\tau^*}^x)]. \end{aligned}$$

PROOF. From Lemma 6.3 and the gradient estimates of harmonic functions, the function $y \mapsto J_{\psi^*}(x, y)$ is locally Lipschitz in an open neighborhood of each $x \in \text{int}(\text{supp } \mu)$. Here, the local Lipschitz constant is uniform in x in a neighborhood, since c is continuous and the function J_{ψ^*} is bounded in $O \times O$ (which follows from the boundedness of $\psi^* \in \mathcal{B}_D$). Combining this with the fact $\|x \mapsto J_{\psi^*}(x, y)\|_{\text{Lip}} \leq C$ (from Lemma 6.2) for each y , we get the function $(x, y) \mapsto J_{\psi^*}$ is locally Lipschitz on an open neighborhood \mathcal{N} of the diagonal set $\{(x, x) | x \in \text{int}(\text{supp } \mu)\}$, contained in $\text{int}(\text{supp } \mu) \times \text{int}(\text{supp } \mu)$. By Rademacher’s theorem J_{ψ^*} is differentiable a.e. in the same set. By Fubini’s theorem and $\mu \ll \text{Leb}$, this implies that, for μ -a.e. x , the function $h \mapsto J_{\psi^*}(x + h, y)$ is differentiable at $h = 0$ for a.e. y in an open neighborhood of x . Because of the assumptions $\mu(\partial \text{supp } \mu) = 0$ and $\mu \ll \text{Leb}$, we can, without loss of generality, assume that $x \in \text{int}(\text{supp } \mu)$, and the ϵ -ball $V_\epsilon(x) \subset \text{int}(\text{supp } \mu)$ and for all sufficiently small h , the function $y \in V_\epsilon(x) \mapsto J(x + h, y)$ is harmonic.

Choose a stopping time ξ such that $\xi \leq \sigma^\epsilon$ for the first hitting time σ^ϵ to the sphere $\partial V_\epsilon(x)$ and $B_\xi^x \sim \rho \ll \text{Leb}$. Then, from the bound $\|x \mapsto J_{\psi^*}(x, y)\|_{\text{Lip}} \leq C$ and the dominated convergence theorem, we see that

$$\left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x + h, B_\xi^x)] = \mathbb{E} \left[\left. \frac{d}{dh} \right|_{h=0} J_{\psi^*}(x + h, B_\xi^x) \right].$$

In particular, the derivative exists. We now use the harmonicity of $y \in V_\epsilon(x) \mapsto J_{\psi^*}(x + h, y)$ for sufficiently small h to see

$$J_{\psi^*}(x + h, x) = \mathbb{E}[J_{\psi^*}(x + h, B_\xi^x)].$$

So, $h \mapsto J_{\psi^*}(x + h, x)$ is differentiable at $h = 0$.

As $\mu \wedge \nu = 0$ and $V_\epsilon(x) \in \text{int}(\text{supp } \mu)$, we see that $\sigma^\epsilon \leq \tau^*$, thus $\xi \leq \tau^*$. This order implies

$$\begin{aligned} J_{\psi^*}(x+h, x) &= \mathbb{E}[J_{\psi^*}(x+h, B_\xi^x)] \\ &\geq \mathbb{E}[J_{\psi^*}(x+h, B_{\tau^*}^x)] \\ &\geq \mathbb{E}[\psi^*(B_{\tau^*}^x) - c(x+h, B_{\tau^*}^x)]. \end{aligned}$$

Note that from Theorem 4.7 (in particular, (34)) we have

$$J_{\psi^*}(x, x) = \mathbb{E}[J_{\psi^*}(x, B_\xi^x)] = \mathbb{E}[\psi^*(B_{\tau^*}^x) - c(x, B_{\tau^*}^x)].$$

Therefore, at $h = 0$ the above inequalities become equalities. Since $J_{\psi^*}(x+h, x) = \mathbb{E}[J_{\psi^*}(x+h, B_\xi^x)]$ and $\mathbb{E}[\psi^*(B_{\tau^*}^x) - c(x+h, B_{\tau^*}^x)]$ are both differentiable at $h = 0$, we see that the function

$$h \mapsto \mathbb{E}[J_{\psi^*}(x+h, B_{\tau^*}^x)] \quad \text{is differentiable at } h = 0,$$

and all of these have the same derivatives,

$$\begin{aligned} \left. \frac{d}{dh} \right|_{h=0} J_{\psi^*}(x+h, x) &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x+h, B_{\tau^*}^x)] = \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[\psi^*(B_{\tau^*}^x) - c(x+h, B_{\tau^*}^x)] \\ &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[-c(x+h, B_{\tau^*}^x)] = \mathbb{E}[-\nabla_x c(x, B_{\tau^*}^x)]. \end{aligned}$$

For ζ it is not clear whether $\xi \leq \zeta$. Therefore, to examine the derivative of the function $\mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)]$, we notice that

$$J_{\psi^*}(x+h, x) \geq \mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)] \geq \mathbb{E}[J_{\psi^*}(x+h, B_{\tau^*}^x)]$$

with equality at $h = 0$ (because of (34)). This shows that $h \mapsto \mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)]$ is differentiable at $h = 0$ and

$$\left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)] = \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x+h, B_{\tau^*}^x)].$$

Moreover,

$$\mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)] \geq \mathbb{E}[\psi^*(B_\zeta^x) - c(x+h, B_\zeta^x)],$$

and from the assumption on ζ we have equality at $h = 0$. This verifies

$$\begin{aligned} \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[J_{\psi^*}(x+h, B_\zeta^x)] &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[\psi^*(B_\zeta^x) - c(x+h, B_\zeta^x)] \\ &= \left. \frac{d}{dh} \right|_{h=0} \mathbb{E}[-c(x+h, B_\zeta^x)] = \mathbb{E}[-\nabla_x c(x, B_\zeta^x)]. \end{aligned}$$

All together these complete the proof. \square

We now use the above lemmata to give the proof of Theorem 6.1.

PROOF OF THEOREM 6.1. Let π denote the probability measure on $\overline{O} \times \overline{O}$ corresponding to B_η , that is, $B_\eta^x \sim \pi_x$ for μ -a.e. x , where π_x is the disintegration $\pi(dx, dy) = \pi_x(dy)\mu(dx)$. Fix a pair (x, y) chosen π -a.e., in particular, for x to satisfy the results of

Lemma 6.4 with $\zeta = \eta$. Then, consider a small ball $V_\epsilon(y)$ of radius $\epsilon > 0$ around y . Define a stopping time ζ_ϵ as

$$\zeta_\epsilon = \begin{cases} \eta & \text{if } B_\eta^x \in V_\epsilon(y), \\ \tau^* & \text{otherwise.} \end{cases}$$

Notice that ζ_ϵ satisfies $\eta \leq \zeta_\epsilon \leq \tau^*$ and (41) (e.g., from Lemma 3.6(1)) so that we can apply Lemma 6.4. Then, (42) gives

$$\mathbb{E}[\nabla_x c(x, B_{\zeta_\epsilon}^x)] = \mathbb{E}[\nabla_x c(x, B_{\tau^*}^x)].$$

From this we see

$$\begin{aligned} 0 &= \frac{1}{\mathbb{P}[B_\eta^x \in V_\epsilon(y)]} [\mathbb{E}[\nabla_x c(x, B_{\tau^*}^x)] - \mathbb{E}[\nabla_x c(x, B_{\zeta_\epsilon}^x)]] \\ &= \mathbb{E}[\nabla_x c(x, B_{\tau^*}^x) | B_\eta^x \in V_\epsilon(y)] - \mathbb{E}[\nabla_x c(x, B_\eta^x) | B_\eta^x \in V_\epsilon(y)]. \end{aligned}$$

Letting $\epsilon \rightarrow 0$, we see that, for $\xi = \tau^* - \eta$ restricted to the paths where $y = B_\eta^x$,

$$\mathbb{E}[\nabla_x c(x, B_\xi^y)] - \nabla_x c(x, y) = 0 \quad \text{for } \pi\text{-a.e. } (x, y).$$

We apply the stochastic twist condition (ST) in Definition 5.1, and we get $\xi = 0$. Since this holds for π -a.e. (x, y) , this implies $\tau^* = \eta$, completing the proof. \square

7. The case of the distance function. We now consider the distance cost $c(x, y) = |x - y|$. We focus on the multidimensional case $d \geq 2$, because for the one-dimensional case ($d = 1$) our problem is equivalent to the martingale optimal transport and the uniqueness and structure of the optimal stopping is well known [5, 32–34]. We first get the following theorem as a corollary of Theorem 6.1, where we assume the strict separation assumption $\text{supp } \mu \cap \text{supp } \nu = \emptyset$ to ignore the singularity at $x = y$. Then, a localization argument allows us to remove this disjointness of supports in Theorem 7.2, where we show that there is a unique optimal randomized stopping time τ^* given by the hitting time of a barrier whenever $\tau^* > 0$, assuming that μ and ν have densities $f \in C(\overline{O})$ and $g \in C(\overline{O})$, respectively.

THEOREM 7.1. *Use the same assumptions as in Theorem 6.1 except that $c(x, y) = |x - y|$. Assume further that $d \geq 2$ and $\text{supp } \mu \cap \text{supp } \nu = \emptyset$. Then, the following hold:*

1. *There exists a constant D and $\psi^* \in \mathcal{B}_D$ such that (ψ^*, J_{ψ^*}) maximize the dual problem.*
2. *There is a unique optimal stopping time that is given by*

$$(43) \quad \eta = \inf\{t; J_{\psi^*}(B_0, B_t) = \psi^*(B_t) - |B_0 - B_t|\}.$$

PROOF. Let τ^* be an optimal stopping time for the cost $c(x, y) = |x - y|$. We let $\epsilon > 0$ be such that $|x - y| \geq \epsilon$ for all $x \in \text{supp}(\mu)$ and $y \in \text{supp}(\nu)$. Then, we consider a smooth subharmonic function $c^\epsilon(x, y) \leq |x - y|$ such that $c^\epsilon(x, y) = |x - y|$ whenever $|x - y| \geq \epsilon$. This can be easily constructed since for $d \geq 2$, $\Delta|x - y| > 0$ whenever $|x - y| \neq 0$. Let D be the constant with $0 \leq \Delta_y c^\epsilon(x, y) \leq D$.

First, observe that, by construction of c^ϵ (and the separation of $\text{supp } \mu$ and $\text{supp } \nu$ by ϵ), τ^* is also an optimal stopping time for the cost c^ϵ , and

$$\mathbb{E}[c^\epsilon(B_0, B_\tau)] = \mathbb{E}[|B_0 - B_\tau|].$$

We now consider the dual optimizers for the cost c^ϵ , namely, $(\psi^*, J_{\psi^*}^\epsilon)$ with $\psi^* \in \mathcal{B}_D$ obtained from Theorem 4.1. Here, $J_{\psi^*}^\epsilon$ is the value function with respect to c^ϵ , that is,

$$J_{\psi^*}^\epsilon(x, y) := \sup_{\sigma} \mathbb{E}[\psi^*(B_\sigma^y) - c^\epsilon(x, B_\sigma^y)].$$

Then, from Theorem 6.1, $\tau^* = \eta^\epsilon$, for η^ϵ given in (39) with respect to c^ϵ and $J_{\psi^*}^\epsilon$; this also proves uniqueness of τ^* .

On the other hand, notice that, because $c^\epsilon(x, y) \leq |x - y|$,

$$(44) \quad J_{\psi^*}(x, y) := \sup_{\sigma} \mathbb{E}[\psi^*(B_\sigma^y) - |x - B_\sigma^y|] \leq J_{\psi^*}^\epsilon(x, y).$$

Therefore,

$$\begin{aligned} \int_O \psi^*(y) \nu(dy) - \int_O J_{\psi^*}(x, x) \mu(dx) &\geq \int_O \psi^*(y) \nu(dy) - \int_O J_{\psi^*}^\epsilon(x, x) \mu(dx) \\ &= \mathbb{E}[c^\epsilon(B_0, B_{\tau^*})] = \mathbb{E}[|B_0 - B_{\tau^*}|]. \end{aligned}$$

This proves that the pair (ψ^*, J_{ψ^*}) is a dual optimizer for the cost $|x - y|$, and the above inequality is, in fact, an equality; thus, applying (44), we get

$$(45) \quad J_{\psi^*}(x, y) = J_{\psi^*}^\epsilon(x, y) \quad \text{for } \pi^*\text{-a.e. } (x, y),$$

where π^* is the optimal subharmonic martingale measure corresponding to τ^* . For μ -a.e. x , $y \mapsto J_{\psi^*}^\epsilon(x, y)$ is harmonic for y satisfying $|y - x| \leq \epsilon$ by Lemma 6.3, so (45) and superharmonicity of J_{ψ^*} , imply that $J_{\psi^*}(x, y) \geq J_{\psi^*}^\epsilon(x, y)$ for all y satisfying $|y - x| \leq \epsilon$. Then, we see that $\tau^* = \eta^\epsilon = \eta$ satisfies (43). This completes the proof. \square

THEOREM 7.2. *For $c(x, y) = |y - x|$ and $d \geq 2$, if $\mu \prec_{\text{SH}} \nu$ and μ and ν have densities $f \in C(\overline{O})$ and $g \in C(\overline{O})$, then there is a unique optimal stopping time τ^* that is randomized only at time 0. The optimal stopping time is given by $\tau^* = 0$ with density $g \wedge f$ and otherwise τ^* is the hitting time η ,*

$$\eta = \inf\{t > 0; (B_0, B_t) \in R\}$$

for some $R \subset \overline{O} \times \overline{O}$ measurable.

We will first show that the overlapping mass, if any, of the probability measures μ and ν stays put under any optimal solution τ^* . This was already shown in [24] by using the monotonicity principle of [3]. We shall give here a direct proof without using that principle.

LEMMA 7.3 (See [24]). *We use the assumptions and notation of Theorem 7.2, except we suppose that the cost function c satisfies following conditions:*

- c is continuous and $c(x, x) = 0, \forall x$;
- c satisfies the triangle inequality: $c(x, y) \leq c(x, z) + c(z, y), \forall x, y, z$, while equality holds only when y is on the unique geodesic (line segment in \mathbb{R}^d) connecting x and z .

Then, any optimal randomized stopping time τ^* stops at time 0 with density $f \wedge g$, that is,

$$\mathbb{E}[\mathbf{1}\{\tau^* = 0\}] = \int_O f(x) \wedge g(x) dx.$$

PROOF. Given an optimal randomized stopping time τ^* , we let h be the density it stops at time 0, that is,

$$\mathbb{E}[\mathbf{1}\{\tau^* = 0\}] = \int_O h(x) dx.$$

Notice that $h \leq f \wedge g$. We will prove the equality. In the following we use the convention in definitions that the value of a quotient is 1 if the denominator vanishes.

First, define a randomized stopping time σ from the initial distribution μ so that σ follows the stopping rule of τ^* with probability density $f - f \wedge g$ (of course, it is possible to stop at time 0 for τ^* with a certain density) and stops at time 0 otherwise. Namely:

- the initial distribution of σ is $B_0 \sim \mu$;
- σ is randomized at the initial point B_0 as

$$\sigma = \begin{cases} \tau^* & \text{with probability } 1 - \frac{f(B_0) \wedge g(B_0)}{f(B_0)}, \\ 0 & \text{with probability } \frac{f(B_0) \wedge g(B_0)}{f(B_0)}. \end{cases}$$

In particular, for any continuous function ϕ , we have

$$(46) \quad \mathbb{E}[\phi(B_\sigma)] = \int \mathbb{E}[\phi(B_{\tau^*}^x)](f(x) - f(x) \wedge g(x)) dx + \int \phi(x)f(x) \wedge g(x) dx.$$

Let $\hat{\mu}$ be the final distribution of σ , that is, $B_\sigma \sim \hat{\mu}$. As $B_{\tau^*} \sim \nu$ has density, by the construction of σ the distribution $\hat{\mu}$ also has density, say \hat{f} . Observe that $\hat{f} \geq f \wedge g$. We now define another randomized stopping time ξ from the initial distribution $\hat{\mu}$ so that it follows the stopping rule of τ^* with probability density $f \wedge g$ and, otherwise, it stops at time 0. More precisely:

- the initial distribution of ξ is $B_0 \sim \hat{\mu}$;
- ξ is randomized at the initial point B_0 as

$$\xi = \begin{cases} \tau^* & \text{with probability } \frac{f(B_0) \wedge g(B_0)}{\hat{f}(B_0)}, \\ 0 & \text{with probability } 1 - \frac{f(B_0) \wedge g(B_0)}{\hat{f}(B_0)}. \end{cases}$$

We verify that B_ξ has the same final distribution ν as B_{τ^*} . To see this, note that

$$(47) \quad \mathbb{E}[\phi(B_\xi)] = \int \mathbb{E}[\phi(B_{\tau^*}^x)]f(x) \wedge g(x) dx + \int \phi(x)(\hat{f}(x) - f(x) \wedge g(x)) dx.$$

Here, realizing $\int \phi(x)\hat{f}(x) dx = \mathbb{E}[\phi(B_\sigma)]$ and using (46), we see that this equation implies

$$\mathbb{E}[\phi(B_\xi)] = \int \mathbb{E}[\phi(B_{\tau^*}^x)]f(x) dx = \mathbb{E}[\phi(B_{\tau^*})],$$

verifying the claim $B_\xi \sim B_{\tau^*} \sim \nu$.

We now let τ be the randomized stopping time from the initial distribution μ that follows first σ then ξ . That is, τ is the randomized stopping time on the random paths obtained by concatenating the random paths following the stopping rule of σ with the random paths following that of ξ . We have $\sigma \leq \tau$ and that the final distribution of τ is the same as the final distribution of ξ . Therefore, $B_\tau \sim \nu$.

Observe, using the fact $c(x, x) = 0$ and (46), that

$$\mathbb{E}[c(B_0, B_\sigma)] = \int \mathbb{E}[c(x, B_{\tau^*}^x)](f(x) - f(x) \wedge g(x)) dx.$$

Similarly, from the construction of τ and (47),

$$\mathbb{E}[c(B_\sigma, B_\tau)] = \mathbb{E}[c(B_0, B_\xi)] = \int \mathbb{E}[c(x, B_{\tau^*}^x)]f(x) \wedge g(x) dx.$$

This shows that

$$(48) \quad \mathbb{E}[c(B_0, B_\sigma)] + \mathbb{E}[c(B_\sigma, B_\tau)] = \mathbb{E}[c(B_0, B_{\tau^*})].$$

On the other hand, from the optimality of τ^* (and the fact that τ^* and τ have the same initial and final distributions) and the triangle inequality of c , we have

$$\mathbb{E}[c(B_0, B_{\tau^*})] \leq \mathbb{E}[c(B_0, B_\tau)] \leq \mathbb{E}[c(B_0, B_\sigma)] + \mathbb{E}[c(B_\sigma, B_\tau)].$$

Here, all these inequalities become equalities due to (48). In particular, we have the equality case of the triangle inequality where B_σ is on the geodesic connecting B_0 and B_τ which holds only if

$$(49) \quad \sigma = 0 \quad \text{or} \quad \sigma = \tau \quad \text{almost surely.}$$

Otherwise, we would find that where $\sigma > 0$, $B_\tau = B_\sigma + \lambda(B_\sigma - B_0)$ for $\lambda > 0$, and hence

$$\mathbb{E}[(B_\tau - B_\sigma) \cdot (B_\sigma - B_0) | \mathcal{F}_\sigma] > 0$$

which cannot hold do to the martingale property of Brownian motion.

We will analyze (49) to draw our conclusion. For the random paths with $\sigma > 0$, (49) and the definition of σ implies $\sigma = \tau = \tau^*$ so the point B_σ lands at where $\xi = 0$. Note that $\xi = 0$ with probability

$$\frac{f(B_{\tau^*}) \wedge g(B_{\tau^*})}{\hat{f}(B_{\tau^*})} \frac{h(B_{\tau^*})}{f(B_{\tau^*})} + \left(1 - \frac{f(B_{\tau^*}) \wedge g(B_{\tau^*})}{\hat{f}(B_{\tau^*})}\right) = 1.$$

This implies that if $\tau^* > 0$, then $h(B_{\tau^*}) = f(B_{\tau^*})$. Therefore, whenever $f(x) > h(x)$, it must be the case that $h(x) = g(x)$ because $B_{\tau^*} = x$ only when $\tau^* = 0$. This then implies that $h(x) = f(x) \wedge g(x)$, completing the proof. \square

PROOF OF THEOREM 7.2. We first reduce the problem to the case where the measures μ and ν have disjoint supports, then apply Theorem 7.1. Indeed, from Lemma 7.3 we have that the optimal stopping time τ^* stops at time 0 with density $g \wedge f$. Thus, we only need to characterize τ^* when $\tau^* > 0$.

We now show that if $\tau^* > 0$, then τ^* is given by η which is the hitting time of a barrier. First, on the subset of the probability space where $\tau^* > 0$, τ^* is optimal for transporting the mass μ^+ with density $(f - g)^+$ to ν^+ with density $(g - f)^+$. These measures satisfy $\mu^+ \wedge \nu^+ = 0$ and $\mu^+(\partial \text{supp } \mu^+) = 0$.

We require one more step to reduce our setting to the case where the measures have strictly disjoint support so that we can apply Theorem 7.1. We introduce μ_ϵ^+ as the restriction of μ^+ to points where the distance to $\partial \text{supp } \mu^+$ is greater than ϵ . Then, we let ν_ϵ^+ be the stopping distribution of the restriction of τ^* to the initial distribution μ_ϵ^+ . Notice that τ^* is still optimal to this restriction. Applying Theorem 7.1, we have that this restricted problem has a unique optimal hitting time (equal to τ^*) given by (43). This defines the set R whenever $|y - x| \geq \epsilon$. Taking ϵ to zero, we get that $\tau^* = \eta$ whenever $\tau^* > 0$.

To prove that τ^* is unique, we suppose that τ_1 and τ_2 are both optimal randomized stopping times. By the argument above we have that both stop at $t = 0$ with density $g \wedge f$ and for $t > 0$ are given by the hitting times η_1 and η_2 . We form the randomized stopping time τ' that stops at τ_1 with probability $\frac{1}{2}$ and at τ_2 with probability $\frac{1}{2}$, that is, with Brownian martingale measure $\pi' = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2$. The same argument applies to show that if $\tau' > 0$, then $\tau' = \eta'$ which is the hitting time to a barrier. It follows that $\tau_1 = \tau_2$ because, otherwise, there is a finite probability of finding $(B_0, B_t) \in R_1$ but $(B_0, B_t) \notin R_2$ (or vice versa) which contradicts that τ' is the hitting time of a barrier. \square

REMARK 7.4. Notice that in Theorem 7.2, the condition $f, g \in C(\overline{O})$ can be relaxed. For example, it suffices that g belongs to $H^{-1}(O)$ outside the support of μ .

APPENDIX A: APPROXIMATIONS

We give a couple of approximation (via inf-convolution) results discussed in [14, 21, 38] which are used in Lemma 2.3 and Proposition 3.2. For $h \in \text{LSC}(\overline{O})$, we define the inf-convolution as

$$h_\epsilon(y) = \inf_{z \in \overline{O}} \left\{ h(z) + \frac{1}{2\epsilon} |y - z|^2 \right\}.$$

LEMMA A.1. *Given $h \in \text{LSC}(\overline{O})$, we have that h_ϵ is Lipschitz and semiconcave. If $\Delta h \leq D$ on O in the sense of viscosity and $\text{dist}(x, \partial O)^2 > 4\epsilon \|h\|_\infty$, then $\Delta h_\epsilon(x) \leq D$ in the sense of viscosity. Furthermore, h_ϵ has a distributional Hessian $\nabla^2 h_\epsilon$, that is, a matrix-valued measure and satisfies $\Delta h_\epsilon \leq D$ in the weak sense for the open set $\{y \in O \mid \text{dist}(y, \partial O)^2 > 4\epsilon \|h\|_\infty\}$.*

In particular, there is a sequence of functions $h_i \in C^\infty(\overline{O})$ such that $h_i(x) \leq h(x)$ for $x \in O$ and for each $x \in O$ and $\delta > 0$, there exists I such that for $i \geq I$ we have $\Delta h_i(x) \leq D$ and $h(x) - h_i(x) \leq \delta$.

PROOF. For each z , $y \mapsto h(z) + \frac{1}{2\epsilon} |y - z|^2$ is Lipschitz with constant $\epsilon^{-1} \text{diam} O$ and if we subtract $\frac{1}{2\epsilon} |y|^2$, it becomes concave. These properties are inherited by the infimum.

Suppose now that $\Delta h(y) \leq D$ in the sense of viscosity for all $y \in O$, and $\text{dist}(x, \partial O)^2 > 4\epsilon \|h\|_\infty$. Then, there is $z \in O$ such that $h_\epsilon(x) = h(z) + \frac{1}{2\epsilon} |x - z|^2$. Now suppose that w touches $h_\epsilon(x)$ from below at x . Then, we define

$$\phi(y) := w(x + y - z) - \frac{1}{2\epsilon} |x - z|^2.$$

Then, we have that

$$\phi(z) = w(x) - \frac{1}{2\epsilon} |x - z|^2 = h_\epsilon(x)$$

and

$$\phi(y) \leq h_\epsilon(x + y - z) - \frac{1}{2\epsilon} |x - z|^2 \leq h(y).$$

Thus, using the viscosity property $\Delta h \leq D$ of h , we have $\Delta \phi(z) \leq D$, and since $\Delta \phi(z) = \Delta w(x)$, this shows that $\Delta h_\epsilon(x) \leq D$ in the sense of viscosity.

It follows from semiconcavity that the distributional Hessian of h_ϵ is a matrix-valued measure. Furthermore, it decomposes as $\nabla^2 h_\epsilon = M + T$, where $M \leq 0$, $T \in L^\infty$ and $M \perp T$. If there was a point with density at x where $\text{tr}(T) > D$, we could construct a C^2 function near x , satisfying $D < \Delta w(x) < \text{tr}(T)(x)$ such that $w(x) = h_\epsilon(x)$ and $w(y) \leq h_\epsilon(y)$. Thus, for x satisfying $\text{dist}(x, \partial O)^2 > 4\epsilon \|h\|_\infty$, we have $\Delta h \leq D$ in the weak sense.

Finally, for the smooth approximation we can define $\hat{h}_\epsilon(x)$ by extending h_ϵ outside of the domain O and convolving with a smooth mollifier whose support shrinks as $\delta \rightarrow 0$. Note that $\Delta \hat{h}_\epsilon \leq D$. We can then subtract a small constant so that $\hat{h}_\epsilon \leq h_\epsilon$ but is converging uniformly as $\delta \rightarrow 0$ (from Lipschitz property of h_ϵ). The result on smooth approximation follows by choosing both ϵ and δ sufficiently small, where we use the fact that $h_\epsilon \leq h$ and from the lower semicontinuity of h , $\liminf_{\epsilon \rightarrow 0} h_\epsilon \geq h$ for each x . This completes the proof. \square

For functions in \mathcal{B}_D , the boundary condition gives a cleaner smooth approximation as follows:

LEMMA A.2. *There is an open-bounded convex set \tilde{O} (depending only on O and D) such that $O \subset \tilde{O}$ and for any $0 < \epsilon \leq 1$ and $h \in \mathcal{B}_D$, h_ϵ satisfies $h_\epsilon(x) \leq 0$ and $\Delta h_\epsilon(x) \leq D$ for $x \in \tilde{O}$ and $h_\epsilon(y) = 0$ for $y \in \partial \tilde{O}$. It follows that each $h \in \mathcal{B}_D$ can be approximated pointwise by smooth functions satisfying the properties above in some open-bounded convex set containing O .*

PROOF. For each $h \in \mathcal{B}_D$, we extend it to \mathbb{R}^d by giving zero value on $\mathbb{R}^d \setminus O$. Notice that this extended function h (using the same notation) still satisfies $\Delta h \leq D$ in \mathbb{R}^d in the sense of viscosity. Define $h_\epsilon(y) = \inf_{z \in \mathbb{R}^d} h(z) + \frac{1}{2\epsilon}|y - z|^2$. As h is bounded, we can apply the same reasoning as in the proof of Lemma A.1 to see that $\Delta h_\epsilon(x) \leq D$ for all $x \in \mathbb{R}^d$ in the sense of viscosity. Also, it clearly holds $h_\epsilon \leq h \leq 0$.

Consider a uniform bound M for functions on B_D , that is, $\sup_{x \in O} |u_O(x)|$, as in (25). From this, observe that, for y satisfying $\text{dist}(y, O)^2 \geq 4\epsilon M$, we have $h_\epsilon(y) = 0$. A further convolution by a smooth bump function allows us to approximate h_ϵ uniformly with smooth functions h_ϵ^i that also satisfy $\Delta h_\epsilon^i(x) \leq D$ and $h_\epsilon^i \leq 0$, with support in a fixed convex-bounded open domain \tilde{O} (independent of individual h) which contains O and $|y - z|^2 \geq 5M$ for all $y \in \partial \tilde{O}$ and $z \in O$. By also letting $\epsilon \rightarrow 0$, pointwise convergence of h_ϵ^i to h follows. This completes the proof. \square

APPENDIX B: A PROOF OF THE MONOTONICITY PRINCIPLE

As a simple application of our dual attainment (Theorems 4.1 and 4.7), we now provide an alternative proof of the following version of the monotonicity principle of Beiglböck, Cox and Huesmann [3] adapted to our setting.

THEOREM B.1 (See [3]). *With the same notation as in Theorems 4.1 and 4.7, in particular, we consider the disintegrated probability measure $\sigma(dx, dy) = \sigma_x(dy)\mu(dx)$ such that $0 \prec_{\text{SH}} \sigma_x \prec_{\text{SH}} \pi_x^*$ for each μ -a.e. x , and its corresponding randomized stopping time ξ (note that $\xi \leq \tau^*$).*

For π^ -a.e. (x, y) and σ -a.e. (x', y') , it holds that if $y = y'$, then the stopping time $\tau^* - \xi$ restricted to paths with $B_\xi = y$ satisfies*

$$c(x, y) + \mathbb{E}[c(x', B_{\tau^*-\xi}^y)] \leq c(x', y) + \mathbb{E}[c(x, B_{\tau^*-\xi}^y)].$$

PROOF. Notice that from Theorem 4.7 we have for π^* -a.e. (x, y) and σ -a.e. (x', y') ,

$$(50) \quad J_{\psi^*}(x, y) = \psi^*(y) - c(x, y),$$

$$(51) \quad \begin{aligned} J_{\psi^*}(x', B_{\tau^*-\xi}^{y'}) &= J_{\psi^*}(x', B_{\tau^*}^{x'}) \\ &= \psi^*(B_{\tau^*}^{x'}) - c(x', B_{\tau^*}^{x'}) = \psi^*(B_{\tau^*-\xi}^{y'}) - c(x', B_{\tau^*-\xi}^{y'}). \end{aligned}$$

From the dynamic programming formulation of J_{ψ^*} , we have

$$(52) \quad J_{\psi^*}(x, y') \geq \mathbb{E}[\psi^*(B_{\tau^*-\xi}^{y'}) - c(x, B_{\tau^*-\xi}^{y'})].$$

It also follows from (35) that

$$(53) \quad \mathbb{E}[J(x', B_{\tau^*-\xi}^{y'})] = J_{\psi^*}(x', y') \geq \psi^*(y') - c(x', y'),$$

where the first equality follows from (34) and the inequality from the dynamic programming principle. Taking the expectation in (51), we see that, under the condition $y = y'$, the left-hand sides of (50), (51) are equal to the left-hand sides of the inequalities (52) and (53). Now

subtract the sum of the two equations from the sum of the two inequalities and cancel the terms $\psi^*(y)$ and $\mathbb{E}[\psi^*(B_{\tau^*-\xi}^y)]$, to obtain

$$0 \geq -c(x', y) - \mathbb{E}[c(x, B_{\tau^*-\xi}^{y'})] + c(x, y) + \mathbb{E}[c(x', B_{\tau^*-\xi}^{y'})],$$

hence completing the proof. \square

REMARK B.2. The condition $y = y'$ together with $\sigma \leq \tau^*$ gives a version of the stop-go pair of [3]. Our proof is similar in spirit to that of [29] where weak duality is used, while we use the strong duality (dual attainment) under the additional assumption $0 \leq \Delta_y c(x, y) \leq D$ among others. Because of this last condition, our monotonicity result does not completely replace that of [3] for the distance cost $|x - y|$.

Acknowledgments. The first two authors are partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

- [1] AMBROSIO, L. (2003). Lecture notes on optimal transport problems. In *Mathematical Aspects of Evolving Interfaces (Funchal, 2000). Lecture Notes in Math.* **1812** 1–52. Springer, Berlin. MR2011032 https://doi.org/10.1007/978-3-540-39189-0_1
- [2] BAXTER, J. R. and CHACON, R. V. (1974). Potentials of stopped distributions. *Illinois J. Math.* **18** 649–656. MR0358960
- [3] BEIGLBÖCK, M., COX, A. M. G. and HUESMANN, M. (2017). Optimal transport and Skorokhod embedding. *Invent. Math.* **208** 327–400. MR3639595 <https://doi.org/10.1007/s00222-016-0692-2>
- [4] BEIGLBÖCK, M., HENRY-LABORDÈRE, P. and PENKNER, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.* **17** 477–501. MR3066985 <https://doi.org/10.1007/s00780-013-0205-8>
- [5] BEIGLBÖCK, M. and JUILLET, N. (2016). On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44** 42–106. MR3456332 <https://doi.org/10.1214/14-AOP966>
- [6] BEIGLBÖCK, M., LIM, T. and OBLÓJ, J. (2019). Dual attainment for the martingale transport problem. *Bernoulli* **25** 1640–1658. MR3961225 <https://doi.org/10.3150/17-BEJ1015>
- [7] BEIGLBÖCK, M., LIM, T. and OBLÓJ, J. (2019). Dual attainment for the martingale transport problem. *Bernoulli* **25** 1640–1658. MR3961225 <https://doi.org/10.3150/17-BEJ1015>
- [8] BEIGLBÖCK, M., NUTZ, M. and TOUZI, N. (2017). Complete duality for martingale optimal transport on the line. *Ann. Probab.* **45** 3038–3074. MR3706738 <https://doi.org/10.1214/16-AOP1131>
- [9] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163 <https://doi.org/10.1007/s002110050002>
- [10] BERNARD, P. and BUFFONI, B. (2007). Optimal mass transportation and Mather theory. *J. Eur. Math. Soc. (JEMS)* **9** 85–121. MR2283105 <https://doi.org/10.4171/JEMS/74>
- [11] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 <https://doi.org/10.1002/cpa.3160440402>
- [12] CAFFARELLI, L. A. (1996). Allocation maps with general cost functions. In *Partial Differential Equations and Applications. Lecture Notes in Pure and Applied Mathematics* **177** 29–35. Dekker, New York. MR1371577
- [13] CAFFARELLI, L. A., FELDMAN, M. and MCCANN, R. J. (2002). Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** 1–26. MR1862796 <https://doi.org/10.1090/S0894-0347-01-00376-9>
- [14] CRANDALL, M. G., ISHII, H. and LIONS, P.-L. (1992). User’s guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)* **27** 1–67. MR1118699 <https://doi.org/10.1090/S0273-0979-1992-00266-5>
- [15] DELLACHERIE, C. and MEYER, P.-A. (1978). *Probabilities and Potential. North-Holland Mathematics Studies* **29**. North-Holland, Amsterdam. MR0521810
- [16] DOLINSKY, Y. and SONER, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Related Fields* **160** 391–427. MR3256817 <https://doi.org/10.1007/s00440-013-0531-y>

- [17] DUBINS, L. E. (1968). On a theorem of Skorohod. *Ann. Math. Stat.* **39** 2094–2097. MR0234520 <https://doi.org/10.1214/aoms/1177698036>
- [18] EDGAR, G. A., MILLET, A. and SUCHESTON, L. (1982). On compactness and optimality of stopping times. In *Martingale Theory in Harmonic Analysis and Banach Spaces (Cleveland, Ohio, 1981)*. *Lecture Notes in Math.* **939** 36–61. Springer, Berlin. MR0668536
- [19] EVANS, L. C. and GANGBO, W. (1999). *Differential Equations Methods for the Monge–Kantorovich Mass Transfer Problem* **653**. Am. Math. Soc., Providence.
- [20] FALKNER, N. (1980). On Skorohod embedding in n -dimensional Brownian motion by means of natural stopping times. In *Seminar on Probability, XIV (Paris, 1978/1979) (French)*. *Lecture Notes in Math.* **784** 357–391. Springer, Berlin. MR0580142
- [21] FUKUDA, I., ISHII, H. and TSUTSUMI, M. (1993). Uniqueness of solutions to the Cauchy problem for $u_t - u\Delta u + \gamma|\nabla u|^2 = 0$. *Differential Integral Equations* **6** 1231–1252. MR1235190
- [22] GANGBO, W. (1995). Habilitation thesis, Université de Metz. Available at <http://people.math.gatech.edu/gangbo/publications/habilitation.pdf>.
- [23] GANGBO, W. and MCCANN, R. J. (1996). The geometry of optimal transportation. *Acta Math.* **177** 113–161. MR1440931 <https://doi.org/10.1007/BF02392620>
- [24] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2017). Optimal Brownian Stopping between radially symmetric marginals in general dimensions. ArXiv e-prints.
- [25] GHOUSSOUB, N., KIM, Y.-H. and LIM, T. (2019). Structure of optimal martingale transport plans in general dimensions. *Ann. Probab.* **47** 109–164. MR3909967 <https://doi.org/10.1214/18-AOP1258>
- [26] GHOUSSOUB, N., KIM, Y.-H. and PALMER, A. Z. (2018). Optimal transport with controlled dynamics and free end times. *SIAM J. Control Optim.* **56** 3239–3259. MR3853603 <https://doi.org/10.1137/18M1178037>
- [27] GHOUSSOUB, N., KIM, Y.-H. and PALMER, A. Z. (2019). PDE methods for optimal Skorokhod embeddings. *Calc. Var. Partial Differential Equations* **58** 113. MR3959935 <https://doi.org/10.1007/s00526-019-1563-7>
- [28] GRIGOR'YAN, A. (1999). Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds. *Bull. Amer. Math. Soc. (N.S.)* **36** 135–249. MR1659871 <https://doi.org/10.1090/S0273-0979-99-00776-4>
- [29] GUO, G., TAN, X. and TOUZI, N. (2016). On the monotonicity principle of optimal Skorokhod embedding problem. *SIAM J. Control Optim.* **54** 2478–2489. MR3549873 <https://doi.org/10.1137/15M1025268>
- [30] HENRY-LABORDÈRE, P. (2017). *Model-Free Hedging: A Martingale Optimal Transport Viewpoint*. *Chapman & Hall/CRC Financial Mathematics Series*. CRC Press, Boca Raton, FL. MR3699668
- [31] HENRY-LABORDÈRE, P. and TOUZI, N. (2016). An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.* **20** 635–668. MR3519164 <https://doi.org/10.1007/s00780-016-0299-x>
- [32] HOBSON, D. (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris–Princeton Lectures on Mathematical Finance 2010*. *Lecture Notes in Math.* **2003** 267–318. Springer, Berlin. MR2762363 https://doi.org/10.1007/978-3-642-14660-2_4
- [33] HOBSON, D. and KLIMMEK, M. (2012). Model-independent hedging strategies for variance swaps. *Finance Stoch.* **16** 611–649. MR2972236 <https://doi.org/10.1007/s00780-012-0190-3>
- [34] HOBSON, D. and KLIMMEK, M. (2015). Robust price bounds for the forward starting straddle. *Finance Stoch.* **19** 189–214. MR3292129 <https://doi.org/10.1007/s00780-014-0249-4>
- [35] KANTOROVITCH, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201. MR0009619
- [36] KINDERLEHRER, D. and STAMPACCHIA, G. (2000). *An Introduction to Variational Inequalities and Their Applications*. *Classics in Applied Mathematics* **31**. SIAM, Philadelphia, PA. MR1786735 <https://doi.org/10.1137/1.9780898719451>
- [37] LEVIN, V. (1999). Abstract cyclical monotonicity and Monge solutions for the general Monge–Kantorovich problem. *Set-Valued Anal.* **7** 7–32. MR1699061 <https://doi.org/10.1023/A:1008753021652>
- [38] LIONS, P.-L. (1983). Optimal control of diffusion processes and Hamilton–Jacobi–Bellman equations. II. Viscosity solutions and uniqueness. *Comm. Partial Differential Equations* **8** 1229–1276. MR0709162 <https://doi.org/10.1080/03605308308820301>
- [39] MA, X.-N., TRUDINGER, N. S. and WANG, X.-J. (2005). Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* **177** 151–183. MR2188047 <https://doi.org/10.1007/s00205-005-0362-9>
- [40] MONGE, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de L'Académie Royale des Sciences de Paris*.
- [41] OBLÓJ, J. (2004). The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1** 321–392.

- [42] PRATELLI, A. (2007). On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Ann. Inst. Henri Poincaré Probab. Stat.* **43** 1–13. MR2288266 <https://doi.org/10.1016/j.anihpb.2005.12.001>
- [43] ROOT, D. H. (1969). The existence of certain stopping times on Brownian motion. *Ann. Math. Stat.* **40** 715–718. MR0238394 <https://doi.org/10.1214/aoms/1177697749>
- [44] ROST, H. (1971). The stopping distributions of a Markov Process. *Invent. Math.* **14** 1–16. MR0346920 <https://doi.org/10.1007/BF01418740>
- [45] SKOROKHOD, A. V. (1965). *Studies in the Theory of Random Processes. Translated from the Russian by Scripta Technica, Inc.* Addison-Wesley, Reading, MA. MR0185620
- [46] STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423–439. MR0177430 <https://doi.org/10.1214/aoms/1177700153>
- [47] SYLVESTRE, L. (2015). Viscosity solutions of elliptic equations. *Lect. Notes Second Chicago Summer Sch. Anal.* Available at <http://math.uchicago.edu/~luis/preprints/viscosity-solutions.pdf>.
- [48] TRUDINGER, N. S. and WANG, X.-J. (2001). On the Monge mass transfer problem. *Calc. Var. Partial Differential Equations* **13** 19–31. MR1854255 <https://doi.org/10.1007/PL00009922>