Probability, Uncertainty and Quantitative Risk
Vol. 7, No. 3, 2022, 247–274
https://www.aimsciences.org/journal/2095-9672
DOI:10.3934/puqr.2022015

Mean field games of controls: Propagation of monotonicities

Chenchen Mou[1], Jianfeng Zhang[2,*]

[1]*Department of Mathematics, City University of Hong Kong, Hong Kong SAR, China*
[2]*Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA*

Email: chencmou@cityu.edu.hk, jianfenz@usc.edu

Dedicated to Professor Alain Bensoussan on the occasion of his 80th birthday.

Abstract The theory of Mean Field Game of Controls considers a class of mean field games where the interaction is through the joint distribution of the state and control. It is well known that, for standard mean field games, certain monotonicity conditions are crucial to guarantee the uniqueness of mean field equilibria and then the global wellposedness for master equations. In the literature the monotonicity condition could be the Lasry–Lions monotonicity, the displacement monotonicity, or the anti-monotonicity conditions. In this paper, we investigate these three types of monotonicity conditions for Mean Field Games of Controls and show their propagation along the solutions to the master equations with common noises. In particular, we extend the displacement monotonicity to semi-monotonicity, whose propagation result is new even for standard mean field games. This is the first step towards the global wellposedness theory for master equations of Mean Field Games of Controls.

Keywords Mean field game of controls, Master equation, Lasry–Lions monotonicity, Displacement semi-monotonicity, Anti-monotonicity

2020 Mathematics Subject Classification 35R15, 49N80, 60H30, 91A16, 93E20

1. Introduction

The theory of Mean Field Games (MFGs) was introduced independently by Huang–Caines–Malhamé [9] and Lasry–Lions [32]. Since then, its literature has witnessed an increase in many directions and the theory is extremely rich in applications, including economics [1, 33], engineering [10, 11], finance [30, 31], social science [4, 22], and many others. We refer to Lions [34], Cardaliaguet [12] and Bensoussan–Frehse–Yam [5] for the an introduction to the subject in its early stages and Camona–Delarue [16, 17] and Cardaliaguet–Porretta [14] for more recent developments. Such problems consider the limit behavior of large systems where the agents interact with each other in some symmetric way, with systemic risk as a notable application. The master equation, introduced by Lions [34], characterizes the value of the MFG provided there is a unique mean field equilibrium. This plays the role of the PDE in the standard

Received 26 May 2022; Accepted 11 August 2022; Early access 28 September 2022
*Corresponding author

literature of controls/games, and is a powerful tool in the mean field framework. The main feature of the master equation is that its state variables include a probability measure μ, representing the distribution of the population, so it can be viewed as a PDE on the Wasserstein space of probability measures.

In a standard MFG, the interaction is only through the law of the state. In many applications, however, the interaction could be through the joint law of the state and the control. Such a game is called a Mean Field Game of Controls (MFGCs), which was also called extended MFGs in the early stages. To be precise, let B and B^0 stand for the idiosyncratic and common noises, respectively. Given an $\mathbb{F}^{B^0}$-adapted stochastic measure flow $\{\nu_\cdot\} = \{\nu_t\}_{t\in[0,T]} \subset \mathcal{P}_2(\mathbb{R}^{2d})$, we denote its first marginal by $\mu_t := \pi_{1\#}\nu_t \in \mathcal{P}_2(\mathbb{R}^d)$, where $\pi_1(x,a) = x$ for any $(x,a) \in \mathbb{R}^d \times \mathbb{R}^d$ is a projection. Given the above $\{\nu_\cdot\}$, we minimize the following cost functional over all admissible controls $\alpha : [0,T] \times \mathbb{R}^d \times C([0,T];\mathbb{R}^d) \to \mathbb{R}^d$: for any $\xi \in \mathbb{L}^2(\mathcal{F}_0, \mathbb{R}^d)$,

$$J(\xi, \{\nu_\cdot\}; \alpha) := \mathbb{E}\Big[G(X_T^{\xi,\{\nu_\cdot\},\alpha}, \mu_T) + \int_0^T f(X_t^{\xi,\{\nu_\cdot\},\alpha}, \alpha(t, X_t^{\xi,\{\nu_\cdot\},\alpha}, B^0_{[0,t]}), \nu_t)\mathrm{d}t\Big], \tag{1.1}$$

where, for a constant $\beta \geq 0$,

$$X_t^{\xi,\{\nu_\cdot\},\alpha} = \xi + \int_0^t b(X_s^{\xi,\{\nu_\cdot\},\alpha}, \alpha(s, X_s^{\xi,\{\nu_\cdot\},\alpha}, B^0_{[0,s]}), \nu_s)\mathrm{d}s + B_t + \beta B_t^0. \tag{1.2}$$

Here the running drift and cost b, f depend on the joint law of the state and control, while the terminal cost G depends on the law of the state only. We call $(\alpha^*, \{\nu^*_\cdot\})$ a Nash equilibrium if

$$\alpha^* \in \arg\min_\alpha J(\xi, \{\nu^*_\cdot\}; \alpha), \quad \text{and} \quad \nu_t^* = \mathcal{L}_{(X_t^{\xi,\{\nu^*_\cdot\},\alpha^*}, \alpha_t^*)|\mathcal{F}_t^{B^0}}.$$

Introduce the Hamiltonian H as

$$H(x,p,\nu) := \inf_{a\in\mathbb{R}^d} \big[p \cdot b(x,a,\nu) + f(x,a,\nu)\big], \text{ with an optimal argument } a^* = \phi(x,p,\nu). \tag{1.3}$$

The above problem leads to the following MFGC system of forward-backward stochastic partial differential equations (FBSPDEs) with a solution $(\{\mu_\cdot\}, \{\nu_\cdot\}, u, v)$:

$$\begin{aligned}
&\mathrm{d}\mu_t(x) = \Big[\frac{\hat\beta^2}{2}\mathrm{tr}(\partial_{xx}\mu_t(x)) - div\big(\mu_t(x)\partial_p H(x, \partial_x u(t,x), \nu_t)\big)\Big]\mathrm{d}t - \beta\partial_x\mu_t(x)\cdot \mathrm{d}B_t^0;\\
&\mathrm{d}u(t,x) = v(t,x)\cdot \mathrm{d}B_t^0 - \Big[\frac{\hat\beta^2}{2}\mathrm{tr}(\partial_{xx}u(t,x)) + \beta\mathrm{tr}(\partial_x v^\top(t,x)) + H(x, \partial_x u(t,x), \nu_t)\Big]\mathrm{d}t;\\
&\nu_t = (id, \phi(\cdot, \partial_x u(t,\cdot), \nu_t))_{\#}\mu_t; \qquad \hat\beta^2 = 1 + \beta^2;\\
&\mu_0 = \mathcal{L}_\xi, \quad u(T,x) = G(x, \mu_T).
\end{aligned} \tag{1.4}$$

The wellposedness of the above MFGC system has been investigated by many authors in recent years, essentially in the case $\beta = 0$ and $b(x,a,\nu) = a$. For example, Gomes–Patrizi–Voskanyan [24], Kobeissi [29], and Graber–Mayorga [27] investigated the system under some smallness conditions. The global wellposedness (especially the uniqueness) was studied by Gomes–Voskanyan [25, 26], Carmona–Lacker [18], Carmona–Delarue [16], Cardaliaguet–Lehalle [13], and Kobeissi [28], under the crucial Lasry–Lions monotonicity condition. We also refer to Djete [20] for some convergence analysis from N-player games to MFGCs and Achdou–Kobeissi [2] for some numerical studies of MFGCs, without requiring the uniqueness of the equilibria. However, to the best of our knowledge, the wellposedness of master equations for MFGCs remains completely open. We recall that the master equation is the PDE to characterize the value function V of the MFGC, provided the equilibrium is unique, and it also serves as the decoupling function V of the MFGC system (1.4):

$$u(t,x) = V(t,x,\mu_t).$$

The monotonicity condition is used to guarantee the uniqueness of the mean field equilibria, and then the global wellposedness of MFG master equations. There are three types of monotonicity conditions in the literature for master equations of standard MFGs: the Lasry–Lions monotonicity, the displacement monotonicity, and the anti-monotonicity. The Lasry–Lions monotonicity, introduced by Lions [34] and used extensively in the literature, can be formulated as follows: for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1;\mathbb{R}^d)$ and their independent copies $\tilde{\xi}, \tilde{\eta}$ in the probability space $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}})$ (see their definitions in section 2),

$$\tilde{\mathbb{E}}\Big[\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \tilde{\xi})\tilde{\eta}, \eta\rangle\Big] \geq 0. \tag{1.5}$$

The displacement monotonicity, originating in Ahuja [3], is

$$\tilde{\mathbb{E}}\Big[\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \tilde{\xi})\tilde{\eta}, \eta\rangle + \langle \partial_{xx} G(\xi, \mathcal{L}_\xi)\eta, \eta\rangle\Big] \geq 0, \tag{1.6}$$

which can be further weakened to the displacement semi-monotonicity: for some constant $\lambda \geq 0$,

$$\tilde{\mathbb{E}}\Big[\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \tilde{\xi})\tilde{\eta}, \eta\rangle + \langle \partial_{xx} G(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \lambda|\eta|^2\Big] \geq 0. \tag{1.7}$$

See, e.g., Bensoussan–Graber–Yam [6] and Gangbo–Meszaros–Mou–Zhang [21]. Note that if G is Lasry–Lions monotone and $\partial_{xx} G$ is bounded, then G is displacement semi-monotone. The anti-monotonicity, recently introduced by the authors [35], takes the following form:

$$\begin{aligned}\tilde{\mathbb{E}}\Big[&\lambda_0\langle \partial_{xx} G(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \lambda_1\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \tilde{\xi})\tilde{\eta}, \eta\rangle \\ &+ |\partial_{xx} G(\xi, \mathcal{L}_\xi)\eta|^2 + \lambda_2\Big|\tilde{\mathbb{E}}[\partial_{x\mu} G(\xi, \mathcal{L}_\xi, \tilde{\xi})\tilde{\eta}]\Big|^2 - \lambda_3|\eta|^2\Big] \leq 0,\end{aligned} \tag{1.8}$$

for some appropriate constants $\lambda_0 > 0$, $\lambda_1 \in \mathbb{R}$, $\lambda_2 > 0$, and $\lambda_3 \geq 0$.

In [21, 35] we made a simple but crucial observation: the propagation of a monotonicity is crucial for the global wellposedness of the (standard) MFG master equations. That is, provided the terminal condition G satisfies one of the above three types of monotonicity conditions, if one can show a priori that any classical solution V of the master equation satisfies the same type of monotonicity for all time t, then one can establish the global wellposedness of the master equation, which in turn will imply the uniqueness of mean field equilibria and the convergence from the N-player game to the MFG. Our goal is to extend all of these results to MFGCs, however, in this paper we focus only on the propagation of these three types of monotonicities. That is, we follow the approach in [21, 35] to find sufficient conditions on the Hamiltonian H (or alternatively on b and f) so that the monotonicity of G can be propagated along $V(t,\cdot,\cdot)$, provided the master equation has a classical solution V. We leave the global wellposedness of the master equations and the convergence of the N-player games to an accompanying paper.

The Lasry–Lions monotonicity condition has been used to study the MFGC system (1.4), as mentioned earlier. It is observed in [21] that, for standard MFGs with non-separable f, the Lasry–Lions monotonicity can hardly be propagated. The extra dependence on the law of the control actually helps for propagating the Lasry–Lions monotonicity, in particular, the separability of f is not required anymore.

The displacement semi-monotonicity condition was introduced in [21], however, only the propagation of displacement monotonicity is established there. In this paper, we manage to propagate the displacement semi-monotonicity for MFGCs, so it improves the result of [21] even for standard MFGs. In particular, by combining with the arguments in [21], we easily obtain the global wellposedness result of standard MFG master equations under displacement semi-monotonicity conditions. We remark again that the displacement semi-monotonicity is weaker than both displacement monotonicity and Lasry–Lions monotonicity (provided $\partial_{xx}V$ is bounded, which is typically the case), so in this sense our result provides a unified framework for the wellposedness theory of master equations under Lasry–Lions monotonicity and displacement monotonicity conditions.

Another feature of our results is that we allow for a general form of the drift b. In the literature, one typically sets $b(\cdot,a,\cdot)=a$ (or slightly more general forms), and then focuses on appropriate monotonicity conditions of f to ensure the uniqueness of the mean field equilibria and/or the wellposedness of the master equations. However, for a general b, especially when b depends on the law (of the state and/or the control), it does not make sense to propose monotonicity condition on f alone. A conceivable notion of monotonicity on the general b has never been studied, to our best knowledge. Our approach works on the Hamiltonian H directly, which has the mixed impacts of b and f together. Again, our results are new in this aspect even for standard MFGs.

The rest of the paper is organized as follows. In section 2, we introduce MFGCs. In section 3, we introduce the master equation and the notions of monotonicities. In sections 4, 5, and 6, we propagate each of the three types of monotonicities in each section, respectively. In particular, in subsection 5.1, we also establish the global wellposedness of standard MFG master equations under displacement semi-monotonicity conditions. Finally, some technical proofs are included in the appendix.

2. Mean field games of controls

We consider the setting in [21]. Let d be a dimension and $[0,T]$ a fixed finite-time horizon. Let $(\Omega_0,\mathbb{F}^0,\mathbb{P}_0)$ and $(\Omega_1,\mathbb{F}^1,\mathbb{P}_1)$ be two filtered probability spaces, on which are defined d-dimensional Brownian motions B^0 and B, respectively. For $\mathbb{F}^i=\{\mathcal{F}^i_t\}_{0\le t\le T}$, $i=0,1$, we assume $\mathcal{F}^0_t=\mathcal{F}^{B^0}_t$, $\mathcal{F}^1_t=\mathcal{F}^1_0\vee\mathcal{F}^B_t$, and $\mathbb{P}_1$ has no atom in $\mathcal{F}^1_0$ so it can support any measure on $\mathbb{R}^d$ with finite second-order moment. Consider the product spaces

$$\Omega:=\Omega_0\times\Omega_1,\quad \mathbb{F}=\{\mathcal{F}_t\}_{0\le t\le T}:=\{\mathcal{F}^0_t\otimes\mathcal{F}^1_t\}_{0\le t\le T},\quad \mathbb{P}:=\mathbb{P}_0\otimes\mathbb{P}_1,\quad \mathbb{E}:=\mathbb{E}^{\mathbb{P}}. \tag{2.1}$$

In particular, $\mathcal{F}_t:=\sigma(A_0\times A_1: A_0\in\mathcal{F}^0_t, A_1\in\mathcal{F}^1_t\}$ and $\mathbb{P}(A_0\times A_1)=\mathbb{P}_0(A_0)\mathbb{P}_1(A_1)$. We automatically extend $B^0,B,\mathbb{F}^0,\mathbb{F}^1$ to the product space in the obvious sense, but use the same notation. Note that B^0 and B^1 are independent $\mathbb{P}$-Brownian motions and are independent of $\mathcal{F}_0$.

For convenience, we introduce another filtered probability space $(\tilde\Omega_1,\tilde{\mathbb{F}}^1,\tilde B,\tilde{\mathbb{P}}_1)$ in the same manner as $(\Omega_1,\mathbb{F}^1,B,\mathbb{P}_1)$, and consider the larger filtered probability space given by

$$\tilde\Omega:=\Omega\times\tilde\Omega_1,\quad \tilde{\mathbb{F}}=\{\tilde{\mathcal{F}}_t\}_{0\le t\le T}:=\{\mathcal{F}_t\otimes\tilde{\mathcal{F}}^1_t\}_{0\le t\le T},\quad \tilde{\mathbb{P}}:=\mathbb{P}\otimes\tilde{\mathbb{P}}_1,\quad \tilde{\mathbb{E}}:=\mathbb{E}^{\tilde{\mathbb{P}}}. \tag{2.2}$$

Given an $\mathcal{F}_t$-measurable random variable $\xi=\xi(\omega^0,\omega^1)$, we say $\tilde\xi=\tilde\xi(\omega^0,\tilde\omega^1)$ is a conditionally independent copy of ξ if, for each ω^0, the $\mathbb{P}_1$-distribution of $\xi(\omega^0,\cdot)$ is equal to the $\tilde{\mathbb{P}}_1$-distribution of $\tilde\xi(\omega^0,\cdot)$. That is, conditional on $\mathcal{F}^0_t$, by extending to $\tilde\Omega$ the random variables ξ and $\tilde\xi$ are conditionally independent and have the same conditional distribution under $\tilde{\mathbb{P}}$. Note that, for any appropriate deterministic function φ,

$$\tilde{\mathbb{E}}_{\mathcal{F}_t^0}\big[\varphi(\xi,\tilde{\xi})\big](\omega^0) = \mathbb{E}^{\mathbb{P}_1\otimes\tilde{\mathbb{P}}_1}\Big[\varphi\big(\xi(\omega^0,\cdot),\tilde{\xi}(\omega^0,\tilde{\cdot})\big)\Big], \quad \mathbb{P}_0\text{-a.e. } \omega^0;$$
$$\tilde{\mathbb{E}}_{\mathcal{F}_t}\big[\varphi(\xi,\tilde{\xi})\big](\omega^0,\omega^1) = \mathbb{E}^{\tilde{\mathbb{P}}_1}\Big[\varphi\big(\xi(\omega^0,\omega^1),\tilde{\xi}(\omega^0,\tilde{\cdot})\big)\Big], \quad \mathbb{P}\text{-a.e. } (\omega^0,\omega^1). \tag{2.3}$$

Here, $\mathbb{E}^{\tilde{\mathbb{P}}_1}$ is the expectation on $\tilde{\omega}^1$, and $\mathbb{E}^{\mathbb{P}_1\times\tilde{\mathbb{P}}_1}$ is on $(\omega^1,\tilde{\omega}^1)$. Throughout this paper, we use the probability space $(\Omega,\mathbb{F},\mathbb{P})$. However, when conditionally independent copies of random variables or processes are needed, we implicitly use the extension to the larger space $(\tilde{\Omega},\tilde{\mathbb{F}},\tilde{\mathbb{P}})$.

When we need two conditionally independent copies, we further introduce $(\bar{\Omega}_1,\bar{\mathbb{F}}^1,\bar{B},\bar{\mathbb{P}}_1)$ and the product space $(\bar{\Omega},\bar{\mathbb{F}},\bar{\mathbb{P}},\bar{\mathbb{E}})$ as in (2.2), and set the joint product space

$$\bar{\tilde{\Omega}} := \Omega\times\tilde{\Omega}_1\times\bar{\Omega}_1,\ \bar{\tilde{\mathbb{F}}} = \{\bar{\tilde{\mathcal{F}}}_t\}_{0\le t\le T} := \{\mathcal{F}_t\otimes\tilde{\mathcal{F}}_t^1\otimes\bar{\mathcal{F}}_t^1\}_{0\le t\le T},\ \bar{\tilde{\mathbb{P}}} := \mathbb{P}\otimes\tilde{\mathbb{P}}_1\otimes\bar{\mathbb{P}}_1,\ \bar{\tilde{\mathbb{E}}} := \mathbb{E}^{\bar{\tilde{\mathbb{P}}}}. \tag{2.4}$$

For any dimension k and any constant $p\ge 1$, let $\mathcal{P}(\mathbb{R}^k)$ denote the set of probability measures on $\mathbb{R}^k$, and $\mathcal{P}_p(\mathbb{R}^k)$ the subset of $\mu\in\mathcal{P}(\mathbb{R}^k)$ with finite p-th moment, equipped with the p-Wasserstein distance W_p. Moreover, for any sub-σ-algebra $\mathcal{G}\subset\mathcal{F}_T$, $\mathbb{L}^p(\mathcal{G})$ denotes the set of $\mathbb{R}^k$-valued, $\mathcal{G}$-measurable, and p-integrable random variables; and for any $\mu\in\mathcal{P}_p(\mathbb{R}^k)$, $\mathbb{L}^p(\mathcal{G};\mu)$ denotes the set of $\xi\in\mathbb{L}^p(\mathcal{G})$ with law $\mathcal{L}_\xi=\mu$. Similarly, for any sub-filtration $\mathbb{G}\subset\mathbb{F}$, $\mathbb{L}(\mathbb{G};\mathbb{R}^k)$ denotes the set of $\mathbb{G}$-progressively measurable $\mathbb{R}^k$-valued processes.

For a continuous function $U:\mathcal{P}_2(\mathbb{R}^k)\to\mathbb{R}$, we recall its linear functional derivative $\frac{\delta U}{\delta\mu}:\mathcal{P}_2(\mathbb{R}^k)\times\mathbb{R}^k\to\mathbb{R}$ and Lions derivative $\partial_\mu U:\mathcal{P}_2(\mathbb{R}^k)\times\mathbb{R}^k\to\mathbb{R}^k$. We say $U\in\mathcal{C}^1(\mathcal{P}_2(\mathbb{R}^k))$ if $\partial_\mu U$ exists and is continuous on $\mathcal{P}_2(\mathbb{R}^k)\times\mathbb{R}^k$, and we note that $\partial_\mu U(\mu,\tilde{x}) = \partial_{\tilde{x}}\frac{\delta U}{\delta\mu}(\mu,\tilde{x})$. Similarly, we define the second-order derivative $\partial_{\mu\mu}U(\mu,\tilde{x},\bar{x})$, and we say $U\in\mathcal{C}^2(\mathcal{P}_2(\mathbb{R}^k))$ if $\partial_\mu U$, $\partial_{\tilde{x}\mu}U$ and $\partial_{\mu\mu}U$ exist and are continuous. We refer to [16, Chapter 5] or [23] for further details.

Our mean field game of controls (MFGC) depend on the following data:

$$b:\mathbb{R}^{2d}\times\mathcal{P}_2(\mathbb{R}^{2d})\to\mathbb{R}^d;\quad f:\mathbb{R}^{2d}\times\mathcal{P}_2(\mathbb{R}^{2d})\to\mathbb{R};\quad G:\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R};\ \text{and } \beta\in[0,\infty).$$

We always assume appropriate technical conditions so that all of the equations in this section are well posed and all of the involved random variables are integrable. Given $t_0\in[0,T]$, denote $B_t^{t_0} := B_t - B_{t_0}$, $B_t^{0,t_0} := B_t^0 - B_{t_0}^0$, $t\in[t_0,T]$. Let $\mathcal{A}_{t_0}$ denote the set of admissible controls $\alpha:[t_0,T]\times\mathbb{R}^d\times C([t_0,T];\mathbb{R}^d)\to\mathbb{R}^d$ which are progressively measurable and adapted in the path variable and square integrable; and $\mathbb{L}^2(\mathbb{F}^{B^{0,t_0}};\mathcal{P}_2(\mathbb{R}^{2d}))$ the set of $\mathbb{F}^{B^{0,t_0}}$-progressively measurable stochastic measure flows $\{\nu_\cdot\} = \{\nu_t\}_{t\in[t_0,T]}\subset\mathcal{P}_2(\mathbb{R}^{2d})$. Here, for notational simplicity, we assume that the controls also take values in $\mathbb{R}^d$, and b and f do not depend on time, but one can remove these constraints without any difficulty.

Given $t_0\in[0,T]$, $x\in\mathbb{R}^d$, $\alpha\in\mathcal{A}_{t_0}$, and $\{\nu_\cdot\}\in\mathbb{L}^2(\mathbb{F}^{B^{0,t_0}};\mathcal{P}_2(\mathbb{R}^{2d}))$, the state of the agent satisfies the following controlled SDE on $[t_0,T]$:

$$X_t^{\{\nu_\cdot\},\alpha} = x + \int_{t_0}^t b(X_s^{\{\nu_\cdot\},\alpha},\alpha_s,\nu_s)\mathrm{d}s + B_t^{t_0} + \beta B_t^{0,t_0};$$
$$\text{where}\quad X^{\{\nu_\cdot\},\alpha} = X^{t_0,\{\nu_\cdot\};x,\alpha},\quad \alpha_s := \alpha(s,X_s^{\{\nu_\cdot\},\alpha},B^{0,t_0}_{[t_0,s]}). \tag{2.5}$$

Consider the expected cost for the MFGC: denoting by $\pi_{1\#}\nu_t$ the first component of ν_t,

$$J(t_0,x;\{\nu_\cdot\},\alpha) := \inf_{\alpha\in\mathcal{A}_{t_0}}\mathbb{E}\Big[G(X_T^{\{\nu_\cdot\},\alpha},\pi_{1\#}\nu_T) + \int_{t_0}^T f(X_t^{\{\nu_\cdot\},\alpha},\alpha_t,\nu_t)\mathrm{d}t\Big]. \tag{2.6}$$

Definition 2.1 *For any $(t,\mu)\in[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$, we say $(\alpha^*,\{\nu_\cdot^*\})\in\mathcal{A}_t\times\mathbb{L}^2(\mathbb{F}^{B^{0,t_0}};\mathcal{P}_2(\mathbb{R}^{2d}))$ is a mean field equilibrium (MFE) at (t,μ) if*

$$
\begin{aligned}
&J(t,x;\{\nu^*_\cdot\},\alpha^*) = \inf_{\alpha\in\mathcal{A}_t} J(t,x;\{\nu^*_\cdot\},\alpha), \quad \textit{for } \mu\textit{-a.e. } x\in\mathbb{R}^d;\\
&\pi_{1\#}\nu^*_t = \mu, \quad \nu^*_s := \mathcal{L}_{(X^*_s,\alpha^*(s,X^*_s,B^{0,t_0}_{[t_0,s]}))|\mathcal{F}^0_s}, \quad \textit{where}\\
&X^*_t = \xi + \int_{t_0}^t b(X^*_s,\alpha^*(s,X^*_s,B^{0,t_0}_{[t_0,s]}),\nu^*_s)\mathrm{d}s + B^{t_0}_t + \beta B^{0,t_0}_t, \quad \xi\in\mathbb{L}^2(\mathcal{F}^1_t,\mu).
\end{aligned}
\tag{2.7}
$$

When there is a unique MFE for each $(t,\mu)\in[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$, denoted as $(\alpha^*(t,\mu;\cdot),\{\nu^*_\cdot(t,\mu)\})$, then the game problem leads to the following value function for the agent:

$$
V(t,x,\mu) := J(t,x;\{\nu^*_\cdot(t,\mu)\},\alpha^*(t,\mu;\cdot)) \quad \text{for any } x\in\mathbb{R}^d. \tag{2.8}
$$

We note that, by (2.7), the above V is well defined only for μ-a.e. x. However, for each t, its continuous extension to $\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)$ is unique, and we always consider this continuous extension. Our goal is to study the master equation for the value function $V(t,x,\mu)$.

For this purpose, we introduce the Hamiltonian: for $(x,p,\nu)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^{2d})$,

$$
H(x,p,\nu) := \inf_{a\in\mathbb{R}^d} h(x,p,\nu,a), \quad h(x,p,\nu,a) := p\cdot b(x,a,\nu) + f(x,a,\nu). \tag{2.9}
$$

Note that H depends on ν, while V depends only on $\mu=\pi_{1\#}\nu$. We also remark that the Hamiltonian in [21, 35] is $-H$. To introduce the master equation, which we do in the next section, we need the following fixed point.

Assumption 2.2 (i) *The Hamiltonian H has a unique minimizer $a^*=\phi(x,p,\nu)$, namely,*

$$
H(x,p,\nu) = h(x,p,\nu,\phi(x,p,\nu)). \tag{2.10}
$$

(ii) *For any $\xi\in\mathbb{L}^2(\mathcal{F})$ and $\eta\in\mathbb{L}^2(\sigma(\xi))$, the following mapping on $\mathcal{P}_2(\mathbb{R}^{2d})$:*

$$
\mathcal{I}^{\xi,\eta}(\nu) := \mathcal{L}_{(\xi,\phi(\xi,\eta,\nu))} \tag{2.11}
$$

has a unique fixed point ν^*: $\mathcal{I}^{\xi,\eta}(\nu^*)=\nu^*$, denoted as $\Phi(\mathcal{L}_{(\xi,\eta)})$.

We refer to [16, Lemma 4.60] for some sufficient conditions on the existence of Φ. By (2.10), one can easily check that

$$
b(x,\phi(x,p,\nu),\nu) = \partial_p H(x,p,\nu), \quad f(x,\phi(x,p,\nu),\nu) = H(x,p,\nu) - p\cdot\partial_p H(x,p,\nu). \tag{2.12}
$$

As in the standard MFG theory, provided V is smooth, p corresponds to $\partial_x V(t,x,\mu)$. Consequently, later on the above fixed point is applied as follows: given (t,μ) and $\xi\in\mathbb{L}^2(\mathcal{F}^1_t,\mu)$,

$$
\eta = \partial_x V(t,\xi,\mu), \quad \nu^* := \Phi(\mathcal{L}_{\xi,\partial_x V(t,\xi,\mu)}), \quad \alpha^* := \phi(\xi,\partial_x V(t,\xi,\mu),\Phi(\mathcal{L}_{\xi,\partial_x V(t,\xi,\mu)})). \tag{2.13}
$$

Pluging these into (2.7), we obtain the following McKean–Vlasov SDE (Stochastic Differential Equation): recalling (2.12),

$$
\begin{aligned}
&X^*_t = \xi + \int_{t_0}^t \partial_p H\big(X^*_s,\partial_x V(s,X^*_s,\mu^*_s),\nu^*_s)\mathrm{d}s + B^{t_0}_t + \beta B^{0,t_0}_t,\\
&\text{where} \quad \mu^*_s := \mathcal{L}_{X^*_s|\mathcal{F}^0_s}, \quad \nu^*_s := \Phi\big(\mathcal{L}_{(X^*_s,\partial_x V(s,X^*_s,\mu^*_s))|\mathcal{F}^0_s}\big).
\end{aligned}
\tag{2.14}
$$

That is, if V is smooth, then under Assumption 2.2 we may obtain the unique MFE α^* through (2.13) and (2.14) (by abusing the notation α^*): given (t_0,μ) and $\xi\in\mathbb{L}^2(\mathcal{F}^1_{t_0},\mu)$,

$$
\alpha^*(s,x,B^{0,t_0}_{[t_0,s]}) = \phi(x,\partial_x V(s,x,\mu^*_s),\nu^*_s). \tag{2.15}
$$

Here we used the fact that μ^*_s,ν^*_s are actually adapted to the shifted filtration generated by B^{0,t_0}.

Assumption 2.2 (i) is generally standard in the literature, for example, when h in (2.9) is convex in a. In particular, when $b(x,a,\nu)=a$, which is often the case, we have $\phi=\partial_p H$. We now provide two examples for Assumption 2.2 (ii).

Example 2.3 *Assume* b, f *are separable in the following sense:*

$$b(x,a,\nu) = b_0(x,a,\pi_{1\#}\nu) + b_1(x,\nu), \quad f(x,a,\nu) = f_0(x,a,\pi_{1\#}\nu) + f_1(x,\nu). \tag{2.16}$$

In this case, (2.9) *becomes*

$$H(x,p,\nu) = H_0(x,p,\pi_{1\#}\nu) + H_1(x,p,\nu),$$

$$\text{where } H_0(x,p,\mu) := \inf_{a\in\mathbb{R}^d}\big[p\cdot b_0(x,a,\mu) + f_0(x,a,\mu)\big], \quad H_1(x,p,\nu) := p\cdot b_1(x,\nu) + f_1(x,\nu).$$

Assume Assumption 2.2 (i) holds, and clearly in this case we have $a^* = \phi(x,p,\mu)$, *with the dependence on* ν *only through its first component* $\mu = \pi_{1\#}\nu$. *Then* $\mathcal{I}^{\xi,\eta}(\nu) := \mathcal{L}_{(\xi,\phi(\xi,\eta,\pi_{1\#}\nu))}$. *Further note that the fixed point requires* $\pi_{1\#}\nu^* = \mathcal{L}_\xi$. *Now, Assumption 2.2 (ii) holds with* $\Phi(\mathcal{L}_{(\xi,\eta)}) = \mathcal{L}_{(\xi,\phi(\xi,\eta,\mathcal{L}_\xi))}$.

We note that the above f satisfies the conditions in [16, Lemma 4.60], while the drift b is more general. The next example, however, is out of the scope of [16, Lemma 4.60].

Example 2.4 *Assume* $d = 1$ *and, by writing* $\mathbb{E}_\nu[\alpha]$ *to indicate expectation under law* $\mathcal{L}_\alpha = \pi_{2\#}\nu$,

$$b(x,a,\nu) = -b_0(x,\pi_{1\#}\nu)a + b_1(x,\nu), \quad f(x,a,\nu) = \frac{|a|^2}{2} - af_0(x,\pi_{1\#}\nu,\mathbb{E}_\nu[\alpha]) + f_1(x,\nu). \tag{2.17}$$

One sees that $\phi(x,p,\nu) = f_0(x,\pi_{1\#}\nu,\mathbb{E}_\nu[\alpha]) + pb_0(x,\pi_{1\#}\nu)$ *and thus*

$$\mathcal{I}^{\xi,\eta}(\nu) := \mathcal{L}_{(\xi,\ f_0(\xi,\pi_{1\#}\nu,\mathbb{E}_\nu[\alpha]) + b_0(\xi,\pi_{1\#}\nu)\eta)}.$$

Then, $\mathcal{I}^{\xi,\eta}$ *has a fixed point if and only if the following mapping has a fixed point:*

$$m \in \mathbb{R} \to \psi^{\xi,\eta}(m) := \mathbb{E}\Big[f_0(\xi,\mathcal{L}_\xi,m) + b_0(\xi,\mathcal{L}_\xi)\eta\Big]. \tag{2.18}$$

Assume $\partial_m f_0 \le 1-\varepsilon$ *for some* $\varepsilon > 0$, *in particular, if* f_0 *is decreasing in* m, *then* $\partial_m\psi^{\xi,\eta} \le 1-\varepsilon$ *and thus* $\psi^{\xi,\eta}$ *has a unique fixed point* $m^* = \varphi(\mathcal{L}_{(\xi,\eta)})$. *Therefore,* $\mathcal{I}^{\xi,\eta}$ *has a unique fixed point:*

$$\Phi(\mathcal{L}_{(\xi,\eta)}) = \mathcal{L}_{(\xi,\ f_0(\xi,\mathcal{L}_\xi,\varphi(\mathcal{L}_{(\xi,\eta)})) + b_0(\xi,\mathcal{L}_\xi)\eta)}. \tag{2.19}$$

2.1 Derivatives of measure valued functions

Note that Φ is a mapping from $\mathcal{P}_2(\mathbb{R}^{2d})$ to $\mathcal{P}_2(\mathbb{R}^{2d})$. Consider an arbitrary dimension k. In this subsection we introduce the linear functional derivative of functions mapping from $\mathcal{P}_2(\mathbb{R}^k)$ to $\mathcal{P}_2(\mathbb{R}^k)$, which is interesting in its own right. We refer to [16, equation (5.52)] for the linear functional derivative of functions mapping from $\mathcal{P}_2(\mathbb{R}^k)$ to $\mathbb{R}$. Let $\mathcal{S}(\mathbb{R}^k)$ denote the Schwartz space, namely, the set of smooth functions $u \in \mathcal{C}^\infty(\mathbb{R}^k;\mathbb{R})$ such that u and all its derivatives decrease rapidly when $|x| \to \infty$; and let $\mathcal{S}'(\mathbb{R}^k)$ denote its dual space, namely, the space of tempered distributions.

Definition 2.5 *Consider a mapping* $\Phi : \mathcal{P}_2(\mathbb{R}^k) \to \mathcal{P}_2(\mathbb{R}^k)$. *We say* $\frac{\delta\Phi}{\delta\rho} : \mathcal{P}_2(\mathbb{R}^k) \times \mathbb{R}^k \to \mathcal{S}'(\mathbb{R}^k)$ *is the linear functional derivative of* Φ *if, for any* $\psi \in \mathcal{S}(\mathbb{R}^k)$,

$$\frac{\delta\Psi}{\delta\rho}(\rho,x) = \Big\langle \frac{\delta\Phi}{\delta\rho}(\rho,x),\ \psi\Big\rangle, \quad \text{where} \quad \Psi(\rho) := \int_{\mathbb{R}^k}\psi(x)\Phi(\rho;\mathrm{d}x). \tag{2.20}$$

We note that $\frac{\delta\Phi}{\delta\rho}(\rho,x)$ is well defined for ρ-a.e. x.

For later applications we require $\frac{\delta\Phi}{\delta\rho}$ to have stronger properties. For this purpose, let $\mathcal{SM}_2(\mathbb{R}^k)$ denote the set of the square integrable signed measures of bounded variation on $\mathbb{R}^k$.

That is, m has the unique decomposition $m = m_1 - m_2$ and $\int_{\mathbb{R}^k}(1+|y|^2)|m|(\mathrm{d}y) < \infty$, where m_1, m_2 are mutually singular non-negative measures on $\mathbb{R}^k$, and $|m|(\mathrm{d}y) := m_1(\mathrm{d}y) + m_2(\mathrm{d}y)$, see, e.g., [7] for details. Moreover, for any $n \geq 0$, let $\mathcal{DSM}_2^n(\mathbb{R}^k) \subset \mathcal{S}'(\mathbb{R}^k)$ denote the linear span of generalized derivatives of signed measures in $\mathcal{SM}_2(\mathbb{R}^k)$ up to order n, namely, the span of terms taking the form $\partial_{y_1}^{j_1}\cdots\partial_{y_k}^{j_k} m$, where $m \in \mathcal{SM}_2(\mathbb{R}^k)$ and $\sum_{i=1}^k j_i \leq n$. Alternatively, let $\mathcal{C}_2^n(\mathbb{R}^k)$ denote the set of functions $\psi : \mathbb{R}^k \to \mathbb{R}$ such that ψ has continuous derivatives up to order n and

$$\|\psi\|_n := \sup_{y\in\mathbb{R}^k} \sum_{j_1+\cdots+j_k\leq n} \frac{|\partial_{y_1}^{j_1}\cdots\partial_{y_k}^{j_k}\psi(y)|}{1+|y|^2} < \infty. \tag{2.21}$$

Then, clearly $\mathcal{DSM}_2^n(\mathbb{R}^k)$ is in the dual space of $\mathcal{C}_2^n(\mathbb{R}^k)$ in the sense that

$$\Big\langle \partial_{x_1}^{j_1}\cdots\partial_{x_k}^{j_k} m,\ \psi \Big\rangle = (-1)^{\sum_{i=1}^k j_i} \int_{\mathbb{R}^k} \partial_{y_1}^{j_1}\cdots\partial_{y_k}^{j_k}\psi(y) m(\mathrm{d}y). \tag{2.22}$$

Now, if $\frac{\delta\Phi}{\delta\rho}(\rho,x) \in \mathcal{DSM}_2^n(\mathbb{R}^k)$, then we extend (2.20) to all $\psi \in \mathcal{C}_2^n(\mathbb{R}^k)$, and we write

$$\int_{\mathbb{R}^k} \psi(y)\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y) := \big\langle \frac{\delta\Phi}{\delta\rho}(\rho,x),\ \psi\big\rangle, \quad \forall\psi \in \mathcal{C}_2^n(\mathbb{R}^k),$$

where the right side is in the sense of (2.22).

We now show two examples.

Example 2.6 *Let $\Phi(\rho) = \rho$ for any $\rho \in \mathcal{P}_2(\mathbb{R}^k)$. Then $\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y) = \delta_x(\mathrm{d}y)$, namely, $\frac{\delta\Phi}{\delta\rho}(\rho,x) \in \mathcal{SM}_2(\mathbb{R}^k) = \mathcal{DSM}_2^0(\mathbb{R}^k)$ for all $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ and $x \in \mathbb{R}^k$.*

Proof For any $\psi \in \mathcal{S}(\mathbb{R}^k)$, by (2.20) we have $\Psi(\rho) = \int_{\mathbb{R}^d}\psi(x)\rho(\mathrm{d}x)$. Then $\frac{\delta\Psi}{\delta\rho}(\rho,x) = \psi(x) = \int_{\mathbb{R}^k}\psi(y)\delta_x(\mathrm{d}y)$, and thus $\frac{\delta\Phi}{\delta\rho}(\rho,x) = \delta_x \in \mathcal{SM}_2(\mathbb{R}^k)$. □

Example 2.7 *Set $\Phi(\mathcal{L}_{(\xi,\eta)}) := \mathcal{L}_{(\xi,\eta+c\mathbb{E}[\eta])}$, $\forall\xi,\eta \in \mathbb{L}^2(\mathcal{F};\mathbb{R}^d)$, for some constant $c \in \mathbb{R}$. Then $\frac{\delta\Phi}{\delta\rho}(\rho,x,p) \in \mathcal{DSM}_2^1(\mathbb{R}^{2d})$. More precisely, letting $\mathbb{E}_\rho$ denote expectation under law $\rho = \mathcal{L}_{(\xi,\eta)}$,*

$$\frac{\delta\Phi}{\delta\rho}(\rho,x,p;\mathrm{d}\tilde{x},\mathrm{d}\tilde{p}) = \delta_x(\mathrm{d}\tilde{x})\delta_{p+c\mathbb{E}_\rho[\eta]}(\mathrm{d}\tilde{p}) - c\,\partial_{\tilde{p}}\Phi(\rho)(\mathrm{d}\tilde{x},\mathrm{d}\tilde{p})\cdot p. \tag{2.23}$$

Proof For any $\psi \in \mathcal{S}(\mathbb{R}^{2d})$, we have $\Psi(\rho) = \mathbb{E}_\rho\big[\psi(\xi,\eta+c\mathbb{E}_\rho[\eta])\big]$. Then

$$\begin{aligned}\frac{\delta\Psi}{\delta\rho}(\rho,x,p) &= \psi\big(x,p+c\mathbb{E}_\rho[\eta]\big) + c\mathbb{E}_\rho\big[\partial_p\psi(\xi,\eta+c\mathbb{E}_\rho[\eta])\big]\cdot p\\ &= \psi\big(x,p+c\mathbb{E}_\rho[\eta]\big) + cp\cdot\int_{\mathbb{R}^k}\partial_{\tilde{p}}\psi(\tilde{x},\tilde{p})\Phi(\rho)(\mathrm{d}\tilde{x},\mathrm{d}\tilde{p}).\end{aligned}$$

Compare this with (2.20), we obtain (2.23) immediately. □

Our main result of this part is the following chain rule. We use the notation $\nu = \Phi(\rho)$.

Proposition 2.8 *Let $\Phi : \mathcal{P}_2(\mathbb{R}^k) \to \mathcal{P}_2(\mathbb{R}^k)$, $U : \mathcal{P}_2(\mathbb{R}^k) \to \mathbb{R}$. Assume*

(i) Φ has a linear functional derivative $\frac{\delta\Phi}{\delta\rho}(\rho,x) \in \mathcal{DSM}_2^n(\mathbb{R}^k)$ for all $(\rho,x) \in \mathcal{P}_2(\mathbb{R}^k)\times\mathbb{R}^k$; $\frac{\delta\Phi}{\delta\rho}(\rho,x)$ is continuous in (ρ,x) under the weak topology, that is, for any $\psi \in \mathcal{C}_2^n(\mathbb{R}^k)$, the mapping $(\rho,x) \to \big\langle\frac{\delta\Phi}{\delta\rho}(\rho,x),\psi\big\rangle$ is continuous (under $\mathcal{W}_2$ for ρ); and, for any compact set $K \subset \mathcal{P}_2(\mathbb{R}^k)$, there exists a constant $C_K > 0$ such that

$$\sup_{\rho\in K}\Big|\langle\frac{\delta\Phi}{\delta\rho}(\rho,x),\ \psi\rangle\Big|\le C_K\|\psi\|_n[1+|x|^2],\quad \forall\psi\in\mathcal{C}_2^n(\mathbb{R}^k). \tag{2.24}$$

(ii) *U has a linear functional derivative $\frac{\delta U}{\delta\nu}$; for each $\nu\in\mathcal{P}_2(\mathbb{R}^k)$, $\frac{\delta U}{\delta\nu}(\nu,\cdot)\in\mathcal{C}_2^n(\mathbb{R}^k)$; and, by equipping $\mathcal{C}_2^n(\mathbb{R}^k)$ with the norm $\|\cdot\|_n$ in (2.21), the mapping $\nu\to\frac{\delta U}{\delta\nu}(\nu,\cdot)$ is continuous.*

Then the composite function $\widehat{U}:=U\circ\Phi:\mathcal{P}_2(\mathbb{R}^k)\to\mathbb{R}$ has a linear functional derivative:

$$\frac{\delta\widehat{U}}{\delta\rho}(\rho,x)=\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}(\Phi(\rho),y)\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y). \tag{2.25}$$

Proof Fix $\rho,\rho'\in\mathcal{P}_2(\mathbb{R}^k)$. For $0<\varepsilon<1$, denote $\rho_\varepsilon:=\rho+\varepsilon(\rho'-\rho)$. By the definition of $\frac{\delta U}{\delta\nu}$ we have

$$\widehat{U}(\rho_\varepsilon)-\widehat{U}(\rho)=U(\Phi(\rho_\varepsilon))-U(\Phi(\rho))=\int_0^1\Big[\Psi_\theta(\rho_\varepsilon)-\Psi_\theta(\rho)\Big]\mathrm{d}\theta,$$

where $\psi_\theta(x):=\frac{\delta U}{\delta\nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),x\Big),\quad \Psi_\theta(\tilde\rho):=\int_{\mathbb{R}^k}\psi_\theta(x)\Phi(\tilde\rho;\mathrm{d}x),\ \forall\tilde\rho\in\mathcal{P}_2(\mathbb{R}^k).$

Then, by (2.20), we have

$$\frac{\delta\Psi_\theta}{\delta\rho}(\tilde\rho,x)=\int_{\mathbb{R}^k}\psi_\theta(y)\frac{\delta\Phi}{\delta\rho}(\tilde\rho,x;\mathrm{d}y)=\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\tilde\rho,x;\mathrm{d}y).$$

Note that $\rho+\tilde\theta(\rho_\varepsilon-\rho)=\rho_{\tilde\theta\varepsilon}$, then

$$\begin{aligned}
&\frac{1}{\varepsilon}\big[\widehat{U}(\rho_\varepsilon)-\widehat{U}(\rho)\big]=\frac{1}{\varepsilon}\int_0^1\int_0^1\int_{\mathbb{R}^k}\frac{\delta\Psi_\theta}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x)(\rho_\varepsilon-\rho)(\mathrm{d}x)\mathrm{d}\tilde\theta\mathrm{d}\theta\\
&=\int_0^1\int_0^1\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)(\rho'-\rho)(\mathrm{d}x)\mathrm{d}\tilde\theta\mathrm{d}\theta\\
&=I_1(\varepsilon)+I_2(\varepsilon),
\end{aligned}$$

where

$$\begin{aligned}
I_1(\varepsilon)&:=\int_0^1\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)(\rho'-\rho)(\mathrm{d}x)\mathrm{d}\tilde\theta;\\
I_2(\varepsilon)&:=\int_0^1\int_0^1\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\Big[\frac{\delta U}{\delta\nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),y\Big)-\frac{\delta U}{\delta\nu}\big(\Phi(\rho),y\big)\Big]\\
&\qquad\times\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)(\rho'-\rho)(\mathrm{d}x)\mathrm{d}\tilde\theta\mathrm{d}\theta.
\end{aligned}$$

Clearly, $\lim_{\varepsilon\to0}W_2(\rho_{\tilde\theta\varepsilon},\rho)=0$. By the continuity of $\frac{\delta\Phi}{\delta\rho}$, we have

$$\lim_{\varepsilon\to0}\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)=\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y),\quad\forall\tilde\theta,x.$$

Moreover, note that $K:=\{\rho_\varepsilon:0\le\varepsilon\le1\}\subset\mathcal{P}_2(\mathbb{R}^k)$ is compact. Then, by (2.24), we have

$$\Big|\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)\Big|\le C\|\frac{\delta U}{\delta\nu}\big(\Phi(\rho),\cdot\big)\|_n[1+|x|^2].$$

Now it follows from the dominated convergence theorem that

$$\lim_{\varepsilon\to0}I_1(\varepsilon)=\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\frac{\delta U}{\delta\nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y)(\rho'-\rho)(\mathrm{d}x). \tag{2.26}$$

Moreover, by (2.24), we again have

$$\Big|\int_{\mathbb{R}^k}\Big[\frac{\delta U}{\delta \nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),y\Big)-\frac{\delta U}{\delta \nu}\big(\Phi(\rho),y\big)\Big]\frac{\delta\Phi}{\delta\rho}(\rho_{\tilde\theta\varepsilon},x;\mathrm{d}y)\Big|$$
$$\le C\|\frac{\delta U}{\delta \nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),\cdot\Big)-\frac{\delta U}{\delta \nu}\big(\Phi(\rho),\cdot\big)\|_n[1+|x|^2].$$

Then

$$|I_2(\varepsilon)|\le C\int_0^1\int_{\mathbb{R}^k}\|\frac{\delta U}{\delta \nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),\cdot\Big)-\frac{\delta U}{\delta \nu}\big(\Phi(\rho),\cdot\big)\|_n[1+|x|^2](\rho'+\rho)(\mathrm{d}x)\mathrm{d}\theta$$
$$\le C\int_0^1\|\frac{\delta U}{\delta \nu}\Big(\theta\Phi(\rho_\varepsilon)+(1-\theta)\Phi(\rho),\cdot\Big)-\frac{\delta U}{\delta \nu}\big(\Phi(\rho),\cdot\big)\|_n\mathrm{d}\theta\to 0,\ \text{as } \varepsilon\to 0,$$

thanks to the continuity of $\frac{\delta U}{\delta \nu}$ in ν under $\|\cdot\|_n$. This, together with (2.26), leads to

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\big[\widehat U(\rho_\varepsilon)-\widehat U(\rho)\big]=\int_{\mathbb{R}^k}\int_{\mathbb{R}^k}\frac{\delta U}{\delta \nu}\Big(\Phi(\rho),y\Big)\frac{\delta\Phi}{\delta\rho}(\rho,x;\mathrm{d}y)(\rho'-\rho)(\mathrm{d}x),$$

which implies (2.25) immediately. □

Remark 2.9 *By considering generalized derivatives in appropriate dual space, we can define higher-order derivatives of* Φ*, including the Lions derivative* $\partial_\rho\Phi(\rho,x):=\partial_x\frac{\delta\Phi}{\delta\rho}(\rho,x)$*. Alternatively, because we later always consider the composite function* $\widehat U$*, we can define higher-order derivatives through the left side of* (2.25).

3. The master equation and the monotonicities

Throughout the paper, Assumption 2.2 is always in force. Denote

$$\widehat H(x,p,\rho):=H(x,p,\Phi(\rho)),\quad (x,p,\rho)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^{2d}). \tag{3.1}$$

The derivatives of $\widehat H$ with respect to ρ are understood as in Proposition 2.8 and Remark 2.9. Then (2.14) becomes

$$\begin{gathered}X_t^*=\xi+\int_{t_0}^t\partial_p\widehat H\Big(X_s^*,\partial_xV(s,X_s^*,\mu_s^*),\rho_s^*\Big)\mathrm{d}s+B_t^{t_0}+\beta B_t^{0,t_0},\\ \text{where}\quad \mu_s^*:=\mathcal{L}_{X_s^*|\mathcal{F}_s^0},\quad \rho_s^*:=\mathcal{L}_{(X_s^*,\partial_xV(s,X_s^*,\mu_s^*))|\mathcal{F}_s^0}.\end{gathered} \tag{3.2}$$

On the other hand, it follows from the standard stochastic control theory that, for given t_0,μ, the optimization (2.7) is associated with the following backward SDE: recalling (2.12),

$$\begin{aligned}Y_t^*&=G(X_T^*,\mu_T^*)-\int_t^T Z_s^*\mathrm{d}B_s-\int_t^T Z_s^{0,*}\mathrm{d}B_s^0\\&\quad+\int_t^T\Big[\widehat H(\cdot)-\partial_xV(s,X_s^*,\mu_s^*)\cdot\partial_p\widehat H(\cdot)\Big]\Big(X_s^*,\partial_xV(s,X_s^*,\mu_s^*),\rho_s^*\Big)\mathrm{d}s,\end{aligned} \tag{3.3}$$

which, together with (3.2), form the MFGC system. We note that this is the SDE counterpart of the MFGC system (1.4). In particular, we have

$$Y_t^*=V\big(t,X_t^*,\mu_t^*\big). \tag{3.4}$$

Then, by applying Itô's formula (see [17, Theorem 4.17], [8, 19]), we obtain

$$\begin{aligned}
\mathrm{d}V(t,X_t^*,\mu_t^*) &= \Big[\partial_t V + \partial_x V\cdot\partial_p \widehat{H}(X_t^*,\partial_x V,\rho_t^*) + \frac{1+\beta^2}{2}\mathrm{tr}\big(\partial_{xx}V\big)\Big](t,X_t^*,\mu_t^*)\mathrm{d}t\\
&\quad + \partial_x V(t,X_t^*,\mu_t^*)\cdot \mathrm{d}B_t + \beta\Big[\partial_x V(t,X_t^*,\mu_t^*) + \tilde{\mathbb{E}}_{\mathcal{F}_t}\big[\partial_\mu V(t,X_t^*,\mu_t^*,\tilde X_t^*)\big]\Big]\cdot \mathrm{d}B_t^0\\
&\quad + \mathrm{tr}\Big(\tilde{\mathbb{E}}_{\mathcal{F}_t}\big[\partial_\mu V(t,X_t^*,\mu_t^*,\tilde X_t^*)(\partial_p\widehat{H}(t,\tilde X_t^*,\partial_x V(t,\tilde X_t,\mu_t^*),\rho_t^*))^\top\big]\Big)\mathrm{d}t\\
&\quad + \mathrm{tr}\Big(\beta^2\tilde{\mathbb{E}}_{\mathcal{F}_t}\big[\partial_x\partial_\mu V(t,X_t^*,\mu_t^*,\tilde X_t^*) + \frac{1+\beta^2}{2}\partial_{\tilde x}\partial_\mu V(t,X_t^*,\mu_t^*,\tilde X_t^*)\big]\\
&\quad + \frac{\beta^2}{2}\bar{\tilde{\mathbb{E}}}_{\mathcal{F}_t}\big[\partial_{\mu\mu}V(t,X_t^*,\mu_t^*,\tilde X_t^*,\bar X_t^*)\big]\Big)\mathrm{d}t. \qquad (3.5)
\end{aligned}$$

Here, as usual, $\tilde X^*,\bar X^*$ are conditionally independent copies of X^*, conditional on $\mathbb{F}^0$. Comparing this with (3.3), we derive the master equation: for independent copies $\xi,\tilde\xi,\bar\xi$ with law μ,

$$\begin{aligned}
&\mathcal{L}V(t,x,\mu) := \partial_t V + \frac{\widehat{\beta}^2}{2}\mathrm{tr}(\partial_{xx}V) + \widehat{H}(x,\partial_x V,\mathcal{L}_{(\xi,\partial_x V(t,\xi,\mu))}) + \mathcal{M}V = 0,\\
&V(T,x,\mu) = G(x,\mu), \quad \text{where}\\
&\mathcal{M}V(t,x,\mu) := \mathrm{tr}\Big(\bar{\tilde{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2}\partial_{\tilde x}\partial_\mu V(t,x,\mu,\tilde\xi) + \beta^2\partial_x\partial_\mu V(t,x,\mu,\tilde\xi) + \frac{\beta^2}{2}\partial_{\mu\mu}V(t,x,\mu,\bar\xi,\tilde\xi)\\
&\qquad + \partial_\mu V(t,x,\mu,\tilde\xi)(\partial_p\widehat{H})^\top(\tilde\xi,\partial_x V(t,\tilde\xi,\mu),\mathcal{L}_{(\xi,\partial_x V(t,\xi,\mu))})\Big]\Big), \quad \text{and} \quad \widehat{\beta}^2 := 1+\beta^2. \qquad (3.6)
\end{aligned}$$

In addition to Assumption 2.2, we assume the following.

Assumption 3.1 *$\widehat{H}\in\mathcal{C}^2(\mathbb{R}^{2d}\times\mathcal{P}_2(\mathbb{R}^{2d}))$ with bounded $\partial_{xp}\widehat{H},\partial_{xx}\widehat{H},\partial_{pp}\widehat{H},\partial_{x\mu}\widehat{H},\partial_{p\mu}\widehat{H}$.*

Since we work on the master equation, here we impose our conditions directly on $\widehat{H}$, rather than on b,f. It is straightforward to find some sufficient conditions on b and f to ensure these.

3.1 The monotonicities

In this subsection, we introduce three types of monotonicity conditions: Lasry–Lions monotonicity, displacement semi-monotonicity, and anti-monotonicity.

Definition 3.2 *Assume $U\in\mathcal{C}^1(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d))$ and $\partial_\mu U(\cdot,\mu,\tilde x)\in\mathcal{C}^1(\mathbb{R}^d)$ for all $(\mu,\tilde x)\in\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d$. We say U is Lasry–Lions monotone if*

$$MON^{LL}U(\xi,\eta) := \tilde{\mathbb{E}}\Big[\big\langle \partial_{x\mu}U(\xi,\mathcal{L}_\xi,\tilde\xi)\tilde\eta,\eta\big\rangle\Big] \ge 0, \quad \forall \xi,\eta\in\mathbb{L}^2(\mathcal{F}_T^1). \qquad (3.7)$$

We note that, since (ξ,η) is $\mathcal{F}_T^1$-measurable, here $(\tilde\xi,\tilde\eta)$ is an independent copy (instead of a conditionally independent copy).

Definition 3.3 *Assume $U,\partial_x U\in\mathcal{C}^1(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d))$. For any $\lambda\ge 0$, we say U is displacement λ-monotone if, for all $\xi,\eta\in\mathbb{L}^2(\mathcal{F}_T^1)$,*

$$MON_\lambda^{disp}U(\xi,\eta) := \tilde{\mathbb{E}}\Big[\big\langle\partial_{x\mu}U(\xi,\mathcal{L}_\xi,\tilde\xi)\tilde\eta,\eta\big\rangle + \big\langle\partial_{xx}U(\xi,\mathcal{L}_\xi)\eta,\eta\big\rangle + \lambda|\eta|^2\Big] \ge 0. \qquad (3.8)$$

In particular, we say U is displacement monotone when $\lambda=0$, and displacement semi-monotone if it is displacement λ-monotone for some $\lambda>0$.

Moreover, denote

$$D_4 := \Big\{\vec\lambda = (\lambda_0,\lambda_1,\lambda_2,\lambda_3) : \lambda_0>0, \lambda_1\in\mathbb{R}, \lambda_2>0, \lambda_3\ge 0\Big\}. \qquad (3.9)$$

Definition 3.4 *Let $U \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ and $\vec{\lambda} \in D_4$. We say U is $\vec{\lambda}$-anti-monotone if*

$$MON^{anti}_{\vec{\lambda}} U(\xi,\eta) := \tilde{\mathbb{E}}\Big[\lambda_0 \langle \partial_{xx} U(\xi,\mathcal{L}_\xi)\eta,\eta\rangle + \lambda_1 \langle \partial_{x\mu} U(\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta},\eta\rangle + |\partial_{xx} U(\xi,\mathcal{L}_\xi)\eta|^2 + \lambda_2 \big|\tilde{\mathbb{E}}_{\mathcal{F}}[\partial_{x\mu} U(\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta}]\big|^2 - \lambda_3 |\eta|^2\Big] \le 0, \quad \forall \xi,\eta \in \mathbb{L}^2(\mathcal{F}^1_T). \tag{3.10}$$

Remark 3.5 (i) *By [21, Remark 2.4], Lasry–Lions monotonicity and displacement monotonicity are, respectively, equivalent to the following forms which are commonly seen in the literature:*

$$\mathbb{E}\Big[U(\xi_1,\mathcal{L}_{\xi_1}) + U(\xi_2,\mathcal{L}_{\xi_2}) - U(\xi_1,\mathcal{L}_{\xi_2}) - U(\xi_2,\mathcal{L}_{\xi_1})\Big] \ge 0, \quad \forall \xi_1,\xi_2 \in \mathbb{L}^2(\mathcal{F}^1_T);$$
$$\mathbb{E}\Big[\big\langle \partial_x U(\xi_1,\mathcal{L}_{\xi_1}) - \partial_x U(\xi_2,\mathcal{L}_{\xi_2}),\ \xi_1 - \xi_2\big\rangle\Big] \ge 0, \quad \forall \xi_1,\xi_2 \in \mathbb{L}^2(\mathcal{F}^1_T).$$

(ii) *Consider the case that $\partial_x U(x,\mu) = \partial_\mu \mathcal{U}(\mu,x)$ for some $\mathcal{U} \in \mathcal{C}^2(\mathcal{P}_2(\mathbb{R}^d))$. Then the Lasry–Lions monotonicity of U is equivalent to the convexity of the mapping $\mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathcal{U}(\mu)$, and the displacement monotonicity of U is equivalent to the convexity of the mapping $\xi \in \mathbb{L}^2(\mathcal{F}^1_T) \mapsto \mathcal{U}(\mathcal{L}_\xi)$ (see, e.g., [15, 16]).*

(iii) *Both the Lasry–Lions monotonicity (provided $\partial_{xx} U$ is bounded) and the displacement monotonicity imply displacement semi-monotonicity. However, Lasry–Lions monotonicity and displacement monotonicity do not imply each other, see [21, Remark 2.5].*

(iv) *By setting $\lambda_0 = \lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 0$,* (3.10) *implies*

$$\tilde{\mathbb{E}}\Big[\langle \partial_{xx} U(\xi,\mathcal{L}_\xi)\eta,\eta\rangle + \langle \partial_{x\mu} U(\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta},\eta\rangle\Big] \le 0,$$

which is in the opposite direction of (3.8) *with $\lambda = 0$. Moreover, if $\partial_{xx} U$ is non-negative definite, we then further have*

$$\tilde{\mathbb{E}}\Big[\langle \partial_{x\mu} U(\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta},\eta\rangle\Big] \le 0,$$

which is in the opposite direction of (3.7). *That is why we call* (3.10) *anti-monotonicity.*

(v) *If U satisfies* (3.8) *for some $\lambda \ge 0$, then $\partial_{xx} U + \lambda I$ is non-negative definite, see [21, Lemma 2.6].*

3.2 A road map towards the global wellposedness

Our ultimate goal is to establish the global wellposedness of the master equation (3.6). We adopt the strategy in [21, 35], which consists of three steps:

Step 1 Introduce an appropriate monotonicity condition on the data which ensures the propagation of a monotonicity of one of the three types introduced in the previous subsection, along any classical solution to the master equation.

Step 2 Show that the monotonicity of $V(t,\cdot,\cdot)$ implies an (a priori) uniform Lipschitz continuity of V in the measure variable μ.

Step 3 Combine the local wellposedness of classical solutions and the above uniform Lipschitz continuity to obtain the global wellposedness of classical solutions.

Moreover, following [15], we continue to investigate the convergence problem:

Step 4 Use the classical solution V to prove the convergence of the related N-player game.

In this paper, we focus on Step 1 only, and we leave the remaining three steps to future research.

We emphasize that Step 1 (and Step 2) considers prior estimates, and thus also we assume the following.

Assumption 3.6 *$V \in \mathcal{C}^{1,2,2}([0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d))$ is a classical solution of the master equation* (3.6) *such that $\partial_{xx}V(t,\cdot,\cdot)\in\mathcal{C}^2(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d))$, $\partial_{x\mu}V(t,\cdot,\cdot,\cdot)\in\mathcal{C}^2(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d)$, and all of the second- and higher-order derivatives of V involved above are uniformly bounded and continuous in t.*

Note that we do not require V or its first-order derivatives to be bounded. Moreover, since $G = V(T,\cdot,\cdot)$, the above assumption also ensures the regularity of G. We also remark that Assumption 3.6 considers the existence of classical solutions of the master equation (3.6), which implies the uniqueness of the mean field equilibrium (see [21, Remark 2.10 (ii)]). The uniqueness of classical solutions satisfying the desired Lipschitz continuity is standard, see, e.g., the arguments in [21, Theorem 6.3].

4. Propagation of Lasry–Lions monotonicity

To propagate the Lasry–Lions monotonicity of V, we impose the following assumption on $\widehat{H}$.

Assumption 4.1 *For any $\xi,\eta,\gamma,\zeta\in\mathbb{L}^2(\mathcal{F}_T^1)$ and $\varphi:\mathbb{R}^d\to\mathbb{R}^d$ Lipschitz continuous,*

$$\begin{aligned}\tilde{\mathbb{E}}\Big[&\Big\langle\zeta,\widehat{H}_{pp}(\xi)\zeta\Big\rangle-\Big\langle\eta,\ \widehat{H}_{x\rho_1}(\xi,\tilde{\xi})\tilde{\eta}+\widehat{H}_{x\rho_2}(\xi,\tilde{\xi})[\tilde{\gamma}+\tilde{\zeta}]\Big\rangle\\&-\Big\langle\gamma-\zeta,\ \widehat{H}_{p\rho_1}(\xi,\tilde{\xi})\tilde{\eta}+\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)[\tilde{\gamma}+\tilde{\zeta}]\Big\rangle\Big]\le 0,\end{aligned}\tag{4.1}$$

where $\widehat{H}_{pp}(x):=\partial_{pp}\widehat{H}\Big(x,\varphi(x),\mathcal{L}_{(\xi,\varphi(\xi))}\Big)$, $\widehat{H}_{x\rho}(x,\tilde{x}):=\partial_{x\rho}\widehat{H}\Big(x,\varphi(x),\mathcal{L}_{(\xi,\varphi(\xi))},\tilde{x},\varphi(\tilde{x})\Big)$, and similarly for $\widehat{H}_{p\rho}(x,\tilde{x})$.

The main result of this section follows.

Theorem 4.2 *Let Assumptions 2.2, 3.1, 3.6, and 4.1 hold. If G satisfies the Lasry–Lions monotonicity* (3.7)*, then $V(t,\cdot,\cdot)$ satisfies* (3.7) *for all $t\in[0,T]$.*

Proof Without loss of generality, we prove the theorem only for $t=0$.

For any $\xi,\eta\in\mathbb{L}^2(\mathcal{F}_0)$, inspired by (3.2), we consider the following system of McKean–Vlasov SDEs, which clearly has a unique solution $(X,\delta X)$ under Assumptions 3.1 and 3.6:

$$\begin{aligned}&X_t=\xi+\int_0^t\widehat{H}_p(X_s,\partial_xV(s,X_s,\mu_s),\rho_s)\mathrm{d}s+B_t+\beta B_t^0;\\&\delta X_t=\eta+\int_0^t\Big[\widehat{H}_{px}(X_s)\delta X_s+\widehat{H}_{pp}(X_s)[\Gamma_s+\Upsilon_s]+N_s\Big]\mathrm{d}s;\quad\text{where}\\&\mu_t:=\mathcal{L}_{X_t|\mathcal{F}_t^0},\quad\rho_t:=\mathcal{L}_{(X_t,\partial_xV(t,X_t,\mu_t))|\mathcal{F}_t^0};\\&\Gamma_t:=\partial_{xx}V(X_t)\delta X_t,\quad\Upsilon_t:=\tilde{\mathbb{E}}_{\mathcal{F}_t}[\partial_{x\mu}V(X_t,\tilde{X}_t)\delta\tilde{X}_t];\\&N_t:=\tilde{\mathbb{E}}_{\mathcal{F}_t}\Big[\widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t+\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\big[\tilde{\Gamma}_t+\tilde{\Upsilon}_t\big]\Big].\end{aligned}\tag{4.2}$$

Here, $(\tilde{X},\delta\tilde{X},\tilde{\Gamma},\tilde{\Upsilon})$ is a conditionally independent copy of $(X,\delta X,\Gamma,\Upsilon)$, conditional on $\mathbb{F}^0$. Moreover, here and in what follows, for simplicity of notation, we omit the variables (t,μ_t)

inside V and its derivatives, and omit ρ_t and $\partial_x V$ inside $\widehat{H}$ and its derivatives, for example,

$$\begin{aligned}&\partial_{x\mu}V(X_t,\tilde{X}_t)=\partial_{x\mu}V(t,X_t,\mu_t,\tilde{X}_t),\quad \widehat{H}_p(X_t):=\partial_p\widehat{H}(X_t,\partial_x V(t,X_t,\mu_t),\rho_t),\\ &\widehat{H}_{p\rho}(X_t,\tilde{X}_t):=(\widehat{H}_{p\rho_1},\widehat{H}_{p\rho_2})(X_t,\tilde{X}_t):=\partial_{p\rho}\widehat{H}(X_t,\partial_x V(t,X_t,\mu_t),\rho_t,\tilde{X}_t,\partial_x V(t,\tilde{X}_t,\mu_t)).\end{aligned}\tag{4.3}$$

Introduce,

$$I(t):=\mathbb{E}[\langle\Upsilon_t,\delta X_t\rangle]=MON^{LL}V(t,\cdot,\cdot)(X_t,\delta X_t).\tag{4.4}$$

By applying Itô's formula (3.5) and since V satisfies the master equation (3.6), we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}I(t)=\tilde{\mathbb{E}}\Big[&\Big\langle\Upsilon_t,\ \widehat{H}_{pp}(X_t)\Upsilon_t\Big\rangle-\Big\langle\delta X_t,\ \widehat{H}_{x\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t+\widehat{H}_{x\rho_2}(X_t,\tilde{X}_t)[\tilde{\Gamma}_t+\tilde{\Upsilon}_t]\Big\rangle\\ &-\Big\langle\Gamma_t-\Upsilon_t,\ \widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t+\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\big[\tilde{\Gamma}_t+\tilde{\Upsilon}_t]\Big\rangle\Big].\end{aligned}\tag{4.5}$$

The calculation is lengthy but quite straightforward, we postpone the details to the appendix. Taking the conditional expectation on $\mathcal{F}^0_t$ and then by the desired conditional independence, we apply (4.1) to obtain:

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t)\le 0.\tag{4.6}$$

Note that, by the Lasry–Lions monotonicity of $G=V(T,\cdot,\cdot)$, we have $I(T)\ge 0$. Then (4.6) clearly implies $I(0)\ge 0$, and hence $V(0,\cdot,\cdot)$ satisfies the Lasry–Lions monotonicity (3.7). □

Remark 4.3 *In* (4.2), X *is the agent's state process along the (unique) mean field equilibrium, and* δX *is the gradient of* X *when its initial condition* ξ *is perturbed along the direction* η.

Remark 4.4 *Note that* (4.5) *is an equality, so our condition* (4.1) *is essentially sharp for the propagation of Lasry–Lions monotonicity, in particular, for* (4.6). *In [13, 16, 28] the uniqueness of the mean field game system is obtained when* $b(\cdot,a,\cdot)=a$ *(or a slightly more general form), and* f *satisfies the Lasry–Lions monotonicity in the following sense: for any* $\xi_i,\alpha_i\in\mathcal{L}^2(\mathcal{F})$, $i=1,2$,

$$\mathbb{E}\big[f(\xi^1,\alpha^1,\mathcal{L}_{(\xi^1,\alpha^1)})+f(\xi^2,\alpha^2,\mathcal{L}_{(\xi^2,\alpha^2)})-f(\xi^1,\alpha^1,\mathcal{L}_{(\xi^2,\alpha^2)})-f(\xi^2,\alpha^2,\mathcal{L}_{(\xi^1,\alpha^1)})\big]\ge 0.\tag{4.7}$$

We claim that in this case (4.6) *holds true, and hence the Lasry–Lions monotonicity propagates. We postpone its proof to the appendix.*

Remark 4.5 *For the standard MFG with* $b(x,a,\nu)=a$ *(and* $f=f(x,a,\mu)$*), it is observed in [21] that it is hard to propagate Lasry–Lions monotonicity unless* f *is separable:* $f(x,a,\mu)=f_0(x,a)+f_1(x,\mu)$. *Dependence on the law of* α *in MFGC actually helps for the propagation of the Lasry–Lions monotonicity. In particular, in this case we do not require* f *to be separable.*

We now provide an example with a more general b, which does not seem to be covered by the analysis of mean field game systems (or master equations) in the literature.

Example 4.6 *We consider a special case of* (2.17) *with* $d=1$:

$$\begin{aligned}b(x,a,\mathcal{L}_{(\xi,\alpha)})&=-a+b_1(\mathbb{E}[\xi],\mathbb{E}[\alpha])+b_2(x),\\ f(x,a,\mathcal{L}_{(\xi,\alpha)})&=\tfrac{|a|^2}{2}-c_1a\mathbb{E}[\alpha]+c_2x\mathbb{E}[\xi]+c_3x\mathbb{E}[\alpha]+f_1(x),\end{aligned}\tag{4.8}$$

where $0<c_1<1$ *and* $c_2,c_3>0$ *are constants. Assume the matrix*

$$\begin{bmatrix} 1-[\bar c_1\partial_{m_2}b_1-\hat c_1] & 0 & \frac{1}{2}[\hat c_3-\partial_{m_1}b_1] \\ 0 & [\bar c_1\partial_{m_2}b_1-\hat c_1] & \frac{1}{2}[\hat c_3+\partial_{m_1}b_1] \\ \frac{1}{2}[\hat c_3-\partial_{m_1}b_1] & \frac{1}{2}[\hat c_3+\partial_{m_1}b_1] & c_2 \end{bmatrix} \geq 0, \tag{4.9}$$

where $\hat c_1 := \frac{c_1}{1-c_1}$, $\bar c_1 := \frac{1}{1-c_1}$, $\hat c_3 := \frac{c_3}{1-c_1}$, *and* m_1, m_2 *stand for* $\mathbb{E}[\xi], \mathbb{E}[\alpha]$. *Then* (4.1) *holds true.*

Proof By Example 2.4, we see that

$$\Phi(\mathcal{L}_{(\xi,\eta)}) = \mathcal{L}_{(\xi,\ \hat c_1\mathbb{E}[\eta]+\eta)},$$
$$H(x,p,\mathcal{L}_{(\xi,\alpha)}) = -\frac{1}{2}\big|c_1\mathbb{E}[\alpha]+p\big|^2 + p\big[b_1(\mathbb{E}[\xi],\mathbb{E}[\alpha]) + b_2(x)\big] + c_2x\mathbb{E}[\xi] + c_3x\mathbb{E}[\alpha] + f_1(x).$$

Note that $\mathbb{E}[\alpha] = [1+\hat c_1]\mathbb{E}[\eta] = \bar c_1\mathbb{E}[\eta]$. Then

$$\widehat H(x,p,\mathcal{L}_{(\xi,\eta)}) = -\frac{1}{2}\big|\hat c_1\mathbb{E}[\eta]+p\big|^2 + p\big[b_1(\mathbb{E}[\xi],\bar c_1\mathbb{E}[\eta]) + b_2(x)\big] + c_2x\mathbb{E}[\xi] + \hat c_3x\mathbb{E}[\eta] + f_1(x).$$

One may compute that

$$\widehat H_{pp} = -1, \quad \widehat H_{x\rho_1} = c_2, \quad \widehat H_{x\rho_2} = \hat c_3, \quad \widehat H_{p\rho_1} = \partial_{m_1}b_1, \quad \widehat H_{p\rho_2} = \bar c_1\partial_{m_2}b_1 - \hat c_1. \tag{4.10}$$

Then, noting that $\partial_{m_1}b_1$ and $\partial_{m_1}b_2$ are deterministic,

$$\begin{aligned}
&\tilde{\mathbb{E}}\Big[-\widehat H_{pp}(\xi)|\zeta|^2 + \eta\big[\widehat H_{x\rho_1}(\xi,\tilde\xi)\tilde\eta + \widehat H_{x\rho_2}(\xi,\tilde\xi)[\tilde\gamma+\tilde\zeta]\big] \\
&\quad + [\gamma-\zeta]\big[\widehat H_{p\rho_1}(\xi,\tilde\xi)\tilde\eta + \widehat H_{p\rho_2}(X_t,\tilde X_t)[\tilde\gamma+\tilde\zeta]\big]\Big] \\
&= \tilde{\mathbb{E}}\Big[|\zeta|^2 + c_2\eta\tilde\eta + \hat c_3\eta[\tilde\gamma+\tilde\zeta]\big] + [\gamma-\zeta]\big[\partial_{m_1}b_1\tilde\eta + [\bar c_1\partial_{m_2}b_1-\hat c_1][\tilde\gamma+\tilde\zeta]\big]\Big] \\
&= \mathbb{E}[|\zeta|^2] + c_2\big|\mathbb{E}[\eta]\big|^2 + \hat c_3\mathbb{E}[\eta]\big[\mathbb{E}[\gamma]+\mathbb{E}[\zeta]\big] + \partial_{m_1}b_1\mathbb{E}[\eta]\big[\mathbb{E}[\gamma]-\mathbb{E}[\zeta]\big] \\
&\quad + [\bar c_1\partial_{m_2}b_1-\hat c_1]\big[|\mathbb{E}[\gamma]|^2 - |\mathbb{E}[\zeta]|^2\big] \\
&\geq \big[1-[\bar c_1\partial_{m_2}b_1-\hat c_1]\big]\big|\mathbb{E}[\zeta]\big|^2 + [\bar c_1\partial_{m_2}b_1-\hat c_1]\big|\mathbb{E}[\gamma]\big|^2 + c_2\big|\mathbb{E}[\eta]\big|^2 \\
&\quad + \big[\hat c_3+\partial_{m_1}b_1\big]\mathbb{E}[\eta]\mathbb{E}[\gamma] + \big[\hat c_3-\partial_{m_1}b_1\big]\mathbb{E}[\eta]\mathbb{E}[\zeta].
\end{aligned}$$

This, together with (4.9), clearly implies (4.1). □

5. Propagation of displacement λ-monotonicity

In this section we fix a constant $\lambda \geq 0$.

Assumption 5.1 *For any* $\xi,\eta,\gamma,\zeta \in \mathbb{L}^2(\mathcal{F}_T^1)$ *and* $\varphi:\mathbb{R}^d\to\mathbb{R}^d$ *Lipschitz continuous,*

$$\begin{aligned}
&\tilde{\mathbb{E}}\Big[\Big\langle \gamma+\zeta, \widehat H_{pp}(\xi)[\gamma+\zeta]\Big\rangle - \Big\langle \eta, [\widehat H_{xx}(\xi) - 2\lambda\widehat H_{px}(\xi)]\eta\Big\rangle \\
&+ \Big\langle \gamma+\zeta,\ [\widehat H_{p\rho_1}(\xi,\tilde\xi) + \widehat H_{\rho_2 x}(\tilde\xi,\xi) + 2\lambda\widehat H_{\rho_2 p}(\tilde\xi,\xi)]\tilde\eta + 2\lambda\widehat H_{pp}(\xi)\eta\Big\rangle \\
&+ \Big\langle \gamma+\zeta, \widehat H_{p\rho_2}(\xi,\tilde\xi)[\tilde\gamma+\tilde\zeta]\Big]\Big\rangle - \Big\langle \eta,\ [\widehat H_{x\rho_1}(\xi,\tilde\xi) - 2\lambda\widehat H_{p\rho_1}(\xi,\tilde\xi)]\tilde\eta\Big\rangle\Big] \leq 0,
\end{aligned} \tag{5.1}$$

where $\widehat{H}_{pp}, \widehat{H}_{x\rho}, \widehat{H}_{p\rho}$ *are as in Assumption 4.1.*

Theorem 5.2 *Let Assumptions 2.2, 3.1, 3.6, and 5.1 hold. If* G *satisfies the displacement* λ-*monotonicity* (3.8)*, then* $V(t,\cdot,\cdot)$ *satisfies* (3.8) *for all* $t\in[0,T]$.

Proof Without loss of generality, we prove the theorem only for $t=0$. We continue to use the notation in the proof of Theorem 4.2.

Introduce

$$\bar I(t) := \mathbb{E}[\langle \Gamma_t, \delta X_t\rangle], \quad \text{and thus} \quad I(t)+\bar I(t)+\lambda\mathbb{E}[|\delta X_t|^2] = MON_\lambda^{disp}V(t,\cdot,\cdot)(X_t,\delta X_t). \tag{5.2}$$

Similarly to (4.5), we show that (see the appendix for more details),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\bar I(t) = \tilde{\mathbb{E}}\Big[&\Big\langle \widehat{H}_{pp}(X_t)\Gamma_t,\Gamma_t\Big\rangle + 2\Big\langle \widehat{H}_{pp}(X_t)\Gamma_t,\Upsilon_t\Big]\Big\rangle\\ &+2\Big\langle \Gamma_t, \widehat{H}_{p\rho_1}(X_t,\tilde X_t)\delta\tilde X_t + \widehat{H}_{p\rho_2}(X_t,\tilde X_t)[\tilde\Gamma_t+\tilde\Upsilon_t]\Big\rangle - \Big\langle \widehat{H}_{xx}(X_t)\delta X_t,\delta X_t\Big\rangle\Big].\end{aligned} \tag{5.3}$$

Moreover, by (4.2), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[|\delta X_t|^2\right] = 2\mathbb{E}\Big[\Big\langle \widehat{H}_{px}(X_t)\delta X_t + \widehat{H}_{pp}(X_t)[\Upsilon_t+\Gamma_t]+N_t,\ \delta X_t\Big\rangle\Big]. \tag{5.4}$$

Combining (4.5), (5.3), and (5.4) and recalling the N in (4.2), we deduce that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\Big[MON_\lambda^{disp}V(t,\cdot,\cdot)(X_t,\delta X_t)\Big] = \frac{\mathrm{d}}{\mathrm{d}t}\Big[I(t)+\bar I(t)+\lambda\mathbb{E}[|\delta X_t|^2]\Big]\\ &=\tilde{\mathbb{E}}\Big[\Big\langle \Upsilon_t+\Gamma_t, \widehat{H}_{pp}(X_t)[\Upsilon_t+\Gamma_t]+\widehat{H}_{p\rho_2}(X_t,\tilde X_t)[\tilde\Gamma_t+\tilde\Upsilon_t]\Big\rangle\\ &\quad+\Big\langle \Upsilon_t+\Gamma_t, [\widehat{H}_{p\rho_1}(X_t,\tilde X_t)+\widehat{H}_{\rho_2 x}(\tilde X_t,X_t)+2\lambda\widehat{H}_{\rho_2 p}(\tilde X_t,X_t)]\delta\tilde X_t + 2\lambda\widehat{H}_{pp}(X_t)\delta X_t\Big\rangle\\ &\quad-\Big\langle \delta X_t, [\widehat{H}_{x\rho_1}(X_t,\tilde X_t)-2\lambda\widehat{H}_{p\rho_1}(X_t,\tilde X_t)]\delta\tilde X_t + [\widehat{H}_{xx}(X_t)-2\lambda\widehat{H}_{px}(X_t)]\delta X_t\Big\rangle\Big].\end{aligned} \tag{5.5}$$

Then, by the desired conditional independence of the involved processes above, conditional on $\mathcal{F}_t^0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big[MON_\lambda^{disp}V(t,\cdot,\cdot)(X_t,\delta X_t)\Big]\le 0. \tag{5.6}$$

Note that $V(T,\cdot,\cdot)=G$ satisfies (3.8), then clearly $V(0,\cdot,\cdot)$ also satisfies (3.8). □

We next provide a sufficient condition for Assumption 5.1. Denote, for any $A\in\mathbb{R}^{d\times d}$,

$$\begin{aligned}|A| &:= \sup_{|x|=|y|=1}\langle Ax,y\rangle, \quad \underline{\kappa}(A) := \inf_{|x|=1}\langle Ax,x\rangle = \text{the smallest eigenvalue of } \tfrac12[A+A^\top],\\ &\qquad\qquad \overline{\kappa}(A) := \sup_{|x|=1}\langle Ax,x\rangle = -\underline{\kappa}(-A).\end{aligned} \tag{5.7}$$

Proposition 5.3 *Assume there exists a constant* $c_0\ge 0$ *such that* $|\partial_{p\rho_2}\widehat{H}|\le c_0$, *and* $\widehat{H}_{pp} < -c_0 I_d$, *where* I_d *denotes the* $d\times d$ *identity matrix. Then the following condition implies* (5.1)*:*

$$\begin{aligned}&\tilde{\mathbb{E}}\Big[\Big\langle \eta, [\widehat{H}_{xx}(\xi)-2\lambda\widehat{H}_{px}(\xi)]\eta\Big\rangle + \Big\langle \eta,\ [\widehat{H}_{x\rho_1}(\xi,\tilde\xi)-2\lambda\widehat{H}_{p\rho_1}(\xi,\tilde\xi)]\tilde\eta\Big\rangle - \tfrac{|\Lambda(\xi,\eta)|^2}{4}\Big]\ge 0,\\ &\textit{where}\quad \Lambda(\xi,\eta) := (-\widehat{H}_{pp}(\xi)-c_0 I_d)^{-\frac12}\\ &\qquad\qquad\times\Big[\tilde{\mathbb{E}}_{\mathcal{F}_T^1}\big[[\widehat{H}_{p\rho_1}(\xi,\tilde\xi)+\widehat{H}_{\rho_2 x}(\tilde\xi,\xi)+2\lambda\widehat{H}_{\rho_2 p}(\tilde\xi,\xi)]\tilde\eta\big]+2\lambda\widehat{H}_{pp}(\xi)\eta\Big],\end{aligned} \tag{5.8}$$

for all $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1)$. *In particular, when* $\lambda = 0$, *the above reduces to*

$$\begin{aligned}&\tilde{\mathbb{E}}\Big[\Big\langle \eta, \widehat{H}_{xx}(\xi)\eta\Big\rangle + \Big\langle \eta, \widehat{H}_{x\rho_1}(\xi,\tilde{\xi})\tilde{\eta}\Big\rangle - \tfrac{|\Lambda(\xi,\eta)|^2}{4}\Big] \geq 0,\\ \textit{where}\quad &\Lambda(\xi,\eta) := (-\widehat{H}_{pp}(\xi) - c_0 I_d)^{-\frac{1}{2}}\tilde{\mathbb{E}}_{\mathcal{F}_T^1}\Big[[\widehat{H}_{p\rho_1}(\xi,\tilde{\xi}) + \widehat{H}_{\rho_2 x}(\tilde{\xi},\xi)]\tilde{\eta}\Big].\end{aligned} \tag{5.9}$$

Proof Denote $\Xi := \big[\widehat{H}_{p\rho_1}(\xi,\tilde{\xi}) + \widehat{H}_{\rho_2 x}(\tilde{\xi},\xi) + 2\lambda \widehat{H}_{\rho_2 p}(\tilde{\xi},\xi)\big]\tilde{\eta}$. Note that

$$\begin{aligned}&\tilde{\mathbb{E}}\Big[\Big\langle \gamma+\zeta, \widehat{H}_{pp}(\xi)[\gamma+\zeta]\Big\rangle + \Big\langle \gamma+\zeta, \Xi + 2\lambda\widehat{H}_{pp}(\xi)\eta\Big\rangle + \Big\langle \gamma+\zeta, \widehat{H}_{p\rho_2}(\xi,\tilde{\xi})[\tilde{\gamma}+\tilde{\zeta}]\Big]\Big\rangle\Big]\\ &\leq \tilde{\mathbb{E}}\Big[\Big\langle \gamma+\zeta, \widehat{H}_{pp}(\xi)[\gamma+\zeta]\Big\rangle + \Big\langle \gamma+\zeta, \Xi + 2\lambda\widehat{H}_{pp}(\xi)\eta\Big\rangle + \frac{c_0}{2}[|\gamma+\zeta|^2 + |\tilde{\gamma}+\tilde{\zeta}|^2]\Big]\Big]\\ &= \mathbb{E}\Big[\Big\langle \gamma+\zeta, [\widehat{H}_{pp}(\xi) + c_0 I_d][\gamma+\zeta]\Big\rangle + \Big\langle \gamma+\zeta, \tilde{\mathbb{E}}_{\mathcal{F}_T^1}[\Xi] + 2\lambda\widehat{H}_{pp}(\xi)\eta\Big\rangle\Big]\\ &= \mathbb{E}\Big[-\Big|\Big[(-\widehat{H}_{pp}(\xi) - c_0 I_d)^{\frac{1}{2}}[\gamma+\zeta] - \frac{1}{2}\Lambda(\xi,\eta)\Big|^2 + \frac{1}{4}|\Lambda(\xi,\eta)|^2\Big]\Big]\\ &\leq \frac{1}{4}\mathbb{E}\big[|\Lambda(\xi,\eta)|^2\big].\end{aligned}$$

Then, clearly (5.8) implies (5.1). □

Remark 5.4 *For standard MFGs where* b, f *do not depend on the law of* α, *we have* $\widehat{H}(x,p,\rho) = H(x,p,\mu)$ *where* $\mu = \pi_{1\#}\rho$, *and thus* $\partial_{\rho_1}\widehat{H} = \partial_\mu H$, $\partial_{\rho_2}\widehat{H} = 0$, $c_0 = 0$. *Note that* H *is concave in* p. *We assume it is strictly concave and thus* $H_{pp} < 0$. *Then,* (5.8) *reduces to*

$$\begin{aligned}&\tilde{\mathbb{E}}\Big[\Big\langle \eta, [H_{xx}(\xi) - 2\lambda H_{px}(\xi)]\eta\Big\rangle + \Big\langle \eta,\ [H_{x\mu}(\xi,\tilde{\xi}) - 2\lambda H_{p\mu}(\xi,\tilde{\xi})]\tilde{\eta}\Big\rangle\\ &\quad - \frac{1}{4}\Big|(-H_{pp}(\xi))^{-\frac{1}{2}}\big[\tilde{\mathbb{E}}_{\mathcal{F}_T^1}[H_{p\mu}(\xi,\tilde{\xi})\tilde{\eta}] + 2\lambda H_{pp}(\xi)\eta\big]\Big|^2\Big] \geq 0.\end{aligned} \tag{5.10}$$

Moreover, when $\lambda = 0$, (5.10) *(and* (5.9)*) further reduces to*

$$\tilde{\mathbb{E}}\Big[\Big\langle \eta, H_{xx}(\xi)\eta\Big\rangle + \Big\langle \eta,\ H_{x\mu}(\xi,\tilde{\xi})\tilde{\eta}\Big\rangle - \frac{1}{4}\Big|(-H_{pp}(\xi))^{-\frac{1}{2}}\tilde{\mathbb{E}}_{\mathcal{F}_T^1}[H_{p\mu}(\xi,\tilde{\xi})\tilde{\eta}]\Big|^2\Big] \geq 0. \tag{5.11}$$

This is exactly the condition in [21, Definition 3.4], except that [21] uses $-H$ *instead of* H.

We now present an example which satisfies (5.9), and hence (5.1) with $\lambda = 0$.

Example 5.5 *We consider a special case of* (2.17) *with* $d = 1$*: for some constant* $0 < c < 1$,

$$b(x,a,\mathcal{L}_{(\xi,\alpha)}) = -a + b_1(\mathcal{L}_\xi, \mathbb{E}[\alpha]), \quad f(x,a,\mathcal{L}_{(\xi,\alpha)}) = \frac{|a|^2}{2} - ca\mathbb{E}[\alpha] + f_1(x,\mathcal{L}_{(\xi,\alpha)}). \tag{5.12}$$

Assume there exist constants $0 \leq c_0 < 1$ *and* $\kappa > 0$ *such that*

$$|\bar{c}\partial_{m_2}b_1 - \hat{c}| \leq c_0, \quad \partial_{xx}f_1 \geq \kappa \geq \|\partial_{x\nu_1}f_1\| + \frac{1}{4(1-c_0)}\Big(\|\partial_{m_1}b_1\| + [1+\hat{c}]\|\partial_{x\nu_2}f_1\|\Big)^2, \tag{5.13}$$

where $\hat{c} := \dfrac{c}{1-c}$, $\bar{c} := \dfrac{1}{1-c}$, m_1, m_2 *stand for* $\mathbb{E}[\xi]$ *and* $\mathbb{E}[\alpha]$, *respectively, and* $\|\cdot\|$ *denotes the* *supremum norm of the function over all variables. Then* (5.9) *holds true.*

Proof By Example 2.4, we see that

$$\Phi(\mathcal{L}_{(\xi,\eta)}) = \mathcal{L}_{(\xi,\ \hat{c}\mathbb{E}[\eta]+\eta)},$$
$$H(x,p,\mathcal{L}_{(\xi,\alpha)}) = -\frac{1}{2}\Big|c\mathbb{E}[\alpha]+p\Big|^2 + pb_1(\mathcal{L}_\xi,\mathbb{E}[\alpha]) + f_1(x,\mathcal{L}_{(\xi,\alpha)}),$$
$$\widehat{H}(x,p,\mathcal{L}_{(\xi,\eta)}) = -\frac{1}{2}\Big|\hat{c}\mathbb{E}[\eta]+p\Big|^2 + pb_1(\mathcal{L}_\xi,\bar{c}\mathbb{E}[\eta]) + f_1(x,\mathcal{L}_{(\xi,\ \hat{c}\mathbb{E}[\eta]+\eta)}).$$

By applying Proposition 2.8 and Example 2.7, we have, for $\widehat{f}_1(x,\rho) := f_1(x,\Phi(\rho))$ where $\rho = \mathcal{L}_{(\xi,\eta)}$,

$$\partial_{\rho_1}\widehat{f}_1(x,\rho,\tilde{x},\tilde{p}) = \partial_{\nu_1} f_1(x,\Phi(\rho),\tilde{x},\tilde{p}+\hat{c}\mathbb{E}_\rho[\eta]),$$
$$\partial_{\rho_2}\widehat{f}_1(x,\rho,\tilde{x},\tilde{p}) = \partial_{\nu_2} f_1(x,\Phi(\rho),\tilde{x},\tilde{p}+\hat{c}\mathbb{E}_\rho[\eta]) + \hat{c}\mathbb{E}_\rho\Big[\partial_{\nu_2} f_1(x,\Phi(\rho),\xi,\eta+\hat{c}\mathbb{E}_\rho[\eta])\Big].$$

Then, one may compute that

$$\widehat{H}_{pp} = -1, \quad \widehat{H}_{xx} = \partial_{xx} f_1, \quad \widehat{H}_{x\rho_1} = \partial_{x\nu_1} f_1, \quad \widehat{H}_{p\rho_1} = \partial_{m_1} b_1,$$
$$\widehat{H}_{x\rho_2} = \partial_{x\nu_2} f_1 + \hat{c}\mathbb{E}_\rho[\partial_{x\nu_2} f_1], \quad \widehat{H}_{p\rho_2} = \bar{c}\partial_{m_2} b_1 - \hat{c}.$$

Then $|\widehat{H}_{p\rho_2}| \le c_0$, and (5.9) becomes

$$\tilde{\mathbb{E}}_\rho\Big[\partial_{xx} f_1|\eta|^2 + \partial_{x\nu_1} f_1\eta\tilde{\eta} - \frac{|\Lambda(\xi)|^2}{4(1-c_0)}\Big] \ge 0,$$
$$\text{where } \Lambda(x) := \bar{\tilde{\mathbb{E}}}_\rho\Big[\tilde{\eta}\Big[\partial_{m_1} b_1\big(\mathcal{L}_\xi,\tilde{\xi},\bar{c}\mathbb{E}_\rho[\eta]\big) + \partial_{x\nu_2} f_1\big(\tilde{\xi},\Phi(\rho),x,\varphi(x)+\hat{c}\mathbb{E}_\rho[\varphi(\xi)]\big) \tag{5.14}$$
$$+\hat{c}\partial_{x\nu_2} f_1(\tilde{\xi},\Phi(\rho),\bar{\xi},\varphi(\bar{\xi})+\hat{c}\mathbb{E}_\rho[\varphi(\xi)])\Big]\Big].$$

Clearly, (5.13) implies (5.14), and hence (5.9). □

5.1 Global wellposedness for master equations of standard MFGs

For standard MFGs, by combining Proposition 5.3 and the strategy in [21], see also subsection 3.2, one establishes the following global wellposedness result for the master equation under displacement semi-monotonicity, which generalizes [21, Theorem 6.3]. We remark again that, for MFGC master equations, we shall investigate their global wellposedness in future research.

Theorem 5.6 *Assume* $\lambda \ge 0$, $b(x,a,\nu) = a$ *and* $f(x,a,\nu) = f(x,a,\mu)$. *Further assume:*

(i) H *and* G *have the regularity:*

$$H, \partial_{xx}H, \partial_{xp}H, \partial_{pp}H, \partial_{xxp}H, \partial_{xpp}H, \partial_{ppp}H \in \mathcal{C}^2(\mathbb{R}^{2d}\times\mathcal{P}_2(\mathbb{R}^d)),$$
$$\partial_{x\mu}H, \partial_{p\mu}H, \partial_{xp\mu}H, \partial_{pp\mu}H \in \mathcal{C}^2(\mathbb{R}^{2d}\times\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d),$$
$$G, \partial_{xx}G \in \mathcal{C}^2(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)), \partial_{x\mu}G \in \mathcal{C}^2\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d,$$

and all of the second- and higher-order derivatives of H *and* G *involved above are uniformly bounded.*

(ii) H *is uniformly concave in* p: $\partial_{pp}H \le -c_0 I_d$ *for some constant* $c_0 > 0$.

(iii) (5.10) *holds for* H *and* (3.8) *holds for* G.

Then the master equation (3.6) *on* $[0,T]$ *admits a unique classical solution* V *with bounded* $\partial_{xx}V$ *and* $\partial_{x\mu}V$.

Proof We follow the road map given in section 3.2 to show the global wellposedness. Since the arguments are very similar to those in [21, 35], we only outline a proof below.

Step 1 We apply Theorem 5.2 to show that, if V is a classical solutions of the master equation (3.6) and V satisfies Assumptions 3.6, then V propagates the displacement λ-monotonicity, i.e., $V(t,\cdot,\cdot)$ satisfies (3.8) for all $t \in [0,T]$.

Step 2 We follow the same proof as in [21, Theorem 5.1] to show an a priori uniform $\mathcal{W}_2$-Lipschitz continuity of $\partial_x V$ in μ, uniformly in $(t,x) \in [0,T] \times \mathbb{R}^d$. We note that V might not be uniformly $\mathcal{W}_2$-Lipschitz continuous in μ under our (weaker) assumptions. The key assumption used from [21, Theorem 5.1] is the boundedness of $\partial_{xx} V$, which was proved using the first-order derivatives of H and G in [21, Proposition 6.1]. This is no longer the case here. To show this, we first apply Theorem 5.2 to prove that $V(t,\cdot,\cdot)$ satisfies (3.8) for all $t \in [0,T]$. By Remark 3.5(iv), $\partial_{xx} V$ is uniformly semi-convex in x, uniformly in $(t,\mu) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$. It is standard to obtain the uniform semi-concavity of V in x from the boundedness of the second-order derivatives of H and G by the classical control theory. Thus, we obtain the a priori boundedness of $\partial_{xx} V$. Then we obtain the uniform $\mathcal{W}_2$-Lipschitz continuity of V in μ.

By [21, Proposition 6.2], we further strengthen the above a priori $\mathcal{W}_2$-Lipschitz continuity to an a priori $\mathcal{W}_1$-Lipschitz continuity for $\partial_x V$ in μ.

Step 3 We follow the same proof as the one in [35, Theorem 7.1] to show the global wellposedness of the master equation (3.6). The desired regularity of solution V is a byproduct of Step 2. However, we cannot directly show the wellposedness of the master equation due to the lack of the a priori Lipschitz continuity of V in x and μ, we thus use the approach in [35, section 7]. That is, we first use the a priori Lipschitz estimate of $\partial_x V$ constructed in Step 2 to show the wellposedness of the vectorial master equation for $\vec{U} := \partial_x V$. We then utilize the solution to the vectorial master equation to establish the wellposedness of the master equation (3.6). $\square$

Remark 5.7 *If G satisfies the Lasry–Lions monotonicity and $\partial_{xx} G$ is bounded by λ, then G is displacement semi-monotone. Therefore, we obtain that, if H and G satisfy the assumptions (i) and (ii) in Theorem 5.6, H satisfies* (5.10) *and G is Lasry–Lions monotone, then the master equation is well posed on $[0,T]$. In this sense, Theorem 5.6 unifies the wellposedness results under the Lasry–Lions monotonicity and the displacement monotonicity.*

We remark though, even when G is Lasry–Lions monotone, V propagates the displacement semi-monotonicity, not necessarily the Lasry–Lions monotonicity (when f is non-separable).

6. Propagation of anti-monotonicity

In this section we fix $\vec{\lambda} \in D_4$. Recall (5.7).

Assumption 6.1 (i) *$\widehat{H} \in \mathcal{C}^2(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^{2d}))$ and there exist constants $\overline{L}, L_0 > 0$ and $\overline{\gamma} > \underline{\gamma} > 0$ such that*

$$|\partial_{xp}\widehat{H}| \le \overline{\gamma} L_0, \quad |\partial_{xx}\widehat{H}| \le \overline{\gamma} L_0, \quad |\partial_{pp}\widehat{H}|, |\partial_{x\rho_1}\widehat{H}|, |\partial_{x\rho_2}\widehat{H}|, |\partial_{p\rho_1}\widehat{H}|, |\partial_{p\rho_2}\widehat{H}| \le \overline{L}; \tag{6.1}$$

$$\underline{\kappa}(-\partial_{xp}\widehat{H}) \ge L_0, \quad \underline{\kappa}(-\partial_{xx}\widehat{H}) \ge \underline{\gamma} L_0. \tag{6.2}$$

(ii) *There exists a constant $L^u_{xx} > 0$ such that*

$$\theta_1 := \frac{\overline{\gamma}[1 + L^u_{xx}]}{\sqrt{4(\underline{\gamma}\lambda_0 + 2\lambda_3)}} < 1, \quad \text{and} \quad \overline{L}\,\overline{\kappa}(A_1^{-1}A_2) \le L_0, \tag{6.3}$$

where

$$A_1 := \begin{bmatrix} 4[1-\theta_1] & 0 & 0 \\ 0 & 2\lambda_2 & 0 \\ 0 & 0 & [1-\theta_1][\lambda_0\underline{\gamma}+2\lambda_3] \end{bmatrix},$$
$$A_2 := B_1 L^u_{xx} + B_2 := \begin{bmatrix} 2 & 2+\lambda_2 & 1 \\ 2+\lambda_2 & 4\lambda_2 & \lambda_2 \\ 1 & \lambda_2 & 0 \end{bmatrix} L^u_{xx}$$
$$+ \begin{bmatrix} \lambda_0 + 2|\lambda_0-\lambda_1| & \lambda_0 + |\lambda_0 - \frac{1}{2}\lambda_1| + \frac{1}{2}|\lambda_1| + \lambda_2 & |\lambda_0 - \frac{1}{2}\lambda_1| + \frac{1}{2}|\lambda_1| + 2\lambda_3 \\ \lambda_0 + |\lambda_0 - \frac{1}{2}\lambda_1| + \frac{1}{2}|\lambda_1| + \lambda_2 & 2|\lambda_1| + 2\lambda_2 & |\lambda_1| + \lambda_2 + 2\lambda_3 \\ |\lambda_0 - \frac{1}{2}\lambda_1| + \frac{1}{2}|\lambda_1| + 2\lambda_3 & |\lambda_1| + \lambda_2 + 2\lambda_3 & |\lambda_1| + 2\lambda_3 \end{bmatrix}. \tag{6.4}$$

Theorem 6.2 *Let Assumptions 2.2, 3.1, 3.6, and 6.1 hold. Assume further that, for the constant L^u_{xx} in Assumption 6.1(ii),*

$$|\partial_{xx}V| \le L^u_{xx}. \tag{6.5}$$

If G satisfies the $\vec{\lambda}$-anti-monotonicity (3.10)*, then $V(t,\cdot,\cdot)$ satisfies* (3.10) *for all $t\in[0,T]$.*

We remark that the bound L^u_{xx} of $\partial_{xx}V$ can be estimated a priori by using the Hamilton–Jacobi–Bellman equation or the backward SDE in the mean field game system, see [35, section 6] for more details.

Proof Without loss of generality, we prove the theorem only for $t=0$. We continue to use the notation as in the proofs of Theorem 4.2 and 5.2. Introduce

$$\Xi_t := \lambda_0 \bar I(t) + \lambda_1 I(t) + \mathbb{E}\big[|\Gamma_t|^2 + \lambda_2|\Upsilon_t|^2 - \lambda_3|\delta X_t|^2\big] = MON^{anti}_{\vec\lambda} V(t,\cdot,\cdot)(X_t,\delta X_t).$$

Then it is sufficient to show that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Xi_t \ge 0. \tag{6.6}$$

Following the calculation in [35, Theorem 4.1], we have

$$\begin{aligned} \mathrm{d}\Upsilon_t &= \big[-K_1(t)\Upsilon_t - K_2(t)\big]\mathrm{d}t + (\mathrm{d}B_t)^\top K_3(t) + \beta(\mathrm{d}B^0_t)^\top K_4(t);\\ \mathrm{d}\Gamma_t &= \big[-2\widehat H_{xp}(X_t)\Gamma_t + \partial_{xx}V(X_t)\widehat H_{pp}(X_t)\Upsilon_t - \bar K_1(t)\big]\mathrm{d}t + (\mathrm{d}B_t)^\top \bar K_2(t) + \beta(\mathrm{d}B^0_t)^\top \bar K_3(t), \end{aligned} \tag{6.7}$$

where

$$\begin{aligned} K_1(t) &:= \widehat H_{xp}(X_t) + \partial_{xx}V(X_t)\widehat H_{pp}(X_t),\\ K_2(t) &:= \tilde{\mathbb{E}}_{\mathcal F_t}\Big[\big[\widehat H_{x\rho_1}(X_t,\tilde X_t)\delta\tilde X_t + \widehat H_{x\rho_2}(X_t,\tilde X_t)[\tilde\Gamma_t+\tilde\Upsilon_t]\big]\\ &\qquad + \partial_{xx}V(X_t)\big[\widehat H_{p\rho_1}(X_t,\tilde X_t)\delta\tilde X_t + \widehat H_{p\rho_2}(X_t,\tilde X_t)[\tilde\Gamma_t+\tilde\Upsilon_t]\big]\Big],\\ K_3(t) &:= \tilde{\mathbb{E}}_{\mathcal F_t}\big[\partial_{xx\mu}V(X_t,\tilde X_t)\delta\tilde X_t\big],\\ K_4(t) &:= K_3(t) + \bar{\tilde{\mathbb{E}}}_{\mathcal F_t}\Big[\big[(\partial_{\mu x\mu}V)(X_t,\bar X_t,\tilde X_t) + \partial_{\tilde x x\mu}V(X_t,\tilde X_t)\big]\delta\tilde X_t\Big],\\ \bar K_1(t) &:= [\widehat H_{xx}(X_t) - \partial_{xx}V(X_t)\widehat H_{px}(X_t)]\delta X_t\\ &\qquad - \partial_{xx}V(X_t)\tilde{\mathbb{E}}_{\mathcal F_t}\big[\widehat H_{p\rho_1}(X_t,\tilde X_t)\delta\tilde X_t + \widehat H_{p\rho_2}(X_t,\tilde X_t)[\tilde\Gamma_t+\tilde\Upsilon_t]\big],\\ \bar K_2(t) &:= \partial_{xxx}V(X_t)\delta X_t,\\ \bar K_3(t) &:= \bar K_2(t) + \tilde{\mathbb{E}}_{\mathcal F_t}\Big[(\partial_{\mu xx}V)(X_t,\tilde X_t)\delta\tilde X_t\Big]. \end{aligned} \tag{6.8}$$

In particular, this implies that

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[|\Upsilon_t|^2] \geq 2\mathbb{E}\Big[\big\langle \Upsilon_t,\ -K_1(t)\Upsilon_t - K_2(t)\big\rangle\Big];\\&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[|\Gamma_t|^2] \geq 2\mathbb{E}\Big[\big\langle \Gamma_t,\ -2\widehat{H}_{xp}(X_t)\Gamma_t + \partial_{xx}V(X_t)\widehat{H}_{pp}(X_t)\Upsilon_t - \bar{K}_1(t)\big\rangle\Big].\end{aligned} \tag{6.9}$$

Thus, combining (4.5), (5.3), and (5.4) and recalling N in (4.2), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Xi_t \geq{}& \lambda_0\tilde{\mathbb{E}}\Big[\big\langle \widehat{H}_{pp}(X_t)\Gamma_t, \Gamma_t\big\rangle + 2\big\langle \widehat{H}_{pp}(X_t)\Gamma_t, \Upsilon_t\big\rangle\\&+ 2\big\langle \Gamma_t, \widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t + \widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)[\tilde{\Gamma}_t + \tilde{\Upsilon}_t]\big\rangle - \big\langle \widehat{H}_{xx}(X_t)\delta X_t, \delta X_t\big\rangle\Big]\\&+ \lambda_1\tilde{\mathbb{E}}\Big[\big\langle \widehat{H}_{pp}(X_t)\Upsilon_t,\ \Upsilon_t\big\rangle - \big\langle \widehat{H}_{x\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t + \widehat{H}_{x\rho_2}(X_t,\tilde{X}_t)[\tilde{\Gamma}_t + \tilde{\Upsilon}_t], \delta X_t\big\rangle\\&- \big\langle \widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t + \widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)[\tilde{\Gamma}_t + \tilde{\Upsilon}_t], \Gamma_t - \Upsilon_t\big\rangle\Big]\\&+ 2\tilde{\mathbb{E}}\Big[\big\langle \Gamma_t, \big[-2\widehat{H}_{xp}(X_t)\Gamma_t + \partial_{xx}V(X_t)\widehat{H}_{pp}(X_t)\Upsilon_t - \bar{K}_1(t)\big]\big\rangle + \lambda_2\big\langle \Upsilon_t, \big[-K_1(t)\Upsilon_t - K_2(t)\big]\big\rangle\Big]\\&- 2\lambda_3\tilde{\mathbb{E}}\Big[\big\langle \widehat{H}_{px}(X_t)\delta X_t + \widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\delta\tilde{X}_t + \widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)[\tilde{\Gamma}_t + \tilde{\Upsilon}_t] + \widehat{H}_{pp}(X_t)[\Upsilon_t + \Gamma_t], \delta X_t\big\rangle\Big]\\={}& \tilde{\mathbb{E}}\Big[\Big\langle \big[\lambda_0\widehat{H}_{pp}(X_t) - 4\widehat{H}_{xp}(X_t)\big]\Gamma_t,\ \Gamma_t\Big\rangle + \Big\langle \big[\lambda_1\widehat{H}_{pp}(X_t) - 2\lambda_2K_1(t)\big]\Upsilon_t, \Upsilon_t\Big\rangle\\&+ \Big\langle \big[-\lambda_0\widehat{H}_{xx}(X_t) - 2\lambda_3\widehat{H}_{px}(X_t)\big]\delta X_t, \delta X_t\Big\rangle + \Big\langle \big[2\lambda_0 - \lambda_1 + 2\partial_{xx}V(X_t)\big]\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\tilde{\Gamma}_t, \Gamma_t\Big\rangle\\&+ \Big\langle \big[[\lambda_1 - 2\lambda_2\partial_{xx}V(X_t)]\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t) - 2\lambda_2\widehat{H}_{x\rho_2}(X_t,\tilde{X}_t)\big]\tilde{\Upsilon}_t, \Upsilon_t\Big\rangle\\&- \Big\langle \big[\lambda_1\widehat{H}_{x\rho_1}(X_t,\tilde{X}_t) + 2\lambda_3\widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)\big]\delta\tilde{X}_t, \delta X_t\Big\rangle\\&+ \Big\langle 2[\lambda_0\widehat{H}_{pp}(X_t) + \partial_{xx}V(X_t)\widehat{H}_{pp}(X_t)]\Upsilon_t + [2\lambda_0 - \lambda_1 + 2\partial_{xx}V(X_t)]\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\tilde{\Upsilon}_t, \Gamma_t\Big\rangle\\&+ \Big\langle \big[[\lambda_1 - 2\lambda_2\partial_{xx}V(X_t)]\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t) - 2\lambda_2\widehat{H}_{x\rho_2}(X_t,\tilde{X}_t)\big]\tilde{\Gamma}_t, \Upsilon_t\Big\rangle\\&+ \Big\langle \big[[2\lambda_0 - \lambda_1 + 2\partial_{xx}V(X_t)]\widehat{H}_{p\rho_1}(X_t,\tilde{X}_t) - \lambda_1\widehat{H}_{\rho_2x}(\tilde{X}_t,X_t) - 2\lambda_3\widehat{H}_{\rho_2p}(\tilde{X}_t,X_t)\big]\delta\tilde{X}_t\\&+ 2\big[-\widehat{H}_{xx}(X_t) + \partial_{xx}V(X_t)\widehat{H}_{px}(X_t) - \lambda_3\widehat{H}_{pp}(X_t)\big]\delta X_t, \Gamma_t\Big\rangle\\&+ \Big\langle \big[[\lambda_1 - 2\lambda_2\partial_{xx}V(X_t)]\widehat{H}_{p\rho_1}(X_t,\tilde{X}_t) - \lambda_1\widehat{H}_{\rho_2x}(\tilde{X}_t,X_t)\\&- 2\lambda_2\widehat{H}_{x\rho_1}(X_t,\tilde{X}_t) - 2\lambda_3\widehat{H}_{\rho_2p}(\tilde{X}_t,X_t)\big]\delta\tilde{X}_t - 2\lambda_3\widehat{H}_{pp}(X_t)\delta X_t, \Upsilon_t\Big\rangle\Big].\end{aligned}$$

Recall (3.9), (5.7), and (6.8), by (6.1) and (6.2) we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\Xi_t \geq{}& \big[4L_0 - \lambda_0\overline{L}\big]\mathbb{E}[|\Gamma_t|^2] + \big[2\lambda_2 L_0 - |\lambda_1|\overline{L} - 2\lambda_2 L^u_{xx}\overline{L}\big]\mathbb{E}[|\Upsilon_t|^2] \\
&+ \big[\lambda_0\underline{\gamma}L_0 + 2\lambda_3 L_0\big]\mathbb{E}[|\delta X_t|^2] - \big[|2\lambda_0 - \lambda_1| + 2L^u_{xx}\big]\overline{L}\big(\mathbb{E}[|\Gamma_t|]\big)^2 \\
&- \big[|\lambda_1| + 2\lambda_2 L^u_{xx} + 2\lambda_2\big]\overline{L}\big(\mathbb{E}[|\Upsilon_t|]\big)^2 - \big[|\lambda_1| + 2\lambda_3\big]\overline{L}\big(\mathbb{E}[|\delta X_t|]\big)^2 \\
&- \big[2[\lambda_0 + L^u_{xx}] + |2\lambda_0 - \lambda_1| + 2L^u_{xx} + |\lambda_1| + 2\lambda_2 L^u_{xx} + 2\lambda_2\big]\overline{L}\mathbb{E}[|\Gamma_t|]\mathbb{E}[|\Upsilon_t|] \\
&- \big[[|2\lambda_0 - \lambda_1| + 2L^u_{xx} + |\lambda_1| + 2\lambda_3]\overline{L} + 2[\overline{\gamma}L_0 + L^u_{xx}\overline{\gamma}L_0 + \lambda_3\overline{L}]\big]\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Gamma_t|] \\
&- \big[[|\lambda_1| + 2\lambda_2 L^u_{xx} + |\lambda_1| + 2\lambda_2 + 2\lambda_3 + 2\lambda_3]\overline{L}\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Upsilon_t|] \\
\geq{}& \Big[4L_0 - \big[\lambda_0 + |2\lambda_0 - \lambda_1| + 2L^u_{xx}\big]\overline{L}\Big]\big(\mathbb{E}[|\Gamma_t|]\big)^2 \\
&+ \Big[2\lambda_2 L_0 - \big[2|\lambda_1| + 4\lambda_2 L^u_{xx} + 2\lambda_2\big]\overline{L}\Big]\big(\mathbb{E}[|\Upsilon_t|]\big)^2 \\
&+ \Big[[\lambda_0\underline{\gamma} + 2\lambda_3]L_0 - \big[|\lambda_1| + 2\lambda_3\big]\overline{L}\Big]\big(\mathbb{E}[|\delta X_t|]\big)^2 \\
&- 2\Big[\lambda_0 + |\lambda_0 - \frac{\lambda_1}{2}| + \frac{|\lambda_1|}{2} + \lambda_2 + [2+\lambda_2]L^u_{xx}\Big]\overline{L}\mathbb{E}[|\Gamma_t|]\mathbb{E}[|\Upsilon_t|] \\
&- 2\Big[\overline{\gamma}[1 + L^u_{xx}]L_0 + \big[|\lambda_0 - \frac{\lambda_1}{2}| + \frac{|\lambda_1|}{2} + 2\lambda_3 + L^u_{xx}\big]\overline{L}\Big]\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Gamma_t|] \\
&- 2\Big[|\lambda_1| + \lambda_2 + \lambda_2 L^u_{xx} + 2\lambda_3\Big]\overline{L}\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Upsilon_t|] \\
\geq{}& \Big[4[1-\theta_1]L_0 - \big[\lambda_0 + |2\lambda_0 - \lambda_1| + 2L^u_{xx}\big]\overline{L}\Big]\big(\mathbb{E}[|\Gamma_t|]\big)^2 \\
&+ \Big[2\lambda_2 L_0 - \big[2|\lambda_1| + 4\lambda_2 L^u_{xx} + 2\lambda_2\big]\overline{L}\Big]\big(\mathbb{E}[|\Upsilon_t|]\big)^2 \\
&+ \Big[(1-\theta_1)[\lambda_0\underline{\gamma} + 2\lambda_3]L_0 - \big[|\lambda_1| + 2\lambda_3\big]\overline{L}\Big]\big(\mathbb{E}[|\delta X_t|]\big)^2 \\
&- 2\Big[\lambda_0 + |\lambda_0 - \frac{\lambda_1}{2}| + \frac{|\lambda_1|}{2} + \lambda_2 + [2+\lambda_2]L^u_{xx}\Big]\overline{L}\mathbb{E}[|\Gamma_t|]\mathbb{E}[|\Upsilon_t|] \\
&- 2\Big[\big[|\lambda_0 - \frac{\lambda_1}{2}| + \frac{|\lambda_1|}{2} + 2\lambda_3 + L^u_{xx}\big]\overline{L}\Big]\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Gamma_t|] \\
&- 2\Big[|\lambda_1| + \lambda_2 + \lambda_2 L^u_{xx} + 2\lambda_3\Big]\overline{L}\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Upsilon_t|],
\end{aligned}
$$

where in the last step we use the fact that: recalling the θ_1 in (6.3),

$$2\overline{\gamma}[1 + L^u_{xx}]\mathbb{E}[|\delta X_t|]\mathbb{E}[|\Gamma_t|] \leq 4\theta_1\big(\mathbb{E}[|\Gamma_t|]\big)^2 + \theta_1[\lambda_0\underline{\gamma} + 2\lambda_3]\big(\mathbb{E}[|\delta X_t|]\big)^2,$$

Then, recalling (6.4) and denoting $e := \big(\mathbb{E}[|\Gamma_t|],\ \mathbb{E}[|\Upsilon_t|],\ \mathbb{E}[|\delta X_t|]\big)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Xi_t \geq e\big[A_1 L_0 - A_2\overline{L}\big]e^\top \geq 0,$$

thanks to (6.3) and the fact that $A_1>0$. □

Example 6.3 *Again, we consider a special case of* (2.17) *with* $d=1$*:*

$$b(x,a,\mathcal{L}_{(\xi,\alpha)})=-a-L_0x+b_1(\mathcal{L}_\xi,\mathbb{E}[\alpha]),$$
$$f(x,a,\mathcal{L}_{(\xi,\alpha)})=\frac{|a|^2}{2}-ca\mathbb{E}[\alpha]-\frac{\gamma L_0}{2}x^2+f_1(x,\mathcal{L}_{(\xi,\alpha)}),$$

for some constants $0<c<1$*,* $\gamma>0$*, and* $L_0>0$*. For any* $L^u_{xx}>0$*, when* L_0 *is large enough, there exist appropriate* $\overline{L}>0$*,* $\overline{\gamma}>\underline{\gamma}>0$*, and* $\vec{\lambda}\in D_4$ *such that Assumption 6.1 holds true.*

Proof By Example 2.4 and recalling the notations $\hat{c},\bar{c}$ in Example 5.5, we find that

$$\Phi(\mathcal{L}_{(\xi,\eta)})=\mathcal{L}_{(\xi,\ \hat{c}\mathbb{E}[\eta]+\eta)},$$
$$H(x,p,\mathcal{L}_{(\xi,\alpha)})=-\frac{1}{2}\Big|c\mathbb{E}[\alpha]+p\Big|^2-L_0xp+pb_1(\mathcal{L}_\xi,\mathbb{E}[\alpha])-\frac{\gamma L_0}{2}x^2+f_1(x,\mathcal{L}_{(\xi,\alpha)}),$$
$$\widehat{H}(x,p,\mathcal{L}_{(\xi,\eta)})=-\frac{1}{2}\Big|\hat{c}\mathbb{E}[\eta]+p\Big|^2-L_0xp+pb_1(\mathcal{L}_\xi,\bar{c}\mathbb{E}[\eta])-\frac{\gamma L_0}{2}x^2+f_1(x,\mathcal{L}_{(\xi,\ \hat{c}\mathbb{E}[\eta]+\eta)}).$$

Following the same calculation as in Example 5.5,

$$\widehat{H}_{pp}=-1,\quad \widehat{H}_{xp}=-L_0,\quad \widehat{H}_{xx}=-\gamma L_0+\partial_{xx}f_1,\quad \widehat{H}_{x\rho_1}=\partial_{x\nu_1}f_1,$$
$$\widehat{H}_{x\rho_2}=\partial_{x\nu_2}f_1+\hat{c}\mathbb{E}[\partial_{x\nu_2}f_1],\quad \widehat{H}_{p\rho_1}=\partial_{m_1}b_1,\quad \widehat{H}_{p\rho_2}=\bar{c}\partial_{m_2}b_1-\hat{c}.$$

For given functions f_1,b_1, clearly there exists a fixed constant $\overline{L}>0$ such that

$$|\partial_{pp}\widehat{H}|,\ |\partial_{x\rho_1}\widehat{H}|,\ |\partial_{x\rho_2}\widehat{H}|,\ |\partial_{p\rho_1}\widehat{H}|,\ |\partial_{p\rho_2}\widehat{H}|\le\overline{L}.$$

Set $\underline{\gamma}:=\frac{\gamma}{2}$, $\overline{\gamma}:=\gamma+1$, for L_0 sufficiently large, we have

$$\underline{\kappa}(-\widehat{H}_{xp})=L_0,\qquad |\widehat{H}_{xp})|=L_0\le\overline{\kappa}L_0;$$
$$|\partial_{xx}\widehat{H}|-\overline{\gamma}L_0\le\gamma L_0+\overline{L}-\overline{\gamma}L_0=\overline{L}-L_0\le 0;$$
$$\underline{\kappa}(-\partial_{xx}\widehat{H})-\underline{\gamma}L_0\ge\gamma L_0-\overline{L}-\underline{\gamma}L_0=\tfrac{\gamma}{2}L_0-\overline{L}\ge 0.$$

That is, (6.1) and (6.2) hold true.

We now fix arbitrary $\lambda_0,\lambda_2>0$ and $\lambda_1\in\mathbb{R}$. Choosing $\lambda_3>0$ sufficiently large, we have that $\theta_1\le\frac{1}{2}$. Finally, set L_0 sufficiently large such that $L_0\ge\overline{L}\overline{\kappa}(A_1^{-1}A_2)$, we verify (6.3) as well. □

We point out though, by (6.5), L^u_{xx} may in turn depend on L_0, so extra effort is needed to ensure full compatibility of our conditions. This, however, requires the a priori estimate for $\partial_{xx}V$ which is not carried out in this paper. We thus leave it to our accompanying paper on global wellposedness of MFGC master equations. We remark that we have a complete result in [35] for standard MFG master equations.

7. Appendix

Proof of (4.5) We first apply the Itô's formula (3.5) on $\partial_{x\mu}V(t,X_t,\mathcal{L}_{X_t|\mathcal{F}^0_t},\tilde{X}_t)$ to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t)=I_1+I_2+I_3,\tag{7.1}$$

where, by using $\hat{X}$ to denote another conditionally independent copy,

$$
\begin{aligned}
I_1 := \bar{\tilde{\mathbb{E}}}\Big[\Big\langle\Big\{&\partial_{tx\mu}V(X_t,\tilde{X}_t)+\frac{\widehat{\beta}^2}{2}((\mathrm{tr}\partial_{xx})\partial_{x\mu}V)(X_t,\tilde{X}_t)+\widehat{H}_p(X_t)^\top\partial_{xx\mu}V(X_t,\tilde{X}_t)\\
&+\beta^2(\mathrm{tr}(\partial_{x\mu})\partial_{x\mu}V)(X_t,\bar{X}_t,\tilde{X}_t)+\beta^2(\mathrm{tr}(\partial_{\tilde{x}\mu})\partial_{x\mu}V)(X_t,\bar{X}_t,\tilde{X}_t)\\
&+\beta^2(\mathrm{tr}(\partial_{\tilde{x}x})\partial_{x\mu}V)(X_t,\tilde{X}_t)+\frac{\beta^2}{2}(\mathrm{tr}(\partial_{\mu\mu})\partial_{x\mu}V)(X_t,\hat{X}_t,\bar{X}_t,\tilde{X}_t)\\
&+\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{\bar{x}\mu})\partial_{x\mu}V)(X_t,\bar{X}_t,\tilde{X}_t)+\widehat{H}_p(\bar{X}_t)^\top\partial_{\mu x\mu}V(X_t,\bar{X}_t,\tilde{X}_t)\\
&+\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{\tilde{x}\tilde{x}})\partial_{x\mu}V)(X_t,\tilde{X}_t)+\widehat{H}_p(\tilde{X}_t)^\top\partial_{\tilde{x}x\mu}V(X_t,\tilde{X}_t)\Big\}\delta\tilde{X}_t,\ \delta X_t\Big\rangle\Big];\\
I_2 := -\hat{\bar{\tilde{\mathbb{E}}}}\Big[\Big\langle&\partial_{\mu x}V(X_t,\tilde{X}_t)\Big\{\big[\widehat{H}_{px}(X_t)+\widehat{H}_{pp}(X_t)\partial_{xx}V(X_t)\big]\delta X_t\\
&+\Big[\widehat{H}_{p\rho_1}(X_t,\bar{X}_t)+\widehat{H}_{p\rho_2}(X_t,\bar{X}_t)\partial_{xx}V(\bar{X}_t)+\widehat{H}_{pp}(X_t)\partial_{x\mu}V(X_t,\bar{X}_t)\Big]\delta\bar{X}_t\\
&+\widehat{H}_{p\rho_2}(X_t,\bar{X}_t)\partial_{x\mu}V(\bar{X}_t,\hat{X}_t)\delta\hat{X}_t\Big\},\ \delta\tilde{X}_t\Big\rangle\Big];\\
I_3 := -\hat{\bar{\tilde{\mathbb{E}}}}\Big[\Big\langle&\partial_{x\mu}V(X_t,\tilde{X}_t)\Big\{\big[\widehat{H}_{px}(\tilde{X}_t)+\widehat{H}_{pp}(\tilde{X}_t)\partial_{xx}V(\tilde{X}_t)\big]\delta\tilde{X}_t\\
&+\Big[\widehat{H}_{p\rho_1}(\tilde{X}_t,\bar{X}_t)+\widehat{H}_{p\rho_2}(\tilde{X}_t,\bar{X}_t)\partial_{xx}V(\bar{X}_t)+\widehat{H}_{pp}(\tilde{X}_t)\partial_{x\mu}V(\tilde{X}_t,\bar{X}_t)\Big]\delta\bar{X}_t\\
&+\widehat{H}_{p\rho_2}(\tilde{X}_t,\bar{X}_t)\partial_{x\mu}V(\bar{X}_t,\hat{X}_t)\delta\hat{X}_t\Big\},\delta X_t\Big\rangle\Big].
\end{aligned}
$$

On the other hand, applying $\partial_{x\mu}$ to (3.6) we obtain

$$
0=(\partial_{x\mu}\mathscr{L}V)(t,x,\mu,\tilde{x})=J_1+J_2+J_3, \tag{7.2}
$$

where

$$
\begin{aligned}
J_1 :=\ &\partial_{tx\mu}V(x,\tilde{x})+\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{xx})\partial_{x\mu}V)(x,\tilde{x})+\widehat{H}_{x\rho_1}(x,\tilde{x})\\
&+\widehat{H}_{x\rho_2}(x,\tilde{x})\partial_{xx}V(\tilde{x})+\bar{\mathbb{E}}[\widehat{H}_{x\rho_2}(x,\bar{\xi})\partial_{x\mu}V(\bar{\xi},\tilde{x})]\\
&+\partial_{xx}V(x)\Big[\widehat{H}_{p\rho_1}(x,\tilde{x})+\widehat{H}_{p\rho_2}(x,\tilde{x})\partial_{xx}V(\tilde{x})+\bar{\mathbb{E}}[\widehat{H}_{p\rho_2}(x,\bar{\xi})\partial_{x\mu}V(\bar{\xi},\tilde{x})]\Big]\\
&+\Big[\widehat{H}_{xp}(x)+\partial_{xx}V(x)\widehat{H}_{pp}(x)\Big]\partial_{x\mu}V(x,\tilde{x})+\widehat{H}_p(x)^\top\partial_{xx\mu}V(x,\tilde{x});\\
J_2 :=\ &\frac{\widehat{\beta}^2}{2}(\partial_{x\tilde{x}}\mathrm{tr}(\partial_{\tilde{x}\mu})V)(x,\tilde{x})+\partial_{x\mu}V(x,\tilde{x})\Big[\widehat{H}_{px}(\tilde{x})+\widehat{H}_{pp}(\tilde{x})\partial_{xx}V(\tilde{x})\Big]\\
&+\widehat{H}_p(\tilde{x})^\top\partial_{\tilde{x}x\mu}V(x,\tilde{x})+\beta^2(\partial_{x\tilde{x}}\mathrm{tr}(\partial_{x\mu})V)(x,\tilde{x})+\beta^2\bar{\mathbb{E}}\big[(\partial_{x\tilde{x}}\mathrm{tr}(\partial_{\mu\mu})V)(x,\bar{\xi},\tilde{x})\big];\\
J_3 :=\ &\hat{\bar{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{\bar{x}\mu})\partial_{x\mu}V)(x,\tilde{x},\bar{\xi})+\widehat{H}_p(\bar{\xi})^\top\partial_{\mu x\mu}V(x,\tilde{x},\bar{\xi})\\
&+\partial_{x\mu}V(x,\bar{\xi})\Big[\widehat{H}_{p\rho_1}(\bar{\xi},\tilde{x})+\widehat{H}_{p\rho_2}(\bar{\xi},\tilde{x})\partial_{xx}V(\tilde{x})+\widehat{H}_{p\rho_2}(\bar{\xi},\hat{\xi})\partial_{x\mu}V(\hat{\xi},\tilde{x})+\widehat{H}_{pp}(\bar{\xi})\partial_{x\mu}V(\bar{\xi},\tilde{x})\Big]\\
&+\beta^2(\mathrm{tr}(\partial_{x\mu})\partial_{x\mu}V)(x,\tilde{x},\bar{\xi})+\frac{\beta^2}{2}(\mathrm{tr}(\partial_{\mu\mu})\partial_{x\mu}V)(x,\tilde{x},\hat{\xi},\bar{\xi})\Big].
\end{aligned}
$$

Then, evaluate (7.2) along $(X_t,\mu_t,\tilde{X}_t)$ and plug into (7.1). As in [21, Theorem 4.1], we obtain (4.5). □

Proof of Remark 4.4 Given $\xi_i \in \mathbb{L}^2(\mathcal{F}_0)$, $i=1,2$, let X^i solve the McKean–Vlasov SDE:

$$X_t^i = \xi_i + \int_0^t \partial_p \widehat{H}(X_s^i, \partial_x V(s, X_s^i, \mu_s^i), \rho_s^i)\mathrm{d}s + B_t + \beta B_t^0,\\ \text{where} \quad \mu_t^i := \mathcal{L}_{X_t^i|\mathcal{F}_t^0}, \quad \rho_t^i := \mathcal{L}_{(X_t^i, \partial_x V(t, X_t^i, \mu_t^i))|\mathcal{F}_t^0}. \tag{7.3}$$

It is standard that the optimal control is $\alpha_s^i := \partial_p \widehat{H}(X_s^i, \partial_x V(s, X_s^i, \mu_s^i), \rho_s^i)$, and thus

$$\mathbb{E}[V(t, X_t^i, \mu_t^i)] = \mathbb{E}\Big[V(t_\delta, X_{t_\delta}^i, \mu_{t_\delta}^i) + \int_t^{t_\delta} f(X_s^i, \alpha_s^i, \rho_s^i)\mathrm{d}s\Big], \tag{7.4}$$

where $t_\delta := t+\delta$. Let α^δ be any admissible control in $\mathcal{A}_{t_\delta}$. Consider

$$\alpha^{i,\delta}(s,x) := \begin{cases} \partial_p \widehat{H}(x, \partial_x V(s, x, \mu_s^i), \rho_s^i), & s \in [t, t_\delta); \\ \alpha^\delta(s,x), & s \in [t_\delta, T]. \end{cases}$$

$$X_s^{i,\delta} = X_t^i + \int_t^s \alpha^{i,\delta}(s, X_s^{i,\delta})\mathrm{d}s + B_s^t + \beta B_s^{0,t}, \quad s \in [t,T].$$

Since $b(\cdot, a, \cdot) = a$, we have $X_s^{i,\delta} = X_s^i$ and $\alpha_s^i = \alpha^{i,\delta}(s, X_s^{i,\delta})$ for any $s \in [t, t_\delta]$. Moreover,

$$X_s^{i,\delta} = X_{t_\delta}^i + \int_{t_\delta}^s \alpha^\delta(s, X_s^{i,\delta})\mathrm{d}s + B_s^{t_\delta} + \beta B_s^{0,t_\delta}, \quad s \in [t_\delta, T].$$

Thus, for $i,j=1,2$ with $i \neq j$,

$$\begin{aligned}\mathbb{E}[V(t, X_t^j, \mu_t^i)] &\le \mathbb{E}\Big[G(X_T^{j,\delta}, \mu_T^i) + \int_t^T f(X_s^{j,\delta}, \alpha^{j,\delta}(s, X_s^{j,\delta}), \rho_s^i)\mathrm{d}s\Big] \\ &= \mathbb{E}\Big[G(X_T^{j,\delta}, \mu_T^i) + \int_{t_\delta}^T f(X_s^{j,\delta}, \alpha^\delta(s, X_s^{j,\delta}), \rho_s^i)\mathrm{d}s + \int_t^{t_\delta} f(X_s^j, \alpha_s^j, \rho_s^i)\mathrm{d}s\Big].\end{aligned}$$

Taking infimum over all admissible controls α^δ in $\mathcal{A}_{t_\delta}$ above, we have

$$\mathbb{E}[V(t, X_t^j, \mu_t^i)] \le \mathbb{E}\Big[V(t_\delta, X_{t_\delta}^j, \mu_{t_\delta}^i) + \int_t^{t_\delta} f(X_s^j, \alpha_s^j, \rho_s^i)\mathrm{d}s\Big]. \tag{7.5}$$

Therefore, by (7.4), (7.5), and (4.7),

$$\begin{aligned}&\mathbb{E}\Big[V(t_\delta, X_{t_\delta}^1, \mu_{t_\delta}^1) + V(t_\delta, X_{t_\delta}^2, \mu_{t_\delta}^2) - V(t_\delta, X_{t_\delta}^1, \mu_{t_\delta}^2) - V(t_\delta, X_{t_\delta}^2, \mu_{t_\delta}^1)\Big] \\ &- \mathbb{E}\Big[V(t, X_t^1, \mu_t^1) + V(t, X_t^2, \mu_t^2) - V(t, X_t^1, \mu_t^2) - V(t, X_t^2, \mu_t^1)\Big] \\ \le\; &- \mathbb{E}\Big[\int_t^{t+\delta} \big[f(X_s^1, \alpha_s^1, \rho_s^1) + f(X_s^2, \alpha_s^2, \rho_s^2) - f(X_s^1, \alpha_s^1, \rho_s^2) - f(X_s^2, \alpha_s^2, \rho_s^1)\big]\mathrm{d}s\Big] \le 0.\end{aligned}$$

Dividing both sides by δ and then sending $\delta \to 0$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[V(t, X_t^1, \mu_t^1) + V(t, X_t^2, \mu_t^2) - V(t, X_t^1, \mu_t^2) - V(t, X_t^2, \mu_t^1)\Big] \le 0,$$

which implies that, denoting $\Delta X_t := X_t^2 - X_t^1$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbb{E}}\Big[\int_0^1 \big\langle \partial_{x\mu} V(t, X_t^1 + \theta\Delta X_t, \mathcal{L}_{(X_t^1+\theta\Delta X_t)|\mathcal{F}_t^0},\ \tilde{X}_t^1 + \theta\Delta\tilde{X}_t)\Delta\tilde{X}_t,\ \Delta X_t\big\rangle\Big]\mathrm{d}\theta \le 0. \tag{7.6}$$

Now fix $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_0)$ and set $\xi_1 := \xi$, $\xi_2 := \xi + \varepsilon\eta$. Then X^1 identifies the X in (4.2), and by denoting $X^\varepsilon = X^2$, one can verify that $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[X_t^\varepsilon - X_t] = \delta X_t$, where the limit is in the $\mathbb{L}^2$ sense and δX is defined in (4.2). Then, by dividing (7.6) with ε^2 and sending $\varepsilon \to 0$, it follows from the regularity of V that

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\mathbb{E}}\Big[\big\langle \partial_{x\mu}V(t,X_t,\mathcal{L}_{X_t|\mathcal{F}^0_t},\tilde{X}_t)\tilde{\delta}X_t,\ \delta X_t\big\rangle\Big]\le 0.$$

This is exactly (4.6). □

Proof of (5.3) We first apply the Itô formula (3.5) to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{I}(t)=\bar{I}_1+\bar{I}_2+\bar{I}_3, \tag{7.7}$$

where

$$\begin{aligned}
\bar{I}_1 &:= \tilde{\mathbb{E}}\Big[\Big\langle\Big\{\partial_{txx}V(X_t)+\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{xx})\partial_{xx}V)(X_t)+\widehat{H}_p(X_t)^\top\partial_{xxx}V(X_t)\\
&\quad +\beta^2(\mathrm{tr}(\partial_{x\mu})\partial_{xx}V)(X_t,\tilde{X}_t)\Big\}\delta X_t,\delta X_t\Big\rangle\Big],\\
\bar{I}_2 &:= \bar{\tilde{\mathbb{E}}}\Big[\Big\langle\Big\{\frac{\beta^2}{2}(\mathrm{tr}(\partial_{\mu\mu})\partial_{xx}V)(X_t,\tilde{X}_t,\bar{X}_t)\\
&\quad +\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{\tilde{x}\mu})\partial_{xx}V)(X_t,\tilde{X}_t)+\widehat{H}_p(\tilde{X}_t)^\top\partial_{\mu xx}V(X_t,\tilde{X}_t)\Big\}\delta X_t,\delta X_t\Big\rangle\Big],\\
\bar{I}_3 &:= 2\bar{\tilde{\mathbb{E}}}\Big[\Big\langle\partial_{xx}V(X_t)\Big\{\big[\widehat{H}_{px}(X_t)+\widehat{H}_{pp}(X_t)\partial_{xx}V(X_t)\big]\delta X_t\\
&\quad \big[\widehat{H}_{p\rho_1}(X_t,\tilde{X}_t)+\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\partial_{xx}V(\tilde{X}_t)+\widehat{H}_{pp}(X_t)\partial_{x\mu}V(X_t,\tilde{X}_t)\big]\delta\tilde{X}_t\\
&\quad +\widehat{H}_{p\rho_2}(X_t,\tilde{X}_t)\partial_{x\mu}V(\tilde{X}_t,\bar{X}_t)\delta\bar{X}_t\Big\},\ \delta X_t\Big\rangle\Big].
\end{aligned}$$

On the other hand, applying ∂_{xx} to (3.6), we obtain

$$0=(\partial_{xx}\mathscr{L}V)(t,x,\mu)=\bar{J}_1+\bar{J}_2, \tag{7.8}$$

where

$$\begin{aligned}
\bar{J}_1 &:= \partial_{txx}V+\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{xx})\partial_{xx}V)+\widehat{H}_{xx}(x)+2\widehat{H}_{xp}(x)\partial_{xx}V(x)\\
&\quad +\partial_{xx}V(x)\widehat{H}_{pp}(x)\partial_{xx}V(x)+\widehat{H}_p(x)^\top\partial_{xxx}V(x),\\
\bar{J}_2 &:= \bar{\tilde{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2}(\mathrm{tr}(\partial_{\tilde{x}\mu})\partial_{xx}V)(x,\tilde{\xi})+\widehat{H}_p(\tilde{\xi})^\top\partial_{\mu xx}V(x,\tilde{\xi})\\
&\quad +\beta^2(\mathrm{tr}(\partial_{x\mu})\partial_{xx}V)(x,\tilde{\xi})+\frac{\beta^2}{2}(\mathrm{tr}(\partial_{\mu\mu})\partial_{xx}V)(x,\bar{\xi},\tilde{\xi})\Big].
\end{aligned}$$

Evaluating (7.8) along (X_t,μ_t) and then plugging into (7.7), we obtain (5.3). □

Acknowledgements

We would like to thank Gangbo and Mészáros for helpful discussions on Subsection 5.1, concerning the related wellposedness result for standard MFGs under displacement semi-monotonicity conditions.

Chenchen Mou is supported in part by CityU Start-up(Grant No. 7200684) and Hong Kong RGC(Grant No. ECS 9048215). Jianfeng Zhang is supported in part by NSF (Grant Nos. DMS-1908665 and DMS-2205972).

References

[1] Achdou, Y., Han, J., Lasry, J. M., Lions, P. L. and Moll, B., Income and wealth distribution in macroeconomics: a countinuous-time approach, The Review of Economic Studies, 2022, 89(1): 45–86.

[2] Achdou, Y. and Kobeissi, Z., Mean field games of controls: finite difference approximations, Math. Eng., 2021, 3(3): 1–35.

[3] Ahuja, S., Wellposedness of mean field games with common noise under a weak monotonicity condition, SIAM J. Control Optim., 2016, 54(1): 30–48.

[4] Bauso, D., Tembine, H. and Basar, T., Opinion dynamics in social networks through mean-field games, SIAM J. Control Optim., 2016, 54(6): 3225–3257.

[5] Bensoussan, A., Frehse, J. and Yam, S. C. P., Mean Field Games and Mean Field Type Control Theory, Springer Briefs in Mathematics, Springer, New York, 2013.

[6] Bensoussan, A., Graber, P. J. and Yam, S. C. P., Control on Hilbert spaces and application to some mean field type control problems, arXiv: 2005.10770, 2020.

[7] Bogachev, V. I., Measure Theory, Springer, Berlin, Heidelberg, 2007.

[8] Buckdahn, R., Li, J., Peng, S. and Rainer, C., Mean-field stochastic differential equations and associated PDEs, Ann. Probab., 2017, 45(2): 824–878.

[9] Huang, M., Caines, P. E. and Malhamé, R. P., Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, Commun. Inf. Syst., 2006, 6(3): 221–251.

[10] Huang, M., Caines, P. E. and Malhamé, R. P., Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε-Nash equilibria, IEEE Transactions on Automatic Control, 2007, 52(9): 1560–1571.

[11] Huang, M., Caines, P. E. and Malhamé, R. P., The Nash certainty equivalence principle and McKean-Vlasov systems: an invariance principle and entry adaptation, 2007 46th IEEE Conference on Decision and Control, New Orleans, LA, USA, 2007, 121–126.

[12] Cardaliaguet, P., Notes on mean field games, Lectures by Lions, P. L., Collège de France, 2010.

[13] Cardaliaguet, P. and Lehalle, C. A., Mean field game of controls and an application to trade crowding, Math. Financ. Econ., 2018, 12(3): 335–363.

[14] Cardaliaguet, P. and Porretta, A., An introduction to mean field game theory, In: Cardaliaguet, P., Porretta, A. (eds), Mean Field Games, Lecture Notes in Mathematics, vol. 2281, Springer, Cham, 2020.

[15] Cardaliaguet, P., Delarue, F., Lasry, J. M. and Lions, P. L., The Master Equation and the Convergence Problem in Mean Field Games, Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 2019.

[16] Carmona, R. and Delarue, F., Probabilistic Theory of Mean Field Games with Applications I-Mean Field FBSDEs, Control, and Games, Probability Theory and Stochastic Modeling, vol. 83, Springer, Cham, 2018.

[17] Carmona, R. and Delarue, F., Probabilistic Theory of Mean Field Games with Applications II-Mean Field Games with Common Noise and Master Equations, Probability Theory and Stochastic Modeling, vol. 84, Springer, Cham, 2018.

[18] Carmona, R. and Lacker, D., A probabilistic weak formulation of mean field games and applications, Ann. Appl. Probab., 2015, 25(3): 1189–1231.

[19] Chassagneux, J. F., Crisan, D. and Delarue, F., A probabilistic approach to classical solutions of the master equation for large population equilibria, accepted in Mem. Amer. Math. Soc., arXiv: 1411.3009.

[20] Djete, M. F., Large population games with interactions through controls and common noise: convergence results and equivalence between open-loop and closed-loop controls, arXiv: 2108.02992, 2021.

[21] Gangbo, W., Mészáros, A. R., Mou, C. and Zhang, J., Mean field games master equations with non-separable Hamiltonians and displacement monotonicity, accepted in Ann. Probab., arXiv: 2101.12362.

[22] Gelfand, M. J., Explaining the puzzle of human diversity, Science, 2019, 366(6466): 686–687.

[23] Gangbo, W. and Tudorascu, A., On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations, J. Math. Pures Appl., 2019, 125: 119–174.

[24] Gomes, D. A., Patrizi, S. and Voskanyan, V. K., On the existence of classical solutions for stationary extended mean field games, Nonlinear Anal. -Theory Methods Appl., 2014, 99: 49–79.

[25] Gomes, D. A. and Voskanyan, V. K., Extended mean field games, Izv. Nats. Akad. Nauk Armenii Mat., 2013, 48(2): 63–76.

[26] Gomes, D. A. and Voskanyan, V. K., Extended deterministic mean-field games, SIAM J. Control Optim., 2016, 54(2): 1030−1055.

[27] Graber, J. and Mayorga, S., A note on mean field games of controls with state constraints: existence of mild solutions, arXiv: 2109.11655, 2021.

[28] Kobeissi, Z., Mean field games with monotonous interactions through the law of states and controls of the agents, arXiv: 2006.12949, 2020.

[29] Kobeissi, Z., On classical solutions to mean field game system of controls, Commun. Partial Differ. Equ., 2022, 47(3): 453−488.

[30] Lacker, D. and Soret, A., Many-player games of optimal consumption and investment under relative performance criteria, Math. Financ. Econ., 2020, 14(2): 263−281.

[31] Lacker, D. and Zariphopoulou, T., Mean field and N-agent games for optimal investment under relative performance criteria, Math. Financ., 2019, 29(4): 1003−1038.

[32] Lasry, J. M. and Lions, P. L., Mean field games, Jpn. J. Math., 2007, 2(1): 229−260.

[33] Lasry, J. M., Lions, P. L. and Guéant, O., Application of mean field games to growth theory, HAL Id: hal-00348376, 2008.

[34] Lions, P. L., Cours au Collège de France, http://www.college-de-france.fr.

[35] Mou, C. and Zhang, J., Mean field game master equations with anti-monotonicity conditions, arXiv: 2201.10762, 2022.