THEORY PROBAB. APPL.
Vol. 66, No. 4, pp. 613–639

CONTROLLED DIFFUSION MEAN FIELD GAMES WITH COMMON NOISE AND MCKEAN–VLASOV SECOND ORDER BACKWARD SDEs*

A. BARRASSO† AND N. TOUZI‡

Abstract. We consider a mean field game with common noise in which the diffusion coefficients may be controlled. We prove existence of a weak relaxed solution under some continuity conditions on the coefficients. We then show that, when there is no common noise, the solution of this mean field game is characterized by a McKean–Vlasov type second order backward SDE.

Key words. stochastic control, Nash equilibrium, mean field game, exterior noise

DOI. 10.1137/S0040585X97T990654

1. Introduction. In this paper, we consider a Mean Field Game (MFG) with common noise in which the diffusion coefficients may be controlled. Mean field games have been introduced by Lasry and Lions [25] and Huang, Malhamé, and Caines [19] and generated a very extended literature. In the present paper, we address an extension which allows for diffusion control and the presence of common noise.

The problem is defined as a Nash equilibrium within a crowd of players who solve, given a fixed random measure M, the individual maximization problem

$$\sup_{\alpha} \mathbf{E}\biggl[\xi(X^{\alpha,M}) + \int_0^T f_r(X^{\alpha,M}, \alpha_r, M)\, dr\biggr], \tag{1.1}$$

where $X^{\alpha,M}$ is the solution of the controlled non-Markovian SDE

$$dX_t^{\alpha,M} = b_t(X^{\alpha,M}, \alpha_t, M)\, dt + \sigma_t^1(X^{\alpha,M}, \alpha_t, M)\, dW_t^1 + \sigma_t^0(X^{\alpha,M}, \alpha_t, M)\, dW^0 \tag{1.2}$$

and α is the control process of a typical player. Here, $X^{\alpha,M}$ is the state process of a typical player, with dynamics controlled by α, and governed by the *individual noise* W^1 and the *common noise* W^0. The individual noise W^1 only impacts the dynamics of one specific player, while the common noise W^0 impacts the dynamics of all players.

The coefficients of the state equation depend on the random distribution M, which represents a distribution on the canonical space of the state process conditional on the common noise W^0, and is intended to model the empirical distribution of the states of the interacting crowd of players.

A solution of the MFG is then a random measure M such that the corresponding optimal diffusion $X^{*,M}$ induced by the problem (1.1) satisfies

$$M = \mathbf{P} \circ (X^{*,M} \mid W^0)^{-1} \quad \text{a.s.}, \tag{1.3}$$

where $\mathbf{P} \circ (X^{*,M} \mid W^0)^{-1}$ denotes the conditional law of $X^{*,M}$ given W^0.

*Received by the editors June 30, 2021. This work was supported by ANR project PACMAN, the joint lab FiME, the Chaires FiME-FDD, and Financial Risks of the Louis Bachelier Institute. Originally published in the Russian journal *Teoriya Veroyatnostei i ee Primeneniya*, 66 (2021), pp. 774–805.
https://doi.org/10.1137/S0040585X97T990654
†Université d'Évry Val d'Essonne, Évry, France (adrien.barrasso@univ-evry.fr).
‡Ecole Polytechnique, France (nizar.touzi@polytechnique.edu).

We prove the existence of a weak relaxed solution of this problem under some continuity conditions on the coefficients. By a *weak solution* we mean that we work with a controlled martingale problem instead of a controlled SDE intended in the strong sense, and that we find a weaker fixed point of type $M = \mathbf{P} \circ (X^{*,M} \mid W^0, M)^{-1}$ a.s. instead of (1.3), a notion introduced by Carmona, Delarue, and Lacker [7]. By *relaxed solution* we mean that we allow relaxed controls, also called mixed strategies, which is the standard framework in stochastic control theory in order to guarantee existence of optimal controls; see [16], [27]. If the control process α takes values in a subset A of a finite dimensional space, then relaxed controls q take values q_t in the space $\mathfrak{M}^1_+(A)$ of probability measures on A.

In the relaxed formulation, the state process $X^{q,M}$ is controlled by the relaxed control q, and the cost functional takes the relaxed form

$$\mathbf{E}\bigg[\xi(X^{q,M}) + \int_0^T \int_A f_r(X^{q,M}, a, M) q_r(da)\, dr\bigg].$$

The first main result of this paper is the existence of a weak relaxed solution of the MFG in the context where the state dynamics exhibit both common noise and controlled diffusion coefficients.

The second part of the paper specializes to the no common noise setting. In this context, our second main result is a characterization of the solution of this MFG by means of a McKean–Vlasov second order backward SDE of the form

$$\begin{aligned} Y_t = \xi &+ \int_t^T F_r(X, Z_r, \widehat{\sigma}^2_r, m)\, dr \\ &- \int_t^T Z_r\, dX_r + U_T - U_t, \qquad t \in [0,T], \quad \mathcal{P}^m\text{-q.s.}, \end{aligned} \tag{1.4}$$

whose precise meaning will be made explicit in section 5. Here, $\mathcal{P}^m$-q.s. means $\mathbf{P}$-a.s. for all $\mathbf{P} \in \mathcal{P}^m$. This extends the previous results by Carmona and Delarue [5], [6] characterizing the solution of a MFG by McKean–Vlasov backward SDEs in the uncontrolled diffusion setting. We believe that the present paper is the first instance of interest in such McKean–Vlasov second order backward SDEs.

Literature review. MFG have been introduced by the pioneering works of Lasry and Lions [25] and Huang, Malhamé, and Caines [19]. Their works were the first to consider the limit of a symmetric game of N players when N tends to infinity, and to link it to a fixed point problem of McKean–Vlasov type, which in its most simple form may be described as follows.

1. For any probability measure m on the space of continuous paths, find the optimal control α^m which minimizes the cost functional

$$\mathbf{E}\bigg[g(X^\alpha_T) + \int_0^T f_r(X^\alpha_r, \alpha_r, m)\, dr\bigg], \tag{1.5}$$

where X^α is the controlled diffusion of dynamics

$$dX^\alpha_t = \alpha_t\, dt + dW_t. \tag{1.6}$$

2. Find an equilibrium measure verifying $m^* = \mathcal{L}\big(X^{\alpha^{m^*}}\big)$.

The idea is that m^* models the behavior of a population of individuals. Each one of these individuals controls a diffusion of type (1.6), where W is a Brownian motion "observed" only by this specific individual and optimizes the cost (1.5).

During the following decade, this topic generated a huge literature with results based on PDE methods on the one hand (see, for instance, [25]), and on probabilistic methods on the other hand, namely, through McKean–Vlasov forward-backward SDEs; see [5] for an overview.

The extension of MFGs to the common noise situation (i.e., with an additional noise W^0 in (1.6)) was addressed recently, motivated by a strong need from applications so as to introduce a source of randomness observed by all players. One may, for example, refer to [6].

The first part of the present paper concerns the continuity of a recent sequence of papers due to Carmona, Delarue, and Lacker. In particular, [23] proves existence of a weak relaxed solution for an MFG with controlled diffusion coefficient but without common noise under merely continuity assumptions on the coefficients, and [7] shows existence of a weak solution of an MFG with common noise but without control in the diffusion coefficient, under similar continuity assumptions on the coefficients. The present paper fills the gap between these two works by extending this existence result in the situation with common noise and allowing for diffusion control.

While MFGs with a control in the drift are connected to McKean–Vlasov backward SDEs, one naturally expects that the control in the diffusion coefficient will, in some way, link the MFG to the second order extension of backward SDEs. The latter is a notion of a Sobolev type solution for path-dependent PDEs, introduced by Soner, Touzi, and Zhang [31] as a representation of diffusion control problems (in contrast with backward SDEs which are related to drift control). A first existence result was obtained in [30], and such second order backward SDEs proved very useful in the study of fully nonlinear second order PDEs, as an extension of the links between backward SDEs and semilinear PDEs; see [12], [13]. We also refer the reader to [29] for a more general existence result and to [26] for the extension to a random terminal time.

The paper is organized in two parts. Sections 2 and 3 concern mean field games with common noise and controlled diffusion coefficient; sections 4, 5, and 6 develop the links between MFGs and McKean–Vlasov second order backward SDEs.

Section 2 provides the precise formulation of our MFG; see in particular Definition 2.2. Section 3 is devoted to the proof of existence of a weak relaxed solution (see Theorem 3.1) under Assumption 3.1. The proof is divided into three parts. We start by showing some preliminary topological results in subsection 3.2. Then, in subsection 3.3, we introduce, as in [7], the notion of discretized strong equilibria (see Definition 3.2) and prove existence of such equilibria; see Proposition 3.2. Finally, in subsection 3.4, we conclude the proof of existence of a weak relaxed solution of the MFG by considering the limit of discretized strong equilibria.

In section 4, we introduce the notion of McKean–Vlasov 2BSDE (see Definition 4.2) and state the main result of the paper, which is that the solution of an MFG with controlled diffusion coefficients provides a solution of such a McKean–Vlasov 2BSDE; see Theorem 4.1. This theorem relies strongly on the representation of relaxed control problems with controlled diffusion coefficient through (classical) 2BSDEs, the proof of which we postpone to section 5. See Proposition 5.2. Section 6 contains the proof of Theorem 4.1.

2. Formulation of the mean field game.

2.1. Notation. A topological space E will always be considered as a measurable space equipped with its Borel σ-field, which will sometimes be denoted $\mathcal{B}(E)$. We denote by $\mathfrak{M}_+^1(E)$ and $\mathfrak{M}(E)$ the spaces of probability measures and of bounded signed measures on $(E, \mathcal{B}(E))$, respectively. These spaces are naturally equipped with the topology of weak convergence and the corresponding Borel σ-field.

Throughout this paper, we fix a maturity date $T > 0$, positive integers $d, p_1, p_0 \in \mathbf{N}^*$, and a compact Polish space A, and we denote by $\Omega := \mathcal{X} \times \mathcal{Q} \times \mathcal{W} \times \mathfrak{M}^1_+(\mathcal{X})$ the canonical space, where $\mathcal{X} := \mathcal{C}([0,T], \mathbf{R}^d)$ is the path space of the state process; $\mathcal{Q}$ is the set of relaxed controls, i.e., of measures q on $[0,T] \times A$ such that $q(\,\cdot \times A)$ is equal to the Lebesgue measure. Each $q \in \mathcal{Q}$ may be identified with a measurable function $t \mapsto q_t$ from $[0,T]$ to $\mathfrak{M}^1_+(A)$ determined a.e. by $q(dt, da) = q_t(da)\,dt$; $\mathcal{W} := \mathcal{W}^1 \times \mathcal{W}^0$, where $\mathcal{W}^i := \mathcal{C}([0,T], \mathbf{R}^{p_i})$, $i \in \{1, 0\}$, denote the path space of the individual noise and that of the common noise, respectively, and we denote by $\mathbb{W}^i$ the Wiener measure on $\mathcal{W}^i$.

Each of these spaces is equipped with its Borel σ field. We also denote by $\mathcal{F} := \mathcal{B}(\Omega)$ and (X, Q, W, M) the identity (or canonical) map on Ω, with $W := (W^1, W^0)$.

On $\mathcal{X}$ (respectively, $\mathcal{Q}$, $\mathcal{W}^1$, $\mathcal{W}^0$), the canonical process X (respectively, Q, W^1, W^0) generates a natural filtration $\mathbb{F}^X$ (respectively, $\mathbb{F}^Q$, $\mathbb{F}^{W^1}$, $\mathbb{F}^{W^0}$). We use similar notation on product spaces.

$\mathfrak{M}^1_+(\mathcal{X})$ is equipped with a filtration $\mathbb{F}^M$ defined by $\mathcal{F}^M_t := \sigma(M(F) \colon F \in \mathcal{F}^X_t)$. We can similarly define a filtration $\mathbb{F}^{X,Q,W,M}$ on Ω, which we denote by $\mathbb{F}$.

Let $\mathbf{P} \in \mathfrak{M}^1_+(\Omega)$, Y be a random variable (r.v.) on $(\Omega, \mathcal{F})$ with values in a measurable space $(E, \mathcal{E})$, and $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$. We denote by $\mathbf{P} \circ (Y \mid \mathcal{G})^{-1}$ the random measure which, to some $F \in \mathcal{E}$, maps $\mathbf{P}[Y \in F \mid \mathcal{G}]$.

Moreover, if $(\mathbf{P}^{\mathcal{G}}_\omega)_{\omega \in \Omega}$ is a regular conditional probability distribution of $\mathbf{P}$ given $\mathcal{G}$, we have $\mathbf{P} \circ (Y \mid \mathcal{G})^{-1} \colon (F, \omega) \mapsto \mathbf{P}^{\mathcal{G}}_\omega(Y \in F)$ $\mathbf{P}$-a.s.

2.2. Controlled state process. The controlled state process is defined as a weak solution of the following relaxed SDE, whose precise meaning will be made clear in Definition 2.1(ii):

$$X_t = X_0 + \int_0^t \int_A b_r(a, M)\, Q_r(da)\, dr + \int_0^t \int_A \sigma_r(a, M)\, N^W(da, dr). \tag{2.1}$$

Here X_0 is an initial value with the given law μ_0, $N^W := (N^{W^1}, N^{W^0})$ is a pair of orthogonal martingale measures with intensity $Q_t\, dt$ (see, e.g., [15]), $M \colon \Omega \to \mathfrak{M}^1_+(\mathcal{X})$ is a random probability measure on $\mathcal{X}$, and

$$\sigma := (\sigma^1 \| \sigma^0), \quad (b, \sigma^i) \colon [0,T] \times \mathcal{X} \times A \times \mathfrak{M}^1_+(\mathcal{X}) \to \mathbf{R}^d \times \mathbb{M}_{d,p_i}(\mathbf{R}), \qquad i = 0, 1,$$

are progressively measurable in the sense that for all $t \leqslant T$, their restriction to $[0,t] \times \mathcal{X} \times A \times \mathfrak{M}^1_+(\mathcal{X})$ is $\mathcal{B}([0,t)) \otimes \mathcal{F}^X_t \otimes \mathcal{B}(A) \otimes \mathcal{F}^M_t$-measurable.

In order to introduce the precise meaning of (2.1), we fix an initial law $\mu_0 \in \mathfrak{M}^1_+(\mathbf{R}^d)$, denote $p := p^1 + p^0$, $\overline{b} := (b \,|\, 0_p)$, $\overline{\sigma} := (\sigma \,|\, I_p)$, and introduce the generator of the controlled pair (X, W), defined for $(t, x, a, m) \in [0,T] \times \mathcal{X} \times A \times \mathfrak{M}^1_+(\mathcal{X})$ by

$$\mathcal{A}^{a,x,m}_t \phi := \overline{b}_t(x, a, m) \cdot \mathbf{D}\phi + \frac{1}{2} \overline{\sigma}\, \overline{\sigma}^\top_t(x, a, m) : \mathbf{D}^2 \phi \quad \text{for all } \phi \in \mathcal{C}^2_b(\mathbf{R}^d \times \mathbf{R}^p),$$

where ":" denotes the scalar product of matrices, and D, D^2 denote the partial gradient and Hessian with respect to the space variables.

Definition 2.1. (i) Π^0 *denotes the set of all measures* $\pi^0 \in \mathfrak{M}^1_+(\mathcal{W}^0 \times \mathfrak{M}^1_+(\mathcal{X}))$ *such that* W^0 *is a* $(\pi^0, \mathbb{F}^{W^0,M})$*-Brownian motion and such that* $\pi_0 \circ (M \circ X_0^{-1})^{-1} = \delta_{\mu_0}$.

(ii) *For* $\pi^0 \in \Pi^0$, *a* π^0-admissible control *is a probability measure* $\mathbf{P} \in \mathfrak{M}_+^1(\Omega)$ *with marginal* $\mathbf{P} \circ (W^0, M)^{-1} = \pi^0$ *satisfying that*

(1) *for all* $\phi \in \mathcal{C}_b^2(\mathbf{R}^d \times \mathbf{R}^p)$, *the following process is a* $(\mathbf{P}, \mathbb{F})$*-martingale*:

$$\phi(X_t, W_t) - \int_0^t \int_A \mathcal{A}_r^{a,X,M} \phi(X_r, W_r)\, Q_r(da)\, dr, \qquad t \in [0, T];$$

(2) M *is* $\mathbf{P}$*-independent of* W^1;

(3) *for all* $t \in [0, T]$, $\mathcal{F}_t^Q$ *is* $\mathbf{P}$*-independent of* $\mathcal{F}_T^W$ *conditionally on* $\mathcal{F}_t^W$, *i.e.*,

$$\mathbf{P}[A_t \cap A_T \mid \mathcal{F}_t^W] = \mathbf{P}[A_t \mid \mathcal{F}_t^W]\mathbf{P}[A_T \mid \mathcal{F}_t^W] \qquad \textit{for all } (A_t, A_T) \in \mathcal{F}_t^Q \times \mathcal{F}_T^W. \tag{2.2}$$

We denote by $\mathcal{P}(\pi^0)$ *the set of* π^0*-admissible controls, and we introduce the set of* admissible controls $\mathcal{P}(\Pi^0)$.

The controlled dynamics $\mathcal{P}(\pi^0)$ are defined as a further relaxation of those considered in [7] by introduction of the martingale measure N^W in order to handle the additional control of the diffusion coefficient in this paper. This relaxation follows the classical approach of [15] so as to have better compactness results, which will be crucial for our subsequent search of the fixed point defining the MFG equilibrium.

The restriction (2.2) is a slight weakening of the compatibility condition of [7]. Loosely speaking, it expresses that the control randomization underlying our approximation is external to the information of the representative agent consisting of the noise process W, in the sense that it is conditionally independent of future information given the current one. The main difference with the compatibility condition of [7] is that this restriction allows for the possible dependence on some W-independent event of the future of M. For this reason, we shall refer to (2.2) as a causality condition.

2.3. The mean field game. Let $f\colon [0, T] \times \mathcal{X} \times A \times \mathfrak{M}_+^1(\mathcal{X}) \to \mathbf{R}$ be a progressively measurable map, let $\xi\colon \mathcal{X} \to \mathbf{R}$ be a Borel map, and define the functional

$$J(\mathbf{P}) := \mathbf{E}^{\mathbf{P}}\left[\xi + \int_0^T \int_A f_r(a, M)\, Q_r(da)\, dr\right], \qquad \mathbf{P} \in \mathfrak{M}_+^1(\Omega). \tag{2.3}$$

A solution of the MFG is defined by the following two steps.

1. Given the joint law $\pi^0 \in \Pi^0$ of the pair (W^0, M), the individual optimization problem consists of the maximization of the functional J over all weak solutions $\mathbf{P} \in \mathcal{P}(\pi^0)$ of (2.1) in the sense of Definition 2.1(ii). The corresponding set of optimal solutions

$$\mathcal{P}^*(\pi^0) := \operatorname*{Argmax}_{\mathbf{P} \in \mathcal{P}(\pi^0)} J(\mathbf{P}), \quad \text{for all } \pi^0 \in \Pi^0,$$

defines a correspondence $\mathcal{P}^*$ from Π^0 to $\mathcal{P}(\Pi^0)$.

2. A strong solution of the MFG is an optimal probability $\mathbf{P}^* \in \mathcal{P}^*(\pi^0)$ such that $M = \mathbf{P}^* \circ (X \mid \mathcal{F}^{W^0})^{-1}$ a.s., i.e., under $\mathbf{P}^*$, M is the conditional law of the state process X given the common noise W^0.

For a technical reason explained below, we need to consider the following weaker notion.

Definition 2.2 (see [7]). *A weak relaxed solution of the MFG is a probability* $\mathbf{P} \in \mathfrak{M}_+^1(\Omega)$ *such that the following properties hold*:

- individual optimality: $\mathbf{P} \in \mathcal{P}^*(\pi^0)$ *for some* $\pi^0 \in \Pi^0$;
- weak equilibrium: $M = \mathbf{P} \circ (X \mid \mathcal{F}^{M,W^0})^{-1}$ $\mathbf{P}$*-a.s.*

Observe that the weak equilibrium condition in the last definition is indeed weaker than the strong equilibrium requirement $M = \mathbf{P} \circ (X \mid \mathcal{F}^{W^0})^{-1}$ a.s., which is thus named strong solution of the MFG by Carmona and Delarue [6]. The reason for introducing this weak notion of solution in [7] is recalled in Remark 3.2 in what follows.

We also remark that if $\mathbf{P}$ is a weak equilibrium, then $\mathbf{P} \circ (X_0 \mid \mathcal{F}^{M,W^0})^{-1} = M \circ X_0^{-1} = \mu_0$ a.s., where μ_0 is deterministic, hence $\mathbf{P} \circ X_0^{-1} = \mu_0$, so that μ_0 is indeed an initial law.

3. Weak relaxed Nash equilibrium.

3.1. Assumptions and main results. The following assumption will be needed to prove the existence of weak relaxed solutions of the MFG.

Assumption 3.1. (i) The coefficients b, σ, f are bounded and continuous in (x, a, m) for all t, and ξ is bounded continuous;

(ii) for every probability measure $\mathbf{Q}$ on $\mathcal{Q} \times \mathcal{W} \times \mathfrak{M}_+^1(\mathcal{X})$ under which W is a Brownian motion, there exists a unique $\mathbf{P} \in \mathfrak{M}_+^1(\Omega)$ with marginal $\mathbf{P} \circ (Q, W, M)^{-1} = \mathbf{Q}$ and satisfying condition (1) of Definition 2.1(ii).

Assumption 3.1(ii) is an existence and uniqueness condition for the SDE (2.1). It is verified, for instance, when b, σ are bounded and locally Lipschitz in x, uniformly in (t, a, m). This can be seen by considering the strong solution of the controlled SDE, which is then driven by martingale measures; see [15] for basic results concerning such SDEs.

We may now state the main result of this section.

THEOREM 3.1. *Under Assumption* 3.1, *there exists at least one weak relaxed solution of the MFG in the sense of Definition* 2.2.

The proof of this theorem will mainly rely on the Kakutani–Fan–Glicksberg fixed point theorem. The appendix of the present paper provides an introduction to set valued functions (or correspondences) which will be used extensively in this paper; we refer the reader to [1, Chapter 17].

Remark 3.1. The proof of Theorem 3.1 may be adapted (and simplified) to a context without common noise. In this context, one would impose M to be a deterministic parameter $m \in \mathfrak{M}_+^1(\mathcal{X})$. The canonical space would then be the space $\mathcal{X} \times \mathcal{Q} \times \mathcal{W}^1$, and the set of admissible controls would be the subset of $\mathbf{P} \in \mathfrak{M}_+^1(\mathcal{X} \times \mathcal{Q} \times \mathcal{W}^1)$ satisfying conditions (1) and (3) in Definition 2.2, with W replaced by W^1, and M by m.

A solution of the MFG with no common noise would be an admissible control $\mathbf{P}$ maximizing the cost functional J at fixed m and satisfying the equilibrium $\mathbf{P} \circ X^{-1} = m$.

3.2. Preliminary topological results. The aim of this subsection is to prove the following topological results.

PROPOSITION 3.1. (i) Π^0 *is a closed convex subset of* $\mathfrak{M}_+^1(\mathcal{W}^0 \times \mathfrak{M}_+^1(\mathcal{X}))$ *and, consequently, of* $\mathfrak{M}(\mathcal{W}^0 \times \mathfrak{M}_+^1(\mathcal{X}))$;

(ii) $\mathcal{P}$ *is a continuous correspondence with nonempty compact convex values*;

(iii) $\mathcal{P}^*$ *is an upper hemicontinuous correspondence with nonempty compact convex values*; *moreover*, $\mathcal{P}^*(\Pi^0)$ *is closed.*

Remark 3.2. Recall that a strong solution of the MFG is a probability measure $\mathbf{P}^* \in \mathcal{P}^*(\pi^*)$, for some $\pi^* \in \Pi^0$ such that $M = \mathbf{P}^* \circ (X \mid \mathcal{F}^{W^0})^{-1}$ $\mathbf{P}^*$ a.s., or, equivalently,

$$\pi^* \in \Phi \circ \mathcal{P}^*(\pi^*), \quad \text{where} \quad \Phi(\mathbf{P}) := \mathbb{W}^0 \circ \big(W^0, \mathbf{P} \circ \big(X \mid \mathcal{F}^{W^0}\big)^{-1}\big)^{-1}. \tag{3.1}$$

If the map Φ were continuous, then we would conclude from Proposition 3.1 that such a fixed point exists by the Kakutani fixed point theorem; see Theorem A.3. Unfortunately, the conditional expectation operator is not continuous in general. For this reason, the proof strategy used in [7] consists of introducing a discretization of the common noise W^0, so as to reduce the fixed point problem to the context of a finite σ-field, where the conditional expectation is indeed continuous. The weak solution of the MFG is then obtained as a limiting point of the solutions of the MFG problems with finite approximation of the common noise. See section 3.3 in what follows.

Proof of Proposition 3.1(i). We start by remarking that the property $\pi_0 \circ (M \circ X_0^{-1})^{-1} = \delta_{\mu_0}$ is stable under convex combinations. Then, by the Lévy characterization, W^0 is an $\mathbb{F}^{W^0,M}$-Brownian motion if and only if W^0 and $W_t^0(W_t^0)^\top - tId_{p_0}$ are martingales. Since the set of solutions of a martingale problem is convex (see Corollary 11.10 in [20]), we immediately deduce that Π^0 is convex.

We now show that Π^0 is closed. The property $\pi_0 \circ (M \circ X_0^{-1})^{-1} = \delta_{\mu_0}$ is also stable under convergence since $\{\delta_{\mu_0}\}$ is closed and $m \mapsto m \circ X_0^{-1}$ is continuous. Assume that a sequence $(\pi_n^0)_{n\in\mathbf{N}}$ of elements of Π^0 converges weakly to some π^0. By the Lévy criterion, we have, for all $s \leqslant t \in [0,T]$ and all bounded continuous $\mathcal{F}_s^{W^0,M}$-measurable ϕ_s,

$$\begin{gathered}\mathbf{E}^{\pi_n^0}[(W_t^0 - W_s^0)\phi_s] = 0,\\ \mathbf{E}^{\pi_n^0}[(W_t^0(W_t^0)^\top - W_s^0(W_s^0)^\top - (t-s)Id_m)\phi_s] = 0.\end{gathered} \tag{3.2}$$

Since $(W_t^0 - W_s^0)\phi_s$ and $(W_t^0(W_t^0)^\top - W_s^0(W_s^0)^\top - (t-s)Id_m)\phi_s$ are continuous uniformly integrable r.v.'s under $(\pi_n^0)_n$, we may send n to infinity in (3.2) and obtain that W^0 is a $(\pi^0, \mathbb{F}^{W^0,M})$-Brownian motion (see [3, Theorem 3.5]).

Proof of Proposition 3.1(iii). We now show that (iii) is a consequence of (ii), whose proof is postponed. Since f, ξ are bounded continuous, the map J introduced in (2.3) is continuous on $\mathfrak{M}_+^1(\Omega)$. As a result, since $\mathcal{P}$ is continuous with nonempty compact values, it follows directly by Theorem A.1 that $\mathcal{P}^*$ is upper hemicontinuous and takes nonempty compact values.

Further, we show that it takes convex values. Let $\pi^0 \in \Pi^0$, $\mathbf{P}^1, \mathbf{P}^2$ be elements of $\mathcal{P}^\star(\pi^0)$, i.e., maximizers of $\mathbf{E}^{\mathbf{P}}[J]$ within $\mathcal{P}(\pi^0)$, and let $\alpha \in [0,1]$. Since $\mathcal{P}$ takes convex values, we have $\alpha\mathbf{P}^1 + (1-\alpha)\mathbf{P}^2 \in \mathcal{P}(\pi^0)$, and since $\mathbf{E}^{\mathbf{P}^1}[J] = \mathbf{E}^{\mathbf{P}^2}[J] = \max_{\mathbf{P}\in\mathcal{P}(\pi^0)} \mathbf{E}^{\mathbf{P}}[J]$, it follows that $\mathbf{E}^{\alpha\mathbf{P}^1+(1-\alpha)\mathbf{P}^2}[J] = \max_{\mathbf{P}\in\mathcal{P}(\pi^0)}$. Hence $\alpha\mathbf{P}^1 + (1-\alpha)\mathbf{P}^2$ also is a maximizer of $\mathbf{E}^{\mathbf{P}}[J]$ within $\mathcal{P}(\pi^0)$, and, therefore, belongs to $\mathcal{P}^*(\pi^0)$.

It remains to prove that $\mathcal{P}^*(\Pi^0)$ is closed. Since $\mathcal{P}^*$ is upper hemicontinuous and compact valued, it has a closed graph; see Proposition A.1(1). Now let $\mathbf{P}^n \to \mathbf{P}$ with $\mathbf{P}^n \in \mathcal{P}^*(\Pi^0)$ for all n. By construction of $\mathcal{P}^*$, we have, for all n, $\mathbf{P}^n \in \mathcal{P}^*(\mathbf{P}^n \circ (M, W^0)^{-1})$, and by continuity of marginals, it follows that $\mathbf{P}^n \circ (M, W^0)^{-1}$ tends to $\mathbf{P}\circ(M, W^0)^{-1}$, which belongs to Π^0 by the closedness property established in (i) of the present proof. So by the closed graph property, $\mathbf{P} \in \mathcal{P}^*(\mathbf{P} \circ (M, W^0)^{-1}) \subset \mathcal{P}^*(\Pi^0)$, completing the proof.

The rest of this section is dedicated to the proof of Proposition 3.1(ii). We start with an immediate consequence of Proposition 3.1(i).

Corollary 3.1. *Let* $\Pi := \{\pi := \mathbb{W}^1 \otimes \pi^0 \colon \pi^0 \in \Pi^0\}$. *Then*

(i) Π *is a closed convex subset of* $\mathfrak{M}^1_+(\mathcal{W} \times \mathfrak{M}^1_+(\mathcal{X}))$;

(ii) *the map* $\mathbf{T} \colon \pi^0 \in \Pi^0 \mapsto \pi := \mathbb{W}^1 \otimes \pi^0 \in \Pi$ *is a homeomorphism*;

(iii) *if* $\mathcal{K}^0$ *is a compact* (*respectively, convex*) *subset of* Π^0, *then* $\mathcal{K} := \mathbf{T}(\mathcal{K}^0)$ *is a compact* (*respectively, convex*) *subset of* Π.

We next consider a further extension of the probability measures $\pi \in \Pi$:

$$\mathfrak{Q}_{\mathrm{c}}(\pi) := \big\{\mathbf{Q} \in \mathfrak{M}^1_+\big(\mathcal{Q} \times \mathcal{W} \times \mathfrak{M}^1_+(\mathcal{X})\big) \colon \mathbf{Q} \circ (W, M)^{-1} = \pi \text{ and } \mathbf{Q} \text{ satisfies } (2.2)\big\},$$

where the subscript "c" stands for the causality condition (2.2).

Lemma 3.1. (i) *The set* $\mathfrak{Q}_c(\Pi)$ *is closed convex.*

(ii) *Let* $\mathcal{K}^0$ *be a compact* (*respectively, convex*) *subset of* Π^0, *and set* $\mathcal{K} := \mathbf{T}(\mathcal{K}^0)$; *then* $\mathfrak{Q}_c(\mathcal{K})$ *is a compact* (*respectively, convex*) *subset of* $\mathfrak{Q}_c(\Pi)$.

(iii) *The correspondence* $\mathfrak{Q}_c \colon \pi \in \Pi \mapsto \mathfrak{Q}_c(\pi)$ *is continuous.*

Proof. Throughout this proof, we denote $\mathfrak{Q} := \big\{\mathbf{Q} \in \mathfrak{M}^1_+\big(\mathcal{Q} \times \mathcal{W} \times \mathfrak{M}^1_+(\mathcal{X})\big) \colon \mathbf{Q} \circ (W, M)^{-1} \in \Pi \text{ and } \mathbf{Q} \text{ satisfies } (2.2)\big\} = \mathfrak{Q}_c(\Pi)$.

(i) Since Π is itself convex and closed by Corollary 3.1, the first item above is stable by convergence or convex combinations. By Theorem 3.11 in [24], since W has independent increments (with respect to its own filtration), the second item above holds if and only if for all $t \leqslant s$, $W_t - W_s$ is $\mathbf{Q}$-independent of $\mathcal{F}^{Q,W}_s$. This condition is also stable under convergence or convex combinations, so $\mathfrak{Q}$ is closed and convex.

(ii) The closeness and convexity of $\mathfrak{Q}_c(\mathcal{K})$ follow from the same arguments as above. Its tightness (hence, relative compactness) follows from the compactness of $\mathcal{Q}$ and the tightness of $\{\mathbf{Q} \circ (W, M)^{-1} \colon \mathbf{Q} \in \mathfrak{Q}_c(\mathcal{K})\} = \mathcal{K}$.

(iii) We decompose $\mathfrak{Q}_c$ as the composition of two continuous correspondences Γ_1, Γ_2 which we now introduce. Denote $\mathcal{K}' := \{\mathbf{Q} \circ (Q, W)^{-1} \colon \mathbf{Q} \in \mathfrak{Q}\}$, i.e., the set of laws in $\mathfrak{M}^1_+(\mathcal{Q} \times \mathcal{W})$ for which W is an $\mathbb{F}^{Q,W}$-Brownian motion. With arguments similar to what we have seen for Π^0 or Π, it is easy to see that $\mathcal{K}'$ is closed convex and is even compact thanks to the compactness of $\mathcal{Q}$.

We define the correspondence Γ_1, which to any $\pi \in \Pi$, maps the subset $\{\pi\} \times \mathcal{K}'$ of $\Pi \times \mathcal{K}'$. We also define Γ_2, which to any (π, π') in $\Pi \times \mathcal{K}'$, maps the set

$$\{\mathbf{Q} \in \mathfrak{Q} \colon \mathbf{Q} \circ (Q, W)^{-1} = \pi', \ \mathbf{Q} \circ (W, M)^{-1} = \pi\}.$$

It is clear that $\mathfrak{Q}_c = \Gamma_2 \circ \Gamma_1$.

The set Γ_1 is the product of the continuous function $\pi \to \pi$ and of the correspondence $\pi \to \mathcal{K}'$, which is compact valued and constant and hence continuous; apply Proposition 3.1(iii), for instance. So Γ_1 is continuous as the product of continuous compact valued correspondences; see Theorem 17.28 in [1].

The set Γ_2 is the restriction on $\Pi \times \mathcal{K}'$ of the inverse ψ^{-1} of the mapping $\psi \colon \mathbf{Q} \to (\mathbf{Q} \circ (Q, W)^{-1}, \mathbf{Q} \circ (W, M)^{-1})$. Adapting Theorem 3 in [11], for example, we see that ψ is an open mapping. Then, by Theorem 17.7 in [1], ψ^{-1} (or its restriction Γ_2) is lower hemicontinuous. It is immediate that Γ_2 has a closed graph; however, its range is not compact, so we cannot conclude immediately that it is upper hemicontinuous.

Let us fix some compact subset $\mathcal{K}^0$ of Π^0 and $\Gamma_2^{\mathcal{K}^0}$ the restriction of Γ_2 on $\mathcal{K}^0$. Then $\Gamma_2^{\mathcal{K}^0}$ is still low hemicontinuous with closed graph but this time has compact range and hence is upper hemicontinuous by the closed graph theorem; see Proposition 3.1(ii).

Therefore, it is continuous. Γ_2 is compact valued and hence can also be seen as a function with values in the metric space of compact subsets of $\mathfrak{M}_+^1(\mathcal{Q}\times\mathcal{W}\times\mathfrak{M}_+^1(\mathcal{X}))$, equipped with the Hausdorff metric. By Proposition 3.1(iii), Γ_2 is continuous on a certain set as a correspondence if and only if it is continuous as a function for the Hausdorff metric. What we have seen is that Γ_2 is in fact continuous on every compact subset of $\Pi \times \mathcal{K}'$, and in a metric space, a function which is continuous on every compact set is continuous everywhere. So Γ_2 is continuous everywhere, and hence $\mathfrak{Q}_c$ is continuous as the composition of continuous correspondences; see Proposition A.1(4). The proof is complete.

Finally, we lift the set $\mathfrak{Q}_c(\Pi)$ by the map

$$\mathbf{Q} \in \mathfrak{Q}_c(\Pi) \to \Psi(\mathbf{Q}) := \mathbf{P} \in \mathcal{P}(\Pi^0) \quad \text{if and only if} \quad \mathbf{P} \circ (Q, W, M)^{-1} = \mathbf{Q},$$

where the existence and uniqueness of $\mathbf{P}$ is guaranteed by Assumption 3.1.

LEMMA 3.2. (i) $\mathcal{P}(\Pi^0)$ *is a closed convex subset of* $\mathfrak{M}_+^1(\Omega)$, *and* $\mathcal{P}(\mathcal{K}^0)$ *is compact* (*respectively*, *convex*) *for all compact* (*respectively*, *convex*) *subset* $\mathcal{K}^0$ *of* Π^0.

(ii) Ψ *is a homeomorphism from* $\mathfrak{Q}_c(\Pi)$ *to* $\mathcal{P}(\Pi^0)$.

Proof. (i) By definition, $\mathbf{P} \in \mathfrak{M}_+^1(\Omega)$ belongs to $\mathcal{P}(\Pi^0)$ if and only if

(a) $\mathbf{P} \circ (Q, W, M)^{-1}$ belongs to $\mathfrak{Q}_c(\Pi)$;

(b) for all $\phi \in \mathcal{C}_b^2(\mathbf{R}^d \times \mathbf{R}^p)$,

$$\phi(X_t, W_t) - \int_0^t \int_A \mathcal{A}_r^{a,X,M} \phi(X_r, W_r)\, Q_r(da)\, dr, \qquad t \in [0, T],$$

is a $(\mathbf{P}, \mathbb{F})$-martingale.

Since $\mathfrak{Q}_c(\Pi)$ is convex and closed by Lemma 3.1, it is clear that the set of $\mathbf{P}$ verifying condition (a) above is convex and closed. Then, since the set of solutions of a martingale problem is convex (see Corollary 11.10 in [20]), and since the coefficients b, σ are bounded and continuous in (x, a, m) for fixed t, the set of probability measures verifying the above condition (b) is also closed convex. This shows that $\mathcal{P}(\Pi^0)$ is closed convex.

Further, we prove the second part of (i). We fix some compact convex subset $\mathcal{K}^0$ of Π^0. It is immediate by construction that $\mathcal{P}(\mathcal{K}^0)$ remains closed convex, so we are left to prove that it is relatively compact. By boundedness of b, σ, the set $\{\mathbf{P} \circ X^{-1} \colon \mathbf{P} \in \mathcal{P}(\mathcal{K}^0)\}$ is tight (see Theorem 1.4.6 in [32], for instance), and by the compactness of $\mathcal{Q}$ and the tightness of $\mathbb{W}^1 \otimes \mathcal{K}^0$ we see that $\{\mathbf{P} \circ (Q, W, M)^{-1}\colon \mathbf{P} \in \mathcal{P}(\mathcal{K}^0)\}$ is tight. So $\mathcal{P}(\mathcal{K}^0)$ is tight and, therefore, relatively compact, which concludes the proof.

(ii) It is clear that Ψ is a bijection, and that its reciprocal Ψ^{-1} (defined by $\Psi^{-1}(\mathbf{P}) = \mathbf{P} \circ (Q, W, M)^{-1}$) is continuous.

Let $\mathbf{P}_n \to \mathbf{P}$ in $\mathfrak{Q}_c(\Pi)$; then we also have $\mathbf{P}_n \circ (M, W^0)^{-1} \to \mathbf{P} \circ (M, W^0)^{-1}$, so the measures $(\mathbf{P}_n \circ (M, W^0)^{-1})_n$ and $\mathbf{P} \circ (M, W^0)^{-1}$ belong to some compact subset $\mathcal{K}^0$ of Π^0 and the measures $(\mathbf{P}_n)_n$ and $\mathbf{P}$ belong to $\mathfrak{Q}_c(\mathcal{K})$, where $\mathcal{K} := \mathbb{W}^1 \otimes \mathcal{K}^0$. So it is enough to show that Ψ is continuous on $\mathfrak{Q}_c(\mathcal{K})$ for any compact subset $\mathcal{K}^0$ of Π^0.

We fix $\mathcal{K}^0$ and $\mathcal{K} := \mathbb{W}^1 \otimes \mathcal{K}^0$. By construction, the restriction of Ψ induces a bijection $\Psi_{\mathcal{K}^0}$ from $\mathfrak{Q}_c(\mathcal{K})$ onto $\mathcal{P}(\mathcal{K}^0)$ which are both compact; see Lemma 3.1 and part (i) of the present lemma. $\Psi_{\mathcal{K}^0}^{-1}$ is the marginal mapping $\mathbf{P} \to \mathbf{P} \circ (Q, W, M)^{-1}$ restricted on $\mathcal{P}(\mathcal{K}^0)$ and hence is continuous. So $\Psi_{\mathcal{K}^0}^{-1}$ is a continuous bijection between compact sets and hence a homeomorphism. $\Psi_{\mathcal{K}^0}^{-1}$ is, therefore, continuous, meaning that Ψ is continuous on $\mathfrak{Q}_c(\mathcal{K})$, and the proof is complete.

We can now conclude the proof of Proposition 3.1.

Proof of Proposition 3.1(ii). The set $\mathcal{P}$ may now be written as the composition $\Psi\circ\mathfrak{Q}_c\circ\mathbf{T}$, where $\mathbf{T}$, $\mathfrak{Q}_c$, and Ψ are introduced, respectively, in Corollary 3.1 and Lemmas 3.1 and 3.2. So thanks to these three results, $\mathcal{P}$ is a continuous correspondence as the composition of two continuous functions and a continuous correspondence; see Proposition A.1(4).

For every $\pi^0\in\Pi^0$, $\mathcal{P}(\pi^0)$ is compact convex by Lemma 3.2(i). Finally, $\mathcal{P}$ takes nonempty values thanks to Assumption 3.1(1).

3.3. Discretized strong equilibria. This section follows the proof strategy of [7] as mentioned earlier in Remark 3.2. The main novelty in what follows is our reformulation of the problem given in (3.1). Under this perspective, all our analysis is made on the space $\mathfrak{M}^1_+(\Omega)$. We believe that this point of view simplifies some technical issues and is the key ingredient for allowing the control in the diffusion coefficient.

DEFINITION 3.1. *For each $n\geqslant 1$, let $t^n_i := i2^{-n}T$ for $i=0,\dots,2^n$. For every n, we fix a partition $c_n := \{C^n_1,\dots,C^n_n\}$ of $\mathbf{R}^{p_0}$ into n Borel sets of strictly positive Lebesgue measure, such that for all n, c_{n+1} is a refinement of c_n, and $\mathcal{B}(\mathbf{R}^{p_0}) = \sigma\big(\bigcup_n c_n\big)$. For a given n, and $I=(i_1,\dots,i_{2^n})\in\{1,\dots,n\}^{2^n}$, $k\leqslant 2^n$, we define $S^{n,k}_I$ as the set of paths with increments up to time k in $C^n_{i_1},\dots,C^n_{i_k}$, i.e.,*

$$S^{n,k}_I := \{\omega^0\in\mathcal{W}^0\colon \omega^0_{t^n_j}-\omega^0_{t^n_{j-1}}\in C^n_{i_j} \text{ for all } j=1,\dots,k\}.$$

We also denote $S^n_I := S^{n,2^n}_I$. The S^n_I's, $I\in\{1,\dots,n\}^{2^n}$, form a finite partition of $\mathcal{W}^0$, where each S^n_I has a strictly positive $\mathbb{W}^0$-measure.

For all n we denote $\mathcal{F}^{n,W^0} := \sigma(S^n_I\colon I\in\{1,\dots,n\}^{2^n})$, and for all $t\in[0,T]$ we denote $\mathcal{F}^{n,W^0}_t := \sigma(S^{n,j}_I\colon I\in\{1,\dots,n\}^{2^n},\ j\leqslant k^n_t)$, where k^n_t is the largest integer such that $t^n_{k^n_t}\leqslant t$.

Finally, for all n, we introduce the mapping $\widehat{X}^n\colon\mathcal{X}\to\mathcal{X}$ such that for all $k<2^n$ and $t\in[t^n_k,t^n_{k+1}[$, $\widehat{X}^n_t = (2^n/T)(t-t^n_k)X_{t^n_k} + (2^n/T)(t^n_{k+1}-t)X_{t^n_{k-1}}$.

The following facts may be found in subsection 2.4.2 and the proof of Lemma 3.6 (second step) in [7].

Remark 3.3. (i) $\mathcal{F}^{W^0}_t = \sigma\big(\bigcup_n\mathcal{F}^{n,W^0}_t\big)$ for all $t\in[0,T]$;

(ii) $(\mathcal{F}^{n,W^0}_t)_{t\geqslant 0}$ is a subfiltration of $\mathbb{F}^{W^0}$;

(iii) for all n, $\widehat{X}^n$ is continuous, and $\widehat{X}^n\to X$ as $n\to\infty$ uniformly on the compact sets of $\mathcal{X}$.

DEFINITION 3.2. *A discretized strong Nash equilibrium of order n is a probability measure $\mathbf{P}\in\mathcal{P}^*(\Pi^0)$ such that*

$$M = \mathbf{P}\circ(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\quad \mathbf{P}\text{-}a.s. \tag{3.3}$$

PROPOSITION 3.2. *For every n, there exists a discretized strong Nash equilibrium of order n.*

We will prove this first existence result by applying the Kakutani fixed point theorem, thanks to the regularity of the correspondence $\mathcal{P}^*$. However, such a fixed point theorem holds in a compact convex set, and our set Π^0 is not compact, so we now construct a smaller (and compact) set, in which this theorem can be applied.

DEFINITION 3.3. *If* $\mathbf{P} \in \mathfrak{M}_+^1(\mathcal{X})$ *is such that* X *is a* $\mathbf{P}$*-semimartingale, we denote by* $A^{\mathbf{P}}$ *and* $M^{\mathbf{P}}$ *the bounded variation and the martingale components of* X *under* $\mathbf{P}$.

$\mathcal{K}^X$ *denotes the closure of the space of elements of* $\mathfrak{M}_+^1(\mathcal{X})$ *under which* X *is a semimartingale for which* $|A^{i,\mathbf{P}}|$, $i \leqslant d$, *and* $\mathrm{Tr}(\langle M^{\mathbf{P}}\rangle)$ *are absolutely continuous with derivatives bounded by* C $dt \otimes d\mathbf{P}$ *a.e., where* C *is a fixed constant bounding* b *and* $\overline{\sigma}\,\overline{\sigma}^\top$ *for the sup norm.*

LEMMA 3.3. *The set* $\mathcal{K}^X$ *is a compact subset of* $\mathfrak{M}_+^1(\mathcal{X})$.

Proof. It is well known that any family of laws of continuous diffusions with bounded coefficients is tight (see, for instance, [32, Theorem 1.4.6]), so $\mathcal{K}^X$ is the closure of a tight set and hence of a relatively compact set by Prokhorov's theorem. The proof is complete.

For all $n \in \mathbf{N} \setminus \{0\}$, we also set $\mathcal{K}_n^X := \{\mathbf{P} \circ (\widehat{X}^n)^{-1} \colon \mathbf{P} \in \mathcal{K}^X\}$. By the tightness of $\mathcal{K}^X$ we may introduce an increasing sequence of compact subsets $(K_k^\infty)_{k\in\mathbf{N}^*}$ of $\mathcal{X}$ such that

$$\mathbf{P}[X \in K_k^\infty] \geqslant 1 - \frac{1}{k} \quad \text{for all} \quad k > 0 \quad \text{and} \quad \mathbf{P} \in \mathcal{K}^X.$$

Finally, we denote

$$K_k^n := \widehat{X}^n(K_k^\infty) \quad \text{and} \quad \overline{K}_k := \bigcup_{n\in\mathbf{N}\cup\{\infty\}} K_k^n \quad \text{for all} \quad k, n \in \mathbf{N}.$$

LEMMA 3.4. *For all* k, n, K_k^n, *and* $\overline{K}_k$ *are compact, and* $\mathcal{K}_n^X$ *is tight.*

Proof. Compactness of K_k^n follows from the continuity of $\widehat{X}^n$ which, therefore, maps compact sets onto compact sets.

We next prove that $\mathcal{K}_n^X$ is tight. Let $\mathbf{Q} = \mathbf{P} \circ (\widehat{X}^n)^{-1} \in \mathcal{K}_n^X$ for some $\mathbf{P} \in \mathcal{K}^X$. Then, for all k, we have $\mathbf{Q}[K_k^n] = \mathbf{P}[\widehat{X}^n \in K_k^n] \geqslant \mathbf{P}[X \in K_k^\infty] \geqslant 1 - 1/k$. Since this holds for any $\mathbf{Q} \in \mathcal{K}_n^X$, the announced tightness is shown.

It remains to prove that $\overline{K}_k$ is compact. For a fixed sequence $(x_n)_{n\geqslant 0}$ in $\overline{K}_k$, either there exists some $(i_1, \ldots, i_N) \in \overline{\mathbf{N}}^N$ such that $(x_n)_{n\geqslant 0}$ remains in the compact set $\bigcup_{j\leqslant N} K_k^{i_j}$, in which case this sequence admits a converging subsequence, or we can assume (up to an extraction which we omit) that there exists a strictly increasing sequence $(p_n)_n$ such that for all n, $x_n \in K_k^{p_n}$.

Now, for all n, we may consider some $y_n \in K_k^\infty$ such that $x_n = \widehat{X}^n(y_n)$, and since K_k^∞ is compact, we may assume (again up to the extraction of a subsequence) that y_n converges to some y in K_k^∞. We now conclude the proof by showing that x_n also tends to y, and hence any such sequence of $\overline{K}_k$ admits a converging subsequence in $\overline{K}_k$. Indeed, we have

$$|x_n - y| = |\widehat{X}^{p_n}(y_n) - y| \leqslant |\widehat{X}^{p_n}(y_n) - y_n| + |y_n - y|.$$

The second term on the right tends to zero, and since p_n is strictly increasing, $\widehat{X}^{p_n}$ tends uniformly to X on compact sets and, in particular, on K_k^∞ (see Remark 3.3(iii)), so $|\widehat{X}^{p_n}(y_n) - y_n|$ tends to zero, and the proof is complete.

We now introduce the set in which we will find the discretized equilibrium:

$$\Pi_c^0 := \left\{\pi^0 \in \Pi^0 \colon \pi^0(\overline{K}_k) \geqslant 1 - \frac{1}{k} \text{ for all } k > 0\right\}. \tag{3.4}$$

LEMMA 3.5. *For all* n, Π_c^0 *is a compact convex set.*

Proof. We fix n. It is immediate by construction that Π_c^0 is tight and hence relatively compact. Moreover, Π^0 is convex (see Proposition 3.1(i) and (3.4)) and is stable by convex combination, so Π_c^0 is also convex.

Let us now show that Π_c^0 is closed. Since Π^0 is closed (see Proposition 3.1(i)), it is enough to show that (3.4) is stable under convergence. We fix a converging sequence $\pi^j \to \pi$, where $\pi^j \in \Pi_c^0$ for all j.

By the Skorokhod representation theorem (see, for instance, [3, Theorem 6.7]), there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbf{P}})$ on which there exist random measures M^j of law $\pi^j \circ M^{-1}$ and $M^{\lim}$ of law $\pi \circ M^{-1}$, and a $\widetilde{\mathbf{P}}$-null set $\mathcal{N}$ such that for all ω in $\mathcal{N}^c$, $M^j(\omega) \to M^{\lim}(\omega)$ weakly. Since the sets $\overline{K}_k$ are closed, a consequence of the Portmanteau theorem (see [3, Theorem 2.1], for instance), is that for all k and $\omega \in \mathcal{N}^c$,

$$M^{\lim}(\omega)(\overline{K}_k) \geqslant \limsup_j M^j(\omega)(\overline{K}_k). \tag{3.5}$$

Now by taking the expectation in (3.5) and applying the reversed Fatou's lemma, we get that, for all k,

$$\mathbf{E}^{\widetilde{\mathbf{P}}}[M^{\lim}(\overline{K}_k)] \geqslant \mathbf{E}^{\widetilde{\mathbf{P}}}[\limsup_j M^j(\overline{K}_k)] \geqslant \limsup_j \mathbf{E}^{\widetilde{\mathbf{P}}}[M^j(\overline{K}_k)], \tag{3.6}$$

and hence $\mathbf{E}^{\pi}[M(\overline{K}_k)] \geqslant \limsup_j \mathbf{E}^{\pi_j}[M(\overline{K}_k)] \geqslant 1 - 1/k$. So (3.4) holds under π, and the proof is complete.

We may now prove the main result of this subsection.

Proof of Proposition 3.2. We first note that $M = \mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}$, $\mathbf{P}$-a.s. is equivalent to having

$$\mathbf{P}\circ(W^0, M)^{-1} = \mathbf{P}\circ\big(W^0, \mathbf{P}\circ(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1} = \mathbb{W}^0\circ\big(W^0, \mathbf{P}\circ(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1}.$$

We introduce on $\mathcal{P}(\Pi^0)$ the mapping

$$\Phi_n \colon \mathbf{P} \to \mathbb{W}^0 \circ \big(W^0, \mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1}$$

and show that it is continuous on this set.

We fix a converging sequence $\mathbf{P}^k \to \mathbf{P}$ in $\mathcal{P}(\Pi^0)$. By Theorem 4.11 in [21], in order to show that

$$\mathbb{W}^0 \circ \big(W^0, \mathbf{P}^k(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1} \to \mathbb{W}^0 \circ \big(W^0, \mathbf{P}^k(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1},$$

it is enough to show that, for all bounded continuous ϕ,

$$\mathbb{W}^0 \circ \big(W^0, \mathbf{E}^k[\phi(\widehat{X}^n) \mid \mathcal{F}^{n,W^0}]\big)^{-1} \to \mathbb{W}^0 \circ \big(W^0, \mathbf{E}[\phi(\widehat{X}^n) \mid \mathcal{F}^{n,W^0}]\big)^{-1}.$$

Since $\mathbf{E}^k[\phi(\widehat{X}^n) \mid \mathcal{F}^{n,W^0}] = \sum_I \big(\mathbf{E}^k[\phi(\widehat{X}^n)\mathbf{1}_{S_I^n}(W^0)]/\mathbb{W}^0[S_I^n]\big)\mathbf{1}_{S_I^n}(W^0)$, for all k, we are reduced to proving, for all $\phi \in \mathcal{C}_b(\mathcal{X})$, $\psi \in \mathcal{C}_b(\mathbf{R})$, and $\zeta \in \mathcal{C}_b(\mathcal{W}^0)$, that

$$\begin{aligned} &\mathbf{E}^{\mathbb{W}^0}\left[\psi\left(\sum_I \frac{\mathbf{E}^k[\phi(\widehat{X}^n)\mathbf{1}_{S_I^n}(W^0)]}{\mathbb{W}^0[S_I^n]}\mathbf{1}_{S_I^n}(W^0)\right)\zeta(W^0)\right] \\ &\quad \underset{k}{\to} \mathbf{E}^{\mathbb{W}^0}\left[\psi\left(\sum_I \frac{\mathbf{E}[\phi(\widehat{X}^n)\mathbf{1}_{S_I^n}(W^0)]}{\mathbb{W}^0[S_I^n]}\mathbf{1}_{S_I^n}(W^0)\right)\zeta(W^0)\right]. \end{aligned} \tag{3.7}$$

Since $\mathbf{P}$ and the $\mathbf{P}^k$ all have the same first marginal $\mathbb{W}^0$, the convergence of $\mathbf{P}^k$ to $\mathbf{P}$ is a stable convergence in the sense that, for all bounded continuous f and bounded Borel g, we have that $\mathbf{E}^k[f(X)g(W^0)]$ tends to $\mathbf{E}[f(X)g(W^0)]$; see Lemma 2.1 in [24], for instance. In particular, by continuity of ϕ and $\widehat{X}^n$, we have for all I,

$$\mathbf{E}^k[\phi(\widehat{X}^n)\mathbf{1}_{S_I^n}(W^0)]\mathbf{1}_{S_I^n}(W^0) \to \mathbf{E}[\phi(\widehat{X}^n)\mathbf{1}_{S_I^n}(W^0)]\mathbf{1}_{S_I^n}(W^0) \quad \mathbb{W}^0\text{-a.s.},$$

and by the dominated convergence theorem, (3.7) holds for any ϕ, ψ, ζ, implying the desired continuity of the mapping Φ_n.

We now show that Φ_n takes values in Π_c^0, as introduced in (3.4). Let $\mathbf{P} \in \mathcal{P}(\Pi^0)$ and $\mathbf{Q} := \Phi_n(\mathbf{P}) = \mathbb{W}^0 \circ \big(W^0, \mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1}$. It is immediate that W^0 is an $\mathbb{F}^{W^0}$-Brownian motion under $\mathbf{Q}$; however, in order to fit the definition of Π_c^0 which is included in Π^0, we need to show that W^0 is an $\mathbb{F}^{M,W^0}$-Brownian motion. Since M is $\mathbf{Q}$-a.s. equal to the $\mathcal{F}^{W^0}$-measurable random measure $\mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}$, in order to show that W^0 is indeed an $\mathbb{F}^{M,W^0}$-Brownian motion, it is enough to show that $\mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}$ is $\mathbb{F}^{W^0}$-adapted in the sense that, for any $F \in \mathcal{F}_t^X$, $\mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}(F)$ is $\mathcal{F}_t^{W^0}$-measurable.

We fix some $k < 2^n$, $t \in [t_k, t_{k+1}[$, and $F \in \mathcal{F}_t^X$. By construction of $\widehat{X}^n$, we have

$$\{\widehat{X}^n \in F\} \in \mathcal{F}_{t_k}^X. \tag{3.8}$$

Now by the definition of $\mathcal{P}(\Pi^0)$ (see Definition 2.1) W^0 is under $\mathbf{P}$ and $\mathbb{F}$-Brownian motion, so for all t, $\mathcal{F}_t^X$ is conditionally independent of $\mathcal{F}_T^{W^0}$ given $\mathcal{F}_t^{W^0}$, and, in particular, combining (3.8) and Theorem 3.11 in [24] we have

$$\mathbf{P} \circ \big(\widehat{X}^n \in F \,\big|\, \mathcal{F}_T^{W^0}\big)^{-1} = \mathbf{P} \circ \big(\widehat{X}^n \in F \,\big|\, \mathcal{F}_{t_k}^{W^0}\big)^{-1} \text{ a.s.} \tag{3.9}$$

We have

$$\begin{aligned}
\mathbf{P} \circ \big(\widehat{X}^n \,\big|\, \mathcal{F}^{n,W^0}\big)^{-1}[F] &:= \mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_T^{n,W^0}\big] \\
&= \mathbf{E}\Big[\mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_T^{W^0}\big] \,\Big|\, \mathcal{F}_T^{n,W^0}\Big] = \mathbf{E}\Big[\mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_{t_k}^{W^0}\big] \,\Big|\, \mathcal{F}_T^{n,W^0}\Big] \\
&= \mathbf{E}\Big[\mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_{t_k}^{W^0}\big] \,\Big|\, \mathcal{F}_{t_k}^{n,W^0}\Big] = \mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_{t_k}^{n,W^0}\big] \\
&= \mathbf{P}\big[\widehat{X}^n \in F \,\big|\, \mathcal{F}_t^{n,W^0}\big],
\end{aligned} \tag{3.10}$$

where the third equality holds by (3.9) and the fourth by the independence of the increments of W^0 and the construction of $\mathbb{F}^{n,W^0}$. So we indeed see that $\mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}(F)$ is $\mathcal{F}_t^{W^0}$-measurable, and, therefore, W^0 is under $\mathbf{Q}$ an $\mathbb{F}^{M,W^0}$-Brownian motion so that $\mathbf{Q} \in \Pi^0$.

We conclude by showing that $\mathbf{Q}$ verifies (3.4). For a fixed integer k, we have

$$\begin{aligned}
\mathbf{E}^{\mathbf{Q}}\big[M[\overline{K}_k]\big] &= \mathbf{E}^{\mathbf{Q}}\big[\mathbf{P} \circ (\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}[\overline{K}_k]\big] = \mathbf{E}^{\mathbf{P}}\Big[\mathbf{P}\big[\widehat{X}^n \in \overline{K}_k \,\big|\, \mathcal{F}^{n,W^0}\big]\Big] \\
&= \mathbf{P}[\widehat{X}^n \in \overline{K}_k] \geqslant \mathbf{P}[\widehat{X}^n \in K_k^n] \geqslant \mathbf{P}[X \in K_k^\infty] \geqslant 1 - \frac{1}{k},
\end{aligned} \tag{3.11}$$

where the last inequality holds since $\mathbf{P} \in \mathcal{P}(\Pi^0)$, hence $\mathbf{P} \circ X^{-1} \in \mathcal{K}^X$, and by construction of the sets $\overline{K}_k$, K_k^n, and K_k.

We may now conclude with a version of Kakutani's theorem. We consider the restriction of $\mathcal{P}^*$ on Π_c^0, namely, $\mathcal{P}^*\colon \Pi_c^0 \twoheadrightarrow \mathcal{P}(\Pi_c^0)$, which defines an upper hemicontinuous correspondence taking nonempty compact convex values (see Proposition 3.1(iii)).

We recall that $\Phi_n\colon \mathcal{P}(\Pi_c^0) \to \Pi_c^0$ is a continuous mapping, and that Π_c^0 is a convex compact subset of a locally convex topological space (see Lemma 3.5), so by Theorem A.3 and Lemma A.1, there exists, in Π_c^0, a fixed point $\pi_n^* \in \Phi_n \circ \mathcal{P}^*(\pi_n^*)$.

We conclude this proof by showing that if we make set $\mathbf{P}_n^*$ the element of $\mathcal{P}^*(\pi_n^*)$ such that $\pi_n^* = \Phi(\mathbf{P}_n^*)$, then $\mathbf{P}_n^*$ is a discretized strong Nash equilibrium of order n; see Definition 3.2.

The probability measure $\mathbf{P}_n^*$ belongs to $\mathcal{P}(\Pi^0)$ and $\mathcal{P}^*(\Pi^0)$. Moreover, it verifies $\mathbf{P}_n^*\circ(W^0, M)^{-1} = \pi_n^* = \mathbb{W}^0\circ\big(W^0, \mathbf{P}_n^*(\widehat{X}^n \mid \mathcal{F}^{n,W^0})^{-1}\big)^{-1}$, hence $M = \mathbf{P}_n^*(X \mid \mathcal{F}^{n,W^0})^{-1}$ $\mathbf{P}_n^*$-a.s., meaning that (3.3) holds, and $\mathbf{P}_n^*$ is a discretized strong Nash equilibrium of order n.

3.4. Existence of a weak Nash equilibrium. We conclude this section by proving Theorem 3.1, i.e., we will verify the existence of a weak Nash equilibrium.

Proof of Theorem 3.1. For every $n \in \mathbf{N}$, we consider $\mathbf{P}_n^*$ a discretized strong Nash equilibrium of order n whose existence is ensured by Proposition 3.2. Every $\mathbf{P}_n^*$ belongs to $\mathcal{P}(\Pi_c^0)$, which is compact since Π_c^0 is (see Lemmas 3.2(i) and 3.5). So we may consider an accumulation point $\mathbf{P}^* \in \mathcal{P}(\Pi_c^0)$ of the sequence $(\mathbf{P}_n^*)_n$. We will now show that $\mathbf{P}^*$ is a weak solution of the MFG in the sense of Definition 2.2.

We first remark that, since every $\mathbf{P}_n^*$ belongs to $\mathcal{P}^*(\Pi^0)$ which is closed (see Proposition 3.1(iii)), it follows that $\mathbf{P}^*$ also belongs to $\mathcal{P}^*(\Pi^0)$, which means that $\mathbf{P}^*$ satisfies the individual optimality condition of Definition 2.2. We are left to show that $\mathbf{P}^*$ satisfies the weak equilibrium condition of Definition 2.2. In what follows, $(\mathbf{P}_n^*)_n$ denote the subsequence converging to $\mathbf{P}^*$.

We need to show that $M = \mathbf{P}^* \circ (X \mid \mathcal{F}^{M,W^0})^{-1}$, $\mathbf{P}^*$-a.s. This means that, for all $F \in \mathcal{F}^X$, $M(F) = \mathbf{P}^*[X \in F \mid \mathcal{F}^{M,W^0}]$ $\mathbf{P}^*$-a.s. By approximation it is enough to show that $M(\phi) = \mathbf{P}^*[\phi(X) \mid \mathcal{F}^{M,W^0}]$, $\mathbf{P}^*$-a.s. for any bounded continuous ϕ, and by the functional monotone class theorem (see [10, Chap. I, Theorem 19]), it is enough to show that, for any N, $t_1, \dots, t_N$, $\phi_1, \dots, \phi_N \in \mathcal{C}_b(\mathbf{R}^d)$, $\psi \in \mathcal{C}_b(\mathfrak{M}_+^1(\mathcal{X}))$, and $F \in \mathcal{F}^{W^0}$, we have

$$\text{(3.12)} \qquad \mathbf{E}^{\mathbf{P}^*}\Big[M\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big] = \mathbf{E}^{\mathbf{P}^*}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big].$$

For every n, $M = \mathbf{P}_n^*[\widehat{X}^n \mid \mathcal{F}^{n,W^0}]$. In particular, M is a.s. equal to an $\mathcal{F}^{n,W^0}$-measurable random measure, and $M = \mathbf{P}_n^*(\widehat{X}^n \mid \mathcal{F}^{n,W^0} \vee \mathcal{F}^M)^{-1}$ $\mathbf{P}_n^*$-a.s., implying that for all $n \geqslant 0$ and $F \in \mathcal{F}^{n,W^0}$,

$$\text{(3.13)} \qquad \mathbf{E}^{\mathbf{P}_n^*}\Big[M\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big] = \mathbf{E}^{\mathbf{P}_n^*}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(\widehat{X}^n_{t_i})\Big].$$

Since $\mathcal{F}^{n,W^0}$ is increasing in n, for fixed $F \in \mathcal{F}^{n,W^0}$, it follows that (3.13) also holds under $\mathbf{P}_k^*$ for all $k \geqslant n$. By the stable convergence of $\mathbf{P}_k^*$ to $\mathbf{P}^*$, the left-hand side of (3.13) tends to that of (3.12). So in order to show that (3.12) holds for this specific

$F \in \mathcal{F}^{n,W^0}$, we will show that

$$\mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i})\Big] \underset{k}{\to} \mathbf{E}^{\mathbf{P}^*}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big]. \tag{3.14}$$

We fix $\varepsilon > 0$. Since $(\mathbf{P}^*_k)_k$ is tight, we may fix a compact subset K_ε of $\mathcal{X}$ such that $\mathbf{P}^*_k(\mathcal{X} \setminus K_\varepsilon) \leqslant \varepsilon$ for all k, and such that $\widehat{X}^k$ converges uniformly to X on K_ε. Eventually, X and all the $\widehat{X}^n$ are uniformly bounded by some constant $C > 0$ on this K_ε, and all the ϕ_i are uniformly continuous on the closed ball $\overline{B}(0, C)$. In particular, there exists k_0 such that for all $k \geqslant k_0$, and $\omega \in K_\varepsilon$,

$$\Big|\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i}(\omega)) - \prod_{i\leqslant N}\phi_i(\omega(t_i))\Big| \leqslant \varepsilon. \tag{3.15}$$

This implies

$$\begin{aligned}
&\Big|\mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i})\Big] - \mathbf{E}^{\mathbf{P}^*}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big]\Big| \\
&\leqslant \Big|\mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i})\Big] - \mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big]\Big| \\
&\quad + \Big|\mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big] - \mathbf{E}^{\mathbf{P}^*}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big]\Big|.
\end{aligned} \tag{3.16}$$

It is immediate that the second term tends to zero, and for the first one we have, for all $k \geqslant k_0$,

$$\begin{aligned}
&\Big|\mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i})\Big] - \mathbf{E}^{\mathbf{P}^*_k}\Big[\psi(M)\mathbf{1}_F(W^0)\prod_{i\leqslant N}\phi_i(X_{t_i})\Big]\Big| \\
&\leqslant \|\psi\|_\infty \mathbf{E}^{\mathbf{P}^*_k}\Big[\Big|\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i}) - \prod_{i\leqslant N}\phi_i(X_{t_i})\Big|\Big] \\
&\leqslant \|\psi\|_\infty \mathbf{E}^{\mathbf{P}^*_k}\Big[\mathbf{1}_{K_\varepsilon}\Big|\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i}) - \prod_{i\leqslant N}\phi_i(X_{t_i})\Big|\Big] \\
&\quad + \|\psi\|_\infty \mathbf{E}^{\mathbf{P}^*_k}\Big[\mathbf{1}_{\mathcal{X}\setminus K_\varepsilon}\Big|\prod_{i\leqslant N}\phi_i(\widehat{X}^k_{t_i}) - \prod_{i\leqslant N}\phi_i(X_{t_i})\Big|\Big] \\
&\leqslant 2^N \varepsilon \,\|\psi\|_\infty \prod_{i\leqslant N}\|\phi_i\|_\infty + \varepsilon\,\|\psi\|_\infty.
\end{aligned} \tag{3.17}$$

Since we may choose ε as small as we want, we see that (3.14) holds and, therefore, (3.13) holds, for any $F \in \mathcal{F}^{n,W^0}$. Since this is true for any n, relation (3.12) holds for any $F \in \bigcup_n \mathcal{F}^{n,W^0}$.

Notice that $\bigcup_n \mathcal{F}^{n,W^0}$ is stable by finite intersection and hence forms a π-system; see Definition 4.9 in [1]. The sets of $F \in \mathcal{F}^{W^0}$ verifying (3.12) form a monotone class (also called a λ-system; see Definition 4.9 in [1] again), so by the monotone class theorem (or Dynkin's lemma; see 4.11 in [1]), we see that (3.12) holds for all $F \in \sigma\big(\bigcup_n \mathcal{F}^{n,W^0}\big)$, which is equal to $\mathcal{F}^{W^0}$ (see Remark 3.3(i)), completing the proof.

4. McKean–Vlasov second order backward SDEs. From now on, we specialize the discussion to the no common noise context, i.e., $p_0 = 0$ and $W = W^1$. Consequently the distribution of X is now deterministic since it is no longer conditioned on the common noise. We shall work on the smaller canonical space $\Omega = \mathcal{X} \times \mathcal{Q}$ by appropriate projection of $\mathcal{W}$.

This section contains the second main results of the paper. Our objective is to provide a characterization of the solution of the MFG in the no common noise context by means of a McKean–Vlasov second order backward SDE (2BSDE). This requires a nondegeneracy condition obtained by separating the control of the drift and that of the diffusion coefficient. We therefore introduce two control sets A and B in which the drift control process and the diffusion control process take values, respectively.

We denote by $\mathcal{Q}^A$ the set of relaxed controls, i.e., of measures q on $[0,T] \times A$ such that $q(\cdot \times A)$ is equal to the Lebesgue measure. Each $q \in \mathcal{Q}^A$ may be identified with a measurable function $t \to q_t$ from $[0,T]$ to $\mathfrak{M}^1_+(A)$ determined a.e. by $q(dt, da) = q_t(da)\, dt$.

We define similarly the set of relaxed controls $\mathcal{Q}^B$ by replacing the space A with B, and we denote $\mathcal{Q} := \mathcal{Q}^A \times \mathcal{Q}^B$ with corresponding canonical process $Q := (Q^A, Q^B)$.

As in the previous section, we equip these spaces with their natural filtrations. We also introduce the right-continuous filtration $\mathbb{F}^{X,+}$ defined for all $t \in [0,T]$ by $\mathcal{F}^{X,+}_t := \bigcap_{n \geqslant 0} \mathcal{F}^X_{t+1/n}$.

We denote by $\mathfrak{SM}$ the set of all $\mathbf{P} \in \mathfrak{M}^1_+(\mathcal{X})$ such that X is a $\mathbf{P}$-semimartingale with absolutely continuous bracket process. According to Karandikar [22], there exists an $\mathbb{F}^X$-progressively measurable process, denoted by $\langle X \rangle$, which coincides with the quadratic variation of X $\mathbf{P}$-a.s. for every $\mathbf{P} \in \mathfrak{SM}$. We may then introduce the process $\widehat{\sigma}^2$ defined by

$$\widehat{\sigma}^2_t := \limsup_{\varepsilon \searrow 0} \frac{\langle X \rangle_t - \langle X \rangle_{t-\varepsilon}}{\varepsilon}, \qquad t \in [0,T].$$

This process is progressively measurable and takes values in the set $\mathbb{S}^+_d$ of $d \times d$ nonnegative symmetric matrices.

We now fix $\mathcal{P} \subset \mathfrak{SM}$. For all $\mathbf{P} \in \mathcal{P}$ and $t \in [0,T]$ we denote by $\mathcal{F}^{X,+,\mathbf{P}}_t$ the σ-field $\mathcal{F}^{X,+}_t$ augmented with $\mathbf{P}$-null sets, and we denote by $\mathbb{F}^{X,+,\mathcal{P}}$ the filtration given by

$$\mathcal{F}^{X,+,\mathcal{P}}_t := \bigcap_{\mathbf{P} \in \mathcal{P}} \mathcal{F}^{X,+,\mathbf{P}}_t, \qquad t \in [0,T].$$

We say that a property holds $\mathcal{P}$-quasi surely (abbreviated $\mathcal{P}$-q.s.) if it holds $\mathbf{P}$-a.s. for all $\mathbf{P} \in \mathcal{P}$. We also denote by $\mathbb{S}^2(\mathcal{P})$ the collection of all càdlàg $\mathbb{F}^{X,+,\mathcal{P}}$-adapted processes S with

$$\|S\|^2_{\mathbb{S}^2(\mathcal{P})} := \sup_{\mathbf{P} \in \mathcal{P}} \mathbf{E}^{\mathbf{P}}\Big[\sup_{t \leqslant T} S^2_t\Big] < \infty.$$

Finally, we denote by $\mathbb{H}^2(\mathcal{P})$ the collection of all $\mathbb{F}^{X,+,\mathcal{P}}$-progressively measurable processes H with

$$\|H\|^2_{\mathbb{H}^2(\mathcal{P})} := \sup_{\mathbf{P} \in \mathcal{P}} \mathbf{E}^{\mathbf{P}}\left[\int_0^T H_t^\top \, d\langle X \rangle_t \, H_t\right] = \sup_{\mathbf{P} \in \mathcal{P}} \mathbf{E}^{\mathbf{P}}\left[\int_0^T H_t^\top \widehat{\sigma}^2_t H_t \, dt\right] < \infty.$$

4.1. Controlled state process. For a fixed $m \in \mathfrak{M}_+^1(\mathcal{X})$, the controlled state is defined by the relaxed SDE

$$
\begin{aligned}
X_t = X_0 &+ \int_0^t \int_{A\times B} (\sigma_r\lambda_r)(X, m, a, b)\, Q_r(da, db)\, dr \\
&+ \int_B \sigma_r(X, m, b)\, N^B(db, dr),
\end{aligned} \tag{4.1}
$$

where N^B is a martingale measure with intensity $Q_t^B\, dt$, and

$$
\begin{aligned}
&\lambda\colon [0,T]\times\mathcal{X}\times\mathfrak{M}_+^1(\mathcal{X})\times A \to \mathbf{R}^d,\\
&\sigma\colon [0,T]\times\mathcal{X}\times\mathfrak{M}_+^1(\mathcal{X})\times B \to \mathbb{M}_{p,d}(\mathbf{R})
\end{aligned}
$$

are progressively measurable maps (in the sense detailed in subsection 2.2). The generator of our controlled martingale problem is defined for $\phi \in \mathcal{C}_b^2(\mathbf{R}^d)$, $(a,b) \in A\times B$, and $(t,x,y) \in [0,T]\times\mathcal{X}\times\mathbf{R}^d$ by

$$
\mathcal{A}_{t,x}^{a,b,m}\phi(y) := (\sigma_t\lambda_t)(x,m,a,b)\cdot D\phi(y) + \frac{1}{2}\sigma_t\sigma_t^\top(x,m,b) : D^2\phi(y).
$$

Definition 4.1. *Fix some $q_0 \in A\times B$, and denote by Q^0 the measure defined by $Q_t^0 = \delta_{q_0}$, $t\in[0,T]$. For $(s,x)\in[0,T]\times\mathcal{X}$ and $m\in\mathfrak{M}_+^1(\mathcal{X})$,*

(i) *let $\overline{\mathcal{P}}_{s,x}^m$ be the subset of all $\mathbf{P}\in\mathfrak{M}_+^1(\Omega)$ such that*

$$
\mathbf{P}[(X_{\wedge s}, Q_{\wedge s}) = (x_{\wedge s}, Q_{\wedge s}^0)] = 1,
$$

and let

$$
\phi(X_t) - \int_s^t \int_{A\times B} \mathcal{A}_{r,X}^{a,b,m}\phi(X_r)\, Q_r(da,db)\, dr, \qquad t\in[s,T],
$$

be a $(\mathbf{P},\mathbb{F})$-martingale for all $\phi\in\mathcal{C}_b^2(\mathbf{R}^d)$;

(ii) *let $\overline{\mathcal{M}}_{s,x}^m$ be the subset of all $\mathbf{P}\in\mathfrak{M}_+^1(\Omega)$ such that*

$$
\mathbf{P}[(X_{\wedge s}, Q_{\wedge s}) = (x_{\wedge s}, Q_{\wedge s}^0] = 1,
$$

and let

$$
\phi(X_t) - \frac{1}{2}\int_s^t \int_B \sigma_t\sigma_t^\top(x,m,b) : D^2\phi(X_r)\, Q_r^B(db)\, dr, \qquad t\in[s,T],
$$

be a $(\mathbf{P},\mathbb{F})$-martingale for all $\phi\in\mathcal{C}_b^2(\mathbf{R}^d)$.

For any s, x, m, we set $\mathcal{P}_{s,x}^m := \{\mathbf{P}\circ X^{-1}\colon \mathbf{P}\in\overline{\mathcal{P}}_{s,x}^m\}$ and $\mathcal{M}_{s,x}^m := \{\mathbf{P}\circ X^{-1}\colon \mathbf{P}\in\overline{\mathcal{M}}_{s,x}^m\}$.

Finally, we set $\overline{\mathcal{M}}^m := \overline{\mathcal{M}}_{0,0}^m$, $\overline{\mathcal{P}}^m := \overline{\mathcal{P}}_{0,0}^m$, $\mathcal{M}^m := \mathcal{M}_{0,0}^m$, and $\mathcal{P}^m := \mathcal{P}_{0,0}^m$.

4.2. Solving a McKean–Vlasov 2BSDE. As in the previous sections, let $\xi\colon\mathcal{X}\to\mathbf{R}$ be an r.v., let $f\colon[0,T]\times\mathcal{X}\times\mathfrak{M}_+^1(\mathcal{X})\times A\times B\to\mathbf{R}$ a progressively measurable process, and denote the dynamic version of the value function of the individual optimization problem for all $(t,x,m)\in[0,T]\times\mathcal{X}\times\mathfrak{M}_+^1(\mathcal{X})$ by

$$
V_t^m(x) := \sup_{\mathbf{P}\in\overline{\mathcal{P}}_{t,x}^m} \mathbf{E}^{\mathbf{P}}\Big[\xi + \int_t^T \int_{A\times B} f_r(m,a,b)\, Q_r(da,db)\, dr\Big].
$$

The backward SDE characterization of the solution of the MFG requires introducing the following nonlinearity:

$$
\begin{aligned}
F_t(x,z,\Sigma,m) &:= \sup_{q\in\mathbf{Q}_t(x,\Sigma,m)} H_t(x,z,m,q),\\
H_\cdot(\,\cdot\,,z,\,\cdot\,,q) &:= \int_{A\times B}(f+z\cdot\sigma\lambda)\,dq
\end{aligned}
\tag{4.2}
$$

for all $(t,x,z,\Sigma,m)\in[0,T]\times\mathcal{X}\times\mathbf{R}^d\times\mathbb{S}_d^+\times\mathfrak{M}_+^1(\mathcal{X})$, where

$$
\mathbf{Q}_t(x,\Sigma,m) := \left\{q\in\mathfrak{M}_1^+(A)\otimes\mathfrak{M}_1^+(B)\colon \int_B \sigma_t\sigma_t^\top(x,m,b)\,q^B(db)=\Sigma\right\}. \tag{4.3}
$$

The following condition is a restatement of Assumption 3.1 in the present context, with a sufficient condition for the well-posedness of the controlled SDE.

Assumption 4.1. The following assertions hold:

(1) the functions ξ, f, λ, σ are bounded;

(2) the functions ξ and f_t, λ_t, σ_t for all t are continuous;

(3) the functions λ, σ are locally Lipschitz continuous in x uniformly in (t,a) at fixed m.

We are now ready for our main characterization of a solution of the MFG from Theorem 3.1 in terms of the McKean–Vlasov second order backward SDE.

DEFINITION 4.2. *We say that* $(m,Y,Z)\in\mathfrak{M}_+^1(\mathcal{X})\times\mathbb{S}^2(\mathcal{P}^m)\times\mathbb{H}^2(\mathcal{P}^m)$ *solves the McKean–Vlasov* 2*BSDE*

$$
Y_t=\xi+\int_t^T F_r(X,Z_r,\widehat{\sigma}_r^2,m)\,dr-\int_t^T Z_r\,dX_r+U_T-U_t,\qquad t\in[0,T],\quad \mathcal{P}^m\text{-q.s.} \tag{4.4}
$$

if the following conditions are met:

(1) *the process* $U:=Y_\cdot-Y_0+\int_0^\cdot F_r(Z_r,\widehat{\sigma}_r^2,m)\,dr-\int_0^\cdot Z_r\,dX_r$ *is a* $\mathbf{P}$*-cádlág supermartingale, orthogonal to* X *for every* $\mathbf{P}\in\mathcal{P}^m$;

(2) $m\in\mathcal{P}^m$ *and* U *is an* m*-martingale.*

Notice that (4.4) differs from the notion introduced in [30] and further developed in [29], [26] by the fact that both the nonlinearity and the set of probability measures depend on the law of X, denoted m. We emphasize that m should not be understood as the law of X under arbitrary $\mathbf{P}\in\mathcal{P}^m$. Instead, m denotes the "optimal" measure in $\mathcal{P}^m$, i.e., the one under which U is a martingale. In other words, the law m which parametrizes the 2BSDE coincides with the optimal law for X within the set of measures under which the 2BSDE holds.

We now state the main result of this second part of the paper, whose proof is postponed to section 6.

THEOREM 4.1. *Let Assumption* 4.1 *be met. Then, there exists a solution* m *of the MFG with coefficients* $\sigma\lambda$, σ, f, ξ, *which induces a solution* (m,Y,Z) *of the McKean–Vlasov* 2*BSDE* (4.4).

Moreover, $Y=V^m$ *means that* $Y_t(x)=V_t^m(x)$ *for all* $(t,x)\in[0,T]\times\mathcal{X}$.

This result provides a connection between the MFG problem and the corresponding second order backward SDE. It would be interesting to analyze directly the well-posedness of such SDEs so as to induce a solution of the MFG equilibrium, we leave this question for future research.

5. 2BSDE representation of relaxed controlled problems. The aim of this section is to introduce the tools needed for the proof of Theorem 4.1. We keep working with the spaces introduced at the beginning of the previous section. However, since marginal distribution m is fixed throughout, we shall drop the dependence on this parameter throughout this section.

5.1. Controlled state process, optimization problem, and value function. The controlled state process is defined by the relaxed SDE (4.1), and the dynamic version of the value function of this control problem is defined by setting, for any $(s,x) \in [0,T] \times \mathcal{X}$,

$$\begin{aligned} &V_s(x) := \sup_{\mathbf{P} \in \overline{\mathcal{P}}_{s,x}} J_s(\mathbf{P}), \\ &J_s(\mathbf{P}) := \mathbf{E}^{\mathbf{P}}\left[\xi + \int_s^T \int_{A\times B} f_r(X,a,b)\, Q_r(da,db)\, dr\right], \end{aligned} \tag{5.1}$$

where ξ, f are jointly measurable, with f progressively measurable in (t,x), and the spaces of probability measures $\overline{\mathcal{P}}_{s,x}$, $\overline{\mathcal{P}}$, $\mathcal{M}_{s,x}$, $\mathcal{M}$, $\mathcal{P}_{s,x}$, $\mathcal{P}$ are defined as in Definition 4.1, with dependence on m dropped throughout.

PROPOSITION 5.1. *Under Assumption* 4.1, *the set-valued map* $(s,x) \to \overline{\mathcal{P}}_{s,x}$ *is a compact valued continuous correspondence*, V *is continuous on* $[0,T]\times\mathcal{X}$, *and the existence holds for problem* (5.1).

Proof. The compactness of $\overline{\mathcal{P}}_{s,x}$ is a consequence of Proposition 3.1(ii). Notice that the correspondence $\Gamma\colon (s,x) \in [0,T]\times\mathcal{X} \to \{(s,x)\}\times\mathfrak{M}^1_+(\mathcal{Q})$ is continuous as the product of the continuous mapping $(s,x)\to(s,x)$ and of the constant compact valued (hence continuous) correspondence $(s,x)\to\mathfrak{M}^1_+(\mathcal{Q})$; see [1, Theorem 17.28].

Since λ, σ are locally Lipschitz in x uniformly in (t,a,b), for any $\mathbf{Q}\in\mathfrak{M}^1_+(\mathcal{Q})$, there exists a unique weak solution of the corresponding SDE, i.e., a unique $\mathbf{P}\in\overline{\mathcal{P}}_{s,x}$ such that $\mathbf{P}\circ Q^{-1} = \mathbf{Q}$.

We denote by $\phi(s,x,\mathbf{Q})$ this unique $\mathbf{P}$. It is clear that $(s,x)\to\overline{\mathcal{P}}_{s,x}$ is equal to $\phi\circ\Gamma$, so by continuity of the composition of continuous correspondences (see Proposition A.1(4)), we are left to show that ϕ is continuous.

We fix a converging sequence $(s_n,x_n,\mathbf{Q}_n)\to(s,x,\mathbf{Q})$ in $[0,T]\times\mathcal{X}\times\mathfrak{M}^1_+(\mathcal{Q})$. Since $(x_n)_n$ converges, it is included in a compact subset C of $\mathcal{X}$. For all n, $\phi(s_n,x_n,\mathbf{Q}_n)\circ X^{-1}$ is the law of a process which coincides with $x_n\in C$ on $[0,s_n]$ and which is a semimartingale with bounded (uniformly in n) characteristics on $[s_n,T]$. Hence, adapting the proof of Proposition 6.2 in [2], we see that $(\phi(s_n,x_n,\mathbf{Q}_n)\circ X^{-1})_n$ is tight. Since A, B are compact sets, it follows that $(\phi(s_n,x_n,\mathbf{Q}_n))_n$ is also tight. We now show that $\phi(s,x,\mathbf{Q})$ is its only possible limiting point, and the proof of the first assertion will be complete. Assume (along some subsequence) that $\phi(s_n,x_n,\mathbf{Q}_n)$ tends to some $\mathbf{P}\in\mathfrak{M}^1_+(\Omega)$. Clearly, $\mathbf{P}\circ Q^{-1}=\mathbf{Q}$. Since $\phi(s,x,\mathbf{Q})$ is unique $\mathbf{P}\in\overline{\mathcal{P}}_{s,x}$ such that $\mathbf{P}\circ Q^{-1}=\mathbf{Q}$, in order to show that $\mathbf{P}=\phi(s,x,\mathbf{Q})$ and to conclude, it is enough to show that $\mathbf{P}\in\overline{\mathcal{P}}_{s,x}$. This is shown exactly as in Proposition 6.3 of [2]. This shows the continuity of $(s,x)\to\overline{\mathcal{P}}_{s,x}$.

It remains to show that V is continuous. We remark that for all (s,x), we have $V_s(x) = \sup_{\mathbf{P}\in\overline{\mathcal{P}}_{s,x}} J_0(\mathbf{P}) - \int_0^s f_r(x,q_0)\,dr$. Since ξ, f are bounded and ξ and f_t for all t are continuous, J_0 is continuous. Since $(s,x)\to\overline{\mathcal{P}}_{s,x}$ is continuous and compact valued, the above supremum is in fact the maximum, and the Berge maximum theorem (see Theorem A.1) states that $(s,x)\to\max_{\mathbf{P}\in\overline{\mathcal{P}}_{s,x}} J_0(\mathbf{P})$ is continuous. Finally, the dominated convergence theorem permits us to show that $(s,x) \to \int_0^s f_r(x,q_0)\,dr$ is continuous, and hence V is continuous.

5.2. 2BSDE solved by the value function. We recall F, H, and $\mathbf{Q}$ were introduced in (4.2), (4.3); we will again drop the parameter m.

LEMMA 5.1. (i) *F is jointly measurable and uniformly Lipschitz in z;*

(ii) *there exists a measurable mapping $\widehat{q}\colon [0,T]\times\mathcal{X}\times\mathbf{R}^d\times\mathbb{S}_d^+ \to \mathfrak{M}_1^+(A)\otimes\mathfrak{M}_1^+(B)$ such that, for all $(t,x,z,\Sigma)\in[0,T]\times\mathcal{X}\times\mathbf{R}^d\times\mathbb{S}_d^+$,*

$$\widehat{q}_t(x,z,\Sigma)\in\mathbf{Q}_t(x,\Sigma) \quad \textit{and} \quad F_t(x,z,\Sigma) = H_t\big(x,z,\widehat{q}_t(x,z,\Sigma)\big).$$

Proof. (i) The joint measurability of f follows from (ii), which will be proved below, together with the measurability of f,λ,σ (hence of H), and that of $\widehat{q}$. Further, observe that $H_t(x,\cdot\,,q)$ is an affine mapping with slope $\int_{A\times B}\sigma_r(x,b)\lambda_r(x,a)\,q(da,db)$. In particular, $F_t(x,\cdot\,,\Sigma)$ is convex as the supremum of affine mappings. Denoting by $\partial F_t(x,\cdot\,,\Sigma)$ its subgradient, since $\mathbf{Q}_t(x,\Sigma)$ is compact and $q\to H_t(x,z,q)$ is continuous for all z, we have (see [17, section D, Theorem 4.4.2]), for all z,

$$\partial F_t(x,\cdot\,,\Sigma)(z)\subset \mathrm{co}\left(\left\{\int_{A\times B}\sigma_r(x,b)\lambda_r(x,a)\,q(da,db)\colon q\in\mathbf{Q}_t(x,\Sigma)\right\}\right),$$

where co denotes the convex hull. In particular, $\partial F_t(x,\cdot\,,\Sigma)(z)$ is included in the centered closed ball of radius $\|\sigma\lambda\|_\infty$. This implies that the semidirectional derivatives of $F_t(x,\cdot\,,\Sigma)$ exist at all z and are bounded by $\|\sigma\lambda\|_\infty$ and, therefore, that this mapping is $\|\sigma\lambda\|_\infty$-Lipschitz.

(ii) Our aim is to show the existence of a measurable selector for the correspondence $(t,x,z,\Sigma)\to\arg\max_{q\in\mathbf{Q}_t(x,\Sigma)}H_t(x,z,q)$. Theorems 18.19 and 18.10 in [1] state that if H is continuous in q for fixed (t,x,z) and measurable in (t,x,z) for fixed q, and if $\mathbf{Q}$ is a measurable correspondence with compact values, then such a measurable selector indeed exists.

By the boundedness and continuity of f_t,λ_t,σ_t for all t, it is immediate that H verifies the conditions mentioned above. It is also clear that $\mathbf{Q}_t(x,\Sigma)$ is a compact subset of $\mathfrak{M}_1^+(A)\otimes\mathfrak{M}_1^+(B)$ for all t, x, Σ. So we are left to show that $\mathbf{Q}$ is a measurable correspondence.

Finally, since $\mathbf{Q}_t(x,\Sigma)=\{q\in\mathfrak{M}_1^+(A)\otimes\mathfrak{M}_1^+(B)\colon h(t,x,\Sigma,q)=0\}$ with $\mathfrak{M}_1^+(A)\otimes\mathfrak{M}_1^+(B)$ compact and $h\colon(t,x,\Sigma,q)\to\int_B\sigma\sigma_t^\top(x,b)\,q^B(db)-\Sigma$, which is measurable in (t,x,Σ) at fixed q and continuous in q at fixed (t,x,Σ), then by Corollary 18.8 in [1], $\mathbf{Q}$ is indeed measurable, and the proof is complete.

Further, we recall the definition of a solution for the 2BSDE:

$$Y_t=\xi+\int_t^T F_r(Z_r,\widehat{\sigma}_r^2)\,dr-\int_t^T Z_r\,dX_r+U_T-U_t,\quad \mathcal{P}\text{-q.s.} \tag{5.2}$$

(see, for instance, [26, Definition 3.9] in which the terminal time may be random). We introduce the additional notation

$$\mathcal{P}_{t,\mathbf{P}}:=\{\mathbf{P}'\in\mathcal{P}\colon \mathbf{P}' \text{ coincides with } \mathbf{P} \text{ on } \mathcal{F}_t^{X,+}\}. \tag{5.3}$$

DEFINITION 5.1. *A pair of processes* $(Y, Z) \in \mathbb{S}^2(\mathcal{P}) \times \mathbb{H}^2(\mathcal{P})$ *is a solution of the* 2*BSDE* (5.2) *if the process*

$$U_t := Y_t - Y_0 + \int_0^t F_r(Z_r, \widehat{\sigma}_r^2)\, dr - \int_0^t Z_r\, dX_r, \qquad t \in [0, T],$$

is a $\mathbf{P}$*-càdlàg supermartingale orthogonal to* X *for all* $\mathbf{P} \in \mathcal{P}$, *and if it satisfies the minimality condition*

$$U_t = \operatorname*{ess\,inf}_{\mathbf{P}' \in \mathcal{P}_{t,\mathbf{P}}}{}^{\mathbf{P}} \mathbf{E}^{\mathbf{P}'}[U_T \mid \mathcal{F}_t^{X,+,\mathbf{P}}], \qquad t \in [0, T], \quad \mathbf{P}\text{-}a.s.$$

Remark 5.1. We recall that under the continuum hypothesis, the stochastic integral $\int_t^T Z_r\, dX_r$ may be defined for all ω independently of the choice of the probability in $\mathcal{P}$ (see [28]).

The aim of this subsection is to show the following representation result for the value function.

THEOREM 5.1. *Under Assumption* 4.1, $V \in \mathbb{S}^2(\mathcal{P})$, *and there exists* $Z \in \mathbb{H}^2(\mathcal{P})$ *such that* (V, Z) *solves the* 2*BSDE* (5.2).

To prove this result, we follow the same argument as in [31] by introducing

$$\widehat{\mathcal{Y}}_t(x) := \sup_{\mathbf{P} \in \mathcal{M}_{t,x}} \mathbf{E}^{\mathbf{P}}[Y_t^{t,x,\mathbf{P}}] \quad \text{for all } (t, x) \in [0, T] \times \mathcal{X}, \tag{5.4}$$

where $(Y^{t,x,\mathbf{P}}, Z^{t,x,\mathbf{P}})$ is a unique solution of the following (well-posed) BSDE on the space $(\mathcal{X}, \mathcal{F}^X, \mathbb{F}^{X,+}, \mathbf{P})$:

$$Y_s^{t,x,\mathbf{P}} = \xi + \int_s^T F_r(Z_r^{t,x,\mathbf{P}}, \widehat{\sigma}_r^2)\, dr - Z_r^{t,x,\mathbf{P}}\, dX_r - dM_r^{t,x,\mathbf{P}}, \qquad s \in [t, T], \tag{5.5}$$

for some martingale $M^{t,x,\mathbf{P}}$, with $\langle X, M^{t,x,\mathbf{P}} \rangle = 0$, $\mathbf{P}$-a.s.

PROPOSITION 5.2. *We have* $V = \widehat{\mathcal{Y}}$.

Proof. Denote

$$f_r^Q := \int_{A \times B} f_r(a, b)\, Q_r(da, db),$$
$$b_r^Q := \int_{A \times B} \sigma_r(b) \lambda_r(a)\, Q_r(da, db),$$

and fix $(t, x) \in [0, T] \times \mathcal{X}$.

1. We first prove that $V_t(x) \leqslant \widehat{\mathcal{Y}}_t(x)$. For an arbitrary $\mathbf{P} \in \overline{\mathcal{P}}_{t,x}$, by Theorem 2.7 in [27] there exists an $\mathbb{F}^X$-progressively measurable process $\overline{q}$ such that the *feedback control* $\mathbf{P} \circ (X, \overline{q}(X))^{-1}$ belongs to $\overline{\mathcal{P}}_{t,x}$ and

$$\mathbf{E}^{\mathbf{P}}\left[\xi + \int_t^T f_r^Q\, dr\right] = \mathbf{E}^{\mathbf{P}}\left[\xi + \int_t^T f_r^{\overline{q}(X)}\, dr\right].$$

We now work on the filtered space $(\mathcal{X}, \mathcal{F}^X, \mathbb{F}^{X,+})$. Even though $\mathbf{P}$ is defined on the larger space $(\Omega, \mathcal{F})$, we will often write $\mathbf{P}$ instead of $\mathbf{P} \circ X^{-1}$ when there can be no confusion. By Theorem IV-2 in [15], on a bigger space there exists a martingale measure N^B with intensity $\overline{q}^B(X)_t\, dt$ such that

$$dX_s = b_s^{\overline{q}(X)}\, ds + \int_B \sigma_s(X,b)\, N^B(db, ds).$$

Notice that the process

$$L_s := -\int_t^s \left(\int_A \lambda_r(X,a)\, \overline{q}_r^A(X)(da)\right) \int_B N^B(db, dr), \qquad s \in [t,T],$$

is a continuous martingale with bounded quadratic variation. Now we may introduce the probability measure $G(\mathbf{P})$ by

$$\frac{dG(\mathbf{P})}{d\mathbf{P}} = \mathcal{E}(L) := e^{L - \langle L \rangle/2}.$$

Since

$$\langle X, L\rangle = -\left\langle \int_t^\cdot \int_B \sigma_r(b)\, N^B(db,dr), L\right\rangle = \int_t^\cdot b_r^{\overline{q}(X)}\, dr,$$

it follows from the Girsanov theorem that X is a $G(\mathbf{P})$-martingale with unchanged quadratic variation

$$\langle X\rangle = \int_t^\cdot \int_B \sigma\sigma_r^\top(X,b)\, \overline{q}(X)_r^B(db)\, dr, \qquad G(\mathbf{P})\text{-a.s.}$$

Hence $G(\mathbf{P}) \in \mathcal{M}_{t,x}$.

Considering on $(\mathcal{X}, \mathcal{F}^X, \mathbb{F}^{X,+}, G(\mathbf{P}))$ the BSDE

$$\overline{Y}_s^{t,x,G(\mathbf{P})} = \xi + \int_s^T \big(f_r^{\overline{q}(X)} + \overline{Z}_r^{t,x,G(\mathbf{P})} b_r^{\overline{q}(X)}\big)\, dr - \overline{Z}_r^{t,x,G(\mathbf{P})}\, dX_r - d\overline{M}_r^{t,x,G(\mathbf{P})}, \tag{5.6}$$

for $s \in [t,T]$, we will now show that we have

$$\mathbf{E}^{\mathbf{P}}\left[\xi + \int_t^T f_r^Q\, dr\right] = \mathbf{E}^{G(\mathbf{P})}[\overline{Y}_t^{t,x,G(\mathbf{P})}] \leqslant \mathbf{E}^{G(\mathbf{P})}[Y_t^{t,x,G(\mathbf{P})}], \tag{5.7}$$

and this implies that $V_t(x) \leqslant \widehat{\mathcal{Y}}_t(x)$. In order to show that the equality in (5.7) holds, we consider under $\mathbf{P}$ the solution $(\widetilde{Y}, \widetilde{Z}, \widetilde{M})$ of the BSDE

$$\widetilde{Y}_s = \xi + \int_s^T f_r^{\overline{q}(X)}\, dr - \widetilde{Z}_r\, dX_r + \widetilde{Z}_r b_r^{\overline{q}(X)}\, dr - d\widetilde{M}_r, \qquad s \in [t,T].$$

Since $X - \int_t^\cdot b_r^{\overline{q}(X)} dr$ is a $\mathbf{P}$-martingale, it follows that $\widetilde{Y} + \int_t^\cdot f_r^{\overline{q}(X)} dr$ is also a $\mathbf{P}$-martingale, and hence by the Girsanov theorem, $\widetilde{Y} + \int_t^\cdot f_r^{\overline{q}(X)}\, dr - \langle \widetilde{Y}, L\rangle$ is a $G(\mathbf{P})$-martingale. Since X is a $G(\mathbf{P})$-martingale, we obtain by standard decomposition that

$$\widetilde{Y}_s = \xi + \int_s^T f_r^{\overline{q}(X)}\, d_r - \int_s^T d\langle \widetilde{Y}, L\rangle_r - \int_s^T Z_r'\, dX_r + (M_T' - M_s'), \qquad s \in [t,T],$$

for some process Z' and some martingale M' orthogonal to X. Hence $\langle \widetilde{Y}, L\rangle = \int_t^{\cdot} Z'_r\, d\langle X, L\rangle = -\int_t^{\cdot} Z'_r b_r^{\overline{q}(X)}\, dr$, and, therefore,

$$\widetilde{Y}_s = \xi + \int_s^T (f_r^{\overline{q}(X)} + Z'_r b_r^{\overline{q}(X)})\, dr - \int_s^T Z'_r\, dX_r + (M'_T - M'_s), \qquad s \in [t, T],$$

which implies that $\widetilde{Y} = \overline{Y}^{t,x,G(\mathbf{P})}$ $G(\mathbf{P})$-a.s. by the uniqueness of the solution of a BSDE. In particular,

$$\begin{aligned}\mathbf{E}^{\mathbf{P}}\left[\xi + \int_t^T f_r^Q\, dr\right] &= \mathbf{E}^{\mathbf{P}}\left[\xi + \int_t^T f_r^{\overline{q}(X)}\, dr\right] = \mathbf{E}^{\mathbf{P}}[\widetilde{Y}_t] \\ &= \mathbf{E}^{G(\mathbf{P})}[\widetilde{Y}_t] = \mathbf{E}^{G(\mathbf{P})}[\overline{Y}_t^{t,x,G(\mathbf{P})}].\end{aligned}$$

By the comparison theorem for BSDEs (see Theorem 2.2 in [14], for instance) and the definition of F and $\widehat{\sigma}^2$ we have $\mathbf{E}^{G(\mathbf{P})}[\overline{Y}_t^{t,x,G(\mathbf{P})}] \leqslant \mathbf{E}^{G(\mathbf{P})}[Y_t^{t,x,G(\mathbf{P})}]$, and, therefore, the inequality in (5.7) holds.

2. Further, we prove the converse inequality $V_t(x) \geqslant \widehat{\mathcal{Y}}_t(x)$. Recall that the maximizer $\widehat{q}$ was introduced in Lemma 5.1, and denote $\widehat{q}_r := \widehat{q}_r(X, Z_r^{t,x,\mathbf{P}}, \widehat{\sigma}_r^2)$. Then, for all $\mathbf{P} \in \mathcal{M}_{t,x}$, we have

$$\begin{aligned} Y_s^{t,x,\mathbf{P}} &= \xi + \int_s^T H_r(X, Z_r^{t,x,\mathbf{P}}, \widehat{q}_r)\, dr \\ &\qquad - Z_r^{t,x,\mathbf{P}}\, dX_r + dM_r^{t,x,\mathbf{P}}, \qquad s \in [t, T], \quad \mathbf{P}\text{-a.s.} \end{aligned} \tag{5.8}$$

Proceeding as in the first part of this proof, we consider the change of measure $d\mathbf{Q}/d\mathbf{P} := \mathcal{E}(\widehat{L})$, where $\widehat{L} := -\int_t^{\cdot}\int_{A\times B} \lambda_r(X, a)\widehat{q}_r^A(da)\, dN^B(db, dr)$. Since $\langle X, \widehat{L}\rangle = -\int_t^{\cdot} b_r^{\widehat{q}}\, dr$ $\mathbf{P}$-a.s., it follows from (5.8) that $\langle Y^{t,x,\mathbf{P}}, \widehat{L}\rangle = -\int_t^{\cdot} Z_r^{t,x,\mathbf{P}} b_r^{\widehat{q}}\, dr$, and we conclude from the Girsanov theorem that $Y^{t,x,\mathbf{P}}$ is a $\mathbf{Q}$-martingale. Finally, let $\widehat{G}(\mathbf{P}) := \mathbf{Q} \circ (X, \widehat{q})^{-1}$. By construction, $\widehat{G}(\mathbf{P})$ belongs to $\overline{\mathcal{P}}_{t,x}$, and we have

$$\mathbf{E}^{\widehat{G}(\mathbf{P})}\left[\xi + \int_t^T f_r^Q\, dr\right] = \mathbf{E}^{\mathbf{Q}}\left[\xi + \int_t^T f_r^{\widehat{q}}\, dr\right] = \mathbf{E}^{\mathbf{Q}}[Y_t^{t,x,\mathbf{P}}] = \mathbf{E}^{\mathbf{P}}[Y_t^{t,x,\mathbf{P}}].$$

By the arbitrariness of $\mathbf{P} \in \mathcal{M}_{t,x}$, and the fact that $\widehat{G}(\mathbf{P})$ belongs to $\overline{\mathcal{P}}_{t,x}$, this implies that $V_t(x) \geqslant \widehat{\mathcal{Y}}_t(x)$.

Proof of Theorem 5.1. By Proposition 5.2, we have $V = \widehat{\mathcal{Y}}$. Moreover, $(t, x) \to V_t(x)$ is continuous by Proposition 5.1, so $t \to \widehat{\mathcal{Y}}_t(X_{\wedge t})$ is a continuous process. The present theorem now follows from Theorem 4.6 in [30] or section 4.4 of [29], where we do not have to consider the path regularization of $t \to \widehat{\mathcal{Y}}_t(X_{\wedge t})$, as we have shown that it is continuous in the present setup.

6. Proof of Theorem 4.1. We will make use of Theorem 3.1 in a setup with no common noise. In particular, as explained in Remark 3.1, we have $p_0 = 0$, $W = W^1$, and M is deterministic.

By Theorem 3.1 and Remark 3.1, there exist $m \in \mathfrak{M}_+^1(\mathcal{X})$ and $\widehat{\mathbf{P}}^* \in \mathfrak{M}_+^1(\mathcal{X} \times \mathcal{Q} \times \mathcal{W})$, which maximizes $\mathbf{E}^{\mathbf{P}}[\xi + \int_0^T f_r^Q\, dr]$ within all elements $\mathbf{P} \in \mathfrak{M}_+^1(\mathcal{X} \times \mathcal{Q} \times \mathcal{W})$ satisfying Definition 2.1(1), with m replacing M, and such that $\widehat{\mathbf{P}}^* \circ X^{-1} = m$.

Let $\mathbf{P}^* := \widehat{\mathbf{P}}^* \circ (X, Q)^{-1}$. We have $m = \mathbf{P}^* \circ X^{-1}$ and $\mathbf{P}^* \in \overline{\mathcal{P}}^m$. In particular $m \in \mathcal{P}^m$, as required in Definition 4.2.

We remark that ξ, f do not depend on W. For any $\mathbf{Q} \in \overline{\mathcal{P}}^m$, there exists $\widehat{\mathbf{Q}} \in \mathfrak{M}_+^1(\mathcal{X} \times \mathcal{Q} \times \mathcal{W})$ satisfying Definition 2.1(1) and such that $\mathbf{Q} := \widehat{\mathbf{Q}} \circ (X, Q)^{-1}$, and hence such that

$$\mathbf{E}^{\mathbf{Q}}\left[\xi + \int_0^T f_r^Q \, dr\right] = \mathbf{E}^{\widehat{\mathbf{Q}}}\left[\xi + \int_0^T f_r^Q \, dr\right] \leqslant \mathbf{E}^{\widehat{\mathbf{P}}^*}\left[\xi + \int_0^T f_r^Q \, dr\right] = \mathbf{E}^{\mathbf{P}^*}\left[\xi + \int_0^T f_r^Q \, dr\right].$$

This shows that $m = \mathbf{P}^* \circ X^{-1}$ and

$$V_0^m(0) = \mathbf{E}^{\mathbf{P}^*}\left[\xi + \int_0^T f_r^Q \, dr\right] = \sup_{\mathbf{P} \in \overline{\mathcal{P}}^m} \mathbf{E}^{\mathbf{P}}\left[\xi + \int_0^T f_r^Q \, dr\right], \tag{6.1}$$

meaning that m is a solution of the MFG on the restricted canonical space $\Omega = \mathcal{X} \times \mathcal{Q}$.

We set $Y_t = V_t^m(X_{\wedge t})$, $t \in [0, T]$. By Theorem 5.1, $Y \in \mathbb{S}^2(\mathcal{P}^m)$, and there exists a process $Z \in \mathbb{H}^2(\mathcal{P}^m)$ such that the process U defined by

$$U := Y_\cdot - Y_0 + \int_0^\cdot F_r(Z_r, \widehat{\sigma}_r^2, m) \, dr - \int_0^\cdot Z_r \, dX_r \tag{6.2}$$

is a càdlàg $\mathbf{P}$-supermartingale orthogonal to X for all $\mathbf{P} \in \mathcal{P}^m$. Consider the Doob–Meyer decomposition of the m-supermartingale $U = M - K$ into an m-martingale M orthogonal to X, and an m-a.s. nondecreasing process K. We define $\overline{q}$, N^B, L, and $G(\mathbf{P}^*)$ as in the proof of Proposition 5.2. Since M is orthogonal to X, it follows that N^B can be taken orthogonal to M (see Proposition III-9 in [15]), so L is orthogonal to M. By the Girsanov theorem, M is also a $G(\mathbf{P}^*)$-martingale. Now it follows from (6.2) that (Y, Z) solves the BSDE

$$Y_t = \xi + \int_t^T F_r(Z_r, \widehat{\sigma}_r^2, m) \, dr + dK_r - Z_r \, dX_r - dM_r, \qquad t \in [0, T], \quad G(\mathbf{P}^*)\text{-a.s.}$$

with orthogonal martingale M. Since K is $G(\mathbf{P}^*)$-a.s. nondecreasing and positive, by the standard comparison result of BSDEs we have $Y_0 \geqslant \mathbf{E}^{G(\mathbf{P}^*)}[Y_0^{G(\mathbf{P}^*)}]$, where $(Y^{\mathbf{P}^*}, Z^{\mathbf{P}^*})$ is defined as in (5.5) by

$$Y_t^{\mathbf{P}^*} = \xi + \int_t^T F_r(Z_r^{\mathbf{P}^*}, \widehat{\sigma}_r^2, m) \, dr - Z_r^{\mathbf{P}^*} \, dX_r - dM_r^{\mathbf{P}^*}, \qquad t \in [0, T], \quad \mathbf{P}^*\text{-a.s.}$$

Moreover, the requirement that U is an m-martingale is equivalent to $K \equiv 0$, $G(\mathbf{P}^*)$-a.s. which in turn is equivalent to $Y_0 = \mathbf{E}^{G(\mathbf{P}^*)}[Y_0^{G(\mathbf{P}^*)}]$, which we prove. Since m satisfies (6.1), it follows from (5.7) and Proposition 5.2 that

$$Y_0 = V_0^m(0) = \mathbf{E}^{\mathbf{P}^*}\left[\xi + \int_0^T f_r^Q \, dr\right] \leqslant \mathbf{E}^{G(\mathbf{P}^*)}[Y_0^{G(\mathbf{P}^*)}], \tag{6.3}$$

and the required result follows from the fact that

$$Y_0 = V_0^m(0) = \max_{\mathbf{P} \in \overline{\mathcal{P}}^m} \mathbf{E}^{\mathbf{P}}\left[\xi + \int_0^T f_r^Q \, dr\right] = \sup_{\mathbf{P} \in \mathcal{M}^m} \mathbf{E}^{\mathbf{P}}[Y_0^{\mathbf{P}}] \geqslant \mathbf{E}^{G(\mathbf{P}^*)}[Y_0^{G(\mathbf{P}^*)}].$$

Appendix A. Basic results concerning correspondences.

Definition A.1. *Let E, F be two Hausdorff topological spaces. A mapping T from E into the subsets of F is called a correspondence from E into F, which we summarize with the notation $T\colon E \twoheadrightarrow F$.*

T is called upper hemicontinuous if, for every $x \in E$ and any neighborhood U of $T(x)$, there is a neighborhood V of x such that $z \in V$ implies $T(z) \subset U$.

T is lower hemicontinuous if, for every $x \in E$ and any open set U that meets $T(x)$, there is a neighborhood V of x such that $z \in V$ implies $T(z) \cap U \neq \varnothing$.

We say that T is continuous if it is both upper hemicontinuous and low hemicontinuous. Finally, T is said to have a closed graph if its graph $\mathrm{Gr}(T) := \{(x, y)\colon x \in E, y \in T(x)\}$ is a closed subset of $E \times F$.

We collect in the following proposition some classical results which can be found in [1, Theorems 17.10, 17.11, 17.15, 17.23; Lemma 17.8].

Proposition A.1. (1) *If T is an upper hemicontinuous correspondence with compact values, then it has a closed graph*;

(2) *conversely, if T has a closed graph and F is compact, then T is upper hemicontinuous*;

(3) *if F is a metric space and T is compact valued, then T may be seen as a function from E to $\mathrm{Comp}(F)$, the set of nonempty compact sets of F, which may be equipped with a metric called the Hausdorff metric such that T is continuous as a correspondence if and only if it is continuous as a function for that metric*;

(4) *the composition of upper hemicontinuous* (*respectively, low hemicontinuous, continuous*) *correspondences is upper hemicontinuous* (*respectively, low hemicontinuous, continuous*);

(5) *the image of a compact set under a compact valued upper hemicontinuous correspondence is compact.*

We now recall the Berge maximum theorem (see [1, Theorem 17.31]).

Theorem A.1. *Let $T\colon E \twoheadrightarrow F$ be a continuous nonempty compact valued correspondence between topological spaces. Let $J\colon F \to \mathbf{R}$ be a continuous function. Then the correspondence $T^*\colon E \twoheadrightarrow F$ defined, for all $x \in E$, by $T^*(x) := \operatorname{argmax}_{y \in T(x)} J(y)$, is upper hemicontinuous and nonempty compact valued.*

Moreover, the mapping $m\colon E \to \mathbf{R}$ given for all $x \in E$ by $m(x) := \max_{y \in T(x)} J(y)$ is continuous.

In [18], Horvath extended the ε-approximate selection theorem obtained by Cellina in [9]. Although it was stated in a framework of generalized convex structures [18, Theorem 6], the lines after its proof imply the following.

Assumption A.1. E is a subset of a locally convex topological vector space such that there exists a distance d_E metrizing the induced topology of E and such that all open balls are convex and that any neighborhood $\{y \in E\colon d_E(y, C) < r\}$ of a convex set C is convex.

Theorem A.2. *Let (K, d_K) be a compact metric space, and let (E, d_E) verify Assumption* A.1. *We denote by d the distance $d_K + d_E$ on $K \times E$.*

Let T be an upper hemicontinuous correspondence taking nonempty compact convex values from K to E. Then, for any $\varepsilon > 0$, there exists a continuous function $f_\varepsilon\colon K \to E$ such that, for all $x \in K$,

$$d((x, f_\varepsilon(x)), \mathrm{Gr}(T)) := \inf\{d((x, f_\varepsilon(x)), (y, z))\colon y \in E,\ z \in T(y)\} < \varepsilon.$$

The following theorem is a generalization of Kakutani's theorem adapted from Proposition 7.4 in [8] which itself adapts a result of Cellina; see [9, Theorem 1].

THEOREM A.3. *Let* (K, d) *be a compact convex subset of a locally convex topological vector space* (E, d_E) *verifying Assumption* A.1, *let* T *be an upper hemicontinuous correspondence taking nonempty compact convex values from* K *to* E, *and let* ϕ *be a continuous function from* E *to* K.

Then there exists some $x \in K$ *such that* $x \in \phi \circ T(x)$.

Proof. Let $\mathrm{Gr}(T) := \{(x, y) \in K \times E : y \in T(x)\}$. By Theorem A.2, for every $n \in \mathbf{N}$, there exists a continuous $f_n : K \to E$ such that, for all $x \in K$,

$$\inf\{d((x, f_n(x)), \mathrm{Gr}(T)\} < \frac{1}{n}.$$

Since $\phi \circ f_n : K \to K$ is continuous, by Schauder's fixed point theorem there exists some $x_n \in K$ such that $x_n = \phi(f_n(x_n))$. By Proposition A.1(1), (5), since T is upper hemicontinuous and compact valued, it follows that $T(K) := \bigcup_{x\in K} T(x)$ is compact and $\mathrm{Gr}(T)$ is closed. Thus $\mathrm{Gr}(T) \subset K \times T(K)$ is compact. Since $d((x_n, f_n(x_n)), \mathrm{Gr}(T)) \to 0$ and $\mathrm{Gr}(T)$ is compact, there exists a subsequence x_{n_k} and a point $(x, y) \in \mathrm{Gr}(T)$ such that $(x_{n_k}, f_{n_k}(x_{n_k})) \to (x, y)$. Now by continuity of ϕ we have

$$x = \lim x_{n_k} = \lim \phi(f_{n_k}(x_{n_k})) = \phi(y),$$

with $y \in T(x)$, so the proof is complete.

LEMMA A.1. *Let* S *be a Polish space and* E *be a convex subset of* $\mathfrak{M}^1_+(S)$ *equipped with the topology of weak convergence. Then there exists on* E *a distance* d_E *such that* (E, d_E) *verifies Assumption* A.1.

In particular, *Theorem* A.3 *applies for such a choice of space* E.

Proof. It is immediate that in a normed space the distance induced by the norm satisfies Assumption A.1. This implies that if we consider a convex subset E of a normed space $(F, \|\cdot\|)$ and equip E with the distance d_E defined by $d_E(x, y) := \|x - y\|$, then (E, d_E) verifies Assumption A.1.

We now recall that $\mathfrak{M}^1_+(S)$ is a convex subset of the vector space $\mathfrak{M}(S)$ which can be equipped with the Kantorovic–Rubinshtein norm (see [4, section 8.3] for an introduction) and that on $\mathfrak{M}^1_+(S)$ this norm induces the topology of weak convergence; see [4, Theorem 8.3.2]. This concludes the proof.

REFERENCES

[1] C. D. ALIPRANTIS AND K. C. BORDER, *Infinite Dimensional Analysis*: *A Hitchhiker's Guide*, 2nd ed., Springer-Verlag, Berlin, 1999, https://doi.org/10.1007/978-3-662-03961-8.
[2] A. BARRASSO AND F. RUSSO, *Path-dependent martingale problems and additive functionals*, Stoch. Dyn., 19 (2019), 1950027, https://doi.org/10.1142/S0219493719500278.
[3] P. BILLINGSLEY, *Probability and Measure*, 2nd ed., Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., John Wiley & Sons, New York, 1986.
[4] V. I. BOGACHEV, *Measure Theory*, Vols. I & II, Springer-Verlag, Berlin, 2007, https://doi.org/10.1007/978-3-540-34514-5.
[5] R. CARMONA AND F. DELARUE, *Probabilistic Theory of Mean Field Games with Applications*, *Vol.* 1: *Mean Field FBSDEs, Control, and Games*, Probab. Theory Stoch. Model. 83, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-58920-6.
[6] R. CARMONA AND F. DELARUE, *Probabilistic Theory of Mean Field Games with Applications*, *Vol.* 2: *Mean Field Games with Common Noise and Master Equations*, Probab. Theory Stoch. Model. 84, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-56436-4.

[7] R. Carmona, F. Delarue, and D. Lacker, *Mean field games with common noise*, Ann. Probab., 44 (2016), pp. 3740–3803, https://doi.org/10.1214/15-AOP1060.

[8] R. Carmona and D. Lacker, *A probabilistic weak formulation of mean field games and applications*, Ann. Appl. Probab., 25 (2015), pp. 1189–1231, https://doi.org/10.1214/14-AAP1020.

[9] A. Cellina, *Approximation of set valued functions and fixed point theorems*, Ann. Mat. Pura Appl. (4), 82 (1969), pp. 17–24, https://doi.org/10.1007/BF02410784.

[10] C. Dellacherie and P.-A. Meyer, *Probabilités et Potentiel, Chapitres* I *à* IV. *Édition entièrement refondue*, Publ. Inst. Math. Univ. Strasbourg XV, Actualités Sci. Indust., No. 1372, Hermann, Paris, 1975.

[11] L. Q. Eifler, *Open mapping theorems for probability measures on metric spaces*, Pacific J. Math., 66 (1976), pp. 89–97, https://doi.org/10.2140/pjm.1976.66.89.

[12] I. Ekren, C. Keller, N. Touzi, and J. Zhang, *On viscosity solutions of path dependent PDEs*, Ann. Probab., 42 (2014), pp. 204–236, https://doi.org/10.1214/12-AOP788.

[13] I. Ekren, N. Touzi, and J. Zhang, *Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part* I, Ann. Probab., 44 (2016), pp. 1212–1253, https://doi.org/10.1214/14-AOP999.

[14] N. El Karoui, S. Peng, and M. C. Quenez, *Backward stochastic differential equations in finance*, Math. Finance, 7 (1997), pp. 1–71, https://doi.org/10.1111/1467-9965.00022.

[15] N. El Karoui and S. Méléard, *Martingale measures and stochastic calculus*, Probab. Theory Related Fields, 84 (1990), pp. 83–101, https://doi.org/10.1007/BF01288560.

[16] U. G. Haussmann, *Existence of optimal Markovian controls for degenerate diffusions*, in Stochastic Differential Systems (Bad Honnef, 1985), Lect. Notes Control Inf. Sci. 78, Springer, Berlin, 1986, pp. 171–186, https://doi.org/10.1007/BFb0041161.

[17] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Grundlehren Text Ed., Springer-Verlag, Berlin, 2001, https://doi.org/10.1007/978-3-642-56468-0.

[18] Ch. D. Horvath, *Extension and selection theorems in topological spaces with a generalized convexity structure*, Ann. Fac. Sci. Toulouse Math. (6), 2 (1993), pp. 253–269, https://doi.org/https://doi.org/10.5802/afst.766.

[19] M. Huang, R. P. Malhamé, and P. E. Caines, *Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst., 6 (2006), pp. 221–252, https://doi.org/10.4310/CIS.2006.v6.n3.a5.

[20] J. Jacod, *Calcul Stochastique et Problèmes de Martingales*, Lecture Notes in Math. 714, Springer, Berlin, 1979, https://doi.org/10.1007/BFb0064907.

[21] O. Kallenberg, *Random Measures, Theory and Applications*, Probab. Theory Stoch. Model. 77, Springer, Cham, 2017, https://doi.org/10.1007/978-3-319-41598-7.

[22] R. L. Karandikar, *On pathwise stochastic integration*, Stochastic Process. Appl., 57 (1995), pp. 11–18, https://doi.org/10.1016/0304-4149(95)00002-O.

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

[24] D. Lacker, *Dense Sets of Joint Distributions Appearing in Filtration Enlargements, Stochastic Control, and Causal Optimal Transport*, preprint, https://arxiv.org/abs/1805.03185v1, 2018.

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

[26] Y. Lin, Z. Ren, N. Touzi, and J. Yang, *Second order backward SDE with random terminal time*, Electron. J. Probab., 25 (2020), 99, https://doi.org/10.1214/20-EJP498; preprint version available at https://arxiv.org/abs/1802.02260, 2018.

[27] N. El Karoui, N. Du'hŪŪ, and M. Jeanblanc-Picqué, *Compactification methods in the control of degenerate diffusions: Existence of an optimal control*, Stochastics, 20 (1987), pp. 169–219, https://doi.org/10.1080/17442508708833443.

[28] M. Nutz, *Pathwise construction of stochastic integrals*, Electron. Commun. Probab., 17 (2012), 24, https://doi.org/10.1214/ECP.v17-2099.

[29] D. Possamaï, X. Tan, and C. Zhou, *Stochastic control for a class of nonlinear kernels and applications*, Ann. Probab., 46 (2018), pp. 551–603, https://doi.org/10.1214/17-AOP1191.

[30] H. M. Soner, N. Touzi, and J. Zhang, *Wellposedness of second order backward SDEs*, Probab. Theory Related Fields, 153 (2012), pp. 149–190, https://doi.org/10.1007/s00440-011-0342-y.

[31] H. M. Soner, N. Touzi, and J. Zhang, *Dual formulation of second order target problems*, Ann. Appl. Probab., 23 (2013), pp. 308–347, https://doi.org/10.1214/12-AAP844.

[32] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, reprint of the 1997 ed., Classics Math., Springer-Verlag, Berlin, 2006, https://doi.org/10.1007/3-540-28999-2.