The Annals of Applied Probability
2024, Vol. 34, No. 5, 4237–4267
https://doi.org/10.1214/24-AAP2064

FROM FINITE POPULATION OPTIMAL STOPPING TO MEAN FIELD OPTIMAL STOPPING

BY MEHDI TALBI[1,a], NIZAR TOUZI[2,b] AND JIANFENG ZHANG[3,c]

[1]*Laboratoire de Probabilités, Statistiques et Modélisation, Université Paris-Cité,* [a]*talbi@lpsm.paris*
[2]*Tandon School of Engineering, New York University,* [b]*nt2635@nyu.edu*
[3]*Department of Mathematics, University of Southern California,* [c]*jianfenz@usc.edu*

This paper analyzes the convergence of the finite population optimal stopping problem towards the corresponding mean field limit. Building on the viscosity solution characterization of the mean field optimal stopping problem of our previous papers (*SIAM J. Control Optim.* **61** (2023) 1712–1736, 2140–2164), we prove the convergence of the value functions by adapting the Barles–Souganidis (*Asymptot. Anal.* **4** (1991) 271–283) monotone scheme method to our context. We next characterize the optimal stopping policies of the mean field problem by the accumulation points of the finite population optimal stopping strategies. In particular, if the limiting problem has a unique optimal stopping policy, then the finite population optimal stopping strategies do converge towards this solution. As a by-product of our analysis, we provide an extension of the standard propagation of chaos to the context of stopped McKean–Vlasov diffusions.

1. Introduction. This paper is a continuation of our previous papers [54, 55], where we established the well-posedness, especially in the viscosity sense, for an obstacle equation on Wasserstein space derived from a mean field optimal stopping problem. In this paper we focus on the convergence of the corresponding N-player optimal stopping problem, the latter was investigated by Kobylanski, Quenez and Rouy–Mironescu [39] in a more general setting. We shall obtain both the convergence of the value functions and the propagation of chaos in terms of the (approximate) optimal trajectories.

In the context of mean field games, initiated independently by Lasry and Lions [45] and Caines, Huang and Malhamé [37], the convergence of the N-player game to the mean field game is one of the core issues in the field. When the mean field equilibrium is unique, typically under certain monotonicity conditions, one may use the popular master equation approach to obtain the convergence, see Cardaliaguet, Delarue, Lasry and Lions [8], followed by Bayraktar and Cohen [4], Bayraktar, Cecchin, Cohen and Delarue [3], Cardaliaguet [6], Cecchin and Pelino [15], Delarue–Lacker–Ramanan [18, 19], Gangbo and Meszaros [35] and Mou and Zhang [47], to mention a few. When there are multiple equilibria, there are numerous publications on the convergence issue, mainly on the propagation of chaos; see, for example, Camona and Delarue [10], Cecchin, Dai Pra, Fisher and Pelino [16], Djete [20, 21], Doncel, Gast and Gaujal [23], Feleqi [28], Fischer [29], Fischer and Silva [31], Lacker [41, 43], Lacker and Flem [44], Lasry and Lions [45], Lauriere and Tangpi [46], Nutz, San Martin and Tan [48] and Possamai and Tangpi [50]. We would also like to mention the set valued approach in Iseri and Zhang [38], where the object is the set of values over all equilibria.

In the context of mean field controls, an important stream of works focuses on the convergence issue; see, for example, Bayraktar and Chakraborty [2], Cardaliaguet, Daudin, Jackson and Souganidis [7], Cardaliaguet and Souganidis [9], Cavagnari, Lisini, Orrieri and Savare

Received October 2022; revised November 2023.
MSC2020 subject classifications. Primary 60G40, 35Q89; secondary 49N80, 49L25, 65K15.
Key words and phrases. Mean field optimal stopping, obstacle problems, viscosity solutions, propagation of chaos.

[13], Cecchin [14], Daudin, Delarue and Jackson [17], Djete, Possamai and Tan [22], Fischer and Livieri [30], Fornasier, Lisini, Orrieri and Savare [32], Germain, Pham and Warin [36] and Lacker [42]. In particular, by utilizing some strong regularity of the value function, [7, 36] obtained certain rate of convergence. [17] also obtained rates of convergence by mollifying the value function of the mean field problem in order to obtain "almost" classical subsolutions, which are then projected on $\mathbb{R}^N$ to derive estimates for the N-particle problem. The work closest to ours is Gangbo, Mayorga and Swiech [34], which uses the uniqueness of viscosity solution (in a certain sense) to prove the convergence of the value function in a setting without idiosyncratic but with common noise. We also mention the paper Talbi [53] by the first author which applies the same approach of this paper to finite-dimensional approximation of mean field control problems, including the path dependent case.

The first goal of this paper is to apply the Barles and Sounganids [1] approach to prove the convergence of the value functions. To our best knowledge, this is the first work in the mean field literature to use the Barles–Sounganids approach. Roughly speaking, denoting by V the value function of the mean field optimal stopping problem and V^N the value function of the corresponding N-player problem, we shall prove that $\underline{V} := \liminf_{N\to\infty} V^N$ and $\overline{V} := \limsup_{N\to\infty} V^N$ are viscosity supersolution and subsolution, respectively, of our mean field obstacle equation. Then, it follows from the comparison principle of viscosity solutions, which was established in our previous paper [55], that $\underline{V} = \overline{V}$ and hence $\lim_{N\to\infty} V^N = V$. Note that our viscosity solution approach allows us to deal with value functions which are merely continuous. Consequently, besides the obvious difference that we consider mean field optimal stopping problem instead of mean field control problem, we require weaker regularity conditions for the coefficients than [7, 36].

Our convergence relies on the viscosity solution property of V^N to a finite-dimensional PDE, as deduced from the general solution of [39] expressed in terms of a recursive sequence of Snell envelopes. We shall refer to this equation as a cascade obstacle problem due to its remarkable structure. One subtle issue is the choice of test functions for viscosity solutions of this finite-dimensional PDE. To adapt to our notion of viscosity solution for our mean field obstacle equation in [55], we only require test functions to be tangent to the super/subsolution through the mean, whereas the tangency is pointwise in the standard literature. Our definition is inspired by the the viscosity theory developed for path-dependent PDEs; see, for example, the series of papers by Ekren, Keller, Ren, Touzi and Zhang [24–27, 51].

Our second main result is the propagation of chaos for stopped McKean–Vlasov diffusion (see, e.g., Sznitman [52] for the case of classical diffusions). By using the convergence of the value functions, we establish three results: roughly speaking,

- Any optimal strategy of the mean field problem can be approximated by (approximate) optimal strategies of the N-player problems;
- Any limit point of the (approximate) optimal strategies of the N-player problem is optimal for the mean field problem;
- If the optimal strategy of the mean field problem is unique, then any (approximate) optimal strategies of the N-player problem converges to that unique optimal strategy of the mean field problem.

These results are consistent with the convergence results for mean field equilibria in the mean field game literature; see, for example, [41]. We shall remark that, for mean field games, the N-player problems are quite different between closed loop controls and open loop controls. However, for our mean field optimal stopping problem (and for mean field control problems), the open loop and the closed loop stopping strategies typically lead to the same value function. In particular, for the convenience of establishing the regularity of the value functions, we consider open loop stopping times for the N-player problems.

The paper is organized as follows. In Sections 2 and 3 we introduce the mean field optimal stopping problem and the finite population optimal stopping problem, respectively. In Section 4 we present the main results of the paper. Sections 5 and 6 are devoted to the proofs of the main results, one on the convergence of the value functions, and the other on the propagation of chaos. In Section 7 we study the multiple optimal stopping problem in detail. Finally, we provide some technical proofs in the Appendix.

Notation. We denote by $\mathcal{P}(\Omega, \mathcal{F})$ the set of probability measures on a measurable space $(\Omega, \mathcal{F})$. When Ω is a topological space and $\mathcal{F}$ is the Borel σ-field, we simply denote them as $\mathcal{P}(\Omega)$ and, for $p \geq 1$, $\mathcal{P}_p(\Omega)$ the subset of probability measures in $\mathcal{P}(\Omega)$ with finite pth order moment, equipped with the p-Wasserstein distance $\mathcal{W}_p$. For a random variable Z on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $\mathbb{P}_Z := \mathbb{P} \circ Z^{-1}$ the law of Z under $\mathbb{P}$. The space of $d \times d$-symmetric matrice is denoted by $\mathcal{S}_d$, with $\mathcal{S}_d^+$ the subset of nonnegative matrices. For vectors $x, y \in \mathbb{R}^d$ and matrices $A, B \in \mathbb{R}^{d \times N}$, 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") "upper (resp. lower) semicontinuous". Moreover, denote

$$\mathbf{S} := \mathbb{R}^d \times \{0, 1\}, \qquad \mathbf{Q}_t := [t, T) \times \mathcal{P}_2(\mathbf{S}), \qquad \overline{\mathbf{Q}}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad t \in [0, T);$$

$$\|m\|_p^p := \int_{\mathbf{S}} |x|^p m(dy), \quad m \in \mathcal{P}_p(\mathbf{S}) \text{ for any } p \geq 1;$$

and, for any $N \geq 1$,

$$\begin{aligned} \Lambda_t^N &:= [t, T) \times \mathbf{S}^N, \qquad \overline{\Lambda}_t^N := [t, T] \times \mathbf{S}^N, \\ m^N(\mathbf{y}) &:= \frac{1}{N} \sum_{k=1}^N \delta_{y_k} \in \mathcal{P}(\mathbb{R}^d) \quad \text{for } \mathbf{y} = (y_1, \ldots, y_N) \in \mathbf{S}^N, \\ \mathcal{P}^N(\mathbf{S}) &:= \{m_N(\mathbf{y}) : \mathbf{y} \in \mathbf{S}^N\}, \qquad \mathbf{Q}_t^N := [t, T) \times \mathcal{P}^N(\mathbf{S}), \qquad \overline{\mathbf{Q}}_t^N := [t, T] \times \mathcal{P}^N(\mathbf{S}), \\ \|\mathbf{x}\|_p^p &:= \frac{1}{N} \sum_{k=1}^N |x_k|^p, \quad \mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^{d \times N}. \end{aligned} \tag{1.1}$$

2. The mean field optimal stopping problem. In this section we introduce the mean field optimal stopping problem and the related obstacle equation investigated in our previous works [54, 55].

2.1. *The mean field optimal stopping problem.* Let $T < \infty$, and denote $\Omega := C^0([-1, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$ the canonical space, with:

- $C^0([-1, T], \mathbb{R}^d)$ 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 recall that Ω is Polish under the Skorokhod distance. We also denote $Y := (X, I)$ the canonical process taking values in the state space $\mathbf{S}$, $\mathbb{F} = (\mathcal{F}_t)_{t \in [-1, T]}$ its canonical filtration, and the corresponding jump time of the survival process I:

$$\iota := \inf\{t \geq 0 : I_t = 0\} \quad \text{so that } I_t := I_{0-} \mathbf{1}_{t < \iota} \text{ for all } t \in [-1, T]. \tag{2.1}$$

Since I is càdlàg, ι is an $\mathbb{F}$-stopping time. We emphasise on the fact that the extension to -1 is arbitrary, as the extension of time to the left of the origin is only needed to allow for an immediate stop at time $t = 0$.

Let $(b, \sigma, f) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) \to \mathbb{R}^d \times \mathbb{S}_d^+ \times \mathbb{R}$, and $g : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$.

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

$$F(t,m) := \int_{\mathbb{R}^d} f(t,x,m)m(dx,1) \quad \text{is continuous on } [0,T] \times \mathcal{P}_2(\mathbf{S}). \tag{2.2}$$

(iii) g is upper-semicontinuous and locally bounded.

We also extend g to $\mathcal{P}_2(\mathbf{S})$ by $g(m) := g(m(\cdot, \{0,1\}))$. Let V be the dynamic value function:

$$\begin{aligned} V(t,m) &:= \sup_{\mathbb{P} \in \mathcal{P}(t,m)} J(t,\mathbb{P}), \quad (t,m) \in \overline{\mathbf{Q}}_0, \\ J(t,\mathbb{P}) &:= \int_t^T F(r, \mathbb{P}_{Y_r})\, dr + g(\mathbb{P}_{Y_T}), \end{aligned} \tag{2.3}$$

with $\mathcal{P}(t,m)$ the set of probability measures $\mathbb{P}$ on $(\Omega, \mathcal{F}_T)$ s.t. $\mathbb{P}_{Y_{t-}} = m$ and the processes:

$$M_\cdot := X_\cdot - \int_t^\cdot b(r, X_r, \mathbb{P}_{Y_r}) I_r\, dr \quad \text{and} \quad M_\cdot M_\cdot^\intercal - \int_t^\cdot \sigma^2(r, X_r, \mathbb{P}_{Y_r}) I_r\, dr \tag{2.4}$$

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

$$X_s = X_t + \int_t^s b(r, X_r, \mathbb{P}_{Y_r}) I_r\, dr + \sigma(r, X_r, \mathbb{P}_{Y_r}) I_r\, dW_r^{\mathbb{P}}, \quad I_s = I_{t-} \mathbf{1}_{s<t}, \mathbb{P}\text{-a.s.} \tag{2.5}$$

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

$$X_s = X_t + \int_t^s b(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r\, dr + \int_t^s \sigma(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r\, dW_r^{\bar{\mathbb{P}}}, \quad I_s = I_{t-} \mathbf{1}_{t \le s < T}, \bar{\mathbb{P}}\text{-a.s.} \tag{2.6}$$

Note that $Y_\cdot = Y_{\cdot \wedge t}$, and in particular $Y_T = Y_t$, $\mathbb{P}$-a.s. Moreover, from the definition of F in (2.2), we have $\int_t^T F(r, \mathbb{P}_{Y_r})\, dr = \mathbb{E}^{\mathbb{P}} \int_t^t f(r, X_r, \mathbb{P}_{Y_r})\, dr$.

We observe that (2.3) admits an optimal control as $\mathcal{P}(t,m)$ is $\mathcal{W}_2$-compact by [54].

2.2. *Obstacle equation on Wasserstein space.* We first 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

$$U(\tilde{m}) - U(m) = \int_0^1 \int_{\mathbf{S}} \delta_m U\big(l\tilde{m} + (1-l)m, y\big)(\tilde{m} - m)(dy)\, dl \quad \text{for all } m, \tilde{m} \in \mathcal{P}_2(\mathbf{S}),$$

and $\delta_m U(m, \cdot)$ has quadratic growth, locally uniformly in m, so as to guarantee integrability in the last expression. As in [54], we denote

$$\delta_m U_i(t,m,x) := \delta_m U(t,m,x,i) \quad \text{for } i \in \{0,1\}, \qquad D_I U := \delta_m U_1 - \delta_m U_0,$$

and we introduce the measure flow generator of X

$$\mathbb{L}U(t,m) := \partial_t U(t,m) + \int_{\mathbb{R}^d} \mathcal{L}_x \delta_m U_1(t,m,x) m(dx,1),$$

where $\mathcal{L}_x$ is the generator of X:

$$\mathcal{L}_x \delta_m U_1 := b \cdot \partial_x \delta_m U_1 + \frac{1}{2} \sigma^2 : \partial_{xx}^2 \delta_m U_1.$$

We may write $\partial_\mu U_{1,1}(t,m,x) := \partial_x \delta_m U(t,m,x,1)$ in the spirit of the Lions derivative, and we also introduce $\partial_{\mu\mu}^2 U_{1,1}(t,m,x,\tilde{x}) := \partial_{\tilde{x}} \delta_m \partial_x \delta_m U(t,m,x,1,\tilde{x},1) \in \mathbb{R}^{d \times d}$.

We next introduce a partial order $\preceq$ on $\mathcal{P}_2(\mathbf{S})$. We say that $m' \preceq m$ if $m'(dx,1)$ is absolutely continuous with respect to $m(dx,1)$ with density $p:\mathbb{R}^d \to [0,1]$, that is,

$$\begin{aligned} &m'(dx,1) = p(x)m(dx,1) \\ \text{and therefore}\quad &m'(dx,0) = \big[1-p(x)\big]m(dx,1) + m(dx,0). \end{aligned} \tag{2.7}$$

Roughly speaking, $m' \preceq m$ if the distribution m' can be obtained by immediately stopping particles from the distribution m. 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 $p(x) = p_t(x) := \mathbb{P}(I_t = 1 | X_t = x, I_{t-} = 1)$. When $p = \mathbf{1}_A$, with $A \in \mathcal{B}(\mathbb{R}^d)$, we denote m^A the probability measure defined by (2.7). We recall from [55] that the set $\{m' : m' \preceq m\}$ is compact. Moreover, we have the following simple lemma whose proof is postponed to the Appendix.

LEMMA 2.2. *Let $m \in \mathcal{P}_2(\mathbf{S})$ such that $x \in \mathbb{R}^d \to m((-\infty, x_1] \times \cdots \times (-\infty, x_d],\ 1)$ is continuous. Then $\{m^A : A \in \mathcal{B}(\mathbb{R}^d)\}$ is $\mathcal{W}_2$-dense in $\{m' : m' \preceq m\}$.*

In [54, 55], we see that V is associated with the obstacle problem on Wasserstein space:

$$\min\Big\{ \min_{m' \in C_U(t,m)} \big[-(\mathbb{L}U + F)(t,m')\big],\ (\mathbb{D}_I U)_*(t,m) \Big\} = 0, \quad (t,m) \in \mathbf{Q}_0, \tag{2.8}$$

with boundary condition $U|_{t=T} = g$. 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 upper semicontinuous, but may not be continuous in general, and the set

$$C_U(t,m) := \big\{ m' \preceq m : U(t,m') \geq U(t,m) \big\}, \quad (t,m) \in \mathbf{Q}_0,$$

indicates the set of positions at t which improve U by stopping the corresponding particles.

2.3. *Viscosity solutions.* 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 $\mathcal{N}_\delta(t,m)$ is compact, as the closure of a càdlàg $\mathcal{P}_2(\mathbf{S})$-valued graph. For a locally bounded function $U : \overline{\mathbf{Q}}_0 \longrightarrow \mathbb{R}$, we introduce its LSC and USC envelopes relatively to $\mathcal{P}(t,m)$, U_* and U^* respectively:

$$\begin{aligned} U_*(t,m) &:= \lim \inf_{(s,\tilde{m}) \to (t,m)} U(s,\tilde{m}), \\ U^*(t,m) &:= \lim \sup_{(s,\tilde{m}) \to (t,m)} U(s,\tilde{m}) \quad \text{for all } (t,m) \in \overline{\mathbf{Q}}_0, \end{aligned}$$

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. Moreover, the smooth functions are in the following sense.

DEFINITION 2.3. Let $C_2^{1,2}(\overline{\mathbf{Q}}_t)$ denote the set of continuous functions $U : \overline{\mathbf{Q}}_t \to \mathbb{R}$ such that:

- $\partial_t U$, $\delta_m U$, $\partial_x \delta_m U_1$, $\partial^2_{xx} \delta_m U_1$ exist, and are continuous in all variables,
- $\partial^2_{xx} \delta_m U_1$ is bounded in x, locally uniformly in (s,m),
- $\partial^2_{\mu\mu} U_{1,1}$ exists and $(s,m,x) \mapsto \partial^2_{\mu\mu} U_{1,1} U(s,m,x,x)$ is continuous and has quadratic growth in x, locally uniformly in (s,m).

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 2.4. Let $U : \overline{\mathbf{Q}}_0 \to \mathbb{R}$ be locally bounded.

(i) U is a viscosity supersolution of (2.8) if, for any $(t,m) \in \mathbf{Q}_0$ and $m' \preceq m$,

$$U_*(t,m) \geq U_*(t,m') \quad \text{and} \quad -(\mathbb{L}\varphi + F)(t,m) \geq 0 \quad \text{for all } \varphi \in \overline{\mathcal{A}}U(t,m). \tag{2.9}$$

(ii) U is a viscosity subsolution of (2.8) if, for any $(t,m) \in \mathbf{Q}_0$ s.t. $C_{U^*}(t,m) = \{m\}$,

$$\min\big\{-(\mathbb{L}\varphi + F), (\mathbb{D}_I\varphi)_*\big\}(t,m) \leq 0 \quad \text{for all } \varphi \in \underline{\mathcal{A}}U(t,m). \tag{2.10}$$

(iii) U is a viscosity solution of (2.8) if it is a viscosity supersolution and subsolution.

REMARK 2.5. In the present paper, the set $C_2^{1,2}(\overline{\mathbf{Q}}_t)$ has a different definition from the one we used in our previous work [55], as we additionally require the existence of the second order derivative $\partial^2_{\mu\mu}U_{1,1}$. However, we can see that the viscosity theory implied by requiring such regularity for the test functions is equivalent to the one developed in [55]; in particular, the test functions involved in the proof of our comparison result [55], Theorem 3.11, are infinitely differentiable, so that the validity of this result does not depend on the definition of $C_2^{1,2}(\overline{\mathbf{Q}}_t)$ we set. □

Recall the assumptions for our comparison result [55], Theorem 3.11. Together with [55], Theorem 3.9, and Remark 2.5, this shows that V is the unique viscosity solution of (2.8).

ASSUMPTION 2.6.

(i) b and σ can be extended to $\mathcal{P}_1(\mathbf{S})$ under $\mathcal{W}_1$ continuously, and are uniformly Lipschitz continuous in (x,m) under $\mathcal{W}_1$;

(ii) f and g extend continuously to $\mathcal{P}_1(\mathbf{S})$ and have linear growth in (x,m), under $\mathcal{W}_1$;

(iii) f is uniformly continuous in (t,x,m), under $\mathcal{W}_2$ for m; and all first and second order derivatives of $\sigma(t,x,m)$ with respect to x and those of $\delta_m\sigma(t,x,m,\tilde{x},\tilde{i})$ with respect to $(x,\tilde{x})$ are bounded, continuous in t, and Lipschitz continuous in $(x,\tilde{x},m)$, under $\mathcal{W}_1$ for m.

3. The N-player optimal stopping problem. In this section we introduce a general N-player optimal stopping problem, following Kobylanski, Quenez and Rouy-Mironescu [39]. In the symmetric case, this reduces to the N-player problem corresponding to the mean field problem (2.3), as we will see in Section 3.3.

3.1. *The general N-player optimal stopping problem.* Let $(\Omega^0, \mathcal{F}_T^0, \mathbb{F}^0, \mathbb{P}^0)$ be a filtered probability space, equipped with a sequence of d-dimensional Brownian motions $\{W^k\}_{k\geq 1}$. For each $N \geq 1$, let $\mathbb{F}^N$ denote the canonical filtration of $\{W^k\}_{1\leq k\leq N}$, $\mathcal{T}_{t,T}^N$ the set of $[t,T]$-valued $\mathbb{F}^N$-stopping times for all $t \in [0,T]$, $\boldsymbol{\mathcal{T}}_{t,T}^N := (\mathcal{T}_{t,T}^N)^N$, $[N] := \{1,\dots,N\}$. We emphasize that, unlike the weak formulation in the previous section, here we use strong formulation. In particular, the stopping times in $\mathcal{T}_{t,T}^N$ are adapted to the Brownian motion filtration, in the spirit of open loop controls. Moreover, in this section N is fixed, but in the next section we will send N to ∞.

Let $(b_k, \sigma_k, f_k) : [0,T] \times \mathbf{S}^N \longrightarrow \mathbb{R}^d \times \mathcal{S}_d^+(\mathbb{R}) \times \mathbb{R}$, $k \in [N]$, and $g_N : \mathbb{R}^{d\times N} \to \mathbb{R}$ satisfy:

ASSUMPTION 3.1.

(i) The functions b_k, σ_k, $k \in [N]$, are continuous in t and uniformly Lipschitz continuous in $\mathbf{x}$ (with the Lipschitz constant possibly depending on N);

(ii) The functions f_k, $k \in [N]$, and g_N are continuous in t and uniformly continuous in $\mathbf{x}$ (with the uniform continuity possibly depending on N).

Fix $t \in [0,T)$ and $\mathbf{y} = (\mathbf{x}, \mathbf{i}) \in \mathbf{S}^N$. For $\boldsymbol{\tau} := (\tau_1, \dots, \tau_N) \in \mathcal{T}_{t,T}^N$, let $\mathbf{X}^{\boldsymbol{\tau}} = \mathbf{X}^{N,\boldsymbol{\tau}} = (X^{\boldsymbol{\tau},1}, \dots, X^{\boldsymbol{\tau},N})$ and $\mathbf{I}^{\boldsymbol{\tau}} = \mathbf{I}^{N,\boldsymbol{\tau}} = (I^{\boldsymbol{\tau},1}, \dots, I^{\boldsymbol{\tau},N})$ solve the following system of SDEs:

$$X_s^{\boldsymbol{\tau},k} = x_k + \int_t^s b_k\big(r, \mathbf{Y}_r^{\boldsymbol{\tau}}\big) I_r^{\boldsymbol{\tau},k}\, dr + \int_t^s \sigma_k\big(r, \mathbf{Y}_r^{\boldsymbol{\tau}}\big) I_r^{\boldsymbol{\tau},k}\, dW_r^k, \quad I_s^{\boldsymbol{\tau},k} = i_k \mathbf{1}_{\tau_k > s}, \ \mathbb{P}^0\text{-a.s.} \tag{3.1}$$

for $k \in [N]$, where $\mathbf{Y}^{\boldsymbol{\tau}} = \mathbf{Y}^{N,\boldsymbol{\tau}} := (\mathbf{X}^{\boldsymbol{\tau}}, \mathbf{I}^{\boldsymbol{\tau}})$. Here and throughout the paper, the dependence of $\mathbf{Y}$ on N is omitted for notational simplicity. Moreover, in light of (2.6), let $\bar{\mathbf{Y}} = (\bar{\mathbf{X}}, \bar{\mathbf{I}})$ denote the unstopped version, namely $\tau_k = T$, or equivalently $\bar{I}_s^k = i_k$, for all $k \in [N]$.

We now consider the problem

$$\begin{aligned} v^N(t,\mathbf{y}) &:= \sup_{\boldsymbol{\tau} \in \mathcal{T}_{t,T}^N} J_N(t,\mathbf{y},\boldsymbol{\tau}), \quad (t,\mathbf{y}) \in \overline{\Lambda}_0^N, \\ \text{where} \quad J_N(t,\mathbf{y},\boldsymbol{\tau}) &:= \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N \int_t^T f_k\big(r, \mathbf{Y}_r^{\boldsymbol{\tau}}\big) I_r^{\boldsymbol{\tau},k}\, dr + g_N\big(\mathbf{Y}_T^{\boldsymbol{\tau}}\big)\bigg]. \end{aligned} \tag{3.2}$$

Here the subscript $(t,\mathbf{y})$ in $\mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0}[\cdot] := \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0}[\cdot | \mathbf{Y}_{t-}^{\boldsymbol{\tau}} = \mathbf{y}]$ indicates the initial condition.

3.2. *Viscosity solutions with tangency in mean.* The function v^N is associated with the following partial differential equation on $\overline{\Lambda}_0^N$:

$$\min\Big\{-\partial_t u(t,\mathbf{y}) - \mathbf{i}\cdot(\mathcal{L}u + \mathbf{f})(t,\mathbf{y}), u(t,\mathbf{y}) - \max_{\mathbf{i}'<\mathbf{i}} u\big(t,\mathbf{x},\mathbf{i}'\big)\Big\} = 0, \tag{3.3}$$

with $\mathbf{y} = (\mathbf{x},\mathbf{i})$ and boundary condition $u|_{t=T} = u|_{\mathbf{i}=0} = g_N$, where $\mathbf{i}' < \mathbf{i}$ means $i_k' \le i_k$, $k \in [N]$, and $\mathbf{i} \ne \mathbf{i}'$, $\mathbf{f} := (f_1, \dots, f_N)$, and

$$\mathcal{L}\varphi := \Big(b_k \cdot \partial_{x_k}\varphi + \frac{1}{2}\sigma_k^2 : \partial_{x_k x_k}^2 \varphi\Big)_{k\in[N]} \quad \text{for all } \varphi \in C^2\big(\mathbb{R}^{d\times N}\big). \tag{3.4}$$

We call (3.3) *cascade obstacle equation*. We consider the notion of viscosity solutions introduced by Ekren, Keller, Touzi and Zhang [25] where, for any function $u : \overline{\Lambda}_0^N \to \mathbb{R}$, the corresponding set of test functions at some point $(t,\mathbf{y}) \in \overline{\Lambda}_0^N$ is defined by a tangency mean rather than the standard pointwise tangency of Crandall and Lions:

$$\begin{aligned} &\overline{\mathcal{A}}u(t,\mathbf{y}) \\ &\quad := \Big\{\phi \in C^{1,2}\big(\overline{\Lambda}_t^N\big) : \exists \mathrm{H} \in \mathcal{T}_{t,T}^+,\ (\phi - u)(t,\mathbf{y}) = \max_{\theta \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0}\big[(\phi - u)(\theta \wedge \mathrm{H}, \bar{\mathbf{Y}}_{\theta\wedge\mathrm{H}})\big]\Big\}, \\ &\underline{\mathcal{A}}u(t,\mathbf{y}) \\ &\quad := \Big\{\phi \in C^{1,2}\big(\overline{\Lambda}_t^N\big) : \exists \mathrm{H} \in \mathcal{T}_{t,T}^+,\ (\phi - u)(t,\mathbf{y}) = \min_{\theta \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0}\big[(\phi - u)(\theta \wedge \mathrm{H}, \bar{\mathbf{Y}}_{\theta\wedge\mathrm{H}})\big]\Big\}, \end{aligned} \tag{3.5}$$

with $\mathcal{T}_{t,T}^+ := \{\mathrm{H} \in \mathcal{T}_{t,T}^N : \mathrm{H} > t\}$, and $C^{1,2}(\overline{\Lambda}_t^N)$ denoting the set of continuous test functions ϕ such that: for any $\mathbf{i} = (i_1, \dots, i_N)$ and for all $k \in [N]$ with $i_k = 1$, the derivatives $\partial_t\phi(s,\mathbf{x},\mathbf{i})$, $\partial_{x_k}\phi(s,\mathbf{x},\mathbf{i})$, $\partial_{x_k,x_k}\phi(s,\mathbf{x},\mathbf{i})$ exist and are continuous for $(s,\mathbf{x}) \in [t,T] \times \mathbb{R}^{d\times N}$.

DEFINITION 3.2. Let $u : \overline{\Lambda}_0^N \longrightarrow \mathbb{R}$.

(i) u is a viscosity supersolution of (3.3) if, for all $(t, \mathbf{y}) \in \overline{\Lambda}_0^N$,

$$\min\Big\{-\partial_t\phi(t,\mathbf{y}) - (\mathcal{L}\phi + \mathbf{f})(t,\mathbf{y})\cdot\mathbf{i}, u(t,\mathbf{y}) - \max_{\mathbf{i}'<\mathbf{i}} u\big(t,\mathbf{x},\mathbf{i}'\big)\Big\} \geq 0 \quad \text{for all } \phi \in \overline{\mathcal{A}}u(t,\mathbf{y}).$$

(ii) u is a viscosity subsolution of (3.3) if, for all $(t, \mathbf{y}) \in \Lambda_0^N$,

$$\min\Big\{-\partial_t\phi(t,\mathbf{y}) - (\mathcal{L}\phi + \mathbf{f})(t,\mathbf{y})\cdot\mathbf{i}, u(t,\mathbf{y}) - \max_{\mathbf{i}'<\mathbf{i}} u\big(t,\mathbf{x},\mathbf{i}'\big)\Big\} \leq 0 \quad \text{for all } \phi \in \underline{\mathcal{A}}u(t,\mathbf{y}).$$

(iii) u is a viscosity solution of (3.3) if it is a viscosity supersolution and subsolution.

We then have the following result, whose proof is postponed to §7.

THEOREM 3.3. *Under Assumption 3.1, the function v^N in (3.2) is the unique continuous viscosity solution of (3.3) with boundary conditions $v^N|_{t=T} = v^N|_{\mathbf{i}=0} = g$.*

REMARK 3.4. We note that a viscosity solution in our sense is always a viscosity solution in the Crandall–Lions sense, as the latter involves a smaller class of test functions. In particular, the comparison principle and the uniqueness of viscosity solution in the latter sense implies the same results in our sense. Our definition turns out to be more convenient for our proof of the convergence results in the next section. □

3.3. *The symmetric case.* Recall the functions b, σ, f, g in Section 2 and the notation m^N in (1.1). In this subsection we consider the symmetric case:

$$\varphi_k(t,\mathbf{y}) = \varphi\big(t, x_k, m^N(\mathbf{y})\big) \quad \text{for } \varphi = b, \sigma, f, \quad \text{and} \quad g_N(\mathbf{y}) = g\big(m^N(\mathbf{y})\big). \tag{3.6}$$

We first have the following simple lemma, whose proof is postponed to Section 7.

LEMMA 3.5.

(i) *Under Assumption 2.1(i) and (3.6), there is a constant $C \geq 0$, independent of N, such that, for $(t, \mathbf{y}) \in \Lambda_0^N$, $\boldsymbol{\tau} \in \mathcal{T}_{t,T}^N$, $k \in [N]$, and $\theta_1, \theta_2 \in \mathcal{T}_{[t,T]}^N$ with $\theta_1 \leq \theta_2$,*

$$\text{and} \quad \mathbb{E}\Big[\sup_{t\leq s\leq T}\big|X_s^{\boldsymbol{\tau},k}\big|^2\Big] \leq C\big(1 + |x_k|^2 + \|\mathbf{x}\|_2^2\big), \tag{3.7}$$

$$\text{and} \quad \mathbb{E}\Big[\sup_{\theta_1\leq s\leq\theta_2}\big|X_s^{\boldsymbol{\tau},k} - X_{\theta_1}^{\boldsymbol{\tau},k}\big|^2\Big] \leq C\big(1 + |x_k|^2 + \|\mathbf{x}\|_2^2\big)\|\theta_2 - \theta_1\|_\infty. \tag{3.8}$$

(ii) *If we assume further that Assumption 2.6(i) holds, then*

$$\text{and} \quad \mathbb{E}\Big[\sup_{t\leq s\leq T}\big|X_s^{\boldsymbol{\tau},k}\big|\Big] \leq C\big(1 + |x_k| + \|\mathbf{x}\|_1\big), \tag{3.9}$$

$$\text{and} \quad \mathbb{E}\Big[\sup_{\theta_1\leq s\leq\theta_2}\big|X_s^{\boldsymbol{\tau},k} - X_{\theta_1}^{\boldsymbol{\tau},k}\big|\Big] \leq C\big(1 + |x_k| + \|\mathbf{x}\|_1\big)\sqrt{\|\theta_2 - \theta_1\|_\infty}. \tag{3.10}$$

Next, under (3.6) one can easily see that v^N is also symmetric in $\mathbf{y}$, that is, there exists a function $V^N : \overline{\mathbf{Q}}_0^N \to \mathbb{R}$ such that

$$v^N(t,\mathbf{y}) = V^N\big(t, m^N(\mathbf{y})\big). \tag{3.11}$$

DEFINITION 3.6. Under (3.6), we call a function $U^N : \overline{\mathbf{Q}}_0^N \to \mathbb{R}$ a symmetric viscosity solution (resp. subsolution, supersolution) of (3.3) if $u^N(t,\mathbf{y}) := U^N(t, m^N(\mathbf{y}))$ is a viscosity solution (resp. subsolution, supersolution) of (3.3) in the sense of Definition 3.2.

The following result establishes stronger regularity than Theorem 3.3 and its proof is also postponed to §7.

THEOREM 3.7. *Let Assumptions 2.1, 2.6(i)–(ii) and (3.6) hold.*

(i) *The function v^N is the unique continuous symmetric viscosity solution of (3.3).*

(ii) *There exist a constant C, independent of N, such that*

$$\big|V^N(t,m_N)\big| \le C\big(1+\|m_N\|_1\big). \tag{3.12}$$

(iii) *For any $R>0$, there exist a modulus of continuity function ρ_R, which may depend on R but not on N, such that, for any $(t,m_N),(\tilde t,\tilde m_N)\in\overline{\mathbf{Q}}_0^N$ with $\|m_N\|_2,\|\tilde m_N\|_2\le R$,*

$$\big|V^N(t,m_N)-V^N(\tilde t,\tilde m_N)\big| \le \rho_R\big(\mathcal{W}_2(m_N,\tilde m_N)+|\tilde t-t|\big). \tag{3.13}$$

Moreover, similar to Carmona and Delarue ([11], Vol. 1, Propositions 5.35 and 5.91), we have the following result, and we shall provide a proof in the Appendix.

LEMMA 3.8. *Assume $\varphi\in C_2^{1,2}(\overline{\mathbf{Q}}_0)$ and $\phi(t,\mathbf{y}):=\varphi(t,m^N(\mathbf{y}))$ for all $(t,\mathbf{y})=(t,\mathbf{x},\mathbf{i})\in[0,T]\times\mathbf{S}^N$. Then, for all $(s,\mathbf{y})\in\overline{\Lambda}_0^N$ and $k\in[N]$ such that $i_k=1$,*

$$\begin{aligned}\partial_{x_k}\phi(s,\mathbf{y}) &= \frac{1}{N}\partial_x\delta_m\varphi_1\big(s,m^N(\mathbf{y}),x_k\big),\\ \partial^2_{x_kx_k}\phi(s,\mathbf{y}) &= \frac{1}{N}\partial^2_{xx}\delta_m\varphi_1\big(s,m^N(\mathbf{y}),x_k\big)+\frac{1}{N^2}\partial^2_{\mu\mu}\varphi_{1,1}\big(s,m^N(\mathbf{y}),x_k,x_k\big).\end{aligned} \tag{3.14}$$

4. Main results. Our first main result states the convergence of the value function of the multiple optimal stopping problem to the corresponding mean field problem.

THEOREM 4.1. *Let Assumptions 2.1 and 2.6 hold, and V^N and V be respectively the value functions of the multiple and mean field optimal stopping problems (3.2)–(3.6)–(3.11) and (2.3). Then V^N converges to V, that is,*

$$\lim_{\substack{N\to\infty,s\to t\\ m_N\in\mathcal{P}^N(\mathbf{S})\xrightarrow{\mathcal{W}_2}m}} V^N(s,m_N)=V(t,m) \quad \text{for all } (t,m)\in\overline{\mathbf{Q}}_0. \tag{4.1}$$

The above theorem relies heavily on the following result. For a sequence of functions $U^N:\overline{\mathbf{Q}}_0^N\to\mathbb{R}$ which are locally bounded functions, uniformly in N, we denote

$$\begin{aligned}\underline{U}(t,m) &:= \lim \inf_{\substack{N\to\infty,s\to t\\ m_N\in\mathcal{P}^N(\mathbf{S})\xrightarrow{\mathcal{W}_2}m}} U^N(s,m_N) \quad \text{and}\\ \overline{U}(t,m) &:= \lim \sup_{\substack{N\to\infty,s\to t\\ m_N\in\mathcal{P}^N(\mathbf{S})\xrightarrow{\mathcal{W}_2}m}} U^N(s,m_N),\end{aligned} \tag{4.2}$$

for all $(t,m)\in\overline{\mathbf{Q}}_0$. We then have

THEOREM 4.2. *Let Assumption 2.1 and (3.6) hold and $U^N:\overline{\mathbf{Q}}_0^N\to\mathbb{R}$ be a sequence of continuous and locally bounded functions, uniformly in N.*

(i) *If each U^N is a viscosity subsolution of (3.3), then $\overline{U}$ is a viscosity subsolution of (2.8).*

(ii) *If each U^N is a viscosity supersolution of* (3.3), *and* $\underline{U}$ *is continuous, then* $\underline{U}$ *is a viscosity supersolution of* (2.8).

The above two theorems are proved in Section 5. We now turn to the convergence of the optimal strategies. Our second main result is as follows.

THEOREM 4.3. *Let Assumptions* 2.1 *and* 2.6 *hold. Fix* $(t,m) \in \mathbf{Q}_0$ *and* $\mathbf{y}^N \in \mathbf{S}^N$, $N \geq 1$, *such that* $m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m$. *Recall* (3.1) *and assume* (3.6) *holds.*

(i) *Denote* $m^N(\mathbf{Y}) := \frac{1}{N}\sum_{k=1}^N \delta_{Y^k} \in \mathcal{P}_2(\Omega)$. *Assume* $\hat{\boldsymbol{\tau}}^N \in \boldsymbol{\mathcal{T}}_{t,T}^N$ *is* ε_N*-optimal for* (3.2) *for all* $N \geq 1$, *with* $\varepsilon_N \to 0$. *Then,* $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}))^{-1}\}_{N\geq 1}$ *is tight, and all its accumulation points are supported on the set of the optimal controls for* (2.3).

(ii) *Let* $\mathbb{P}^* \in \mathcal{P}(t,m)$ *be optimal for* (2.3). *Then there exist* $\varepsilon_N \to 0$ *and* $\hat{\boldsymbol{\tau}}^N \in \boldsymbol{\mathcal{T}}_{t,T}^N$ *which is* ε_N*-optimal for* (3.2) *such that* $m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}) \xrightarrow{\mathcal{W}_2} \mathbb{P}^*$, $\mathbb{P}^0$*-a.s.*

(iii) *Assume further that* (2.3) *has a unique optimal control* $\mathbb{P}^*$. *Then for any sequence of* ε_N*-optimal stopping strategies* $\{\hat{\boldsymbol{\tau}}^N\}_{N\geq 1}$ *for* (3.2), $m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}) \xrightarrow{\mathcal{W}_2} \mathbb{P}^*$, $\mathbb{P}^0$*-a.s.*

Since N varies here, we use the superscript ^{N} in $\hat{\boldsymbol{\tau}}^N$ to indicate its dependence on N.

Theorem 4.3 relies heavily on the following result concerning the propagation of chaos for arbitrary stopping strategies. Both results are proved in Section 6.

THEOREM 4.4. *Let Assumption* 2.1 *hold, and fix* $(t,m) \in \mathbf{Q}_0$ *and* $\mathbf{y}^N \in \mathbf{S}^N$, $N \geq 1$, *such that* $m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m$. *Recall* (3.1) *and assume* (3.6) *holds.*

(i) *For any* $\boldsymbol{\tau}^N \in \boldsymbol{\mathcal{T}}_{t,T}^N$, *the sequence* $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}))^{-1}\}_{N\geq 1}$ *is tight, and all its accumulation points are supported on* $\mathcal{P}(t,m)$.

(ii) *Assume further that* b, σ *are uniformly Lipschitz continuous in* m *under* $\mathcal{W}_1$. *Then, for any* $\mathbb{P} \in \mathcal{P}(t,m)$, *there exists* $\boldsymbol{\tau}^N \in \boldsymbol{\mathcal{T}}_{t,T}^N$, $N \geq 1$, *such that* $m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}) \xrightarrow{\mathcal{W}_2} \mathbb{P}$, $\mathbb{P}^0$*-a.s.*

5. Convergence of the value function. In this section we prove Theorems 4.1 and 4.2.

PROOF OF THEOREM 4.1. (given Theorem 4.2). First, for any $(t,m) \in \overline{\mathbf{Q}}_0$ and $(t^N, m_N(\mathbf{y}^N)) \to (t,m)$ as $N \to \infty$, the sequence $\{V^N(t^N, m_N(\mathbf{y}^N))\}_{N\geq 1}$ is bounded by (3.12). Then the corresponding functions $\overline{V}$, $\underline{V}$ defined as in (4.2) are finite, and are continuous by Theorem 3.7(iii). In particular, $\overline{V}(T,\cdot) = \underline{V}(T,\cdot) = g$. Applying Theorem 4.2 we see that $\overline{V}$ and $\underline{V}$ are a viscosity subsolution and supersolution of (2.8). Then it follows from the comparison principle [55], Theorem 3.11, that $\underline{V} \geq \overline{V}$. As the converse inequality holds by definition, this shows that $\underline{V} = \overline{V}$, and by uniqueness, they are equal to V and thus (4.1) holds. □

PROOF OF THEOREM 4.2. Without loss of generality, we prove the theorem only at $t = 0$.

(i) Fix $m \in \mathcal{P}_2(\mathbf{S})$ and $\varphi \in \underline{\mathcal{A}}\overline{U}(0,m)$ with corresponding $\delta_0 \in (0,T)$. We assume w.l.o.g. (see [55], Remark 3.2) that

(5.1) $\quad (0,m)$ is a strict minimum of $(\varphi - \overline{U})$ on $\mathcal{N}_{\delta_0}(0,m)$, with $C_{\overline{U}}(0,m) = \{m\}$.

Let $(t^N, \mathbf{y}^N)$ be such that

(5.2) $\quad (t^N, m^N(\mathbf{y}^N)) \longrightarrow (0,m), \quad$ and $\quad U^N(t^N, m^N(\mathbf{y}^N)) \longrightarrow \overline{U}(0,m) \quad$ as $N \to \infty$.

We also introduce the functions $\phi^N(s,\mathbf{y}) := \varphi(s, m^N(\mathbf{y}))$ for all $N \geq 1$ and $(s,\mathbf{y}) \in \overline{\Lambda}_{t^N}^N$.

Step 1: We prove that, without loss of generality, $\{\mathbf{y}^N = (\mathbf{x}^N, \mathbf{i}^N)\}_{N\geq 1}$ may be taken s.t.

$$U^N(t^N, \mathbf{y}^N) - \max_{\mathbf{i}' < \mathbf{i}^N} U^N(t^N, \mathbf{x}^N, \mathbf{i}') > 0 \quad \text{for all } N \geq 1. \tag{5.3}$$

Indeed, for each $N \geq 1$, if $\mathbf{i}^N = 0$, then the maximum is equal to $-\infty$ and thus (5.3) is trivially satisfied. Now assume $\mathbf{i}^N \neq 0$. Then there exists $\mathbf{i}^{N,*} \leq \mathbf{i}^N$ attaining the maximum $\max_{\mathbf{i}' \leq \mathbf{i}^N} U^N(t^N, \mathbf{x}^N, \mathbf{i}')$. Note that $\mathbf{i}^{N,*}$ can be chosen to be minimal in $\operatorname{argmax}_{\mathbf{i}' \in \mathbf{i}^N} U^N(t^N, \mathbf{x}^N, \mathbf{i}')$ for the partial order on $\{0,1\}^N$. This implies that $\mathbf{y}^{N,*} := (\mathbf{x}^N, \mathbf{i}^{N,*})$ satisfies (5.3).

It remains to show that $\mathbf{y}^{N,*}$ also satisfies (5.2). First, as $m^N(\mathbf{y}^N)$ converges, the first marginal of $m^N(\mathbf{y}^{N,*})$ converges. Next, since its second marginal is a measure on $\{0,1\}$, the sequence $\{m^N(\mathbf{y}^{N,*})\}_{N\geq 1}$ is tight, and we may thus extract a subsequence (still denoted the same) converging to some $m^* \preceq m$. Then, noting that $U^N(t^N, \mathbf{y}^N) \leq U^N(t^N, \mathbf{y}^{N,*})$, by taking the $\liminf_{N\to\infty}$ we obtain

$$\begin{aligned}\overline{U}(0,m) &= \lim_{N\to\infty} U^N\big(t^N, m^N(\mathbf{y}^N)\big) \leq \liminf_{N\to\infty} U^N\big(t^N, m(\mathbf{y}^{N,*})\big) \\ &\leq \limsup_{N\to\infty} U^N\big(t^N, m(\mathbf{y}^{N,*})\big) \leq \overline{U}(0,m^*).\end{aligned}$$

Then $m^* \in C_{\overline{U}}(0,m)$, and therefore $m^* = m$ by (5.1), so that equality holds everywhere and $U^N(t^N, m^N(\mathbf{y}^{N,*})) \longrightarrow_{N\to\infty} \overline{U}(0,m)$. Consequently, $\{\mathbf{y}^{N,*}\}_{N\geq 1}$ satisfies (5.2), and we may replace $\{\mathbf{y}^N\}_{N\geq 1}$ with $\{\mathbf{y}^{N,*}\}_{N\geq 1}$, which satisfies in addition (5.3).

Step 2: Fix $\delta \in (0, \frac{\delta_0}{3})$. By standard optimal stopping theory, there exists $\theta_\delta^N \in \mathcal{T}_{t^N,T}$ s.t.

$$\mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[(\phi^N - U^N)(\theta_\delta^N \wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}^N_{\theta_\delta^N \wedge \mathrm{H}_\delta^N})\big] = \inf_{\theta \in \mathcal{T}_{t^N,T}} \mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[(\phi^N - U^N)(\theta \wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}^N_{\theta \wedge \mathrm{H}_\delta^N})\big],$$

where $\bar{\mathbf{Y}}^N := (\bar{\mathbf{X}}^N, \bar{\mathbf{i}}^N)$ is the unstopped version of (3.1) starting from $(t^N, \mathbf{y}^N)$, and

$$\mathrm{H}_\delta^N := \inf\big\{s \geq t^N : \mathcal{W}_2\big(m^N(\bar{\mathbf{Y}}_s), m^N(\mathbf{y}^N)\big) = 2\delta\big\} \wedge \big(t^N + 2\delta\big).$$

Since $\{U^N\}_{N\geq 1}$ is locally bounded, we may assume w.l.o.g. that N is large enough and δ is small enough so that $\{(\phi^N - U^N)(\theta_\delta^N \wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}^N_{\theta_\delta^N \wedge \mathrm{H}_\delta^N})\}_{N\geq 1}$ is uniformly bounded.

Step 3: We next show that, for any δ,

$$\lim_{N\to\infty} \sup \mathbb{P}^0\big(\theta_\delta^N < \mathrm{H}_\delta^N\big) > 0. \tag{5.4}$$

Indeed, assume to the contrary that $\limsup_{N\to\infty} \mathbb{P}^0(\theta_\delta^N < \mathrm{H}_\delta^N) = 0$. Then

$$\begin{aligned}(\varphi - \overline{U})(0,m) &= \lim_{N\to\infty} (\varphi - U^N)\big(t^N, m^N(\mathbf{y}^N)\big) \\ &\geq \limsup_{N\to\infty} \mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[(\varphi - U^N)\big(\theta_\delta^N \wedge \mathrm{H}_\delta^N, m^N\big(\bar{\mathbf{Y}}^N_{\theta_\delta^N \wedge \mathrm{H}_\delta^N}\big)\big)\big] \\ &\geq \liminf_{N\to\infty} \mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[(\varphi - U^N)\big(\theta_\delta^N \wedge \mathrm{H}_\delta^N, m^N\big(\bar{\mathbf{Y}}^N_{\theta_\delta^N \wedge \mathrm{H}_\delta^N}\big)\big)\big] \\ &= \liminf_{N\to\infty} \mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[\big\{(\varphi - U^N)\big(\mathrm{H}_\delta^N, m^N\big(\bar{\mathbf{Y}}^N_{\mathrm{H}_\delta^N}\big)\big)\big(1 - \mathbf{1}_{\{\theta_\delta^N < \mathrm{H}_\delta^N\}}\big) \\ &\quad + (\varphi - U^N)\big(\theta_\delta^N, m^N\big(\bar{\mathbf{Y}}^N_{\theta_\delta^N}\big)\big)\mathbf{1}_{\{\theta_\delta^N < \mathrm{H}_\delta^N\}}\big\}\big] \\ &= \liminf_{N\to\infty} \mathbb{E}^{\mathbb{P}^0}_{t^N,\mathbf{y}^N}\big[(\varphi - \mathrm{U}^N)\big(\mathrm{H}_\delta^N, m^N\big(\bar{\mathbf{Y}}^N_{\mathrm{H}_\delta^N}\big)\big)\big],\end{aligned} \tag{5.5}$$

where we use the uniform boundedness of $\{(\phi^N - U^N)(\theta \wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}^N_{\theta \wedge \mathrm{H}_\delta^N})\}_{N\geq 1}$.

Recall that $m^N(\bar{\mathbf{Y}}^N) \in \mathcal{P}_2(\Omega)$, $\mathbb{P}^0$-a.s. Since $(t^N, m^N(\mathbf{y}^N)) \longrightarrow_{N\to\infty} (0,m)$ and $\bar{\mathbf{Y}}^N$ is an unstopped diffusion whose coefficients satisfy the usual Lipschitz conditions, by classical propagation of chaos (see Oelschläger [49]) and compactness of $[0,T]$, again after possibly passing to a subsequence, we see that $\nu^N := \mathbb{P}^0 \circ (\mathrm{H}_\delta^N, m^N(\bar{\mathbf{Y}}^N))^{-1}$ converges weakly to some element of $\nu \in \mathcal{P}_2([0,T]\times\mathcal{P}_2(\Omega))$. Let (τ,λ) denote the canonical mapping on $[0,T]\times\mathcal{P}_2(\Omega)$. Recall that Y is the canonical mapping on Ω, and since $\bar{\mathbf{Y}}^N$ is unstopped, then Y is unstopped, $m^N(\bar{\mathbf{Y}}^N)(\omega^0)$-a.s., for $\mathbb{P}^0$-a.e. $\omega^0 \in \Omega^0$. This implies that Y is unstopped, $\lambda(\overline{\omega})$-a.s., for ν-a.e. $\overline{\omega} \in \mathcal{P}_2([0,T]\times\mathcal{P}_2(\Omega))$, and thus the mapping $(\tau,\lambda) \mapsto \lambda_{Y_\tau}$ (the λ-distribution of the random variable Y_τ) is continuous, ν-a.s. Then, by the definition of $\overline{U}$, together with Fatou's lemma, we deduce from (5.5) that

$$(\varphi - \overline{U})(0,m) \ge \lim \inf_{N\to\infty} \mathbb{E}^{\nu^N}\big[(\varphi - U^N)(\tau,\lambda_{Y_\tau})\big] \ge \mathbb{E}^{\nu}\big[(\varphi-\overline{U})(\tau,\lambda_{Y_\tau})\big]. \tag{5.6}$$

Thus there exists $\bar{\omega} \in [0,T]\times\mathcal{P}_2(\Omega)$ such that Y is unstopped under $\lambda(\bar{\omega})$, and

$$(\varphi - \overline{U})(0,m) \ge (\varphi-\overline{U})\big(\tau(\bar{\omega}), [\lambda(\bar{\omega})]_{Y_{\tau(\bar{\omega})}}\big).$$

Moreover, we may choose $\bar{\omega}$ so that

$$\mathcal{W}_2\big([\lambda(\bar{\omega})]_{Y_{\tau(\bar{\omega})}}, m\big) \vee \tau(\bar{\omega}) \le 3\delta < \delta_0, \qquad \mathcal{W}_2\big([\lambda(\bar{\omega})]_{Y_{\tau(\bar{\omega})}}, m\big) + \tau(\bar{\omega}) \ge \delta,$$

as we have by the definition of H_δ^N, for all $N \ge 1$ that

$$\mathcal{W}_2\big(\lambda_{Y_\tau}, m^N(\mathbf{y}^N)\big) \vee \big(\tau - t^N\big) \le 2\delta \le \mathcal{W}_2\big(\lambda_{Y_\tau}, m^N(\mathbf{y}^N)\big) + \big(\tau - t^N\big), \quad \nu^N\text{-a.s.},$$

and for sufficiently large N:

$$\mathcal{W}_2(\lambda_{Y_\tau}, m) \vee \tau \le 3\delta \quad \text{and} \quad \delta \le \mathcal{W}_2(\lambda_{Y_\tau}, m) + \tau, \quad \nu^N\text{-a.s.} \tag{5.7}$$

By the continuity of $\mathcal{W}_2$ and $\bar{\omega} \longmapsto \lambda_{Y_\tau}(\bar{\omega})$, the last inequalities exhibit a fixed closed support for ν^N, which is then inherited by the weak limit ν. Note further $\lambda_{Y_{t-}} = m$, we deduce that $(0,m) \ne (\tau(\bar{\omega}), [\lambda(\bar{\omega})]_{Y_{\tau(\bar{\omega})}}) \in \mathcal{N}_{\delta_0}(0,m)$. This contradicts with (5.1) that $(0,m)$ is a strict minimum on $\mathcal{N}_{\delta_0}(0,m)$. Thus, (5.4) holds true.

Step 4: We prove that ϕ^N is a test function of U^N at some point. Introduce

$$\mathcal{Z}_s^N := \operatorname*{ess\,inf}_{\theta\in\mathcal{T}_{t^N,T}^N} \mathbb{E}^{\mathbb{P}^0}\big[(\phi^N - U^N)(s, \bar{\mathbf{Y}}_s^{t^N,\mathbf{y}^N})|\mathcal{F}_s^N\big] \quad \text{for all } s \in [t^N, T],$$

that is, the $\mathbb{P}^0$-lower Snell envelope of $s \mapsto (\phi^N - U^N)(s, \bar{\mathbf{Y}}_s^{t^N,\mathbf{y}^N})$. Note that $\mathcal{Z}^N$ is a $\mathbb{P}^0$-submartingale. By (5.4), for all $\delta > 0$ and $N \ge 1$, we may find $\omega^{\delta,N} \in \Omega^0$ such that $\theta_\delta^N(\omega^{\delta,N}) < \mathrm{H}_\delta^N(\omega^{\delta,N})$ and, denoting $(t_\delta^N, \mathbf{y}_\delta^N) := (\theta^{\delta,N}(\omega^{\delta,N}), \bar{\mathbf{Y}}^{t^N,\mathbf{y}^N}_{\theta_\delta^N(\omega^{\delta,N})}(\omega^{\delta,N}))$,

$$(\phi^N - U^N)(t_\delta^N, \mathbf{y}_\delta^N) = \mathcal{Z}_{t_\delta^N}^N(\omega^{\delta,N}) \le \mathbb{E}^{\mathbb{P}^0}\big[\mathcal{Z}^N_{\theta\wedge \mathrm{H}_\delta^N}\big] \le \mathbb{E}^{\mathbb{P}^0}\big[(\phi^N - U^N)(\theta\wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}^{t^N,y^N}_{\theta\wedge \mathrm{H}_\delta^N})\big],$$

for all $\theta \in \mathcal{T}_{t_\delta^N,T}$. Thus we have

$$(\phi^N - U^N)(t_\delta^N, \mathbf{y}_\delta^N) = \min_{\theta\,\in\,\mathcal{T}_{t_\delta^N,T}} \mathbb{E}^{\mathbb{P}^0}_{t_\delta^N,\mathbf{y}_\delta^N}\big[(\phi^N - U^N)(\theta\wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}_{\theta\wedge \mathrm{H}_\delta^N})\big]. \tag{5.8}$$

Therefore, since $\phi^N \in C^{1,2}(\overline{\Lambda}_{t^N}^N)$ by Lemma 3.8, we have $\phi^N \in \underline{\mathcal{A}}^N U^N(t_\delta^N, \mathbf{y}_\delta^N)$, and thus the subsolution property of U^N provides,

$$\begin{aligned}\min\Big\{&-\partial_t\phi^N(t_\delta^N, \mathbf{y}_\delta^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t_\delta^N, \mathbf{y}_\delta^N)\cdot \mathbf{i}^N, U^N(t_\delta^N, \mathbf{y}_\delta^N)\\ &- \max_{\mathbf{i}'<\mathbf{i}^N} U^N(t_\delta^N, \mathbf{x}_\delta^N, \mathbf{i}')\Big\} \le 0.\end{aligned} \tag{5.9}$$

Step 5: We finally derive the viscosity subsolution property of $\overline{U}$. First observe that, since $t_\delta^N = \theta_\delta^N(\omega^{\delta,N}) < \mathrm{H}_\delta^N(\omega^{\delta,N})$, we have $(t_\delta^N, \mathbf{y}_\delta^N) \longrightarrow_{\delta\to 0} (t^N, \mathbf{y}^N)$. Since ϕ^N, $\mathcal{L}\phi^N + \mathbf{f}^N$ and U^N are continuous, sending δ to 0 in (5.9) provides

$$\min\Big\{-\partial_t\phi^N(t^N,\mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N,\mathbf{y}^N)\cdot\mathbf{i}^N,\ U^N(t^N,\mathbf{y}^N) - \max_{\mathbf{i}'<\mathbf{i}^N} U^N(t^N,\mathbf{x}^N,\mathbf{i}')\Big\} \le 0.$$

By (5.3), this implies

$$-\partial_t\phi^N(t^N,\mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N,\mathbf{y}^N)\cdot\mathbf{i}^N \le 0 \quad \text{for all } N \ge 1. \tag{5.10}$$

Thus, noting that i_k^N takes only values 0 and 1, by (3.14) and (3.6) we have

$$\begin{aligned}
&-\partial_t\phi^N(t^N,\mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N,\mathbf{y}^N)\cdot\mathbf{i}^N \\
&\quad = -\mathbb{L}\varphi(t^N, m^N(\mathbf{y}^N)) \\
&\quad\quad - \frac{1}{N}\sum_{k=1}^N f(t^N, x_k^N, m^N(\mathbf{y}^N))i_k^N - \frac{1}{N^2}\sum_{k=1}^N \sigma^2 : \partial^2_{\mu\mu}\varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N)i_k^N \\
&\quad = -[\mathbb{L}\varphi + F](t^N, m^N(\mathbf{y}^N)) - \frac{1}{N}\int_{\mathbb{R}^d} \sigma^2 : \partial^2_{\mu\mu}\varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x, x)m^N(\mathbf{y}^N)(dx, 1),
\end{aligned} \tag{5.11}$$

where $\sigma(s,\tilde m, x, x) := \sigma(s, x, \tilde m)$. Note that, as $N \to \infty$,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \sigma^2 : \partial^2_{\mu\mu}\varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x, x)m^N(\mathbf{y}^N)(dx, 1) \\
&\qquad \to \int_{\mathbb{R}^d} \sigma^2 : \partial^2_{\mu\mu}\varphi_{1,1}(0, m, x, x)m(dx, 1),
\end{aligned}$$

since $(t^N, m^N(\mathbf{y}^N)) \to (0, m)$ and the mapping $(s,\tilde m) \mapsto \int_{\mathbb{R}^d} \sigma^2 : \partial^2_{\mu\mu}\varphi_{1,1}(s,\tilde m, x, x)\tilde m(dx, 1)$ is continuous. Then (5.10) and (5.11) imply that

$$-(\mathbb{L}\varphi + F)(t^N, m^N(\mathbf{y}^N)) + \mathcal{O}\Big(\frac{1}{N}\Big) \le 0.$$

Sending $N \to \infty$ and noticing that $\mathbb{L}\varphi + F$ is continuous, we conclude the viscosity subsolution property of $\overline{U}$ at $(0, m)$.

(ii) We now prove the viscosity supersolution property. Fix $\varphi \in \overline{\mathcal{A}}\underline{U}(0, m)$ with corresponding δ_0 and s.t. $(0, m)$ is a strict maximizer of $\varphi - \underline{U}$ on $\mathcal{N}_{\delta_0}(0, m)$. With the same procedure and notation as in (i) (in particular, we rewrite (5.5) with reversed inequalities and switch lim sup and lim inf), we may find $(t_\delta^N, m^N(\mathbf{y}_\delta^N)) \longrightarrow_{(\delta,N)\to(0,\infty)} (0, m)$ s.t.

$$(\phi^N - U^N)(t_\delta^N, \mathbf{y}_\delta^N) = \max_{\theta\in\mathcal{T}_{t_\delta^N,T}} \mathbb{E}^{\mathbb{P}^0}_{t_\delta^N,\mathbf{y}_\delta^N}\big[(\phi^N - U^N)(\theta\wedge \mathrm{H}_\delta^N, \bar{\mathbf{Y}}_{\theta\wedge \mathrm{H}_\delta^N})\big].$$

Thus $\phi^N \in \overline{\mathcal{A}}^N U^N(t_\delta^N, \mathbf{y}_\delta^N)$ and the viscosity supersolution property provides

$$\begin{aligned}
\min\Big\{&-(\mathbb{L}\varphi + F)(t_\delta^N, m^N(\mathbf{y}_\delta^N)) + \mathcal{O}\Big(\frac{1}{N}\Big), \\
&U^N(t_\delta^N, m^N(\mathbf{y}_\delta^N)) - \max_{\mathbf{i}'<\mathbf{i}_\delta^N} U^N(t_\delta^N, m^N(\mathbf{x}_\delta^N, \mathbf{i}'))\Big\} \ge 0.
\end{aligned} \tag{5.12}$$

As $(t_\delta^N, m^N(\mathbf{y}_\delta^N)) \longrightarrow_{(\delta,N)\to(0,\infty)} (0, m)$, sending $(\delta, N) \longrightarrow (0, \infty)$ in (5.12) provides the first part of the viscosity supersolution property:

$$-(\mathbb{L}\varphi + F)(0, m) \ge 0.$$

It remains to show that $\underline{U}$ is nondecreasing for $\preceq$. We first assume that

(5.13) $m(dx,1)$ satisfies the required regularity in Lemma 2.2.

For any $A \in \mathcal{B}(\mathbb{R}^d)$, by considering the corresponding cdf, we easily see that

$$(m^N)^A(\mathbf{y}^N) := \frac{1}{N}\sum_{j=1}^N \delta_{(x_k^N, i_k^N \mathbf{1}_A(x_k^N))} \underset{N\to\infty}{\longrightarrow} m^A.$$

By (5.12), we have $U^N(t^N, m^N(\mathbf{y}^N)) \geq U^N(t^N, (m^N)^A(\mathbf{y}^N))$ for all $N \geq 1$. Taking the lim inf and using the fact that $U^N(t^N, m^N(\mathbf{y}^N)) \longrightarrow_{N\to\infty} \underline{U}(0,m)$, we deduce $\underline{U}(0,m) \geq \underline{U}(0,m^A)$. Given the arbitrariness of A and by Lemma 2.2, we deduce from the continuity of $\underline{U}$ that $\underline{U}(0,m) \geq \underline{U}(0,m')$ for all $m' \preceq m$, under the condition (5.13).

Now for general m, there exist $m_n \in \mathcal{P}_2(\mathbf{S})$ such that $\lim_{n\to\infty} W_2(m_n, m) = 0$ and each m_n satisfies (5.13). Then $\underline{U}(0,m_n) \geq \underline{U}(0,m_n')$ for all $m_n' \preceq m_n$. Recall (2.7) and let $m' \preceq m$ with transition function p. We first assume p is continuous. Then, letting $m_n' \preceq m_n$ have the same p, we have $\lim_{n\to\infty} W_2(m_n', m') = 0$. Thus, the continuity of $\underline{U}$ implies $\underline{U}(0,m) \geq \underline{U}(0,m')$. Finally, by [55], Lemma 3.8, we extend the inequality to all $m' \preceq m$, with possibly discontinuous p. □

6. Propagation of chaos for stopped diffusions. In this section we prove Theorems 4.3 and 4.4. We first use the latter to prove Theorem 4.3. Recall (2.3), (3.2) and that λ denotes the canonical map on $\mathcal{P}_2(\Omega)$.

PROOF OF THEOREM 4.3. (i) Let $\hat{\nu}^N := \mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}))^{-1}$. By Theorem 4.4(i) and by otherwise considering a subsequence, we may assume $\hat{\nu}^N$ converges weakly to some $\hat{\nu}$ supported on $\mathcal{P}(t,m)$. For each $N \geq 1$, since $\hat{\boldsymbol{\tau}}^N$ is ε_N-optimal for (3.2), we have

$$v^N(t,\mathbf{y}^N) \leq J_N(t,\mathbf{y}^N,\hat{\boldsymbol{\tau}}^N) + \varepsilon_N = \mathbb{E}^{\mathbb{P}^0}\big[J\big(t, m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N})\big)\big] + \varepsilon_N = \mathbb{E}^{\hat{\nu}^N}\big[J(t,\lambda)\big] + \varepsilon_N. \tag{6.1}$$

For any $R > 0$, let $J_R(t,\lambda)$ be defined by truncating F and g with R in (2.3), and define $J_{N,R}$ similarly by truncation in (3.2). Then

$$\lim_{N\to\infty} \mathbb{E}^{\hat{\nu}^N}\big[J_R(t,\lambda)\big] = \mathbb{E}^{\hat{\nu}}\big[J_R(t,\lambda)\big]. \tag{6.2}$$

By (3.7), we have the following uniform estimate:

$$\begin{aligned}
&\big|\mathbb{E}^{\hat{\nu}^N}\big[J_R(t,\lambda)\big] - \mathbb{E}^{\hat{\nu}^N}\big[J(t,\lambda)\big]\big| \\
&\quad = \big|J_{N,R}(t,\mathbf{y}^N,\hat{\boldsymbol{\tau}}^N) - J_N(t,\mathbf{y}^N,\hat{\boldsymbol{\tau}}^N)\big| \\
&\quad \leq C\mathbb{E}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N \int_t^T \big(1 + |X_r^{\hat{\boldsymbol{\tau}}^N,k}| + \|\mathbf{X}_r^{\hat{\boldsymbol{\tau}}^N}\|_1\big)\mathbf{1}_{\{C(1+|X_r^{\hat{\boldsymbol{\tau}}^N,k}|+\|\mathbf{X}_r^{\hat{\boldsymbol{\tau}}^N}\|_1)\geq R\}}\,dr \\
&\qquad + \|\mathbf{X}_T^{\hat{\boldsymbol{\tau}}^N}\|_1 \mathbf{1}_{\{C\|\mathbf{X}_T^{\hat{\boldsymbol{\tau}}^N}\|_1\geq R\}}\bigg] \\
&\quad \leq \frac{C}{R}\mathbb{E}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N \int_t^T \big[1 + |X_r^{\hat{\boldsymbol{\tau}}^N,k}|^2 + \|\mathbf{X}_r^{\hat{\boldsymbol{\tau}}^N}\|_1^2\big]\,dr + \|\mathbf{X}_T^{\hat{\boldsymbol{\tau}}^N}\|_1^2\bigg] \\
&\quad \leq \frac{C}{R}\bigg[\frac{1}{N}\sum_{k=1}^N \big[1 + |x_k^N|^2\big] + \|\mathbf{x}^N\|_1^2\bigg] \leq \frac{C}{R}\big[1 + \|\mathbf{x}^N\|_2^2\big].
\end{aligned}$$

Thus it is clear that

$$\lim_{R\to\infty}\sup_{N\geq 1}\big|\mathbb{E}^{\hat{\nu}^N}\big[J_R(t,\lambda)\big]-\mathbb{E}^{\hat{\nu}^N}\big[J(t,\lambda)\big]\big|=\lim_{R\to\infty}\big|\mathbb{E}^{\hat{\nu}}\big[J_R(t,\lambda)\big]-\mathbb{E}^{\hat{\nu}}\big[J(t,\lambda)\big]\big|=0.$$

Then by (6.2) we have $\lim_{N\to\infty}\mathbb{E}^{\hat{\nu}^N}[J(t,\lambda)]=\mathbb{E}^{\hat{\nu}}[J(t,\lambda)]$. Send $N\to 0$ in (6.1), note that $\varepsilon_N\to 0$ and by Theorem 4.1 $v^N(t,\mathbf{y}^N)\longrightarrow_{N\to\infty} V(t,m)$, we obtain $V(t,m)\leq\mathbb{E}^{\hat{\nu}}[J(t,\lambda)]$. Since $\hat{\nu}$ is supported on $\mathcal{P}(t,m)$, we have $V(t,m)=\sup_{\mathbb{P}\in\mathcal{P}(t,m)}J(t,\mathbb{P})\geq J(t,\lambda)$, $\hat{\nu}$-a.s., and therefore $V(t,m)=J(t,\lambda)$, $\hat{\nu}$-a.s. This proves that $\hat{\nu}$ is supported on the set of optimal controls.

(ii) By Theorem 4.4(ii), there exists a sequence $\hat{\boldsymbol{\tau}}^N$, $N\geq 1$, such that $m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N})\xrightarrow{\mathcal{W}_2}\mathbb{P}^*$, $\mathbb{P}^0$-a.s. Then, by the continuity of F and g and following similar arguments as in (i),

$$\mathbb{E}^{\mathbb{P}^0}\big[J\big(t,m^N\big(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}\big)\big)\big]\to J\big(t,\mathbb{P}^*\big)=V(t,m). \tag{6.3}$$

Introduce

$$\varepsilon_N:=V^N\big(t,m^N\big(\mathbf{y}^N\big)\big)-\mathbb{E}^{\mathbb{P}^0}\big[J\big(t,m^N\big(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}\big)\big)\big].$$

By definition $\hat{\boldsymbol{\tau}}^N$ is ε_N-optimal for (3.2), and by Theorem 4.1 and (6.3), we get $\varepsilon_N\to 0$.

(iii) Assume by contradiction that there exist $\varepsilon_N\to 0$ but $m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N})$ does not converge to $\mathbb{P}^*$. Then there exists $c>0$ such that, by otherwise choosing a subsequence, $\mathbb{P}_0(\mathcal{W}_2(m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}),\mathbb{P}^*)>c)>c$. By Theorem 4.3(i), again by otherwise choosing a subsequence, we may assume $\mathbb{P}^0\circ(m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N}))^{-1}$ converges weakly to some $\nu\in\mathcal{P}_2(\mathcal{P}_2(\Omega))$ whose support is on the singleton $\{\mathbb{P}^*\}$, then we must have $m^N(\mathbf{Y}^{\hat{\boldsymbol{\tau}}^N})\to\mathbb{P}^*$, in probability $\mathbb{P}^0$. This is a desired contradiction. □

PROOF OF THEOREM 4.4. Again for simplicity we assume $t=0$.

(i) For each $N\geq 1$, denote $\nu^N:=\mathbb{P}^0\circ(m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}))^{-1}\in\mathcal{P}_2(\mathcal{P}_2(\Omega))$.

Step 1. We first show that $\{\nu^N\}_{N\geq 1}$ is uniformly integrable, that is,

$$\lim_{R\to\infty}\sup_{N\geq 1}\mathbb{E}^{\nu^N}\big[\|\lambda\|_2\mathbf{1}_{\{\|\lambda\|_2\geq R\}}\big]=0. \tag{6.4}$$

Indeed, for any $R>0$, by (3.7) we have

$$\begin{aligned}\mathbb{E}^{\nu^N}\big[\|\lambda\|_2\mathbf{1}_{\{\|\lambda\|_2\geq R\}}\big]&\leq\frac{1}{R}\mathbb{E}^{\nu^N}\big[\|\lambda\|_2^2\big]=\frac{1}{R}\mathbb{E}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N\big|\sup_{0\leq t\leq T}\big|Y_t^{\boldsymbol{\tau}^N,k}\big|^2\bigg]\\&\leq\frac{C}{R}\bigg[1+\frac{1}{N}\sum_{k=1}^N\big[\big|x_k^N\big|^2+\big\|\mathbf{x}^N\big\|_2^2\big]\bigg]\leq\frac{C}{R}\big[1+\big\|\mathbf{y}^N\big\|_2^2\big].\end{aligned}$$

Since $\mathcal{W}_2(m^N(\mathbf{y}^N),m)\longrightarrow_{N\to\infty}0$, we have $\|\mathbf{y}^N\|_2\to\|m\|_2$, and thus $\{\|\mathbf{y}^N\|_2\}_{N\geq 1}\}$ is bounded. This implies (6.4) immediately.

Step 2. We next prove the tightness of $\{\nu^N\}_{N\geq 1}$. By Lacker [40], Corollary B.1, given (6.4), it suffices to show that the mean measures $\{\mathbb{E}^{\mathbb{P}^0}[m^N(\mathbf{Y}^{\boldsymbol{\tau}^N})]\}_{N\geq 1}\subset\mathcal{P}_2(\Omega)$ is tight. Here, for a random measure $\tilde{m}:\Omega^0\to\mathcal{P}_2(\Omega)$, the mean measure $\mathbb{E}^{\mathbb{P}^0}[\tilde{m}]$ is defined by

$$\big\langle\mathbb{E}^{\mathbb{P}^0}[\tilde{m}],\varphi\big\rangle:=\mathbb{E}^{\mathbb{P}^0}\big[\langle\tilde{m},\varphi\rangle\big]\quad\text{for all }\varphi\in C_b^0(\Omega).$$

To see this, we first remark that, the tightness of the joint measure is equivalent to the tightness of the marginals. For the first marginal, by (3.8) we have

$$\sup_{N\geq 1}\sup_{\tau\in\mathcal{T}_{0,T}}\big\langle\mathbb{E}^{\mathbb{P}^0}\big[m^N\big(\mathbf{Y}^{\boldsymbol{\tau}^N}\big)\big],|X_{(\tau+\delta)\wedge T}-X_\tau|^2\big\rangle$$

$$= \sup_{N\ge 1}\sup_{\tau\in\mathcal{T}_{0,T}} \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}^{\mathbb{P}^0}\big[\big|X^{\boldsymbol{\tau}^N,k}_{(\tau+\delta)\wedge T} - X^{\boldsymbol{\tau}^N,k}_{\tau}\big|^2\big] \le C\delta.$$

Then it follows from Aldous' criterion (see Billingsley [5], Theorem 16.10) that the first marginal of $\{\mathbb{E}^{\mathbb{P}^0}[m^N(\mathbf{Y}^{\boldsymbol{\tau}^N})]\}_{N\ge1}$ is tight. Moreover, since the second marginals are measures on $\mathbb{I}^0([-1,T])$, which are in continuous bijection with measures on $[0,T]$, and thus are tight. Therefore, $\{\mathbb{E}^{\mathbb{P}^0}[m^N(\mathbf{Y}^{\boldsymbol{\tau}^N})]\}_{N\ge1}$ and hence $\{\nu^N\}_{N\ge1}$ are tight.

Step 3. We now show that ν is supported on $\mathcal{P}(0,m)$, or equivalently, that $\lambda\in\mathcal{P}(0,m)$, ν-a.s. Since $\mathcal{W}_2(m^N(\mathbf{y}^N),m)\longrightarrow_{N\to\infty}0$, we have $\lambda_{0-}=m$, ν-a.s. Recall (2.4), we may consider equivalently the martingale problem. That is, for any $\psi\in C^2_b(\mathbb{R}^d)$, we want to show that

$$M^{\psi}_s := \psi(X_s) - \int_0^s \mathcal{L}_x\psi(r,X_r,\lambda_r)I_r\,dr \quad \text{is a } \lambda\text{-martingale, for } \nu\text{-a.e. } \lambda. \tag{6.5}$$

For this purpose, fix $0\le s_1<s_2\le T$ and $h_{s_1}\in C^0_b(\Omega)$ that is $\mathcal{F}_{s_1}$-measurable. Note that

$$\begin{aligned}
\mathbb{E}^{\nu^N}\big[|\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|\wedge1\big] &\le \mathbb{E}^{\nu^N}\big[|\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|^2\big]\\
&= \mathbb{E}^{\mathbb{P}^0}\big[|\langle m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}),h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|^2\big]\\
&= \mathbb{E}^{\mathbb{P}^0}\Bigg[\bigg|\frac{1}{N}\sum_{k=1}^{N} h_{s_1}(Y^{\boldsymbol{\tau}^N,k})\Big[\psi(X^{\boldsymbol{\tau}^N,k}_{s_2})\\
&\qquad - \psi(X^{\boldsymbol{\tau}^N,k}_{s_1}) - \int_{s_1}^{s_2}\mathcal{L}_x\psi\big(r,X^{\boldsymbol{\tau}^N,k}_r,m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}_r)\big)I^{\boldsymbol{\tau}^N,k}_r\,dr\Big]\bigg|^2\Bigg].
\end{aligned}$$

By (3.1) and (3.6), it follows from the Itô formula that

$$\begin{aligned}
&\mathbb{E}^{\nu^N}\big[|\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|\wedge1\big]\\
&\le \mathbb{E}^{\mathbb{P}^0}\Bigg[\bigg|\frac{1}{N}\sum_{k=1}^{N} h_{s_1}(Y^{\boldsymbol{\tau}^N,k})\int_{s_1}^{s_2}\partial_x\psi(X^{\boldsymbol{\tau}^N,k}_r)\sigma\big(r,X^{\boldsymbol{\tau}^N,k}_r,m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}_r)\big)I^{\boldsymbol{\tau}^N,k}_r\,dW_r\bigg|^2\Bigg]\\
&= \frac{1}{N^2}\sum_{k=1}^{N}\mathbb{E}^{\mathbb{P}^0}\Bigg[|h_{s_1}(Y^{\boldsymbol{\tau}^N,k})|^2\int_{s_1}^{s_2}\big|\partial_x\psi(X^{\boldsymbol{\tau}^N,k}_r)\sigma\big(r,X^{\boldsymbol{\tau}^N,k}_r,m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}_r)\big)I^{\boldsymbol{\tau}^N,k}_r\big|^2\,dr\Bigg]\\
&\le \frac{C}{N^2}\sum_{k=1}^{N}\mathbb{E}^{\mathbb{P}^0}\Bigg[\int_{s_1}^{s_2}\big|\sigma\big(r,X^{\boldsymbol{\tau}^N,k}_r,m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}_r)\big)\big|^2\,dr\Bigg] \le \frac{C}{N},
\end{aligned}$$

where the last inequality thanks to (3.7). Note that ν^N converges to ν weakly, and the mapping $\lambda\in\mathcal{P}_2(\Omega)\longmapsto|\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|\wedge1$ is bounded and continuous. Then

$$\mathbb{E}^{\nu}\big[|\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle|\wedge1\big]=0.$$

This implies that, for any desired ψ, h_{s_1} and $s_1<s_2$,

$$\langle\lambda,h_{s_1}[M^{\psi}_{s_2}-M^{\psi}_{s_1}]\rangle=0 \quad \text{for } \nu\text{-a.e. } \lambda. \tag{6.6}$$

Since M^{ψ} is continuous in s, and the spaces of the above ψ and h_{s_1} are separable, then (6.6) implies (6.5). Therefore, $\lambda\in\mathcal{P}(0,m)$, ν-a.s.

(ii) Without loss of generality, assume $(\Omega^0,\mathcal{F}^0_T,\mathbb{F}^0,\mathbb{P}^0)$ is sufficiently large so that there exist a sequence of i.i.d. processes $\{(I^{0,k},W^k)\}_{k\ge1}$ which are independent of $\mathcal{F}^0_{0-}$ (under $\mathbb{P}^0$)

and $\mathbb{P}^0_{(I^{0,k},W^k)} = \mathbb{P}_{(I,W^{\mathbb{P}})|I_{0-}=1}$. In particular, W^k is a $\mathbb{P}^0$-Brownian motion and $I^{0,k}_{0-} = 1$. Now for given $\mathbf{y}^N$, define

$$I^{N,k}_t := I^{0,k}_t \mathbf{1}_{\{i_k=1\}}, \qquad \tau^N_k := \inf\{t \geq 0 : I^{N,k}_t = 0\}, \qquad \boldsymbol{\tau}^N := (\tau^N_1, \dots, \tau^N_N). \tag{6.7}$$

We shall prove in four steps that the above $\boldsymbol{\tau}^N$ satisfies the convergence requirement. Since these stopping times are adapted to the filtration $\mathbb{F}^0$ which might be larger than the Brownian filtration, we finally show in Step 5 that they can be approximated by pure stopping strategies, and that we can derive from this approximation the desired result.

Step 1. We first assume

$$\int_{\mathbf{S}} |x|^3 m(dy) < \infty, \quad \text{and} \quad \lim_{N\to\infty} \mathcal{W}_3(m^N(\mathbf{y}^N), m) = 0. \tag{6.8}$$

Denote $m_t := \mathbb{P}_{Y_t}$ and introduce

$$\tilde{X}^{N,k}_t = x^N_k + \int_0^t b(r, \tilde{X}^{N,k}_r, m_r) I^{N,k}_r \, dr + \int_0^t \sigma(r, \tilde{X}^{N,k}_r, m_r) I^{N,k}_r \, dW^k_r. \tag{6.9}$$

Note that $(W^k, I^{N,k}, \tilde{X}^{N,k})$, $k \in [N]$, are independent. In this step we show that

$$\lim_{N\to\infty} \sup_{0\leq t\leq T} \mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}^N_t), m_t)\big] = 0 \quad \text{for all } t, \text{ where } \tilde{\mathbf{Y}}^{N,k} := (\tilde{X}^{N,k}, I^{N,k}). \tag{6.10}$$

To see this, let $E^N_k \in \mathcal{F}^0_{0-}$, $k \in [N]$ be a partition of Ω^0 such that $\mathbb{P}^0(E^N_k) = \frac{1}{N}$. Denote

$$(\hat{W}^N, \hat{X}^N_t, \hat{I}^N_t) := \sum_{k=1}^N (W^k, \tilde{X}^{N,k}_t, I^{N,k}_t) \mathbf{1}_{E^N_k}, \qquad \hat{Y}^N := (\hat{X}^N_t, \hat{I}^N_t). \tag{6.11}$$

Then it follows from the arguments in [47], Lemma 8.4, that

$$\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}^N_t), \mathbb{P}^0_{\hat{Y}^N_t})\big] \leq \frac{C}{N^{\frac{2}{(d+1)\vee 3}}} \big(\mathbb{E}^{\mathbb{P}^0}\big[|\hat{Y}^N_t|^3\big]\big)^{\frac{2}{3}}. \tag{6.12}$$

We shall point out that [47], Lemma 8.4, assumes Brownian filtration, which is not the case here. However, we emphasize that this assumption is due to the setting in [47] and is never used there. In fact, the arguments provided in the proof of [47], Lemma 8.4, used only the independence between $\{(\tilde{X}^{N,k}, I^{N,k})\}_k$ and $\{E^N_k\}_k$, which holds true here. Then [47] refers to [33], Lemma 5, Lemma 6 and Theorem 1, which do not require the Brownian filtration. Moreover, by standard SDE estimate and noting that $\mathcal{W}_3(m^N(\mathbf{y}^N), m) \to 0$, we may estimate $\mathbb{E}^{\mathbb{P}^0}[|\hat{Y}^N_t|^3]$ and thus derive from (6.12) that

$$\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}^N_t), \mathbb{P}^0_{\hat{Y}^N_t})\big] \leq \frac{C(1 + \|\mathbf{x}^N\|_3^2)}{N^{\frac{2}{(d+1)\vee 3}}} \leq \frac{C_m}{N^{\frac{2}{(d+1)\vee 3}}}, \tag{6.13}$$

for some constant C_m, which may depend on m, but not on N.

Next, for any bounded function $\varphi \in C^0(\Omega)$. Denote

$$v(i) := \mathbb{E}^{\mathbb{P}}\big[\varphi(W^{\mathbb{P}}, I)|I_{0-} = i\big], \quad i = 0, 1.$$

Note that $(W^k, \tilde{X}^{N,k}_t, I^{N,k}_t)$ is independent of $\mathcal{F}^0_{0-}$. Then, by (6.7),

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^0}\big[\varphi(\hat{W}^N, \hat{I}^N)\big] &= \frac{1}{N}\sum_{k=1}^N \mathbb{E}^{\mathbb{P}^0}\big[\varphi(W^k, I^{N,k})\big] \\ &= \frac{1}{N}\sum_{k=1}^N \big[v(1)\mathbf{1}_{\{i_k=1\}} + v(0)\mathbf{1}_{\{i_k=0\}}\big] \\ &\longrightarrow v(1)\mathbb{P}(I_{0-}=1) + v(0)\mathbb{P}(I_{0-}=0) = \mathbb{E}^{\mathbb{P}}\big[\varphi(W^{\mathbb{P}}, I)\big], \end{aligned} \tag{6.14}$$

as $N \to \infty$. That is, $\mathbb{P}^0 \circ (\hat{W}^N, \hat{I}^N)^{-1} \to \mathbb{P} \circ (W^{\mathbb{P}}, I)^{-1}$ weakly. Since I is bounded, by [11], Theorem 5.5, we have

$$\lim_{N\to\infty} \mathcal{W}_2\big(\mathbb{P}^0 \circ (\hat{W}^N, \hat{I}^N)^{-1}, \mathbb{P} \circ (W^{\mathbb{P}}, I)^{-1}\big) = 0.$$

Moreover, since $\lim_{N\to\infty} \mathcal{W}_2(m^N(\mathbf{y}^N), m)) = 0$, then

$$\lim_{N\to\infty} \mathcal{W}_2\big(\mathbb{P}^0 \circ (\hat{X}_0^N, \hat{W}^N, \hat{I}^N)^{-1}, \mathbb{P} \circ (X_0, W^{\mathbb{P}}, I)^{-1}\big) = 0.$$

One may easily verify that

$$\hat{X}_t^N = \hat{X}_0^N + \int_0^t b(r, \hat{X}_r^N, m_r)\hat{I}_r^N \, dr + \int_0^t \sigma(r, \hat{X}_r^N, m_r)\hat{I}_r^N \, d\hat{W}_r^N.$$

Then, by comparing this with (2.5), $\lim_{N\to\infty} \sup_{0\le t\le T} \mathcal{W}_2(\mathbb{P}^0_{\hat{Y}_t^N}, m_t) = 0$. This, together with (6.13), implies (6.10) immediately.

Step 2. In this step we show that, again under the additional condition (6.8),

$$\lim_{N\to\infty} \sup_{0\le t\le T} \mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m_t)\big] = 0. \tag{6.15}$$

Recall (6.7) and (3.6), and compare (3.1) and (6.9). By the Lipschitz continuity of b, σ, especially the $\mathcal{W}_1$-Lipschitz continuity in m, it follows from standard SDE estimates that

$$\mathbb{E}^{\mathbb{P}^0}\big[|X_t^{\boldsymbol{\tau}^N,k} - \tilde{X}_t^{N,k}|^2\big] \le C\mathbb{E}^{\mathbb{P}^0}\bigg[\int_0^t \mathcal{W}_1^2(m^N(\mathbf{Y}_s^{\tau^N}), m_s)\, ds\bigg].$$

Then

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m_t)\big] &\le C\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m^N(\tilde{\mathbf{Y}}_t^N)) + \mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}_t^N), m_t))\big] \\
&\le C\mathbb{E}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N |X_t^{\boldsymbol{\tau}^N,k} - \tilde{X}_t^{N,k}|^2 + \mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}_t^N), m_t))\bigg] \\
&\le C\mathbb{E}^{\mathbb{P}^0}\bigg[\int_0^t \mathcal{W}_1^2(m^N(\mathbf{Y}_s^{\tau^N}), m_s)\, ds + \mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}_t^N), m_t))\bigg].
\end{aligned}$$

Applying the Gronwall inequality, we obtain

$$\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m_t)\big] \le C\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\tilde{\mathbf{Y}}_t^N), m_t))\big].$$

This, together with (6.10), implies (6.15) immediately.

Step 3. We now prove the result without assuming (6.8). For any $R > 0$, let $\phi_R(x)$ denote the truncation function such that x is truncated by R. Denote $m^R := m \circ (\phi_R(X_0), I_{0-})^{-1}$, $\mathbf{y}^{N,R} := (\phi_R(x_k^N), i_k^N)_{1\le k\le N}$. Then it is clear that

$$\lim_{R\to\infty} \mathcal{W}_2(m^R, m) = 0, \qquad \lim_{R\to\infty} \sup_{N\ge 1} \frac{1}{N}\sum_{k=1}^N |\phi_R(x_k^N) - x_k^N|^2 = 0, \tag{6.16}$$

$$\lim_{N\to\infty} \mathcal{W}_3\big(m^N(\mathbf{y}^{N,R}), m^R\big) = 0 \quad \text{for all } R > 0. \tag{6.17}$$

Introduce $\mathbb{P}^R$ and $\mathbf{Y}^{R,\boldsymbol{\tau}}$ in an obvious way. By (6.15) and (6.17) we have

$$\lim_{N\to\infty} \sup_{0\le t\le T} \mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{R,\boldsymbol{\tau}^N}), \mathbb{P}^R_{Y_t})\big] = 0. \tag{6.18}$$

Moreover, by standard SDE estimates one can easily see that

$$\mathcal{W}_2(\mathbb{P}^R_{Y_t}, m_t) \le C\mathcal{W}_2(m^R, m),$$

$$\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_2^2\big(m^N(\mathbf{Y}_t^{R,\boldsymbol{\tau}^N}), m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N})\big)\big] \le \frac{1}{N}\sum_{k=1}^N \mathbb{E}^{\mathbb{P}^0}\big[|X_t^{R,\boldsymbol{\tau}^N,k} - X_t^{\boldsymbol{\tau}^N,k}|^2\big]$$

$$\le \frac{C}{N}\sum_{k=1}^N |\phi_R(x_k^N) - x_k^N|^2.$$

Then it follows from (6.16) that, for any $R > 0$,

$$\begin{aligned}
&\limsup_{N\to\infty}\sup_{0\le t\le T}\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m_t)\big]\\
&\le \limsup_{N\to\infty} C\sup_{0\le t\le T}\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{\boldsymbol{\tau}^N}), m^N(\mathbf{Y}_t^{R,\boldsymbol{\tau}^N}))\\
&\quad + \mathcal{W}_1^2(m^N(\mathbf{Y}_t^{R,\boldsymbol{\tau}^N}), \mathbb{P}^R_{Y_t}) + \mathcal{W}_2^2(\mathbb{P}^R_{Y_t}, m_t))\big]\\
&= C\sup_{0\le t\le T}\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1^2(m^N(\mathbf{Y}_t^{R,\boldsymbol{\tau}^N}), \mathbb{P}^R_{Y_t})\big].
\end{aligned}$$

This, together with (6.18), implies (6.15), without assuming (6.8).

Step 4. Finally we prove the convergence of the distribution of the processes. First, compare (3.1) and (6.9) again, by (6.15) we obtain immediately that

$$\lim_{N\to\infty}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{0\le t\le T}|X_t^{\boldsymbol{\tau}^N,k} - \tilde{X}_t^{N,k}|^2\Big] = 0. \tag{6.19}$$

Next, for any bounded function $\varphi \in C^0(\Omega)$, similar to (6.14), we can show

$$\lim_{N\to\infty}\mathbb{E}^{\mathbb{P}^0}\big[\mathbb{E}^{m^N(\tilde{\mathbf{Y}}^N)}[\varphi(Y)]\big] = \lim_{N\to\infty}\mathbb{E}^{\mathbb{P}^0}\bigg[\frac{1}{N}\sum_{k=1}^N[\varphi(\tilde{Y}^{N,k})]\bigg] = \mathbb{E}^{\mathbb{P}}[\varphi(Y)].$$

This, together with (6.19), implies

$$\lim_{N\to\infty}\mathbb{E}^{\mathbb{P}^0}\big[\mathbb{E}^{m^N(\mathbf{Y}^{\boldsymbol{\tau}^N})}[\varphi(Y)]\big] = \mathbb{E}^{\mathbb{P}}[\varphi(Y)]. \tag{6.20}$$

On the other hand, by (i) of this theorem, we know $\{\mathbb{P}^0\circ(m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}))^{-1}\}_{N\ge1}$ is tight. By (6.20) clearly $\mathbb{P}$ is its unique accumulation point. Then, similar to the arguments in Theorem 4.3(iii), we obtain $m^N(\mathbf{Y}^{\boldsymbol{\tau}^N}) \xrightarrow{\mathcal{W}_2} \mathbb{P}$, $\mathbb{P}^0$-a.s.

Step 5: It remains to approximate the sequence $\{\boldsymbol{\tau}^N\}_{\{N\ge1\}}$ with a sequence of $\tilde{\boldsymbol{\tau}}^N \in \mathcal{T}_{t,T}^N$, that is, of stopping times adapted to the Brownian filtration. Indeed, the stopping times constructed in the previous steps are adapted to a larger filtration. Fixing $N \ge 1$, we have by Carmona, Delarue and Lacker [12], Theorem 6.4, the existence of a sequence $\{\tilde{\boldsymbol{\tau}}^{N,k}\}_{\{k\ge1\}}$ in $\mathcal{T}_{t,T}^N$ such that $(\tilde{\boldsymbol{\tau}}^{N,k}, \boldsymbol{W}^N$ converges in distribution to $(\boldsymbol{\tau}^N, \boldsymbol{W}^N)$ as $k\to\infty$, with $\boldsymbol{W}^N := (W^1, \dots, W^N)$. This implies that $\mathbf{Y}^{N,\tilde{\boldsymbol{\tau}}^{N,k}}$ converges in distribution to $\mathbf{Y}^{N,\boldsymbol{\tau}^N}$, and therefore that $m^N(\mathbf{Y}^{N,\tilde{\boldsymbol{\tau}}^{N,k}})$ converges in distribution to $m^N(\mathbf{Y}^{N,\boldsymbol{\tau}^N})$.

It is also obvious from the previous steps that $m^N(\mathbf{Y}^{N,\boldsymbol{\tau}^N})$ converges in distribution to $\mathbb{P}$ as $N\to\infty$. Then there exists a subsequence such that $m^N(\mathbf{Y}^{N,\tilde{\boldsymbol{\tau}}^{N,k_N}})$ converges in distribution to $\mathbb{P}$ as $N\to\infty$. Note that $\mathbb{P}$ is a constant in the space of measure-valued random variables

(endowed with the distance $\mathcal{W}_2$); therefore the previous convergence also holds in probability. We may then extract a subsequence (still denoted the same) converging to $\mathbb{P}$, $\mathbb{P}^0$-a.s., that is,

$$\mathcal{W}_2\big(m^N(\mathbf{Y}^{N,\tilde{\boldsymbol{\tau}}^{N,k_N}}),\mathbb{P}\big)\longrightarrow_{N\to\infty}0,\quad \mathbb{P}^0\text{-a.s.}$$

This concludes the proof. □

7. Further study on the multiple optimal stopping problem.

7.1. *The viscosity solution property.* In this subsection we prove Theorem 3.3. Recall the notation in Section 3.1 and 3.2, in particular the unstopped process $\bar{\mathbf{Y}}$. As standard, the viscosity property relies on the following dynamic programming principle.

PROPOSITION 7.1. *Under Assumption 3.1, for all $(t,\mathbf{y})\in\Lambda_0^N$, we have*

$$v^N(t,\mathbf{y})=\sup_{\theta\in\mathcal{T}_{t,T}^N}\mathbb{E}_{t,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^N\int_t^\theta f^k(r,\bar{\mathbf{X}}_r)\bar{I}_r^k\,dr+\max_{\mathbf{i}'<\mathbf{i}}v^N\big(\theta,\bar{\mathbf{X}}_\theta,\mathbf{i}'\big)\right].\tag{7.1}$$

In particular, by setting $\theta=t$, we see that v^N is nondecreasing in $\mathbf{i}$:

$$v^N(t,\mathbf{y})\ge v^N\big(t,\mathbf{x},\mathbf{i}'\big)\quad\textit{for all }\mathbf{i}'<\mathbf{i}.\tag{7.2}$$

PROOF. First, by Kobylanski, Quenez and Rouy-Mironescu [39], Theorem 3.1,,[1] we have

$$\begin{aligned}v^N(t,\mathbf{y})&=\sup_{\theta\in\mathcal{T}_{t,T}}\mathbb{E}_{t,\mathbf{y}}\left[\max_{l\in[N]}\sup_{\boldsymbol{\tau}\in\mathcal{T}_{\theta,T}^{N-1}}\mathbb{E}_{\theta,\mathbf{Y}_\theta^{-l}}\left[\frac{1}{N}\sum_{k=1}^N\int_t^T f^k\big(r,\mathbf{X}_r^{\boldsymbol{\tau}\otimes_l\theta}\big)\bar{I}_r^k\,dr+g\big(\mathbf{X}_T^{\boldsymbol{\tau}\otimes_l\theta}\big)\right]\right]\\&=\sup_{\theta\in\mathcal{T}_{t,T}}\mathbb{E}_{t,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^N\int_t^\theta f^k(r,\bar{\mathbf{X}}_r)\bar{I}_r^k\,dr\right.\\&\qquad\left.+\max_{l\in[N]}\sup_{\boldsymbol{\tau}\in\mathcal{T}_{\theta,T}^{N-1}}\mathbb{E}_{\theta,\mathbf{Y}_\theta^{-l}}\left[\frac{1}{N}\sum_{k=1}^N\int_\theta^T f^k\big(r,\mathbf{X}_r^{\boldsymbol{\tau}\otimes_l\theta}\big)\bar{I}_r^k\,dr+g\big(\mathbf{X}_T^{\boldsymbol{\tau}\otimes_l\theta}\big)\right]\right],\end{aligned}$$

where

$$\boldsymbol{\tau}\bigotimes_l\theta:=(\tau_1,\dots,\tau_{l-1},\theta,\tau_{l+1},\dots,\tau_N),\qquad\mathbf{i}^{-l}:=(i_1,\dots,i_{l-1},0,i_{l+1},\dots,i_N),\tag{7.3}$$

and $\mathbf{Y}^{-l}:=(\bar{\mathbf{X}},\mathbf{i}^{-l})$. By the Markov property, we have

$$\sup_{\boldsymbol{\tau}\in\mathcal{T}_{\theta,T}^{N-1}}\mathbb{E}_{\theta,\mathbf{Y}_\theta^{-l}}\left[\frac{1}{N}\sum_{k=1}^N\int_\theta^T f^k\big(r,\mathbf{X}_r^{\boldsymbol{\tau}\otimes_l\theta}\big)\bar{I}_r^k\,dr+g\big(\mathbf{X}_T^{\boldsymbol{\tau}\otimes_l\theta}\big)\right]=v^N\big(\theta,\bar{\mathbf{X}}_\theta,\mathbf{i}^{-l}\big).$$

Finally, by induction v^N is nondecreasing in $\mathbf{i}$ for the partial order on $\{0,1\}^N$, and thus

$$\max_{l\in[N]}v^N\big(\theta,\bar{\mathbf{X}}_\theta,\mathbf{i}^{-l}\big)=\max_{\mathbf{i}'<\mathbf{i}}v^N\big(\theta,\bar{\mathbf{X}}_\theta,\mathbf{i}'\big).$$

□

PROOF OF THEOREM 3.3. We proceed by induction on the cardinality $|A(\mathbf{i})|$ of the set

$$A(\mathbf{i}):=\big\{k\in[N]:i_k=1\big\}.\tag{7.4}$$

[1] Although Kobylanski, Quenez and Rouy-Mironescu assume nonnegative reward processes throughout their paper, the validity of their result in our context is easy to check.

As $(t,\mathbf{x}) \longmapsto g_N(t,\mathbf{x})$ is continuous, it suffices to prove that whenever $v^N(t,\mathbf{x},\mathbf{i})$ is continuous in $(t,\mathbf{x})$ for $|A(\mathbf{i})| \le n$, we may conclude for $|A(\mathbf{i})| = n+1$ that $v^N(\cdot,\mathbf{i})$ is the unique continuous viscosity solution of (3.3). Note that $\bar{I}^k_r \equiv 1$ for $k \in A(\mathbf{i})$ and $\bar{I}^k_r \equiv 0$ for $k \notin A(\mathbf{i})$. To see this, we rewrite (7.1), emphasizing again that the process $\bar{\mathbf{X}}$ is unstopped, thus reducing (7.5) to a standard optimal stopping problem:

$$
\begin{aligned}
v^N(t,\mathbf{x},\mathbf{i}) &= \sup_{\theta\in\mathcal{T}^N_{t,T}} \mathbb{E}_{t,\mathbf{y}}\Big[\frac{1}{N}\sum_{k\in A(\mathbf{i})}\int_t^\theta f^k(r,\bar{\mathbf{X}}_r)\,dr + \hat{v}^N(\theta,\bar{\mathbf{X}}_\theta,\mathbf{i})\Big],\\
\hat{v}^N(t,\mathbf{x},\mathbf{i}) &:= \sup_{\mathbf{i}'<\mathbf{i}} v^N(t,\mathbf{x},\mathbf{i}').
\end{aligned}
\tag{7.5}
$$

Notice that $\hat{v}^N(t,\mathbf{x},\mathbf{i})$ is continuous in $(t,\mathbf{x})$, as $|A(\mathbf{i}')| \le n$. Then it follows from Ekren [24] that $v^N(\cdot,\cdot,\mathbf{i})$ is a continuous viscosity solution of (3.3) in the sense of Definition 3.2. [2] Moreover, note that a viscosity sub/super solution in the sense of Definition 3.2 implies it is a viscosity sub/super solution in the classical Crandall–Lions sense. Then the comparison principle for viscosity semisolutions in the Crandall–Lions sense implies that in our sense, see Remark 3.4, and hence v^N is the unique viscosity solution to (3.3). □

7.2. *The symmetric case.* We now prove Lemma 3.5 and Theorem 3.7 in the symmetric case (3.6).

PROOF OF LEMMA 3.5. (i) Recall (3.1). For $k \in [N]$, by standard SDE estimates we have

$$
\mathbb{E}^{\mathbb{P}^0}\big[|X^{\boldsymbol{\tau},k}_s|^2\big] \le \mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t\le r\le s}|X^{\boldsymbol{\tau},k}_r|^2\Big] \le C\mathbb{E}^{\mathbb{P}^0}\Big[|x_k|^2 + \int_t^s (1+\|\mathbf{X}^\tau_r\|_2^2)\,dr\Big].
\tag{7.6}
$$

Then

$$
\mathbb{E}^{\mathbb{P}^0}\big[\|\mathbf{X}^{\boldsymbol{\tau}}_s\|_2^2\big] \le C\mathbb{E}^{\mathbb{P}^0}\Big[\|\mathbf{x}\|_2^2 + 1 + \int_t^s \|\mathbf{X}^{\boldsymbol{\tau}}_r\|_2^2\,dr\Big].
$$

Applying the Gronwall inequality, we have $\mathbb{E}^{\mathbb{P}^0}[\sup_{t\le s\le T}\|\mathbf{X}^{\boldsymbol{\tau}}_s\|_2 \le C(1+\|\mathbf{x}\|_2)]$. Plug this into the second inequality of (7.6), we obtain (3.7) immediately.

Moreover, for any $\theta_1,\theta_2 \in \mathcal{T}^N_{[t,T]}$ with $\theta_1 \le \theta_2$, by standard SDE estimates we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{\theta_1\le s\le\theta_2}|X^{\boldsymbol{\tau},k}_s - X^{\boldsymbol{\tau},k}_{\theta_1}|^2\Big] &\le C\mathbb{E}^{\mathbb{P}^0}\Big[\int_{\theta_1}^{\theta_2}(1+|X^{\boldsymbol{\tau},k}_r|^2 + \|\mathbf{X}^{\boldsymbol{\tau}}_r\|^2)\,dr\Big]\\
&\le C\|\theta_2-\theta_1\|_\infty\mathbb{E}^{\mathbb{P}^0}\Big[1+\sup_{t\le r\le T}|X^{\boldsymbol{\tau},k}_r|^2 + \frac{1}{N}\sum_{l=1}^N\sup_{t\le r\le T}|X^{\boldsymbol{\tau},l}_r|^2\Big].
\end{aligned}
$$

Plug (3.7) into it, we obtain (3.8).

(ii) We now assume Assumption 2.6(i) also holds true. Let $\delta > 0$ be a small number which will be specified later, and consider a partition $t = t_0 < \cdots < t_n = T$ such that $\Delta t_j \le \delta$ for all j. For each $j = 1,\dots,n$ and $k = 1,\dots,N$, by standard SDE estimates we have, for a generic constant C which is independent of N and δ,

$$
\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}|X^{\boldsymbol{\tau},k}_s|\Big] \le C\mathbb{E}^{\mathbb{P}^0}\Big[|X^{\boldsymbol{\tau},k}_{t_j}| + \int_{t_j}^{t_{j+1}}(1+|X^{\boldsymbol{\tau},k}_s| + \|\mathbf{X}^{\boldsymbol{\tau}}_s\|_1)\,ds
$$

[2] Here again, we notice that our conditions are slightly different from Ekren and that the viscosity solution property is easily checked to be valid in our context.

$$+\left(\int_{t_j}^{t_{j+1}}(1+|X_s^{\boldsymbol{\tau},k}|+\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1)^2\,ds\right)^{\frac{1}{2}}\Big]$$
$$\le C\mathbb{E}^{\mathbb{P}^0}\Big[|X_{t_j}^{\boldsymbol{\tau},k}|+1+\sqrt{\delta}\sup_{t_j\le s\le t_{j+1}}\big[|X_s^{\boldsymbol{\tau},k}|+\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\big]\Big].$$

Set $\delta := \frac{1}{9C^2}$ for the above C, we have

$$\frac{2}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}|X_s^{\boldsymbol{\tau},k}|\Big]\le C\mathbb{E}^{\mathbb{P}^0}\big[|X_{t_j}^{\boldsymbol{\tau},k}|+1\big]+\frac{1}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big]. \tag{7.7}$$

This implies that

$$\frac{2}{3}\mathbb{E}^{\mathbb{P}^0}\left[\frac{1}{N}\sum_{k=1}^N\sup_{t_j\le s\le t_{j+1}}|X_s^{\boldsymbol{\tau},k}|\right]\le C\mathbb{E}^{\mathbb{P}^0}\big[\|\mathbf{X}_{t_j}^{\boldsymbol{\tau}}\|_1+1\big]+\frac{1}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big],$$

and thus

$$\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big]\le\mathbb{E}^{\mathbb{P}^0}\left[\frac{1}{N}\sum_{k=1}^N\sup_{t_j\le s\le t_{j+1}}|X_s^{\boldsymbol{\tau},k}|\right]\le C\mathbb{E}^{\mathbb{P}^0}\big[\|\mathbf{X}_{t_j}^{\boldsymbol{\tau}}\|_1+1\big].$$

By induction on $j=0,\dots,n-1$, one can easily obtain

$$\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t\le s\le T}\|\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big]\le C(1+\|\mathbf{x}\|_1).$$

Plugging this into (7.7) and by induction on $j=0,\dots,n-1$ again, we prove (3.9). Moreover, (3.10) follows similar arguments as for (3.8). □

PROOF OF THEOREM 3.7. It is obvious that Assumptions 2.1, 2.6 and (3.6) imply Assumption 3.1. In particular, the uniform continuity of g^N comes from the fact that g is uniformly continuous on all bounded parts of $\mathcal{P}_2(\mathbb{R}^d)$, which are $\mathcal{W}_1$-compact. Then (i) follows directly from Theorem 3.3.

To see (ii), assume $m_N = m^N(\mathbf{y})$ and recall (3.2) and (3.11). By the linear growth of f and g and by (3.7), for any $\boldsymbol{\tau}$ we have

$$|J_N(t,\mathbf{y},\boldsymbol{\tau})|\le\mathbb{E}^{\mathbb{P}^0}\left[\frac{C}{N}\sum_{k=1}^N\int_t^T(1+|X_r^{\boldsymbol{\tau},k}|+\|\mathbf{X}_r^{\boldsymbol{\tau}}\|_1)\,dr+1+\|\mathbf{X}_T^{\boldsymbol{\tau}}\|_1\right]$$
$$\le C+\frac{C}{N}\sum_{k=1}^N\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t\le r\le T}\|\mathbf{X}_r^{\boldsymbol{\tau}}\|_1\Big]\le C(1+\|\mathbf{x}\|_1)\le C(1+\|m_N\|_1).$$

Then by the arbitrariness of $\boldsymbol{\tau}$ we prove (3.12).

It remains to prove (iii). We proceed in three steps.

Step 1. By Birkhoff's theorem, the Kantorovitch solution of the optimal transport problem in the finite discrete case is a Monge solution, that is, the optimal coupling measure is concentrated on a graph; see, for example, Villani [56] Page 5. Then, for any $t\in[0,T]$, $\mathbf{y},\tilde{\mathbf{y}}\in\mathbf{S}^N$:

$$\mathcal{W}_2^2(m^N(\mathbf{y}),m^N(\tilde{\mathbf{y}}))=\min_{\pi\in\mathfrak{S}_N}\frac{1}{N}\sum_{k=1}^N|y_k-\tilde{y}_{\pi(k)}|^2,$$

where $\mathfrak{S}_N$ is the set of permutations on $[N]$. By possibly reordering the y_k's, we may then assume without loss of generality that

$$\mathcal{W}_2(m^N(\mathbf{y}),m^N(\tilde{\mathbf{y}}))=\|\mathbf{y}-\tilde{\mathbf{y}}\|_2. \tag{7.8}$$

Given $\boldsymbol{\tau}$, let $\mathbf{Y}^{\boldsymbol{\tau}}$ and $\tilde{\mathbf{Y}}^{\boldsymbol{\tau}}$ be defined by (3.1) with initial condition $\mathbf{y}$ and $\tilde{\mathbf{y}}$, respectively, and denote $\Delta\mathbf{Y}^{\boldsymbol{\tau}} := \mathbf{Y}^{\boldsymbol{\tau}} - \tilde{\mathbf{Y}}^{\boldsymbol{\tau}}$, $\Delta\mathbf{y} := \mathbf{y} - \tilde{\mathbf{y}}$. Recall that $\boldsymbol{\tau}$ is adapted to the Brownian filtration, namely in the spirit of open loop controls, then

$$\Delta I_s^{\boldsymbol{\tau},k} = \Delta i_k \mathbf{1}_{\{\tau_k > s\}} \quad \text{and thus} \quad \|\Delta\mathbf{I}_t^{\boldsymbol{\tau}}\|_2 \le \|\Delta\mathbf{i}\|_2 \quad \forall t \in [0,T]. \tag{7.9}$$

In this step we shall prove further that

$$\mathbb{E}^{\mathbb{P}^0}\Big[\frac{1}{N}\sum_{k=1}^N \sup_{t\le s\le T}|\Delta X_s^{\boldsymbol{\tau},k}|\Big] \le C(1+\|\mathbf{x}\|_2)\|\Delta\mathbf{y}\|_2. \tag{7.10}$$

Indeed, when $i_k = 1$ and $\tilde{i}_k = 0$, we have $\tilde{\mathbf{X}}_s^{\boldsymbol{\tau},k} \equiv \tilde{x}_k$, and thus, by (3.10),

$$\begin{aligned}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t\le s\le T}|\Delta X_s^{\boldsymbol{\tau},k}|\Big] &\le C\mathbb{E}^{\mathbb{P}^0}\Big[|\Delta x_k| + \sup_{t\le s\le T}|X_s^{\boldsymbol{\tau},k} - x_k|\Big] \\ &\le C(1+|x_k| + \|\mathbf{x}\|_1 + |\Delta x_k|).\end{aligned} \tag{7.11}$$

Similarly, when $i_k = 0$ and $\tilde{i}_k = 1$, the above still holds true. Moreover, when $\Delta i_k = 0$, we have $\Delta I_s^{\boldsymbol{\tau},k} \equiv 0$. Let δ and $t = t_0 < \cdots < t_n = T$ be as in the proof of Lemma 3.5(ii). Then, for each $j = 0,\dots,n-1$, similar to (7.7) we have

$$\frac{2}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}|\Delta X_s^{\boldsymbol{\tau},k}|\Big] \le \frac{1}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}\|\Delta\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big] + C\mathbb{E}^{\mathbb{P}^0}[|\Delta X_{t_j}^{\boldsymbol{\tau},k}|].$$

This, together with (7.11) which is for all k such that $\Delta i_k \ne 0$, implies that

$$\begin{aligned}\frac{2}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\frac{1}{N}\sum_{k=1}^N \sup_{t_j\le s\le t_{j+1}}|\Delta X_s^{\boldsymbol{\tau},k}|\Big] &\le \frac{1}{3}\mathbb{E}^{\mathbb{P}^0}\Big[\sup_{t_j\le s\le t_{j+1}}\|\Delta\mathbf{X}_s^{\boldsymbol{\tau}}\|_1\Big] + C\mathbb{E}^{\mathbb{P}^0}[\|\Delta X_{t_j}^{\boldsymbol{\tau}}\|_1] \\ &\quad + \frac{C}{N}\sum_{k=1}^N(1+|x_k|+\|\mathbf{x}\|_1+|\Delta x_k|)|\Delta i_k|.\end{aligned}$$

Then one can easily get

$$\mathbb{E}^{\mathbb{P}^0}\Big[\frac{1}{N}\sum_{k=1}^N \sup_{t\le s\le T}|\Delta X_s^{\boldsymbol{\tau},k}|\Big] \le C\big[\|\Delta\mathbf{y}\|_1 + (\|\mathbf{x}\|_2 + \|\Delta\mathbf{x}\|_2)\|\Delta\mathbf{i}\|_2\big],$$

which implies (7.10) immediately.

Step 2. In this step we prove the claimed regularity of v_N in terms of $\mathbf{y}$. Fix $R > 1$ and let $\tilde{R} > R$ be a large number which will be specified later. Denote

$$D_{\tilde{R}} := [0,T] \times \big\{(x,m) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) : |x| + \|m\|_2 \le \tilde{R}\big\}.$$

Note in particular that any bounded subset of $\mathcal{P}_2(\mathbf{S})$ is tight in $\mathcal{P}_1(\mathbf{S})$. Then f and g are uniformly continuous for $\mathcal{W}_1$ on $D_{\tilde{R}}$, and let $\tilde{\rho}_{\tilde{R}}$ denote their common modulus of continuity function. Then, by the linear growth of f and g,

$$\begin{aligned}|f(t,x,m) - f(t,\tilde{x},\tilde{m})| &\le \tilde{\rho}_{\tilde{R}}(|\Delta x| + \mathcal{W}_1(m,\tilde{m})) \\ &\quad + C(1+|x|+\|m\|_1)\mathbf{1}_{\{|x|+\|m\|_2\ge\tilde{R}\}} \\ &\quad + C(1+|\tilde{x}|+\|\tilde{m}\|_1)\mathbf{1}_{\{|\tilde{x}|+\|\tilde{m}\|_2\ge\tilde{R}\}}] \\ &\le \tilde{\rho}_{\tilde{R}}(|\Delta x| + \mathcal{W}_1(m,\tilde{m})) + \frac{C}{\tilde{R}}(1+|x|^2+\|m\|_2^2+|\tilde{x}|^2+\|\tilde{m}\|_2^2),\end{aligned}$$

$$|g(m) - g(\tilde{m})| \le \tilde{\rho}_{\tilde{R}}(\mathcal{W}_1(m,\tilde{m})) + \frac{C}{\tilde{R}}(1+\|m\|_2^2+\|\tilde{m}\|_2^2).$$

We now fix $t \in [0,T]$, $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbf{S}^N$ with $\|\mathbf{y}\|_2, \|\tilde{\mathbf{y}}\|_2 \le R$. By (3.9) we have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}^0}\big[\big|f\big(s, X_s^{\boldsymbol{\tau},k}, m^N(\mathbf{Y}_s^{\boldsymbol{\tau}})\big) - f\big(s, \tilde{X}_s^{\boldsymbol{\tau},k}, m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big|\big] \\
&\quad \le \mathbb{E}^{\mathbb{P}^0}\bigg[\tilde{\rho}_{\tilde{R}}\big(\big|\Delta X_s^{\boldsymbol{\tau},k}\big| + \mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big) \\
&\qquad + \frac{C}{\tilde{R}}\big(1 + |X_s^{\boldsymbol{\tau},k}|^2 + \|m^N(\mathbf{Y}_s^{\boldsymbol{\tau}})\|_2^2 + |\tilde{X}_s^{\boldsymbol{\tau},k}|^2 + \|m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\|_2^2\big)\bigg] \\
&\quad \le \mathbb{E}^{\mathbb{P}^0}\big[\tilde{\rho}_{\tilde{R}}\big(\big|\Delta X_s^{\boldsymbol{\tau},k}\big| + \mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big)\big] + \frac{C(1+R^2)}{\tilde{R}}, \\
&\mathbb{E}^{\mathbb{P}^0}\big[\big|g\big(m^N(\mathbf{Y}_T^{\boldsymbol{\tau}})\big) - g\big(m^N(\tilde{\mathbf{Y}}_T^{\boldsymbol{\tau}})\big)\big|\big] \\
&\quad \le \mathbb{E}^{\mathbb{P}^0}\big[\tilde{\rho}_{\tilde{R}}\big(\mathcal{W}_1\big(m^N(\mathbf{Y}_T^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_T^{\boldsymbol{\tau}})\big)\big)\big] + \frac{C(1+R^2)}{\tilde{R}}.
\end{aligned} \tag{7.12}$$

Then

$$\begin{aligned}
\big|J_N(t,\mathbf{y},\boldsymbol{\tau}) - J_N(t,\tilde{\mathbf{y}},\boldsymbol{\tau})\big| \le \mathbb{E}^{\mathbb{P}^0}\bigg[&\frac{1}{N}\sum_{k=1}^N \int_t^{\tau_k} \tilde{\rho}_{\tilde{R}}\big(\big|\Delta X_s^{\boldsymbol{\tau},k}\big| + \mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big)\,ds \\
&+ \tilde{\rho}_{\tilde{R}}\big(\mathcal{W}_1\big(m^N(\mathbf{Y}_T^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_T^{\boldsymbol{\tau}})\big)\big)\bigg] + \frac{C(1+R^2)}{\tilde{R}}.
\end{aligned}$$

Now for any $\varepsilon > 0$, choose $\tilde{R} = \tilde{R}_\varepsilon := \frac{2C(1+R^2)}{\varepsilon}$. By (7.9) and (7.10), and observing that

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}^0}\big[\tilde{\rho}_{\tilde{R}_\varepsilon}\big(\mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big)\big] \\
&\le \tilde{\rho}_{\tilde{R}_\varepsilon}(\gamma) + \frac{1}{\sqrt{\gamma}}\sqrt{\mathbb{E}^{\mathbb{P}^0}\big[\tilde{\rho}^2_{\tilde{R}_\varepsilon}\big(\mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big)\big]\mathbb{E}^{\mathbb{P}^0}\big[\mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big]}
\end{aligned} \tag{7.13}$$

for all $\gamma > 0$ and $s \in [t,T]$, one can easily see that there exists $\delta > 0$, depending on $\tilde{\rho}_{\tilde{R}_\varepsilon}$ and hence on ε, but not on N or $\boldsymbol{\tau}$, such that, whenever $\|\Delta\mathbf{y}\|_2 < \delta$,

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}^0}\bigg[&\frac{1}{N}\sum_{k=1}^N \int_t^{\tau_k} \tilde{\rho}_{\tilde{R}_\varepsilon}\big(\big|\Delta X_s^{\boldsymbol{\tau},k}\big| + \mathcal{W}_1\big(m^N(\mathbf{Y}_s^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_s^{\boldsymbol{\tau}})\big)\big)\,ds \\
&+ \tilde{\rho}_{\tilde{R}_\varepsilon}\big(\mathcal{W}_1\big(m^N(\mathbf{Y}_T^{\boldsymbol{\tau}}), m^N(\tilde{\mathbf{Y}}_T^{\boldsymbol{\tau}})\big)\big)\bigg] \le \frac{\varepsilon}{2}.
\end{aligned}$$

Thus $|J_N(t,\mathbf{y},\boldsymbol{\tau}) - J_N(t,\tilde{\mathbf{y}},\boldsymbol{\tau})| \le \varepsilon$. By the arbitrariness of $\boldsymbol{\tau}$, we obtain

$$\big|v_N(t,\mathbf{y}) - v_N(t,\tilde{\mathbf{y}})\big| \le \varepsilon \quad \text{whenever } \|\mathbf{y}\|_2, \|\tilde{\mathbf{y}}\|_2 \le R,\ \|\Delta\mathbf{y}\|_2 < \delta.$$

This implies (3.13) when $t = \tilde{t}$, that is, for a modulus of continuity function ρ'_R,

$$\big|v_N(t,\mathbf{y}) - v_N(t,\tilde{\mathbf{y}})\big| \le \rho'_R\big(\mathcal{W}_2\big(m^N(\mathbf{y}), m^N(\tilde{\mathbf{y}})\big)\big) \quad \text{whenever } \|\mathbf{y}\|_2, \|\tilde{\mathbf{y}}\|_2 \le R. \tag{7.14}$$

Step 3. Fix $t < \tilde{t}$ and $\mathbf{y} \in \mathbf{S}^N$ with $\|\mathbf{y}\|_2 \le R$. For any $\boldsymbol{\tau} \in \mathcal{T}^N_{\tilde{t},T}$, by (3.2) we have

$$\begin{aligned}
&J_N(\tilde{t},\mathbf{y},\boldsymbol{\tau}) - v_N(t,\mathbf{y}) \\
&\quad \le J_N(\tilde{t},\mathbf{y},\boldsymbol{\tau}) - J_N(t,\mathbf{y},\boldsymbol{\tau}) \\
&\quad = J_N(\tilde{t},\mathbf{y},\boldsymbol{\tau}) - \mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\bigg[\frac{1}{N}\sum_{k=1}^N \int_t^{\tilde{t}} f\big(r, X_r^{\boldsymbol{\tau},k}, m^N(\mathbf{Y}_r^{\boldsymbol{\tau}})\big) I_r^{\boldsymbol{\tau},k}\,dr + J_N\big(\tilde{t}, \mathbf{Y}_{\tilde{t}}^{\boldsymbol{\tau}}, \boldsymbol{\tau}\big)\bigg]
\end{aligned}$$

$$\le J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[J_N\big(\tilde t,\mathbf{Y}^{\boldsymbol\tau}_{\tilde t},\boldsymbol\tau\big)\big]+\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\bigg[\frac{C}{N}\sum_{k=1}^N\int_t^{\tilde t}\big(1+|X^{\boldsymbol\tau,k}_r|+\|\mathbf{Y}^{\boldsymbol\tau}_r\|_1\big)\,dr\bigg]$$

$$\le J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[J_N\big(\tilde t,\mathbf{Y}^{\boldsymbol\tau}_{\tilde t},\boldsymbol\tau\big)\big]+C(1+R)\sqrt{\tilde t-t},$$

where the last inequality is due to (3.7). Again let $\tilde R>R$ be a large number which will be specified later. By (7.14), (3.12) and (3.7) we have

$$\begin{aligned}
&J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[J_N\big(\tilde t,\mathbf{Y}^{\boldsymbol\tau}_{\tilde t},\boldsymbol\tau\big)\big]\\
&\quad\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\rho'_{\tilde R}\big(\mathcal{W}_2\big(m^N(\mathbf{y}),m^N\big(\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}\big)\big)\big)+\big(\big|J_N(\tilde t,\mathbf{y},\boldsymbol\tau)\big|+\big|J_N\big(\tilde t,\mathbf{Y}^{\boldsymbol\tau}_{\tilde t},\boldsymbol\tau\big)\big|\big)\mathbf{1}_{\{\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}\|_2>\tilde R\}}\big]\\
&\quad\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\bigg[\rho'_{\tilde R}\big(\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big)+\frac{C}{\tilde R}\big(\big|J_N(\tilde t,\mathbf{y},\boldsymbol\tau)\big|^2+\big|J_N\big(\tilde t,\mathbf{Y}^{\boldsymbol\tau}_{\tilde t},\boldsymbol\tau\big)\big|^2+\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}\big\|_2^2\big)\bigg]\\
&\quad\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\bigg[\rho'_{\tilde R}\big(\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big)+\frac{C}{\tilde R}\big(\|\mathbf{y}\|_2^2+\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}\big\|_2^2\big)\bigg]\\
&\quad\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\rho'_{\tilde R}\big(\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big)\big]+\frac{C(1+R^2)}{\tilde R}.
\end{aligned}$$

Then

$$J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-v_N(t,\mathbf{y})\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\rho'_{\tilde R}\big(\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big)\big]+\frac{C(1+R^2)}{\tilde R}+C[1+R]\sqrt{\tilde t-t}.$$

Now for any $\varepsilon>0$, set $\tilde R=\tilde R_\varepsilon:=\frac{3C(1+R^2)}{\varepsilon}$, and $\delta>0$ small enough. Then for $\tilde t-t\le\delta$,

$$J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-v_N(t,\mathbf{y})\le\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\rho'_{\tilde R}\big(\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big)\big]+\frac{2\varepsilon}{3}.$$

Since $\boldsymbol\tau\in\mathcal{T}^N_{\tilde t,T}$, we have $\tau_k\ge\tilde t$ for all k, then $\mathbf{I}^{\boldsymbol\tau}_{\tilde t-}=\mathbf{i}$. Thus, by (3.8),

$$\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\big\|\mathbf{Y}^{\boldsymbol\tau}_{\tilde t}-\mathbf{y}\big\|_2\big]=\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\big[\big\|\mathbf{X}^{\boldsymbol\tau}_{\tilde t}-\mathbf{x}\big\|_2\big]\le C(1+R)(\tilde t-t)\le C(1+R)\delta.$$

Using an estimate similar to (7.13), one can then see that, for $\delta>0$ sufficiently small, which may depend on R and $\tilde R_\varepsilon$ and hence on ε, but not on $\boldsymbol\tau$ or N, we have

$$J_N(\tilde t,\mathbf{y},\boldsymbol\tau)-v_N(t,\mathbf{y})\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.$$

Then, by the arbitrariness of $\boldsymbol\tau\in\mathcal{T}^N_{\tilde t,T}$, we have

$$v_N(\tilde t,\mathbf{y})-v_N(t,\mathbf{y})\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.\tag{7.15}$$

On the other hand, for any $\boldsymbol\tau\in\mathcal{T}^N_{t,T}$, note that $\boldsymbol\tau\vee\tilde t\in\mathcal{T}^N_{\tilde t,T}$. Then similarly we have

$$J_N(t,\mathbf{y},\boldsymbol\tau\vee\tilde t)-v_N(\tilde t,\mathbf{y})\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.\tag{7.16}$$

Note also that,

$$\mathcal{W}_1\big(m^N\big(\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau}\big),m^N\big(\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau\vee\tilde t}\big)\big)\le\big\|\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau}-\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau\vee\tilde t}\big\|_1\le C\sup_{t\le s\le\tilde t}\big\|\mathbf{X}^{\boldsymbol\tau}_s-\mathbf{x}\big\|_1.$$

Then, by (3.9), (3.10) and the last estimate in (7.12), we have

$$\begin{aligned}
&\big|J_N(t,\mathbf{y},\boldsymbol\tau)-J_N(t,\mathbf{y},\boldsymbol\tau\vee\tilde t)\big|\\
&\quad=\bigg|\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\bigg[\frac1N\sum_{k=1}^N\int_{\tau_k}^{\tau_k\vee\tilde t}f\big(r,X^{\boldsymbol\tau,k}_r,m^N\big(\mathbf{Y}^{\boldsymbol\tau}_r\big)\big)\,dr+g\big(m^N\big(\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau}\big)\big)-g\big(m^N\big(\mathbf{X}^{\boldsymbol\tau}_{\boldsymbol\tau\vee\tilde t}\big)\big)\bigg]\bigg|
\end{aligned}$$

$$\le C\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^{N}\int_t^{\tilde t}\big(1+|X_r^{\boldsymbol\tau,k}|+\|\mathbf{Y}_r^{\boldsymbol\tau}\|_1\big)\,dr+\tilde\rho_{\tilde R}\big(\mathcal{W}_1\big(m^N(\mathbf{X}_{\boldsymbol\tau}^{\boldsymbol\tau}),m^N(\mathbf{X}_{\boldsymbol\tau\vee\tilde t}^{\boldsymbol\tau})\big)\right]$$
$$+\frac{C(1+R^2)}{\tilde R}$$
$$\le C\mathbb{E}^{\mathbb{P}^0}_{t,\mathbf{y}}\Big[\tilde\rho_{\tilde R}\Big(C\sup_{t\le s\le\tilde t}\|\mathbf{X}_s^{\boldsymbol\tau}-\mathbf{x}\|_1\Big)\Big]+\frac{C(1+R^2)}{\tilde R}+C[1+R]\sqrt{\tilde t-t}.$$

This, together with (7.16), implies that, for a possibly smaller δ,

$$J_N(t,\mathbf{y},\boldsymbol\tau)-v_N(\tilde t,\mathbf{y})\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.$$

Since $\boldsymbol\tau\in\mathcal{T}_{t,T}^N$ is arbitrary, we have

$$v_N(t,\mathbf{y})-v_N(\tilde t,\mathbf{y})\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.$$

Then by (7.15) we obtain

$$\big|v_N(t,\mathbf{y})-v_N(\tilde t,\mathbf{y})\big|\le\varepsilon\quad\text{whenever }\|\mathbf{y}\|_2\le R,\tilde t-t\le\delta.$$

This, together with (7.14) and (7.8), proves (3.13). □

7.3. *Construction of an optimal stopping strategy.* In this subsection we come back to the general case in Section 3.1. We first recall (7.4), and note that, by (7.2),

$$\max_{\mathbf{i}'<\mathbf{i}}v^N(t,\mathbf{x},\mathbf{i}')=\max_{k\in A(\mathbf{i})}v^N(t,\mathbf{x},\mathbf{i}^{-k}).\tag{7.17}$$

We shall use the above relation to construct an optimal $\hat{\boldsymbol\tau}\in\mathcal{T}_{0,T}^N$ for the problem (3.2). We also refer to [54], Section 4.3, for some heuristic discussions on the optimal stopping policy on the mean field problem (2.3).

Fix $\mathbf{y}\in\mathbf{S}^N$ and assume for simplicity $t=0$. We set

$$\hat{\boldsymbol\tau}:=\boldsymbol\tau^{(|A(\mathbf{i})|)},\tag{7.18}$$

where $\boldsymbol\tau^{(j)}\in\mathcal{T}_{0,T}^N$, $j=0,\dots,|A(\mathbf{i})|$, are constructed as follows.

Step 0 (Initialization). Set $\tau_k^{(0)}:=T\mathbf{1}_{A(\mathbf{i})}(k)$. That is, $\mathbf{Y}^{\boldsymbol\tau^{(0)}}=\bar{\mathbf{Y}}$ is the unstopped process. Moreover, set $A_0:=A(\mathbf{i})$ and $\tau_0^*:=0$.

We then stop the particles in $A(\mathbf{i})$ recursively: for $j=0,\dots,|A(\mathbf{i})|-1$:

Step 1. The next particle is stopped at: denoting $\mathbf{I}^A:=(1_A(1),\dots,1_A(N))$ for $A\subset[N]$,

$$\tau_{j+1}^*:=\inf\Big\{s\ge\tau_j^*:v^N\big(s,\mathbf{Y}_s^{\boldsymbol\tau^{(j)}}\big)=\max_{\mathbf{i}'<\mathbf{I}^{A_j}}v^N\big(s,\mathbf{X}_s^{\boldsymbol\tau^{(j)}},\mathbf{i}'\big)\Big\}.\tag{7.19}$$

We remark that it is possible that $\tau_{j+1}^*=\tau_j^*$, namely two (or more) particles are stopped at the same time.

Step 2. By (7.17), we may pick an index for the particle stopped at τ_{j+1}^*:

$$\begin{aligned}\kappa_{j+1}^*&:=\min\big\{k\in A_j:v^N\big(s,\mathbf{X}_s^{\boldsymbol\tau^{(j)}},(\mathbf{I}^{A_j})^{-k}\big)=v^N\big(s,\mathbf{Y}_s^{\boldsymbol\tau^{(j)}}\big)\big\},\\ A_{j+1}&:=A_j\backslash\{\kappa_{j+1}^*\},\end{aligned}\tag{7.20}$$

where $\mathbf{I}^{-k}$ is defined in (7.3), and $A_{j+1}\in\mathcal{F}_{\tau_{j+1}^*}^N$ is random. Note that there might be multiple k satisfying the above requirement and serving for our purpose. We choose the smallest one just for convenience.

Step 3. We then update the vector of the stopping times:

$$\tau_{\kappa_{j+1}^*}^{(j+1)}:=\tau_{j+1}^*,\quad\text{and}\quad\tau_k^{(j+1)}:=\tau_k^{(j)}\quad\text{for all }k\ne\kappa_{j+1}^*.$$

THEOREM 7.2. *Under Assumption 3.1, the $\hat{\boldsymbol{\tau}}$ constructed in (7.18) is an optimal stopping policy for the problem (3.2).*

PROOF. We prove by induction on $|A(\mathbf{i})|$. When $|A(\mathbf{i})| = 0$, all particles are already stopped. Then it is trivial that $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}^{(0)} = (0, \dots, 0)$ is optimal. Assume the result is true for all $\mathbf{i}'$ with $|A(\mathbf{i}')| \le n$, and now consider $\mathbf{i}$ such that $|A(\mathbf{i})| = n + 1$.

Recall (7.1) and note that $\mathbf{Y}^{\boldsymbol{\tau}^{(0)}} = \bar{\mathbf{Y}}$, $\mathbf{I}^{A_0} = \mathbf{i}$. Then it follows from the standard optimal stopping theory that τ_1^* constructed in (7.19) is an optimal stopping strategy for (7.1) with $t = 0$. That is,

$$\begin{aligned} v^N(0, \mathbf{y}) &= \mathbb{E}_{0,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^{N}\int_0^{\tau_1^*} f^k(r, \bar{\mathbf{X}}_r)\bar{I}_r^k\, dr + \max_{\mathbf{i}'<\mathbf{i}} v^N\big(\tau_1^*, \bar{\mathbf{X}}_{\tau_1^*}, \mathbf{i}'\big)\right] \\ &= \mathbb{E}_{0,\mathbf{y}}\left[\frac{1}{N}\sum_{k\in A_0}\int_0^{\tau_1^*} f^k\big(r, \mathbf{X}_r^{\boldsymbol{\tau}^{(0)}}\big)\, dr + v^N\big(\tau_1^*, \mathbf{X}_{\tau_1^*}^{\boldsymbol{\tau}^{(0)}}, \mathbf{I}^{A_1}\big)\right], \end{aligned}$$

where the last equality follows from (7.20). Our construction also induces that

$$\mathbf{X}_r^{\hat{\tau}} = \mathbf{X}_r^{\boldsymbol{\tau}^{(j)}}, \quad \tau_j^* \le r \le \tau_{j+1}^*, \quad \text{and} \quad \mathbf{I}_r^{\hat{\boldsymbol{\tau}}} = \mathbf{I}^{A_j}, \quad \tau_j^* \le r < \tau_{j+1}^*.$$

Then

$$v^N(0, \mathbf{y}) = \mathbb{E}_{0,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^{N}\int_0^{\tau_1^*} f^k\big(r, \mathbf{X}_r^{\hat{\boldsymbol{\tau}}}\big)I_r^{\hat{\boldsymbol{\tau}},k}\, dr + v^N\big(\tau_1^*, \mathbf{X}_{\tau_1^*}^{\hat{\boldsymbol{\tau}}}, \mathbf{I}^{A_1}\big)\right].$$

Since $|A_1| = n$, by induction assumption we have

$$v^N\big(\tau_1^*, \mathbf{X}_{\tau_1^*}^{\hat{\boldsymbol{\tau}}}, \mathbf{I}^{A_1}\big) = \mathbb{E}_{\tau_1^*, \mathbf{X}_{\tau_1^*}^{\hat{\boldsymbol{\tau}}}, \mathbf{I}^{A_1}}\left[\frac{1}{N}\sum_{k=1}^{N}\int_{\tau_1^*}^{T} f^k\big(r, \mathbf{X}_r^{\hat{\boldsymbol{\tau}}}\big)I_r^{\hat{\boldsymbol{\tau}},k}\, dr + g_N\big(\mathbf{X}_T^{\hat{\boldsymbol{\tau}}}\big)\right],$$

thus proving the optimality of $\hat{\boldsymbol{\tau}}$:

$$v^N(0, \mathbf{y}) = \mathbb{E}_{0,\mathbf{y}}\left[\frac{1}{N}\sum_{k=1}^{N}\int_0^{T} f^k\big(r, \mathbf{X}_r^{\hat{\boldsymbol{\tau}}}\big)I_r^{\hat{\boldsymbol{\tau}},k}\, dr + g_N\big(\mathbf{X}_T^{\hat{\boldsymbol{\tau}}}\big)\right]. \qquad \square$$

APPENDIX

In this appendix, we present a few technical proofs.

PROOF OF LEMMA 2.2. Let $m' \preceq m$ with corresponding $p(x)$ as in (2.7). For any $n \ge 1$, let $\{O_k^n\}_{k\ge1}$ be a partition of $\mathbb{R}^d$ such that, for each $k \ge 1$, $|x - \tilde{x}| \le \frac{1}{n}$ for all $x, \tilde{x} \in O_k^n$. By the required continuity of $m(dx, 1)$, one may easily construct $A_k^n \subset O_k^n$ such that

$$m\big(A_k^n, 1\big) = \int_{O_k^n} p(x)m(dx, 1).$$

Set $A^n := \bigcup_{k\ge1} A_k^n$ and fix an $x_k^n \in A_k^n$. Then, for any $\varphi : \mathbf{S} \to \mathbb{R}$ which is Lipschitz continuous with Lipschitz constant 1, we have: denoting $\Delta\varphi(x) := \varphi(x, 1) - \varphi(x, 0)$,

$$\begin{aligned} &\int_{\mathbf{S}} \varphi(\mathbf{y})m^{A^n}(dy) - \int_{\mathbf{S}} \varphi(\mathbf{y})m'(dy) \\ &\quad= \int_{A^n} \varphi(x, 1)m(dx, 1) + \int_{(A^n)^c} \varphi(x, 0)m(dx, 1) + \int_{\mathbb{R}^d} \varphi(x, 0)m(dx, 0) \\ &\qquad- \int_{\mathbb{R}^d} \varphi(x, 1)p(x)m(dx, 1) - \int_{\mathbb{R}^d} \big(\varphi(x, 0)\big[1 - p(x)\big]m(dx, 1) + \varphi(x, 0)m(dx, 0)\big) \end{aligned}$$

$$
\begin{aligned}
&= \int_{A^n} \Delta\varphi(x) m(dx,1) - \int_{\mathbb{R}^d} \Delta\varphi(x) p(x) m(dx,1) \\
&= \sum_{k\geq 1} \Big(\int_{A_k^n} \Delta\varphi(x) m(dx,1) - \int_{O_k^n} \Delta\varphi(x) p(x) m(dx,1) \Big) \\
&= \sum_{k\geq 1} \Big(\int_{A_k^n} \big(\Delta\varphi(x) - \Delta\varphi(x_k^n)\big) m(dx,1) - \int_{O_k^n} \big(\Delta\varphi(x) - \Delta\varphi(x_k^n)\big) p(x) m(dx,1) \Big) \\
&\leq \frac{C}{n} \sum_{k\geq 1} \Big(\int_{A_k^n} m(dx,1) + \int_{O_k^n} p(x) m(dx,1) \Big) \leq \frac{C}{n}.
\end{aligned}
$$

This implies that $\mathcal{W}_1(m^{A_n}, m') \leq \frac{C}{n}$, and thus $m^{A_n} \to m'$ weakly as $n \to \infty$. Moreover, it is obvious that $\{m^{A^n}\}_{n\geq 1}$ are uniformly square integrable, then it follows from [11], Theorem 5.5, that $\lim_{n\to\infty} \mathcal{W}_2(m^{A_n}, m') = 0$. □

PROOF OF LEMMA 3.8. First, let $h \in \mathbb{R}^d$ and denote $\mathbf{y} +_k h \in \mathbf{S}^N$ as the perturbation of $\mathbf{y}$ by replacing x_k with $x_k + h$. Set $m_h^{l,k} := l m^N(\mathbf{y} +_k h) + (1-l) m^N(\mathbf{y})$. Since $\varphi \in C_2^{1,2}(\mathbf{Q}_0)$ and $i_k = 1$, we have

$$
\begin{aligned}
\phi(s, \mathbf{y} +_k h) - \phi(s, \mathbf{y}) &= \varphi\big(s, m^N(\mathbf{y} +_k h)\big) - \varphi\big(s, m^N(\mathbf{y})\big) \\
&= \int_0^1 \int_{\mathbf{S}} \delta_m \varphi\big(s, m_h^{l,k}, y\big)\, d\big(m^N(\mathbf{y} +_k h) - m^N(\mathbf{y})\big)(y)\, dl \\
&= \frac{1}{N} \int_0^1 \big[\delta_m \varphi\big(s, m_h^{l,k}, (x_k + h, 1)\big) - \delta_m \varphi\big(s, m_h^{l,k}, (x_k, 1)\big)\big]\, dl \\
&= \frac{1}{N} \int_0^1 \int_0^1 h \cdot \partial_x \delta_m \varphi_1\big(s, m_h^{l,k}, x_k + \tilde{l} h\big)\, d\tilde{l}\, dl.
\end{aligned}
$$

Note that $x_k + \tilde{l} h \to x_k$ and $m_h^{l,k} \longrightarrow_{h\to 0} m^N(\mathbf{y})$, uniformly in $\tilde{l}, l \in [0,1]$, as $|h| \to 0$. Then

$$
\phi(s, \mathbf{y} +_k h) - \phi(s, \mathbf{y}) = h \cdot \frac{1}{N} \partial_x \delta_m \varphi_1\big(s, m^N(\mathbf{y}), x_k\big) + o\big(|h|\big).
$$

This implies immediately the first equality in (3.14).

Similarly, one can show that

$$
\partial_x \delta_m \big\{ \varphi_1\big(s, m^N(\mathbf{y} +_k h), x_k\big) - \varphi_1\big(s, m^N(\mathbf{y}), x_k\big) \big\} = \frac{1}{N} \partial^2_{\mu\mu} \varphi_{1,1}\big(s, m^N(\mathbf{y}), x_k, x_k\big) h + o\big(|h|\big).
$$

Then

$$
\begin{aligned}
&\partial_{x_k} \big\{ \phi(s, \mathbf{y} +_k h) - \phi(s, \mathbf{y}) \big\} \\
&\quad = \frac{1}{N} \partial_x \delta_m \big\{ \varphi_1\big(s, m^N(\mathbf{y} +_k h), x_k + h\big) - \varphi_1\big(s, m^N(\mathbf{y}), x_k\big) \big\} \\
&\quad = \frac{1}{N} \partial^2_{xx} \delta_m \varphi_1\big(s, m^N(\mathbf{y}), x_k\big) h + \frac{1}{N^2} \partial^2_{\mu\mu} \varphi_{1,1}\big(s, m^N(\mathbf{y}), x_k, x_k\big) h + o\big(|h|\big).
\end{aligned}
$$

This implies the second equality in (3.14). □

Funding. The first two authors are grateful for the financial support from the Chaires FiME-FDD and Financial Risks of the Louis Bachelier Institute.

The third author is supported in part by NSF Grants DMS-1908665 and DMS-2205972.

REFERENCES

[1] BARLES, G. and SOUGANIDIS, P. E. (1991). Convergence of approximation schemes for fully nonlinear second order equations. *Asymptot. Anal.* **4** 271–283. MR1115933

[2] BAYRAKTAR, E., CECCHIN, A. and CHAKRABORTY, P. (2023). Mean field control and finite agent approximation for regime-switching jump diffusions. *Appl. Math. Optim.* **88** Paper No. 36, 35. MR4595996 https://doi.org/10.1007/s00245-023-10015-3

[3] BAYRAKTAR, E., CECCHIN, A., COHEN, A. and DELARUE, F. (2022). Finite state mean field games with Wright–Fisher common noise as limits of N-player weighted games. *Math. Oper. Res.* **47** 2840–2890. MR4515487 https://doi.org/10.1287/moor.2021.1230

[4] BAYRAKTAR, E. and COHEN, A. (2018). Analysis of a finite state many player game using its master equation. *SIAM J. Control Optim.* **56** 3538–3568. MR3860894 https://doi.org/10.1137/17M113887X

[5] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[6] CARDALIAGUET, P. (2017). The convergence problem in mean field games with local coupling. *Appl. Math. Optim.* **76** 177–215. MR3679342 https://doi.org/10.1007/s00245-017-9434-0

[7] CARDALIAGUET, P., DAUDIN, S., JACKSON, J. and SOUGANIDIS, P. E. (2023). An algebraic convergence rate for the optimal control of McKean–Vlasov dynamics. *SIAM J. Control Optim.* **61** 3341–3369. MR4665660 https://doi.org/10.1137/22M1486789

[8] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. and LIONS, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games*. *Annals of Mathematics Studies* **201**. Princeton Univ. Press, Princeton, NJ. MR3967062 https://doi.org/10.2307/j.ctvckq7qf

[9] CARDALIAGUET, P. and SOUGANIDIS, P. E. (2023). Regularity of the value function and quantitative propagation of chaos for mean field control problems. *NoDEA Nonlinear Differential Equations Appl.* **30** Paper No. 25, 37. MR4530369 https://doi.org/10.1007/s00030-022-00823-x

[10] CARMONA, R. and DELARUE, F. (2013). Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** 2705–2734. MR3072222 https://doi.org/10.1137/120883499

[11] CARMONA, R. and DELARUE, F. (2018). *Probabilistic Theory of Mean Field Games with Applications II*: *Mean Field Games with Common Noise and Master Equations*. *Probability Theory and Stochastic Modelling* **84**. Springer, Cham. MR3753660

[12] CARMONA, R., DELARUE, F. and LACKER, D. (2017). Mean field games of timing and models for bank runs. *Appl. Math. Optim.* **76** 217–260. MR3679343 https://doi.org/10.1007/s00245-017-9435-z

[13] CAVAGNARI, G., LISINI, S., ORRIERI, C. and SAVARÉ, G. (2022). Lagrangian, Eulerian and Kantorovich formulations of multi-agent optimal control problems: Equivalence and gamma-convergence. *J. Differ. Equ.* **322** 268–364. MR4401975 https://doi.org/10.1016/j.jde.2022.03.019

[14] CECCHIN, A. (2021). Finite state N-agent and mean field control problems. *ESAIM Control Optim. Calc. Var.* **27** Paper No. 31, 33. MR4241175 https://doi.org/10.1051/cocv/2021032

[15] CECCHIN, A. and PELINO, G. (2019). Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stochastic Process. Appl.* **129** 4510–4555. MR4013871 https://doi.org/10.1016/j.spa.2018.12.002

[16] CECCHIN, A., PRA, P. D., FISCHER, M. and PELINO, G. (2019). On the convergence problem in mean field games: A two state model without uniqueness. *SIAM J. Control Optim.* **57** 2443–2466. MR3981375 https://doi.org/10.1137/18M1222454

[17] DAUDIN, S., DELARUE, F. and JACKSON, J. (2023). On the optimal rate for the convergence problem in mean field control. Available at arXiv:2305.08423.

[18] DELARUE, F., LACKER, D. and RAMANAN, K. (2019). From the master equation to mean field game limit theory: A central limit theorem. *Electron. J. Probab.* **24** Paper No. 51, 54. MR3954791 https://doi.org/10.1214/19-EJP298

[19] DELARUE, F., LACKER, D. and RAMANAN, K. (2020). From the master equation to mean field game limit theory: Large deviations and concentration of measure. *Ann. Probab.* **48** 211–263. MR4079435 https://doi.org/10.1214/19-AOP1359

[20] DJETE, M. F. (2023). Mean field games of controls: On the convergence of Nash equilibria. *Ann. Appl. Probab.* **33** 2824–2862. MR4612656 https://doi.org/10.1214/22-aap1879

[21] DJETE, M. F. (2023). Large population games with interactions through controls and common noise: Convergence results and equivalence between open-loop and closed-loop controls. *ESAIM Control Optim. Calc. Var.* **29** Paper No. 39, 42. MR4598796 https://doi.org/10.1051/cocv/2023005

[22] DJETE, M. F., POSSAMAÏ, D. and TAN, X. (2022). McKean–Vlasov optimal control: Limit theory and equivalence between different formulations. *Math. Oper. Res.* **47** 2891–2930. MR4515488 https://doi.org/10.1287/moor.2021.1232

[23] DONCEL, J., GAST, N. and GAUJAL, B. (2019). Discrete mean field games: Existence of equilibria and convergence. *J. Dyn. Games* **6** 221–239. MR4026158 https://doi.org/10.3934/jdg.2019016

[24] EKREN, I. (2017). Viscosity solutions of obstacle problems for fully nonlinear path-dependent PDEs. *Stochastic Process. Appl.* **127** 3966–3996. MR3718103 https://doi.org/10.1016/j.spa.2017.03.016

[25] EKREN, I., KELLER, C., TOUZI, N. and ZHANG, J. (2014). On viscosity solutions of path dependent PDEs. *Ann. Probab.* **42** 204–236. MR3161485 https://doi.org/10.1214/12-AOP788

[26] EKREN, I., TOUZI, N. and ZHANG, J. (2016). Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part I. *Ann. Probab.* **44** 1212–1253. MR3474470 https://doi.org/10.1214/14-AOP999

[27] EKREN, I., TOUZI, N. and ZHANG, J. (2016). Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II. *Ann. Probab.* **44** 2507–2553. MR3531674 https://doi.org/10.1214/15-AOP1027

[28] FELEQI, E. (2013). The derivation of ergodic mean field game equations for several populations of players. *Dyn. Games Appl.* **3** 523–536. MR3127148 https://doi.org/10.1007/s13235-013-0088-5

[29] FISCHER, M. (2017). On the connection between symmetric N-player games and mean field games. *Ann. Appl. Probab.* **27** 757–810. MR3655853 https://doi.org/10.1214/16-AAP1215

[30] FISCHER, M. and LIVIERI, G. (2016). Continuous time mean-variance portfolio optimization through the mean field approach. *ESAIM Probab. Stat.* **20** 30–44. MR3528616 https://doi.org/10.1051/ps/2016001

[31] FISCHER, M. and SILVA, F. J. (2021). On the asymptotic nature of first order mean field games. *Appl. Math. Optim.* **84** 2327–2357. MR4304905 https://doi.org/10.1007/s00245-020-09711-1

[32] FORNASIER, M., LISINI, S., ORRIERI, C. and SAVARÉ, G. (2019). Mean-field optimal control as gamma-limit of finite agent controls. *European J. Appl. Math.* **30** 1153–1186. MR4028474 https://doi.org/10.1017/s0956792519000044

[33] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[34] GANGBO, W., MAYORGA, S. and SWIECH, A. (2021). Finite dimensional approximations of Hamilton–Jacobi–Bellman equations in spaces of probability measures. *SIAM J. Math. Anal.* **53** 1320–1356. MR4226237 https://doi.org/10.1137/20M1331135

[35] GANGBO, W. and MÉSZÁROS, A. R. (2022). Global well-posedness of master equations for deterministic displacement convex potential mean field games. *Comm. Pure Appl. Math.* **75** 2685–2801. MR4509653

[36] GERMAIN, M., PHAM, H. and WARIN, X. (2022). Rate of convergence for particle approximation of PDEs in Wasserstein space. *J. Appl. Probab.* **59** 992–1008. MR4507678 https://doi.org/10.1017/jpr.2021.102

[37] HUANG, M., MALHAMÉ, R. P. and CAINES, P. E. (2006). Large population stochastic dynamic games: Closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6** 221–251. MR2346927

[38] ISERI, M. and ZHANG, J. (2021). Set values for mean field games. Available at arXiv:2107.01661.

[39] KOBYLANSKI, M., QUENEZ, M.-C. and ROUY-MIRONESCU, E. (2011). Optimal multiple stopping time problem. *Ann. Appl. Probab.* **21** 1365–1399. MR2857451 https://doi.org/10.1214/10-AAP727

[40] LACKER, D. (2015). Mean field games via controlled martingale problems: Existence of Markovian equilibria. *Stochastic Process. Appl.* **125** 2856–2894. MR3332857 https://doi.org/10.1016/j.spa.2015.02.006

[41] LACKER, D. (2016). A general characterization of the mean field limit for stochastic differential games. *Probab. Theory Related Fields* **165** 581–648. MR3520014 https://doi.org/10.1007/s00440-015-0641-9

[42] LACKER, D. (2017). Limit theory for controlled McKean–Vlasov dynamics. *SIAM J. Control Optim.* **55** 1641–1672. MR3654119 https://doi.org/10.1137/16M1095895

[43] LACKER, D. (2020). On the convergence of closed-loop Nash equilibria to the mean field game limit. *Ann. Appl. Probab.* **30** 1693–1761. MR4133381 https://doi.org/10.1214/19-AAP1541

[44] LACKER, D. and LE FLEM, L. (2023). Closed-loop convergence for mean field games with common noise. *Ann. Appl. Probab.* **33** 2681–2733. MR4612653 https://doi.org/10.1214/22-aap1876

[45] LASRY, J.-M. and LIONS, P.-L. (2007). Mean field games. *Jpn. J. Math.* **2** 229–260. MR2295621 https://doi.org/10.1007/s11537-007-0657-8

[46] LAURIÈRE, M. and TANGPI, L. (2022). Convergence of large population games to mean field games with interaction through the controls. *SIAM J. Math. Anal.* **54** 3535–3574. MR4438040 https://doi.org/10.1137/22M1469328

[47] MOU, C. and ZHANG, J. (2021). Wellposedness of second order master equations for mean field games with nonsmooth data. *Memoirs of the AMS*.

[48] NUTZ, M., SAN MARTIN, J. and TAN, X. (2020). Convergence to the mean field game limit: A case study. *Ann. Appl. Probab.* **30** 259–286. MR4068311 https://doi.org/10.1214/19-AAP1501

[49] OELSCHLÄGER, K. (1984). A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.* **12** 458–479. MR0735849

[50] POSSAMAÏ, D. (2021). Non-asymptotic convergence rates for mean-field games: Weak formulation and McKean–Vlasov BSDEs. Available at arXiv:2105.00484.

[51] REN, Z., TOUZI, N. and ZHANG, J. (2014). An overview of viscosity solutions of path-dependent PDEs. In *Stochastic Analysis and Applications* 2014. *Springer Proc. Math. Stat.* **100** 397–453. Springer, Cham. MR3332721 https://doi.org/10.1007/978-3-319-11292-3_15

[52] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—*1989. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[53] TALBI, M. (2022). A finite-dimensional approximation for partial differential equations on Wasserstein space. Available at arXiv:2211.00719.

[54] TALBI, M., TOUZI, N. and ZHANG, J. (2023). Dynamic programming equation for the mean field optimal stopping problem. *SIAM J. Control Optim.* **61** 2140–2164. MR4613226 https://doi.org/10.1137/21M1404259

[55] TALBI, M., TOUZI, N. and ZHANG, J. (2023). Viscosity solutions for obstacle problems on Wasserstein space. *SIAM J. Control Optim.* **61** 1712–1736. MR4604196 https://doi.org/10.1137/22M1488119

[56] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058