Guanxing $\operatorname{Fu}^{\dagger}$ and Ulrich Horst[‡]

August 4, 2017

Abstract

This paper establishes the existence of relaxed solutions to mean field games (MFGs for short) with singular controls. We also 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 [20, 27], 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 minimizes a cost functional of the form

$$J^{i}(u) = E\left[\int_{0}^{T} f(t, X_{t}^{i}, \bar{\mu}_{t}^{N}, u_{t}^{i})dt + g(X_{T}^{i}, \bar{\mu}_{T}^{N})\right]$$
(1.1)

subject to the state dynamics

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

Here $W^1, ..., W^N$ are independent Brownian motions defined on some underlying filtered probability space, $u = (u^1, \dots, u^N)$, $u^i = (u^i_t)_{t \in [0,T]}$ 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) and the SFB 649 *Economic Risk* is gratefully acknowledged. We thank an anonymous referee for many comments and suggestions that greatly helped to improve the presentation of the results.

[†]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, 20] 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 measurevalued 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 measurevalued process such that the distribution 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 optimization and fixed point problem of the form:

$$\begin{cases} 1. & \text{fix a deterministic function } t \in [0, T] \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d); \\ 2. & \text{solve the corresponding stochastic control problem :} \\ & \inf_u E\left[\int_0^T f(t, X_t, \mu_t, u_t) \, dt + g(X_T, \mu_T)\right], \\ & \text{subject to} \\ & dX_t = b(t, X_t, \mu_t, u_t) \, dt + \sigma(t, X_t, \mu_t, u_t) \, dW_t \\ & X_0 = x_0, \\ 3. & \text{solve the fixed point problem: } Law(X) = \mu, \end{cases}$$

$$(1.3)$$

where $\mathcal{P}(\mathbb{R}^d)$ is the space of probability measures on \mathbb{R}^d and Law(X) denotes the law of the process X.

There are essentially three approaches to solve mean field games. In their original paper [27], 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, 3, 5]. Among them, [3, 5] consider linear-quadratic MFGs, while [1, 9] 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 [26]. 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 [31]. They were later applied to stochastic control in, e.g. [14, 15, 25], to MFGs in [26], and to MFGs with common noise in [6].

Applications of MFGs range from models of optimal exploitation of exhaustible resources [10, 12] to systemic risk [7], and from principal-agent problems [11] to problems of optimal trading under market impact [8, 22]. Motivated by possible applications to optimal portfolio liquidation under strategic interaction that allow for both block trades and absolutely continuous trades as in [16], this paper provides a

¹The idea of decoupling local from global dynamic in large population has been applied to equilibrium models of social interaction in e.g. [17, 18].

probabilistic framework for analyzing MFGs with singular controls. Extending [26], we consider MFGs with singular controls of the form

 $\begin{cases} 1. & \text{fix a deterministic function } t \in [0,T] \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d); \\ 2. & \text{solve the corresponding stochastic singular control problem :} \\ & \inf_{u,Z} E\left[\int_0^T f(t, X_t, \mu_t, u_t) \, dt + g(X_T, \mu_T) + \int_0^T h(t) \, dZ_t\right], \\ & \text{subject to} \\ & dX_t = b(t, X_t, \mu_t, u_t) \, dt + \sigma(t, X_t, \mu_t, u_t) \, dW_t + c(t) \, dZ_t, \\ 3. & \text{solve the fixed point problem: } Law(X) = \mu, \end{cases}$ (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 of a function at given points in time by 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 Whitt [30].

To the best of our knowledge, ours is the first paper to establish the existence of solutions results to MFGs with singular controls². 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 [15]. 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 [15], 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.

Our second main contributions are two approximation results that allow 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 controls 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

 $^{^{2}}$ The recent paper [13] only considers absolutely continuous singular controls. Our notion of singular controls is more general.

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 metric space (E, ϱ) 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 $\mathcal{P}_0(E)$. The set $\mathcal{P}_p(E)$ is endowed with the Wasserstein distance $\mathcal{W}_{p,(E,\varrho)}$; see Definition A.1. For a given interval \mathbb{I} we denote by $\mathcal{D}(\mathbb{I})$ the Skorokhod space of all \mathbb{R}^d -valued càdlàg functions on \mathbb{I} , by $\mathcal{A}(\mathbb{I}) \subset \mathcal{D}(\mathbb{I})$ the subset of nondecreasing functions, by $\mathcal{C}(\mathbb{I}) \subset \mathcal{D}(\mathbb{I})$ the subset of continuous functions, and by $\mathcal{U}(\mathbb{I})$ the set of all measures on $\mathbb{I} \times U$ for some metric space U, whose first marginal is the Lebesgue measure on \mathbb{I} , and whose second marginal belongs to $\mathcal{P}(U)$. 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

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

Likewise, we denote by $\widetilde{\mathcal{A}}_{0,T}(\mathbb{R})$ and $\widetilde{\mathcal{C}}_{0,T}(\mathbb{R})$ the subspace of $\widetilde{\mathcal{D}}_{0,T}(\mathbb{R})$ with nondecreasing and continuous paths, respectively. Moreover, we denote by $\widetilde{\mathcal{U}}_{0,T}(\mathbb{R})$ all measures q(dt, du) on $\mathbb{R} \times U$ whose restriction to [0,T] belongs to $\mathcal{U}(0,T)$, and whose restrictions to $(-\infty,0)$ and (T,∞) are of the form $q(dt, du) = \delta_{\widetilde{u}\widetilde{u}}(du)dt$ and $q(dt, du) = \delta_{\widetilde{u}\widetilde{u}}(du)dt$ for some $\widetilde{u}_0 \in U$ and $\widetilde{u}_T \in U$, respectively. We occasionally drop the subscripts 0 and T if there is no risk of confusion. Throughout this paper, C > 0 denotes a generic constant that may vary from line to line.

2.1 Singular stochastic control problems

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

$$\begin{cases} \inf_{u,Z} E\left[\int_{0}^{T} f(t, X_{t}, u_{t}) dt + g(X_{T}) + \int_{0}^{T} h(t) dZ_{t}\right], \\ \text{subject to} \\ dX_{t} = b(t, X_{t}, u_{t}) dt + \sigma(t, X_{t}, u_{t}) dW_{t} + c(t) dZ_{t}, \\ X_{0-} = 0. \end{cases}$$

$$(2.1)$$

where all parameters are measurable in their respective arguments and are such that the control problem makes sense; see, e.g. [15] for details.³ The *regular control* $u = (u_t)_{t \in [0,T]}$ takes values in a compact metric space U, and the singular control $Z = (Z_t)_{t \in [0,T]}$ takes values in \mathbb{R}^d . For convenience we sometimes write $Z \in \widetilde{\mathcal{A}}(\mathbb{R})$ by which we mean that the sample paths of the stochastic process Z belong to $\widetilde{\mathcal{A}}(\mathbb{R})$. Similarly, we occasionally write $X \in \widetilde{\mathcal{D}}(\mathbb{R})$ and $Y \in \widetilde{\mathcal{C}}(\mathbb{R})$.

2.1.1 Relaxed controls

The existence of optimal *relaxed controls* to stochastic singular control problems has been addressed in [15] 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 [15] where we adopt our convention that all processes are extended to the whole real line.

³Our specific assumptions on the model parameters are introduced in Section 2.2 below.

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

- 1. $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P})$ is a filtered probability space;
- 2. $\mathbb{P}(X_t = 0, Z_t = 0, \underline{Q}_t(du) = \delta_{\widetilde{u}_0}(du) \text{ if } t < 0; X_t = X_T, Z_t = Z_T, \underline{Q}_t(du) = \delta_{\widetilde{u}_T}(du) \text{ if } t > T) = 1,$ for some $\widetilde{u}_0, \widetilde{u}_T \in U;$
- 3. $\underline{Q} : \mathbb{R} \times \Omega \to \mathcal{P}(U)$ is $\{\mathcal{F}_t, t \in \mathbb{R}\}$ progressively measurable, Z is $\{\mathcal{F}_t, t \in \mathbb{R}\}$ progressively measurable and $Z \in \widetilde{\mathcal{A}}(\mathbb{R})$;
- 4. X is a $\{\mathcal{F}_t, t \in \mathbb{R}\}$ adapted stochastic process, $X \in \widetilde{\mathcal{D}}(\mathbb{R})$ and for each $\phi \in \mathcal{C}^2_b(\mathbb{R}^d)$, the space of all continuous and bounded functions with continuous and bounded first- and second-order derivatives, \mathcal{M}^{ϕ} is a well defined \mathbb{P} continuous martingale, where

$$\mathcal{M}_t^{\phi} := \phi(X_t) - \int_0^t \int_U \mathcal{L}\phi(s, X_s, u) \underline{Q}_s(du) ds - \int_0^t (\partial_x \phi(X_{s-}))^\top c(s) dZ_s$$
$$- \sum_{0 \le s \le t} \left(\phi(X_s) - \phi(X_{s-}) - (\partial_x \phi(X_{s-}))^\top \Delta X_s \right), \qquad t \in [0, T]$$

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 $a(t,x,u) = \sigma \sigma^\top(t,x,u)$.

The cost functional corresponding to a relaxed control r is defined by

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

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, \underline{Q}, Z)$ be a relaxed control. If the process \underline{Q} is of the form $\underline{Q}_t(du) = \delta_{u_t}(du)$, for some progressively measurable U-valued process u, then we call $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, u, Z)$ a strict control.⁴ In particular, any strict control corresponds to a relaxed control. Relaxed control can thus be viewed as a form of mixed strategies over strict controls. In particular, both the cost function and the state dynamics (more precisely, the martingale problem) are linear in relaxed controls. Furthermore, compactness w.r.t. relaxed controls is much easier to verify than compactness w.r.t. strict controls. Under suitable convexity conditions on the model data, the optimization problem over the set of relaxed controls is equivalent to the one over strict controls as shown by the following remark.

Remark 2.2. 1. 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, there exists a strict control and a singular control with smaller or equal cost. Indeed, by the proof of [14, Theorem 3.6], for any relaxed control $r = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, X, \underline{Q}, Z)$, there exists a progressively measurable U-valued process \bar{u} and a \mathbb{R}^+ -valued process \bar{v} such that for almost all $(t, \omega) \in [0, T] \times \Omega$,

$$\left(\int_{U} a(t, X_{t}(\omega), u) \underline{Q}_{t}(\omega, du), \int_{U} b(t, X_{t}(\omega), u) \underline{Q}_{t}(\omega, du), \int_{U} f(t, X_{t}(\omega), u) \underline{Q}_{t}(\omega, du)\right) = (a(t, X_{t}(\omega), \bar{u}_{t}(\omega)), b(t, X_{t}(\omega), \bar{u}_{t}(\omega)), f(t, X_{t}(\omega), \bar{u}_{t}(\omega)) + \bar{v}_{t}(\omega)).$$

$$(2.3)$$

Then $\alpha = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, X, \bar{u}, Z)$ is a strict control with smaller or equal cost.

2. If $a(t, x, u) = K_{t,x}u^2$, $b(t, x, u) = K'_{t,x}u^2$, $f(t, x, u) = K''_{t,x}u^2$, where $|K_{t,x}|$, $|K'_{t,x}|$, $|K''_{t,x}| \le K$ for some positive constant K, then the set K(t, x) is convex for each $(t, x) \in [0, T] \times \mathbb{R}^d$.

⁴If there is no risk of confusion, then we call the processes \underline{Q} , respectively u the relaxed, respectively strict control.

2.1.2 Canonical state space and disintegration

In what follows, we always assume that Ω is the canonical path space, i.e.

9

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

and that the filtration $\{\mathcal{F}_t, t \in \mathbb{R}\}$ is generated by the coordinate projections X, Q, Z. More precisely, for each $\omega := (x, q, z) \in \Omega$,

$$X(\omega) = x, \qquad Q(\omega) = q, \qquad Z(\omega) = z.$$

and for $t \in [0,T]$, $\mathcal{F}_t := \mathcal{F}_t^X \times \mathcal{F}_t^Q \times \mathcal{F}_t^Z$, where

$$\mathcal{F}_t^X = \sigma(X_s, s \le t), \qquad \mathcal{F}_t^Q = \sigma(Q(S), S \in \mathcal{B}([0, t] \times U)), \qquad \mathcal{F}_t^Z = \sigma(Z_s, s \le t);$$

if t < 0, then $\mathcal{F}_t := \{\Omega, \emptyset\}$ and if t > T, then $\mathcal{F}_t := \mathcal{F}_T$.

The following argument shows that relaxed controls can be defined in terms of projection mappings. In fact, since [0, T] and U are compact, by the definition of $\widetilde{\mathcal{U}}(\mathbb{R})$, each $q \in \widetilde{\mathcal{U}}(\mathbb{R})$ allows for the disintegration

$$q(dt, du) = q_t(du)dt$$

for some measurable $\mathcal{P}(U)$ -valued function q_t . By [26, Lemma 3.2] and by definition of the space $\widetilde{\mathcal{U}}(\mathbb{R})$ there exists a \mathcal{F}_t^Q -predictable $\mathcal{P}(U)$ -valued process Π such that for each $q \in \widetilde{\mathcal{U}}(\mathbb{R})$,

$$\Pi_t(q) = q_t, \ a.e. \ t \in [0,T]; \qquad \Pi_t(q) \equiv \delta_{\widetilde{u_0}}, \ t < 0; \qquad \Pi_t(q) \equiv \delta_{\widetilde{u_T}}, \ t > T.$$

Hence, the process $Q_t^o := \Pi_t \circ Q$ is \mathcal{F}_t -predictable. As a result, for each $\omega = (x, q, z)$,

$$Q(\omega)(dt, du) = q(dt, du) = q_t(du)dt = \Pi_t(q)(du)dt = \Pi_t \circ Q(\omega)(du)dt = Q_t^o(\omega)(du)dt.$$

This yields an adapted disintegration of Q in terms of the $\{\mathcal{F}_t, t \in \mathbb{R}\}$ progressively measurable process

$$Q^o: \mathbb{R} \times \Omega \to \mathcal{P}(U).$$

and hence allows us to define control rules. We notice that it is not appropriate to replace $\widetilde{\mathcal{U}}(\mathbb{R})$ in the definition of the canonical path space by the space of càdlàg $\mathcal{P}(U)$ -valued functions as the definition of relaxed controls does not assume any path properties of $t \mapsto \underline{Q}_t$.

Definition 2.3. For the canonical path space Ω , the canonical filtration $\{\mathcal{F}_t, t \in \mathbb{R}\}$ and the coordinate projections (X, Q, Z) introduced above, if $r = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, Q^o, Z)$ is a relaxed control in the sense of Definition 2.1, then the probability measure \mathbb{P} is called a *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,

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

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. [15, Proposition 2.6]. In other words, the optimization problems over relaxed controls and control rules are equivalent. It is hence enough to consider control rules. From now on, we let $(Q_t)_{t \in \mathbb{R}} := (Q_t^o)_{t \in \mathbb{R}}$ for simplicity.

Remark 2.4. In [15] - with the choice of different topologies and under suitable assumptions on the cost coefficients - it is 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 [15, Remark 2.2 and Section 4] for details. As a byproduct (see Corollary 3.10) of our analysis of MFGs, under the same assumptions on the coefficients as in [15] we establish the existence of an optimal control rule for terminal cost functions that satisfy a linear growth condition. In Section 3.3 we furthermore outline a generalization of the stochastic singular control problem to problems of McKean-Vlasov-type.

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. Throughout the paper, for each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$, put $\mu_t = \mu \circ \pi_t^{-1}$, where $\pi_t : x \in \widetilde{\mathcal{D}}(\mathbb{R}) \to x_t$. 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 $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$. The canonical path space for MFGs with singular controls is

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

We assume that the spaces $\widetilde{\mathcal{D}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}(\mathbb{R})$ are endowed with the M_1 topology. We define a metric on the space $\mathcal{U}(\mathbb{R})$ induced by the Wasserstein distance on compact time intervals by

$$d_{\mathcal{U}(\mathbb{R})}(q^{1},q^{2}) := \mathcal{W}_{p,[0,T]\times U}\left(\frac{q^{1}}{T},\frac{q^{2}}{T}\right) + \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} \left\{ \mathcal{W}_{p,[-(n+1),-n]\times U}(q^{1},q^{2}) + \mathcal{W}_{p,[T+n,T+n+1]\times U}(q^{1},q^{2}) \right\}.$$
(2.4)

The space $\widetilde{\mathcal{U}}(\mathbb{R})$ endowed with the metric $d_{\widetilde{\mathcal{U}}(\mathbb{R})} := d_{\mathcal{U}(\mathbb{R})}$ is compact. Furthermore, it is well known [30, 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.

Definition 2.5. A probability measure \mathbb{P} is called a control rule with respect to $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ if

- 1. $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P})$ is the canonical probability space and (X, Q, Z) are the coordinate projections;
- 2. for each $\phi \in \mathcal{C}^2_b(\mathbb{R}^d)$, $\mathcal{M}^{\mu,\phi}$ is a well defined \mathbb{P} continuous martingale, where

$$\mathcal{M}_{t}^{\mu,\phi} := \phi(X_{t}) - \int_{0}^{t} \int_{U} \mathcal{L}\phi(s, X_{s}, \mu_{s}, u) Q_{s}(du) ds - \int_{0}^{t} (\partial_{x}\phi(X_{s-}))^{\top} c(s) dZ_{s}$$
$$- \sum_{0 \le s \le t} \left(\phi(X_{s}) - \phi(X_{s-}) - (\partial_{x}\phi(X_{s-}))^{\top} \bigtriangleup X_{s}\right), \quad t \in [0, T]$$
(2.5)

with $\mathcal{L}\phi(t, x, \nu, u) := \frac{1}{2} \sum_{ij} a_{ij}(t, x, \nu, u) \frac{\partial^2 \phi(x)}{\partial x_i \partial x_j} + \sum_i b_i(t, x, \nu, u) \partial_{x_i} \phi(x)$ and $a(t, x, \nu, u) = \sigma \sigma^\top(t, x, \nu, u)$, for each $(t, x, \nu, u) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times U$.

For a fixed measure $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$, the corresponding set of control rules is denoted by $\mathcal{R}(\mu)$, 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 $\mathbb{P} \circ X^{-1}$ or \mathbb{P} or the associated tuple $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, X, Q, Z)$ a relaxed solution to the MFG with singular controls (1.4). Moreover, if $\mathbb{P} \in \mathcal{R}^*(\mathbb{P} \circ X^{-1})$ and $\mathbb{P}(Q(dt, du) = \delta_{\bar{u}_t}(du)dt) = 1$ for some progressively measurable process \bar{u} , then we call $\mathbb{P} \circ X^{-1}$ or \mathbb{P} or the associated tuple $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, X, \bar{u}, Z)$ a strict solution.

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 $\bar{p} > p \ge 1$, we assume that the following conditions are satisfied:

- \mathcal{A}_1 . There exists a positive constant C_1 such that $|b| \leq C_1$ and $|a| \leq C_1$; b and σ are measurable in $t \in [0,T]$ and continuous in $(x,\nu,u) \in \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times U$; moreover, b and σ are Lipschitz continuous in $x \in \mathbb{R}^d$, uniformly in $(t,\nu,u) \in [0,T] \times \mathcal{P}_p(\mathbb{R}^d) \times U$.
- \mathcal{A}_2 . The functions f and g are measurable in $t \in [0,T]$ and are continuous with respect to $(x,\nu,u) \in \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times U$.
- \mathcal{A}_3 . For each $(t, x, \nu, u) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times U$, there exist strictly positive constants C_2 , C_3 and a positive constant C_4 such that

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

and

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

- \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 uniformly in (t, x, u), i.e., for $\varphi = b$, σ and f, there exists $C_5 > 0$ such that for each $(t, x, u) \in [0, T] \times \mathbb{R}^d \times U$ and $\nu^1, \nu^2 \in \mathcal{P}_p(\mathbb{R}^d)$ there holds that

$$|\varphi(t, x, \nu^{1}, u) - \varphi(t, x, \nu^{2}, u)| \le C_{5} \Big(1 + L(\mathcal{W}_{p}(\nu^{1}, \delta_{0}), \mathcal{W}_{p}(\nu^{2}, \delta_{0})) \Big) \mathcal{W}_{p}(\nu^{1}, \nu^{2}),$$

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

 \mathcal{A}_6 . U is a compact metrizable space.

Under assumptions A_1 - A_6 , there exists a relaxed solution to the MFGs with singular controls (1.4).

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

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

where $|\bar{g}(\nu)| \leq \int_{\mathbb{R}^d} |y|^p \nu(dy)$. 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 3.13. 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 $\int |x|^p \nu(dx)$ or $\int |x|^p \nu(dx) \wedge K$ for some fixed constant K if boundedness is required. Remark 2.8. If we assume for each $(t, x, \nu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d)$, $K(t, x, \nu)$ is convex, where

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

a strict solution to our MFG can be constructed from a relaxed solution. Let $r^* = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}^*, X, Q, Z)$ is a relaxed solution to MFG. Let $a^*(t, x, u) = a(t, x, \mu_t^*, u)$, $b^*(t, x, u) = b(t, x, \mu_t^*, u)$ and $f^*(t, x, u) = f(t, x, \mu_t^*, u)$, where $\mu^* = \mathbb{P}^* \circ X^{-1}$. Similar to Remark 2.2, there exist U-valued process \bar{u} and \mathbb{R}^+ -valued process \bar{v} such that (2.3) holds with a, b, f replaced by a^*, b^*, f^* , respectively. Define

$$\alpha^* = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q}^*, X, \bar{u}, Z),$$

where $\mathbb{Q}^* = \mathbb{P}^* \circ (X, \delta_{\bar{u}_t}(du)dt, Z)^{-1}$. Then, α^* is a strict solution. The point is that the marginal distribution μ^* does not change when passing from r^* to α^* .

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, unless stated otherwise, 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 \le 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. The proof uses the definition of the distance |x - [y, z]| of a point x to a line segment [y, z] and the modified strong M_1 oscillation function \tilde{w}_s introduced in (B.4) and (B.11), respectively.

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]:

$$dX_t^n = b_n(t) dt + dM_t^n + dc_n(t), (3.1)$$

where the random coefficients b_n is measurable and bounded uniformly in n, M^n is a continuous martingale with uniformly bounded and absolutely continuous quadratic variation, and c_n is monotone and càdlàg in time a.s. and $\sup_n E^{\mathbb{P}^n}(|c_n(0)| \vee |c_n(T)|)^{\bar{p}} < \infty$. Moreover, assume that $X_t^n = 0$ if t < 0and $X_t^n = X_T^n$ if t > T. Then, the sequence $\{\mathbb{P}^n \circ (X^n)^{-1}\}_{n \geq 1}$ is relatively compact as a sequence in $\mathcal{W}_{p,(\tilde{\mathcal{D}}(\mathbb{R}),d_{M_1})}$.

Proof. By the uniform boundedness of b_n , $E^{\mathbb{P}^n}(|c_n(0)| \vee |c_n(T)|)^{\bar{p}}$ and the quadratic variation of M^n , there exists a constant C that is independent of n, such that

$$E^{\mathbb{P}^n} \sup_{0 \le t \le T} |X^n_t|^{\bar{p}} \le C < \infty.$$
(3.2)

By [29, Definition 6.8(3)] 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 Proposition B.6. 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 $0 \lor (t - \delta) \le t_1 < t_2 < t_3 \le (t + \delta) \land T$, the monotonicity of c_n implies

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

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

$$|X_{t_1}^n - [0, X_{t_2}^n]| \le \left| \int_{t_1}^{t_2} b_n(s) \, ds + M_{t_2}^n - M_{t_1}^n \right|.$$

Therefore,

$$\widetilde{w}_s(X,\delta) \le 3 \sup_t \sup_{t_1,t_2} \left| \int_{t_1}^{t_2} b_n(s) \, ds + M_{t_2}^n - M_{t_1}^n \right|,$$

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

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

for some positive function $k(\delta)$ that is independent of n and m 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 $\bigcup_{\mu \in \mathcal{P}_n(\widetilde{\mathcal{D}}(\mathbb{R}))} \mathcal{R}(\mu)$ is relatively compact in \mathcal{W}_p .

Proof. Let $\{\mu^n\}_{n\geq 1}$ be any sequence in $\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and $\mathbb{P}^n \in \mathcal{R}(\mu^n), n \geq 1$. It is sufficient to show that $\{\mathbb{P}^n \circ X^{-1}\}_{n\geq 1}, \{\mathbb{P}^n \circ Q^{-1}\}_{n\geq 1}$ and $\{\mathbb{P}^n \circ Z^{-1}\}_{n\geq 1}$ are relatively compact. Since U and $\widetilde{\mathcal{A}}^m(\mathbb{R})$ are compact by assumption and Corollary B.5, respectively, $\{\mathbb{P}^n \circ Q^{-1}\}_{n\geq 1}$ and $\{\mathbb{P}^n \circ Z^{-1}\}_{n\geq 1}$ are tight. Since $\widetilde{\mathcal{U}}(\mathbb{R})$ and $\widetilde{\mathcal{A}}^m(\mathbb{R})$ are compact, these sequences are relatively compact in the topology induced by Wasserstein metric; see [29, Definition 6.8(3)].

It remains to prove the relative compactness of $\{\mathbb{P}^n \circ X^{-1}\}_{n\geq 1}$. Since \mathbb{P}^n is a control rule associated with the measure μ^n , for any n, it follows from Proposition A.2 that there exist extensions $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t, t \in \mathbb{R}\}, \mathbb{Q}^n)$ of the canonical path spaces and processes (X^n, Q^n, Z^n, M^n) defined on it, such that

$$dX_t^n = \int_U b(t, X_t^n, \mu_t^n, u) Q_t^n(du) dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n du dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M^n(du, dt) + c(t) dZ_t^n dU dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) dU dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) dU dt + \int_U \sigma($$

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 $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t \in \mathbb{R}\}, \mathbb{Q}^n)$ with intensity Q^n . Relative compactness of $\{\mathbb{P}^n \circ X^{-1}\}_{n \geq 1}$ now reduces to relative compactness of $\{\mathbb{Q}^n \circ (X^n)^{-1}\}_{n \geq 1}$, which is a direct consequence of the preceding Proposition 3.1.

Remark 3.3. For the above result, the assumption c > 0 is not necessary. To see this, we decompose X^n as

$$X_{\cdot}^{n} = \int_{0}^{\cdot} \int_{U} b(t, X_{t}^{n}, \mu_{t}^{n}, u) Q_{t}^{n}(du) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X_{t}^{n}, \mu_{t}^{n}, u) M^{n}(du, dt) + \int_{0}^{\cdot} c^{+}(t) dZ_{t}^{n} - \int_{0}^{\cdot} c^{-}(t) dZ_{t}^{n},$$

where c^+ and c^- are the positive and negative parts of c, respectively. By the boundedness of b and σ , we see that

$$E^{\mathbb{Q}^n} |K_t^n - K_s^n|^4 \le C|t - s|^2,$$

where

$$K^{n}_{\cdot} := \int_{0}^{\cdot} \int_{U} b(t, X^{n}_{t}, \mu^{n}_{t}, u) Q^{n}_{t}(du) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, X^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U} \sigma(t, \chi^{n}_{t}, \mu^{n}_{t}, u) M^{n}(du, dt) dt + \int_{0}^{\cdot} \int_{U}$$

Kolmogorov weak tightness criterion implies that, for each $\epsilon > 0$, there exists a compact set $\mathcal{K}_1 \subseteq \widetilde{\mathcal{C}}(\mathbb{R})$ such that

$$\inf_{n} \mathbb{Q}^{n} \left(K^{n} \in \mathcal{K}_{1} \right) \ge 1 - \epsilon.$$

Define

$$\mathcal{K}_2 := \left\{ \int_0^{\cdot} c^+(s) \, dz_s \in \widetilde{\mathcal{A}}(\mathbb{R}) : \ z_T \in \widetilde{\mathcal{A}}^m(\mathbb{R}) \right\}.$$

and

$$\mathcal{K}_3 := \left\{ -\int_0^{\cdot} c^-(s) \, dz_s \in \widetilde{\mathcal{D}}(\mathbb{R}) : z \in \widetilde{\mathcal{A}}^m(\mathbb{R}) \right\}.$$

Thus,

$$\inf_{n} \mathbb{P}^{n} \{ \omega \in \Omega : X_{\cdot}(\omega) \in \mathcal{K}_{1} + \mathcal{K}_{2} + \mathcal{K}_{3} \}$$

$$\geq \inf_{n} \mathbb{Q}^{n} \left\{ \bar{\omega} \in \bar{\Omega} : K_{\cdot}^{n} \in \mathcal{K}_{1}, \int_{0}^{\cdot} c^{+}(s) \, dZ_{s}^{n} \in \mathcal{K}_{2}, -\int_{0}^{\cdot} c^{-}(s) \, dZ_{s}^{n} \in \mathcal{K}_{3} \right\}$$

$$\geq 1 - \epsilon.$$

$$(3.4)$$

By Corollary B.5, \mathcal{K}_2 and \mathcal{K}_3 are M_1 -compact subsets of $\widetilde{\mathcal{D}}(\mathbb{R})$. By Remark B.2(1), \mathcal{K}_1 is also a M_1 compact subset. Note that the elements of \mathcal{K}_1 do not jump and that $\int_0^{\cdot} c^+(s) dz_s$ and $\int_0^{\cdot} c^-(s) dz_s$ never jump at the same time. Thus, Proposition B.7 implies that $\mathcal{K}_1 + \mathcal{K}_2 + \mathcal{K}_3$ 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

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

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

$$\mathcal{J}(\mu,\omega) = \int_0^T \int_U f(t, x_t, \mu_t, u) \, q_t(du) dt + g(x_T, \mu_T) + \int_0^T h(t) \, dz_t.$$
(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 joint continuity of J will be proved in the final step. 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 $\mu_t^n = \mu^n \circ \pi_t^{-1}$ and $\mu_t = \mu \circ \pi_t^{-1}$, where π is the projection 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 ,

$$\left| \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}^{n}, u) q_{t}(du) dt - \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}, u) q_{t}(du) dt \right|$$

$$\leq C \int_{0}^{T} \left(1 + L \left(\mathcal{W}_{p}(\mu_{t}^{n}, \delta_{0}), \mathcal{W}_{p}(\mu_{t}, \delta_{0}) \right) \right) \mathcal{W}_{p}(\mu_{t}^{n}, \mu_{t}) dt \qquad (3.6)$$

$$\leq C \left(\int_{0}^{T} \left(1 + L \left(\mathcal{W}_{p}(\mu_{t}^{n}, \delta_{0}), \mathcal{W}_{p}(\mu_{t}, \delta_{0}) \right) \right)^{\frac{p}{p-1}} dt \right)^{1-\frac{1}{p}} \left(\int_{0}^{T} \mathcal{W}_{p}(\mu_{t}^{n}, \mu_{t})^{p} dt \right)^{\frac{1}{p}}.$$

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 exists \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

$$\left| \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}^{n}, u) q_{t}(du) dt - \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}, u) q_{t}(du) dt \right|$$

$$\leq C \left(\int_{0}^{T} \left(1 + L \left(\mathcal{W}_{p}(\mathbb{Q} \circ (\bar{X}_{t}^{n})^{-1}, \delta_{0}), \mathcal{W}_{p}(\mathbb{Q} \circ \bar{X}_{t}^{-1}, \delta_{0}) \right) \right)^{\frac{p}{p-1}} dt \right)^{1 - \frac{1}{p}} \left(E^{\mathbb{Q}} \int_{0}^{T} |\bar{X}_{t}^{n} - \bar{X}_{t}|^{p} dt \right)^{\frac{1}{p}}$$

By Remark B.2, we have

$$\int_0^T |\bar{X}_t^n - \bar{X}_t|^p \, dt \to 0 \qquad \text{a.s. } \mathbb{Q}.$$

Moreover, we have

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

On the other hand,

$$E^{\mathbb{Q}}\left(d_{M_{1}}(\bar{X}^{n},0)^{p}+d_{M_{1}}(\bar{X},0)^{p}\right) = \int_{\mathcal{D}[0,T]} d_{M_{1}}(x,0)^{p} \mu^{n}(dx) + \int_{\mathcal{D}[0,T]} d_{M_{1}}(x,0)^{p} \mu(dx)$$
$$\to 2 \int_{\mathcal{D}[0,T]} d_{M_{1}}(x,0)^{p} \mu(dx) < \infty.$$

Therefore, dominated convergence yields

$$E^{\mathbb{Q}} \int_{0}^{T} |\bar{X}_{t}^{n} - \bar{X}_{t}|^{p} dt \to 0.$$
(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

$$\left| \int_0^T \int_U f(t, x_t, \mu_t^n, u) q_t(du) dt - \int_0^T \int_U f(t, x_t, \mu_t, u) q_t(du) dt \right| \to 0, \quad \text{uniformly in } \omega.$$
(3.8)

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 $t \notin Disc(x)$ 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

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

we see that

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

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 $\int_0^T \int_U f(t, x_t, \mu_t, u) q_t(du) dt$. 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

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

$$\leq \int_0^T \sup_{u \in U} |f(t, x_t^n, \mu_t, u) - f(t, x_t, \mu_t, u)| \, dt \to 0.$$

By [29, Definition 6.8], $q^n \to q$ in $d_{\widetilde{\mathcal{U}}(\mathbb{R})}$ implies $q^n \to q$ weakly. Moreover, the first marginal of q^n is Lebesgue measure. Thus, by [21, Corollary 2.9], q^n converges to q in the stable topology (cf. [21, Definition 1.2] for the definition of the stable topology). For fixed $(x, \mu) \in \widetilde{\mathcal{D}}(\mathbb{R}) \times \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$, the compactness of U and the growth condition on f implies the boundedness of f. Hence the definition of stable topology vields that

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

So we get the convergence

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

Step 3: joint continuity of J. Thus far, we have established the separate continuity of the mapping $(\mu, \omega) \to \mathcal{J}(\mu, \omega)$. We are now going to apply [29, Definition 6.8(4)] to prove the joint 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 ,

$$\begin{aligned} \left| \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}, u) q_{t}(du) dt + \int_{0}^{T} h(t) dz_{t} \right| \\ &\leq C \left(1 + \int_{0}^{T} \int_{U} \left(1 + |x_{t}|^{p} + |u|^{p} + \int_{\mathbb{R}^{d}} |y|^{p} \mu_{t}(dy) \right) q_{t}(du) dt + z_{T} \right) \\ &\leq C \left(1 + d_{M_{1}}(x, 0)^{p} + \mathcal{W}_{p,[0,T] \times U} \left(\frac{q}{T}, \delta_{0} \right)^{p} + d_{M_{1}}(z, 0) + \int_{0}^{T} \int_{\mathbb{R}^{d}} |y|^{p} \mu_{t}(dy) dt \right) \\ &\leq C \left(1 + d_{M_{1}}(x, 0)^{p} + \mathcal{W}_{p,[0,T] \times U} \left(\frac{q}{T}, \delta_{0} \right)^{p} + d_{M_{1}}(z, 0)^{p} + \int_{\widetilde{\mathcal{D}}(\mathbb{R})} d_{M_{1}}(y, 0)^{p} \mu(dy) \right). \end{aligned}$$

Hence, using the uniform convergence (3.8), it follows from [29, Definition 6.8] that $(\mu^n, \mathbb{P}^n) \to (\mu, \mathbb{P})$ implies that

$$\begin{aligned} \left| E^{\mathbb{P}^{n}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}^{n}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) - E^{\mathbb{P}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) \right| \\ \leq \left| E^{\mathbb{P}^{n}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}^{n}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) - E^{\mathbb{P}^{n}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) \right| \\ + \left| E^{\mathbb{P}^{n}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) - E^{\mathbb{P}} \left(\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}, u) Q_{t}(du) dt + \int_{0}^{T} h(t) dZ_{t} \right) \right| \\ \rightarrow 0. \end{aligned}$$

$$(3.9)$$

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 (μ, \mathbb{P}) . First, we notice that for each $\tilde{p} > \bar{p}$, by the boundedness of b, σ and Z, we have that

$$\sup_{n} E^{\mathbb{P}^{n}} d_{M_{1}}(X,0)^{\widetilde{p}} \le C < \infty,$$
(3.10)

which implies

$$\lim_{K \to \infty} \sup_{n} \int_{\{x: d_{M_1}(x, 0) > K\}} d_{M_1}(x, 0)^{\bar{p}} \mathbb{P}^n(dx) = 0.$$
(3.11)

By Assumption \mathcal{A}_3 ,

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

Together with (3.11) this implies,

$$E^{\mathbb{P}^n}g(X_T,\mu_T) \to E^{\mathbb{P}}g(X_T,\mu_T).$$
(3.12)

By the tightness of $\{\mathbb{P}^n\}_{n\geq 1}$, for each $\epsilon > 0$, there exists a compact set $K_{\epsilon} \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ such that

$$\begin{aligned} \left| \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T, \mu_T^n) \mathbb{P}^n(dx) - \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T, \mu_T) \mathbb{P}^n(dx) \right| \\ &\leq \int_{K_{\epsilon}} |g(x_T, \mu_T^n) - g(x_T, \mu_T)| \mathbb{P}^n(dx) + \int_{\widetilde{\mathcal{D}}(\mathbb{R})/K_{\epsilon}} |g(x_T, \mu_T^n) - g(x_T, \mu_T)| \mathbb{P}^n(dx) \\ &\leq \sup_{x \in K_{\epsilon}} |g(x_T, \mu_T^n) - g(x_T, \mu_T)| + \left(\int_{\widetilde{\mathcal{D}}(\mathbb{R})/K_{\epsilon}} |g(x_T, \mu_T^n) - g(x_T, \mu_T)|^2 \mathbb{P}^n(dx) \right)^{\frac{1}{2}} \left(\sup_{n} \mathbb{P}^n(\widetilde{\mathcal{D}}(\mathbb{R})/K_{\epsilon}) \right)^{\frac{1}{2}} \\ &\leq \sup_{x \in K_{\epsilon}} |g(x_T, \mu_T^n) - g(x_T, \mu_T)| + C\epsilon^{\frac{1}{2}} \quad (by (3.10)). \end{aligned}$$

$$(3.13)$$

Thus,

$$\left| \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T, \mu_T^n) \mathbb{P}^n(dx) - \int_{\widetilde{\mathcal{D}}(\mathbb{R})} g(x_T, \mu_T) \mathbb{P}^n(dx) \right| \to 0.$$
(3.14)

The convergence (3.9), (3.12) and (3.14) yield the joint continuity of $J(\cdot, \cdot)$.

Remark 3.5. The preceding lemma shows that under a finite fuel constraint the cost functional J is jointly continuous. In general, J is only lower semi-continuous. In fact, for each positive constant K, let $g_K(\cdot) := g(\cdot) \wedge K$ and

$$\mathcal{J}_{K}(\mu,\omega) := \int_{0}^{T} \int_{U} f(t, x_{t}, \mu_{t}, u) q_{t}(du) dt + g_{K}(x_{T}, \mu_{T}) + \int_{0}^{T} h(t) dz_{t}$$

By assumption \mathcal{A}_3 , we have

$$|g_K(x,\mu)| \le 2K + C_2 \left(1 + \int_{\mathbb{R}^d} |y|^p \mu(dy)\right) \le C \left(1 + \int_{\mathbb{R}^d} |y|^p \mu(dy)\right).$$

So (3.12) and (3.13) still hold with g replaced by g_K while (3.9) still holds for f and h. So $(\mu^n, \mathbb{P}^n) \to (\mu, P)$ implies

$$\int_{\Omega} \mathcal{J}_K(\mu^n, \omega) \mathbb{P}^n(d\omega) \to \int_{\Omega} \mathcal{J}_K(\mu, \omega) \mathbb{P}(d\omega).$$

Thus, by monotone convergence theorem, we have

$$\liminf_{n \to \infty} \int_{\Omega} \mathcal{J}(\mu^n, \omega) \mathbb{P}^n(d\omega) \ge \int_{\Omega} \mathcal{J}(\mu, \omega) \mathbb{P}(d\omega).$$

We now recall from [15, 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. Since it is difficult to locate the proof, we give a sketch one in Appendix C.

Proposition 3.6. A probability measure \mathbb{P} is a control rule with respect to the given $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ if and only if there exists an \mathcal{F}_t adapted process $Y \in \mathcal{C}(0,T)$ on the filtered canonical space $(\Omega, \mathcal{F}, \mathcal{F}_t)$ such that

(1)
$$\mathbb{P}(\omega \in \Omega : X_t(\omega) = Y_t(\omega) + \int_0^t c(s) dZ_s(\omega), t \in [0, T]) = 1;$$

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

$$\overline{\mathcal{M}}_{t}^{\mu,\phi} = \phi(Y_{t}) - \int_{0}^{t} \int_{U} \bar{\mathcal{L}}\phi(s, X_{s}, Y_{s}, \mu_{s}, u) Q_{s}(du) ds, \qquad t \in [0, T]$$
(3.15)

with $\bar{\mathcal{L}}\phi(s,x,y,\nu,u) = \sum_i b_i(s,x,\nu,u)\partial_{y_i}\phi(y) + \frac{1}{2}\sum_{ij}a_{ij}(s,x,\nu,u)\frac{\partial^2\phi(y)}{\partial_{y_i}\partial_{y_j}}$ for each $(t,x,y,\nu,u) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_p(\mathbb{R}^d) \times U.$

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

Proposition 3.7. Suppose that \mathcal{A}_1 and \mathcal{A}_4 - \mathcal{A}_6 hold. For any sequence $\{\mu^n\}_{n\geq 1} \subseteq \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ with $\mu^n \to \mu$ in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}),d_{M_1})}$, if $\mathbb{P}^n \in \mathcal{R}(\mu^n)$ and $\mathbb{P}^n \to \mathbb{P}$ in \mathcal{W}_p , then $\mathbb{P} \in \mathcal{R}(\mu)$.

Proof. In order to verify conditions (1) and (2), notice first that, for each n, there exists a stochastic process $Y^n \in \mathcal{C}(0,T)$ such that

$$\mathbb{P}^n\left(X_t = Y_t^n + \int_0^t c(s) \, dZ_s, t \in [0, T]\right) = 1$$

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.2. 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

$$d\bar{X}_{t}^{n} = \int_{U} b(t, \bar{X}_{t}^{n}, \mu_{t}^{n}, u) \, \bar{Q}_{s}^{n}(du) ds + \int_{U} \sigma(t, \bar{X}_{t}^{n}, \mu_{t}^{n}, u) \, M^{n}(du, dt) + c(t) d\bar{Z}_{t}^{n} + c(t) d\bar{Z}_{t}^{n}$$

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

$$E^{\mathbb{P}^{n}}|Y_{t}^{n}-Y_{s}^{n}|^{4} = E^{\mathbb{P}^{n}} \left| \left(X_{t} - \int_{0}^{t} c(r) \, dZ_{r} \right) - \left(X_{s} - \int_{0}^{s} c(r) \, dZ_{r} \right) \right|^{4} \\ = E^{\mathbb{Q}^{n}} \left| \left(\bar{X}_{t}^{n} - \int_{0}^{t} c(r) \, d\bar{Z}_{r}^{n} \right) - \left(\bar{X}_{s}^{n} - \int_{0}^{s} c(r) \, d\bar{Z}_{r}^{n} \right) \right|^{4} \\ = E^{\mathbb{Q}^{n}} \left| \int_{s}^{t} \int_{U} b(r, \bar{X}_{r}^{n}, \mu_{r}^{n}, u) \, \bar{Q}_{r}^{n}(du) dr + \int_{s}^{t} \int_{U} \sigma(r, \bar{X}_{r}^{n}, \mu_{r}^{n}, u) \, M^{n}(du, dr) \right|^{4} \\ \leq C|t-s|^{2}.$$
(3.16)

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 $(\hat{X}, \hat{Q}, \hat{Z}, \hat{Y})$ 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

$$Law(\widetilde{X}^n, \widetilde{Q}^n, \widetilde{Z}^n, \widetilde{Y}^n) = Law(X, Q, Z, Y^n), \quad Law(\widetilde{X}, \widetilde{Q}, \widetilde{Z}, \widetilde{Y}) = Law(\widehat{X}, \widehat{Q}, \widehat{Z}, \widehat{Y})$$

and

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

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

$$\mathbb{Q}\left(\widetilde{X}_t = \widetilde{Y}_t + \int_0^t c(s) \, d\widetilde{Z}_s, t \in [0, T]\right) = 1.$$

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 exists a stochastic process $Y \in \mathcal{C}(0, T)$ such that

$$\mathbb{P}\left(X_t = Y_t + \int_0^t c(s) \, dZ_s, t \in [0, T]\right) = 1$$

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

$$\overline{\mathcal{M}}_{t}^{n,\mu^{n},\phi} = \phi(Y_{t}^{n}) - \int_{0}^{t} \int_{U} \overline{\mathcal{L}}(s, X_{s}, Y_{s}^{n}, \mu_{s}^{n}, u) Q_{s}(du) ds,$$
$$\widetilde{\mathcal{M}}_{t}^{n,\mu^{n},\phi} = \phi(\widetilde{Y}_{t}^{n}) - \int_{0}^{t} \int_{U} \overline{\mathcal{L}}(s, \widetilde{X}_{s}^{n}, \widetilde{Y}_{s}^{n}, \mu_{s}^{n}, u) \widetilde{Q}_{s}^{n}(du) ds,$$

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

$$0 = E^{\mathbb{P}^{n}} \left(\overline{\mathcal{M}}_{t}^{n,\mu^{n},\phi} - \overline{\mathcal{M}}_{s}^{n,\mu^{n},\phi} \right) F(X,Q,Z) = E^{\mathbb{Q}} \left(\widetilde{\mathcal{M}}_{t}^{n,\mu^{n},\phi} - \widetilde{\mathcal{M}}_{s}^{n,\mu^{n},\phi} \right) F(\widetilde{X}^{n},\widetilde{Q}^{n},\widetilde{Z}^{n}) \rightarrow E^{\mathbb{Q}} \left(\widetilde{\mathcal{M}}_{t}^{\mu^{*},\phi} - \widetilde{\mathcal{M}}_{s}^{\mu^{*},\phi} \right) F(\widetilde{X},\widetilde{Q},\widetilde{Z}) = E^{\mathbb{P}} \left(\overline{\mathcal{M}}_{t}^{\mu,\phi} - \overline{\mathcal{M}}_{s}^{\mu,\phi} \right) F(X,Q,Z).$$

$$(3.17)$$

Remark 3.8. Note that the proof of Proposition 3.7, does not require the finite fuel constraint.

The next corollary shows that the correspondence \mathcal{R} is continuous in the sense of [2, Definition 17.2, Theorem 17.20, 17.21].

Corollary 3.9. Suppose that $\mathcal{A}_1, \mathcal{A}_4$ - \mathcal{A}_6 hold. Then, $\mathcal{R} : \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R})) \to 2^{\mathcal{P}_p(\Omega)}$ is continuous and compact-valued.

Proof. The lower hemi-continuity of \mathcal{R} can be dealt with as [26, Lemma 4.4] since b and σ are Lipschitz continuous in x. Lemma 3.2, Proposition 3.7 and [2, Theorem 17.20] imply that \mathcal{R} is upper hemi-continuous and compact-valued.

Corollary 3.10. Under assumptions \mathcal{A}_1 - \mathcal{A}_6 , $\mathcal{R}^*(\mu) \neq \emptyset$ for each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and \mathcal{R}^* is upper hemicontinuous.

Proof. By [23, Section 5.4], for each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ the set $\mathcal{R}(\mu)$ is nonempty. Corollary 3.9 implies that \mathcal{R} is compact-valued and continuous. By Lemma 3.4, $J : Gr\mathcal{R} \to \mathbb{R}$ is jointly continuous. Thus, [2, Theorem 17.31] yields that \mathcal{R}^* is nonempty valued and upper hemi-continuous.

Remark 3.11. Corollary 3.10 in fact shows that the stochastic singular control problem (2.1) admits an optimal control rule in the sense of Definition 2.3. Using our method, we could have obtained Corollary 3.10 under the same assumptions of the coefficients as in [15]. We will generalize it to McKean-Vlasov case at the end of this section.

Theorem 3.12. 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. From inequality (3.3) in the proof of Proposition 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 $\mathbb{P}(\widetilde{w}_s(X,\delta) > \eta) \leq \frac{k(\delta)}{\eta}$ and $\lim_{\delta \to 0} k(\delta) = 0$, where \widetilde{w}_s is the modified oscillation function defined in (B.11).

Let us now define a set-valued map ψ by

$$\psi: \mathcal{P}_{p}(\widetilde{\mathcal{D}}(\mathbb{R})) \to 2^{\mathcal{P}_{p}(\widetilde{\mathcal{D}}(\mathbb{R}))},$$

$$\mu \mapsto \{\mathbb{P} \circ X^{-1} : \mathbb{P} \in \mathcal{R}^{*}(\mu)\}, \qquad (3.18)$$

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 $\psi(\mu) \subseteq S \subseteq \overline{S}$, for each $\mu \in \widetilde{\mathcal{D}}(\mathbb{R})$. Hence, $\psi : \overline{S} \to 2^{\overline{S}}$. Moreover, by Corollary 3.10, ψ is nonempty-valued and upper hemi-continuous. Therefore, [2, Corollary 17.55] 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

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

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 **MFG**^m the mean field games

corresponding to Ω^m . The preceding analysis showed that there exists 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 $\{\mathbb{P}^{m*}\}_{m\geq 1}$ is relatively compact; the subsequent one shows that any accumulation point is a control rule.

Lemma 3.13. Suppose A_1 , A_3 , A_4 and 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 $\{\mathbb{P}^{m*}\}_{m\geq 1}$ is relatively compact in $\mathcal{W}_{p,\widetilde{\mathcal{D}}(\mathbb{R})\times\widetilde{\mathcal{U}}(\mathbb{R})\times\widetilde{\mathcal{A}}(\mathbb{R})}$.

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

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

$$(3.19)$$

and

$$E^{\mathbb{P}^{m^*}} |X_t|^p \le C \left(1 + E^{\mathbb{P}^{m^*}} |Z_T|^p \right), \ t \in [0, T].$$
(3.20)

Moreover,

$$\begin{split} J(\mu^{m*}, \mathbb{P}^{m*}) &= E^{\mathbb{P}^{m*}} \left[\int_{0}^{T} \int_{U} f(t, X_{t}, \mu_{t}^{m*}, u) \, Q_{t}(du) dt + g(X_{T}, \mu_{T}^{m*}) + \int_{0}^{T} h(t) \, dZ_{t} \right] \\ &\geq -C \left(1 + \int_{0}^{T} \int_{\mathbb{R}^{d}} |x|^{p} \, \mu_{t}^{m*}(dx) dt + E^{\mathbb{P}^{m*}} \int_{0}^{T} |X_{t}|^{p} \, dt + E^{\mathbb{P}^{m*}} \int_{0}^{T} \int_{U} |u|^{p} \, Q_{t}(du) dt \\ &- E^{\mathbb{P}^{m*}} |X_{T}|^{\bar{p}} + \int_{\mathbb{R}^{d}} |x|^{p} \, \mu_{T}^{m*}(dx) + E^{\mathbb{P}^{m*}} \left| \int_{0}^{T} h(t) \, dZ_{t} \right| \right) \quad \text{(by assumption } \mathcal{A}_{3}) \\ &\geq -C \left(1 + \int_{0}^{T} \int_{\mathbb{R}^{d}} |x|^{p} \, \mu_{t}^{m*}(dx) dt + E^{\mathbb{P}^{m*}} \int_{0}^{T} |X_{t}|^{p} \, dt + E^{\mathbb{P}^{m*}} \int_{0}^{T} \int_{U} |u|^{p} \, Q_{t}(du) dt \\ &+ \int_{\mathbb{R}^{d}} |x|^{p} \, \mu_{T}^{m*}(dx) + E^{\mathbb{P}^{m*}} \left| \int_{0}^{T} h(t) \, dZ_{t} \right| - E^{\mathbb{P}^{m*}} |Z_{T}|^{\bar{p}} \right) \quad \text{(by (3.19))}. \end{split}$$

Now choose any $\mathbb{P}_0 \in \mathcal{R}^m(\mu^{m*})$ such that $\sup_m J(\mu^{m*}, \mathbb{P}_0) < \infty$ (e.g. $\mathbb{P}_0 \in \mathcal{R}(\mu^{m,*})$ such that $\mathbb{P}_0(Q|_{[0,T]} \equiv \delta_{\widetilde{u}}(du)dt|_{[0,T]}, Z \equiv 0) = 1$ for some $\widetilde{u} \in U$). Then,

$$E^{\mathbb{P}^{m*}}|Z_{T}|^{\bar{p}} \leq J(\mu^{m*}, \mathbb{P}^{m*}) + C\left(1 + E^{\mathbb{P}^{m*}}\left|\int_{0}^{T}h(t)\,dZ_{t}\right| + E^{\mathbb{P}^{m*}}\int_{0}^{T}|X_{t}|^{p}\,dt + E^{\mathbb{P}^{m*}}|X_{T}|^{p}\right)$$

$$\leq J(\mu^{m*}, \mathbb{P}_{0}) + C\left(1 + E^{\mathbb{P}^{m*}}|Z_{T}| + E^{\mathbb{P}^{m*}}|Z_{T}|^{p}\right) \quad (by (3.20) \text{ and the optimality of } \mathbb{P}^{m*})$$

$$\leq C\left(1 + E^{\mathbb{P}^{m*}}|Z_{T}| + E^{\mathbb{P}^{m*}}|Z_{T}|^{p}\right). \tag{3.21}$$

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

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

Together with (3.21) this yields,

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

By [29, Definition 6.8] and Proposition 3.1, 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 3.7 (see Remark 3.8).

Lemma 3.14. Suppose that \mathcal{A}_1 and \mathcal{A}_3 - \mathcal{A}_6 hold, let \mathbb{P}^* be an accumulation point of the sequence $\{\mathbb{P}^{m*}\}_{m>1}$. 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 3.15. 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

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

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.2, 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

$$d\bar{X}_{t} = \int_{U} b(t, \bar{X}_{t}, \mu_{t}^{*}, u) \,\bar{Q}_{t}(du)dt + \int_{U} \sigma(t, \bar{X}_{t}, \mu_{t}^{*}, u) M(du, dt) + c(t) \, d\bar{Z}_{t},$$
(3.22)

where M is a martingale measure with intensity \overline{Q} . Using the same argument as in the proof of Lemma 3.13 we see that,

$$E^{\mathbb{P}}Z_T^{\bar{p}} = E^{\bar{\mathbb{P}}}\bar{Z}_T^{\bar{p}} < \infty.$$
(3.23)

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

$$d\bar{X}_{t}^{m} = \int_{U} b(t, \bar{X}_{t}^{m}, \mu_{t}^{m*}, u) \,\bar{Q}_{t}(du)dt + \int_{U} \sigma(t, \bar{X}_{t}^{m}, \mu_{t}^{m*}, u) M(du, dt) + c(t) \, d\bar{Z}_{t}^{m}, \tag{3.24}$$

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

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

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, we have $Z = Z^m$. Hence,

$$E^{\mathbb{P}} \sup_{0 \le t \le T} \left| \int_{0}^{t} c(s) d\bar{Z}_{s} - \int_{0}^{t} c(s) d\bar{Z}_{s}^{m} \right|$$

$$= E^{\mathbb{P}} \sup_{0 \le t \le T} \left| \int_{0}^{t} c(s) dZ_{s} - \int_{0}^{t} c(s) dZ_{s}^{m} \right|$$

$$= \int_{\widetilde{\mathcal{A}}(\mathbb{R}) \setminus \widetilde{\mathcal{A}}^{m}(\mathbb{R})} \sup_{0 \le t \le T} \left| \int_{0}^{t} c(s) dZ_{s}(\omega) - \int_{0}^{t} c(s) dZ_{s}^{m}(\omega) \right| \mathbb{P}(d\omega).$$
(3.25)

By Hölder's inequality,

$$\begin{split} &\int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R})} \sup_{0\leq t\leq T} \left| \int_{0}^{t} c(s) \, dZ_{s}(\omega) - \int_{0}^{t} c(s) \, dZ_{s}^{m}(\omega) \right| \mathbb{P}(d\omega) \\ &\leq \left| \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R})} \int_{0}^{T} c(t) \, dZ_{t}(\omega) \mathbb{P}(d\omega) + \int_{\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R})} \int_{0}^{T} c(t) \, dZ_{t}^{m}(\omega) \mathbb{P}(d\omega) \right| \\ &\leq C \left(E^{\mathbb{P}} Z_{T}^{p} \right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R}))^{1-\frac{1}{p}} + C \left(E^{\mathbb{P}} (Z_{T}^{m})^{p} \right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R}))^{1-\frac{1}{p}} \\ &\leq C \left(E^{\mathbb{P}} Z_{T}^{p} \right)^{\frac{1}{p}} \mathbb{P}(\widetilde{\mathcal{A}}(\mathbb{R})\setminus\widetilde{\mathcal{A}}^{m}(\mathbb{R}))^{1-\frac{1}{p}}. \end{split}$$

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,

$$E^{\bar{\mathbb{P}}} \sup_{0 \le t \le T} \left| \int_0^t c(s) \, d\bar{Z}_s - \int_0^t c(s) \, d\bar{Z}_s^m \right| \to 0.$$
(3.26)

Similarly,

$$E^{\mathbb{\bar{P}}} \left| \int_0^T h(t) \, d\bar{Z}_t - \int_0^T h(t) \, d\bar{Z}_t^m \right| \to 0.$$
(3.27)

By (3.22), (3.24) and (3.26), the Lipschitz continuity of b and σ in x and μ and the Burkholder-Davis-Gundy inequality, standard estimate of SDE yields that

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

By (3.27), (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

$$E^{\bar{\mathbb{P}}}\left(\int_{0}^{T} f(t, \bar{X}_{t}^{m}, \mu_{t}^{m*}, u) \,\bar{Q}_{t}(du)dt + g(\bar{X}_{T}^{m}, \mu_{T}^{m*}) + \int_{0}^{T} h(t) \, d\bar{Z}_{t}^{m}\right)$$

$$\to E^{\bar{\mathbb{P}}}\left(\int_{0}^{T} f(t, \bar{X}_{t}, \mu_{t}^{*}, u) \,\bar{Q}_{t}(du)dt + g(\bar{X}_{T}, \mu_{T}^{*}) + \int_{0}^{T} h(t) \, d\bar{Z}_{t}\right).$$

This shows that

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

Moreover, by Remark 3.5, $\liminf_{m\to\infty} J(\mu^{m*}, \mathbb{P}^{m*}) \geq J(\mu^*, \mathbb{P}^*)$. Hence,

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

3.3 Related McKean-Vlasov stochastic singular control problem

MFGs and control problems of McKean-Vlasov type are compared in [5]. The literatures on McKean-Vlasov singular control focus on necessary conditions for optimality; the existence of optimal control is typically assumed; see e.g. [19]. With the above method for MFGs, we can also establish the existence of an optimal control to the following McKean-Vlasov stochastic singular control problem:

$$\min_{u,Z} J(u,Z) = \min_{u,Z} E\left[\int_0^T f(t, X_t, Law(X_t), u_t) dt + g(X_T, Law(X_T)) + \int_0^T h(t) dZ_t\right]$$
(3.29)

subject to

$$dX_t = b(t, X_t, Law(X_t), u_t) dt + \sigma(t, X_t, Law(X_t), u_t) dW_t + c(t) dZ_t, \ t \in [0, T].$$
(3.30)

To this end, we need to introduce relaxed controls and control rules similar to Section 2.

Definition 3.16. We call $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, \underline{Q}, Z)$ 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' $(\mathcal{M}^{\mathbb{P},\phi}, \{\mathcal{F}_t, t \geq 0\}, \mathbb{P})$ is a well defined continuous martingale, where

$$\mathcal{M}_{t}^{\mathbb{P},\phi} = \phi(X_{t}) - \int_{0}^{t} \int_{U} \phi'(X_{s}) b(s, X_{s}, \mathbb{P} \circ X_{s}^{-1}, u) \underline{Q}_{s}(du) ds$$

$$- \frac{1}{2} \int_{0}^{t} \int_{U} \phi'(X_{s}) a(s, X_{s}, \mathbb{P} \circ X_{s}^{-1}, u) \underline{Q}_{s}(du) ds$$

$$- \int_{0}^{t} \phi'(X_{s-}) c(s) dZ_{s} - \sum_{0 \le s \le t} (\phi(X_{s}) - \phi(X_{s-}) - \phi'(X_{s-}) \Delta X_{s}), \ t \in [0, T].$$
(3.31)

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

$$J(r) = E^{\mathbb{P}}\left[\int_0^T \int_U f\left(t, X_t, \mathbb{P} \circ X_t^{-1}, u\right) \underline{Q}_t(du)dt + g\left(X_T, \mathbb{P} \circ X_T^{-1}\right) + \int_0^T h(t) \, dZ_t\right].$$
(3.32)

We still denote by $\Omega := \widetilde{\mathcal{D}}(\mathbb{R}) \times \widetilde{\mathcal{U}}(\mathbb{R}) \times \widetilde{\mathcal{A}}(\mathbb{R})$ the canonical space, \mathcal{F}_t the canonical filtration and (X, Q, Z) the coordinate projections with the associated predictable disintegration Q^o , as introduced in Section 2. The notion of control rules can be defined similarly as that 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 [15, Proposition 2.6] we see that our optimization problems over relaxed controls and over control rules are equivalent. Once the optimal control rule is established, under the same additional assumption as in Remark 2.8, we can establish a strict optimal control from the optimal control rule. 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 general unconstraint case. We do not give a formal proof as the arguments are exactly the same as in the preceding subsection.

Theorem 3.17. Suppose A_4 , A_5 hold and A_1 holds without Lipschitz continuity of b and σ on x. Under a finite-fuel constraint on the singular controls, $\mathcal{R} \neq \emptyset$.

Proof. For each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$, there exists a solution to the martingale problem $\mathcal{M}^{\mu,\phi}$, where $\mathcal{M}^{\mu,\phi}$ is defined in (2.5). Thus, we define a set-valued map Φ on $\mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ with non-empty convex images by

$$\Phi: \mu \to \{\mathbb{P} \circ X^{-1} : \mathbb{P} \in \mathcal{R}(\mu)\}$$

where $\mathcal{R}(\mu)$ is the control rule with μ as in the previous section.

The compactness of $\Phi(\mu)$ for each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and the upper hemi-continuity of Φ are results of the compactness of $\mathcal{R}(\mu)$ for each $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ and upper hemi-continuity of $\mathcal{R}(\cdot)$, respectively, which are direct results of Corollary 3.9.⁵ By analogy to the proof of Theorem 3.12 we can define a non-empty, compact, convex set $\overline{S} \subset \mathcal{P}_p(\widetilde{\mathcal{D}}(\mathbb{R}))$ such that $\Phi : \overline{S} \to 2^{\overline{S}}$. Hence, Φ has a fixed point, due to [2, Corollary 17.55].

Theorem 3.18. Suppose \mathcal{A}_3 - \mathcal{A}_6 hold and that \mathcal{A}_1 holds without Lipschitz assumptions on b and σ in x, 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

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

⁵Note that we only need upper hemi-continuity of $\mathcal{R}(\cdot)$, so Lipschtiz assumptions on b and σ are not necessary.

Proof. It is sufficient to prove \mathcal{R} is compact and J is lower semi-continuous. The former one can be achieved by the same way to Corollary 3.9. As for the lower semi-continuity, note that f and g can be approximated by continuous functions f_N and g_N increasingly. For f_N and g_N , by the same way as that in the proof of Lemma 3.4, one has

$$\liminf_{n \to \infty} E^{\mathbb{P}^n} \left[\int_0^T \int_U f_N(t, X_t, \mathbb{P}^n \circ X_t^{-1}, u) \, Q_t(du) dt + g_N(X_T, \mathbb{P}^n \circ X_T^{-1}) + \int_0^T h(t) \, dZ_t \right]$$

$$\to E^{\mathbb{P}} \left[\int_0^T \int_U f_N(t, X_t, \mathbb{P} \circ X_t^{-1}, u) \, Q_t(du) dt + g_N(X_T, \mathbb{P} \circ X_T^{-1}) + \int_0^T h(t) \, dZ_t \right].$$

Thus, monotone convergence implies the lower semi-continuity of J.

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. For the reasons outlined in Remark 4.2 below we restrict ourselves to MFGs without terminal cost or singular control cost. More precisely, we consider MFGs with singular controls of the form:

 $\begin{cases} 1. & \text{fix a deterministic measure } \mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})); \\ 2. & \text{solve the corresponding stochastic singular control problem :} \\ & \inf_{u,Z} E\left[\int_0^T f(t, X_t, \mu_t, u_t)dt\right] \\ & \text{subject to} \\ & dX_t = b(t, X_t, \mu_t, u_t) dt + \sigma(t, X_t, \mu_t, u_t)dW_t + c(t) dZ_t, \ t \in [0, T+\epsilon]; \\ 3. & \text{solve the fixed point problem: } \mu = Law(X), \end{cases}$ (4.1)

for some fixed $\epsilon > 0$ under the finite-fuel constraint $Z \in \widetilde{\mathcal{A}}_{0,T}^m(\mathbb{R})$. 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 $\widetilde{\mathcal{D}}_{0,T}(\mathbb{R})$.

4.1 Solving MFGs with singular controls using MFGs with regular controls

In this section we establish an approximation of (relaxed) solutions results for the MFGs (4.1) under a finite-fuel constraint by (relaxed) solutions to MFGs with only regular controls. To this end, we associate with each singular control $Z \in \widetilde{\mathcal{A}}_{0,T}^m(\mathbb{R})$ the sequence of absolutely continuous controls

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

Then, $Z^{[n]} \in \widetilde{\mathcal{A}}^m_{0,T+\epsilon}(\mathbb{R})$ 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. However, by Proposition B.1 (3.) and the discussion before Proposition B.4 we do know that

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

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

1. fix a deterministic measure
$$\mu \in \mathcal{P}_p(\mathcal{D}_{0,T+\epsilon}(\mathbb{R}));$$

2. solve the corresponding stochastic control problem :
 $\inf_{u,Z} E\left[\int_0^T f(t, X_t^{[n]}, \mu_t, u_t)dt\right]$
subject to
 $dX_t^{[n]} = b(t, X_t^{[n]}, \mu_t, u_t)dt + \sigma(t, X_t^{[n]}, \mu_t, u_t)dW_t + c(t) dZ_t^{[n]}, t \in [0, T+\epsilon]$
 $X_0^{[n]} = 0$
 $Z_t^{[n]} = n \int_{(t-\frac{1}{n})}^t Z_s ds;$
3. solve the fixed point problem: $\mu = Law(X^{[n]})$

Definition 4.1. We call the vector $r^n = (\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, \underline{Q}, Z^{[n]})$ a relaxed control with respect to μ for some $\mu \in \mathcal{P}_p(\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R}))$ if $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}, X, \underline{Q}, Z)$ satisfies 1.-3. in Definition 2.1 with item 4 being replaced by

4'. X is a $\{\mathcal{F}_t, t \in \mathbb{R}\}$ adapted stochastic process and $X \in \widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$ such that for each $\phi \in \mathcal{C}_b^2(\mathbb{R}^d)$, $\mathcal{M}^{[n],\mu,\phi}$ is a well defined \mathbb{P} continuous martingale, where

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

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

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

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

and (X, Q, Z) being the coordinate projections on $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in \mathbb{R}\})$ and Q^o being the disintegration of Q as in Section 2.1.2.

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 by $\mathcal{R}^{[n]}(\mu)$ 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

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

For each fixed n and μ , denote by $\mathcal{R}^{[n]*}(\mu)$ the set of 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 $MFG^{[n]}$.

The proof of the following theorem is very similar to that of Theorem 3.12 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 $MFG^{[n]}$.

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 3.6 there exists, for each n, a $\{\mathcal{F}_t, 0 \leq t \leq T + \epsilon\}$ adapted continuous process Y^n , such that

$$\mathbb{P}^{[n]}\left(X_t = Y_t^n + \int_0^t c(s) \, dZ_s^{[n]}, \ t \in [0, T + \epsilon]\right) = 1.$$

Arguing as in the proof of Proposition 3.7, there exists 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 $(\widetilde{X}^n, \widetilde{Y}^n, \widetilde{Q}^n, \widetilde{Z}^n) \to (\widetilde{X}, \widetilde{Y}, \widetilde{Q}, \widetilde{Z})$ 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

$$\mathbb{Q}\left(\widetilde{X}_t^n = \widetilde{Y}_t^n + \int_0^t c(s) \, d\widetilde{Z}_s^{[n],n}, \ t \in [0, T+\epsilon]\right) = 1,\tag{4.5}$$

where $\widetilde{Z}_t^{[n],n} = n \int_{(t-1/n)}^t \widetilde{Z}_s^n ds$. 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

$$\left| n \int_{t-\frac{1}{n}}^{t} \widetilde{Z}_{s}^{n}(\widetilde{\omega}) \, ds - \widetilde{Z}_{t}(\widetilde{\omega}) \right| \leq n \int_{t-\frac{1}{n}}^{t} \left| \widetilde{Z}_{s}^{n}(\widetilde{\omega}) - \widetilde{Z}_{s}(\widetilde{\omega}) \right| \, ds + n \int_{t-\frac{1}{n}}^{t} \left| \widetilde{Z}_{s}(\widetilde{\omega}) - \widetilde{Z}_{t}(\widetilde{\omega}) \right| \, ds$$
$$\leq \sup_{t-\frac{1}{n} \leq s \leq t} \left| \widetilde{Z}_{s}^{n}(\widetilde{\omega}) - \widetilde{Z}_{s}(\widetilde{\omega}) \right| + \sup_{t-\frac{1}{n} \leq s \leq t} \left| \widetilde{Z}_{s}(\widetilde{\omega}) - \widetilde{Z}_{t}(\widetilde{\omega}) \right|$$
$$\rightarrow 0.$$

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

$$\mathbb{Q}\left(\widetilde{X}_t = \widetilde{Y}_t + \int_0^t c(s) \, d\widetilde{Z}_s, \ t \in [0, T + \epsilon]\right) = 1.$$
(4.6)

The desired result can be obtained by the same proof as Proposition 3.7.

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

$$dX_t^n = \int_U b(t, X_t^n, \mu_t^{[n]}, u) \, \widetilde{Q}_t(du) dt + \int_U \sigma(t, X_t^n, \mu_t^{[n]}, u) \, \widetilde{M}(du, dt) + c(t) \, d\widetilde{Z}_t^{[n]}.$$

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

$$E^{\widetilde{\mathbb{P}}} \int_{0}^{T} |X_{t}^{n} - \widetilde{X}_{t}|^{2} dt \leq CE^{\widetilde{\mathbb{P}}} \int_{0}^{T} \left| \int_{0}^{t} c(s) d\widetilde{Z}_{s}^{[n]} - \int_{0}^{t} c(s) d\widetilde{Z}_{s} \right|^{2} dt + CE^{\widetilde{\mathbb{P}}} \int_{0}^{T} \left(1 + L(W_{p}(\mu_{t}^{[n]}, \delta_{0}), W_{p}(\mu_{t}^{*}, \delta_{0})) \right)^{2} \mathcal{W}_{p}(\mu_{t}^{[n]}, \mu_{t}^{*})^{2} dt.$$

$$(4.7)$$

 $\widetilde{Z}^{[n]} \to \widetilde{Z}$ in M_1 a.s. implies

$$E^{\widetilde{\mathbb{P}}} \int_0^T \left| \int_0^t c(s) \, d\widetilde{Z}_s^{[n]} - \int_0^t c(s) \, d\widetilde{Z}_s \right|^2 \, dt \to 0.$$

By the same arguments leading to (3.7) in the proof of Lemma 3.4,

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

This yields,

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

Hence, up to a subsequence, dominated convergence implies

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

Moreover, by Lemma 3.4,

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

Altogether, this yields,

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

4.2 Approximating a given solutions to MFGs with singular controls

In this subsection, we show how to approximate a *given* solution to a MFG with singular controls of the form (4.1) introduced in the previous subsection by a sequence of admissible control rules of MFGs with only regular controls.

Let \mathbb{P}^* be any solution to the MFG (4.1). Since $(\Omega, \{\mathcal{F}_t, t \in \mathbb{R}\}, \mathbb{P}^*, X, Q, Z)$ satisfies the associated martingale problem, there exists a tuple $(\hat{X}, \hat{Q}, \hat{Z}, M)$ defined on some extension $(\hat{\Omega}, \{\hat{\mathcal{F}}_t, t \in \mathbb{R}\}, \mathbb{Q})$ of the canonical path space, such that

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

and

$$\mathbb{Q}\left(\widehat{X}_{\cdot}=\int_{0}^{\cdot}\int_{U}b(s,\widehat{X}_{s},\mu_{s}^{*},u)\,\widehat{Q}_{s}(du)ds+\int_{0}^{\cdot}\int_{U}\sigma(s,\widehat{X},\mu_{s}^{*},u)\,M(du,ds)+\int_{0}^{\cdot}c(s)\,d\widehat{Z}_{s}\right)=1.$$
(4.9)

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

$$dX_t^{[n]} = \int_U b(t, X_t^{[n]}, \mu_t^{[n]}, u) \,\widehat{Q}_t(du)dt + \int_U \sigma(t, X_t^{[n]}, \mu_t^{[n]}, u) \, M(du, dt) + c(t) \, d\widehat{Z}_t^{[n]}, \tag{4.10}$$

where $\widehat{Z}^{[n]}$ is defined by (4.2) and $\mu^{[n]}$ is any sequence satisfying $\mu^{[n]} \to \mu^*$ in $\mathcal{W}_{p,(\widetilde{\mathcal{D}}(\mathbb{R}),d_{M_1})}$. One checks immediately that

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

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, which relies on the following lemma. Its proof uses the notion of a parameter representation of the thin graph of a function $x \in \mathcal{D}(0,T)$ introduced in Appendix B.

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

$$dX_t^n = \int_U b(t, X_t^n, \mu_t^n, u) Q_t(du) dt + \int_U \sigma(t, X_t^n, \mu_t^n, u) M(du, dt) + dZ_t^n, \ t \in [0, \widetilde{T}]$$
(4.11)

respectively,

$$dX_t = \int_U b(t, X_t, \mu_t, u) Q_t(du) dt + \int_U \sigma(t, X_t, \mu_t, u) M(du, dt) + dZ_t, \ t \in [0, \widetilde{T}]$$
(4.12)

where \widetilde{T} is a fixed positive constant, b and σ satisfy \mathcal{A}_1 and \mathcal{A}_5 . 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

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

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 Proposition B.1(2), for each $\omega \in \underline{\Omega}$, there exist parameter representations $(u(\omega), r(\omega)) \in \Pi_{Z(\omega)}$ and $(u_n(\omega), r_n(\omega)) \in \Pi_{Z^n(\omega)}$ of $Z(\omega)$ and $Z^n(\omega)$ $(n \in \mathbb{N})$, respectively, such that

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

$$(4.13)$$

Parameter representations with the desired convergence properties are constructed in, e.g., [28, Section 4]; see also [28, Theorem 1.2]. A careful inspection of [28, Section 4] shows that the constructions of $(u(\omega), r(\omega))$ and $(u_n(\omega), r_n(\omega))$ only use measurable operations. As a result the mappings $(u(\cdot), r(\cdot))$ and $(u_n(\cdot), r_n(\cdot))$ are measurable.

We now construct parameter representations $(u_{X^n}(\omega), r_{X^n}(\omega))$ and $(u_X(\omega), r_X(\omega))$ of $X^n(\omega)$ and $X(\omega)$, respectively. Since $X(\omega)$ (resp. $X^n(\omega)$) jumps at the same time as $Z(\omega)$ (resp. $Z^n(\omega)$), we can choose

$$r_X(\omega) = r(\omega), \quad r_{X^n}(\omega) = r_n(\omega).$$

In the following, we will drop the dependence on $\omega \in \underline{\Omega}$, if there is no confusion. By [28, equation (3.1)], parameter representations of X^n and X in terms of the parameter representations of Z^n and Z are given by, respectively,

$$u_{X^{n}}(t) = \int_{0}^{r_{n}(t)} \int_{U} b(s, X^{n}_{s}, \mu^{n}_{s}, u) Q_{s}(du) ds + \int_{0}^{r_{n}(t)} \int_{U} \sigma(s, X^{n}_{s}, \mu^{n}_{s}, u) M(du, ds) + u_{n}(t),$$

 $\quad \text{and} \quad$

$$u_X(t) = \int_0^{r(t)} \int_U b(s, X_s, \mu_s, u) Q_s(du) ds + \int_0^{r(t)} \int_U \sigma(s, X_s, \mu_s, u) M(du, ds) + u(t).$$

Hence, by the Lipschitz property of b and σ and BDG's inequality, we get,

$$E \sup_{0 \le t \le \widetilde{T}} |u_{X^{n}}(t) - u_{X}(t)|$$

$$\leq CE \left(\int_{0}^{\widetilde{T}} |X^{n}(s) - X(s)|^{2} ds \right)^{\frac{1}{2}}$$

$$+ C \left(\int_{0}^{\widetilde{T}} (1 + L(\mathcal{W}_{p}(\mu_{s}^{n}, \delta_{0}), \mathcal{W}_{p}(\mu_{s}, \delta_{0})))^{2} \mathcal{W}_{p}^{2}(\mu_{s}^{n}, \mu_{s}) ds \right)^{\frac{1}{2}} + CE \sup_{0 \le t \le \widetilde{T}} |r_{n}(t) - r(t)| \qquad (4.14)$$

$$+ E \sup_{0 \le t \le \widetilde{T}} \left| \int_{0}^{r_{n}(t)} \int_{U} \sigma(s, X_{s}, \mu_{s}, u) M(du, ds) - \int_{0}^{r(t)} \int_{U} \sigma(s, X_{s}, \mu_{s}, u) M(du, ds) \right|$$

$$+ E \sup_{0 \le t \le \widetilde{T}} |u_{n}(t) - u(t)|.$$

The same argument as in the proof of Theorem 4.6 yields that the first two terms on the right hand side of (4.14) converge to 0 while the last three terms converge to 0 due to (4.13). Thus,

$$\lim_{n \to \infty} E \sup_{0 \le t \le \widetilde{T}} |u_{X^n}(t) - u_X(t)| = 0.$$

Corollary 4.8. Under the assumptions of Proposition 4.7, along a subsequence $\mathbb{P}^{[n]} \to \mathbb{P}^*$ in \mathcal{W}_p .

Proof. For each $\tilde{\epsilon} > 0$, we extend the equations (4.9) and (4.10) by

$$\widehat{X}_s = \int_{-\widetilde{\epsilon}}^s \int_U \widetilde{b}(t, \widehat{X}_t, \mu_t^*, u) \, \widehat{Q}_t(du) dt + \int_{-\widetilde{\epsilon}}^s \int_U \widetilde{\sigma}(t, \widehat{X}_t, \mu_t^*, u) \, M(du, dt) + \int_{-\widetilde{\epsilon}}^s \widetilde{c}(t) \, d\widehat{Z}_t,$$

respectively,

$$X_s^{[n]} = \int_{-\widetilde{\epsilon}}^s \int_U \widetilde{b}(t, X_t^{[n]}, \mu_t^{[n]}, u) \, \widehat{Q}_t(du) dt + \int_{-\widetilde{\epsilon}}^s \int_U \widetilde{\sigma}(t, X_t^{[n]}, \mu_t^{[n]}, u) \, M(du, dt) + \int_{-\widetilde{\epsilon}}^s \widetilde{c}(t) \, d\widehat{Z}_t^{[n]},$$

where

$$\widetilde{b}(s,\cdot) = b(s,\cdot), \ \widetilde{\sigma}(s,\cdot) = \sigma(s,\cdot), \ \widetilde{c}(s) = c(s) \text{ when } s \ge 0; \ \widetilde{b}(s,\cdot) = 0, \ \widetilde{\sigma}(s,\cdot) = 0, \ \widetilde{c}(s) = c(0) \text{ when } s < 0.$$

Moreover, we have that

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

where a.s. in $(\mathcal{A}^m(-\tilde{\epsilon}, T+\epsilon), d_{M_1}),$

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

Since $\int_{-\tilde{\epsilon}}^{\cdot} \tilde{c}^+(t) d\hat{Z}_t$ and $\int_{-\tilde{\epsilon}}^{\cdot} \tilde{c}^-(t) d\hat{Z}_t$ never jump at the same time, Proposition B.7 implies that

$$\int_{-\widetilde{\epsilon}}^{\cdot} \widetilde{c}(t) \, d\widehat{Z}_t^{[n]} \to \int_{-\widetilde{\epsilon}}^{\cdot} \widetilde{c}(t) \, d\widehat{Z}_t$$

a.s. in $(\mathcal{A}^m(-\tilde{\epsilon}, T+\epsilon), d_{M_1})$. Hence, by Proposition 4.7,

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

Hence, up to a subsequence,

$$d_{M_1}(X^{[n]}, \widehat{X}) \to 0 \text{ in } \mathcal{D}(-\widetilde{\epsilon}, T+\epsilon); \quad \mathbb{Q}\text{-a.s.},$$

which implies the same convergence holds in $\widetilde{\mathcal{D}}_{0,T+\epsilon}(\mathbb{R})$. For any nonnegative continuous function ϕ satisfying

$$\phi(x,q,z) \le C(1 + d_{M_1}(x,0)^p + \mathcal{W}_p^p(q/T,\delta_0) + d_{M_1}(z,0)^p),$$

the uniform integrability of $d_{M_1}(X^{[n]}, 0)^p$, $\mathcal{W}_p^p(\widehat{Q}/T, \delta_0)$ and $d_{M_1}(\widehat{Z}, 0)^p$ yields

$$E^{\mathbb{Q}}\phi(X^{[n]},\widehat{Q},\widehat{Z}) \to E^{\mathbb{Q}}\phi(\widehat{X},\widehat{Q},\widehat{Z}).$$

This implies $\mathbb{Q} \circ (X^{[n]}, \widehat{Q}, \widehat{Z})^{-1} \to \mathbb{Q} \circ (\widehat{X}, \widehat{Q}, \widehat{Z})^{-1}$ in $\mathcal{W}_{p,\Omega}$ by [29, Definition 6.8], that is, $\mathbb{P}^{[n]} \to \mathbb{P}^*$ in $\mathcal{W}_{p,\Omega}$.

A Wasserstein distance and representation of martingales

Definition A.1. Let (E, ϱ) 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:

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

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

It is well known [24, Theorem III-10] that for every continuous square integrable martingale m with quadratic variation process $\int_0^{\cdot} \int_U a(s, u) v_s(du) ds$, where $a = \sigma \sigma^{\top}$ and σ is a bounded measurable function and v is $\mathcal{P}(U)$ valued stochastic process, on some extension of the original probability space, there exists a martingale measure M with intensity $v_s(du)ds$ such that $m_{\cdot} = \int_0^{\cdot} \int_U \sigma(t, u) M(du, dt)$. This directly leads to the following proposition, which is frequently used in the main text.

Proposition A.2. The existence of solution \mathbb{P} to the martingale problem (2.5) is equivalent to the existence of the weak solution to the following SDE

$$d\bar{X}_{t} = \int_{U} b(t, \bar{X}_{t}, \mu_{t}, u) \,\bar{Q}_{s}(du) ds + \int_{U} \sigma(t, \bar{X}_{t}, \mu_{t}, u) \,\bar{M}(du, dt) + c(t) \,d\bar{Z}_{t}, \tag{A.2}$$

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 [30]. Note that in [30] 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

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

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

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

and the so called strong M_1 oscillation function is defined as

$$w_s(x,t,\delta) = \sup_{0 \lor (t-\delta) \le t_1 < t_2 < t_3 \le (t+\delta) \land T} |x_{t_2} - [x_{t_1}, x_{t_3}]|,$$
(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,

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

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

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

Remark B.2. (1) When restricted to $\mathcal{C}(0,T)$, the uniform topology is equivalent to the M_1 topology. Indeed, when $x^n \in \mathcal{C}(0,T)$ and $||x^n - x|| \to 0$, then $x \in \mathcal{C}(0,T)$. For any $r : [0,1] \to [0,T]$, define

$$u^{n}(t) := x_{r(t)}^{n}, \qquad u(t) := x_{r(t)}.$$

Thus, (u^n, r) and (u, r) can serve as a parameter representation of x^n and x, respectively. Moreover, it holds that $||u^n - u|| \to 0$. So we have $x^n \to x$ in M_1 by Proposition B.1(2). On the other hand, when $x^n, x \in \mathcal{C}(0,T)$ and $x^n \to x$ in M_1 , by Proposition B.1(2), there exist parameter representations (u^n, r^n) and (u, r) of x^n and x, respectively, such that

$$u^{n}(t) = x^{n}_{r^{n}(t)}, \ u(t) = x_{r(t)}, \ \|u^{n} - u\| \to 0 \text{ and } \|r^{n} - r\| \to 0.$$

To show that $||x^n - x|| \to 0$, it is sufficient to prove that $t^n \to t$ implies $|x_{t^n}^n - x_t| \to 0$. Let $r^n(s^n) = t^n$ and r(s) = t. Then we have

$$|r(s^{n}) - r(s)| \le |r^{n}(s^{n}) - r(s^{n})| + |r^{n}(s^{n}) - r(s)| \to 0.$$

So we get

$$|x_{t^n}^n - x_t| = |u^n(s^n) - u(s)| \le |u^n(s^n) - u(s^n)| + |u(s^n) - u(s)| = |u^n(s^n) - u(s^n)| + |x_{r(s^n)} - x_{r(s)}| \to 0.$$

(2) Proposition B.1(3) implies that $(\mathcal{D}(0,T), d_{M_1})$ convergence is stronger than $L^{\alpha}[0,T]$ convergence, for any $\alpha > 0$. In fact, if $x^n \to x$ in M_1 , then $x_t^n \to x_t$ for a.e. $t \in [0,T]$, due to Proposition B.1(3). Moreover,

 $|x_t^n - x_t|^{\alpha} \le 2^{\alpha} \left(d_{M_1}^{\alpha}(x^n, 0) + d_{M_1}^{\alpha}(x, 0) \right) \to 2^{\alpha + 1} d_{M_1}(x, 0) < \infty.$

Thus, the assertion follows from dominated convergence.

Proposition B.3. 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

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

In [30], 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.3 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.3 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 $\{x^n, n \ge 1\}$ converges to x in $\mathcal{D}(\mathbb{R})$ if and only if the sequences $\{x^n|_{[a,b]}, n \ge 1\}$ converge to $x|_{[a,b]}$ for all a < b at which x is continuous; see [30, Chapter 3].

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

Proposition B.4. The sequence $\{x^n, n \ge 1\} \subseteq \widetilde{\mathcal{D}}(\mathbb{R})$ 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

$$\widetilde{w}_s(x,\delta) = w_s(x,\delta) + \sup_{0 \le s < t \le \delta} |x_s - [0, x_t]|.$$
(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 0^6 .

Corollary B.5. Let $A = \{z \in \widetilde{\mathcal{A}}(\mathbb{R}) : z_T \leq K\}$ for some K > 0. Then A is $(\widetilde{\mathcal{D}}(\mathbb{R}), M_1)$ compact.

Proof. This follows from Proposition B.4 as $w_s(z, t, \delta) = 0$ for each $z \in A, t \in \mathbb{R}$ and $\delta > 0$.

Proposition B.6. A sequence of probability measures $\{\mathbb{P}_n\}_{n\geq 1}$ on $\mathcal{D}(\mathbb{R})$ is tight if and only if (1) for each $\epsilon > 0$, there exists 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

$$\sup_{n} \mathbb{P}_{n}(\widetilde{w}_{s}(x,\delta) \ge \eta) < \epsilon.$$
(B.13)

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.7. 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

$$x^n + y^n \to x + y \text{ in } M_1. \tag{B.14}$$

C Sketch Proof of Proposition 3.6

It is sufficient to establish the equivalence of martingale problems in Definition 2.5 and Proposition 3.6. Only the one-dimensional case is proved; the multi-dimensional case is similar.

Proposition 3.6 \Rightarrow Definition 2.5: Without loss of generality (see [23, Proposition 4.11 and Remark 4.12]), we can take $\phi(y) = y$, y^2 and following the proof of [23, Proposition 4.6], we have that M is a continuous martingale with the quadratic variation

$$\langle M \rangle_t = \int_0^t \int_U a(s, X_s, \mu_s, u) Q_s(du) ds,$$

where

$$M_t = Y_t - \int_0^t \int_U b(s, X_s, \mu_s, u) Q_s(du) ds.$$

By applying Itô's formula to $\phi(X_t)$ and noting $X = Y + \int_0 c(s) dZ_s$, the desired result follows from

$$\phi(X_t) = \phi(X_{0-}) + \int_0^t \int_U \phi'(X_s) b(s, X_s, \mu_s, u) Q_s(du) ds + \frac{1}{2} \int_0^t \int_U \phi''(X_s) a(s, X_s, \mu_s, u) Q_s(du) ds + \int_0^t \phi'(X_{s-}) c(s) dZ_s + \sum_{0 \le s \le t} [\phi(X_s) - \phi(X_{s-}) - \phi'(X_{s-}) \triangle X_s] + \int_0^t \phi'(X_s) dM_s.$$

Definition 2.5 \Rightarrow Proposition 3.6: By Proposition A.2, there exits $(\overline{X}, \overline{Q}, \overline{Z})$ and a martingale measure \overline{M} with intensity \overline{Q} on some extension $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$, s.t. (A.2) holds and $\mathbb{P} \circ (X, Q, Z)^{-1} = \overline{\mathbb{P}} \circ (\overline{X}, \overline{Q}, \overline{Z})^{-1}$. Let

$$\overline{Y}_{\cdot} = \overline{X}_{\cdot} - \int_0^{\cdot} c(s) \, d\overline{Z}_s.$$

⁶Due to the right-continuity of the elements in $\widetilde{\mathcal{D}}(\mathbb{R})$ there is no line segment between T and T+.

Then

$$Y_{\cdot} := X_{\cdot} - \int_0^{\cdot} c(s) \, dZ_s \stackrel{d}{=} \overline{Y}.$$

By applying Itô's formula to $\phi(\overline{Y}_t)$,

$$\phi(\overline{Y}_t) - \int_0^t \int_U \phi'(\overline{Y}_s) b(s, \overline{X}_s, \mu_s, u) \,\overline{Q}_s(du) ds - \frac{1}{2} \int_0^t \int_U \phi''(\overline{Y}_s) a(s, \overline{X}_s, \mu_s, u) \,\overline{Q}_s(du) ds$$

is a martingale. Hence the following is also a martingale:

$$\phi(Y_t) - \int_0^t \int_U \phi'(Y_s) b(s, X_s, \mu_s, u) Q_s(du) ds - \frac{1}{2} \int_0^t \int_U \phi''(Y_s) a(s, X_s, \mu_s, u) Q_s(du) ds$$

References

- S. AHUJA, Wellposedness of Mean Field Games with Common Noise under a Weak Monotonicity Condition, SIAM J. Control Optim., 51 (2016), 30-48.
- [2] C. ALIPRANTIS AND K. BORDER, Infinite Dimensional Analysis: a Hitchhikers Guide, Springer, 2006.
- [3] A. BENSOUSSAN, K. SUNG, P. YAM AND S. YUNG, *Linear Quadratic Mean Field Games*, J. Optim. Theory Appl., 169 (2016), 496-529.
- [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), 131-166.
- [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, X. ZHU, A Probabilistic Approach to Mean Field Games with Major and Minor Players, Ann. Appl. Probab., 26 (2016), 1535-1580.
- [10] P. CHAN, R. SIRCAR, Fracking, Renewables & Mean Field Games, to appear in SIAM Rev., (2016).
- [11] R. ELIE, T. MASTROLIA AND D POSSAMAÏ, A Tale of a Principal and Many Many Agents, arXiv:1608.05226, 2016.
- [12] O. GUEANT, J. LASRY AND P. LIONS, Mean Field Games and Applications, Paris-Princeton Lectures on Mathematical Finance, Springer, 205-266, 2010.
- [13] X. GUO AND J. LEE, Mean Field Games with Singular Controls of Bounded Velocity, arXiv:1703.04437, 2017.

- [14] U. HAUSSMANN, J. LEPELTIER, On the Existence of Optimal Controls, SIAM. J. Control Optim. 28 (1990), 851-902.
- [15] U. HAUSSMANN, W. SUO, Singular Optimal Stochastic Controls I: Existence, SIAM. J. Control Optim. 33 (1995), 916-936.
- [16] U. HORST, F. NAUJOKAT, When to Cross the Spread: Trading in Two-Sided Limit Order Books, SIAM J. Financ. Math., 5 (2014), 278-315.
- [17] U. HORST, J. SCHEINKMAN, Equilibria in Systems of Social Interactions, J. Econ. Theory, 130 (2006), 44-77.
- [18] U. HORST, J. SCHEINKMAN, A Limit Theorem for Systems of Social Interactions, J. Math. Econ., 45 (2009), 609-623.
- [19] Y. HU, B. ØKSENDAL, AND A. SULEM, Singular Mean-Field Control Games with Applications to Optimal Harvesting and Investment Problems, arXiv:1406.1863, 2014.
- [20] 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.
- [21] J. JACOD, J. MÉMIN, Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité, Séminaire de probabilités (Strasbourg), 15 (1981), 529-546.
- [22] S. JAIMUNGAL AND M. NOURIAN, Mean-Field Game Strategies for a Major-Minor Agent Optimal Execution Problem, ssrn.2578733, 2015.
- [23] I. KARATZAS, S.E. SHREVE, Brownian Motion and Stocahstic Calculus, Springer-Verlag, New York, 1991.
- [24] N. EL KAROUI, S. MELEARD, Martingale Measures and Stochastic Calculus, Probab. Theory Related Fields, 84 (1990), 83-101.
- [25] 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.
- [26] D. LACKER, Mean Field Games via Controlled Martingale Problems: Existence of Markovian Equilibria, Stoch. Proc. Appl. 125 (2015), 2856-2894.
- [27] J. LASRY AND P. LIONS, Mean Field Games, Jpn. J. Math., 2 (2007), 229-260.
- [28] G. PANG, W. WHITT, Continuity of a Queueing Integral Representation in the M₁ Topology, Ann. Appl. Prob. 20 (2010), 214-237.
- [29] C. VILLANI, Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften, 338, Springer-Verlag, 2009.
- [30] W. WHITT, Stochastic Process Limits: an Introduction to Stochastic Process Limits and Their Application to Queues, Springer-Verlag, New York, 2002.
- [31] 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