Mean Field Games with Singular Controls

Guanxing Fu (Humboldt University Berlin)
Ulrich Horst (Humboldt University Berlin)

Discussion Paper No. 22

March 23, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Mean Field Games with Singular Controls*

Guanxing Fu[†] and Ulrich Horst[‡]

March 21, 2017

Abstract

This paper establishes the existence of relaxed solutions to mean field games (MFGs for short) with singular controls. As a by-product, we obtain an existence of relaxed solutions results for McKean-Vlasov stochastic singular control problems. Finally, we prove approximations of solutions results for a particular class of MFGs with singular controls by solutions, respectively control rules, for MFGs with purely regular controls. Our existence and approximation results strongly hinge on the use of the Skorokhod M_1 topology on the space of càdlàg functions.

AMS Subject Classification: 93E20, 91B70, 60H30.

Keywords: mean field game, singular control, relaxed control, Skorokhod M_1 topology.

1 Introduction and overview

Starting with the seminal papers [25, 33], the analysis of *mean field games* (MFGs) has received considerable attention in the stochastic control and financial mathematics literature. In a standard MFG, each player $i \in \{1, ..., N\}$ chooses an action from a given set of admissible controls that maximizes a cost functional of the form

$$J^i(u) = E\left[\int_0^T f(t, X_t^i, \bar{\mu}_t^N, u_t^i)dt + g(X_T^i, \bar{\mu}_T^N)\right] \tag{1.1}$$

subject to the state dynamics

$$\begin{cases} dX_t^i = b(t, X_t^i, \bar{\mu}_t^N, u_t^i)\, dt + \sigma(t, X_t^i, \bar{\mu}_t^N, u_t^i)\, dW_t^i, \\ X_0^i = x_0 \end{cases} . \tag{1.2}$$

Here $W^1, ..., W^N$ are independent Brownian motions defined on some underlying filtered probability space, $u = (u^1, \cdots, u^N)$, $u^i = (u_t^i)_{t\in[0,T]}$ is an adapted stochastic process, the *action* of player i, and $\bar{\mu}_t^N := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}$ denotes the empirical distribution of the individual players' states at time $t \in [0, T]$. In particular, all players are identical ex ante and each player interacts with the other players only through the empirical distribution of the state processes.

*Financial support by the Berlin Mathematical School (BMS), the SFB 649 *Economic Risk* and the TRR 190 *Rationality and Competition: the Economic Performance of Individuals and Firms* is gratefully acknowledged.

†Department of Mathematics, and Berlin Mathematical School, Humboldt-Universität zu Berlin Unter den Linden 6, 10099 Berlin, Germany; email: fuguanxing725@gmail.com

‡Department of Mathematics, and School of Business and Economics, Humboldt-Universität zu Berlin Unter den Linden 6, 10099 Berlin, Germany; email: horst@math.hu-berlin.de

The existence of *approximate Nash equilibria* in the above game for large populations has been established in [4, 25] using a representative agent approach. In view of the independence of the Brownian motions the idea is to first approximate the dynamics of the empirical distribution by a deterministic measure-valued process, and to consider instead the optimization problem of a representative player that takes the distribution of the states as given, and then to solve the fixed-point problem of finding a measure-valued process such that the distribution $\mathcal{L}(X)$ of the representative player's state process X under her optimal strategy coincides with that process.[1]

Following the representative agent approach, a MFG can then be formally described by a coupled control and fixed point problem of the form:

$$\begin{cases} \inf_u E\left[\int_0^T f(t, X_t, \mu_t, u_t)\, dt + g(X_T, \mu_T)\right], \\ \text{subject to} \\ dX_t = b(t, X_t, \mu_t, u_t)\, dt + \sigma(t, X_t, \mu_t, u_t)\, dW_t \\ X_0 = x_0, \\ \mathcal{L}(X) = \mu. \end{cases} \tag{1.3}$$

There are essentially three approaches to solve mean field games. In their original paper [33], Lasry and Lions followed an analytic approach. They analyzed a coupled forward-backward PDE system, where the backward component is the Hamiltion-Jacobi-Bellman equation arising from the representative agent's optimization problem, and the forward component is a Kolmogorov-Fokker-Planck equation that characterizes the dynamics of the state process.

A second, more probabilistic, approach was introduced by Carmona and Delarue in [4]. Using a maximum principle of Pontryagin type, they showed that the fixed point problem reduces to solving a McKean-Vlasov forward-backward SDEs (FBSDEs for short). Other results based on probabilistic approaches include [1, 2, 3, 5, 10]. Among them, [2, 3, 5] consider linear-quadratic MFGs, while [1, 10] consider MFGs with common noise and with major and minor players, respectively. A class of MFGs in which the interaction takes place both through the state dynamics and the controls has recently been introduced in [8]. In that paper the weak formulation, or martingale optimality principle, is used to prove the existence of a solution.

A *relaxed solution* concept to MFGs was introduced by Lacker in [31]. Considering MFGs from a more game-theoretic perspective, the idea is to search for equilibria in relaxed controls ("mixed strategies") by first establishing the upper hemi-continuity of the representative agent's best response correspondence to a given μ using Berge's maximum theorem, and then to apply the Kakutani-Fan-Glicksberg fixed point theorem in order to establish the existence of some measure-valued process μ^* such that the law of the agent's state process under a best response to μ^* coincides with that process.

Relaxed controls date back to Young [38]. They were later applied to stochastic control in, e.g. [19, 20, 29], to MFGs in [31], and to MFGs with common noise in [6]. In the common noise case, the fixed point is random. This prevents an application of standard fixed-point results. To overcome the problem of randomness, [6] introduced a notion of *weak solution* to MFGs, based on the approximation of MFGs by MFGs with discretized common noise paths. The notion of weak solutions is further supported in [32] where it is shown that the weak limit of ϵ-Nash equilibria for N player games as $N \to \infty$ is a weak solution to MFGs. Moreover, each weak solution to MFGs yields an ϵ-Nash equilibrium for the N player game.

Applications of MFGs range from models of optimal exploitation of exhaustible resources [11, 12, 18]

[1] The idea of decoupling local from global dynamic in large interactive stochastic systems goes back at least to Föllmer [14], see also [15], and has been successfully applied to equilibrium models of social interaction in e.g. [22, 23].

to systemic risk and bank-run models [7, 9, 34], and from principal-agent problems [13] to problems of optimal trading under market impact [8, 16, 26]. Motivated by possible applications to optimal portfolio liquidation under strategic interaction that allow for both block trades and absolutely continuous trades as in [17, 21, 30], this paper provides a probabilistic framework for analyzing MFGs with singular controls. Extending [31], we consider MFGs with singular controls of the form

$$\begin{cases} \inf_{u,Z} E\left[\int_0^T f(t,X_t,\mu_t,u_t)\,dt + g(X_T,\mu_T) + \int_0^T h(t)\,dZ_t\right], \\ \text{subject to} \\ dX_t = b(t,X_t,\mu_t,u_t)\,dt + \sigma(t,X_t,\mu_t,u_t)\,dW_t + c(t)\,dZ_t, \\ \mathcal{L}(X) = \mu \end{cases} \tag{1.4}$$

where $u = (u_t)_{t\in[0,T]}$ is the *regular control*, and $Z = (Z_t)_{t\in[0,t]}$ is the *singular control*. When singular controls are admissible, the state process no longer takes values in the space of continuous functions, but rather in the Skorokhod space $\mathcal{D}(0,T)$ of all càdlàg functions. The key is then to identify a suitable topology on the Skorokhod space with respect to which the compactness and continuity assumptions of the maximum and the fixed-point theorems are satisfied.

There are essentially three possible topologies on the space of càdlàg functions: the (standard) Skorokhod J_1 topology (J_1 topology for short), the Meyer-Zheng topology (or pseudo-path topology), and the Skorokhod M_1 topology (M_1 topology for short). The M_1 topology seems to be the most appropriate one for our purposes. First, the set of bounded singular controls is compact in the M_1 topology but not in the J_1 topology. Second, there is no explicit expression for the metric corresponding to Meyer-Zheng topology. In particular, one cannot bound the value function at given points in time in the Meyer-Zheng topology. Third, the M_1 topology has better continuity properties than the J_1 topology. For instance, it allows for an approximation of discontinuous functions by continuous ones. This enables us to approximate solutions to certain classes of MFGs with singular controls by solutions to MFGs with only regular controls. Appendix B summarizes useful properties of the M_1 topology; for more details, we refer to the textbook of Whit [37].

To the best of our knowledge, ours is the first paper to establish the existence of solutions results to MFGs with singular controls. As a byproduct, we obtain a new proof for the existence of optimal (relaxed) controls for the corresponding class of stochastic singular control problems. A similar control problem, albeit with a trivial terminal cost function has been analyzed in [20]. While the methods and techniques applied therein can be extended to non-trivial terminal cost functions after a modification of the control problem, they cannot be used to prove existence of equilibria in MFGs. In fact, in [20], it is assumed that the state space $\mathcal{D}(0,T)$ is endowed with Meyer-Zheng topology, and that the spaces of admissible singular and regular controls are endowed with the topology of weak convergence and the stable topology, respectively. With this choice of topologies the continuity of cost functional and the upper-hemicontinuity of distribution of the representative agent's state process under the optimal control w.r.t. to a given process μ cannot be established. As a second byproduct we obtain a novel existence of solutions result for a class of McKean-Vlasov stochastic singular control problems. MFGs and control problems of McKean-Vlasov type are compared in [5]. The main difference between these somewhat similar, yet very different problems lies in the order of carrying out the optimization and the fixed point arguments. When optimizing first, the subsequent fixed point problem leads to MFGs, while in McKean-Vlasov control problems one searches for fixed points before solving the optimization problem.

Our second main contributions are two approximation results that allows us to approximate solutions to a certain class of MFGs with singular controls by the solutions to MFGs with only regular controls. The approximation result, too, strongly hinges on the choice of the M_1 topology.

The rest of this paper is organized as follows: in Section 2, we recall the notion of relaxed control for singular stochastic control problems, introduce MFGs with singular controls and state our main existence of solutions result. The proof is given in Section 3. In Section 4, we state and prove two approximation results for MFGs with singular controls by MFGs with regular controls. Appendix A recalls known results and definitions that are used throughout this paper. Append B reviews key properties of the M_1 topology.

2 Assumptions and the main results

In this section we introduce MFGs with singular controls and state our main existence of solutions result. For a given interval $\mathbb{I}$ we denote by $\mathcal{C}(\mathbb{I})$ the space of all $\mathbb{R}^d$-valued continuous functions on $\mathbb{I}$, by $\mathcal{D}(\mathbb{I})$ the Skorokhod space of all $\mathbb{R}^d$-valued càdlàg functions on $\mathbb{I}$, and by $\mathcal{A}(\mathbb{I}) \subset \mathcal{D}(\mathbb{I})$ the subset of nondecreasing functions. For reasons that will become clear later we identify processes on $[0,T]$ with processes on the whole real line. For instance, we identify the space $\mathcal{D}(0,T)$ with the space

$$\widetilde{\mathcal{D}}_{0,T}(\mathbb{R}) = \{x \in \mathcal{D}(\mathbb{R}) : x|_{[0,T]} \in \mathcal{D}\left(0,T\right), x_t = 0 \text{ if } t < 0 \text{ and } x_t = x_T \text{ if } t > T\}.$$

Likewise, we denote by $\widetilde{\mathcal{A}}_{0,T}(\mathbb{R})$ the subspace of $\widetilde{\mathcal{D}}_{0,T}(\mathbb{R})$ with nondecreasing paths. We occasionally drop the subscripts 0 and T if there is no risk of confusion. For a metric space (E,d) we denote by $\mathcal{P}_p(E)$ the class of all probability measures on E with finite moment of p-th order. For $p=0$ we write $\mathcal{P}(E)$ instead of $\mathcal{P}_0(E)$. By $\mathcal{W}_{p,(E,d)}$ we denote the Wasserstein distance on $\mathcal{P}_p(E)$; see Definition A.4. To save notation the set $\mathcal{P}_p(E)$ endowed with the Wasserstein distance is often denoted $\mathcal{W}_{p,(E,d)}$ or $\mathcal{W}_p$ if there is no risk of confusion about the underlying state space.

2.1 Singular stochastic control problems

Before introducing MFGs with singular controls, we briefly review stochastic singular control problems of the form

$$\begin{cases} \inf_{u,Z} E\left[\int_0^T f(t,X_t,u_t)\,dt + g(X_T) + \int_0^T h(t)\,dZ_t\right], \\ \text{subject to} \\ dX_t = b(t,X_t,u_t)\,dt + \sigma(t,X_t,u_t)\,dW_t + c(t)\,dZ_t, \\ X_{0-} = 0, \end{cases} \tag{2.1}$$

where the *regular control* $u = (u_t)_{t\in[0,T]}$ takes values in a compact set U, and that the *singular control* $Z = (Z_t)_{t\in[0,T]}$ takes values in $\widetilde{\mathcal{A}}(\mathbb{R})$. The existence of optimal *relaxed controls* to stochastic singular control problems has been addressed in [20] using the so-called compactification method. We use a similar approach to solve MFGs with singular controls, albeit in different topological setting.

The following notion of relaxed controls follows [20].

Definition 2.1. The tuple $r = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, q, Z)$ is called a relaxed control if

1. $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P})$ is a filtered probability space;
2. $\mathbb{P}(X_t = 0, Z_t = 0 \text{ if } t < 0, X_t = X_T, Z_t = Z_T \text{ if } t > T \text{ and } q_0 = \delta_{u_0}) = 1$, for some $u_0 \in U$;
3. q is a $\mathcal{P}(U)$ valued and $\{\mathcal{F}_t, t \geq 0\}$ progressively measurable stochastic process, and Z is a $\widetilde{\mathcal{A}}(\mathbb{R})$-valued and $\{\mathcal{F}_t, t \geq 0\}$ progressively measurable stochastic process;
4. X is a $\{\mathcal{F}_t, t \geq 0\}$ adapted stochastic process with path in $\widetilde{D}(\mathbb{R})$ such that for each $\phi \in \mathcal{C}_b^2(\mathbb{R}^d)$, the space of all continuous and bounded functions with continuous and bounded first- and second-order

derivatives, $\mathcal{M}^\phi$ is a $\mathbb{P}$ continuous martingale, where

$$\mathcal{M}_t^\phi := \phi(X_t) - \int_0^t \int_U \mathcal{L}\phi(s, X_s, u)\, q_s(du)ds - \int_0^t (\partial_x \phi(X_{s-}))^\top c(s) dZ_s \\ - \sum_{0 \le s \le t} \big(\phi(X_s) - \phi(X_{s-}) - (\partial_x \phi(X_{s-}))^\top \triangle X_s\big),$$

with $\mathcal{L}\phi(t,x,u) := \frac{1}{2}\sum_{ij} a_{ij}(t,x,u)\frac{\partial^2 \phi(x)}{\partial_{x_i}\partial_{x_j}} + \sum_i b_i(t,x,u)\partial_{x_i}\phi(x)$ and $a(t,x,u) = \sigma^\top\sigma(t,x,u)$.

With some abuse of notation we sometimes write $Z \in \widetilde{\mathcal{A}}(\mathbb{R})$ to indicate that Z is a progressively measurable stochastic process taking values in $\widetilde{\mathcal{A}}(\mathbb{R})$ and call the process q the relaxed control; the control u will be referred to as the *strict control*. The class of strict controls can be embedded into the class of relaxed controls via the mapping $u \mapsto \delta_u$. The cost functional corresponding to a relaxed control r is defined by

$$\widetilde{J}(r) = E^{\mathbb{P}}\left[\int_0^T \int_U f(t, X_t, u)\, q_t(du)dt + \int_0^T h(t)\, dZ_t + g(X_T)\right]. \tag{2.2}$$

Remark 2.2. For $(t,x) \in [0,T] \times \mathbb{R}^d$, let

$$K(t,x) = \{(a(t,x,u), b(t,x,u), e) :\ e \ge f(t,x,u),\ u \in U\}.$$

If $K(t,x)$ is convex for each $(t,x) \in [0,T] \times \mathbb{R}^d$, then it can be shown that for each relaxed control r, there exists a strict control u and a singular control Z with equal cost; see [19, Theoerem 3.6] for details.

In what follows, we restrict ourselves to relaxed controls and always assume that Ω is the canonical path space, and that the σ-algebra $\{\mathcal{F}_t, t \ge 0\}$ is generated by the corresponding coordinate processes X, q and Z. More precisely, from now on we restrict ourselves to relaxed controls with

$$\Omega = \widetilde{\mathcal{D}}(\mathbb{R}) \times \mathcal{U}(0,T) \times \widetilde{\mathcal{A}}(\mathbb{R})$$

where $\mathcal{U}(0,T)$ denotes the set of all finite measures on $[0,T] \times U$ whose first marginal is the Lebesgue measure on $[0,T]$ and whose the second marginal belongs to $\mathcal{P}(U)$, and assume that for each $\omega := (x,q,z) \in \Omega$,

$$X(\omega) = x, \quad q(\omega) = q \quad \text{and} \quad Z(\omega) = z.$$

Definition 2.3. Let $r = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \ge 0\}, \mathbb{P}, X, q, Z)$ be a relaxed control where Ω is the canonical path space and (X,q,Z) are the coordinate processes. Then, the probability measure $\mathbb{P}$ is called the *control rule*. The associated cost functional is defined as

$$\widehat{J}(\mathbb{P}) := \widetilde{J}(r).$$

Let us denote by $\mathcal{R}$ the class of all the control rules for the stochastic control problem (2.1). Clearly,

$$\inf_{\mathbb{P} \in \mathcal{R}} \widehat{J}(\mathbb{P}) \ge \inf_{\text{relaxed control } r} \widetilde{J}(r).$$

Conversely, for any relaxed control r one can construct a control rule $\mathbb{P} \in \mathcal{R}$ such that $\widehat{J}(\mathbb{P}) = \widetilde{J}(r)$. The proof is standard; it can be found in, e.g. [20, Proposition 2.6]. In other words, the optimization problems with relaxed controls and control rules are equivalent. It is hence enough to consider control rules.

Remark 2.4. In [20], it is assumed that the space $\mathcal{A}(0,T)$ of singular controls is endowed with the topology of weak convergence, that the state space $\mathcal{D}(0,T)$ is endowed with Meyer-Zheng topology, and that the space of relaxed controls $\mathcal{U}(0,T)$ is endowed with the stable topology. Note that in [20], $\mathcal{D}(0,T)$ and $\mathcal{A}(0,T)$ are càglàd path spaces rather than càdlàg. For this choice of topologies and under suitable assumptions on the cost coefficients it is then shown that an optimal control rule exists if $g \equiv 0$. Their method allows for terminal costs only after a modification of the cost function; see [20, Remark 2.2 and Section 4] for details. As a byproduct (see Proposition 3.9) of our analysis of MFGs, under the same assumptions on the cost coefficients as in [20] we establish the existence of an optimal control rule for terminal cost functions that satisfy linear growth condition. Moreover, we generalize the stochastic singular control problem to problems of McKean-Vlasov-type (see Theorem 3.17).

2.2 Mean field games with singular controls

We are now going to consider MFGs with singular controls of the form (1.4). We again restrict ourselves to relaxed controls. The first step of solving mean field games is to solve the representative agent's optimal control problem

$$\begin{cases} \inf_{u,Z} E\left[\int_0^T f(t,X_t,\mu_t,u_t)\,dt + g(X_T,\mu_T) + \int_0^T h(t)\,dZ_t\right] \\ \text{subject to} \\ dX_t = b(t,X_t,\mu_t,u_t)\,dt + \sigma(t,X_t,\mu_t,u_t)\,dW_t + c(t)\,dZ_t, \\ X_{0-} = 0 \end{cases}$$

for any *fixed* mean field measure μ. As in the stochastic control framework, the canonical path space for MFGs with singular controls is

$$\Omega := \widetilde{\mathcal{D}}(\mathbb{R}) \times \mathcal{U}(0,T) \times \widetilde{\mathcal{A}}(\mathbb{R}).$$

We assume that the spaces $\widetilde{\mathcal{D}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}(\mathbb{R})$ are endowed with the M_1 topology, and that the space $\mathcal{U}(0,T)$ is endowed with the topology induced by the Wasserstein distance

$$\mathcal{W}_{p,[0,T]\times U}(q^1,q^2) := \frac{1}{T}\inf\left\{\left(\int_{([0,T]\times U)^2} |\kappa_1-\kappa_2|^p\,\gamma(d\kappa_1,d\kappa_2)\right)^{\frac{1}{p}} : \text{the marginals of } \gamma \text{ are } q^1 \text{ and } q^2\right\}.$$

It is well known [37, Chapter 3] that the spaces $\widetilde{\mathcal{D}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}(\mathbb{R})$ are Polish spaces when endowed with the M_1 topology, and that the σ-algebras on $\widetilde{\mathcal{D}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}(\mathbb{R})$ coincide with the Kolmogorov σ-algebras generated by the coordinate projections. Moreover, $(\mathcal{U}(0,T), \mathcal{W}_{p,[0,T]\times U})$ is a separable metric space.

Definition 2.5. A probability measure $\mathbb{P}$ is called a control rule with respect to μ if

1. $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P})$ is the canonical probability space and (X,q,Z) are the coordinate mappings on it;
2. $\mathbb{P}(X_t = 0, Z_t = 0 \text{ if } t<0, X_t = X_T, Z_t = Z_T \text{ if } t > T \text{ and } q_0 = \delta_{u_0}) = 1$ for some $u_0 \in U$;
3. q is a $\mathcal{P}_p(U)$ valued and $\{\mathcal{F}_t, t\geq 0\}$ progressively measurable stochastic process. Z is a $\{\mathcal{F}_t, t \geq 0\}$ progressively measurable stochastic process valued in $\widetilde{\mathcal{A}}(\mathbb{R})$;
4. X is a $\{\mathcal{F}_t, t\geq 0\}$ adapted stochastic process with path in $\widetilde{\mathcal{D}}(\mathbb{R})$ such that for each $\phi \in \mathcal{C}_b^2(\mathbb{R}^d)$, $\mathcal{M}^{\mu,\phi}$ is a $\mathbb{P}$ continuous martingale, where

$$\begin{aligned}\mathcal{M}_t^{\mu,\phi} := \phi(X_t) - \int_0^t\int_U \mathcal{L}\phi(s,X_s,\mu_s,u)\,q_s(du)ds - \int_0^t (\partial_x\phi(X_{s-}))^\top c(s)dZ_s \\ - \sum_{0\leq s\leq t}\big(\phi(X_s)-\phi(X_{s-})-(\partial_x\phi(X_{s-}))^\top \triangle X_s\big),\end{aligned} \tag{2.3}$$

with $\mathcal{L}\phi(t,x,\mu,u) := \frac{1}{2}\sum_{ij} a_{ij}(t,x,\mu,u)\frac{\partial^2\phi(x)}{\partial_{x_i}\partial_{x_j}} + \sum_i b_i(t,x,\mu,u)\partial_{x_i}\phi(x)$ and $a(t,x,\mu,u) = \sigma^\top\sigma(t,x,\mu,u)$.

For a fixed measure-valued process μ, the corresponding set of control rules is denoted $\mathcal{R}(\mu)$, the cost functional corresponding to a control rule $\mathbb{P} \in \mathcal{R}(\mu)$ is

$$J(\mu,\mathbb{P}) = E^{\mathbb{P}}\left[\int_0^T \int_U f(t,X_t,\mu_t,u)\, q_t(du)dt + \int_0^T h(t)\, dZ_t + g(X_T,\mu_T)\right],$$

and the (possibly empty) set of optimal control rules is denoted by

$$\mathcal{R}^*(\mu) := \operatorname{argmin}_{\mathbb{P}\in\mathcal{R}(\mu)} J(\mu,\mathbb{P}).$$

If a probability measure $\mathbb{P}$ satisfies the fixed point property

$$\mathbb{P} \in \mathcal{R}^*(\mathbb{P}\circ X^{-1}),$$

then we call $\mathbb{P}$ or $\mathbb{P}\circ X^{-1}$ a *relaxed solution* to the MFG with singular controls (1.4). The following theorem gives sufficient conditions for the existence of a relaxed solution to our MFG. The proof is given in Section 3.

Theorem 2.6. *For some $\bar{p} > p \geq 1$, we assume that the following conditions are satisfied:*

$\mathcal{A}_1$. *There exist a positive constant C_1 such that $|b| \leq C_1$ and $|a| \leq C_1$; moreover, b and σ are measurable in t and are Lipschitz continuous in x, uniformly in t, u and μ.*

$\mathcal{A}_2$. *The functions f and g are measurable in t and are continuous with respect to x, u and μ, uniformly in t.*

$\mathcal{A}_3$. *For each $(t,x,\mu,u) \in [0,T]\times\mathbb{R}^d\times\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))\times U$, there exist strictly positive constants C_2, C_3 and a positive constant C_4 such that*

$$-C_2\left(1 - |x|^{\bar{p}} + \int_{\mathbb{R}^d} |x|^p\,\mu(dx)\right) \leq g(x,\mu) \leq C_3\left(1 + |x|^{\bar{p}} + \int_{\mathbb{R}^d} |x|^p\,\mu(dx)\right),$$

and

$$|f(t,x,\mu,u)| \leq C_4\left(1 + |x|^p + |u|^p + \int_{\mathbb{R}^d} |x|^p\,\mu(dx)\right).$$

$\mathcal{A}_4$. *The functions c and h are continuous and c is strictly positive.*

$\mathcal{A}_5$. *The functions b, σ and f are locally Lipschitz continuous with μ in p-th Wasserstein metric, uniformly in (t,x,u), i.e., for $\varphi = b$, σ and f,*

$$|\varphi(t,x,\mu^1,u) - \varphi(t,x,\mu^2,u)| \leq C\Big(1 + L(\mathcal{W}_p(\mu^1,\delta_0),\mathcal{W}_p(\mu^2,\delta_0))\Big)\mathcal{W}_p(\mu^1,\mu^2),$$

where $L(\mathcal{W}_p(\mu^1,\delta_0),\mathcal{W}_p(\mu^2,\delta_0))$ is locally bounded with $\mathcal{W}_p(\mu^1,\delta_0)$ and $\mathcal{W}_p(\mu^2,\delta_0)$.

$\mathcal{A}_6$. *U is a compact subspace of a Polish space.*

Under assumptions $\mathcal{A}_1$-$\mathcal{A}_6$, there exist a relaxed solution to the MFGs with singular controls (1.4).

Remark 2.7. A typical example where assumption $\mathcal{A}_3$ holds is

$$g(x,\mu) = |x|^{\bar{p}} + \bar{g}(\mu),$$

where $|\bar{g}(\mu)| \leq \int_{\mathbb{R}^d} |y|^p\,\mu(dy)$. This assumption is not needed under a finite fuel constraint on the singular controls. It is needed in order to approximate MFGs with singular controls by MFGs with a finite fuel constraint. The assumption that $c > 0$ is also only needed when passing from finite fuel constrained to unconstrained problems, see Lemma 3.12. Assumption $\mathcal{A}_5$ is needed in order to prove the continuity of the cost function and the correspondence $\mathcal{R}$ in μ. A typical example for $\mathcal{A}_5$ is $\int |x|^p\mu(dx)$ or $\int |x|^p\mu(dx)\wedge K$ for some fixed constant K if boundedness is required.

3 Proof of the main result

The proof of Theorem 2.6 is split into two parts. In Section 3.1 we prove the existence of a solution to our MFG under a finite fuel constraint on the singular controls. The general case is established in Section 3.2 using an approximation argument.

3.1 Existence under a finite fuel constraint

In this section, we prove the existence of a relaxed solution to our MFG under a finite fuel constraint. That is, we restrict the set of admissible singular controls to the set

$$\widetilde{\mathcal{A}}^m(\mathbb{R}) = \{z \in \widetilde{\mathcal{A}}(\mathbb{R}) : z_T \leq m\},$$

for some $m > 0$. By Corollary B.5, the set $\widetilde{\mathcal{A}}^m(\mathbb{R})$ is $(\widetilde{\mathcal{D}}(\mathbb{R}), d_{M_1})$ compact.

We start with the following auxiliary result on the tightness of the distributions of the solutions to a certain class of SDEs.

Proposition 3.1. For each $n \in \mathbb{N}$, on a probability space $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$, let X^n satisfy the following SDE on $[0, T]$:

$$dX^n_t = b_n(t)\,dt + \sigma_n(t)\,dM^n_t + dc_n(t), \tag{3.1}$$

where the random coefficients b_n and σ_n are bounded uniformly in n, M^n is a continuous martingale with bounded quadratic variation, uniformly in n and c_n is monotone, uniformly bounded and càdlàg in time a.s.. Moreover, assume that $X^n_t = 0$ if $t < 0$ and $X^n_t = X^n_T$ if $t > T$. Then, the sequence $\{\mathbb{P}^n \circ (X^n)^{-1}\}_{n\geq 1}$ is relatively compact as a sequence in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}),d_{M_1})}$.

Proof. By the uniform boundedness of b_n, σ_n, c_n and the quadratic variation of M^n, for each κ, there exists a constant C that is independent of n, such that

$$E^{\mathbb{P}^n} \sup_{0\leq t\leq T} |X^n_t|^\kappa \leq C < \infty. \tag{3.2}$$

By Proposition A.5(2) it is thus sufficient to check the tightness of $\{\mathbb{P}^n \circ (X^n)^{-1}\}_{n\geq 1}$. This can be achieved by applying Corollary B.7. Indeed, the condition (B.12) holds, due to (3.2). Hence, one only needs to check that for each $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\sup_n \mathbb{P}^n(\widetilde{w}_s(X^n, \delta) \geq \eta) < \epsilon.$$

To this end, we first notice that for each t and $t_1,\ t_2,\ t_3$ satisfying $0 \vee (t-\delta) \leq t_1 < t_2 < t_3 \leq (t+\delta) \wedge T$, the monotonicity of c_n implies

$$\begin{aligned}
&|X^n_{t_2} - [X^n_{t_1}, X^n_{t_3}]| \\
=\ & \inf_{0\leq\lambda\leq 1} |X^n_{t_2} - \lambda X^n_{t_1} - (1-\lambda)X^n_{t_3}| \\
\leq\ & \left|\int_{t_1}^{t_2} b_n(s)\,ds + \int_{t_1}^{t_2} \sigma_n(s)\,dM^n_s\right| + \left|\int_{t_2}^{t_3} b_n(s)\,ds + \int_{t_2}^{t_3} \sigma_n(s)\,dM^n_s\right| \\
& + \inf_{0\leq\lambda\leq 1} |c_n(t_2) - \lambda c_n(t_1) - (1-\lambda)c_n(t_3)| \\
=\ & \left|\int_{t_1}^{t_2} b_n(s)\,ds + \int_{t_1}^{t_2} \sigma_n(s)\,dM^n_s\right| + \left|\int_{t_2}^{t_3} b_n(s)\,ds + \int_{t_2}^{t_3} \sigma_n(s)\,dM^n_s\right|.
\end{aligned}$$

Similarly, for t_1 and t_2 satisfying $0 \le t_1 < t_2 \le \delta$,

$$|X^n_{t_1} - [0, X^n_{t_2}]| \le \left| \int_{t_1}^{t_2} b_n(s)\, ds + \int_{t_1}^{t_2} \sigma_n(s)\, dM^n_s \right|.$$

Therefore,

$$\widetilde{w}_s(X, \delta) \le \; 3 \sup_t \sup_{t_1, t_2} \left| \int_{t_1}^{t_2} b_n(s)\, ds + \int_{t_1}^{t_2} \sigma_n(s)\, dM^n_s \right|,$$

where the first supremum extends over $0 \le t \le T$ and the second one extends over $0 \vee (t - \delta) \le t_1 \le t_2 \le T \wedge (t + \delta)$. By the Markov inequality and the boundedness of b_n and σ_n, this yields

$$\mathbb{P}^n(\widetilde{w}_s(X^n, \delta) \ge \eta) \le \frac{k(\delta)}{\eta}, \tag{3.3}$$

for some constant C that is independent of n and some function $k(\delta)$ with $\lim_{\delta \to 0} k(\delta) = 0$. □

The next result shows that the class of all possible control rules is relatively compact. In a subsequent step this will allow us to apply Berge's maximum theorem.

Lemma 3.2. *Under assumptions $\mathcal{A}_1$, $\mathcal{A}_4$ and $\mathcal{A}_6$, the set $\bigcup_{\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))} \mathcal{R}(\mu)$ is relatively compact in $W_{p,(\widetilde{\mathcal{D}}(\mathbb{R}), d_{M_1})}$.*

Proof. Let $\{\mu^n\}_{n \ge 1}$ be any sequence in $\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and $\mathbb{P}^n \in \mathcal{R}(\mu^n), n \ge 1$. It is sufficient to show that $\{\mathbb{P}^n|_{\widetilde{\mathcal{D}}(\mathbb{R})}\}_{n \ge 1}$, $\{\mathbb{P}^n|_{\mathcal{U}(0,T)}\}_{n \ge 1}$ and $\{\mathbb{P}^n|_{\widetilde{\mathcal{A}}^m(\mathbb{R})}\}_{n \ge 1}$ are relatively compact. Since U and $\widetilde{\mathcal{A}}^m(\mathbb{R})$ are compact by assumption and Corollary B.5, respectively, $\{\mathbb{P}^n|_{\mathcal{U}(0,T)}\}_{n \ge 1}$ and $\{\mathbb{P}^n|_{\widetilde{\mathcal{A}}^m(\mathbb{R})}\}_{n \ge 1}$ are tight. Since $\mathcal{U}$ and $\widetilde{\mathcal{A}}^m(\mathbb{R})$ are bounded, these sequences are relatively compact in the topology induced by Wasserstein metric; see Proposition A.5(2).

It remains to prove the relative compactness of $\{\mathbb{P}^n|_{\widetilde{\mathcal{D}}(\mathbb{R})}\}_{n \ge 1}$. Since $\mathbb{P}^n$ is a control rule associated with the measure μ^n, for any n, it follows from Proposition A.6 that there exit extensions $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}_{t \ge 0}, \mathbb{Q}^n)$ of the canonical path spaces and processes (X^n, q^n, Z^n, M^n) defined on it, such that

$$dX^n_t = \int_U b(t, X^n_t, \mu^n_t, u)\, q^n_t(du) dt + \int_U \sigma(t, X^n_t, \mu^n_t, u)\, M^n(du, dt) + c(t)\, dZ^n_t$$

and

$$\mathbb{P}^n = \mathbb{P}^n \circ (X, q, Z)^{-1} = \mathbb{Q}^n \circ (X^n, q^n, Z^n)^{-1},$$

where M^n is a martingale measure on $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}_{t \ge 0}, \mathbb{Q}^n)$ with intensity q^n. Relative compactness of $\{\mathbb{P}^n \circ X^{-1}\}_{n \ge 1}$ now reduces to relative compactness of $\{\mathbb{Q}^n \circ (X^n)^{-1}\}_{n \ge 1}$, which is a direct consequence of the preceding Proposition 3.1. □

Remark 3.3. For the above proof, the assumption $c > 0$ is not necessary. To see this, we decompose $\bar{X}$ as

$$\bar{X}_\cdot = \int_0^\cdot \int_U b(t, \bar{X}_t, \mu^n_t, u)\, \bar{q}_t(du) dt + \int_0^\cdot \int_U \sigma(t, \bar{X}_t, \mu^n_t, u)\, M^n(du, dt) + \int_0^\cdot c^+(t)\, d\bar{Z}_t - \int_0^\cdot c^-(t)\, d\bar{Z}_t,$$

where c^+ and c^- are the positive and negative parts of c, respectively. By the above proof, we see that the law of

$$K_\cdot := \int_0^\cdot \int_U b(t, \bar{X}_t, \mu^n_t, u)\, \bar{q}_t(du) dt + \int_0^\cdot \int_U \sigma(t, \bar{X}_t, \mu^n_t, u)\, M^n(du, dt) + \int_0^\cdot c^+(t)\, d\bar{Z}_t$$

is tight. This implies, for each $\epsilon > 0$, the existence of a $(\widetilde{\mathcal{D}}(\mathbb{R}), d_{M_1})$-compact set $\mathcal{K}_1 \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ such that

$$\inf_n \mathbb{P}^n \left(\int_0^{\cdot} \int_U b(t, \bar{X}_t, \mu_t^n, u)\, \bar{q}_t(du)dt + \int_0^{\cdot} \int_U \sigma(t, \bar{X}_t, \mu_t^n, u)\, M^n(du, dt) + \int_0^{\cdot} c^+(t)\, d\bar{Z}_t \in \mathcal{K}_1 \right) \geq 1 - \epsilon.$$

On the other hand, due to the boundedness of Z, for the given $\epsilon > 0$, there exists a positive constant K_2 such that $\inf_n \mathbb{P}^n \left(\int_0^T c^-(s)\, d\bar{Z}_s \leq K_2 \right) \geq 1 - \epsilon$. Thus,

$$\begin{aligned} &\inf_n \mathbb{P}^n\{\omega \in \Omega : X_\cdot(\omega) \in \mathcal{K}_1 + \mathcal{K}_2\} \\ \geq\ &\inf_n \mathbb{P}^n \left\{ \omega \in \Omega : K_\cdot \in \mathcal{K}_1, -\int_0^{\cdot} c^-(s)\, dZ_s \in \mathcal{K}_2 \right\} \\ \geq\ &1 - 2\epsilon. \end{aligned} \tag{3.4}$$

By Corollary B.5, the set $\mathcal{K}_2 := \{z \in \widetilde{\mathcal{D}}(\mathbb{R}) : -z \in \widetilde{\mathcal{A}}(\mathbb{R}),\ z_T \geq -K_2\}$ is a M_1-compact subset of $\widetilde{\mathcal{D}}(\mathbb{R})$. Proposition B.8 implies that $\mathcal{K}_1 + \mathcal{K}_2$ is a M_1-compact subset of $\widetilde{\mathcal{D}}(\mathbb{R})$.

The next result states that the cost functional is continuous on the graph

$$\mathrm{Gr}\mathcal{R} := \{(\mu, \mathbb{P}) \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})) \times \mathcal{P}_p(\Omega) :\ \mathbb{P} \in \mathcal{R}(\mu)\}.$$

of the multi-function $\mathcal{R}$. This, too, will be needed to apply Berge's maximum theorem below.

Lemma 3.4. *Suppose that $\mathcal{A}_1$-$\mathcal{A}_6$ hold. Then $J : Gr\mathcal{R} \to \mathbb{R}$ is continuous.*

Proof. For each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and $\omega = (x, q, z) \in \Omega$, set

$$\mathcal{J}(\mu, \omega) = \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt + g(x_T, \mu_T) + \int_0^T h(t)\, dz_t. \tag{3.5}$$

Thus

$$J(\mu, \mathbb{P}) = \int_\Omega \mathcal{J}(\mu, \omega)\mathbb{P}(d\omega).$$

In a first step we prove that $\mathcal{J}(\cdot, \cdot)$ is continuous in the first variable; in a second step we prove continuity and a polynomial growth condition in the second variable. The two results together will give us the desired continuity of J.

Step 1: continuity in μ. Let $\mu^n \to \mu$ in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}), d_{M_1})}$ and recall that $\mu_t^n = \mu^n \circ X_t^{-1}$ and $\mu_t = \mu \circ X_t^{-1}$, where X is the coordinate process on $\widetilde{\mathcal{D}}(\mathbb{R})$. We consider the first two terms on the r.h.s. in (3.5) separately, starting with the first one. By assumption $\mathcal{A}_5$,

$$\begin{aligned} &\left| \int_0^T \int_U f(t, x_t, \mu_t^n, u)\, q_t(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt \right| \\ \leq\ &C \int_0^T \left(1 + L\left(\mathcal{W}_p(\mu_t^n, \delta_0), \mathcal{W}_p(\mu_t, \delta_0)\right)\right) \mathcal{W}_p(\mu_t^n, \mu_t)\, dt \\ \leq\ &C \left(\int_0^T \left(1 + L\left(\mathcal{W}_p(\mu_t^n, \delta_0), \mathcal{W}_p(\mu_t, \delta_0)\right)\right)^{\frac{p}{p-1}}\, dt \right)^{1 - \frac{1}{p}} \left(\int_0^T \mathcal{W}_p(\mu_t^n, \mu_t)^p\, dt \right)^{\frac{1}{p}}. \end{aligned} \tag{3.6}$$

The convergence $\mu^n \to \mu$ in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}), d_{M_1})}$ implies $\mu^n \to \mu$ weakly. By Skorokhod's representation theorem, there exits $\bar{X}^n$ and $\bar{X}$ defined on some probability space $(\mathbb{Q}, \bar{\Omega}, \bar{\mathcal{F}})$, such that

$$\mu^n = \mathbb{Q} \circ (\bar{X}^n)^{-1}, \quad \mu = \mathbb{Q} \circ \bar{X}^{-1}$$

and

$$d_{M_1}(\bar{X}^n, \bar{X}) \to 0 \quad \mathbb{Q}\text{-}a.s.$$

Hence, (3.6) implies that

$$\left| \int_0^T \int_U f(t, x_t, \mu_t^n, u)\, q_t(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt \right|$$
$$\leq C \left(\int_0^T \left(1 + L\left(\mathcal{W}_p(\mathbb{Q} \circ (\bar{X}_t^n)^{-1}, \delta_0), \mathcal{W}_p(\mathbb{Q} \circ \bar{X}_t^{-1}, \delta_0)\right)\right)^{\frac{p}{p-1}} dt \right)^{1-\frac{1}{p}} \left(E^{\mathbb{Q}} \int_0^T |\bar{X}_t^n - \bar{X}_t|^p \, dt \right)^{\frac{1}{p}}$$

For each $t \in [0, T]$,

$$|\bar{X}_t^n(\bar{\omega}) - \bar{X}_t(\bar{\omega})|^p \leq 2^p \left(d_{M_1}(\bar{X}^n(\bar{\omega}), 0)^p + d_{M_1}(\bar{X}(\bar{\omega}), 0)^p \right),$$

and so,

$$\int_0^T |\bar{X}_t^n(\bar{\omega}) - \bar{X}_t(\bar{\omega})|^p \, dt \leq 2^p T \left(d_{M_1}(\bar{X}^n(\bar{\omega}), 0)^p + d_{M_1}(\bar{X}(\bar{\omega}), 0)^p \right).$$

On the other hand,

$$\begin{aligned} E^{\mathbb{Q}} \left(d_{M_1}(\bar{X}^n, 0)^p + d_{M_1}(\bar{X}, 0)^p \right) &= \int_{\mathcal{D}[0,T]} d_{M_1}(x, 0)^p \, \mu^n(dx) + \int_{\mathcal{D}[0,T]} d_{M_1}(x, 0)^p \, \mu(dx) \\ &\to 2 \int_{\mathcal{D}[0,T]} d_{M_1}(x, 0)^p \, \mu(dx) < \infty. \end{aligned}$$

Therefore, dominated convergence yields

$$E^{\mathbb{Q}} \int_0^T |\bar{X}_t^n - \bar{X}_t|^p \, dt \to 0. \tag{3.7}$$

Since $\sup_n \mathcal{W}_p(\mathbb{Q} \circ (\bar{X}_t^n)^{-1}, \delta_0) < \infty$ it thus follows from the local boundedness of the function L that

$$\left| \int_0^T \int_U f(t, x_t, \mu_t^n, u)\, q_t(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt \right| \to 0.$$

As for the second term on the r.h.s. in (3.5) recall first that $x^n \to x$ in M_1 implies $x_t^n \to x_t$ for each $t \notin Dist(x)$ and $x_T^n \to x_T$. In particular, the mapping $x \mapsto \varphi(x_T)$ is continuous for any continuous real-valued function φ on $\mathbb{R}^d$. Since any continuous positive function φ on $\mathbb{R}^d$ that satisfies $\varphi(x) \leq C(1+|x|^p)$, also satisfies

$$\varphi(x_T) \leq C(1 + |x_T|^p) \leq C(1 + d_{M_1}(x, 0)^p)$$

we see that

$$\left| \int_{\mathbb{R}^d} \varphi(x)\, \mu_T^n(dx) - \int_{\mathbb{R}^d} \varphi(x)\, \mu_T(dx) \right| = \left| \int_{\widetilde{\mathcal{D}}(\mathbb{R})} \varphi(x_T)\, \mu^n(dx) - \int_{\widetilde{\mathcal{D}}(\mathbb{R})} \varphi(x_T)\, \mu(dx) \right| \overset{n \to \infty}{\longrightarrow} 0.$$

More generally, we obtain $\mu_T^n \to \mu_T$ from $\mu^n \to \mu$, which also implies that $g(x_T, \mu_T^n) \to g(x_T, \mu_T)$.

Step 2: continuity in ω. If $\omega^n = (x^n, q^n, z^n) \to \omega = (x, q, z)$, then $x_T^n \to x_T$. In particular,

$$g(x_T^n, \mu_T) \to g(x_T, \mu_T).$$

Moreover, $z^n \to z$ in M_1 implies $z_t^n \to z_t$ for for all continuity points of z and $z_T^n \to z_T$. By the Portmanteau theorem this implies that

$$\int_0^T h(t)\, dz_t^n \to \int_0^T h(t)\, dz_t.$$

Next we show the convergence of $\int_0^T \int_U f(t, x_t^n, \mu_t, u)\, q_t^n(du)dt$ to $\int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt$. We have that,

$$
\begin{aligned}
&\int_0^T \int_U f(t, x_t^n, \mu_t, u)\, q_t^n(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt \\
=& \int_0^T \int_U f(t, x_t^n, \mu_t, u)\, q_t^n(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t^n(du)dt \\
&+ \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t^n(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt.
\end{aligned}
$$

By Assumption $\mathcal{A}_2$ the convergence of x^n to x yields $f(t, x_t^n, \mu_t, u) \to f(t, x_t, \mu_t, u)$ for each $t \notin Disc(x)$. From the compactness of U it follows that $\sup_{u \in U} |f(t, x_t^n, \mu_t, u) - f(t, x_t, \mu_t, u)| \to 0$ for each $t \notin Disc(x)$. Since $Disc(x)$ is at most countable this implies

$$
\begin{aligned}
&\left| \int_0^T \int_U f(t, x_t^n, \mu_t, u)\, q_t^n(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t^n(du)dt \right| \\
\leq& \int_0^T \sup_{u \in U} |f(t, x_t^n, \mu_t, u) - f(t, x_t, \mu_t, u)|\, dt \to 0.
\end{aligned}
$$

The compactness of U, the growth condition on f and the convergence of q^n to q, imply

$$
\lim_{n \to \infty} \left| \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t^n(du)dt - \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt \right| = 0.
$$

Step 3: continuity of J. Thus far, we have established the continuity of the mapping $(\mu, \omega) \to \mathcal{J}(\mu, \omega)$. We are now going to apply Proposition A.5(4) to prove the continuity of J. To this end, notice first that for each fixed $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$, due to Assumption $\mathcal{A}_3$,

$$
\begin{aligned}
&\left| \int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt + \int_0^T h(t)\, dz_t \right| \\
\leq& C \left(1 + \int_0^T \int_U \left(1 + |x_t|^p + |u|^p + \int_{\mathbb{R}^d} |y|^p \mu_t(dy) \right) q_t(du)dt + z_T \right) \\
\leq& C \left(1 + d_{M_1}(x, 0)^p + \mathcal{W}_{p, U \times [0,T]}(q, \delta_0)^p + d_{M_1}(z, 0) + \int_0^T \int_{\mathbb{R}^d} |y|^p\, \mu_t(dy)dt \right) \\
\leq& C \left(1 + d_{M_1}(x, 0)^p + \mathcal{W}_{p, U \times [0,T]}(q, \delta_0)^p + d_{M_1}(z, 0)^p + \int_{\widetilde{\mathcal{D}}(\mathbb{R})} d_{M_1}(y, 0)^p\, \mu(dy) \right).
\end{aligned}
$$

Hence, using the previously established continuity in μ and the local Lipschitz continuity of f in μ, it follows from Proposition A.5 that $(\mu^n, \mathbb{P}^n) \to (\mu, \mathbb{P})$ implies

$$
\begin{aligned}
&E^{\mathbb{P}^n} \left(\int_0^T \int_U f(t, x_t, \mu_t^n, u)\, q_t(du)dt + \int_0^T h(t)\, dz_t \right) \\
\to& E^{\mathbb{P}} \left(\int_0^T \int_U f(t, x_t, \mu_t, u)\, q_t(du)dt + \int_0^T h(t)\, dz_t \right).
\end{aligned}
\tag{3.8}
$$

Since the terminal cost functions is not necessarily Lipschitz continuous we need to argue differently in order to prove the continuous dependence of the expected terminal cost on $(\mu, \mathbb{P})$. First, we notice that for each $\widetilde{p} > \bar{p}$, by the boundedness of b, σ and Z, we have that

$$
\sup_n E^{\mathbb{P}^n} d_{M_1}(X, 0)^{\widetilde{p}} \leq C < \infty, \tag{3.9}
$$

which implies

$$\lim_{K\to\infty}\sup_n \int_{\{x: d_{M_1}(x,0)>K\}} d_{M_1}(x,0)^{\bar{p}}\,\mathbb{P}^n(dx) = 0. \tag{3.10}$$

By Assumption $\mathcal{A}_3$,

$$|g(x_T,\mu_T)| \le C\left(1+|x_T|^{\bar{p}}+\int |y|^p\mu_T(dy)\right) \le C\left(1+|x_T|^{\bar{p}}+\int |y|^{\bar{p}}\mu_T(dy)\right).$$

Together with (3.10) this implies,

$$E^{\mathbb{P}^n} g(X_T,\mu_T) \to E^{\mathbb{P}} g(X_T,\mu_T). \tag{3.11}$$

By the tightness of $\{\mathbb{P}^n\}_{n\ge1}$, for each $\epsilon>0$, there exists a compact set $K_\epsilon \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ such that

$$\begin{aligned}
&\left|\int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T,\mu_T^n)\mathbb{P}^n(dx) - \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T,\mu_T)\mathbb{P}^n(dx)\right| \\
\le\ & \int_{K_\epsilon} |g(x_T,\mu_T^n)-g(x_T,\mu_T)|\mathbb{P}^n(dx) + \int_{\widetilde{\mathcal{D}}(\mathbb{R})/K_\epsilon} |g(x_T,\mu_T^n)-g(x_T,\mu_T)|\mathbb{P}^n(dx) \\
\le\ & \sup_{x\in K_\epsilon} |g(x_T,\mu_T^n)-g(x_T,\mu_T)| + \left(\int_{\widetilde{\mathcal{D}}(\mathbb{R})/K_\epsilon} |g(x_T,\mu_T^n)-g(x_T,\mu_T)|^2\mathbb{P}^n(dx)\right)^{\frac12}\left(\sup_n \mathbb{P}^n(\widetilde{\mathcal{D}}(\mathbb{R})/K_\epsilon)\right)^{\frac12} \\
\le\ & \sup_{x\in K_\epsilon} |g(x_T,\mu_T^n)-g(x_T,\mu_T)| + C\epsilon^{\frac12} \quad \text{(by (3.9))}.
\end{aligned}$$

Thus,

$$\left|\int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T,\mu_T^n)\mathbb{P}^n(dx) - \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T,\mu_T)\mathbb{P}^n(dx)\right| \to 0. \tag{3.12}$$

The convergence (3.8), (3.11) and (3.12) yield the continuity of $J(\cdot,\cdot)$. □

We now recall from [20, Proposition 3.1] an equivalent characterization for the set of control rules $\mathcal{R}(\mu)$. This equivalent characterization allows us to verify the martingale property of the state process by verifying the martingale property of its continuous part.

Proposition 3.5. A probability measure $\mathbb{P}$ is a control rule with respect to the given μ if and only if there exists a $\mathcal{C}(0,T)$-valued $\{\mathcal{F}_t, t\ge 0\}$ adapted process Y on the filtered canonical space $(\Omega,\mathcal{F},\{\mathcal{F}_t,t\ge0\})$ such that

(1) $\mathbb{P}(X_t=0, Z_t=0 \text{ if } t<0, X_t=X_T, Z_t=Z_T \text{ if } t>T \text{ and } q_0=\delta_{u_0})=1$;

(2) $\mathbb{P}(\omega\in\Omega:\ X_\cdot(\omega)=Y_\cdot(\omega)+Z_\cdot(\omega))=1$;

(3) for each $\phi\in\mathcal{C}_b^2(\mathbb{R}^d)$, $\overline{\mathcal{M}}^{\mu,\phi}$ is a continuous $(\mathbb{P},\{\mathcal{F}_t,t\ge0\})$ martingale, where

$$\overline{\mathcal{M}}_t^{\mu,\phi} = \phi(Y_t) - \int_0^t\int_U \bar{\mathcal{L}}\phi(s,X_s,Y_s,\mu_s,u)\,q_s(du)ds, \tag{3.13}$$

with $\bar{\mathcal{L}}\phi(s,x,y,\mu,u)=\sum_i b_i(s,x,\mu,u)\partial_{y_i}\phi(y)+\frac12\sum_{ij}a_{ij}(s,x,\mu,u)\frac{\partial^2\phi(y)}{\partial_{y_i}\partial_{y_j}}$.

The previous characterization of control rules allows us to show that the correspondence $\mathcal{R}$ has a closed graph.

Proposition 3.6. Suppose that $\mathcal{A}_1$ and $\mathcal{A}_4$-$\mathcal{A}_6$ hold. For any sequence $\{\mu^n\}_{n\ge1}\subseteq\mathcal{P}_p(\widetilde{D}(\mathbb{R}))$ and $\mu\in\mathcal{P}_p(\widetilde{D}(\mathbb{R}))$ with $\mu^n\to\mu$ in $\mathcal{W}_{p,(\widetilde{D}(\mathbb{R}),d_{M_1})}$, if $\{\mathbb{P}^n\}_{n\ge1}\subseteq\mathcal{R}(\mu^n)$ and $\mathbb{P}^n\to\mathbb{P}$ in $\mathcal{W}_p$, then $\mathbb{P}\in\mathcal{R}(\mu)$.

Proof. By the Proposition B.1(3) and the Portmanteau theorem the set

$$\widetilde{\Omega}_{0,T} := \{\omega : x_t = 0, z_t = 0 \text{ if } t < 0, x_t = x_T, z_t = z_T \text{ if } t > T \text{ and } q_0 = \delta_{u_0}\}$$

is closed. Hence $\mathbb{P}$ satisfies condition (1) in Proposition 3.5.

In order to verify conditions (2) and (3), notice first that, for each n, there exits a $\mathcal{C}(0,T)$-valued stochastic process Y^n such that

$$\mathbb{P}^n \left(X_\cdot = Y_\cdot^n + \int_0^\cdot c(s)\, dZ_s \right) = 1$$

and such that the corresponding martingale problem is satisfied. In order to show that a similar decomposition and the martingale problem hold under the measure $\mathbb{P}$ we apply Proposition A.6. For each n, there exits a probability space $(\Omega^n, \mathcal{F}^n, \mathbb{Q}^n)$ that supports random variables $(\bar{X}^n, \bar{q}^n, \bar{Z}^n)$ and a martingale measure M^n with intensity $\bar{q}^n$ such that

$$\mathbb{P}^n = \mathbb{Q}^n \circ (\bar{X}^n, \bar{q}^n, \bar{Z}^n)^{-1}$$

and

$$d\bar{X}_t^n = \int_U b(t, \bar{X}_t^n, \mu_t^n, u)\, \bar{q}_s^n(du)ds + \int_U \sigma(t, \bar{X}_t^n, \mu_t^n, u)\, M^n(du, dt) + c(t)d\bar{Z}_t^n.$$

Thus, for each $0 \le s < t \le T$,

$$\begin{aligned} E^{\mathbb{P}^n} |Y_t^n - Y_s^n|^4 &= E^{\mathbb{P}^n} \left| \left(X_t - \int_0^t c(r)\, dZ_r \right) - \left(X_s - \int_0^s c(r)\, dZ_r \right) \right|^4 \\ &= E^{\mathbb{Q}^n} \left| \left(\bar{X}_t^n - \int_0^t c(r)\, d\bar{Z}_r^n \right) - \left(\bar{X}_s^n - \int_0^s c(r)\, d\bar{Z}_r^n \right) \right|^4 \\ &= E^{\mathbb{Q}^n} \left| \int_s^t \int_U b(r, \bar{X}_r^n, \mu_r^n, u)\, \bar{q}_r^n(du)dr + \int_s^t \int_U \sigma(r, \bar{X}_r^n, \mu_r^n, u)\, M^n(du, dr) \right|^4 \\ &\le C|t-s|^2. \end{aligned} \tag{3.14}$$

Hence, Kolmogorov's weak compactness criterion implies the tightness of Y^n. Therefore, taking a subsequence if necessary, the sequence (X, q, Z, Y^n) of random variables taking values in $\Omega \times \mathcal{C}(0,T)$ has weak limit $(\widehat{X}, \widehat{q}, \widehat{Z}, \widehat{Y})$ defined on some probability space.

By Skorokhod's representation theorem, there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{Q})$ that supports random variables $(\widetilde{X}^n, \widetilde{q}^n, \widetilde{Z}^n, \widetilde{Y}^n)$ and $(\widetilde{X}, \widetilde{q}, \widetilde{Z}, \widetilde{Y})$ such that

$$\mathcal{L}(\widetilde{X}^n, \widetilde{q}^n, \widetilde{Z}^n, \widetilde{Y}^n) = \mathcal{L}(X, q, Z, Y^n), \quad \mathcal{L}(\widetilde{X}, \widetilde{q}, \widetilde{Z}, \widetilde{Y}) = \mathcal{L}(\widehat{X}, \widehat{q}, \widehat{Z}, \widehat{Y})$$

where $\mathcal{L}(\cdot)$ stands for the law of a random variable, and such that

$$(\widetilde{X}^n, \widetilde{q}^n, \widetilde{Z}^n, \widetilde{Y}^n) \to (\widetilde{X}, \widetilde{q}, \widetilde{Z}, \widetilde{Y}) \quad \mathbb{Q}\text{-a.s.}$$

In particular, $\tilde{Y}$ is $\mathcal{C}(0,T)$-valued as the uniform limit of a sequence of continuous processes, and

$$\mathbb{Q} \left(\widetilde{X}_\cdot = \widetilde{Y}_\cdot + \int_0^\cdot c(s)\, d\widetilde{Z}_s \right) = 1.$$

Since $\mathbb{P}^n \to \mathbb{P}$, we have $\mathbb{P} \circ (X, q, Z)^{-1} = \mathbb{Q} \circ (\widetilde{X}, \widetilde{q}, \widetilde{Z})^{-1}$. Hence, there exits a $\mathcal{C}(0,T)$-valued stochastic process Y such that

$$\mathbb{P} \left(X_\cdot = Y_\cdot + \int_0^\cdot c(s)\, dZ_s \right) = 1$$

and $\mathbb{P}\circ(X,q,Z,Y)^{-1}=\mathbb{Q}\circ(\widetilde{X},\widetilde{q},\widetilde{Z},\widetilde{Y})^{-1}$. Finally, for each $t\in[0,T]$, define

$$\overline{\mathcal{M}}_t^{n,\mu^n,\phi}=\phi(Y_t^n)-\int_0^t\int_U\bar{\mathcal{L}}(s,X_s,Y_s^n,\mu_s^n,u)\,q_s(du)ds,$$

$$\widetilde{\mathcal{M}}_t^{n,\mu^n,\phi}=\phi(\widetilde{Y}_t^n)-\int_0^t\int_U\bar{\mathcal{L}}(s,\widetilde{X}_s^n,\widetilde{Y}_s^n,\mu_s^n,u)\,\widetilde{q}_s^n(du)ds,$$

and

$$\widetilde{\mathcal{M}}_t^{\mu,\phi}=\phi(\widetilde{Y}_t)-\int_0^t\int_U\bar{\mathcal{L}}(s,\widetilde{X}_s,\widetilde{Y}_s,\mu_s,u)\,\widetilde{q}_s(du)ds.$$

For each $0\le s<t\le T$ and each F that is continuous, bounded and $\mathcal{F}_s$-measurable, we have

$$\begin{aligned}0&=E^{\mathbb{P}^n}\left(\overline{\mathcal{M}}_t^{n,\mu^n,\phi}-\overline{\mathcal{M}}_s^{n,\mu^n,\phi}\right)F(X,q,Z)=E^{\mathbb{Q}}\left(\widetilde{\mathcal{M}}_t^{n,\mu^n,\phi}-\widetilde{\mathcal{M}}_s^{n,\mu^n,\phi}\right)F(\widetilde{X}^n,\widetilde{q}^n,\widetilde{Z}^n)\\&\to E^{\mathbb{Q}}\left(\widetilde{\mathcal{M}}_t^{\mu^*,\phi}-\widetilde{\mathcal{M}}_s^{\mu^*,\phi}\right)F(\widetilde{X},\widetilde{q},\widetilde{Z})=E^{\mathbb{P}}\left(\overline{\mathcal{M}}_t^{\mu,\phi}-\overline{\mathcal{M}}_s^{\mu,\phi}\right)F(X,q,Z).\end{aligned}\tag{3.15}$$

□

Corollary 3.7. Suppose that $\mathcal{A}_1$ and $\mathcal{A}_4$-$\mathcal{A}_6$ hold. Then, for each $\mu\in\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ the set $\mathcal{R}(\mu)$ is compact.

Proof. It suffices to prove for each $\mu\in\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and any sequence $\{\mathbb{P}^n\}_{n\ge1}\subseteq\mathcal{R}(\mu)$, there exists a convergence subsequence $\mathbb{P}^{n_k}$ and that the limit still belongs to $\mathcal{R}(\mu)$. By Lemma 3.2, $\bigcup_n\mathcal{R}(\mu^n)$ is relatively compact. Hence, the assertion follows from Proposition 3.6. □

Corollary 3.8. Suppose that $\mathcal{A}_1$ and $\mathcal{A}_4$-$\mathcal{A}_6$ hold. Then, $\mathcal{R}(\cdot):\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))\to2^{\mathcal{P}_p(\Omega)}$ is hemi-continuous.

Proof. The lower hemi-continuity of $\mathcal{R}$ can be dealt with as [31, Lemma 4.4]. The upper hemi-continuity follows from the closed-graph theorem and Corollary 3.7. □

Corollary 3.9. Under assumptions $\mathcal{A}_1$-$\mathcal{A}_6$, the stochastic singular control problem (2.1) admits an optimal control rule in the sense of Definition 2.3.

Proof. This is a direct corollary of Lemma 3.2, Corollary 3.8, Lemma 3.4 and Theorem A.1. □

Remark 3.10. Using our method, we could have obtained Corollary 3.9 under the same assumptions of the coefficients as in [20]. We will generalize it to McKean-Vlasov case at the end of this section.

Theorem 3.11. *Under assumptions $\mathcal{A}_1$-$\mathcal{A}_6$ and the finite-fuel constraint $Z\in\widetilde{\mathcal{A}}^m(\mathbb{R})$, there exists a relaxed solution to (1.4).*

Proof. By [27, Section 5.4], for each $\mu\in\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ the set $\mathcal{R}(\mu)$ is nonempty. Moreover, by Corollary 3.7, the correspondence $\mathcal{R}$ is compact valued. Therefore, by Corollary 3.8, Lemma 3.4 and Theorem A.1, the argmax-correspondence $\mathcal{R}^*$ is upper hemi-continuous.

From inequality (3.3) in the proof of Lemma 3.1, we see that for each $\mu\in\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and $\mathbb{P}\in\mathcal{R}(\mu)$, there exists a nonnegative function $k(\cdot)$ that is independent of μ, such that $\mathbb{P}(\widetilde{w}_s(X,\delta)>\eta)\le\frac{k(\delta)}{\eta}$ and $\lim_{\delta\to0}k(\delta)=0$, where $\widetilde{w}_s$ is the modified oscillation function defined in (B.11).

Let us now define a set-valued map ψ by

$$\begin{aligned}\psi:\ &\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))\to2^{\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))},\\&\mu\mapsto\{\mathbb{P}|_{\widetilde{\mathcal{D}}(\mathbb{R})}:\mathbb{P}\in\mathcal{R}^*(\mu)\},\end{aligned}\tag{3.16}$$

and let

$$S=\left\{\mathbb{P}\in\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})):\text{for each }\eta>0,\ \mathbb{P}(\widetilde{w}_s(X,\delta)>\eta)\le\frac{k(\delta)}{\eta}\text{ and }E^{\mathbb{P}}\sup_{0\le t\le T}|X_t|^{\bar p}\le C\right\}$$

where $C<\infty$ denotes the upper bound in (3.2). It can be checked that S is non-empty, relatively compact, convex, and that $\psi(\mu)\subseteq S\subseteq\bar S$, for each $\mu\in\widetilde{\mathcal{D}}(\mathbb{R})$. Hence, $\psi:\bar S\to 2^{\bar S}$. Therefore, Theorem A.2 is applicable by embedding $\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ into $\mathcal{M}(\widetilde{\mathcal{D}}(\mathbb{R}))$, the space of all bounded signed measures on $\widetilde{\mathcal{D}}(\mathbb{R})$ endowed with weak convergence topology. □

3.2 Existence in the general case

In this section we establish the existence of a solution to MFGs with singular controls for general singular controls $Z\in\widetilde{\mathcal{A}}(\mathbb{R})$. For each m and μ, define

$$\Omega^m=\widetilde{\mathcal{D}}(\mathbb{R})\times\mathcal{U}\times\widetilde{\mathcal{A}}^m(\mathbb{R})$$

and denote by $\mathcal{R}^m(\mu)$ the control rules corresponding to Ω^m and μ, that is, $\mathcal{R}^m(\mu)$ is the subset of probability measures in $\mathcal{R}(\mu)$ that are supported on Ω^m. Denote by $\mathbf{MFG}^m$ the mean field games corresponding to Ω^m. The preceding analysis showed that there exits a solution $\mathbb{P}^{m*}$ to $\mathbf{MFG}^m$, for each m. In what follows,

$$\mu^{m*}:=\mathbb{P}^{m*}\circ X^{-1}.$$

The next lemma shows that the sequence $\{\mathbb{P}^{m*}\}_{m\ge1}$ is relatively compact; the subsequent one shows that any accumulation point is a control rule.

Lemma 3.12. *Suppose $\mathcal{A}_1$, $\mathcal{A}_3$, $\mathcal{A}_4$ and $\mathcal{A}_6$ hold. Then there exists a constant $K<\infty$ such that*

$$\sup_m E^{\mathbb{P}^{m*}}|Z_T|^{\bar p}\le K<\infty.$$

As a consequence, the sequence $\{\mathbb{P}^{m*}\}_{m\ge1}$ is relatively compact.

Proof. We recall that $c(\cdot)$ is bounded away from 0. Hence, there exists a constant $C<\infty$ such that, for all $m\in\mathbb{N}$,

$$E^{\mathbb{P}^{m*}}|Z_T|^{\bar p}\le C\left(1+E^{\mathbb{P}^{m*}}|X_T|^{\bar p}\right)\tag{3.17}$$

and

$$E^{\mathbb{P}^{m*}}|X_t|^{p}\le C\left(1+E^{\mathbb{P}^{m*}}|Z_T|^{p}\right),\ t\in[0,T].\tag{3.18}$$

Moreover,

$$\begin{aligned}J(\mu^{m*},\mathbb{P}^{m*})=&\ E^{\mathbb{P}^{m*}}\left[\int_0^T\int_U f(t,X_t,\mu_t^{m*},u)\,q_t(du)dt+g(X_T,\mu_T^{m*})+\int_0^T h(t)\,dZ_t\right]\\
\ge&\ -C\Bigg(1+\int_0^T\int_{\mathbb{R}^d}|x|^p\,\mu_t^{m*}(dx)dt+E^{\mathbb{P}^{m*}}\int_0^T|X_t|^p\,dt+E^{\mathbb{P}^{m*}}\int_0^T\int_U|u|^p\,q_t(du)dt\\
&-E^{\mathbb{P}^{m*}}|X_T|^{\bar p}+\int_{\mathbb{R}^d}|x|^p\,\mu_T^{m*}(dx)+E^{\mathbb{P}^{m*}}\left|\int_0^T h(t)\,dZ_t\right|\Bigg)\quad\text{(by assumption }\mathcal{A}_3)\\
\ge&\ -C\Bigg(1+\int_0^T\int_{\mathbb{R}^d}|x|^p\,\mu_t^{m*}(dx)dt+E^{\mathbb{P}^{m*}}\int_0^T|X_t|^p\,dt+E^{\mathbb{P}^{m*}}\int_0^T\int_U|u|^p\,q_t(du)dt\\
&+\int_{\mathbb{R}^d}|x|^p\,\mu_T^{m*}(dx)+E^{\mathbb{P}^{m*}}\left|\int_0^T h(t)\,dZ_t\right|-E^{\mathbb{P}^{m*}}|Z_T|^{\bar p}\Bigg)\quad\text{(by (3.17))}.\end{aligned}$$

Now choose any $\mathbb{P}_0 \in \mathcal{R}^m(\mu^{m*})$ such that $\sup_m J(\mu^{m*}, \mathbb{P}_0) < \infty$. Then,

$$\begin{aligned}
E^{\mathbb{P}^{m*}}|Z_T|^{\bar{p}} &\leq J(\mu^m, \mathbb{P}^{m*}) + C\left(1 + E^{\mathbb{P}^{m*}}\left|\int_0^T h(t)\,dZ_t\right| + E^{\mathbb{P}^{m*}}\int_0^T |X_t|^p\,dt + E^{\mathbb{P}^{m*}}|X_T|^p\right) \\
&\leq J(\mu^m, \mathbb{P}_0) + C\left(1 + E^{\mathbb{P}^{m*}}|Z_T| + E^{\mathbb{P}^{m*}}|Z_T|^p\right) \quad \text{(by (3.18) and the optimality of } \mathbb{P}^{m*}) \\
&\leq C\left(1 + E^{\mathbb{P}^{m*}}|Z_T| + E^{\mathbb{P}^{m*}}|Z_T|^p\right).
\end{aligned} \tag{3.19}$$

Since the measure $\mathbb{P}^{m*}$ is supported on Ω^m, we see that $E^{\mathbb{P}^{m*}}|Z_T|^{\bar{p}}$ is finite, for each m. In order to see that there exists a uniform upper bound on $E^{\mathbb{P}^{m*}}|Z_T|^{\bar{p}}$, notice that, independently of m we can choose $M > 0$ large enough such that

$$E^{\mathbb{P}^{m*}}|Z_T|^{p_0} \leq M + \frac{1}{4C}E^{\mathbb{P}^{m*}}|Z_T|^{\bar{p}} \quad (p_0 = 1, p)$$

Together with (3.19) this yields,

$$E^{\mathbb{P}^{m*}}|Z_T|^{\bar{p}} \leq 2C(1+M) := K.$$

By Proposition A.5 and Lemma 3.2, the relative compactness of $\{\mathbb{P}^{m*}\}_{m\geq 1}$ follows. □

The previous lemma shows that the sequence $\{\mathbb{P}^{m*}\}_{m\geq 1}$ has an accumulation point $\mathbb{P}^*$. Let $\mu^* = \mathbb{P}^* \circ X^{-1}$. Clearly, $\mu^{m*} \to \mu^*$ in $\mathcal{W}_p$ along a subsequence. The following result is an immediate corollary to Proposition 3.6.

Lemma 3.13. *Suppose that $\mathcal{A}_1$ and $\mathcal{A}_3$-$\mathcal{A}_6$ hold, let $\mathbb{P}^*$ be an accumulation point of the sequence $\{\mathbb{P}^{m*}\}_{m\geq 1}$. Then, $\mathbb{P}^* \in \mathcal{R}(\mu^*)$.*

The next theorem establish the existence of relaxed MFGs solution to (1.4) in the general case, i.e. it proves Theorem 2.6.

Theorem 3.14. *Suppose $\mathcal{A}_1$-$\mathcal{A}_6$ hold. Then $\mathbb{P}^* \in \mathcal{R}^*(\mu^*)$, i.e., for each $\mathbb{P} \in \mathcal{R}(\mu^*)$ it holds that*

$$J(\mu^*, \mathbb{P}^*) \leq J(\mu^*, \mathbb{P}).$$

Proof. It is sufficient to prove that $J(\mu^*, \mathbb{P}^*) \leq J(\mu^*, \mathbb{P})$ for each $\mathbb{P} \in \mathcal{R}(\mu^*)$ with $J(\mu^*, \mathbb{P}) < \infty$.

By Proposition A.6, there exists a filtered probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathcal{F}}_t, \bar{\mathbb{P}})$ on which random variables $(\bar{X}, \bar{q}, \bar{Z}, M)$ are defined such that $\mathbb{P} = \bar{\mathbb{P}} \circ (\bar{X}, \bar{q}, \bar{Z})^{-1}$ and

$$d\bar{X}_t = \int_U b(t, \bar{X}_t, \mu_t^*, u)\,\bar{q}_t(du)dt + \int_U \sigma(t, \bar{X}_t, \mu_t^*, u)M(du, dt) + c(t)\,d\bar{Z}_t, \tag{3.20}$$

where M is a martingale measure with intensity $\bar{q}$. Using the same argument as in the proof of Lemma 3.12 we see that,

$$E^{\mathbb{P}} Z_T^{\bar{p}} = E^{\bar{\mathbb{P}}} \bar{Z}_T^{\bar{p}} < \infty. \tag{3.21}$$

Define $\mathbb{P}^m = \bar{\mathbb{P}} \circ (\bar{X}^m, \bar{q}, \bar{Z}^m) \in \mathcal{R}^m(\mu^{m*})$, such that $\bar{X}^m$ is the unique strong solution to

$$d\bar{X}_t^m = \int_U b(t, \bar{X}_t^m, \mu_t^{m*}, u)\,\bar{q}_t(du)dt + \int_U \sigma(t, \bar{X}_t^m, \mu_t^{m*}, u)M(du, dt) + c(t)\,d\bar{Z}_t^m, \tag{3.22}$$

where for each $\bar{\omega} \in \bar{\Omega}$,

$$\bar{Z}_t^m(\bar{\omega}) = \begin{cases} \bar{Z}_t(\bar{\omega}), & \text{if } t < \tau^m(\bar{\omega}) \\ m, & \text{if } t \geq \tau^m(\bar{\omega}), \end{cases}$$

with $\tau^m(\bar{\omega}) = \inf\{t : \bar{Z}_t(\bar{\omega}) > m\}$. Similarly, we can define Z^m. Furthermore, if Z is $\widetilde{\mathcal{A}}^m(\mathbb{R})$ valued, then $Z = Z^m$. Hence,

$$\begin{aligned}
&E^{\bar{\mathbb{P}}} \sup_{0\le t\le T} \left| \int_0^t c(s)\, d\bar{Z}_s - \int_0^t c(s)\, d\bar{Z}_s^m \right| \\
&= E^{\mathbb{P}} \sup_{0\le t\le T} \left| \int_0^t c(s)\, dZ_s - \int_0^t c(s)\, dZ_s^m \right| \\
&= \int_{\widetilde{\mathcal{A}}^m(\mathbb{R})} \sup_{0\le t\le T} \left| \int_0^t c(s)\, dZ_s(\omega) - \int_0^t c(s)\, dZ_s^m(\omega) \right| \mathbb{P}(d\omega) \\
&\quad + \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})} \sup_{0\le t\le T} \left| \int_0^t c(s)\, dZ_s(\omega) - \int_0^t c(s)\, dZ_s^m(\omega) \right| \mathbb{P}(d\omega) \\
&= \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})} \sup_{0\le t\le T} \left| \int_0^t c(s)\, dZ_s(\omega) - \int_0^t c(s)\, dZ_s^m(\omega) \right| \mathbb{P}(d\omega).
\end{aligned} \tag{3.23}$$

By Hölder's inequality,

$$\begin{aligned}
&\int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})} \sup_{0\le t\le T} \left| \int_0^t c(s)\, dZ_s(\omega) - \int_0^t c(s)\, dZ_s^m(\omega) \right| \mathbb{P}(d\omega) \\
&\le \left| \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})} \int_0^T c(t)\, dZ_t(\omega)\mathbb{P}(d\omega) + \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})} \int_0^T c(t)\, dZ_t^m(\omega)\mathbb{P}(d\omega) \right| \\
&\le C\left(E^{\mathbb{P}} Z_T^p\right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R}))^{1-\frac{1}{p}} + C\left(E^{\mathbb{P}} (Z_T^m)^p\right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R}))^{1-\frac{1}{p}} \\
&\le C\left(E^{\mathbb{P}} Z_T^p\right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R}))^{1-\frac{1}{p}}.
\end{aligned}$$

Since $\widetilde{\mathcal{A}}^m(\mathbb{R}) \uparrow \widetilde{\mathcal{A}}(\mathbb{R})$ implies $\mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^m(\mathbb{R})) \to 0$ we get,

$$E^{\bar{\mathbb{P}}} \sup_{0\le t\le T} \left| \int_0^t c(s)\, d\bar{Z}_s - \int_0^t c(s)\, d\bar{Z}_s^m \right| \to 0. \tag{3.24}$$

Similarly,

$$E^{\bar{\mathbb{P}}} \left| \int_0^T h(t)\, d\bar{Z}_t - \int_0^T h(t)\, d\bar{Z}_t^m \right| \to 0. \tag{3.25}$$

By (3.20) and (3.22), the Lipschitz continuity of b and σ in x and μ and the Burkholder-Davis-Gundy inequality,

$$\begin{aligned}
&E^{\bar{\mathbb{P}}} \sup_{0\le t\le T} \left|\bar{X}_t^m - \bar{X}_t\right| \\
&\le E^{\bar{\mathbb{P}}} \int_0^T \left|\bar{X}_t^m - \bar{X}_t\right| dt + \int_0^T C\left(1 + L\left(\mathcal{W}_p(\mu_t^{m*}, \delta_0), \mathcal{W}_p(\mu_t^*, \delta_0)\right)\right) \mathcal{W}_p(\mu_t^{m*}, \mu_t^*)\, dt \\
&\quad + CE^{\bar{\mathbb{P}}} \left(\int_0^T \left|\bar{X}_t^m - \bar{X}_t\right|^2 dt \right)^{\frac{1}{2}} \\
&\quad + C\left(\int_0^T \left(1 + L\left(\mathcal{W}_p(\mu_t^{m*}, \delta_0), \mathcal{W}_p(\mu_t^*, \delta_0)\right)\right)^2 \mathcal{W}_p(\mu_t^{m*}, \mu_t^*)^2\, dt \right)^{\frac{1}{2}} \\
&\quad + E^{\bar{\mathbb{P}}} \sup_{0\le t\le T} \left| \int_0^t c(s)\, d\bar{Z}_s - \int_0^t c(s)\, d\bar{Z}_s^m \right| \\
&\le E^{\bar{\mathbb{P}}} \int_0^T \left|\bar{X}_t^m - \bar{X}_t\right| dt + C\left(\int_0^T \left(1 + L\left(\mathcal{W}_p(\mu_t^{m*}, \delta_0), \mathcal{W}_p(\mu_t^*, \delta_0)\right)\right)^{\frac{p}{p-1}} dt \right)^{1-\frac{1}{p}} \left(\int_0^T \mathcal{W}_p(\mu_t^{m*}, \mu_t^*)^p\, dt \right)^{\frac{1}{p}} \\
&\quad + CT^{\frac{1}{2}} E^{\bar{\mathbb{P}}} \sup_{0\le t\le T} \left|\bar{X}_t^m - \bar{X}_t\right|
\end{aligned}$$

$$
\begin{aligned}
&+C\left(\int_0^T \left(1+L\left(\mathcal{W}_p(\mu_t^{m*},\delta_0),\mathcal{W}_p(\mu_t^*,\delta_0)\right)\right)^{\frac{2p}{p-1}}\,dt\right)^{\frac{1}{2}-\frac{1}{2p}}\left(\int_0^T \mathcal{W}_p(\mu_t^{m*},\mu_t^*)^{2p}\,dt\right)^{\frac{1}{2}}\\
&+E^{\bar{\mathbb{P}}}\sup_{0\le t\le T}\left|\int_0^t c(s)\,d\bar{Z}_s-\int_0^t c(s)\,d\bar{Z}_s^m\right|.
\end{aligned}
\tag{3.26}
$$

Set T_0 such that $CT_0^{\frac{1}{2}}=1/2$. Then (3.26) and Gronwall's inequality yield that

$$
\begin{aligned}
&E^{\bar{\mathbb{P}}}\sup_{0\le t\le T_0}\left|\bar{X}_t^m-\bar{X}_t\right|\\
&\le \widetilde{C}\left(\int_0^T \left(1+L\left(\mathcal{W}_p(\mu_t^{m*},\delta_0),\mathcal{W}_p(\mu_t^*,\delta_0)\right)\right)^{\frac{p}{p-1}}\,dt\right)^{1-\frac{1}{p}}\left(\int_0^T \mathcal{W}_p(\mu_t^{m*},\mu_t^*)^{p}\,dt\right)^{\frac{1}{p}}\\
&\quad+\widetilde{C}\left(\int_0^T \left(1+L\left(\mathcal{W}_p(\mu_t^{m*},\delta_0),\mathcal{W}_p(\mu_t^*,\delta_0)\right)\right)^{\frac{2p}{p-1}}\,dt\right)^{\frac{1}{2}-\frac{1}{2p}}\left(\int_0^T \mathcal{W}_p(\mu_t^{m*},\mu_t^*)^{2p}\,dt\right)^{\frac{1}{2}}\\
&\quad+\widetilde{C}E^{\bar{\mathbb{P}}}\sup_{0\le t\le T}\left|\int_0^t c(s)\,d\bar{Z}_s-\int_0^t c(s)\,d\bar{Z}_s^m\right|,
\end{aligned}
\tag{3.27}
$$

where $\widetilde{C}$ is a constant depending on C and T_0. By (3.24) and the arguments leading to (3.7) in the proof of Lemma 3.4,

$$
\lim_{m\to\infty}E^{\bar{\mathbb{P}}}\sup_{0\le t\le T_0}\left|\bar{X}_t^m-\bar{X}_t\right|=0.
$$

Iterating the same argument, we get

$$
\lim_{m\to\infty}E^{\bar{\mathbb{P}}}\sup_{0\le t\le T}\left|\bar{X}_t^m-\bar{X}_t\right|=0.
\tag{3.28}
$$

By (3.25), (3.28), $\mu^{m*}\to\mu^*$ in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}),d_{M_1})}$ and the same arguments as in the proof of Lemma 3.4, we get

$$
\begin{aligned}
&E^{\bar{\mathbb{P}}}\left(\int_0^T f(t,\bar{X}_t^m,\mu_t^{m*},u)\,\bar{q}_t(du)dt+g(\bar{X}_T^m,\mu_T^{m*})+\int_0^T h(t)\,d\bar{Z}_t^m\right)\\
&\to E^{\bar{\mathbb{P}}}\left(\int_0^T f(t,\bar{X}_t,\mu_t^{*},u)\,\bar{q}_t(du)dt+g(\bar{X}_T,\mu_T^{*})+\int_0^T h(t)\,d\bar{Z}_t\right).
\end{aligned}
$$

This shows that

$$
J(\mu^{m*},\mathbb{P}^m)\to J(\mu^*,\mathbb{P}).
$$

Finally, although g does not have growth of p-th order in x, for each K, by Proposition A.5 we have

$$
\lim_{m\to\infty}\int_\Omega \mathcal{J}(\mu^{m*},\omega)\wedge K\,\mathbb{P}^{m*}(d\omega)=\int_\Omega \mathcal{J}(\mu^*,\omega)\wedge K\,\mathbb{P}^*(d\omega),
$$

from which monotone convergence implies $\liminf_{m\to\infty}J(\mu^{m*},\mathbb{P}^{m*})\ge J(\mu^*,\mathbb{P}^*)$, where $\mathcal{J}(\mu,\omega)$ is defined as in Lemma 3.4. Finally, we obtain that

$$
J(\mu^*,\mathbb{P})=\lim_{m\to\infty}J(\mu^{m*},\mathbb{P}^m)\ge\liminf_{m\to\infty}J(\mu^{m*},\mathbb{P}^{m*})\ge J(\mu^*,\mathbb{P}^*).
$$

□

3.3 Related McKean-Vlasov stochastic singular control problem

The literatures on McKean-Vlasov singular control focuses on necessary conditions for optimality; the existence of optimal control is typically assumed; see e.g. [24]. In this section, we establish the existence of a relaxed control to the following McKean-Vlasov stochastic singular control problem:

$$\min_Z J(Z) = \min_Z E\left[\int_0^T f(t, X_t, \mathcal{L}(X_t), \mathcal{L}(Z_t))\, dt + g(X_T, \mathcal{L}(X_T), \mathcal{L}(Z_T)) + \int_0^T h(t)\, dZ_t\right] \tag{3.29}$$

subject to

$$dX_t = b(t, X_t, \mathcal{L}(X_t), \mathcal{L}(Z_t))\, dt + \sigma(t, X_t, \mathcal{L}(X_t), \mathcal{L}(Z_t))\, dW_t + c(t)\, dZ_t,\ t \in [0, T]. \tag{3.30}$$

The regular control is dropped to simplify notation; it can easily be reintroduced. Without the regular control, the canonical path space is

$$\Omega = \widetilde{\mathcal{D}}(\mathbb{R}) \times \widetilde{\mathcal{A}}(\mathbb{R})$$

Definition 3.15. We call $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, Z)$ a relaxed control to McKean-Vlasov stochastic singular control problem (3.29)-(3.30) if it satisfies items 1, 2 and 3 in Definition 2.1 and

4' $\left(\mathcal{M}^{\mathbb{P},\phi}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}\right)$ is a continuous martingale, where

$$\begin{aligned}\mathcal{M}_t^{\mathbb{P},\phi} = \phi(X_t) &- \int_0^t \phi'(X_s) b(s, X_s, \mathbb{P} \circ X_s^{-1}, \mathbb{P} \circ Z_s^{-1})\, ds \\ &- \frac{1}{2}\int_0^t \phi'(X_s) a(s, X_s, \mathbb{P} \circ X_s^{-1}, \mathbb{P} \circ Z_s^{-1})\, ds \\ &- \int_0^t \phi'(X_{s-}) c(s)\, dZ_s - \sum_{0 \leq s \leq t} (\phi(X_s) - \phi(X_{s-}) - \phi'(X_{s-})\Delta X_s)\,.\end{aligned} \tag{3.31}$$

For each relaxed control $r = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, Z)$, we define the corresponding cost functional by

$$J(r) = E^{\mathbb{P}}\left[\int_0^T f\left(t, X_t, \mathbb{P} \circ X_t^{-1}, \mathbb{P} \circ Z_t^{-1}\right)\, dt + g\left(X_T, \mathbb{P} \circ X_T^{-1}, \mathbb{P} \circ Z_T^{-1}\right) + \int_0^T h(t)\, dZ_t\right]. \tag{3.32}$$

The notion of control rules can be defined as in Definition 2.3. Denote by $\mathcal{R}$ all the control rules. For $\mathbb{P} \in \mathcal{R}$, the corresponding cost functional is defined as in (3.32).

Using straightforward modifications of arguments given in the proof of [20, Proposition 2.6] we see that our optimization problems with relaxed controls and with control rules are equivalent. The next two theorems prove the existence of an optimal control under a finite-fuel constraint. The existence results can then be extended to the the general unconstraint case. We do not give a formal proof as the arguments are exactly the same as in the preceding subsection.

Theorem 3.16. *Suppose $\mathcal{A}_1$, $\mathcal{A}_4$, $\mathcal{A}_5$ hold. Under a finite-fuel constraint on the singular controls, $\mathcal{R} \neq \emptyset$.*

Proof. For each $(\mu^1, \mu^2) \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})) \times \mathcal{P}_p(\widetilde{\mathcal{A}}^m(\mathbb{R}))$, there exists a weak solution to the SDE

$$dX_t = b(t, X_t, \mu_t^1, \mu_t^2)\, dt + \sigma(t, X_t, \mu_t^1, \mu_t^2)\, dW_t + c(t)\, dZ_t,\ t \in [0, T]. \tag{3.33}$$

We define a set-valued map Φ on $\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})) \times \mathcal{P}_p(\widetilde{\mathcal{A}}^m(\mathbb{R}))$ with non-empty convex images by

$$\Phi : (\mu^1, \mu^2) \to \{(\mathbb{P}^{\mu^1,\mu^2} \circ X^{-1}, \mathbb{P}^{\mu^1,\mu^2} \circ Z^{-1}) : (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}^{\mu^1,\mu^2}, X, Z) \text{ is a weak solution to (3.33)}\}.$$

The relatively compactness of $\Phi(\mu^1, \mu^2)$ can be obtained by analogy to Lemma 3.2; closedness follows from Corollary 3.7. The hemi-continuity of Φ can be obtained by analogy to Lemma 3.2 and Lemma 3.8. By analogy to the proof of Theorem 3.11 we can define a non-empty, compact, convex set $\bar{S} \subset \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})) \times \mathcal{P}_p(\widetilde{\mathcal{A}}^m(\mathbb{R}))$ such that $\Phi : \bar{S} \to 2^{\bar{S}}$. Hence, Φ has a fixed point, due to Theorem A.2. This gives us the desired optimal control rule. □

Theorem 3.17. *Suppose $\mathcal{A}_1$, $\mathcal{A}_3$-$\mathcal{A}_5$ hold and that $\mathcal{A}_2$ holds with the continuity of f and g being replaced by lower semi-continuity. Under a finite-fuel constraint, there exist an optimal control rule, that is, there exists $\mathbb{P}^* \in \mathcal{R}$ such that*

$$J(\mathbb{P}^*) \leq J(\mathbb{P}) \quad \textit{for all} \quad \mathbb{P} \in \mathcal{R}.$$

Proof. It is sufficient to prove $\mathcal{R}$ is compact and J is lower semi-continuous. It can be achieved by Corollary 3.7 and by the same way as that in the proof of Lemma 3.4, respectively. □

4 MFGs with regular controls and MFGs with singular controls

In this section we establish two approximation results for a class of MFGs with singular controls under finite-fuel constraints. In Section 4.1 we prove the convergence of (relaxed) solutions to certain MFGs with regular controls to a (relaxed) solution of a related MFG with singular controls, while in Section 4.2 we show how to approximate *any* relaxed solution to a MFG with singular controls by admissible control rules for MFGs with regular controls.

4.1 Solving MFGs with singular controls using MFGs with regular controls

In this section we establish an approximation of (relaxed) solutions for MFGs with singular controls and finite-fuel constraints by (relaxed) solutions to MFGs with regular controls. More precisely, we consider MFGs with singular controls of the form:

$$\begin{cases} \inf_{u,Z} E\left[\int_0^T f(t, X_t, \mu_t, u_t)dt\right] \\ \text{subject to} \\ dX_t = b(t, X_t, \mu_t, u_t)\,dt + \sigma(t, X_t, \mu_t, u_t)dW_t + c(t)\,dZ_t, \;\; t \in [0, T+\epsilon] \\ \mu = \mathcal{L}(X), \end{cases} \tag{4.1}$$

for some fixed $\epsilon > 0$ under the finite-fuel constarint $Z \in \widetilde{\mathcal{A}}^m_{0,T}(\mathbb{R})$. The reason we define the state process on the time interval $[0, T+\epsilon]$ is that we approximate the singular controls by absolutely continuous ones that are most naturally regarded as elements of $\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$ rather than $\widetilde{\mathcal{D}}_{0,T}(\mathbb{R})$. Specifically, we associate with each singular control $Z \in \widetilde{\mathcal{A}}^m_{0,T}(\mathbb{R})$ the sequence of absolutely continuous controls

$$Z^{[n]}_t = n \int_{(t-\frac{1}{n})}^t Z_s\,ds \quad (t \in [0, T], n \in \mathbb{N}). \tag{4.2}$$

These controls take values in $\widetilde{\mathcal{A}}^m_{0,T+\epsilon}(\mathbb{R})$ for all sufficiently large $n \in \mathbb{N}$. Since each $Z^{[n]}$ is absolutely continuous and Z is càdlàg we cannot expect convergence of Z^n to Z in the Skorokhod J_1 topology in general. But we do know that

$$Z^{[n]} \to Z \quad \text{a.s. in} \quad \left(\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R}), d_{M_1}\right).$$

For each n, we consider the following finite-fuel constrained MFGs denoted by $\mathbf{MFG}^{[n]}$:

$$\begin{cases} \inf_{u,Z} E\left[\int_0^T f(t, X_t^{[n]}, \mu_t, u_t)dt\right] \\ \text{subject to} \\ dX_t^{[n]} = b(t, X_t^{[n]}, \mu_t, u_t)dt + \sigma(t, X_t^{[n]}, \mu_t, u_t)dW_t + c(t)\, dZ_t^{[n]},\ t \in [0, T+\epsilon] \\ X_0^{[n]} = 0 \\ Z_t^{[n]} = n \int_{(t-\frac{1}{n})}^t Z_s\, ds \\ \mu = \mathcal{L}(X^{[n]}). \end{cases} \tag{4.3}$$

Definition 4.1. We call the vector $r^n = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, q, Z^{[n]})$ a *relaxed control* if $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, q, Z)$ satisfies 1.-3. in Definition 2.1 with item 4 being replaced by

4′. X is a $\{\mathcal{F}_t, t \geq 0\}$ adapted stochastic process with path in $\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$ such that for each $\phi \in \mathcal{C}_b^2(\mathbb{R}^d)$, $\mathcal{M}^{[n],\mu,\phi}$ is a $\mathbb{P}$ continuous martingale, where

$$\mathcal{M}_t^{[n],\mu,\phi} := \phi(X_t) - \int_0^t \int_U \mathcal{L}\phi(s, X_s, \mu_s, u)\, q_s(du)ds - \int_0^t (\partial_x \phi(X_s))^\top c(s)\, dZ_s^{[n]}, \tag{4.4}$$

with $\mathcal{L}$ defined as in Definition 2.5.

The probability measure $\mathbb{P}$ is called a control rule if $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}, X, q, Z^{[n]})$ is a relaxed control with $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\})$ being the filtered canonical space with

$$\Omega := \widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R}) \times \mathcal{U}(0, T) \times \widetilde{\mathcal{A}}_{0,T}^m(\mathbb{R})$$

and (X, q, Z) being the coordinate processes on $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\})$.

Remark 4.2. If Z is discontinuous at T, then $Z^{[n]}$ may not converge to Z in $\widetilde{\mathcal{D}}_{0,T}(\mathbb{R})$ but only in $\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$. Likewise, the associated sequence of the state processes may only converge in $\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$. The possible discontinuity at the terminal time T is also the reason why there is no terminal cost and no cost from singular control in this section. If we assume that T is always a continuous point, then terminal costs and costs from singular controls are permitted. In this case, one may as well allow unbounded singular controls.

For each fixed n and μ, denote $\mathcal{R}^{[n]}(\mu)$ the set of all the control rules for $\mathbf{MFG}^{[n]}$, and define the cost functional corresponding to the control rule $\mathbb{P} \in \mathcal{R}^{[n]}(\mu)$ by

$$J^{[n]}(\mu, \mathbb{P}) = E^{\mathbb{P}}\left(\int_0^T \int_U f(t, X_t, \mu_t, u)\, q_s(du)dt\right).$$

For each fixed n and μ, denote by $\mathcal{R}^{[n]*}(\mu)$ all the optimal control rules. We can still check that

$$\inf_{\text{relaxed control } r^n} J^{[n]}(\mu, r^n) = \inf_{\mathbb{P} \in \mathcal{R}^{[n]}(\mu)} J^{[n]}(\mu, \mathbb{P}),$$

which implies we can still restrict ourselves to control rules in analyzing $\mathbf{MFG}^{[n]}$.

The proof of the following theorem is very similar to that of Theorem 3.11 and is hence omitted.

Theorem 4.3. *Suppose $\mathcal{A}_1$-$\mathcal{A}_6$ hold. For each n, there exists a relaxed solution $\mathbb{P}^{[n]}$ to $\mathbf{MFG}^{[n]}$.*

By Proposition 3.1, the sequence $\{\mathbb{P}^{[n]}\}_{n\geq 1}$ is relatively compact. Denote its limit (up to a subsequence) by $\mathbb{P}^*$ and set $\mu^* = \mathbb{P}^* \circ X^{-1}$. Then, μ^* is the limit of $\mu^{[n]} := \mathbb{P}^{[n]} \circ X^{-1}$. The following lemma shows that $\mathbb{P}^*$ is admissible.

Lemma 4.4. *Suppose $\mathcal{A}_1$-$\mathcal{A}_2$, $\mathcal{A}_4$-$\mathcal{A}_6$ hold. Then $\mathbb{P}^* \in \mathcal{R}(\mu^*)$.*

Proof. By Proposition 3.5 there exists, for each n, a $(\mathbb{P}^{[n]}, \{\mathcal{F}_t, 0 \le t \le T+\epsilon\})$ continuous process Y^n, such that

$$\mathbb{P}^{[n]}\left(X_\cdot = Y^n_\cdot + \int_0^\cdot c(s)\, dZ^{[n]}_s\right) = 1.$$

Arguing as in the proof of Proposition 3.6, there exits a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{Q})$ supporting random varibales $(\widetilde{X}^n, \widetilde{Y}^n, \widetilde{q}^n, \widetilde{Z}^n)$ and $(\widetilde{X}, \widetilde{Y}, \widetilde{q}, \widetilde{Z})$ such that $(\widetilde{X}^n, \widetilde{Y}^n, \widetilde{q}^n, \widetilde{Z}^n) \to (\widetilde{X}, \widetilde{Y}, \widetilde{q}, \widetilde{Z})$ $\mathbb{Q}$-a.s. and

$$\mathbb{P}^{[n]} \circ (X, Y^n, q, Z)^{-1} = \mathbb{Q} \circ (\widetilde{X}^n, \widetilde{Y}^n, \widetilde{q}^n, \widetilde{Z}^n)^{-1},$$

which implies

$$\mathbb{Q}\left(\widetilde{X}^n_\cdot = \widetilde{Y}^n_\cdot + \int_0^\cdot c(s)\, d\widetilde{Z}^{[n],n}_s\right) = 1, \tag{4.5}$$

where $\widetilde{Z}^{[n],n}_t = n\int_{(t-1/n)}^t \widetilde{Z}^n_s\, ds$. For each fixed $\widetilde{\omega} \in \widetilde{\Omega}$ and for each t which is a continuous point of $\widetilde{Z}(\widetilde{\omega})$, by (B.6) in Proposition B.1, we have

$$\begin{aligned}
\left| n\int_{t-\frac{1}{n}}^t \widetilde{Z}^n_s(\widetilde{\omega})\, ds - \widetilde{Z}_t(\widetilde{\omega})\right| &\le n\int_{t-\frac{1}{n}}^t |\widetilde{Z}^n_s(\widetilde{\omega}) - \widetilde{Z}_s(\widetilde{\omega})|\, ds + n\int_{t-\frac{1}{n}}^t |\widetilde{Z}_s(\widetilde{\omega}) - \widetilde{Z}_t(\widetilde{\omega})|\, ds \\
&\le \sup_{t-\frac{1}{n} \le s \le t} |\widetilde{Z}^n_s(\widetilde{\omega}) - \widetilde{Z}_s(\widetilde{\omega})| + \sup_{t-\frac{1}{n} \le s \le t} |\widetilde{Z}_s(\widetilde{\omega}) - \widetilde{Z}_t(\widetilde{\omega})| \\
&\to 0.
\end{aligned}$$

Then (4.5) and right-continuity of the path yields that

$$\mathbb{Q}\left(\widetilde{X}_\cdot = \widetilde{Y}_\cdot + \int_0^\cdot c(s)\, d\widetilde{Z}_s\right) = 1. \tag{4.6}$$

The desired result can be obtained by the same proof as Proposition 3.6.

□

Remark 4.5. In the above proof, the local uniform convergence near a continuous point is necessary. As stated in Proposition B.1, this is a direct consequence of the convergence in the M_1 topology. Local uniform convergence cannot be guaranteed in the Meyer-Zheng topology. For Meyer-Zheng topology, we only know that convergence is equivalent to convergence in Lebesgue measure but we do not have uniform convergence in general.

We are now ready to state and prove the main result of this section.

Theorem 4.6. *Suppose $\mathcal{A}_1$-$\mathcal{A}_6$ hold. Then $\mathbb{P}^*$ is a relaxed solution to the MFG* (4.1).

Proof. For each $\mathbb{P} \in \mathcal{R}(\mu^*)$ such that $J(\mu^*, \mathbb{P}) < \infty$, on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_t, t \ge 0\}, \widetilde{\mathbb{P}})$ we have,

$$d\widetilde{X}_t = \int_U b(t, \widetilde{X}_t, \mu^*_t, u)\, \widetilde{q}_t(du)dt + \int_U \sigma(t, \widetilde{X}_t, \mu^*_t, u)\, \widetilde{M}(du, dt) + c(t)\, d\widetilde{Z}_t,$$

and $\mathbb{P} = \widetilde{\mathbb{P}} \circ (\widetilde{X}, \widetilde{q}, \widetilde{Z})^{-1}$. Let $\widetilde{Z}^{[n]}_t = n\int_{t-1/n}^t \widetilde{Z}_s\, ds$. By the Lipschitz continuity of the coefficient b and σ, there exists a unique strong solution X^n to the following SDE on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_t, t \ge 0\}, \widetilde{\mathbb{P}})$:

$$dX^n_t = \int_U b(t, X^n_t, \mu^{[n]}_t, u)\, \widetilde{q}_t(du)dt + \int_U \sigma(t, X^n_t, \mu^{[n]}_t, u)\, \widetilde{M}(du, dt) + c(t)\, d\widetilde{Z}^{[n]}_t.$$

For each n, set $\mathbb{P}^n = \widetilde{\mathbb{P}} \circ (X^n, \widetilde{Z})^{-1}$. It is easy to check that $\mathbb{P}^n \in \mathcal{R}^{[n]}(\mu^{[n]})$. Standard estimates yield,

$$\begin{aligned} E^{\widetilde{\mathbb{P}}} \int_0^T |X_t^n - \widetilde{X}_t|^2 \, dt \leq & \; E^{\widetilde{\mathbb{P}}} \int_0^T |Z_t^{[n]} - \widetilde{Z}_t|^2 \, dt \\ & + \; C E^{\widetilde{\mathbb{P}}} \int_0^T \Big(1 + L(W_p(\mu_t^{[n]}, \delta_0), W_p(\mu_t^*, \delta_0))\Big)^2 \mathcal{W}_p(\mu_t^{[n]}, \mu_t^*)^2 \, dt. \end{aligned} \tag{4.7}$$

By Proposition B.1, $Z^{[n]} \to Z$ in M_1 a.s. By the same arguments leading to (3.7) in the proof of Lemma 3.4,

$$E^{\widetilde{\mathbb{P}}} \int_0^T \Big(1 + L(W_p(\mu_t^{[n]}, \delta_0), W_p(\mu_t^*, \delta_0))\Big)^2 \mathcal{W}_p(\mu_t^{[n]}, \mu_t^*)^2 \, dt \to 0.$$

This yields,

$$\lim_{n\to\infty} E^{\widetilde{\mathbb{P}}} \int_0^T |X_t^n - \widetilde{X}_t|^2 \, dt = 0. \tag{4.8}$$

Hence, up to a subsequence, dominated convergence implies

$$\begin{aligned} \lim_{n\to\infty} J^{[n]}(\mu^{[n]}, \mathbb{P}^n) = & \lim_{n\to\infty} E^{\widetilde{\mathbb{P}}} \left[\int_0^T \int_U f(t, X_t^n, \mu_t^{[n]}, u) \, \widetilde{q}_t(du) dt \right] \\ = & E^{\widetilde{\mathbb{P}}} \left[\int_0^T \int_U f(t, X_t, \mu_t^*, u) \, \widetilde{q}_t(du) dt \right] \\ = & J(\mu^*, \mathbb{P}). \end{aligned}$$

Moreover, by Lemma 3.4,

$$\lim_{n\to\infty} J^{[n]}(\mu^{[n]}, \mathbb{P}^{[n]}) = J(\mu^*, \mathbb{P}^*).$$

Altogether, this yields,

$$J(\mu^*, \mathbb{P}) = \lim_{n\to\infty} J^{[n]}(\mu^{[n]}, \mathbb{P}^n) \geq \lim_{n\to\infty} J^{[n]}(\mu^{[n]}, \mathbb{P}^{[n]}) = J(\mu^*, \mathbb{P}^*).$$

□

4.2 Approximating solutions to MFGs with singular controls by control rules of MFGs with regular controls

In this subsection, we show how to approximate a *given* solution to a class of MFGs with only singular controls by a sequence of admissible control rules of MFGs with regular controls. Specifically, we consider the MFG with (only) singular control

$$\begin{cases} \inf_Z E\left[\int_0^T f(t, X_t, \mu_t) dt\right] \\ \text{subject to} \\ dX_t = b(t, X_t, \mu_t) \, dt + \sigma(t, \mu_t) dW_t + c(t) \, dZ_t, \;\; t \in [0, T + \epsilon] \\ X_{0-} = 0 \;\; \text{and} \\ \mu = \mathcal{L}(X). \end{cases} \tag{4.9}$$

where $Z \in \widetilde{\mathcal{A}}^m_{0,T}(\mathbb{R})$ with the canonical path space

$$\Omega := \widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R}) \times \widetilde{\mathcal{A}}^m_{0,T}(\mathbb{R}).$$

Let $\mathbb{P}^*$ be any solution to the above MFG. Since $(\Omega, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P}^*, X, Z)$ satisfies the associated martingale problem, there exist a triplet $(\widehat{X}, \widehat{Z}, B)$ defined on some extension $(\widehat{\Omega}, \{\widehat{\mathcal{F}}_t, t \geq 0\}, \mathbb{Q})$ of the canonical path space, such that

$$\mathbb{P}^* \circ (X, Z)^{-1} = \mathbb{Q} \circ (\widehat{X}, \widehat{Z})^{-1}$$

and

$$\mathbb{Q}\left(\widehat{X}_\cdot = \int_0^\cdot b(s,\widehat{X}_s,\mu_s^*)\,ds + \int_0^\cdot \sigma(s,\mu_s^*)\,dB_s + \int_0^\cdot c(s)\,d\widehat{Z}_s\right) = 1. \tag{4.10}$$

Let $X^{[n]}$ be the unique strong solution of the SDE

$$dX_t^{[n]} = b(t,X_t^{[n]},\mu_t^{[n]})\,dt + \sigma(t,\mu_t^{[n]})\,dB_t + c(t)\,d\widehat{Z}_t^{[n]}, \tag{4.11}$$

where $\widehat{Z}^{[n]}$ is defined by (4.2) and $\mu^{[n]} = \mathbb{Q}\circ(X^{[n]})^{-1}$. One checks immediately that

$$\mathbb{P}^{[n]} := \mathbb{Q}\circ(X^{[n]},\widehat{Z})^{-1} \in \mathcal{R}^{[n]}(\mu^{[n]}).$$

Our goal is to show that the sequence $\{\mathbb{P}^{[n]}\}_{n\geq 1}$ converges to $\mathbb{P}^*$ in $\mathcal{W}_p$ along some subsequence. The proof is based on the following generalization of [35, Theorem 1.1] to McKean-Vlasov case with random noise that extends the second moment convergence (4.7) and (4.8) in the proof of Theorem 4.6.

Proposition 4.7. On some probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t,t\geq 0\},\mathbb{P})$, let X^n and X be the unique strong solution to SDE,

$$dX_t^n = b(t,X_t^n,\mu_t^n)\,dt + \sigma(t,\mu_t^n)\,dB_t + \,dZ_t^n,\ \ t\in[0,\widetilde{T}] \tag{4.12}$$

respectively,

$$dX_t = b(t,X_t,\mu_t)\,dt + \sigma(t,\mu_t)\,dB_t + \,dZ_t,\ \ t\in[0,\widetilde{T}] \tag{4.13}$$

where $\widetilde{T}$ is a fixed positive constant, b and σ satisfy $\mathcal{A}_1$ and $\mathcal{A}_5$, and b is continuous in the time variable. If $Z^n\to Z$ in $(\mathcal{A}^m(0,\widetilde{T}),d_{M_1})$ a.s. and $\mu^n\to\mu$ in $\mathcal{W}_{p,(\mathcal{D}(0,\widetilde{T}),d_{M_1})}$, then

$$\lim_{n\to\infty} E^{\mathbb{P}} d_{M_1}(X^n,X)^p = 0.$$

Proof. By the a.s. convergence of Z^n to Z in M_1, there exists $\underline{\Omega}\subseteq\Omega$ with full measure such that $d_{M_1}(Z^n(\omega),Z(\omega))\to 0$ for each $\omega\in\underline{\Omega}$. Furthermore, by [35, Theorem 1.2], for each $\omega\in\underline{\Omega}$, there exit parameter representations $(u(\omega),r(\omega))\in\Pi_{Z(\omega)}$ and $(u_n(\omega),r_n(\omega))\in\Pi_{Z^n(\omega)}$ for each n, such that

$$\|u_n(\omega)-u(\omega)\|\to 0 \text{ and } \|r_n(\omega)-r(\omega)\|\to 0,$$

where $r_n(\omega)$ and $r(\omega)$ satisfy the following properties:

1. $r_n(\omega,\cdot)$ and $r(\omega,\cdot)$ are absolutely continuous w.r.t. Lebesgue measure with densities $r_n'(\omega,\cdot)$ and $r'(\omega,\cdot)$, respectively;

2. $$\lim_{n\to\infty}\int_0^1 |r_n'(\omega,t)-r'(\omega,t)|\,dt = 0;$$

3. $\|r'(\omega,\cdot)\|\leq 2\widetilde{T}$.

Let $(u_{X^n}(\omega),r_{X^n}(\omega))$ and $(u_X(\omega),r_X(\omega))$ be the parameter representations of $X^n(\omega)$ and $X(\omega)$, respectively. Since $X(\omega)$ (resp. $X^n(\omega)$) jumps at the same time as $Z(\omega)$ (resp. $Z^n(\omega)$), the time change parameter of $X(\omega)$ (resp. $X^n(\omega)$) is $r(\omega)$ (resp. $r_n(\omega)$). In the following, we will drop the dependence on $\omega\in\underline{\Omega}$, if there is no confusion.

By [35, equation (3.1)],

$$u_{X^n}(t) = \int_0^t b(r_n(s),u_{X^n}(s),\mu^n_{r_n(s)})r_n'(s)\,ds + \int_0^{r_n(t)}\sigma(s,\mu_s^n)\,dB_s + u_n(t),$$

and

$$u_X(t) = \int_0^t b(r(s), u_X(s), \mu_{r(s)}) r'(s)\, ds + \int_0^{r(t)} \sigma(s, \mu_s)\, dB_s + u(t).$$

Hence,

$$\begin{aligned}
|u_{X^n}(t) - u_X(t)| \le{} & \left| \int_0^t b(r_n(s), u_{X^n}(s), \mu^n_{r_n(s)}) r_n'(s)\, ds - \int_0^t b(r(s), u_X(s), \mu_{r(s)}) r'(s)\, ds \right| \\
& + \left| \int_0^{r_n(t)} \sigma(s, \mu^n_s)\, dB_s - \int_0^{r(t)} \sigma(s, \mu_s)\, dB_s \right| + |u_n(t) - u(t)| \\
\le{} & \left| \int_0^t b(r_n(s), u_{X^n}(s), \mu^n_{r_n(s)}) r_n'(s)\, ds - \int_0^t b(r_n(s), u_X(s), \mu_{r(s)}) r_n'(s)\, ds \right| \\
& + \left| \int_0^t b(r_n(s), u_X(s), \mu_{r(s)}) r_n'(s)\, ds - \int_0^t b(r(s), u_X(s), \mu_{r(s)}) r_n'(s)\, ds \right| \\
& + \left| \int_0^t b(r(s), u_X(s), \mu_{r(s)}) r_n'(s)\, ds - \int_0^t b(r(s), u_X(s), \mu_{r(s)}) r'(s)\, ds \right| \\
& + \left| \int_0^{r_n(t)} \sigma(s, \mu^n_s)\, dB_s - \int_0^{r(t)} \sigma(s, \mu_s)\, dB_s \right| + |u_n(t) - u(t)| \\
\le{} & 2\widetilde{T} \int_0^t |u_{X^n}(s) - u_X(s)|\, ds \\
& + \int_0^t C \left(1 + L(\mathcal{W}_p(\mu^n_{r_n(s)}, \delta_0), \mathcal{W}_p(\mu_{r(s)}, \delta_0)) \right) \mathcal{W}_p(\mu^n_{r^n(s)}, \mu_{r(s)}) r'(s)\, ds \\
& + 2\widetilde{T} \int_0^t \left| b(r_n(s), u_X(s), \mu_{r(s)}) - b(r(s), u_X(s), \mu_{r(s)}) \right| ds + C \int_0^t |r_n'(s) - r'(s)|\, ds \\
& + \left| \int_0^{r_n(t)} \sigma(s, \mu^n_s)\, dB_s - \int_0^{r_n(t)} \sigma(s, \mu_s)\, dB_s \right| \\
& + \left| \int_0^{r_n(t)} \sigma(s, \mu_s)\, dB_s - \int_0^{r(t)} \sigma(s, \mu_s)\, dB_s \right| + |u_n(t) - u(t)|\,.
\end{aligned} \tag{4.14}$$

Since $\mu^n \to \mu$ in $\mathcal{W}_{p,(\mathcal{D}(0,\widetilde{T}), d_{M_1})}$, Skorokhod's representation theorem yields the existence of a probability space $(\check{\Omega}, \check{\mathcal{F}}, \check{\mathbb{P}})$ carrying random variables $\check{X}^n$ and $\check{X}$, such that $\mu^n = \check{\mathbb{P}} \circ (\check{X}^n)^{-1}$, $\mu = \check{\mathbb{P}} \circ \check{X}^{-1}$ and $d_{M_1}(\check{X}^n, \check{X}) \to 0$ $\check{\mathbb{P}}$ a.s. This yields the following representation for the second term on the right hand side of the above inequality:

$$\begin{aligned}
& \int_0^t \left(1 + L(\mathcal{W}_p(\mu^n_{r_n(s)}, \delta_0), \mathcal{W}_p(\mu_{r(s)}, \delta_0)) \right) \mathcal{W}_p(\mu^n_{r^n(s)}, \mu_{r(s)}) r'(s)\, ds \\
={} & \int_0^t \left[1 + L\left(\left(E^{\check{\mathbb{P}}} |\check{X}^n_{r_n(s)}|^p \right)^{\frac{1}{p}}, \left(E^{\check{\mathbb{P}}} |\check{X}_{r(s)}|^p \right)^{\frac{1}{p}} \right) \right] \left(E^{\check{\mathbb{P}}} |\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p \right)^{\frac{1}{p}} r'(s)\, ds \\
=:{} & K.
\end{aligned}$$

By the definition of M_1 topology,

$$E^{\check{\mathbb{P}}} |\check{X}_{r(s)}|^p \le E^{\check{\mathbb{P}}} d_{M_1}(\check{X}, 0)^p < \infty \tag{4.15}$$

and

$$E^{\check{\mathbb{P}}} |\check{X}^n_{r_n(s)}|^p \le E^{\check{\mathbb{P}}} d_{M_1}(\check{X}^n, 0)^p \to E^{\check{\mathbb{P}}} d_{M_1}(\check{X}, 0)^p < \infty. \tag{4.16}$$

Hence, the local boundedness of the function L yields a constant $\check{C}$ depending on $E^{\check{\mathbb{P}}} d_{M_1}(\check{X}, 0)^p$ such that

$$L\left(E^{\check{\mathbb{P}}} |\check{X}^n_{r_n(s)}|^p, E^{\check{\mathbb{P}}} |\check{X}_{r(s)}|^p \right) \le \check{C}. \tag{4.17}$$

By (4.17), an application of Hölder's inequality yields

$$
\begin{aligned}
K &\le \left[\int_0^t \left[1 + L\left(E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)}|^p, E^{\check{\mathbb{P}}}|\check{X}_{r(s)}|^p\right)\right]^{\underline{p}} r'(s)\,ds\right]^{\frac{1}{\underline{p}}} \left[\int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s)\,ds\right]^{\frac{1}{p}} \\
&\le \widetilde{C}\left[\int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s)\,ds\right]^{\frac{1}{p}},
\end{aligned}
\tag{4.18}
$$

where $\underline{p}$ is the Hölder conjugate of p. Since r_n and r are defined on $\underline{\Omega}$, they can be considered as constants under the measure $\check{\mathbb{P}}$. If $r(s)$ is a jump time of Z (and thus a jump time of X), then $r(s)$ is a constant, thus $r'(s) = 0$. Therefore,

$$
\begin{aligned}
&\int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s)\,ds \\
&= \int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s) 1_{\{X_{r(s)} = X_{r(s)-}\}}\,ds + \int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s) 1_{\{X_{r(s)} \neq X_{r(s)-}\}}\,ds \\
&= \int_0^t E^{\check{\mathbb{P}}}|\check{X}^n_{r_n(s)} - \check{X}_{r(s)}|^p r'(s) 1_{\{X_{r(s)} = X_{r(s)-}\}}\,ds \\
&= E^{\check{\mathbb{P}}}\int_0^{r(t)} |\check{X}^n_{r_n \circ r^{-1}(s)} - \check{X}_s|^p 1_{\{X_s = X_{s-}\}}\,ds \\
&= E^{\check{\mathbb{P}}}\int_0^{r(t)} |\check{X}^n_{r_n \circ r^{-1}(s)} - \check{X}_s|^p 1_{\{X_s = X_{s-}, \check{X}_s = \check{X}_{s-}\}}\,ds.
\end{aligned}
\tag{4.19}
$$

By (4.15) and (4.16), dominated convergence and (B.6) yield that

$$
\begin{aligned}
&E^{\mathbb{P}}E^{\check{\mathbb{P}}}\int_0^{r(t)} |\check{X}^n_{r_n \circ r^{-1}(s)} - \check{X}_s|^p 1_{\{X_s = X_{s-}, \check{X}_s = \check{X}_{s-}\}}\,ds \\
&\le E^{\mathbb{P}}E^{\check{\mathbb{P}}}\int_0^{\widetilde{T}} \left|\check{X}^n_{r_n \circ r^{-1}(s)} - \check{X}_s\right|^p 1_{\{X_s = X_{s-}, \check{X}_s = \check{X}_{s-}\}}\,ds \to 0.
\end{aligned}
$$

By (4.14), (4.18) and (4.19), we have

$$
\begin{aligned}
E^{\mathbb{P}}\|u_{X^n} - u_X\|^p \le{}& CE^{\mathbb{P}}\int_0^1 |u_{X^n}(s) - u_X(s)|^p\,ds + CE^{\mathbb{P}}\left|\int_0^1 |r_n'(s) - r'(s)|\,ds\right|^p \\
&+ CE^{\mathbb{P}}\int_0^1 \left|b(r_n(s), u_X(s), \mu_{r(s)}) - b(r(s), u_X(s), \mu_{r(s)})\right|^p\,ds \\
&+ CE^{\mathbb{P}}\sup_{0\le t\le \widetilde{T}}\left|\int_0^t \sigma(s, \mu^n_s)\,dB_s - \int_0^t \sigma(s, \mu_s)\,dB_s\right|^p \\
&+ CE^{\mathbb{P}}\sup_{0\le t\le 1}\left|\int_0^{r_n(t)} \sigma(s, \mu_s)\,dB_s - \int_0^{r(t)} \sigma(s, \mu_s)\,dB_s\right|^p \\
&+ CE^{\mathbb{P}}\|u_n - u\|^p + CE^{\mathbb{P}}E^{\check{\mathbb{P}}}\int_0^{\widetilde{T}} \left|\check{X}^n_{r_n \circ r^{-1}(s)} - \check{X}_s\right|^p 1_{\{X_s = X_{s-}, \check{X}_s = \check{X}_{s-}\}}\,ds.
\end{aligned}
\tag{4.20}
$$

Therefore, Gronwall's inequality and dominated convergence imply

$$
E^{\mathbb{P}}\|u_{X^n} - u_X\|^p \to 0.
$$

□

Corollary 4.8. Under the assumptions of Proposition 4.7, along a subsequence

$$
\mathbb{P}^{[n]} \to \mathbb{P}^* \quad \text{in } \mathcal{W}_p.
$$

Proof. We have that

$$\int_0^{\cdot} c(t)\, d\widehat{Z}_t^{[n]} = \int_0^{\cdot} c^+(t)\, d\widehat{Z}_t^{[n]} - \int_0^{\cdot} c^-(t)\, d\widehat{Z}_t^{[n]},$$

where a.s. in $(\widetilde{\mathcal{A}}^m_{0,T+\epsilon}(\mathbb{R}), d_{M_1})$,

$$\int_0^{\cdot} c^+(t)\, d\widehat{Z}_t^{[n]} \to \int_0^{\cdot} c^+(t)\, d\widehat{Z}_t \quad \text{and} \quad \int_0^{\cdot} c^-(t)\, d\widehat{Z}_t^{[n]} \to \int_0^{\cdot} c^-(t)\, d\widehat{Z}_t.$$

Since $\int_0^{\cdot} c^+(t)\, d\widehat{Z}_t$ and $\int_0^{\cdot} c^-(t)\, d\widehat{Z}_t$ never jump at the same time, Proposition B.8 implies that

$$\int_0^{\cdot} c(t)\, d\widehat{Z}_t^{[n]} \to \int_0^{\cdot} c(t)\, d\widehat{Z}_t$$

a.s. in $(\widetilde{\mathcal{A}}^m_{0,T+\epsilon}(\mathbb{R}), d_{M_1})$. Hence, by Proposition 4.7,

$$E^{\mathbb{Q}} d_{M_1}(X^{[n]}, \widehat{X})^p \to 0.$$

So up to a subsequence, $d_{M_1}(X^{[n]}, \widehat{X})^p \to 0$ $\mathbb{Q}$ a.s., which implies that $(X^{[n]}, \widehat{Z}) \to (\widehat{X}, \widehat{Z})$ a.s. in the M_1 product topology. For any nonnegative continuous function ϕ satisfying

$$\phi(x, z) \le C(1 + d_{M_1}(x, 0)^p + d_{M_1}(z, 0)^p),$$

the uniform integrability of $d_{M_1}(X^{[n]}, 0)^p$ and $d_{M_1}(\widehat{Z}, 0)^p$ yields $E^{\mathbb{Q}}\phi(X^{[n]}, \widehat{Z}) \to E^{\mathbb{Q}}\phi(\widehat{X}, \widehat{Z})$. This implies $\mathbb{Q} \circ (X^{[n]}, \widehat{Z})^{-1} \to \mathbb{Q} \circ (\widehat{X}, \widehat{Z})^{-1}$ in $\mathcal{W}_p$ by Proposition A.5, that is, $\mathbb{P}^{[n]} \to \mathbb{P}^*$ in $\mathcal{W}_p$. □

A Useful Notions and Propositions

In this appendix we summarize some results on correspondences and weak convergence of measures that are frequently used throughout the paper.

A.1 Maximum and fixed-point theorem

Theorem A.1. *[Berge's Maximum Principle] Let $\psi : X \to 2^Y$ be a continuous set-valued function between topological spaces with nonempty- and compact-values, and $f : Gr\psi \to \mathbb{R}$ be a continuous function. $\nu(x) := arg\max_{y\in\psi(x)} f(x, y)$. Then*

1. *ν has nonempty and compact values*
2. *if Y is Hausdorff, then ν is upper hemicontinuous.*

Theorem A.2. *[Kakutani-Fan-Glicksberg fixed point theorem] Given a locally convex topology vector space Y, S, a subset of Y, is convex, nonempty and compact. Let $\psi : S \to 2^S$ be a set-valued function, which is upper hemi-continuous. If ψ is nonempty-, convex- and compact-valued, then ψ has a fixed point, i.e., $\exists\ y \in S$ such that $y \in \psi(y)$.*

Definition A.3. [hemi-continuity in metric space] Let $\mathcal{A}$ and $\mathcal{B}$ be two metric spaces. The set-valued function $\psi : \mathcal{A} \to 2^{\mathcal{B}}$ is closed valued. Then we say ψ is upper hemi-continuous if whenever $a_n \to a$ in $\mathcal{A}$ and $b_n \in \psi(a_n)$, there exist a subsequence b_{n_k} of b_n such that the limit of b_{n_k} belongs to $\psi(a)$. ψ is called lower hemi-continuous if whenever $a_n \to a$ and $b \in \psi(a)$, there exists $b_{n_k} \in \psi(a_{n_k})$ such that $b_{n_k} \to b$. If ψ is both upper and lower hemi-continuous, we say ψ is hemi-continuous.

A.2 Wasserstein distance and representation of martingales

Definition A.4. Let (E, d) be a metric space. Denote by $\mathcal{P}_p(E)$ the class of all probability measures on E with finite moment of p-th order. The *p-th Wasserstein metric* on $\mathcal{P}_p(E)$ is defined by:

$$\mathcal{W}_{p,(E,d)}(\mathbb{P}_1, \mathbb{P}_2) = \inf \left\{ \left(\int_{E\times E} d(x,y)^p \, \gamma(dx,dy) \right)^{\frac{1}{p}} : \gamma(dx,E) = \mathbb{P}_1(dx), \gamma(E,dy) = \mathbb{P}_2(dy) \right\}. \quad \text{(A.1)}$$

The set $\mathcal{P}_p(E)$ endowed with the Wasserstein distance is denoted by $\mathcal{W}_{p,(E,d)}$ or $\mathcal{W}_{p,E}$ or $\mathcal{W}_p$ if there is no risk of confusion about the underlying state space or distance.

Proposition A.5. ([36, Definition 6.8]) Let (E, d) be a Polish space and $p \geq 1$. Let $\{\mu_n\}_{n\geq 1}$ be a sequence of probability measures in $\mathcal{P}_p(E)$ and $\mu \in \mathcal{P}_p(E)$. Then the following statements are equivalent:

(1). $\mu_n \to \mu$ in $\mathcal{W}_{p,(E,d)}$;

(2). $\mu_n \to \mu$ in weak sense and moreover, for some (and thus for any) $y_0 \in E$,

$$\lim_{K\to\infty} \sup_n \int_{\{y\in E: d(y,y_0)>K\}} d(y,y_0)^p \, \mu_n(dy) = 0; \quad \text{(A.2)}$$

(3). $\mu_n \to \mu$ in weak sense and moreover, for some (and thus for any) $y_0 \in E$,

$$\lim_{n\to\infty} \int_E d(y,y_0)^p \, \mu_n(dy) = \int_E d(y,y_0)^p \, \mu(dy); \quad \text{(A.3)}$$

(4). for each continuous φ satisfies $\varphi(y) \leq C(1 + d(y,y_0)^p)$ for some (and thus for any) $y_0 \in E$, it holds

$$\int_E \varphi(y) \, \mu_n(dy) \to \int_E \varphi(y) \, \mu(dy). \quad \text{(A.4)}$$

It is well known [28, Theorem III-10] that for every continuous square integrable martingale m with quadratic variation process $\int_0^\cdot \int_U a(s,x,\mu,u) \, q_s(du)ds$, on some extension of the original probability space, there exists a martingale measure M with intensity $q_s(du)ds$ such that $m_\cdot = \int_0^\cdot \int_U \sigma(t,x,\mu,u) \, M(du,dt)$. This directly leads to the following proposition, which is frequently used in the main text.

Proposition A.6. The existence of solution $\mathbb{P}$ to the martingale problem (2.3) is equivalent to the existence of the weak solution to the following SDE

$$d\bar{X}_t = \int_U b(t, \bar{X}_t, \mu_t, u) \, \bar{q}_s(du)ds + \int_U \sigma(t, \bar{X}_t, \mu_t, u) \, \bar{M}(du,dt) + c(t) \, d\bar{Z}_t, \quad \text{(A.5)}$$

where $\bar{X}$, $\bar{M}$ and $\bar{Z}$ are defined on some extension $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ and $\bar{M}$ is a martingale measure with intensity $\bar{q}$. Moreover, the two solutions are related by $\mathbb{P} = \bar{\mathbb{P}} \circ (\bar{X}, \bar{q}, \bar{Z})^{-1}$.

B Strong M_1 Topology in Skorokhod Space

In this section, we summarise some definitions and properties about strong Skorokhod M_1 topology. For more details, please refer to Chapter 3, 11 and 12 in [37]. Note that in [37] two M_1 topologies are introduced, the strong one and the weak one. In this paper, we only apply the strong one. So without abuse of terminologies, we just take M_1 topology for short.

For $x \in \mathcal{D}(0,T)$, denote by $Disc(x)$ the set of discontinuous points of x. Note that on $[0,T]$, $Disc(x)$ is at most countable. Define the thin graph of x as

$$G_x = \{(z,t) \in \mathbb{R}^d \times [0,T] : z \in [x_{t-}, x_t]\}, \tag{B.1}$$

where x_{t-} is the left limit of x at t and $[a,b]$ means the line segment between a and b, i.e., $[a,b] = \{\alpha a + (1-\alpha)b : 0 \le \alpha \le 1\}$. On the thin graph, we define an order relation. For each pair $(z_i, t_i) \in G_x$, $i = 1, 2$, $(z_1, t_1) \le (z_2, t_2)$ if either of the following holds: (1) $t_1 < t_2$; (2) $t_1 = t_2$ and $|z_1 - x_{t_1-}| < |z_2 - x_{t_2-}|$.

Now we define the parameter representation, on which the M_1 topology depends. The mapping pair (u, r) is called a parameter representation if $(u, r) : [0,1] \to G_x$, which is continuous and nondecreasing w.r.t. the order relation defined above. Denote by Π_x all the parameter representations of x. Let

$$d_{M_1}(x_1, x_2) = \inf_{(u_i, r_i) \in \Pi_{x_i}, i=1,2} ||u_1 - u_2|| \vee ||r_1 - r_2||. \tag{B.2}$$

It can be shown that d_{M_1} is a metric on $\mathcal{D}(0,T)$ such that $\mathcal{D}(0,T)$ is a Polish space. The topology induced by d_{M_1} is called M_1 topology.

For each $t \in [0,T]$ and $\delta > 0$, the oscillation function around t is defined as

$$\bar{v}(x,t,\delta) = \sup_{0 \vee (t-\delta) \le t_1 \le t_2 (t+\delta) \wedge T} |x_{t_1} - x_{t_2}|, \tag{B.3}$$

and the so called strong M_1 oscillation function is defined as

$$w_s(x,t,\delta) = \sup_{0 \vee (t-\delta) \le t_1 < t_2 < t_3 \le (t+\delta) \wedge T} |x_{t_2} - [x_{t_1}, x_{t_3}]|, \tag{B.4}$$

where $|x_{t_2} - [x_{t_1}, x_{t_3}]|$ is the distance from x_{t_2} to the line segment $[x_{t_1}, x_{t_3}]$. Moreover,

$$w_s(x,\delta) := \sup_{0 \le t \le T} w_s(x,t,\delta). \tag{B.5}$$

Now we present the characterizations of M_1 convergence, continuity of convergence, relative compactness and tightness.

Proposition B.1. The following statements about the characterization of M_1 convergence are equivalent,

1. $x^n \to x$ in M_1 topology;
2. there exist $(u, r) \in \Pi_x$ and $(u^n, r^n) \in \Pi_{x^n}$ for each n such that

$$\lim_{n\to\infty} \|u^n - u\| \vee \|r^n - r\| = 0;$$

3. $x_n(t) \to x(t)$ for each $t \in [0,T] \setminus Disc(x)$ including 0 and T, and

$$\lim_{\delta \to 0} \overline{\lim}_{n\to\infty} w_s(x^n, \delta) = 0.$$

Moreover, each one of the above three items implies the local uniform convergence of x^n to x at each continuous point of x, that is, for each $t \notin Disc(x)$, there holds

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{t-\delta \le s \le t+\delta} |x_n(s) - x(s)| = 0. \tag{B.6}$$

Proposition B.2. A subset A of $(\mathcal{D}(0,T), d_{M_1})$ is relatively compact w.r.t. M_1 topology if and only if

$$\sup_{x\in A} ||x|| < \infty \tag{B.7}$$

and

$$\lim_{\delta\downarrow 0}\sup_{x\in A} w'_s(x,\delta) = 0, \tag{B.8}$$

where

$$w'_s(x,\delta) = w_s(x,\delta) \vee \bar{v}(x,0,\delta) \vee \bar{v}(x,T,\delta). \tag{B.9}$$

In [37], it is assumed that $x_{0-} = x_0$, which implies there is no jump at the initial time. For singular control problems it is natural to admit jumps a the initial time. It is also implied by Proposition B.2 that the terminal time T is a continuous point of $x \in \mathcal{D}(0,T)$. This, too, is not appropriate for singular control problems. In order to adapt the relative compactness criteria stated in Proposition B.2 to functions with jumps at 0 and T, we work on the extended state spaces $\widetilde{\mathcal{D}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}(\mathbb{R})$. Convergence in $\widetilde{\mathcal{D}}(\mathbb{R})$ can be defined as convergence in $\mathcal{D}(\mathbb{R})$, where a sequence $\{x^n, n \geq 1\}$ converges to x in $\mathcal{D}(\mathbb{R})$ if and only if the sequences $\{x^n|_{[a,b]}, n \geq 1\}$ converge to $x|_{[a,b]}$ for all $a < b$ at which x is continuous; see [37, Chapter 3].

Relative compactness of a sequence $\{x^n, n \geq 1\} \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ is equivalent to that of the sequence $\{x^n|_{[a,b]}, n \geq 1\} \subseteq \mathcal{D}[a,b]$ for any $a < 0$ and $b > T$. Specifically, we have the following result.

Proposition B.3. The sequence $\{x^n, n \geq 1\} \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ is relatively compact if and only if

$$\sup_n ||x_n|| < \infty \quad \text{and} \quad \lim_{\delta\downarrow 0}\sup_{x\in A} \widetilde{w}_s(x,\delta) = 0, \tag{B.10}$$

where the modified oscillation function $\widetilde{w}_s$ is defined as

$$\widetilde{w}_s(x,\delta) = w_s(x,\delta) + \sup_{0\leq s<t\leq \delta} |x_s - [0, x_t]|. \tag{B.11}$$

We notice that the modified oscillation function $\widetilde{w}_s$ is defined in terms of the original oscillation function w_s and the line segment (if it exists) between $0-$ and 0[2]. As such the space $\widetilde{\mathcal{D}}(\mathbb{R})$ is isomorphic to the space

$$\mathcal{D}_{0,T} := \{(y, x|_{[0,T]}) \in \mathbb{R}^d \times \mathcal{D}(0,T) : x \in \mathcal{D}(\mathbb{R}), x_{0-} = y\}.$$

On $\mathcal{D}_{0,T}$, we can construct the modified thin graph by taking the segment (if it exists) between $0-$ and 0 into consideration. In the same spirit of M_1 metric on the thin graph, we can define the modified M_1 metric (we call it $\widetilde{M_1}$) on the modified thin graph. Therefore, we have the following characterization of convergence in $(\mathcal{D}_{0,T}, \widetilde{M_1})$.

Lemma B.4. *$(y^n, x^n|_{[0,T]}) \to (y, x|_{[0,T]})$ in $\widetilde{M_1}$ on $\mathcal{D}_{0,T}$ if and only if $x^n_t \to x_t$ for each $t \in [0,T]\backslash Disc(x)$ including T, $y^n \to y$, and*

$$\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} \widetilde{w}_s(x^n, \delta) = 0.$$

For each set $A \subseteq \mathcal{D}_{0,T}$, define

$$\widetilde{A} := \{\widetilde{x} \in \widetilde{\mathcal{D}}(\mathbb{R}) : \text{for some } (y, x|_{[0,T]}) \in A,\ \widetilde{x}|_{[0,T]} = x|_{[0,T]},\ \widetilde{x}_t = y \text{ if } t < 0 \text{ and } \widetilde{x}_t = x_T \text{ if } t > T\}.$$

It is easy to show that the $(\mathcal{D}_{0,T}, \widetilde{M_1})$ relative compactness of A is equivalent to the $(\widetilde{\mathcal{D}}(\mathbb{R}), M_1)$ relative compactness of $\widetilde{A}$. In this way, we could consider $\mathcal{D}_{0,T}$ as the canonical space as well.

[2]Due to the right-continuity of the elements in $\widetilde{\mathcal{D}}(\mathbb{R})$ it is not necessary to consider the line segment between $T-$ and T separately.

Corollary B.5. Let $A = \{z \in \widetilde{\mathcal{A}}(\mathbb{R}) : z_T \leq K\}$ for some positive constant K, then A is $(\widetilde{\mathcal{D}}(\mathbb{R}), M_1)$ compact.

Proof. This follows from Proposition B.3 as $w_s(z, t, \delta) = 0$ for each $z \in A$, $t \in \mathbb{R}$ and $\delta > 0$. □

We now state conditions for the tightness of probability measures on $\mathcal{D}(0, T)$ and $\widetilde{\mathcal{D}}(\mathbb{R})$, respectively.

Proposition B.6. A sequence of probability measures $\{\mathbb{P}_n\}_{n\geq 1}$ on $(\mathcal{D}(0, T), d_{M_1})$ is tight if and only if

(1) for each $\epsilon > 0$, there exits c large enough such that

$$\sup_n \mathbb{P}_n(||x|| > c) < \epsilon; \tag{B.12}$$

(2) for each $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ small enough such that

$$\sup_n \mathbb{P}_n(w'_s(x, \delta) \geq \eta) < \epsilon. \tag{B.13}$$

By Proposition B.6 and Proposition B.3, we have the following tightness criteria for probability measures on $\widetilde{\mathcal{D}}(\mathbb{R})$.

Corollary B.7. A sequence of probability measures $\{\mathbb{P}_n\}_{n\geq 1}$ on $\widetilde{\mathcal{D}}(\mathbb{R})$ is tight if and only if (B.12) holds, and for each $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ small enough such that

$$\sup_n \mathbb{P}_n(\widetilde{w}_s(x, \delta) \geq \eta) < \epsilon. \tag{B.14}$$

The following proposition shows that if two M_1 limits do not jump at the same time, then the M_1 convergence preserves by the addition operation.

Proposition B.8. If $x^n \to x$ and $y^n \to y$ in $(\mathcal{D}(0, T), d_{M_1})$, and $Disc(x) \cap Disc(y) = \emptyset$, then

$$x^n + y^n \to x + y \ in \ M_1. \tag{B.15}$$

References

[1] S. Ahuja, *Wellposedness of Mean Field Games with Common Noise under a Weak Monotonicity Condition*, SIAM J. Control Optim., **51** (2016), 3048.

[2] A. Bensoussan, K. Sung and P. Yam, *Linear Quadratic Time-Inconsistent Mean Field Games*, Dyn. Games Appl., **3** (2013), 537552.

[3] A. Bensoussan, K. Sung, P. Yam and S. Yung, *Linear Quadratic Mean Field Games*, J. Optim. Theory Appl., **169** (2016), 496529.

[4] R. Carmona, F. Delarue, *Probabilistic Analysis of Mean-Field Games*, SIAM J. Control Optim., **51** (2013), 2705-2734.

[5] R. Carmona, F. Delarue, and A. Lachapelle, *Control of McKeanVlasov Dynamics versus Mean Field Games*, Math. Financ. Econ., **7** (2013), 131166.

[6] R. Carmona, F. Delarue and D. Lacker, *Mean Field Games with Common Noise*, to appear in Ann. Probab., (2016).

[7] R. Carmona, J. Fouque, and L. Sun, *Mean Field Games and Systemic Risk*, Commun. Math. Sci., **13** (2015), 911-933.

[8] R. Carmona, D. Lacker, *A Probabilistic Weak Formulation of Mean Field Games and Applications*, Ann. Appl. Probab., **25** (2015), 1189-1231.

[9] R. Carmona, D. Lacker, *Mean Field Games of Timing and Models for Bank Runs*, arXiv:1606.03709v2, 2016.

[10] R. Carmona, X. Zhu, *A Probabilistic Approach to Mean Field Games with Major and Minor Players*, Ann. Appl. Probab., **26** (2016), 1535-1580.

[11] P. Chan, R. Sircar, *Bertrand and Cournot Mean Field Games*, App. Math. Optim., **71** (2015), 533-569.

[12] P. Chan, R. Sircar, *Fracking, Renewables* & *Mean Field Games*, to appear in SIAM Rev., (2016).

[13] R. Elie, T. Mastrolia and D Possama, *A Tale of A Principal and Many Many Agents*, arXiv:1608.05226, 2016.

[14] H. Föllmer, *Macroscopic Convergence of Markov Chains on Infinite Product Spaces*, in: Colloquium on Random Fields, Ungarn. Math. Soc. Janos Bolyai, Esztergom (1979), 363-371.

[15] H. Föllmer, U. Horst, *Convergence of Locally and Globally Interacting Markov Chains*, Stoch. Proc. Appl., **96** (2001) 99-121.

[16] D. Gomes, J. Saúde, *Mean Field Games Models-A Brief Survey*, Dyn. Games Appl., **4** (2014), 110-154.

[17] P. Graewe, U. Horst and E. Séré, *Smooth Solutions to Portfolio Liquidation Problems under Price-Sensitive Market Impact*, arXiv:1309.0474, 2016.

[18] O. Gueant, J. Lasry and P. Lions, *Mean Field Games and Applications*, Paris-Princeton Lectures on Mathematical Finance, Springer, 205-266, 2010.

[19] U. Haussmann, J. Lepeltier, *On the Existence of Optimal Controls*, SIAM. J. Control Optim. **28** (1990), 851-902.

[20] U. Haussmann, W. Suo, *Singular Optimal Stochastic Controls I: Existence*, SIAM. J. Control Optim. **33** (1995), 916-936.

[21] U. Horst, F. Naujokat, *When to Cross the Spread: Trading in Two-Sided Limit Order Books*, SIAM J. Financ. Math., **5** (2014), 278-315.

[22] U. Horst, J. Scheinkman, *Equilibria in Systems of Social Interactions*, J. Econ. Theory, **130** (2006), 44-77.

[23] U. Horst, J. Scheinkman, *A Limit Theorem for Systems of Social Interactions*, J. Math. Econ., **45** (2009), 609-623.

[24] Y. Hu, B. Øksendal, and A. Sulem, *Singular Mean-Field Control Games with Applications to Optimal Harvesting and Investment Problems*, arXiv:1406.1863, 2014.

[25] M. Huang, R. Malhamé, and P. Caines, *Large Population Stochastic Dynamic Games: Closed-Loop McKean-Vlasov Systems and the Nash Certainty Equivalence Principle*, Commun. Inf. Syst., **6** (2006), 221-252.

[26] S. Jaimungal and M. Nourian, *Mean-Field Game Strategies for A Major-Minor Agent Optimal Execution Problem*, Available at SSRN 2578733, 2015.

[27] I. Karatzas, S.E. Shreve, *Brownian Motion and Stocahstic Calculus*, Springer-Verlag, New York, 1991.

[28] N. El Karoui, S. M el eard, *Martingale Measures and Stochastic Calculus*, Probab. Theory Related Fields, **84** (1990), 83-101.

[29] N. El Karoui, D. Huu Nguyen and M. Jeanblanc-Picqué, *Compactification Methods in the Control of Degenerate Diffusions: Exitence of An Optimal Control*, Stochastics, **20** (1987), 196-219.

[30] P. Kratz, *An Explicit Solution of A Nonlinear-Quadratic Constrained Stochastic Control Problem with Jumps: Optimal Liquidation in Dark Pools with Adverse Selection*, Math. Oper. Res. **39** (2014), 1198-1120.

[31] D. Lacker, *Mean Field Games via Controlled Martingale Problems: Existence of Markovian Equilibria*, Stoch. Proc. Appl. **125** (2015), 2856-2894.

[32] D. Lacker, *A General Characterization of the Mean Field Limit for Stochastic Differential Games*, Probab. Theory Related Fields, **165** (2016), 581-648.

[33] J. Lasry and P. Lions, *Mean Field Games*, Jpn. J. Math., **2** (2007), 229-260.

[34] M. Nutz, *A Mean Field Game of Optimal Stopping*, arXiv:1605.09112, 2016.

[35] G. Pang, W. Whitt, *Continuity of a queueing intergral representation in the M_1 topology*, Ann. Appl. Prob. **20** (2010), 214-237.

[36] C. Villani, *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, **338**, Springer-Verlag, 2009.

[37] W. Whitt, *Stochastic Process Limits: An Introduction to Stochastic Process Limits and Their Application to Queues*, Springer-Verlag, New York, 2002.

[38] L. Young, *Generalized Curves and the Existence of An Attained Absolute Minimum in the Calculus of Variations*, CR. Soc. Sci. Letters Varsovie, III **30** (1937):212-234