SIAM J. CONTROL OPTIM.
Vol. 61, No. 3, pp. 1712–1736

VISCOSITY SOLUTIONS FOR OBSTACLE PROBLEMS ON WASSERSTEIN SPACE*

MEHDI TALBI†, NIZAR TOUZI†, AND JIANFENG ZHANG‡

Abstract. This paper is a continuation of our accompanying paper [M. Talbi, N. Touzi, and J. Zhang, *Dynamic Programming Equation for the Mean Field Optimal Stopping Problem*, https://arxiv.org/abs/2103.05736, 2021], where we characterized the mean field optimal stopping problem by an obstacle equation on the Wasserstein space of probability measures, provided that the value function is smooth. Our purpose here is to establish this characterization under weaker regularity requirements. We shall define a notion of viscosity solutions for such an equation and prove existence, stability, and the comparison principle.

Key words. mean field optimal stopping, obstacle problems, viscosity solutions

MSC codes. 60G40, 35Q89, 49N80, 49L25, 60H30

DOI. 10.1137/22M1488119

1. Introduction.

In our previous paper [26], we characterized the so-called mean field optimal stopping problem by a dynamic programming equation on the Wasserstein space, which we call an obstacle equation on Wasserstein space by analogy with the equation corresponding to the standard optimal stopping problem (see, e.g., El Karoui [10] or Shiryaev [25]). More precisely, we proved that the value function of our optimization problem is the unique solution of the obstacle equation on the Wasserstein space, provided it has $C^{1,2}$ regularity (in an appropriate sense). We note that, besides its obvious connection with multiple stopping problems over a large interacting particle system, this obstacle equation provides a convenient tool for many time inconsistent optimal stopping problems. We also remark that our mean field optimal stopping problem has a structure quite different from the mean field games of optimal stopping.

However, as in the case of the standard optimal stopping problems, one can rarely expect a classical solution for the obstacle equations. In particular, the infinite dimensionality of the space of measures makes the regularity requirement even harder to meet. Our goal of this paper is thus to develop a viscosity solution theory for the obstacle problem on the Wasserstein space, which as is well-known requires much weaker regularities.

There have been some serious efforts on viscosity solutions of nonlinear partial differential equations on the Wasserstein space. We first mention the paper by Cardaliaguet and Quincampoix [3], which considered a first order Hamilton–Jacobi–Isaacs equation on Wasserstein space arising from deterministic zero-sum games with

*Received by the editors April 1, 2022; accepted for publication (in revised form) February 10, 2023; published electronically June 21, 2023.
https://doi.org/10.1137/22M1488119
Funding: The first two authors received financial support from the Chaires FiME-FDD and Financial Risks of the Louis Bachelier Institute. The third author is supported in part by NSF grant DMS-1908665.
†CMAP, École polytechnique, 91128 Palaiseau Cedex, France (mehdi.talbi@polytechnique.edu, nizar.touzi@polytechnique.edu).
‡Department of Mathematics, University of Southern California, Los Angeles, CA 90089-2532 USA (jianfenz@usc.edu).

random initial conditions. The comparison principle for viscosity solutions was established by combining the doubling variables argument with Ekeland's variational principle. We may also mention the work of Gangbo, Nguyen, and Tudorascu [13] and Jimenez, Marigonda, and Quincampoix [14], who also define a notion of viscosity solutions for Hamilton–Jacobi equations by using subdifferentials. Another approach followed by several authors (see, e.g., Pham and Wei [22]) consists in exploiting Lions' idea [19] of lifting the functions on the Wasserstein space into functions on the Hilbert space of random variables and then using the existing viscosity theory on Hilbert spaces (see, e.g., Lions [16, 17, 18] and Fabbri, Gozzi, and Swiech [11]). More recently, Cosso et al. [6] defined viscosity solutions for Hamilton–Jacobi equations by requiring the global extrema on the Wasserstein space for the tangency property of the test functions.

In the context of mean field control problems in a path dependent setting, Wu and Zhang [27] proposed a notion of viscosity solutions for parabolic equations on the Wasserstein space, inspired by Ekren, Keller, Ren, Touzi, and Zhang [7, 8, 9, 23]. Note that the natural idea which consists in taking Wasserstein balls for the viscosity neighborhood (as in Carmona and Delarue [4]) leads to serious difficulties as the Wasserstein ball is in general not compact. Instead, [27] restricted the viscosity neighborhood of some point (t, μ) (where t is a time and μ a measure) to the compact set of all possible laws of the controlled state process starting from this point. Another remarkable work by Burzoni et al. [2], in the context of mean field control of jump-diffusions, restricted the viscosity neighborhood in another way, so as to guarantee compactness. They proved a comparison result by the doubling variables argument. To do this, they succeeded in constructing a smooth metric which serves as a test function, but unfortunately restricts the scope of the method to the case when the coefficients of the controlled dynamics do not depend on the space variable.

We shall follow the approach of [27]. We consider the joint law of $(X_{\tau\wedge t}, \mathbf{1}_{\{\tau\geq t\}})$ as the variable of the value function, where X is the state process and τ is the stopping time. As in [27] we define viscosity solutions by using the set of such laws over all stopping times τ. This neighborhood set of laws, for a given initial condition, is compact under Wasserstein distance and thus is desirable for the viscosity theory. We show that, under natural conditions, the value function of the mean field optimal stopping problem is indeed the unique viscosity solution of the corresponding obstacle equation on Wasserstein space. We shall also establish the stability and the comparison principle for the viscosity solutions. To prove the latter, one key ingredient is a smooth mollifier for continuous functions on the Wasserstein space, introduced by Mou and Zhang [20]. However, to obtain some uniform estimates of the smooth mollifier which are needed in our proof of comparison principle, as in [20] we require the coefficients to be Lipschitz continuous under the 1-Wasserstein distance, rather than the more natural 2-Wasserstein distance.

As applications of our viscosity theory, we invest several time inconsistent optimal stopping problems, including problems related to mean variance, probability distortion, and expected shortfall. By considering the law (instead of the value) of the stopped state process as the variable, we show that the value functions are indeed the unique viscosity solution to the corresponding obstacle equation on the Wasserstein space. Moreover, our results can be easily extended to the infinite horizon case.

The paper is organized as follows. In section 2, we present the mean field optimal stopping problem, the corresponding dynamic programming equation, and some of its elementary properties. Section 3 is the main section, where we propose our definition of viscosity solutions and prove the main results. Section 4 is devoted to several applications. Finally, we prove some technical results in the appendix.

Notation. We denote by $\mathcal{P}(\Omega,\mathcal{F})$ the set of probability measures on a measurable space $(\Omega,\mathcal{F})$, and $\mathcal{P}_2(\Omega,\mathcal{F}) := \{m \in \mathcal{P}(\Omega,\mathcal{F}) : \int_\Omega d(x_0,x)^2 m(dx) < \infty\}$ for some $x_0 \in \Omega$, where d is a metric on Ω. $\mathcal{P}_2(\Omega,\mathcal{F})$ is equipped with the corresponding 2-Wasserstein distance $\mathcal{W}_2$. When $(\Omega,\mathcal{F}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we simply denote them as $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_2(\mathbb{R}^d)$. For a random variable Z and a probability $\mathbb{P}$, we denote by $\mathbb{P}_Z := \mathbb{P}\circ Z^{-1}$ the law of Z under $\mathbb{P}$. For vectors $x, y \in \mathbb{R}^n$ and matrices $A, B \in \mathbb{R}^{n\times m}$, denote $x\cdot y := \sum_{i=1}^n x_i y_i$ and $A : B := \mathrm{tr}(AB^\top)$. We shall also write "USC" (resp., "LSC") for "upper (resp., lower) semicontinuous."

2. The obstacle problem on Wasserstein space.

2.1. Formulation.

Let $T < \infty$ be fixed and $\Omega := C^0([-1,T],\mathbb{R}^d) \times \mathbb{I}^0([-1,T])$ be the canonical space, where

- $C^0([-1,T],\mathbb{R}^d)$ is the set of continuous paths from $[-1,T]$ to $\mathbb{R}^d$, constant on $[-1,0)$;
- $\mathbb{I}^0([-1,T])$ is the set of nonincreasing and càdlàg maps from $[-1,T]$ to $\{0,1\}$, constant on $[-1,0)$, and ending with value 0 at T.

We equip Ω with the Skorokhod distance, under which it is a Polish space. Note that the choice of the extension to -1 is arbitrary; the extension of time to the left of the origin is only needed to allow for an immediate stop at time $t=0$.

We denote $Y := (X, I)$ the canonical process, with state space $\mathbf{S} := \mathbb{R}^d \times \{0,1\}$, its canonical filtration $\mathbb{F} = \mathbb{F}^Y = (\mathcal{F}_t)_{t\in[-1,T]}$, and the corresponding jump time of the survival process I:

$$\tau := \inf\{t \ge 0 : I_t = 0\} \ \text{ so that } I_t := I_{0-}\mathbf{1}_{t<\tau} \text{ for all } t \in [-1,T].$$

By the càdlàg property of I, τ is an $\mathbb{F}$-stopping time.

Let $(b,\sigma,f) : [0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbf{S}) \to \mathbb{R}^d\times\mathbb{R}^{d\times d}\times\mathbb{R}$ with σ taking values in nonnegative matrices, and $g : \mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$. In the following assumption, which will always be in force throughout the paper, $\mathcal{P}_2(\mathbf{S})$ is equipped with the $\mathcal{W}_2$ distance.

Assumption 2.1. (i) b, σ are continuous in t and uniformly Lipschitz continuous in (x,m).
(ii) f is Borel measurable and has quadratic growth in $x \in \mathbb{R}^d$, and

$$F(t,m) := \int_{\mathbb{R}^d} f(t,x,m)m(dx,1) \text{ is continuous on } [0,T]\times\mathcal{P}_2(\mathbf{S}). \tag{2.1}$$

(iii) g is USC and locally bounded and is extended to $\mathcal{P}_2(\mathbf{S})$ by $g(m) := g(m(\cdot,\{0,1\}))$.

Introduce the dynamic value function

$$V(t,m) := \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \Big\{ \int_t^T F(r,\mathbb{P}_{Y_r})dr + g(\mathbb{P}_{Y_T}) \Big\}, \quad (t,m)\in[0,T]\times\mathcal{P}_2(\mathbf{S}). \tag{2.2}$$

Here $\mathcal{P}(t,m)$ is the set of probability measures $\mathbb{P}$ on $(\Omega,\mathcal{F}_T)$ s.t. $\mathbb{P}_{Y_{t-}} = m$, the paths $s\in[-1,t)\to Y_s$ are constants, $\mathbb{P}$-a.s., and the processes

$$M_\cdot := X_\cdot - \int_t^\cdot b(r,X_r,\mathbb{P}_{Y_r})I_r dr \quad \text{and} \quad M_\cdot M_\cdot^\intercal - \int_t^\cdot \sigma^2(r,X_r,\mathbb{P}_{Y_r})I_r dr \tag{2.3}$$

are $\mathbb{P}$-martingales on $[t,T]$, that is, for some $\mathbb{P}$-Brownian motion $W^\mathbb{P}$,

$$X_s = X_t + \int_t^s b(r,X_r,\mathbb{P}_{Y_r})I_r dr + \int_t^s \sigma(r,X_r,\mathbb{P}_{Y_r})I_r dW_r^\mathbb{P},\ I_s = I_{t-}\mathbf{1}_{s<\tau},\ \mathbb{P}\text{-a.s.}$$

A special element of $\mathcal{P}(t,m)$ is $\bar{\mathbb{P}} = \bar{\mathbb{P}}^{t,m}$ under which X is unstopped. That is,

(2.4)
$$X_s = X_t + \int_t^s b(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dr + \int_t^s \sigma(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dW_r^{\bar{\mathbb{P}}},\ I_s = I_{t-}\mathbf{1}_{[t,T)}(s),\ \bar{\mathbb{P}}\text{-a.s.}$$

Note that $Y_\cdot = Y_{\cdot\wedge\tau}$, and in particular $Y_T = Y_\tau$, $\mathbb{P}$-a.s. Moreover, from the definition of F in (2.1), we have $\int_t^T F(r, \mathbb{P}_{Y_r}) dr = \mathbb{E}^{\mathbb{P}} \int_t^\tau f(r, X_r, \mathbb{P}_{Y_r}) dr$.

We recall from our first paper [26] that $\mathcal{P}(t,m)$ is compact under the Wasserstein distance $\mathcal{W}_2$, and thus existence holds for the mean field optimal stopping problem (2.2). Furthermore, we have the dynamic programming principle (DPP): for any $s \in [t,T]$,

$$\begin{aligned} V(t,m) &= \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \Big\{ \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}) \Big\} \\ &= \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \Big\{ \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_s}) \Big\}. \end{aligned} \tag{2.5}$$

2.2. Differential calculus. We next recall some differential calculus tools on the Wasserstein space. We say that a function $u : \mathcal{P}_2(\mathbf{S}) \to \mathbb{R}$ has a functional linear derivative $\delta_m u : \mathcal{P}_2(\mathbf{S}) \times \mathbf{S} \to \mathbb{R}$ if

$$u(m') - u(m) = \int_0^1 \int_{\mathbf{S}} \delta_m u(\lambda m' + (1-\lambda)m, y)(m' - m)(dy) d\lambda \text{ for all } m, m' \in \mathcal{P}_2(\mathbf{S}),$$

$\delta_m u$ is continuous for the product topology, with $\mathcal{P}_2(\mathbf{S})$ equipped with the 2-Wasserstein distance, and has quadratic growth in $x \in \mathbb{R}^d$, locally uniformly in $m \in \mathcal{P}_2(\mathbf{S})$, so as to guarantee integrability in the last expression. As in [26], we denote

$$\delta_m u_i(t,m,x) := \delta_m u(t,m,x,i) \text{ for } \quad i \in \{0,1\}, \qquad D_I u := \delta_m u_1 - \delta_m u_0, \tag{2.6}$$

and we introduce the measure flow generator of X

$$\begin{aligned} \mathbb{L}u(t,m) &:= \partial_t u(t,m) + \int_{\mathbb{R}^d} \mathcal{L}_x \delta_m u_1(t,m,x) m(dx,1), \\ \text{where} \quad \mathcal{L}_x \delta_m u_1 &:= b \cdot \partial_x \delta_m u_1 + \tfrac{1}{2}\sigma^2 : \partial^2_{xx} \delta_m u_1. \end{aligned} \tag{2.7}$$

Throughout this paper, we denote

$$\mathbf{Q}_t := [t,T) \times \mathcal{P}_2(\mathbf{S}), \text{ and } \overline{\mathbf{Q}}_t := [t,T] \times \mathcal{P}_2(\mathbf{S}),\ t \in [0,T).$$

DEFINITION 2.2. *Let $C_2^{1,2}(\overline{\mathbf{Q}}_t)$ be the set of functions $u : \overline{\mathbf{Q}}_t \to \mathbb{R}$ s.t.*
- *$\partial_t u, \delta_m u, \partial_x \delta_m u_1, \partial^2_{xx} \delta_m u_1$ exist and are continuous in all variables,*
- *$\partial^2_{xx} \delta_m u_1$ is bounded in x, locally uniformly in (t,m).*

The following Itô's formula is due to [26, section 3]: for any $u \in C_2^{1,2}(\overline{\mathbf{Q}}_0)$ and $\mathbb{P} \in \mathcal{P}(0,m)$,

$$\begin{aligned} u(T, m_{T-}) = u(0,m) \quad &+ \int_0^T \mathbb{L}u(s, m_s) ds \\ &+ \sum_{s \in J_{[0,T)}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] \\ &+ \mathbb{E}^{\mathbb{P}} \Big[\int_{J^c_{[0,T)}(\mathbf{m})} D_I u(s, m_s, X_s) dI_s \Big], \end{aligned} \tag{2.8}$$

where $\mathbf{m} := \{m_s := \mathbb{P}_{Y_s}\}_{s\in[-1,T]}$, $J_{\mathbb{T}}(\mathbf{m}) := \{s \in \mathbb{T} : m_s \neq m_{s-}\}$, for all subsets $\mathbb{T} \subset [0,T]$, and $J^c_{\mathbb{T}}(\mathbf{m})$ its complement set in $\mathbb{T}$.

2.3. The dynamic programming equation. Given two probability measures $m, m' \in \mathcal{P}_2(\mathbf{S})$, we say that $m' \preceq m$ if $m'(\cdot,1)$ is absolutely continuous w.r.t. $m(\cdot,1)$ with density bounded by 1, i.e.,

$$m'(dx,1) = p(x)m(dx,1), \text{ and } m'(dx,0) = [1-p(x)]m(dx,1) + m(dx,0), \tag{2.9}$$

for some measurable $p : \mathbb{R}^d \to [0,1]$. In other words, $m'(dx,1)$ is obtained from m by randomly stopping a proportion $1-p(x)$ of the surviving particles. In our context, $m_{t-} = \mathbb{P}_{(X_t, I_{t-})}$ and $m_t = \mathbb{P}_{(X_t, I_t)}$, with $\mathbb{P} \in \mathcal{P}(t,m)$, so that $m_t \preceq m_{t^-}$ with conditional transition probability $p(x) = p_t(x) := \mathbb{P}(I_t = 1 \mid X_t = x, I_{t-} = 1)$.

The following property (proved in Appendix A) will be used often in this paper.

LEMMA 2.3. *For an arbitrary $m \in \mathcal{P}_2(\mathbf{S})$,*
(i) *the set $\{m' : m' \preceq m\}$ is compact,*
(ii) *any compact subset $\mathcal{K}(m) \subset \{m' : m' \preceq m\}$ has a smallest element for $\preceq$, i.e., there exists $\bar{m} \in \mathcal{K}(m)$ such that for all $m' \in \mathcal{K}(m)$, we have that $m' \preceq \bar{m}$ implies that $m' = \bar{m}$.*

The dynamic programming equation corresponding to our mean field optimal stopping problem is the infinitesimal counterpart of the DPP (2.5) and is defined by

$$\min\Big\{ \min_{m' \in C_u(t,m)} \big[-(\mathbb{L}u + F)(t,m') \big],\ (\mathbb{D}_I u)_*(t,m) \Big\} = 0,\ (t,m) \in \mathbf{Q}_0, \tag{2.10}$$

with boundary condition $u|_{t=T} = g$. Here the function $(\mathbb{D}_I u)_*$ is the LSC envelope of

$$\mathbb{D}_I u : (t,m) \longmapsto \inf_{x \in \mathrm{Supp}(m(\cdot,1))} D_I u(t,m,x),$$

which is USC, but not continuous, in general, and the set

$$C_u(t,m) := \big\{ m' \preceq m : u(t,m') \geq u(t,m) \big\},\ (t,m) \in \mathbf{Q}_0,$$

indicates the set of positions in $\mathbf{Q}_0$ which improve u by stopping the corresponding particles.

For the purpose of the present paper, we note that this equation is slightly different from the obstacle equation introduced in our previous work [26]:

- if u is a classical solution of (2.10), then it is nondecreasing for $\preceq$ (see [26, Lemma 4.3]), and thus $C_u(t,m)$ is characterized by an equality, as in [26];

- despite the remaining differences, the two equations define the same solution, but this does not seem to have an immediate proof; we emphasize, however, that the equivalence between these two equations is a direct consequence of our uniqueness result in [26, Theorem 4.5] and the comparison Theorem 3.13 below.

Our objective in this paper is to develop a notion of viscosity solution for this equation which bypasses the strong regularity requirements of classical solutions. As usual, we start by introducing the notions of the sub- and supersolutions.

DEFINITION 2.4. *Let $u \in C_2^{1,2}(\overline{\mathbf{Q}}_0)$.*
(i) *u is a classical supersolution of* (2.10) *if*

$$\min\big\{ -(\mathbb{L}u + F),\ \mathbb{D}_I u \big\}(t,m) \geq 0 \quad \text{for all } (t,m) \in \mathbf{Q}_0. \tag{2.11}$$

(ii) *u is a classical subsolution of* (2.10) *if*

$$\min\big\{ -(\mathbb{L}u + F), (\mathbb{D}_I u)_* \big\}(t,m) \leq 0 \quad \text{for all } (t,m) \in \mathbf{Q}_0 \text{ s.t. } C_u(t,m) = \{m\}. \tag{2.12}$$

(iii) *u is a classical solution of* (2.10) *if it is a classical supersolution and subsolution.*

3. Viscosity solutions.

3.1. Definition and consistency. For $\delta > 0$ and $(t,m) \in \mathbf{Q}_0$, we introduce the neighborhood

$$\mathcal{N}_\delta(t,m) := \big\{(s,\tilde m) : s \in [t,t+\delta], \mathbb{P} \in \mathcal{P}(t,m),\ \tilde m \in \{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}\big\}.$$

Note that, as the closure of a càdlàg $\mathcal{P}_2(\mathbf{S})$-valued graph, $\mathcal{N}_\delta(t,m)$ is compact, by the compactnesses of $[t,t+\delta]$, $\mathcal{P}(t,m)$, and $\{(\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s})\}_{s\in[t,t+\delta]}$ for any $\mathbb{P} \in \mathcal{P}(t,m)$.

Definition 3.1. *Let $u : \overline{\mathbf{Q}}_0 \longrightarrow \mathbb{R}$. We say that u is $\mathcal{N}$-USC (resp., $\mathcal{N}$-LSC) if*

$$\begin{aligned} u(t,m) &\geq \limsup_{(s,\tilde m)\to(t,m)} u(s,\tilde m) \quad (resp.,\ u(t,m) \\ &\leq \liminf_{(s,\tilde m)\to(t,m)} u(s,\tilde m)) \qquad for\ all\ (t,m) \in \overline{\mathbf{Q}}_0, \end{aligned}$$

where the limits are sequences $(t_n,m_n) \to (t,m)$ with $(t_n,m_n) \in \mathcal{N}_{T-t}(t,m)$.

Note that the standard $\mathcal{W}_2$-semicontinuity implies the $\mathcal{N}$-semicontinuity. For a locally bounded function $u : \overline{\mathbf{Q}}_0 \longrightarrow \mathbb{R}$, we introduce its $\mathcal{N}$-LSC and $\mathcal{N}$-USC envelopes relatively to $\mathcal{P}(t,m)$, u_*, and u^*, respectively:

$$u_*(t,m) := \liminf_{(s,\tilde m)\to(t,m)} u(s,\tilde m), \quad u^*(t,m) := \limsup_{(s,\tilde m)\to(t,m)} u(s,\tilde m), \text{for all} (t,m) \in \overline{\mathbf{Q}}_0,$$

where the limits are taken on all sequences $\{t_n,m_n\}_{n\geq 1}$ converging to (t,m) with $(t_n,m_n) \in \mathcal{N}_{T-t}(t,m)$ for all n. We then introduce the sets of test functions

$$\overline{\mathcal{A}}u(t,m) := \Big\{\varphi \in C_2^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u_*)(t,m) = \max_{\mathcal{N}_\delta(t,m)} (\varphi - u_*) \text{ for some } \delta > 0\Big\},$$

$$\underline{\mathcal{A}}u(t,m) := \Big\{\varphi \in C_2^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u^*)(t,m) = \min_{\mathcal{N}_\delta(t,m)} (\varphi - u^*) \text{ for some } \delta > 0\Big\}.$$

Definition 3.2. *Let $u : \mathbf{Q}_0 \to \mathbb{R}$ be locally bounded.*

(i) *u is a viscosity supersolution of* (2.10) *if, for any $(t,m) \in \mathbf{Q}_0$,*

(3.1)
$$u_*(t,m) \geq u_*(t,m') \text{ for all } m' \preceq m \quad \text{and} \quad -(\mathbb{L}\varphi + F)(t,m) \geq 0 \text{ for all } \varphi \in \overline{\mathcal{A}}u(t,m).$$

(ii) *u is a viscosity subsolution of* (2.10) *if, for any $(t,m) \in \mathbf{Q}_0$ s.t. $C_{u^*}(t,m) = \{m\}$,*

$$\min\{-(\mathbb{L}\varphi + F), (\mathbb{D}_I\varphi)_*\}(t,m) \leq 0 \quad \text{for all } \varphi \in \underline{\mathcal{A}}u(t,m). \tag{3.2}$$

(iii) *u is a viscosity solution of* (2.10) *if it is a viscosity supersolution and subsolution.*

Remark 3.3. Without loss of generality (w.l.o.g.), we may assume that the maximum in the definition of $\overline{\mathcal{A}}u(t,m)$ is strict. Indeed, for $\varphi \in \overline{\mathcal{A}}u(t,m)$, we set

$$\tilde\varphi(s,\tilde m) := \varphi(s,\tilde m) - (s-t)^2 - \big(\tilde m(\mathbb{R}^d,1) - m(\mathbb{R}^d,1)\big)^2 \quad \text{for all } (s,\tilde m) \in \mathbf{Q}_t.$$

It is obvious that $\tilde\varphi \in C_2^{1,2}(\overline{\mathbf{Q}}_t)$. As $\tilde\varphi(s,\tilde m) = \tilde\varphi(t,m)$ if and only if $s = t$ and $\tilde m = m$ (since $\tilde m(\mathbb{R}^d,1) = m(\mathbb{R}^d,1)$, and observing that in this case $\tilde m = \mathbb{P}_{Y_t} \preceq m$ for some $\mathbb{P} \in \mathcal{P}(t,m)$), we deduce that $\tilde\varphi \in \overline{\mathcal{A}}u(t,m)$ and the maximum is strict. Moreover, simple computations show that $\partial_t\tilde\varphi(t,m) = \partial_t\varphi(t,m)$ and $\mathbb{L}\tilde\varphi(t,m) = \mathbb{L}\varphi(t,m)$. An analogous statement holds for $\underline{\mathcal{A}}u(t,m)$. ∎

Our first result shows the consistency between classical and viscosity solutions.

THEOREM 3.4. *Let $u \in C_2^{1,2}(\overline{\mathbf{Q}}_0)$. Then u is a classical sub- (resp., super-) solution of* (2.10) *if and only if it is a viscosity sub- (resp., super-) solution of* (2.10).

Proof. (i) Let $(t,m) \in \mathbf{Q}_0$. If u is a viscosity super/subsolution, then given its smoothness we have $u \in \overline{\mathcal{A}}u(t,m) \cap \underline{\mathcal{A}}u(t,m)$, and we immediately deduce that u is a classical super/subsolution. In particular, by [26, Lemma 4.3], u being nondecreasing for $\preceq$ implies that $D_I u \geq 0$.
(ii) Assume u is a classical supersolution of (2.10). By (2.11) we see that $D_I u \geq 0$; then by [26, Lemma 4.3] again we see that u is nondecreasing for $\preceq$. Now let $\varphi \in \overline{\mathcal{A}}u(t,m)$ with corresponding δ. Introduce $\psi := \varphi - u$ and let $\bar{\mathbb{P}} \in \mathcal{P}(t,m)$ be defined by (2.4) s.t. X is unstopped under $\bar{\mathbb{P}}$. By the definition of $\overline{\mathcal{A}}u(t,m)$, we have $\psi(t,m) \geq \psi(s, \bar{\mathbb{P}}_{Y_s})$ for all $s \in [t, t+\delta]$. Applying Itô's formula (2.8), since the jump terms are equal to zero under $\bar{\mathbb{P}}$, we obtain $-\frac{1}{\delta}\int_t^{t+\delta} \mathbb{L}\psi(s, \bar{\mathbb{P}}_{Y_s})ds \geq 0$. Sending $\delta \to 0$, by the continuity of $s \mapsto \bar{\mathbb{P}}_{Y_s}$ we have $-\mathbb{L}\psi(t,m) \geq 0$, hence $-(\mathbb{L}\varphi + F)(t,m) \geq -(\mathbb{L}u + F)(t,m) \geq 0$ by the supersolution property of u.

Assume now that u is a classical subsolution. Let $\varphi \in \underline{\mathcal{A}}u(t,m)$ with corresponding δ, and assume that $(\mathbb{D}_I)_*\varphi(t,m) > 0$ and $C_u(t,m) = \{m\}$. By definition of $\underline{\mathcal{A}}u(t,m)$, we have

$$[\varphi - u](t,m) \leq [\varphi - u](s, \mathbb{P}_{Y_s}) \quad \text{for all } s \in [t, t+\delta], \mathbb{P} \in \mathcal{P}(t,m). \tag{3.3}$$

Set $s = t$ in (3.3); then it follows from the arbitrariness of $\mathbb{P} \in \mathcal{P}(t,m)$ that $[\varphi - u](t,m) \leq [\varphi - u](t,m')$ for all $m' \preceq m$. Following the arguments of [26, Lemma 4.3], we deduce from above that $D_I[\varphi - u](t,m,\cdot) \leq 0$, and therefore $(\mathbb{D}_I u)_*(t,m) \geq (\mathbb{D}_I \varphi)_*(t,m) > 0$. The subsolution property of u then implies that $-(\mathbb{L}u + F)(t,m) \leq 0$. Using Itô's formula (2.8) under $\bar{\mathbb{P}}$ again on $[t, t+\delta]$, we get from (3.3) that $-(\mathbb{L}\varphi + F)(t,m) \leq -(\mathbb{L}u + F)(t,m) \leq 0$. ∎

3.2. Some regularity results. In this subsection, we present some regularity results which will be used in the rest of this section. Since our main focus is the viscosity properties, we postpone their proofs to Appendix A.

LEMMA 3.5. *Under Assumption* 2.1*, the value function V is USC under $\mathcal{W}_2$.*

THEOREM 3.6. (i) *Assume f and g are uniformly continuous in (t,x,m), under $\mathcal{W}_2$ for m; then V is continuous on $\overline{\mathbf{Q}}_0$, under $\mathcal{W}_2$ for m.*
(ii) *Assume further that b, σ are uniformly Lipschitz continuous in m under $\mathcal{W}_1$, and f, g are uniformly continuous in m under $\mathcal{W}_1$; then V is also continuous in m under $\mathcal{W}_1$.*

Even for the standard optimal stopping problems, one can hardly expect the value function to be smooth. We next establish a regularity result for the value function when X is unstopped. For $(t,m) \in \mathbf{Q}_0$, let $\bar{\mathbb{P}}^{t,m} \in \mathcal{P}(t,m)$ be as by (2.4), and define

$$U(t,m) := g(\bar{\mathbb{P}}^{t,m}_{Y_T}) + \int_t^T F(r, \bar{\mathbb{P}}^{t,m}_{Y_r})dr. \tag{3.4}$$

LEMMA 3.7. *For $\varphi = b, \sigma, f, g$, assume φ is continuous in t and $\partial_x\varphi, \delta_m\varphi, \partial_x\delta_m\varphi, \partial^2_{xx}\delta_m\varphi$ exist and are continuous and bounded and that, for $\varphi = b, \sigma$, all the derivatives of φ are Lipschitz up to order* 2*. Then $U \in C^{1,2}(\overline{\mathbf{Q}}_0)$ with bounded $\partial_x\delta_m U, \partial_{xx}\delta_m U$ and in particular $U \in C_2^{1,2}(\overline{\mathbf{Q}}_0)$. Moreover, if b, σ, f, g are uniformly Lipschitz continuous in m under $\mathcal{W}_1$ with a Lipschitz constant L, then U is uniformly Lipschitz continuous in m under $\mathcal{W}_1$ with a Lipschitz constant C_L.*

Finally, we introduce a smooth mollifier for functions on the Wasserstein space.

LEMMA 3.8. (i) *Let $U : \mathcal{P}_2(\mathbf{S}) \to \mathbb{R}$ be continuous. There exists $\{U_n\}_{n\geq 1}$ in $C^\infty(\mathcal{P}_2(\mathbf{S}))$ such that $\lim_{n\to\infty}\sup_{m\in\mathcal{M}}|U_n(m)-U(m)|=0$ for any compact set $\mathcal{M}\subset\mathcal{P}_2(\mathbf{S})$.* (ii) *Let $U : \mathcal{P}_1(\mathbf{S}) \to \mathbb{R}$ be continuous under $\mathcal{W}_1$. There exists $\{U_n\}_{n\geq 1}$ in $C^\infty(\mathcal{P}_2(\mathbf{S}))\cap C^0(\mathcal{P}_1(\mathbf{S}))$ such that $\lim_{n\to\infty}\sup_{m\in\mathcal{M}}|U_n(m)-U(m)|=0$ for any compact set $\mathcal{M}\subset\mathcal{P}_1(\mathbf{S})$.* (iii) *Assume further that U is Lipschitz continuous under $\mathcal{W}_1$; then we may choose $\{U_n\}_{n\geq 1}$ to be Lipschitz continuous under $\mathcal{W}_1$, uniformly in n.*

The mollifier is adopted from Mou and Zhang [20]. Note that the extension of the state space from $\mathbb{R}^d$ in [20] to $\mathbf{S}$ here is straightforward. We remark that if U is Lipschitz continuous under $\mathcal{W}_2$, in general the Lipschitz continuity of U_n under $\mathcal{W}_2$ is not uniform in n.

3.3. Viscosity property. We first need a simple lemma whose proof is postponed to Appendix A.

LEMMA 3.9. (i) *Let $v:\mathcal{P}_2(\mathbf{S})\longrightarrow\mathbb{R}$ be $\mathcal{N}$-LSC, and $m\in\mathcal{P}_2(\mathbf{S})$ s.t. $v(m)\geq v(m')$ for all $m'\preceq m$ with continuous conditional transition probability. Then $v(m)\geq v(m')$ for all $m'\preceq m$.*
(ii) *Let $\varphi\in C^0(\overline{\mathbf{Q}}_0,\mathbb{R})$ admit a continuous linear functional derivative $\delta_m\varphi$. Assume we have $(\mathbb{D}_I\varphi)_*(t,m)>0$ for some $(t,m)\in\mathbf{Q}_0$. Then φ is nondecreasing for $\preceq$ in a neighborhood of (t,m).*

THEOREM 3.10. *The value function V is a viscosity solution of* (2.10).

Proof. First, by Lemma 3.5, V inherits the local boundedness of g.
(i) We first verify the viscosity supersolution property. Fix (t,m) and $\varphi\in\overline{\mathcal{A}}V(t,m)$. We may assume w.l.o.g. that $[V_*-\varphi](t,m)=0$. Let $\delta>0$ and $(t_n,m_n)_{n\geq 1}\in\mathcal{N}_\delta(t,m)$ converging to (t,m) s.t. $V(t_n,m_n)\underset{n\to\infty}{\longrightarrow}V_*(t,m)$, and denote $\eta_n:=[V-\varphi](t_n,m_n)\geq 0$, as $V\geq V_*$. Thus, we have $\eta_n\underset{n\to\infty}{\longrightarrow}0$. By the DPP (2.5), we have

$$\begin{aligned}\eta_n+\varphi(t_n,m_n)=V(t_n,m_n)&\geq\int_{t_n}^{s_n}F(r,\bar{\mathbb{P}}^{m_n}_{Y_r})dr+V(s_n,\bar{\mathbb{P}}^{m_n}_{Y_{s_n}})\\&\geq\int_{t_n}^{s_n}F(r,\bar{\mathbb{P}}^{m_n}_{Y_r})dr+V_*(s_n,\bar{\mathbb{P}}^{m_n}_{Y_{s_n}})\geq\int_{t_n}^{s_n}F(r,\bar{\mathbb{P}}^{m_n}_{Y_r})dr+\varphi(s_n,\bar{\mathbb{P}}^{m_n}_{Y_{s_n}}),\end{aligned}$$

where $\bar{\mathbb{P}}^{m_n}:=\bar{\mathbb{P}}^{t_n,m_n}\in\mathcal{P}(t_n,m_n)$ is defined by (2.4) such that X is unstopped, and $s_n:=t_n+h_n$ with $h_n:=\sqrt{\eta_n}\vee n^{-1}$. Thus, by Itô's formula, the above gives $h_n+\frac{1}{h_n}\int_{t_n}^{s_n}-(\mathbb{L}\varphi+F)(r,\bar{\mathbb{P}}^{m_n}_{Y_r})dr\geq 0$. Send $n\to\infty$; since $h_n\longrightarrow 0$, we obtain $-(\mathbb{L}\varphi+F)(t,m)\geq 0$.

We now prove the remaining part of the supersolution property. Let $m'\preceq m$ with transition probability p. By Lemma 3.9(i), we may assume w.l.o.g. that p is continuous. For all $n\geq 1$, define $m'_n\preceq m_n$ as the measure obtained from m_n by applying the same p. Given the continuity of p and the compactness $\mathcal{N}_\delta(t,m)$, we see by (2.9) that $\mathcal{W}_2(m'_n,m')\underset{n\to\infty}{\longrightarrow}0$. Let $\bar{\mathbb{P}}^{m_n,m'_n}\in\mathcal{P}(t_n,m_n)$ be s.t. $\bar{\mathbb{P}}^{m_n,m'_n}_{Y_{t_n}}=m'_n$, and $I_s=I_{t_n}$, $\bar{\mathbb{P}}^{m_n,m'_n}$-a.s. for all $s\geq t_n$. By (2.5),

$$V(t_n,m_n)\geq\int_{t_n}^{s}F(r,\bar{\mathbb{P}}^{m_n,m'_n}_{Y_r})dr+V\big(s,\bar{\mathbb{P}}^{m_n,m'_n}_{Y_s}\big)\ \text{ for all }\quad s\geq t_n\text{ and }n\geq 1.\tag{3.5}$$

Taking $s=t_n$ and $\liminf_{n\to\infty}$ in (3.5), we obtain $V_*(t,m)\geq V_*(t,m')$ as $V(t_n,m_n)\to V_*(t,m)$.

(ii) We next verify the viscosity subsolution property. Let (t,m) and $\varphi \in \underline{\mathcal{A}}V(t,m)$ be s.t. $C_{V^*}(t,m)=\{m\}$ and $(\mathbb{D}_I\varphi)_*(t,m)>0$. We may assume w.l.o.g. that $[V^*-\varphi](t,m)=0$. Let $\delta>0$ and $(t_n,m_n)_{n\geq 1}\in\mathcal{N}_\delta(t,m)$ converging to (t,m) such that $V(t_n,m_n)\underset{n\to\infty}{\longrightarrow} V^*(t,m)$, and denote $-\eta_n:=[V-\varphi](t_n,m_n)\leq 0$. Thus $\eta_n\underset{n\to\infty}{\longrightarrow}0$. For $n\geq 1$, since g is USC and $\mathcal{P}(t_n,m_n)$ is compact, there exists $\mathbb{P}^{n,*}\in\mathcal{P}(t_n,m_n)$ s.t. $V(t_n,m_n)=\int_{t_n}^T F(r,\mathbb{P}^{n,*}_{Y_r})dr+g(\mathbb{P}^{n,*}_{Y_T})$. By DPP, we have

$$V(t_n,m_n)\geq\int_{t_n}^{s_n}F(r,\mathbb{P}^{n,*}_{Y_r})dr+V(s_n,\mathbb{P}^{n,*}_{Y_{s_n}})\geq\int_{t_n}^{T}F(r,\mathbb{P}^{n,*}_{Y_r})dr+g(\mathbb{P}^{n,*}_{Y_T}),$$

where $s_n:=t_n+h_n$ with $h_n:=\sqrt{\eta_n}\vee n^{-1}$, and thus,

$$V(t_n,m_n)=\int_{t_n}^{s_n}F(r,\mathbb{P}^{n,*}_{Y_r})dr+V(s_n,\mathbb{P}^{n,*}_{Y_{s_n}}). \tag{3.6}$$

Noting that $\mathbb{P}^{n,*}\in\mathcal{P}(t_n,m_n)\subset\mathcal{P}(t,m)$ for all n, and $\mathcal{P}(t,m)$ is compact, we may extract a subsequence (still denoted the same) s.t. $\mathbb{P}^{n,*}\underset{n\to\infty}{\longrightarrow}\mathbb{P}^*$ for some $\mathbb{P}^*\in\mathcal{P}(t,m)$. As the trajectories $r\mapsto\mathbb{P}^{n,*}_{Y_r}$ are càdlàg and $s_n\downarrow t$, this implies $\mathcal{W}_2(\mathbb{P}^{n,*}_{Y_{s_n}},m^*)\underset{n\to\infty}{\longrightarrow}0$, where $m^*:=\mathbb{P}^*_{Y_t}\preceq m$ as $\mathbb{P}^*\in\mathcal{P}(t,m)$. Thus, taking the $\limsup_{n\to\infty}$ in (3.6) and recalling $V(t_n,m_n)\to V^*(t,m)$, we have $V^*(t,m)\leq V^*(t,m^*)$. As $C_{V^*}(t,m)=\{m\}$, we obtain $m^*=m$. Moreover, (3.6) also implies that

$$-\eta_n+\varphi(t_n,m_n)\leq\int_{t_n}^{s_n}F(r,\mathbb{P}^{n,*}_{Y_r})dr+\varphi(s_n,\mathbb{P}^{n,*}_{Y_{s_n-}})\quad\text{for all } n\geq 1. \tag{3.7}$$

Let $\mathcal{B}_{\mathcal{W}_2}(m,\delta)$ denote the $\mathcal{W}_2$ ball centered in m, with radius δ. By Lemma 3.9(ii), the fact that $(\mathbb{D}_I\varphi)_*(t,m)>0$ implies that φ is (strictly) increasing for $\preceq$ on $[t,t+\delta)\times\mathcal{B}_{\mathcal{W}_2}(m,\delta)$ for a possibly smaller $\delta>0$. By convergence to (t,m), we have $\{\mathbb{P}^{n,*}_{Y_r},t\leq r\leq s_n\}\subset\mathcal{B}_{\mathcal{W}_2}(m,\delta)$ for n large. Then $D_I\varphi(r,\mathbb{P}^{n,*}_{Y_{r-}},\cdot)\geq 0$ and $\varphi(r,\mathbb{P}^{n,*}_{Y_{r-}})\geq\varphi(r,\mathbb{P}^{n,*}_{Y_r})$ for $t\leq r\leq s_n$. Using the fact that the trajectories are càdlàg, by applying Itô's formula on (3.7) we obtain $-(\mathbb{L}\varphi+F)(t,m)\leq 0$. ∎

3.4. Stability.

THEOREM 3.11. *Let $\{F_\varepsilon\}_{\varepsilon>0}$ be a family of functions from $\overline{\mathbf{Q}}_0$ to $\mathbb{R}$ such that $F_\varepsilon\underset{\varepsilon\to0}{\longrightarrow}F$ uniformly on compact subsets of $\overline{\mathbf{Q}}_0$, and let $\{u_\varepsilon\}_{\varepsilon>0}$ and $\{v_\varepsilon\}_{\varepsilon>0}$ be two families of viscosity subsolutions and supersolutions of* (2.10) *with F_ε instead of F, respectively. Assume that the following relaxed semilimits are finite:*

$$\overline{u}(t,m):=\limsup_{(\varepsilon,s,\tilde{m})\to(0,t,m)}u_\varepsilon(s,\tilde{m}),\ \text{and}\ \underline{v}(t,m):=\liminf_{(\varepsilon,s,\tilde{m})\to(0,t,m)}v_\varepsilon(s,\tilde{m}),(t,m)\in\overline{\mathbf{Q}}_0,$$

where the limits are sequences $(\varepsilon_n,t_n,m_n)\to(0,t,m)$ with $(t_n,m_n)\in\mathcal{N}_{T-t}(t,m)$. Then $\overline{u}$ (resp., $\underline{v}$) is an $\mathcal{N}$-USC (resp., $\mathcal{N}$-LSC) viscosity subsolution (resp., supersolution) of (2.10).

Proof. (i) We prove the stability of the supersolution first. Observe that we may assume w.l.o.g. that v_ε is $\mathcal{N}$-LSC as $\underline{v}(t,m)=\liminf_{(\varepsilon,s,\tilde{m})\to(0,t,m)}(v_\varepsilon)_*(s,\tilde{m})$. Also note that $\underline{v}$ is clearly $\mathcal{N}$-LSC in the sense of Definition 3.1.

Fix $(t,m)\in\mathbf{Q}_0$, and $\varphi\in\overline{\mathcal{A}}\underline{v}(t,m)$ with corresponding δ, and s.t. (t,m) is a strict maximizer of $\varphi-\underline{v}$ on $\mathcal{N}_\delta(t,m)$; see Remark 3.3. By definition, there exists a sequence $(\varepsilon_n,t_n,m_n)\to(0,t,m)$ s.t. $v_{\varepsilon_n}(t_n,m_n)\to\underline{v}(t,m)$. Note that $(t_n,m_n)\in\mathcal{N}_\delta(t,m)$ for

all n large; then we can find $\delta' < \delta$ s.t. $\mathcal{N}_{\delta'}(t_n, m_n) \subset \mathcal{N}_\delta(t,m)$. Let $(\hat{t}_n, \hat{m}_n)$ be a maximizer of $\varphi - v_{\varepsilon_n}$ on $\mathcal{N}_{\delta'}(t_n, m_n)$. We first note that

$$(\hat{t}_n, \hat{m}_n) \underset{n\to\infty}{\longrightarrow} (t,m). \tag{3.8}$$

Indeed, $(\hat{t}_n, \hat{m}_n) \in \mathcal{N}_{\delta'}(t_n, m_n) \subset \mathcal{N}_\delta(t,m)$ for all n. Thus, by compactness, there exists a subsequence (still named $\hat{m}_n$) converging to some $(\hat{t}, \hat{m}) \in \mathcal{N}_\delta(t,m)$. Observing that

$$\begin{aligned}[\varphi - \underline{v}](t,m) &= \lim_{n\to\infty} [\varphi - v_{\varepsilon_n}](t_n, m_n) \le \liminf_{n\to\infty} [\varphi - v_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \\ &\le \limsup_{n\to\infty} [\varphi - v_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \le [\varphi - \underline{v}](\hat{t}, \hat{m}),\end{aligned}$$

we conclude from the fact that (t,m) is a strict maximizer of $\varphi - \underline{v}$ on $\mathcal{N}_\delta(t,m)$ that $(\hat{t}, \hat{m}) = (t,m)$, and thus (3.8) holds true. Then, given that (t_n, m_n) and $(\hat{t}_n, \hat{m}_n)$ have the same limit, we have $\mathcal{N}_{\delta''}(\hat{t}_n, \hat{m}_n) \subset \mathcal{N}_{\delta'}(t_n, m_n)$ for some $\delta'' < \delta'$ and n large enough. Then, as $(\hat{t}_n, \hat{m}_n)$ is also a maximizer on $\mathcal{N}_{\delta''}(\hat{t}_n, \hat{m}_n)$, the supersolution property implies $-(\mathbb{L}\varphi + F_{\varepsilon_n})(\hat{t}_n, \hat{m}_n) \ge 0$ for n large enough, and we derive the first part of the supersolution property of $\underline{v}$ by sending $n \to \infty$.

We now prove that $\underline{v}$ is increasing for $\preceq$. By Lemma 3.9(i), it suffices to prove that $\underline{v}(t,m) \ge \underline{v}(t,m')$ for a given $m' \preceq m$ with continuous conditional transition probability p. We define for all n the measure $m'_n \preceq m_n$, obtained from m_n by applying p. As $\mathcal{W}_2(m_n, m) \underset{n\to\infty}{\longrightarrow} 0$ and p is continuous, similarly to the proof of Theorem 3.10, we see that $\mathcal{W}_2(m'_n, m') \underset{n\to\infty}{\longrightarrow} 0$. Moreover, by the supersolution property of v_{ε_n}, we have $v_{\varepsilon_n}(t_n, m_n) \ge v_{\varepsilon_n}(t_n, m'_n)$ for all $n \ge 1$, and we conclude by taking the lim inf that $\underline{v}(t,m) \ge \underline{v}(t,m')$, as the left-hand side of the inequality converges.

(ii) We now prove the stability of the subsolution. Similarly to (i), we may assume that $\{u_\varepsilon\}_{\varepsilon>0}$ is a family of $\mathcal{N}$-USC viscosity subsolutions of (2.10), and observe that $\overline{u}$ is clearly $\mathcal{N}$-USC. Let (t,m) and $\varphi \in \underline{\mathcal{A}}\overline{u}(t,m)$ be such that (t,m) is a strict local minimizer of $\varphi - \overline{u}$. Assume that $C_{\overline{u}}(t,m) = \{m\}$ and $(\mathbb{D}_I\varphi)_*(t,m) > 0$. Following the same argument as in the previous step, replacing maximizers with minimizers, we may construct $(\hat{t}_n, \hat{m}_n)$, converging to some $(\hat{t}, \hat{m})$, and satisfying the inequalities

$$[\varphi - \overline{u}](t,m) \ge \limsup_{n\to\infty} [\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \ge \liminf_{n\to\infty} [\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \ge [\varphi - \overline{u}](\hat{t}, \hat{m}).$$

By the strict minimum property of (t,m), this again implies that $(\hat{t}, \hat{m}) = (t,m)$, and $\lim_{n\to\infty} u_{\varepsilon_n}(\hat{t}_n, \hat{m}_n) = \overline{u}(t,m)$. By Lemma 2.3, we may now take

$$m^*_n \in \underset{\preceq}{\operatorname{argmin}}\, C_{u_{\varepsilon_n}}(\hat{t}_n, \hat{m}_n).$$

By compactness, there is a subsequence $\{m^*_n\}_{n\ge1}$ converging to some m^*. As $u_{\varepsilon_n}(\hat{t}_n, \hat{m}_n) \le u_{\varepsilon_n}(\hat{t}_n, m^*_n)$ for all n, by definition of $C_{u_{\varepsilon_n}}(\hat{t}_n, \hat{m}_n)$, taking the lim sup implies $\overline{u}(t,m) \le \overline{u}(t,m^*)$, hence $m^* = m$ as $C_{\overline{u}}(t,m) = \{m\}$. As $(\mathbb{D}_I\varphi)_*(t,m) > 0$, $\hat{m}_n$ and m^*_n are both in a neighborhood where φ is strictly increasing for n large enough, and thus $[\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \ge [\varphi - u_{\varepsilon_n}](\hat{t}_n, m^*_n)$, which implies equality by definition of $(\hat{t}_n, \hat{m}_n)$ and the fact that $(\hat{t}_n, m^*_n) \in \mathcal{N}_{\delta''}(\hat{t}_n, \hat{m}_n)$. Then $\varphi \in \underline{\mathcal{A}}u_{\varepsilon_n}(\hat{t}_n, m^*_n)$. As $C_{u_{\varepsilon_n}}(\hat{t}_n, m^*_n) = \{m^*_n\}$, the viscosity subsolution implies $-(\mathbb{L}\varphi + F_{\varepsilon_n})(\hat{t}_n, m^*_n) \le 0$ for n large enough, and we conclude by letting $n \longrightarrow \infty$. ∎

Remark 3.12. A natural extension of the stability result is to allow the perturbation of b and σ. However, this would change the definition of $\mathcal{P}(t,m)$ in (2.3), and therefore our viscosity neighborhoods $\mathcal{N}_\delta(t,m)$. Although we expect the stability property to remain true, this would require extending $\mathcal{P}(t,m)$ in some sense, which would go beyond the scope of the present paper.

3.5. Comparison.

THEOREM 3.13. (i) *Let u be an $\mathcal{N}$-USC viscosity subsolution of* (2.10) *satisfying $u|_{t=T} \le g$. Assume further that f is uniformly continuous in (t,x,m) under $\mathcal{W}_2$. Then $u \le V$.*
(ii) *Let v be an $\mathcal{N}$-LSC viscosity supersolution of* (2.10) *satisfying $v|_{t=T} \ge g$. Assume further that b, σ, f, g can be extended to $\mathcal{P}_1(\mathbf{S})$ under $\mathcal{W}_1$ continuously, b is uniformly Lipschitz continuous in (x,m) under $\mathcal{W}_1$, and σ has the regularity required in Lemma* 3.7. *Then $v \ge V$.*

Proof. (i) We first compare V and u. Assume by contradiction that $u(t,m) > V(t,m)$ for some (t,m). Then, for $\varepsilon > 0$ small enough,

$$u(t,m) - \varphi_\varepsilon(t,m) > \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \Big\{ \int_t^T F(r,\mathbb{P}_{Y_r})dr + g(\mathbb{P}_{Y_T}) \Big\}, \tag{3.9}$$

where $\varphi_\varepsilon(s,\tilde{m}) := \varepsilon[(T-t) + m(\mathbb{R}^d,1)]$. Let $(t^*,\mathbb{P}^*)$ be s.t.

$$\begin{aligned}(u-\varphi_\varepsilon)(t^*,m^*) + \int_t^{t^*} F(r,\mathbb{P}^*_{Y_r})dr \\ = \max_{\substack{(s,\,m,\,\mathbb{P})\,\in\,\mathcal{N}_{T-t}(t,\,m)\,\times\,\mathcal{P}(t,\,m):\\ m\,\in\,\{\mathbb{P}_{Y_{s-}},\,\mathbb{P}_{Y_s}\}}} \Big\{ (u-\varphi_\varepsilon)(s,m) + \int_t^s F(r,\mathbb{P}_{Y_r})dr \Big\},\end{aligned} \tag{3.10}$$

where m^* is the optimal argument in $\{\mathbb{P}^*_{Y_{t^*-}}, \mathbb{P}^*_{Y_{t^*}}\}$. Clearly $t^* < T$. Indeed, if $t^* = T$, then $(T,m^*) \in \mathcal{N}_{T-t}(t,m)$, and by (3.10) and (3.9) we have

$$\begin{aligned}u(T,m^*) - \varepsilon m^*(\mathbb{R}^d,1) + \int_t^T F(r,\mathbb{P}^*_{Y_r})dr \ge (u-\varphi_\varepsilon)(t,m) \\ > \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \Big\{ \int_t^T F(r,\mathbb{P}_{Y_r})dr + g(\mathbb{P}_{Y_T}) \Big\} \ge \int_t^T F(r,\mathbb{P}^*_{Y_r})dr + u(T,m^*),\end{aligned}$$

as $u(T,\cdot) \le g$. This is the desired contradiction. Moreover, by Lemma 2.3, we may choose m^* to be the smallest one which keeps the same value $(u-\varphi_\varepsilon)(t^*,m^*)$. Note that this change is only at t^* and thus has no impact on the value of $\int_t^{t^*} F(r,\mathbb{P}^*_{Y_r})dr$. Then

$$(u-\varphi_\varepsilon)(t^*,m^*) > (u-\varphi_\varepsilon)(t^*,m') \quad \text{for all } m^* \ne m' \preceq m^*. \tag{3.11}$$

Furthermore, we note that since $m \mapsto m(\mathbb{R}^d,1)$ is increasing, by (3.11) actually we have

$$u(t^*,m^*) > u(t^*,m') \quad \text{for all } \ m^* \ne m' \preceq m^*, \quad \text{namely} \quad C_u(t^*,m^*) = \{m^*\}. \tag{3.12}$$

Next, let f^+, f^- denote the positive and negative parts of f, respectively, and ρ_0 the modulus of continuity function of f. Introduce

$$\begin{aligned}\underline{f}^+(s,x,\tilde{m}) &:= f^+(s,x,\tilde{m}) - \rho_0\big(|\tilde{m}(\mathbb{R}^d,1) - m^*(\mathbb{R}^d,1)|^{\frac{1}{2}}\big);\\ \overline{f}^-(s,x,\tilde{m}) &:= f^-(s,x,\tilde{m}) + \rho_0\big(|\tilde{m}(\mathbb{R}^d,1) - m^*(\mathbb{R}^d,1)|^{\frac{1}{2}}\big).\end{aligned}$$

It is clear that $\underline{f}^+, \overline{f}^-$ are also uniformly continuous in $(s,x,\tilde{m})$ (under $\mathcal{W}_2$). For $\varepsilon > 0$, by Lemma 3.8(i) let $\underline{f}^+_n, \overline{f}^-_n$ be a smooth mollifier (under $\mathcal{W}_2$) such that

$$|\underline{f}^+_n - \underline{f}^+| \le \frac{\varepsilon}{6}, \quad |\overline{f}^-_n - \overline{f}^-| \le \frac{\varepsilon}{6} \quad \text{on} \quad \mathcal{P}(t^*,m^*).$$

Then, for all $(s,\tilde m)\in\mathcal{N}_{T-t^*}(t^*,m^*)$ with corresponding $\mathbb{P}\in\mathcal{P}(t^*,m^*)$, and $t^*\le r<s$, considering the case $\tilde m=\mathbb{P}_{Y_{s-}}$, we have

$$\mathbb{E}^{\mathbb{P}}\Big[\Big|f^+(s,X_s,\mathbb{P}_{(X_s,I_r)})I_{s-}-f^+(s,X_s,\tilde m)I_{s-}\Big|\Big]\le\rho_0\Big(\mathcal{W}_2\big(\mathbb{P}_{(X_s,I_r)},\mathbb{P}_{(X_s,I_{s-})}\big)\Big)$$
$$\le\rho_0\Big(\sqrt{\mathbb{E}^{\mathbb{P}}[|I_r-I_{s-}|^2]}\Big)\le\rho_0\Big(\sqrt{\mathbb{E}^{\mathbb{P}}[|I_{t^*}-I_{s-}|^2]}\Big)=\rho_0\big(|\tilde m(\mathbb{R}^d,1)-m^*(\mathbb{R}^d,1)|^{\frac12}\big).$$

Then $\mathbb{E}^{\mathbb{P}}[f^+(s,X_s,\mathbb{P}_{(X_s,I_r)})I_{s-})]\ge\mathbb{E}^{\mathbb{P}}[\underline{f}^+(s,X_s,\tilde m)I_{s-}]$ and, similarly,$\mathbb{E}^{\mathbb{P}}[f^-(t^*,X_{t^*},\mathbb{P}_{(X_s,I_r)})I_{t^*}]\le\mathbb{E}^{\mathbb{P}}[\overline{f}^-(t^*,X_{t^*},\tilde m)I_{t^*}]$. Thus, by (2.1) and the regularity of f, we have

$$\begin{aligned}
&F(r,\mathbb{P}_{Y_r})=\mathbb{E}^{\mathbb{P}}\Big[f(r,X_r,\mathbb{P}_{Y_r})I_r\Big]=\mathbb{E}^{\mathbb{P}}\Big[f^+(r,X_r,\mathbb{P}_{Y_r})I_r-f^-(r,X_r,\mathbb{P}_{Y_r})I_r\Big]\\
&\ge\mathbb{E}^{\mathbb{P}}\Big[f^+(r,X_r,\mathbb{P}_{Y_r})I_{s-}-f^-(r,X_r,\mathbb{P}_{Y_r})I_{t^*}\Big]\\
&\ge\mathbb{E}^{\mathbb{P}}\Big[f^+(s,X_s,\mathbb{P}_{(X_s,I_r)}))I_{s-}-f^-(t^*,X_{t^*},\mathbb{P}_{(X_s,I_r)})I_{t^*}\Big]-\rho(s-t^*)\\
&\ge\mathbb{E}^{\mathbb{P}}\Big[\underline{f}^+(s,X_s,\tilde m)I_{s-}-\overline{f}^-(t^*,X_{t^*},\tilde m)I_{t^*}\Big]-\rho(s-t^*)\\
&\ge\mathbb{E}^{\mathbb{P}}\Big[\underline{f}^+_n(s,X_s,\tilde m)I_{s-}-\overline{f}^-_n(t^*,X_{t^*},\tilde m)I_{t^*}\Big]-\frac{\varepsilon}{3}-\rho(s-t^*)
\end{aligned}$$

for some modulus of continuity ρ which can be chosen to be smooth on $(0,\infty)$. That is,

$$F(r,\mathbb{P}_{Y_r})\ge\int\underline{f}^+_n(s,x,\tilde m)i\,\tilde m(dx,di)-\int\overline{f}^-_n(t^*,x,\tilde m)i\;m^*(dx,di)-\frac{\varepsilon}{3}-\rho(s-t^*).$$

In the case $\tilde m=\mathbb{P}_{Y_s}$, following similar arguments we see the above still holds true. Denote

$$\begin{aligned}
\phi^n_\varepsilon(s,\tilde m)&:=\int f^+_n(s,x,\tilde m)i\;\tilde m(dx,di)-\int f^-_n(t^*,x,\tilde m)i\;m^*(dx,di),\\
\psi^n_\varepsilon(s,\tilde m)&:=\varphi_\varepsilon(s,\tilde m)-(s-t^*)\Big[\phi^n_\varepsilon(s,\tilde m)-\frac{\varepsilon}{3}-\rho(s-t^*)\Big],
\end{aligned}$$

which are obviously in $C^{1,2}_2(\overline{\mathbf{Q}}_0)$. Then, by (3.10),

$$\begin{aligned}
&(u-\psi^n_\varepsilon)(t^*,m^*)=(u-\varphi_\varepsilon)(t^*,m^*)\ge(u-\varphi_\varepsilon)(s,\tilde m)+\int_{t^*}^s F(r,\mathbb{P}_{Y_r})dr\\
&\ge(u-\varphi_\varepsilon)(s,\tilde m)+(s-t^*)\Big[\phi^n_\varepsilon(s,\tilde m)-\frac{\varepsilon}{3}-\rho(s-t^*)\Big]=(u-\psi^n_\varepsilon)(s,\tilde m).
\end{aligned}$$

Thus $\psi^n_\varepsilon\in\underline{\mathcal{A}}u(t^*,m^*)$. Note that

$$\begin{aligned}
&[\mathbb{L}\psi^n_\varepsilon+F](t^*,m^*)=\Big[\mathbb{L}\varphi_\varepsilon-\phi^n_\varepsilon+\frac{\varepsilon}{3}+F\Big](t^*,m^*)=-\varepsilon-\phi^n_\varepsilon(t^*,m^*)+\frac{\varepsilon}{3}+F(t^*,m^*)\\
&\quad=-\frac{2\varepsilon}{3}+\int\big[\underline{f}^+_n(t^*,x,m^*)-\overline{f}^-_n(t^*,x,m^*)\big]i\;m^*(dx,di)+F(t^*,m^*)\\
&\quad\le-\frac{\varepsilon}{3}+\int\big[\underline{f}^+(t^*,x,m^*)-\overline{f}^-(t^*,x,m^*)\big]i\;m^*(dx,di)+F(t^*,m^*)\\
&\quad\le-\frac{\varepsilon}{3}+\int\big[f^+(t^*,x,m^*)-f^-(t^*,x,m^*)\big]i\;m^*(dx,di)+F(t^*,m^*)=-\frac{\varepsilon}{3}<0;\\
&(D_I\psi^n_\varepsilon)(s,\tilde m)=\varepsilon-(s-t^*)(D_I\phi^n_\varepsilon)(s,\tilde m),\quad\text{and thus}\quad(D_I\psi^n_\varepsilon)_*(t^*,m^*)=\varepsilon>0.
\end{aligned}$$

Recalling (3.12), this contradicts the viscosity subsolution property of u.

(ii) We next compare V and v. Fix $\varepsilon > 0$. For each $n \geq 1$, denote $t_i := t_i^{(n)} := \frac{iT}{n}$, $0 \leq i \leq n$. First, note that, for $(t,m) \in \mathbf{Q}_0$, it follows from the continuity of the coefficients that

$$V(t,m) := \lim_{n\to\infty} V_n(t,m), \quad \text{where}$$
$$V_n(t,m) := \sup_{\mathbb{P}\in\mathcal{P}_n(t,m)} \Big\{ \int_t^T F(r,\mathbb{P}_{Y_r})dr + g(\mathbb{P}_{Y_T}) \Big\},$$
$$\mathcal{P}_n(t,m) := \Big\{ \mathbb{P} \in \mathcal{P}(t,m) : \tau \text{ takes values in } \{t_1,\dots,t_n\} \cap [t,T], \mathbb{P}\text{-a.s.} \Big\}.$$

Step 1: We show that $(V_n - v)(t_{n-1},\cdot) \leq \frac{\varepsilon}{n}$. Assume to the contrary that there exists m_{n-1} such that $(V_n - v)(t_{n-1},m_{n-1}) > \frac{\varepsilon}{n}$. By the definition of $\mathcal{P}_n(t_{n-1},m_{n-1})$, we have $V_n(t,m) = \int_t^T F(r,\bar{\mathbb{P}}^{t,m}_{Y_r})dr + g(\bar{\mathbb{P}}^{t,m}_{Y_T})$, $t \in (t_{n-1},T]$, where $\bar{\mathbb{P}}^{t,m} \in \mathcal{P}(t,m)$ is defined by (2.4).

Let $\delta_1,\delta_2 > 0$ be small numbers which will be specified later. Applying Lemma 3.8(ii), (iii), let (g_k,f_k,b_k) be the smooth mollifier of (g,f,b) (under $\mathcal{W}_1$), where b_k is also mollified in (t,x) in a standard way, such that $\|g_k - g\|_\infty + \|f_k - f\|_\infty \leq \delta_1$, $\|b_k - b\|_\infty \leq \delta_2$, and g_k is Lipschitz continuous under $\mathcal{W}_1$ with a Lipschitz constant L_k depending on k, and b_k is uniformly Lipschitz continuous in (x,m) under $\mathcal{W}_1$ with a Lipschitz constant L independent of k. By otherwise choosing a larger L we assume σ is also uniformly Lipschitz continuous in (x,m) under $\mathcal{W}_1$ with Lipschitz constant L. Let U^{k_1,k_2} be defined by (3.4) corresponding to $(b_{k_2},\sigma,g_{k_1},f_{k_1})$. Then, by Lemma 3.7,

(3.13)
$$\partial_t U^{k_1,k_2}(t,m) + \int_{\mathbb{R}^d} \Big[b_{k_2}\cdot\partial_x\delta_m U_1^{k_1,k_2} + \frac{1}{2}\sigma^2 : \partial^2_{xx}\delta_m U_1^{k_1,k_2} + f_{k_1} \Big] m(dx,1) = 0,$$

and U^{k_1,k_2} is Lipschitz continuous in m under $\mathcal{W}_1$ with a Lipschitz constant $C_{L,L_{k_1}}$ independent of k_2. Here, $\delta_m U_1^{k_1,k_2}$ is in the sense of (2.6). This, in particular, implies $|\partial_x\delta_m U_1^{k_1,k_2}| \leq C_{L,L_{k_1}}$ for all k_2. Then, we deduce from (3.13) that

(3.14)
$$\Big|(\mathbb{L}U^{k_1,k_2} + F_{k_1})(t,m)\Big| = \Big| \int (b - b_{k_2})\cdot\partial_x\delta_m U_1^{k_1,k_2} m(dx,1) \Big| \leq C_{L,L_{k_1}}\delta_2 \text{ for all } k_2 \geq 1,$$

where $F_{k_1}(t,m) := \int_{\mathbb{R}^d} f_{k_1}(t,x,m)m(dx,1)$ as in (2.1). Moreover, since

$$U^{k_1,k_2}(t,m) = g_{k_1}(\bar{\mathbb{P}}^{t,m,k_2}_{Y_T}) + \int_t^T F_{k_1}(r,\bar{\mathbb{P}}^{t,m,k_2}_{Y_r})dr,$$

where $\bar{\mathbb{P}}^{t,m,k_2}$ is s.t. X is unstopped with drift coefficient b_{k_2} instead of b, one can easily show that

$$\left| U^{k_1,k_2}(t,m) - \left(g(\bar{\mathbb{P}}^{t,m}_{Y_T}) + \int_t^T F(r,\bar{\mathbb{P}}^{t,m}_{Y_r})dr \right) \right| \leq C[\delta_1+\delta_2] \leq \frac{\varepsilon}{4n}$$

for δ_1,δ_2 small enough. Then

$$\begin{aligned} V_n(t_{n-1},m_{n-1}) &= \sup_{m'\preceq m_{n-1}} \Big\{ \int_{t_{n-1}}^T F(r,\bar{\mathbb{P}}^{t_{n-1},m'}_{Y_r})dr + g(\bar{\mathbb{P}}^{t_{n-1},m'}_{Y_T}) \Big\} \\ &\leq \sup_{m'\preceq m_{n-1}} U^{k_1,k_2}(t_{n-1},m') + \frac{\varepsilon}{4n}. \end{aligned}$$

By the supersolution property, v is nondecreasing for $\preceq$, hence

$$\frac{\varepsilon}{n} \le (V_n - v)(t_{n-1}, m_{n-1}) \le \sup_{m' \preceq m_{n-1}} (U_{k_1,k_2} - v)(t_{n-1}, m') + \frac{\varepsilon}{4n}.$$

This implies that

$$\max_{(s,\tilde{m}) \in \mathcal{N}_{\frac{T}{n}}(t_{n-1}, m_{n-1})} \left\{ (U^{k_1,k_2} - v)(s,\tilde{m}) - \frac{T-s}{n} \right\} \ge \frac{3\varepsilon}{4n} - \frac{T}{n^2} \ge \frac{\varepsilon}{2n} \tag{3.15}$$

for n sufficiently large. Note that $(U^{k_1,k_2} - v)(T, \mathbb{P}_{Y_T}) \le (g_{k_1} - g)(\mathbb{P}_{Y_T}) \le \frac{\varepsilon}{4n}$ for all $\mathbb{P} \in \mathcal{P}(t_{n-1}, m_{n-1})$ and v is $\mathcal{N}$-LSC; then by compactness of $\mathcal{N}_{\frac{T}{n}}(t_{n-1}, m_{n-1})$ there exists an optimal argument (t^*, m^*), $t^* < T$, to the above maximum. Thus $\varphi(s,\tilde{m}) := U^{k_1,k_2}(s,\tilde{m}) - \frac{T-s}{n} \in \overline{\mathcal{A}} v(t^*, m^*)$, and therefore,

$$\begin{aligned} 0 &\le -(\mathbb{L}\varphi + F)(t^*, m^*) = -(\mathbb{L}U^{k_1,k_2} + F_{k_1})(t^*, m^*) + (F_{k_1} - F)(t^*, m^*) - \frac{1}{n} \\ &\le C_{L,L_{k_1}} \delta_2 + (F_{k_1} - F)(t^*, m^*) - \frac{1}{n}, \end{aligned}$$

where the last inequality is thanks to (3.14). Fixing k_1 so that $(F_{k_1} - F)(t^*, m^*) \le \frac{1}{2n}$ and setting δ_2 small enough, we obtain the desired contradiction.

Step 2: We show that $(V_n - v)(t_{n-2}, \cdot) \le \frac{2\varepsilon}{n}$. Assume to the contrary that there exists m_{n-2} such that $(V_n - v)(t_{n-2}, m_{n-2}) > \frac{2\varepsilon}{n}$. By the DPP, we have

$$V_n(t_{n-2}, m_{n-2}) = \sup_{\mathbb{P} \in \mathcal{P}_n(t_{n-2}, m_{n-2})} \left\{ \int_{t_{n-2}}^{t_{n-1}} F(r, \mathbb{P}_{Y_r}) dr + V_n(t_{n-1}, \mathbb{P}_{Y_{(t_{n-1})-}}) \right\},$$

Observe the fact that v being a viscosity supersolution of (2.10) also implies that $v + \frac{\varepsilon}{n}$ is a viscosity supersolution. Moreover, by Step 1, we have $(v + \frac{\varepsilon}{n})(t_{n-1}, \cdot) \ge V_n(t_{n-1}, \cdot)$. Thus, using the same procedure as in Step 1 (where V_n replaces g on $(t_{n-2}, t_{n-1}]$), it follows that

$$\left(V_n - \left(v + \frac{\varepsilon}{n} \right) \right)(t_{n-2}, \cdot) \le \frac{\varepsilon}{n}.$$

Finally, by backward induction, we have $(V_n - v)(t_{n-j}, \cdot) \le \frac{j\varepsilon}{n}$ for all $j \in \{0, \dots, n\}$, and thus $(V_n - v)(t, \cdot) \le \varepsilon$, which implies by the arbitrariness of n and ε that $v \ge V$. □

3.6. Infinite horizon case. As in [26, section 6.1], we may formulate the problem in infinite horizon (i.e., in the case $T = \infty$) by replacing Assumption 2.1 with the following conditions.

Assumption 3.14. (i) Assumption 2.1 holds true on $[0, \infty)$;
(ii) $\int_0^\infty \sup_{m \in \mathcal{P}_2(\mathbf{S})} |F(t,m)| dt < \infty$;
(iii) for any (t,m) and $\mathbb{P} \in \mathcal{P}(t,m)$, $X_\infty := \lim_{t \to \infty} X_t$ exists, $\mathbb{P}$-a.s.

We remark that one sufficient condition of (ii) above is that $|f(t,x,m)| \le Ce^{-\lambda t}$ for some constants $C, \lambda > 0$, and a special case of (iii) is

$$d = 1, \quad b = b_0 x, \quad \sigma = \sigma_0 x, \quad b_0 - \frac{1}{2}\sigma_0^2 < 0; \tag{3.16}$$

see, e.g., Pedersen and Peskir [21] and Xu and Zhou [28]. The last condition implies that, under $\bar{\mathbb{P}}^{t,m}$ in (2.4), the unstopped process X is a geometric Brownian motion vanishing at infinity.

Assumption 3.14 allows us to include the case $\tau = \infty$ in our framework and to preserve the compactness of $\mathcal{P}(t,m)$ in the infinite horizon setting, so that all our previous results extend immediately.

Remark 3.15. A study of the general infinite horizon would of course be of very relevant interest. In standard optimal stopping, this is addressed by adding a discount factor to the reward function $\mathbb{E}[e^{-r\tau}\psi(X_\tau)]$, assuming $F = 0$ for simplicity. However, embedding this in our formulation is more involved as $\mathbb{E}[e^{-r\tau}\psi(X_\tau)] = \mathbb{E}[e^{-r\int_0^T I_s ds}\psi(X_\tau)]$, which is a function of the joint law of X_∞ and the path of I. We therefore leave it for further research.

4. Examples. In this section we revisit the three examples studied in [26] and add a new example concerning probability distortion. Note that in [26] we assumed that the value functions are smooth, which is hard to verify. In this section we show that they are the unique continuous viscosity solution of the corresponding obstacle problem. Note that we shall allow both $T < \infty$ and $T = \infty$, and correspondingly we always assume Assumption 2.1 or 3.14, and we shall report the detailed arguments in the case $T < \infty$ only. Moreover, for simplicity in this section we always assume $f = 0$.

4.1. Connection with standard optimal stopping. Assume for this example that b and σ do not depend on the measure variable m. For a measurable function $\psi : \mathbb{R}^d \to \mathbb{R}$, we define the optimal stopping problem

$$V(t,m) := \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \mathbb{E}^{\mathbb{P}}\Big[\psi(X_T)\Big], (t,m) \in \overline{\mathbf{Q}}_0. \tag{4.1}$$

That is, $g(\mu) := \int_{\mathbb{R}^d} \psi(x)\mu(dx)$ for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. We also introduce $v(t,x) := V(t,\delta_{(x,1)})$, which is related to the standard obstacle problem: recalling (2.7),

$$\min\{-(\partial_t + \mathcal{L})v, v - \psi\} = 0, \quad v(T,\cdot) = \psi, \quad \text{where} \quad \mathcal{L}v := b\cdot\partial_x v + \frac{1}{2}\sigma^2 : \partial^2_{xx} v. \tag{4.2}$$

Proposition 4.1. *Assume b, σ do not depend on m, σ satisfies the regularities required in Lemma* 3.7*, and ψ is uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation* (2.10)*, and it holds that*

$$V(t,m) = \int_{\mathbf{S}} \big[v(t,x)i + \psi(x)(1-i)\big] m(dx,di). \tag{4.3}$$

Moreover, there exists a pure strategy optimal stopping time.

Proof. First, by the uniform continuity of ψ one can easily show that g is uniformly continuous in m under $\mathcal{W}_1$. Then by Theorem 3.6 V is continuous in t and uniformly continuous in m under $\mathcal{W}_1$. Thus it follows from Theorems 3.10 and 3.13 that V is the unique viscosity solution of (2.10).

It remains to verify (4.3). Let $\mathbb{P}^* \in \mathcal{P}(t,m)$ be such that

$$\tau = \inf\{s \geq t : v(s, X_s) = \psi(X_s)\}, \quad \mathbb{P}^*\text{-a.s. on } \{I_{t-} = 1\}. \tag{4.4}$$

By the standard optimal stopping problem (see, e.g., Karatzas and Shreve [15, Appendix D]), v is continuous and $\mathbb{P}^*$ is optimal. Then by (4.1) we derive (4.3):

$$\begin{aligned} V(t,m) &= \int_{\mathbb{R}^d} \psi(x) m(dx,0) + \int_{\mathbb{R}^d} \mathbb{E}^{\mathbb{P}^*}\Big[\psi(X_T)\Big|X_t = x\Big] m(dx,1) \\ &= \int_{\mathbb{R}^d} \psi(x) m(dx,0) + \int_{\mathbb{R}^d} v(t,x) m(dx,1). \end{aligned}$$

Moreover, clearly the optimal stopping time determined by (4.4) is a pure strategy. ∎

We remark that, by utilizing (4.3), it is possible to prove the uniqueness of the viscosity solution under weaker requirement on σ. We leave the details to the interested reader.

4.2. A generalization of the mean variance problem. Consider the optimal stopping problem:

$$V(t,m) := \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \varphi\Big(\mathbb{E}^{\mathbb{P}}[\psi(X_T)]\Big), \tag{4.5}$$

where $\psi : \mathbb{R}^d \to \mathbb{R}^k$ for some $k \geq 1$ and $\varphi : \mathbb{R}^k \to \mathbb{R}$. That is, $g(\mu) = \varphi(\int_{\mathbb{R}^d} \psi(x)\mu(dx))$.

PROPOSITION 4.2. *Let b, σ satisfy the conditions in Theorem* 3.13(ii)*, ψ be uniformly continuous, and φ be continuous. Assume further that either ψ is bounded or φ is uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation* (2.10)*.*

Proof. Note that when $|\psi| \leq C$, we have $|\mathbb{E}^{\mathbb{P}}[\psi(X_T)]| \leq C$ and thus in (4.5) we may replace φ with the truncated function $\varphi_C(z) := \varphi(\frac{C}{|z|\vee C} z)$, $z \in \mathbb{R}^k$, which is uniformly continuous. Then in both cases, we may assume w.l.o.g. that φ is uniformly continuous, and therefore, g is uniformly continuous in μ under $\mathcal{W}_1$. Then the result follows from Theorems 3.6, 3.10, and 3.13. ∎

Remark 4.3. (i) In the case that φ is convex, $\varphi(z) := \sup_\alpha[\alpha z - \varphi^*(\alpha)]$, we have

$$V(t,m) = \sup_\alpha\, [V_\alpha(t,m) - \varphi^*(\alpha)], \quad \text{where} \quad V_\alpha(t,m) := \sup_{\mathbb{P}\in\mathcal{P}(t,m)} \mathbb{E}^{\mathbb{P}}[\alpha \cdot \psi(X_T)].$$

Let $\alpha^*(t,m)$ be the optimal argument; then the optimal $\mathbb{P}^*$ for $V_{\alpha^*(t,m)}(t,m)$ is also optimal for $V(t,m)$, and thus by Proposition 4.1 there exists a pure optimal strategy for $V(t,m)$.

Moreover, let $\mathbb{P}^*$ be the optimal control for $V(0,m)$ and $V_{\alpha^*(0,m)}(0,m)$ as above, and denote $m_t^* := \mathbb{P}^*_{Y_t}$. Then, by the DPP for V and for $V_{\alpha^*(0,m)}$, we have

$$\begin{aligned} V(t, m^*_{t-}) = V(0,m) &= V_{\alpha^*(0,m)}(0,m) - \varphi^*(\alpha^*(0,m)) \\ &= V_{\alpha^*(0,m)}(t, m^*_{t-}) - \varphi^*(\alpha^*(0,m)). \end{aligned}$$

That is, $\alpha^*(0,m)$ is optimal for $\sup_\alpha\, [V_\alpha(t, m^*_{t-}) - \varphi^*(\alpha)]$ or, say, $\alpha^*(t, m^*_{t-}) = \alpha^*(0,m)$ for all t.

(ii) A more special case is the mean variance problem: for some constant $\lambda > 0$,

$$d = 1, \quad k = 2, \quad \psi_1(x) = x, \quad \psi_2(x) = x^2, \quad \varphi(z_1, z_2) = z_1 + \frac{\lambda}{2} z_1^2 - \frac{\lambda}{2} z_2. \tag{4.6}$$

In the homogeneous case (3.16) with $T = \infty$, Pedersen and Peskir [21] solved the problem $V(\delta_{(x,1)})$ and the optimal stopping time is a pure strategy. We are in a much more general framework. However, we should point out that (4.6) does not satisfy the technical conditions in Proposition 4.2.

4.3. Expected shortfall. Let $d = 1$, and fix some $\alpha \in (0,1)$; we consider the mean field optimal stopping problem

$$V(t,m) := \inf_{\mathbb{P}\in\mathcal{P}(t,m)} \mathrm{ES}^{\mathbb{P}}_\alpha(X_T) \quad \text{for all } (t,m) \in \overline{\mathbf{Q}}_0,$$

where $\mathrm{ES}_\alpha^{\mathbb{P}}$ denotes the expected shortfall under $\mathbb{P}$, i.e., for any r.v. Z with law μ,

$$g(\mu) := \mathrm{ES}_\alpha^{\mathbb{P}}(Z) := \frac{1}{\alpha}\int_0^\alpha q_\gamma(Z)d\gamma = \inf_{\beta\in\mathbb{R}}\Big\{\beta + \frac{1}{1-\alpha}\int_{\mathbb{R}}(x-\beta)^+\mu(dx)\Big\}, \tag{4.7}$$
$$\text{where} \quad q_\gamma(Z) := \inf\{z : \mu(Z \le z) > \gamma\}.$$

Here the second equality has been established by Rockafellar and Uryasev [24].

PROPOSITION 4.4. *V is the unique continuous viscosity solution of the corresponding equation.*

Proof. Clearly, $x \mapsto (x-\beta)^+$ is Lipschitz continuous with Lipschitz constant 1. By (4.7), this implies that g is Lipschitz continuous, and, given our assumptions on the coefficients, we conclude similarly to Proposition 4.2 that the required claim follows. ∎

Note further that

$$V(t,m) = \inf_{\beta\in\mathbb{R}}\Big\{\beta + \frac{1}{1-\alpha}V_\beta(t,m)\Big\}, \quad \text{where} \quad V_\beta(t,m) := \inf_{\mathbb{P}\in\mathcal{P}(t,m)}\mathbb{E}^{\mathbb{P}}[(X_T-\beta)^+].$$

One can easily show that $\lim_{\beta\to\infty}[\beta + \frac{1}{1-\alpha}V_\beta(t,m)] = \lim_{\beta\to-\infty}[\beta + \frac{1}{1-\alpha}V_\beta(t,m)] = \infty$, where the second equality is due to $\alpha \in (0,1)$. Then there exists optimal $\beta^* = \beta^*(t,m) \in \mathbb{R}$ such that $V(t,m) = \beta^* + \frac{1}{1-\alpha}V_{\beta^*}(t,m)$. Therefore, similar to Remark 4.3(i), $V(t,m)$ and $V_{\beta^*}(t,m)$ share an optimal $\mathbb{P}^* \in \mathcal{P}(t,m)$, which is a pure optimal strategy as in Proposition 4.1.

Moreover, in the homogeneous case with (3.16) and $T = \infty$, one can easily show that V and V_β are independent of t, and $V_\beta(m) = \beta^- m(\mathbb{R}_+, 1) + \int_0^\infty (x-\beta)^+ m(dx, 0)$ whenever $m(\mathbb{R}_+, \{0,1\}) = 1$.

4.4. Probability distortion. Consider the following optimal stopping problem under probability distortion:

$$V(t,m) := \sup_{\mathbb{P}\in\mathcal{P}(t,m)}\int_0^\infty \varphi\big(\mathbb{P}(\psi(X_T) \ge z)\big)dz, \tag{4.8}$$

where $\psi : \mathbb{R}^d \to [0,\infty)$ is a utility function, $\varphi : [0,1] \to [0,1]$ is a probability distortion function $\varphi(0) = 0, \varphi(1) = 1$, and φ is strictly increasing. That is, $g(\mu) = \int_0^\infty \varphi(\mu(\{\psi \ge z\}))dz$.

PROPOSITION 4.5. *Let b, σ satisfy the conditions in Theorem* 3.13(ii)*, φ be a uniformly Lipschitz continuous probability distortion function, and ψ be uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation* (2.10)*.*

Proof. As in the previous examples, it suffices to show that g is uniformly continuous in m under $\mathcal{W}_1$. Assume arbitrary $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and, for $i = 1, 2$, let ξ_i be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}_{\xi_i} = \mu_i$ and $\mathbb{E}^{\mathbb{P}}[|\xi_1 - \xi_2|] = \mathcal{W}_1(\mu_1, \mu_2)$. Then

$$\begin{aligned}\big|g(\mu_1) - g(\mu_2)\big| &\le \int_0^\infty \Big|\varphi\big(\mathbb{P}(\psi(\xi_1) \ge z)\big) - \varphi\big(\mathbb{P}(\psi(\xi_2) \ge z)\big)\Big|dz \\ &\le C\int_0^\infty \mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{1}_{\{\psi(\xi_1)\ge z\}} - \mathbf{1}_{\{\psi(\xi_2)\ge z\}}\big|\Big]dz = C\mathbb{E}^{\mathbb{P}}\Big[\big|\psi(\xi_1) - \psi(\xi_2)\big|\Big].\end{aligned}$$

Since ψ is uniformly continuous, we see that g is uniformly continuous in μ under $\mathcal{W}_1$. ∎

Remark 4.6. (i) In the homogeneous case (3.16) with $T=\infty$, Xu and Zhou [28] solved the optimal stopping problem $V(\delta_{(x,1)})$ for appropriate φ,ψ, and the optimal stopping time is a pure strategy.
(ii) The mean variance and probability distortion problems are typically viewed as time inconsistent, as the DPP does not hold for value function $v(t,x):=V(t,\delta_{(x,1)})$. However, we emphasize that, by viewing m as our variable, V satisfies the DPP and the problem is hence time consistent. ∎

Appendix A. Technical results.

Proof of Lemma 2.3. (i) The set $\{m' : m' \preceq m\}$ is in continuous bijection with the compact set $\{\hat{m} \in \mathcal{P}_2(\mathbf{S}\times\{0,1\}) : \hat{m}\circ(\mathbf{x},\mathbf{i})^{-1} = m\}$, with $(\mathbf{x},\mathbf{i},\mathbf{i}')$ the projection coordinates on $\mathbf{S}\times\{0,1\}$. This shows the compactness of $\{m' : m' \preceq m\}$. (ii) As the map $m' \in \mathcal{K}(m) \longmapsto m'(\mathbb{R}^d,1)$ is continuous and $\mathcal{K}(m)$ is compact, there exists $\bar{m}\in\mathcal{K}(m)$ s.t. $\bar{m}\in \operatorname{argmin}_{m'\in\mathcal{K}(m)} m'(\mathbb{R}^d,1)$. Let $m'\in\mathcal{K}(m)$ be such that $m'\preceq\bar{m}$ with some corresponding transition probability p; see the definition in (2.9). Then, clearly $m'(\mathbb{R}^d,1)\le\bar{m}(\mathbb{R}^d,1)$ and thus equality holds by minimality of $\bar{m}(\mathbb{R}^d,1)$. As $p\le 1$, we conclude that $m'=\bar{m}$. ∎

Proof of Lemma 3.5. For each $(t,m)\in\overline{\mathbf{Q}}_0$ and $\mathbb{P}\in\mathcal{P}(t,m)$, we extend $\mathbb{P}$ to $(\Omega,\mathcal{F}_T)$ as follows: denote $\hat{\mathbb{P}}\in\hat{\mathcal{P}}(t,m)$, $X_s=X_r$, $I_s=I_{t-}$, $s\in[-1,t)$, $\hat{\mathbb{P}}$-a.s. We prove the lemma in two steps.

Step 1. For any compact $\mathcal{M}\subset\mathcal{P}_2(\mathbf{S})$, denote $\hat{\mathcal{P}}_{\mathcal{M}}:=\bigcup_{(t,m)\in[0,T]\times\mathcal{M}}\hat{\mathcal{P}}(t,m)$. For each (t,m), $\mathbb{P}\in\mathcal{P}(t,m)$, and $R>1$, following the proof of [26, Proposition 2.2] we have

$$\mathbb{E}^{\hat{\mathbb{P}}}\big[|X_T^*|^2\big]\le C\int_{\mathbb{R}^d}|x|^2 m(dx,\{0,1\});$$
$$\mathbb{E}^{\hat{\mathbb{P}}}\big[|X_T^*|^2\mathbf{1}_{\{|X_T^*|\ge R\}}\big]\le C\int_{\mathbf{S}}[1+|x|^2]\Big[\mathbf{1}_{\{|x|\ge\sqrt{R}-1\}}+\frac{1}{\sqrt{R}}\Big]m(dx,\{0,1\}),$$

where $X_T^*:=\sup_{0\le s\le T}|X_s|$. By the compactness of $\mathcal{M}$, one can easily see that

$$\sup_{\hat{\mathbb{P}}\in\hat{\mathcal{P}}_{\mathcal{M}}}\mathbb{E}^{\hat{\mathbb{P}}}\big[|X_T^*|^2\big]<\infty,\quad \lim_{R\to\infty}\sup_{\hat{\mathbb{P}}\in\hat{\mathcal{P}}_{\mathcal{M}}}\mathbb{E}^{\hat{\mathbb{P}}}\big[|X_T^*|^2\mathbf{1}_{\{|X_T^*|\ge R\}}\big]=0.$$

Then the set $\hat{\mathcal{P}}_{\mathcal{M}}$ is compact. That is, for any $(t_n,m_n)\in[0,T]\times\mathcal{M}$ and $\mathbb{P}^n\in\mathcal{P}(t_n,m_n)$, there exists a subsequence, still denoted the same, such that $\hat{\mathbb{P}}^n\to\hat{\mathbb{P}}^*$ under $\mathcal{W}_2$, for some $\mathbb{P}^*\in\mathcal{P}_2(\Omega,\mathcal{F}_T)$.

We may assume w.l.o.g. that $(t_n,m_n)\to(t^*,m^*)$ under $\mathcal{W}_2$ for some $(t^*,m^*)\in[0,T]\times\mathcal{M}$. We next show that $\hat{\mathbb{P}}^*\in\hat{\mathcal{P}}(t^*,m^*)$. Indeed, for any $\delta>0$, we have $t^*-\delta<t_n<t^*+\delta$ for all n large enough. By the required convergence, it is obvious that $X_s=X_{t^*-\delta}, I_s=I_{t^*-\delta}, s\le t^*-\delta, \hat{\mathbb{P}}^*$-a.s., and $\hat{\mathbb{P}}^*_{Y_{t-\delta}}=m^*$. Thus, by sending $\delta\to0$, $X_s=X_{t^*}$, $I_s=I_{t^*-}$, $s<t^*$, $\hat{\mathbb{P}}^*$-a.s. and $\hat{\mathbb{P}}^*_{Y_{t^*-}}=m^*$. Here we used the fact that X has continuous paths. Moreover, following the arguments in [26, Proposition 2.2] again, we see that the processes M and $MM^\top$ in (2.3) are $\hat{\mathbb{P}}^*$-martingales on $[t^*+\delta,T]$ for all $\delta>0$, and hence also on $[t^*,T]$ (again since X is continuous). That is, $\hat{\mathbb{P}}^*\in\hat{\mathcal{P}}(t^*,m^*)$.

Step 2. We now show that V is USC. Fix (t,m) and choose $(t_n,m_n)\to(t,m)$ such that $\lim_{n\to\infty}V(t_n,m_n)=\limsup_{(\tilde{t},\tilde{m})\to(t,m)}V(\tilde{t},\tilde{m})$. For each n, let $\mathbb{P}^n\in\mathcal{P}(t_n,m_n)$ be optimal: $V(t_n,m_n)=\int_{t_n}^T F(r,\mathbb{P}^n_{Y_r})dr+g(\mathbb{P}^n_{Y_T})$. Note that $\mathcal{M}:=\{m,m_n,n\ge1\}\subset$

$\mathcal{P}_2(\mathbf{S})$ is compact. By Step 1, we may assume w.l.o.g. that $\hat{\mathbb{P}}^n \to \hat{\mathbb{P}} \in \hat{\mathcal{P}}(t,m)$. Then, since F is continuous and g is USC in m, we have

$$\lim_{n\to\infty} V(t_n,m_n) = \lim_{n\to\infty} \Big[\int_{t_n}^T F(r,\hat{\mathbb{P}}^n_{Y_r})dr + g(\hat{\mathbb{P}}^n_{Y_T}) \Big] \le \int_t^T F(r,\hat{\mathbb{P}}_{Y_r})dr + g(\hat{\mathbb{P}}_{Y_T}) \le V(t,m).$$

This means that V is USC. ∎

Proof of Theorem 3.6. (i) follows similar but easier arguments than (ii), so we prove (ii) only. Let ρ_0 denote the modulus of continuity of f,g under $\mathcal{W}_1$. We proceed in two steps.

Step 1. Fix $t\in[0,T]$ and $m,\tilde{m}\in\mathcal{P}_2(\mathbf{S})$. For any $\mathbb{P}\in\mathcal{P}(t,m)$, by possibly enlarging the space, there exists $(\tilde{X}_t,\tilde{I}_{t-})$ on the space $(\Omega,\mathcal{F},\mathbb{P})$ such that

$$\mathbb{P}_{(\tilde{X}_t,\tilde{I}_{t-})} = \tilde{m},\ \mathbb{E}^{\mathbb{P}}\Big[|\tilde{X}_t - X_t| + |\tilde{I}_{t-} - I_{t-}|\Big] = \mathcal{W}_1(m,\tilde{m}).$$

Consider the following SDE on the space $(\Omega,\mathcal{F},\mathbb{P})$: for $\tilde{Y} := (\tilde{X},\tilde{I})$,

(A.1)
$$\tilde{X}_s = \tilde{X}_t + \int_t^s b(r,\tilde{X}_r,\mathbb{P}_{\tilde{Y}_r})\tilde{I}_r dr + \int_t^s \sigma(r,\tilde{X}_r,\mathbb{P}_{\tilde{Y}_r})\tilde{I}_r dW_r^{\mathbb{P}}, \quad \tilde{I}_r := I_r\tilde{I}_{t-}, \quad \mathbb{P}\text{-a.s.}$$

Denote $\Delta Y := \tilde{Y} - Y$. Note that $I_r = I_r I_{t-}$, then

$$\text{(A.2)}\quad \sup_{t\le r\le T}|\Delta I_r| = I_r|\Delta I_{t-}| \le |\Delta I_{t-}|, \quad \text{and thus} \quad \mathbb{E}^{\mathbb{P}}\Big[\sup_{t\le r\le T}|\Delta I_r|\Big] \le \mathcal{W}_1(m,\tilde{m}).$$

Moreover, for $\varphi = b,\sigma$, by the desired Lipschitz continuity under $\mathcal{W}_1$, we have

$$\Big|\varphi(r,\tilde{X}_r,\mathbb{P}_{\tilde{Y}_r})\tilde{I}_r - \varphi(r,X_r,\mathbb{P}_{Y_r})I_r\Big| \le C\Big[|\Delta X_r| + \mathcal{W}_1(\mathbb{P}_{\tilde{Y}_r},\mathbb{P}_{Y_r})\Big] + C[1+|X_r|]|\Delta I_{t-}|.$$

By standard estimates, one can show that

$$\mathbb{E}^{\mathbb{P}}_t\Big[\sup_{t\le s\le T}|X_s|^2\Big] \le C[1+|X_t|^2];$$

$$\mathbb{E}^{\mathbb{P}}_t\Big[|\Delta X_s|^2\Big] \le C\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r},\mathbb{P}_{Y_r})dr + C|\Delta X_t|^2 + C\sup_{t\le s\le T}\mathbb{E}^{\mathbb{P}}_t[1+|X_s|^2]|\Delta I_{t-}|^2;$$

(A.3)
$$\mathbb{E}^{\mathbb{P}}_t\big[|\Delta X_s|\big] \le C\Big(\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r},\mathbb{P}_{Y_r})dr\Big)^{\frac12} + C|\Delta X_t| + C[1+|X_t|]|\Delta I_{t-}|.$$

This implies that

$$\mathbb{E}^{\mathbb{P}}\big[|\Delta X_s|\big] \le C\Big(\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r},\mathbb{P}_{Y_r})dr\Big)^{\frac12} + C\mathbb{E}^{\mathbb{P}}\Big[|\Delta X_t| + [1+|X_t|]|\Delta I_{t-}|\Big];$$

$$\mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_s},\mathbb{P}_{Y_s}) \le C\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r},\mathbb{P}_{Y_r})dr + C\Big(\mathbb{E}^{\mathbb{P}}\Big[|\Delta X_t| + [1+|X_t|]|\Delta I_{t-}|\Big]\Big)^2.$$

By Gronwall's inequality we have, for any $R>0$,

$$\begin{aligned}\sup_{t\le s\le T}\mathcal{W}_1(\mathbb{P}_{\tilde{Y}_s},\mathbb{P}_{Y_s}) &\le C\mathbb{E}^{\mathbb{P}}\Big[|\Delta X_t| + [1+|X_t|]|\Delta I_{t-}|\Big]\\ &\le CR\mathcal{W}_1(m,\tilde{m}) + C\mathbb{E}^{\mathbb{P}}\Big[|X_t|\mathbf{1}_{\{|X_t|\ge R\}}\Big] =: \delta_R. \qquad \text{(A.4)}\end{aligned}$$

Notice that $\tilde{\mathbb{P}} := \mathbb{P} \circ (\tilde{X}, I)^{-1} \in \mathcal{P}(t, \tilde{m})$. Then

$$\begin{aligned}
&\int_t^T \hat{F}(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}) - V(t, \tilde{m}) \\
&\le \int_t^T \mathbb{E}^{\mathbb{P}}\big[f(r, X_r, \mathbb{P}_{Y_r}) I_r - f(r, \tilde{X}_r, \mathbb{P}_{\tilde{Y}_r}) \tilde{I}_r\big] dr + \big[g(\mathbb{P}_{Y_T}) - g(\mathbb{P}_{\tilde{Y}_T})\big] \\
&\le \rho_0\Big(\mathcal{W}_1(\mathbb{P}_{\tilde{Y}_T}, \mathbb{P}_{Y_T})\Big) + \int_t^T \mathbb{E}^{\mathbb{P}}\Big[\rho_0\big(|\Delta X_r|\big) + \rho_0\big(\mathcal{W}_1(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r})\big) + |f(r, X_r, \mathbb{P}_{Y_r})||\Delta I_r|\Big].
\end{aligned} \tag{A.5}$$

The uniform regularity of f implies that

$$|f(r, X_r, \mathbb{P}_{Y_r})| \le |f(r, 0, \mathbb{P}_{Y_r})| + C|X_r| \le C_m[1 + |X_r|],$$

where the constant C_m may depend on m. Then, by (A.2), (A.3), and (A.4), we have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\Big[|f(r, X_r, \mathbb{P}_{Y_r})||\Delta I_r|\Big] \le C_m \mathbb{E}^{\mathbb{P}}\Big[[1 + |X_r|]|\Delta I_{t-}|\Big] \\
&\le C_m \mathbb{E}^{\mathbb{P}}\Big[|\Delta I_{t-}| + \rho_0(\delta_R) + |\Delta X_t| + [1 + |X_t|]|\Delta I_{t-}|\Big] \le C_m[\mathcal{W}_1(m, \tilde{m}) + \rho_0(\delta_R)\big].
\end{aligned}$$

Plugging this into (A.5), we have

$$\begin{aligned}
&\int_t^T \hat{F}(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}) - V(t, \tilde{m}) \\
&\quad \le C_m\big[\mathcal{W}_1(m, \tilde{m}) + \rho_0(\delta_R)\big] + \int_t^T \mathbb{E}^{\mathbb{P}}\big[\rho_0(|\Delta X_r|)\big] dr.
\end{aligned}$$

Since $\mathbb{P} \in \mathcal{P}(t, m)$ is arbitrary, for some appropriate modulus of continuity ρ we have $V(t, m) - V(t, \tilde{m}) \le C_m \rho(\delta_R)$. Switching $m, \tilde{m}$, and noticing that we may still use X_t in δ_R, we have

$$|V(t, m) - V(t, \tilde{m})| \le C_m \rho(\delta_R). \tag{A.6}$$

Fixing m and send $\tilde{m} \to m$ under $\mathcal{W}_1$, we see that

$$\limsup_{\tilde{m} \to m} |V(t, m) - V(t, \tilde{m})| \le C_m \rho\Big(C\mathbb{E}^{\mathbb{P}}[|X_t| \mathbf{1}_{\{|X_t| \ge R\}}]\Big)$$

for any $R > 0$. Now sending $R \to \infty$, we see that $\lim_{\tilde{m} \to m} V(t, \tilde{m}) = V(t, m)$.

Step 2. Let $t < \tilde{t}$ and $m \in \mathcal{P}_2(\mathbf{S})$. By DPP we have

$$\begin{aligned}
V(t, m) &= \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \Big\{ \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr + V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}-}}) \Big\} \\
&= \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \Big\{ \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr + V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}}}) \Big\}, \\
V(\tilde{t}, m) &= \sup_{m' \preceq m} V(\tilde{t}, m').
\end{aligned} \tag{A.7}$$

First, for any $\mathbb{P} \in \mathcal{P}(t, m)$, note that $m' := \mathbb{P} \circ (X_t, I_{\tilde{t}-})^{-1} \preceq m$; then

$$V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}-}}) - V(\tilde{t}, m) \le V(\tilde{t}, \mathbb{P}_{(X_{\tilde{t}}, I_{\tilde{t}-})}) - V(\tilde{t}, \mathbb{P}_{(X_t, I_{\tilde{t}-})}) \le C_m \rho(\delta_R),$$

thanks to (A.6) and (A.4), where, following similar arguments as in Step 1,

$$\begin{aligned}\delta_R &:= CR\mathbb{E}^{\mathbb{P}}[|X_{\tilde t}-X_t|]+C\mathbb{E}^{\mathbb{P}}\Big[|X_t|\mathbf{1}_{\{|X_t|\ge R\}}\Big]\\ &\le CR\mathbb{E}^{\mathbb{P}}[1+|X_t|]\sqrt{\tilde t-t}+C\mathbb{E}^{\mathbb{P}}\big[|X_t|\mathbf{1}_{\{|X_t|\ge R\}}\big].\end{aligned}\tag{A.8}$$

Since $\mathbb{P}\in\mathcal{P}(t,m)$ is arbitrary, by (A.7) we have

$$V(t,m)-V(\tilde t,m)\le\sup_{\mathbb{P}\in\mathcal{P}(t,m)}\int_t^{\tilde t}F(r,\mathbb{P}_{Y_r})dr+C_m\rho(\delta_R)\le C_m\rho(\delta_R).$$

Next, for $m'\preceq m$, choose $\mathbb{P}\in\mathcal{P}(t,m)$ s.t. $I_s=I_{t-}$, $t\le s<\tilde t$, and $\mathbb{P}\circ(X_t,I_{\tilde t})^{-1}=m'$. Then

$$V(\tilde t,m')-V(t,m)\le V(\tilde t,\mathbb{P}_{(X_t,I_{\tilde t})})-V(\tilde t,\mathbb{P}_{(X_{\tilde t},I_{\tilde t})})-\int_t^{\tilde t}F(r,\mathbb{P}_{Y_r})dr\le C_m\rho(\delta_R).$$

Since $m'\preceq m$ is arbitrary, by (A.7) we have

$$V(\tilde t,m)-V(t,m)\le C_m\rho(\delta_R),\quad\text{and thus}\quad\big|V(t,m)-V(\tilde t,m)\big|\le C_m\rho(\delta_R).$$

This, together with (A.8), implies the desired regularity immediately. ∎

Proof of Lemma 3.7. We shall apply the results in Buckdahn et al. [1]. For this purpose, we extend functions on $\mathcal{P}_2(\mathbf{S})$ to $\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R})$. Let $\phi:\mathbb{R}\to\mathbb{R}$ be a smooth function with bounded derivatives s.t. $0\le\phi\le1$, $\phi(0)=0$, $\phi(1)=1$, and $\Phi:\hat m\in\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R})\mapsto m\in\mathcal{P}_2(\mathbf{S})$, with

$$m(A,1):=\int_{\mathbb{R}}\phi(y)\hat m(A,dy),\quad m(A,0):=\int_{\mathbb{R}}[1-\phi(y)]\hat m(A,dy)\quad\text{for all } A\in\mathcal{B}(\mathbb{R}^d).$$

Now for $\varphi=b,\sigma,f,g$, define $\hat\varphi(t,x,\hat m):=\varphi(t,x,\Phi(\hat m))$. $\hat\varphi$ inherits the regularity of φ on $\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R})$.

Next, fix a filtered probability space $(\hat\Omega,\hat{\mathcal{F}}_T,\hat{\mathbb{F}},\hat{\mathbb{P}})$ on which is defined a d-dimensional Brownian motion W. For any $(t,\hat m)$, let $\xi\in\mathbb{L}^2(\mathcal{F}_t;\mathbb{R}^d)$, $\eta\in\mathbb{L}^2(\mathcal{F}_t;\mathbb{R})$ be such that $\hat{\mathbb{P}}_{(\xi,\eta)}=\hat m$. Consider the following SDE on $[t,T]$ with solution $\hat Y=(\hat X,\hat I)$:

$$\begin{aligned}\hat X_s&=\xi+\int_t^s\hat b(r,\hat X_r,\hat{\mathbb{P}}_{\hat Y_r})\phi(\hat I_r)dr+\int_t^s\hat\sigma(r,\hat X_r,\hat{\mathbb{P}}_{\hat Y_r})\phi(\hat I_r)dW_r;\\ \hat I_s&=\eta\mathbf{1}_{[t,T)}(s),\quad\hat{\mathbb{P}}\text{-a.s.}\end{aligned}$$

We then define, recalling (2.1),

$$\hat U(t,\hat m):=\hat g(\hat{\mathbb{P}}_{\hat Y_T})+\int_t^T\hat F(r,\hat{\mathbb{P}}_{Y_r})dr,\quad\text{where}\quad\hat F(r,\hat m):=\int_{\mathbb{R}^{d+1}}\hat f(r,x,\hat m)i\,\hat m(dx,di).$$

We remark that, since b and σ are not necessarily bounded, the coefficients of the SDE for $\hat X$ are not Lipschitz continuous in $\hat I$. However, since $\hat I$ is already given, such Lipschitz continuity is not needed. In particular, we can apply [1, Lemmas 6.2 and 7.1] so that $\partial_t\hat U,\partial_{\hat m}\hat U,\partial_{\hat y\hat m}\hat U$ exist and are continuous and bounded. Here $\partial_{\hat m}\hat U$ is the Lions derivative and satisfies $\partial_{\hat m}\hat U(t,\hat m,\hat y):=\partial_{\hat y}\delta_{\hat m}\hat U(t,\hat m,\hat y)$; see, e.g., Carmona and Delarue [5, Volume 1, Chapter 5]. We also remark that in [1] the function $\hat U$ takes the form $\hat U(t,x,\hat m)$ while here $\hat U$ does not have the x-variable. Moreover, note

that each $m \in \mathcal{P}(\mathbf{S})$ can be viewed as an element of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R})$ with support included in $\mathbf{S}$. Since $\phi(0) = 0, \phi(1) = 1$, one can easily see that $U(t,m) = \hat{U}(t,m)$. Then clearly $U \in C_2^{1,2}(\mathbf{Q}_0)$. Finally, the $\mathcal{W}_1$ Lipschitz continuity of U follows arguments similar to that of Theorem 3.6, and we thus omit the proof here. ∎

Proof of Lemma 3.8. (ii) and (iii) follow directly from [20, Theorem 3.1], after the straightforward extension to $\mathcal{P}_2(\mathbf{S})$, as we will do next. Thus we shall only prove (i). For ease of presentation, we assume $d = 1$.

Fixing $n \geq 1$, we construct U_n as follows. First, let $H_n, \phi_j^n \in C^\infty(\mathbb{R})$, $j \in \mathbb{Z}$, satisfy

$$0 \leq H_n \leq 1, \quad \text{Supp}\,(H_n) \subset \Big[-\frac{3n}{2}, \frac{3n}{2}\Big], \quad H_n(x) = 1 \text{ for } |x| \leq n, \quad |\partial_x H_n| \leq \frac{3}{n};$$
$$0 \leq \phi_j^n \leq 1, \quad \text{Supp}\,(\phi_j^n) \subset \Big[\frac{j-1}{n}, \frac{j+1}{n}\Big], \quad \phi_j^n(x) + \phi_{j+1}^n(x) = 1 \text{ for all } x \in \Big[\frac{j}{n}, \frac{j+1}{n}\Big].$$

See [27, (3.3)] for a construction of ϕ_j^n. Next, for each $j \in \mathbb{Z}$, define

$$\psi_j^n(\mu) := \int_{\mathbb{R}} \phi_j^n(x) H_n(x) \mu(dx) + \mathbf{1}_{\{j=0\}} \int_{\mathbb{R}} [1 - H_n(x)] \mu(dx) \tag{A.9}$$

for all finite measure μ on $\mathbb{R}$. We emphasize that, slightly different from [27], here the μ will be $m(\cdot, i)$ whose total measure is less than 1 and thus it is not a probability measure. Note that $\psi_j^n \geq 0$ and $\sum_{j \in \mathbb{Z}} \psi_j^n = \mu(\mathbb{R})$. Moreover, denote $\mathbb{Z}_n := \{j \in \mathbb{Z} : |j| \leq 2n^2\}$ with size $N_n := 4n^2 + 1$, and

$$\Delta_n := \Big\{ \vec{z} = \{z_j\}_{j \in \mathbb{Z}_n} : |z_j| \leq N_n^{-3} \text{ for all } j \neq 0, \quad \text{and} \quad z_0 := -\sum_{j \in \mathbb{Z}_n \setminus \{0\}} z_j \Big\}.$$

We now define, for each $\vec{z} \in \Delta_n$ and $m \in \mathcal{P}_2(\mathbf{S})$, $i = 0, 1$,

$$\begin{aligned} m_n(dx, i, \vec{z}) &:= \sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(m(\cdot, i), \vec{z}) \delta_{\frac{j}{n}}(dx), \quad \hat{\psi}_j^n(\mu, \vec{z}) \\ &:= \frac{N_n}{N_n + 1} \Big[\psi_j^n(\mu) + \mu(\mathbb{R}) \Big[\frac{1}{N_n^2} + z_j \Big] \Big]. \end{aligned} \tag{A.10}$$

Note that $|z_0| \leq N_n^{-2}$, and thus $\hat{\psi}_j^n(\mu, \vec{z}) \geq 0$. One may easily verify that

$$\sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(\mu, \vec{z}) = \frac{N_n}{N_n + 1} \Big[\sum_{j \in \mathbb{Z}_n} \psi_j^n(\mu) + \frac{\mu(\mathbb{R})}{N_n} \Big] = \frac{N_n}{N_n + 1} \Big[\mu(\mathbb{R}) + \frac{\mu(\mathbb{R})}{N_n} \Big] = \mu(\mathbb{R});$$
$$m_n(\mathbb{R}, i, \vec{z}) = \sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(m(\cdot, i), \vec{z}) = m(\mathbb{R}, i), \quad \text{and thus} \quad m_n(\mathbf{S}, \vec{z}) = 1.$$

In particular, this implies that $m_n(\cdot, \vec{z}) \in \mathcal{P}_2(\mathbf{S})$ for every $\vec{z} \in \Delta_n$, where the square integrability follows from the fact that $\text{Supp}\,(m_n(\cdot, \vec{z}))$ is finite. Finally, let ζ_n be a smooth density function with support Δ_n, and we construct

$$U_n(m) := \int_{\Delta_n} U(m_n(\cdot, \vec{z})) \zeta_n(\vec{z}) d\vec{z}, \quad m \in \mathcal{P}_2(\mathbf{S}). \tag{A.11}$$

The smoothness of U_n follows from the same arguments as in [20, Theorem 3.1]. However, we note that [20] uses the $\mathcal{W}_1$-distance and requires $\mathcal{M}$ to be a compact subset of $\mathcal{P}_1(\mathbf{S})$. This is mainly for the uniform Lipschitz continuity of U_n which holds

only under $\mathcal{W}_1$. Here we provide a proof for the uniform convergence of U_n under $\mathcal{W}_2$. We first show that

$$\overline{\mathcal{M}} := \Big\{ m_n(\cdot,\vec{z}) : m \in \mathcal{M}, n \geq 1, \vec{z} \in \Delta_n \Big\} \subset \mathcal{P}_2(\mathbf{S}) \quad \text{is compact.} \tag{A.12}$$

Indeed, fix $R>0$. Denote $\mathbb{Z}_n^R := \{j \in \mathbb{Z}_n : |j| \geq nR\}$ for $n > \frac{R}{2}$. Then

$$\begin{aligned}\int_{\{|x|>R\}} |x|^2 m_n(dx,i,\vec{z}) &= \sum_{j\in\mathbb{Z}_n^R} \frac{j^2}{n^2} \hat{\psi}_j^n(m(\cdot,i),\vec{z}) \\ &= \sum_{j\in\mathbb{Z}_n^R} \frac{j^2}{n^2} \frac{N_n}{N_n+1} \Big[\psi_j^n(m(\cdot,i)) + m(\mathbb{R},i)\Big[\frac{1}{N_n^2} + z_j\Big]\Big].\end{aligned}$$

From the construction of $\psi_{\vec{j}}$, one can easily verify that

$$\sum_{\vec{j}\in\mathbb{Z}_n^R} \frac{j^2}{n^2} \psi_j^n(m(\cdot,i)) \leq 2 \int_{\{|x|>R\}} |x|^2 m(dx,i).$$

Moreover, note that $|z_j| \leq N_n^{-3}$ for all $j \in \mathbb{Z}_n^R$. Then, for $n > \frac{R}{2}$,

$$\begin{aligned}&\int_{\{|x|>R\}} |x|^2 m_n(dx,i,\vec{z}) \leq 2 \int_{\{|x|>R\}} |x|^2 m(dx,i) + \sum_{j\in\mathbb{Z}_n^R} \frac{j^2}{n^2} m(\mathbb{R},i) \frac{C}{N_n^2} \\ &\leq 2 \int_{\{|x|>R\}} |x|^2 m(dx,i) + \frac{C m(\mathbb{R},i)}{N_n} \leq 2 \int_{\{|x|>R\}} |x|^2 m(dx,i) + \frac{C m(\mathbb{R}^d,i)}{R^2}.\end{aligned}$$

On the other hand, when $n < \frac{R}{2}$, we have $\int_{\{|x|>R\}} |x|^2 m_n(dx,i,\vec{z}) = 0$. Thus,

$$\sup_{m\in\mathcal{M}, n\geq 1, \vec{z}\in\Delta_n} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m_n(dx,i,\vec{z}) \leq 2 \sup_{m\in\mathcal{M}} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m(dx,i) + \frac{C}{R^2}.$$

Since $\mathcal{M} \subset \mathcal{P}_2(\mathbf{S})$ is compact, we have $\lim_{R\to\infty} \sup_{m\in\mathcal{M}} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m(dx,i) = 0$. Then

$$\lim_{R\to\infty} \sup_{m\in\mathcal{M}, n\geq 1, \vec{z}\in\Delta_n} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m_n(dx,i,\vec{z}) = 0.$$

This proves that $\overline{\mathcal{M}}$ is uniformly square integrable, and therefore compact in $\mathcal{P}_2(\mathbf{S})$.

Next, note that $\mathcal{M}$ is also compact in $\mathcal{P}_1(\mathbf{S})$; by [20, (3.15)] we have

$$\lim_{n\to\infty} \sup_{m\in\mathcal{M}, \vec{z}\in\Delta_n} \mathcal{W}_1(m_n(\cdot,\vec{z}),m) = 0. \tag{A.13}$$

Then, for any $R>0$,

$$\mathcal{W}_2^2(m_n(\cdot,\vec{z}),m) \leq R\mathcal{W}_1(m_n(\cdot,\vec{z}),m) + C \sum_{i=0,1} \int_{|x|\geq\frac{R}{2}} |x|^2 \big[m_n(dx,i,\vec{z}) + m(dx,i)\big].$$

This, together with the uniform integrability of $\overline{\mathcal{M}}$ and (A.13), implies immediately that

$$\lim_{n\to\infty} \sup_{m\in\mathcal{M}, \vec{z}\in\Delta_n} \mathcal{W}_2(m_n(\cdot,\vec{z}),m) = 0. \tag{A.14}$$

Finally, by the compactness (A.12), we see that U is uniformly continuous on $\overline{\mathcal{M}}$. Then it follows from (A.11) and (A.14) that $\lim_{n\to\infty}\sup_{m\in\mathcal{M}}|U_n(m)-U(m)|=0$. ∎

Remark A.1. While not used in the paper, the following property is interesting in its own right: if U is monotone under $\preceq$, then so is the U_n constructed in (A.11). Indeed, assume U is increasing, and let $m'\preceq m$ with transition probability p. For each $\vec{z}\in\Delta_n$, by (A.9) and (A.10), it is clear that

$$0<\hat{\psi}_j^n(m'(\cdot,1),\vec{z})\le\hat{\psi}_j^n(m(\cdot,1),\vec{z}),\quad\text{and thus}$$
$$\hat{p}_j(\vec{z}):=\hat{\psi}_j^n(m'(\cdot,1),\vec{z})\,/\,\hat{\psi}_j^n(m'(\cdot,1),\vec{z})\in(0,1].$$

Since $m'(dx,\{0,1\})=m(dx,\{0,1\})$, it is also obvious that $\sum_{i=0,1}\hat{\psi}_j^n(m'(\cdot,i),\vec{z})=\sum_{i=0,1}\hat{\psi}_j^n(m(\cdot,i),\vec{z})$. Then $m_n'(\cdot,\vec{z})\le m_n(\cdot,\vec{z})$ for each $\vec{z}\in\Delta_n$, with transition probability $\hat{p}(\cdot,\vec{z})$ satisfying $\hat{p}(\frac{j}{n},\vec{z})=\hat{p}_j(\vec{z})$ for all $j\in\mathbb{Z}_n$. Then, since U is increasing, by (A.11) we see that $U_n(m')\le U_n(m)$.

Proof of Lemma 3.9. (i) Let $m'\preceq m$ with transition probability p. As m is a probability measure on $(\mathbf{S},\mathcal{B}(\mathbf{S}))$, it is a Radon measure. Then, by Lusin's theorem (see Folland [12, 7.10]), we may find for all $k\ge1$ a continuous $p_k:\mathbb{R}^d\longrightarrow[0,1]$ s.t.

$$m\big(\{x:p(x)\neq p_k(x)\},\{0,1\}\big)\le\frac{1}{k}.$$

Let $\{m_k'\}_{k\ge1}$ be the measures obtained from m by applying the transition probabilities $\{p_k\}_{k\ge1}$, and ϕ a bounded and continuous function. Then

$$\Big|\int_{\mathbb{R}^d}\phi(x)p_k(x)m(dx,1)-\int_{\mathbb{R}^d}\phi(x)p(x)m(dx,1)\Big|\le\frac{2}{k}\|\phi\|_\infty,$$

and thus $m_k'(dx,1)$ converges weakly to $m'(dx,1)$. We do similarly with $m_k'(dx,0)$, and thus m_k' converges weakly to m'. As $\{m_k'\}_{k\ge1}$ is uniformly integrable, we have $\lim_{k\to\infty}\mathcal{W}_2(m_k',m')=0$. As v is nondecreasing for $\preceq$, we have $v(m)\ge v(m_k')$ for all $k\ge1$. Then, as v is $\mathcal{N}$-LSC, we have $v(m)\ge\liminf_{k\to\infty}v(m_k')\ge v(m')$.
(ii) As $(\mathbb{D}_I\varphi)_*$ is LSC, there exists $\delta>0$ s.t. $(\mathbb{D}_I\varphi)_*\ge0$ on $[t,t+\delta]\times\mathcal{B}_{\mathcal{W}_2}(m,\delta)$. Let $(s,m_0),(s,m_1)$ be in this neighborhood, s.t. $m_1\preceq m_0$ with transition probability p. Then, we have

$$\varphi(s,m_0)-\varphi(s,m_1)=\int_0^1\int_{\mathbb{R}^d}D_I\varphi(t,\lambda m_0+(1-\lambda)m_1,x)(1-p(x))m(dx,1)d\lambda.$$

By convexity of $\mathcal{B}_{\mathcal{W}_2}(m,\delta)$, we have $D_I\varphi(t,\lambda m_0+(1-\lambda)m_1,\cdot)\ge0$, hence the desired result. ∎

REFERENCES

[1] R. Buckdahn, J. Li, S. Peng, and C. Rainer, *Mean-field stochastic differential equations and associated PDEs*, Ann. Probab., 45 (2017), pp. 824–878.

[2] M. Burzoni, I. Vincenzo, A. M. Reppen, and H. M. Soner, *Viscosity solutions for controlled McKean-Vlasov jump-diffusions*, SIAM J. Control Optim., 58 (2020), pp. 1676–1699.

[3] P. Cardaliaguet and M. Quincampoix, *Deterministic differential games under probability knowledge of initial condition*, Int. Game Theory Rev., 10 (2008), pp. 1–16.

[4] R. Carmona and F. Delarue, *The master equation for large population equilibriums*, in *Stochastic Analysis and Applications 2014*, Springer, New York, 2014, pp. 77–128.

[5] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications* I-II, Springer, New York, 2018.

[6] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato, *Master Bellman Equation in the Wasserstein Space: Uniqueness of Viscosity Solutions*, https://arxiv.org/abs/2107.10535, 2021.
[7] I. Ekren, C. Keller, N. Touzi, and J. Zhang, *On viscosity solutions of path dependent PDEs*, Ann. Probab., 42 (2014), pp. 204–236.
[8] I. Ekren, N. Touzi, and J. Zhang, *Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part* I, Ann. Probab., 44 (2016), pp. 1212–1253.
[9] I. Ekren, N. Touzi, and J. Zhang, *Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II*, Ann. Probab., 44 (2016), pp. 2507–2553.
[10] N. El Karoui, *Les aspects probabilistes du contrôle stochastique*, in Ecole d'Eté de Probabilités de Saint-Flour IX-1979, Springer, New York, 1981, pp. 73–238.
[11] G. Fabbri, F. Gozzi, and A. Swiech, *Stochastic Optimal Control in Infinite Dimension*, Probab. Theory Stoch. Model. 82, Springer, New York, 2017.
[12] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, Pure Appl. Math. 40, John Wiley & Sons, New York, 1999.
[13] W. Gangbo, T. Nguyen, and A. Tudorascu, *Hamilton-Jacobi equations in the Wasserstein space*, Methods Appl. Anal., 15 (2008), pp. 155–184.
[14] C. Jimenez, A. Marigonda, and M. Quincampoix, *Optimal control of multiagent systems in the Wasserstein space*, Calc. Var. Partial Differential Equations, 59 (2020), pp. 1–45.
[15] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Appl. Math. 39, Springer, New York, 1998.
[16] P.-L. Lions, *Viscosity solutions of fully nonlinear second-order equations and optimal stochastic control in infinite dimensions. Part* I*: The case of bounded stochastic evolutions*, Acta Math., 161 (1988), pp. 243–278.
[17] P.-L. Lions, *Viscosity solutions of fully nonlinear second order equations and optimal stochastic control in infinite dimensions. Part* II*: Optimal control of Zakai's equation*, in *Stochastic Partial Differential Equations and Applications* II, Springer, New York, 1989, pp. 147–170.
[18] P.-L. Lions, *Viscosity solutions of fully nonlinear second-order equations and optimal stochastic control in infinite dimensions. Part* III*: Uniqueness of viscosity solutions for general second-order equations*, J. Funct. Anal., 86 (1989), pp. 1–18.
[19] P.-L. Lions, *Cours au Collège de France*, www. college-de-france.fr (2007).
[20] C. Mou and J. Zhang, *Wellposedness of second order master equations for mean field games with nonsmooth data*, Mem. Amer. Math. Soc., to appear.
[21] J. L. Pedersen and G. Peskir, *Optimal mean-variance selling strategies*, Math. Financ. Econ., 10 (2016), pp. 203–220.
[22] H. Pham and X. Wei, *Bellman equation and viscosity solutions for mean-field stochastic control problem*, ESAIM Control Optim. Calc. Var., 24 (2018), pp. 437–461.
[23] Z. Ren, N. Touzi, and J. Zhang, *Comparison of viscosity solutions of fully nonlinear degenerate parabolic path-dependent PDEs*, SIAM J. Math. Anal., 49 (2017), pp. 4093–4116.
[24] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, J. Risk, 2 (2000), pp. 21–42.
[25] A. N. Shiryaev, *Optimal Stopping Rules*, Stoch. Model. Appl. Probab. 8, Springer, New York, 2007.
[26] M. Talbi, N. Touzi, and J. Zhang, *Dynamic Programming Equation for the Mean Field Optimal Stopping Problem*, https://arxiv.org/abs/2103.05736, 2021.
[27] C. Wu and J. Zhang, *Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls*, Ann. Appl. Probab., 30 (2020), pp. 936–986.
[28] Q. Zuoquan and X. Y. Zhou, *Optimal stopping under probability distortion*, Ann. Appl. Probab., 23 (2013), pp. 251–282.