J. Eur. Math. Soc. 27, 4469–4499 (2025) DOI 10.4171/JEMS/1455

Chenchen Mou · Jianfeng Zhang

Mean field game master equations with anti-monotonicity conditions

Received 25 January 2022; revised 23 November 2022

Abstract. It is well known that the monotonicity condition, either in Lasry–Lions sense or in displacement sense, is crucial for the global well-posedness of mean field game master equations, as well as for the uniqueness of mean field equilibria and solutions to mean field game systems. In the literature, the monotonicity conditions are always taken in a fixed direction. In this paper, we propose a new type of monotonicity condition in the opposite direction, which we call the anti-monotonicity condition, and establish the global well-posedness for mean field game master equations with non-separable Hamiltonians. Our anti-monotonicity condition allows our data to violate both the Lasry–Lions monotonicity and the displacement monotonicity conditions.

Keywords: master equation, mean field games, Lasry–Lions monotonicity, displacement monotonicity, anti-monotonicity.

1. Introduction

In this paper, we consider the following second-order master equation, arising from mean field games with common noise, with terminal condition $V(T,x,\mu) = G(x,\mu)$:

$$\mathscr{L}V(t,x,\mu) := -\partial_t V - \frac{\widehat{\beta}^2}{2}\operatorname{tr}(\partial_{xx}V) + H(x,\mu,\partial_x V) - \mathscr{N}V = 0, \tag{1.1}$$

where

$$\begin{aligned}\mathscr{N}V(t,x,\mu) := \operatorname{tr}\Big(\bar{\widetilde{\mathbb{E}}}\Big[&\frac{\widehat{\beta}^2}{2}\partial_{\tilde{x}}\partial_\mu V(t,x,\mu,\tilde{\xi}) - \partial_\mu V(t,x,\mu,\tilde{\xi})(\partial_p H)^\top(\tilde{\xi},\mu,\partial_x V(t,\tilde{\xi},\mu))\\ &+ \beta^2\partial_x\partial_\mu V(t,x,\mu,\tilde{\xi}) + \frac{\beta^2}{2}\partial_{\mu\mu}V(t,x,\mu,\bar{\xi},\tilde{\xi})\Big]\Big),\\ &(t,x,\mu)\in[0,T)\times\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d).\end{aligned}$$

Chenchen Mou: Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong, P. R. China; chencmou@cityu.edu.hk

Jianfeng Zhang: Department of Mathematics, University of Southern California, 3620 S. Vermont Avenue, KAP 104, Los Angeles, CA 90089-2532, USA; jianfenz@usc.edu

Mathematics Subject Classification 2020: 35R15 (primary); 49N80, 49Q22, 60H30, 91A16, 93E20 (secondary).

Here $\beta \geq 0$ is a constant, $\widehat{\beta}^2 := 1 + \beta^2$, ∂_t, ∂_x, ∂_{xx} are standard temporal and spatial derivatives, ∂_μ, $\partial_{\mu\mu}$ are W_2-Wasserstein derivatives, $\widetilde{\xi}$ and $\bar{\xi}$ are independent random variables with the same law μ and $\bar{\widetilde{\mathbb{E}}}$ is the expectation with respect to their joint law. The theory of mean field games (MFGs, for short), initiated independently by Caines–Huang–Malhamé [33] and Lasry–Lions [40], studies the asymptotic behavior of stochastic differential games with a large number of players interacting in certain symmetric way. We refer to Lions [41], Cardaliaguet [15], Bensoussan–Frehse–Yam [7], Carmona–Delarue [20,21] and Cardaliaguet–Porretta [18] for a comprehensive exposition of the subject. First introduced by Lions [41], the master equation characterizes the value of the MFG, provided there is a unique mean field equilibrium. Roughly speaking, it plays the role of the HJB equation in the stochastic control theory.

The master equation (1.1) admits a unique local (in time) classical solution when the data H and G are sufficiently smooth, see, e.g., Gangbo–Swiech [32], Bensoussan–Yam [10], Mayorga [42], Carmona–Delarue [21] and Cardaliaguet–Cirant–Porretta [16]. In particular, [16] studied the local well-posedness of the master equations not only for MFGs involving homogeneous minor players but also for MFGs with a major player. It is much more challenging to obtain a global classical solution, we refer to Buckdahn–Li–Peng–Rainer [14], Chassagneux–Crisan–Delarue [24], Cardaliaguet–Delarue–Lasry–Lions [17], Carmona–Delarue [21], Gangbo–Meszaros–Mou–Zhang [31] and, in the realm of potential MFGs, Bensoussan–Graber–Yam [8, 9], Gangbo–Meszaros [30]. We also refer to Mou–Zhang [43], Bertucci [12], and Cardaliaguet–Souganidis [19] for global weak solutions which require much weaker regularity on the data, and Bayraktar–Cohen [5], Bertucci–Lasry–Lions [13], Cecchin–Delarue [23], Bertucci [11] for classical or weak solutions of finite state mean field game master equations. All the above global well-posedness results, with the exception [14] that considers linear master equations and thus no control or game is involved, require certain monotonicity condition, which we explain next.

One typical condition, extensively used in the literature [5, 11–13, 17, 19, 21, 24, 43], is the following well-known Lasry–Lions monotonicity condition. For a function $G: \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$,

$$\mathbb{E}[G(\xi_1, \mathcal{L}_{\xi_1}) + G(\xi_2, \mathcal{L}_{\xi_2}) - G(\xi_1, \mathcal{L}_{\xi_2}) - G(\xi_2, \mathcal{L}_{\xi_1})] \geq 0 \tag{1.2}$$

for any square integrable random variables ξ_1, ξ_2. Another type of monotonicity condition, originating in Ahuja [1] and later sparsely used in the literature, see Ahuja–Ren–Yang [2] and [8, 9, 30, 31], is the displacement (or weak) monotonicity,

$$\mathbb{E}[[\partial_x G(\xi_1, \mathcal{L}_{\xi_1}) - \partial_x G(\xi_2, \mathcal{L}_{\xi_2})][\xi_1 - \xi_2]] \geq 0. \tag{1.3}$$

When G is regular enough with bounded $\partial_{xx} G$, $\partial_{x\mu} G$, (1.2) and (1.3) are equivalent to the following inequalities, respectively, for all square integrable random variables ξ, η:

$$\begin{aligned} \widetilde{\mathbb{E}}[\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] &\geq 0, \\ \widetilde{\mathbb{E}}[\langle \partial_{x\mu} G(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] + \mathbb{E}[\langle \partial_{xx} G(\xi, \mathcal{L}_\xi)\eta, \eta\rangle] &\geq 0, \end{aligned} \tag{1.4}$$

where $(\widetilde{\xi}, \widetilde{\eta})$ is an independent copy of (ξ, η). These monotonicity conditions are crucial for the uniqueness of mean field equilibria and the well-posedness of the master equations.

When none of the monotonicity conditions holds, the MFG could have multiple equilibria, see, e.g., Foguen Tchuendom [29], Cecchin–Dai Pra–Fisher–Pelino [22], Bayraktar–Zhang [6]. In this case, one approach is to consider a special type of equilibria, see, e.g., [22], Delarue–Foguen Tchuendom [25], Cecchin–Delarue [23], Bayraktar–Cecchin–Cohen–Delarue [3, 4]. A larger literature is on the possible convergence of the equilibria for the N-player game, which is quite often unique because the corresponding Nash system is non-degenerate due to the presence of the individual noises, to the mean field equilibria (which may or may not be unique), see, e.g., [17, 21, 43], Delarue–Lacker–Ramanan [26, 27], Djete [28], Lacker [35–38], Lacker–Flem [39], Nuts–San Martin–Tan [44]. Finally, we note that Iseri–Zhang [34] takes a quite different approach by investigating the set of game values over all mean field equilibria and establishes the dynamic programming principle and the convergence from the N-player game to the MFG.

We emphasize that the two inequalities in (1.4) share the same direction. Our goal of this paper is to propose a new type of monotonicity condition in the opposite direction, which we call anti-monotonicity condition, and establish the global well-posedness for the master equation (1.1), with possibly non-separable Hamiltonian H. We remark that the mean field equilibrium is a fixed point, and the monotonicity conditions (1.4) were used to ensure the uniqueness of the fixed point. To motivate our anti-monotonicity condition, let us use a very simple example to illustrate the idea. Suppose that $f:\mathbb{R}^1 \to \mathbb{R}^1$ is a continuously differentiable function and we are interested in its fixed point x^*: $f(x^*) = x^*$. When f is decreasing, i.e., $f' \le 0$, clearly f admits a unique fixed point x^*. When f is increasing, in general neither the existence nor the uniqueness of x^* is guaranteed. However, if f is sufficiently monotone, in the sense that $f' \ge 1 + \varepsilon$ for some $\varepsilon > 0$, then again f has a unique fixed point x^*. While in complete different contexts, our conditions follow the same spirit. Roughly speaking, the standard monotonicity conditions (1.4) correspond to the case that f is decreasing, while our new anti-monotonicity condition corresponds to the case f is increasing, and for the same reason we will need to require our data to be sufficiently anti-monotone in appropriate sense.

To be precise, our anti-monotonicity condition takes the following form:

$$\begin{aligned}\widetilde{\mathbb{E}}\big[&\lambda_0\langle \partial_{xx}G(\xi,\mathcal{L}_\xi)\eta,\eta\rangle + \lambda_1\langle \partial_{x\mu}G(\xi,\mathcal{L}_\xi,\widetilde{\xi})\widetilde{\eta},\eta\rangle + |\partial_{xx}G(\xi,\mathcal{L}_\xi)\eta|^2 \\ &+ \lambda_2|\widetilde{\mathbb{E}}_{\mathcal{F}^1_T}[\partial_{x\mu}G(\xi,\mathcal{L}_\xi,\widetilde{\xi})\widetilde{\eta}]|^2 - \lambda_3|\eta|^2\big] \le 0 \end{aligned} \tag{1.5}$$

for some appropriate constants $\lambda_0 > 0$, $\lambda_1 \in \mathbb{R}$, $\lambda_2 > 0$, $\lambda_3 \ge 0$. We remark that the inequality here takes the opposite direction to those in (1.4). In particular, the displacement monotonicity requires the convexity of G in x, while here G is typically concave in x, due to the first term in (1.5). This justifies the name of anti-monotonicity (and to have a better comparison with (1.4), we may also set $\lambda_1 = 1$). We also note that, considering the case $\lambda_3 = 0$, the second line of (1.5) is positive, this means that the first line of (1.5) should be sufficiently negative, which is exactly in the spirit that G to be sufficiently anti-monotone.

To establish the global well-posedness of the master equation (1.1), we follow the strategy in [31], which consists of three steps. The key step of this approach is to show a priori that the anti-monotonicity propagates along the solution V. That is, under appropriate conditions, as long as $V(T,\cdot) = G$ is anti-monotone, then $V(t,\cdot)$ is anti-monotone for all t. The second step is to show that the anti-monotonicity of V implies $\partial_x V$ is uniformly Lipschitz continuous in (x,μ), under W_2 in μ. This, together with a representation formula established in [43], implies further the Lipschitz continuity under W_1. In the final step, we show that the uniform Lipschitz continuity under W_1 enables us to extend a local classical solution to a global one.

There is a major technical difference from [31] though. The assumptions we impose for the propagation of anti-monotonicity prevents us from assuming uniform Lipschitz continuity of the data G and H. Instead, we can only assume $\partial_x G$, $\partial_x H$ are uniformly Lipschitz. This has two consequences. First, the a priori estimate for the boundedness of $\partial_{xx} V$, which is crucial for the global well-posedness of the master equation and is pretty easy to obtain under the conditions in [31], becomes very subtle. In fact, we need some serious efforts to obtain this estimate. Moreover, unlike in [31], under our conditions the solution V will not be Lipschitz continuous. Instead, we can only expect the Lipschitz continuity of $\partial_x V$. Therefore, we will actually consider the vector master equation of $\vec{U} := \partial_x V$ and establish its global well-posedness first. Once we obtain $\vec{U}$, then it is immediate to solve the original master equation (1.1) for V.

The rest of the paper is organized as follows. In Section 2, we review the setting in [31] and introduce our problem. In Section 3, we introduce the new notion of anti-monotonicity and present the technical conditions used in the paper. In Section 4, we show a priori the crucial propagation of the anti-monotonicity. Section 5 is devoted to the a priori uniform Lipschitz estimate of $\partial_x V$ in μ, first under W_2 and then under W_1. In Section 6, we provide the a priori estimate for $\partial_{xx} V$. Finally, in Section 7 we establish the global well-posedness of the master equation (1.1).

2. The setting

Throughout the paper, we will use the setting in [31]. We review it briefly in this section and refer to [31] for more details.

We consider the following product filtered probability space on $[0,T]$:

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

Here, for $\omega = (\omega^0, \omega^1) \in \Omega$, $B^0(\omega) = B^0(\omega^0)$ and $B(\omega) = B(\omega^1)$ are independent d-dimensional Brownian motions; $\mathbb{F}^0 = \{\mathcal{F}_t^0\}$ is generated by B^0, and $\mathbb{F}^1 = \{\mathcal{F}_t^1\}$ is generated by B and $\mathcal{F}_0^1$, where we assume $\mathcal{F}_0^1$ has no atom. Let $(\tilde{\Omega}_1, \tilde{\mathcal{F}}^1, \tilde{B}, \tilde{\mathbb{P}}_1)$ be a copy of the filtered probability space $(\Omega_1, \mathbb{F}^1, B, \mathbb{P}_1)$ and define the larger filtered probability space by

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

Given an $\mathcal{F}_t$-measurable random variable $\xi(\tilde{\omega}) = \varphi(\omega^0, \omega^1)$, $\tilde{\omega} = (\omega^0, \omega^1, \tilde{\omega}^1) \in \tilde{\Omega}$, we see that $\tilde{\xi}(\tilde{\omega}) := \varphi(\omega^0, \tilde{\omega}^1)$ is a conditionally independent copy of ξ, conditional on $\mathcal{F}_t^0$ under $\tilde{\mathbb{P}}$. When two conditionally independent copies are needed, we let $(\bar{\Omega}_1, \bar{\mathbb{F}}^1, \bar{B}, \bar{\mathbb{P}}_1)$ be another copy of $(\Omega_1, \mathbb{F}^1, B, \mathbb{P}_1)$, and enlarge the joint product space further,

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

Throughout the paper, we will use the probability space $(\Omega, \mathbb{F}, \mathbb{P})$. However, when conditionally independent copies of random variables or processes are needed, we will tacitly use their extensions to the larger spaces $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde{\mathbb{E}})$ and $(\bar{\tilde{\Omega}}, \bar{\tilde{\mathbb{F}}}, \bar{\tilde{\mathbb{P}}}, \bar{\tilde{\mathbb{E}}})$ without mentioning.

We next introduce the Wasserstein space and differential calculus on Wasserstein space. Let $\mathcal{P} := \mathcal{P}(\mathbb{R}^d)$ be the set of all probability measures on $\mathbb{R}^d$ and, for any $q \ge 1$, let $\mathcal{P}_q$ denote the set of $\mu \in \mathcal{P}$ with finite q-th moment. For any sub-σ-field $\mathcal{G} \subset \mathcal{F}_T$ and $\mu \in \mathcal{P}_q$, we denote the set of $\mathbb{R}^d$-valued, $\mathcal{G}$-measurable, and q-integrable random variables ξ by $\mathbb{L}^q(\mathcal{G})$; and the set of $\xi \in \mathbb{L}^q(\mathcal{G})$ such that the law $\mathcal{L}_\xi = \mu$ by $\mathbb{L}^q(\mathcal{G}; \mu)$. For any $\mu, \nu \in \mathcal{P}_q$, the W_q-Wasserstein distance between them is defined as follows:

$$W_q(\mu, \nu) := \inf\{(\mathbb{E}[|\xi - \eta|^q])^{\frac{1}{q}} : \text{for all } \xi \in \mathbb{L}^q(\mathcal{F}_T; \mu),\ \eta \in \mathbb{L}^q(\mathcal{F}_T; \nu)\}.$$

For a W_2-continuous function $U\colon \mathcal{P}_2 \to \mathbb{R}$, its Wasserstein gradient, also called Lions-derivative, takes the form

$$\partial_\mu U\colon (\mu, \tilde{x}) \in \mathcal{P}_2 \times \mathbb{R}^d \to \mathbb{R}^d$$

and satisfies

$$U(\mathcal{L}_{\xi+\eta}) - U(\mu) = \mathbb{E}[\langle \partial_\mu U(\mu, \xi), \eta \rangle] + o(\|\eta\|_2) \tag{2.1}$$

for all $\xi \in \mathbb{L}^2(\mathcal{F}_T; \mu)$, $\eta \in \mathbb{L}^2(\mathcal{F}_T)$. Let $\mathcal{C}^0(\mathcal{P}_2)$ denote the set of W_2-continuous functions $U\colon \mathcal{P}_2 \to \mathbb{R}$. For $k = 1, 2$, we introduce $\mathcal{C}^k(\mathcal{P}_2)$, which are referred to as functions of *full* $\mathcal{C}^k$ regularity in [20, Theorem 4.17], as follows. By $\mathcal{C}^1(\mathcal{P}_2)$, we mean the space of functions $U \in \mathcal{C}^0(\mathcal{P}_2)$ such that $\partial_\mu U$ exists and is continuous on $\mathcal{P}_2 \times \mathbb{R}^d$, it is uniquely determined by (2.1). Similarly, $\mathcal{C}^2(\mathcal{P}_2)$ stands for the set of functions $U \in \mathcal{C}^1(\mathcal{P}_2)$ such that $\partial_{\tilde{x}\mu} U$, $\partial_{\mu\mu} U$ exist and are continuous on $\mathcal{P}_2 \times \mathbb{R}^d$ and $\mathcal{P}_2 \times \mathbb{R}^{2d}$, respectively. Let $\mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$ denote the set of continuous functions $U\colon \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$ satisfying $\partial_x U$, $\partial_{xx} U$ exist and are joint continuous on $\mathbb{R}^d \times \mathcal{P}_2$, $\partial_\mu U$, $\partial_{x\mu} U$, $\partial_{\tilde{x}\mu} U$ exist and are continuous on $\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d$, and $\partial_{\mu\mu} U$ exists and is continuous on $\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^{2d}$. Finally, we fix the state space

$$\Theta := [0, T] \times \mathbb{R}^d \times \mathcal{P}_2$$

for our master equation, and let $\mathcal{C}^{1,2}(\Theta)$ denote the set of continuous functions $U \in \Theta \to \mathbb{R}$ which have the following continuous derivatives: $\partial_t U$, $\partial_x U$, $\partial_{xx} U$, $\partial_\mu U$, $\partial_{x\mu} U$, $\partial_{\tilde{x}\mu} U$, $\partial_{\mu\mu} U$.

One crucial property of functions $U \in \mathcal{C}^{1,2}(\Theta)$ is the Itô formula. For $i = 1, 2$, let $dX_t^i := b_t^i dt + \sigma_t^i dB_t + \sigma_t^{i,0} dB_t^0$, where $b^i: [0,T] \times \Omega \to \mathbb{R}^d$ and $\sigma^i, \sigma^{i,0}: [0,T] \times \Omega \to \mathbb{R}^{d\times d}$ are $\mathbb{F}$-progressively measurable and bounded (for simplicity) processes, and $\rho_t := \mathcal{L}_{X_t^2|\mathcal{F}_t^0}$, then we have (cf. [21, Theorem 4.17], [14, 24])

$$\begin{aligned} dU(t, X_t^1, \rho_t) = \Big[&\partial_t U + \partial_x U \cdot b_t^1 + \frac{1}{2}\mathrm{tr}(\partial_{xx} U[\sigma_t^1(\sigma_t^1)^\top + \sigma_t^{1,0}(\sigma_t^{1,0})^\top])\Big](t, X_t^1, \rho_t)dt \\ &+ \partial_x U(t, X_t^1, \rho_t) \cdot \sigma_t^1 dB_t + (\sigma_t^{1,0})^\top \partial_x U(t, X_t^1, \rho_t) \cdot dB_t^0 \\ &+ \mathrm{tr}(\widetilde{\mathbb{E}}_{\mathcal{F}_t}[\partial_\mu U(t, X_t^1, \rho_t, \widetilde{X}_t^2)(\widetilde{b}_t^2)^\top])dt \\ &+ \widetilde{\mathbb{E}}_{\mathcal{F}_t}[(\widetilde{\sigma}_t^{2,0})^\top \partial_\mu U(t, X_t^1, \rho_t, \widetilde{X}_t^2)] \cdot dB_t^0 \\ &+ \mathrm{tr}\Big(\bar{\widetilde{\mathbb{E}}}_{\mathcal{F}_t}\Big[\frac{1}{2}\partial_{\tilde{x}}\partial_\mu U(t, X_t^1, \rho_t, \widetilde{X}_t^2)[\widetilde{\sigma}_t^2(\widetilde{\sigma}_t^2)^\top + \widetilde{\sigma}_t^{2,0}(\widetilde{\sigma}_t^{2,0})^\top] \\ &\qquad + \partial_x \partial_\mu U(t, X_t^1, \rho_t, \widetilde{X}_t^2)\sigma_t^{1,0}(\widetilde{\sigma}_t^{2,0})^\top \\ &\qquad + \frac{1}{2}\partial_{\mu\mu} U(t, X_t^1, \rho_t, \widetilde{X}_t^2, \bar{X}_t^2)\widetilde{\sigma}_t^{2,0}(\bar{\sigma}_t^{2,0})^\top\Big]\Big)dt. \end{aligned} \tag{2.2}$$

Here $\mathcal{L}_{X_t^2|\mathcal{F}_t^0}$ stands for the conditional law of X_t^2 given $\mathcal{F}_t^0$, and $\widetilde{\mathbb{E}}_{\mathcal{F}_t}$ and $\bar{\widetilde{\mathbb{E}}}_{\mathcal{F}_t}$ are the conditional expectations given $\mathcal{F}_t$ corresponding to the probability measures $\widetilde{\mathbb{P}}$ and $\bar{\widetilde{\mathbb{P}}}$, respectively. Throughout the paper, the elements of $\mathbb{R}^d$ are viewed as column vectors; $\partial_x U, \partial_\mu U \in \mathbb{R}^d$ are also column vectors; $\partial_{x\mu} U := \partial_x \partial_\mu U := \partial_x[(\partial_\mu U)^\top] \in \mathbb{R}^{d\times d}$, where $^\top$ denotes the transpose, and similarly for the other second-order derivatives; both the notations "·" and $\langle \cdot, \cdot \rangle$ denote the inner product of column vectors.

We finally introduce the mean field game system related to the master equation (1.1). Given $t_0 \in [0,T]$ and $\xi \in \mathbb{L}^2(\mathcal{F}_{t_0})$, it either takes the form of forward backward McKean–Vlasov SDEs on $[t_0, T]$, denoting $B_t^{t_0} := B_t - B_{t_0}$, $B_t^{0,t_0} := B_t^0 - B_{t_0}^0$,

$$\begin{aligned} X_t^\xi &= \xi - \int_{t_0}^t \partial_p H(X_s^\xi, \rho_s, Z_s^\xi)ds + B_t^{t_0} + \beta B_t^{0,t_0}, \\ Y_t^\xi &= G(X_T^\xi, \rho_T) + \int_t^T \widehat{L}(X_s^\xi, \rho_s, Z_s^\xi)ds - \int_t^T Z_s^\xi \cdot dB_s - \int_t^T Z_s^{0,\xi} \cdot dB_s^0, \end{aligned} \tag{2.3}$$

where

$$\widehat{L}(x, \mu, p) := p \cdot \partial_p H(x, \mu, p) - H(x, \mu, p), \quad \rho_t := \rho_t^\xi := \mathcal{L}_{X_t^\xi|\mathcal{F}_t^0},$$

or takes the form of forward backward stochastic PDE system on $[t_0, T]$,

$$\begin{aligned} d\rho(t,x) &= \Big[\frac{\widehat{\beta}^2}{2}\mathrm{tr}(\partial_{xx}\rho(t,x)) + \mathrm{div}(\rho(t,x)\partial_p H(x, \rho(t,\cdot), \partial_x u(t,x)))\Big]dt \\ &\quad - \beta \partial_x \rho(t,x) \cdot dB_t^0, \\ du(t,x) &= -\Big[\mathrm{tr}\Big(\frac{\widehat{\beta}^2}{2}\partial_{xx} u(t,x) + \beta \partial_x v^\top(t,x)\Big) - H(x, \rho(t,\cdot), \partial_x u(t,x))\Big]dt \\ &\quad + v(t,x) \cdot dB_t^0, \\ \rho(t_0, \cdot) &= \mathcal{L}_\xi, \quad u(T,x) = G(x, \rho(T,\cdot)), \end{aligned} \tag{2.4}$$

where $\widehat{\beta}^2 := 1 + \beta^2$, the solution triple (ρ, u, v) is $\mathbb{F}^0$-progressively measurable and $\rho(t, \cdot, \omega)$ is a (random) probability measure. Systems (2.3) and (2.4) connect to the master equation (1.1) as follows: provided all the equations are well posed and in particular (1.1) has a classical solution V, then

$$\begin{aligned} Y_t^\xi &= V(t, X_t^\xi, \rho_t), \\ Z_t^\xi &= \partial_x V(t, X_t^\xi, \rho_t), \\ u(t, x, \omega) &= V(t, x, \rho(t, \cdot, \omega)). \end{aligned} \tag{2.5}$$

It is already well known that, cf. [21], if the master equation (1.1) has a classical solution V with bounded derivatives, then we can get existence and uniqueness of the mean field equilibrium, and the equilibrium of the corresponding N-player game will converge to the mean field equilibrium. Therefore, we shall only focus on the global well-posedness of the master equation (1.1).

We conclude this section with the strategy from [31] for the global well-posedness of (1.1). We will follow the same strategy in this paper, except that we shall replace the monotonicity condition with the anti-monotonicity condition:

Step 1. Introduce appropriate monotonicity condition on data which ensure the propagation of the monotonicity 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 well-posedness of classical solutions and the above uniform Lipschitz continuity to obtain the global well-posedness of classical solutions.

3. Assumptions and anti-monotonicity conditions

In this section, we introduce the following notations. For any $A \in \mathbb{R}^{d \times d}$,

$$\begin{aligned} \underline{\kappa}(A) &:= \inf_{|x|=1} \langle Ax, x \rangle = \text{the smallest eigenvalue of } \frac{1}{2}[A + A^\top], \\ \overline{\kappa}(A) &:= \sup_{|x|=1} \langle Ax, x \rangle, \\ \underline{\kappa}'(A) &:= \text{the smallest real part of eigenvalues of } A, \\ |A| &:= \sup_{|x|=|y|=1} \langle Ax, y \rangle. \end{aligned} \tag{3.1}$$

It is obvious that, for any $A, A_1, A_2 \in \mathbb{R}^{d \times d}$ and $x \in \mathbb{R}^d$,

$$\begin{aligned} &|\cdot| \text{ is a norm on } \mathbb{R}^{d \times d}, \quad |A_1 A_2| \le |A_1||A_2|, \quad |Ax| \le |A||x|, \\ &\text{and, when } A \text{ is symmetric,} \quad \underline{\kappa}'(A) = \underline{\kappa}(A), \quad |A| = |\underline{\kappa}(A)| \vee |\overline{\kappa}(A)|. \end{aligned} \tag{3.2}$$

3.1. Regularity assumptions

We first specify some technical assumptions on G and H.

Assumption 3.1. (i) $H \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d)$ and there exist constants $\overline{L}^H_{xp}, \overline{L}^H_{xx}, L^H_2 > 0$ such that

$$|\partial_{xp} H| \le \overline{L}^H_{xp}, \quad |\partial_{xx} H| \le \overline{L}^H_{xx}, \quad |\partial_{pp} H|, |\partial_{x\mu} H|, |\partial_{p\mu} H| \le L^H_2. \tag{3.3}$$

(ii) $H \in \mathcal{C}^3(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d)$, and

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

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

Assumption 3.2. (i) $G \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$, and there exist constants $\overline{L}^G_{xx}, L^G_2 > 0$ such that

$$|\partial_{xx} G| \le \overline{L}^G_{xx}, \quad |\partial_{x\mu} G| \le L^G_2. \tag{3.4}$$

(ii) $\partial_x G, \partial_{xx} G \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$, and $\partial_\mu G, \partial_{x\mu} G \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d)$, and all the second- and higher-order derivatives of G involved here are uniformly bounded.

Here the spaces $\mathcal{C}^2$, $\mathcal{C}^3$ are defined in the same manner as $\mathcal{C}^{1,2}(\Theta)$. Note that at above we do not require the first-order derivatives to be uniformly bounded. In fact, condition (3.14) below does not allow $\partial_x H$ to be bounded.

Remark 3.3. Under Assumption 3.2 (i), we see that $\partial_x G$ is uniformly Lipschitz continuous in μ under W_1 on $\mathbb{R}^d \times \mathcal{P}_2$ with Lipschitz constant L^G_2. This implies further the Lipschitz continuity of $\partial_x G$ in μ under W_2 on $\mathbb{R}^d \times \mathcal{P}_2$, and we denote the Lipschitz constant by $\widetilde{L}^G_2 \le L^G_2$,

$$\widetilde{\mathbb{E}}[[\partial_{x\mu} G(x, \mu, \widetilde{\xi})\widetilde{\eta}]] \le \widetilde{L}^G_2 (\mathbb{E}[|\eta|^2])^{\frac{1}{2}} \quad \forall\, \xi \in \mathbb{L}^2(\mathcal{F}^1_T, \mu),\ \eta \in \mathbb{L}^2(\mathcal{F}^1_T).$$

3.2. Monotonicity and anti-monotonicity conditions

Under the above regularity conditions on the data G and H, the MFG may still have multiple mean field equilibria over a long time duration and thus the global well-posedness of classical solutions for the master equations can fail. Therefore, some structural conditions on G, H are needed in order to guarantee its global well-posedness. The typical structural conditions assumed in the literature are two types of monotonicity conditions, i.e., the Lasry–Lions monotonicity condition and the displacement monotonicity condition.

Definition 3.4. Let $U : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$ be such that $U \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$.

(i) U is called *Lasry–Lions monotone* if for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}^1_T)$,

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] \ge 0. \tag{3.5}$$

(ii) U is called *displacement monotone* if for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1)$,

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle + \langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle] \geq 0. \tag{3.6}$$

(iii) U is called *displacement semi-monotone* if for some $\lambda \in \mathbb{R}$ and for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1)$,

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle + \langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle] - \lambda \mathbb{E}[|\eta|^2] \geq 0. \tag{3.7}$$

Here, as in Section 2, $(\widetilde{\xi}, \widetilde{\eta})$ is an independent copy of (ξ, η). We remark that the displacement semi-monotonicity is obviously weaker than the displacement monotonicity (3.6), and when $\partial_{xx} U$ is bounded, it is also weaker than the Lasry–Lions monotonicity (3.5).

Remark 3.5. The above formulations of the monotonicity conditions are convenient for our purpose. For $U \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$, (3.5) and (3.6) are equivalent to (1.2) and (1.3), respectively, which appear more often in the literature. See [31, Remark 2.4].

We next turn to the monotonicity conditions for the Hamiltonian H. In the literature, the Lasry–Lions monotonicity has only been proposed for the separable Hamiltonians, i.e., $H(x, \mu, p) = H_0(x, p) - F(x, \mu)$ and F satisfies (1.2). In [31], a notion of displacement monotonicity for non-separable H was proposed to study the well-posedness of the master equation (1.1).

Definition 3.6. Let H be a Hamiltonian satisfying Assumption 3.1 (i) and H be strictly convex in p. We say that H is displacement monotone if for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1)$ and any bounded Lipschitz continuous function $\varphi \in C^1(\mathbb{R}^d; \mathbb{R}^d)$,

$$\begin{aligned}\widetilde{\mathbb{E}}\Big[&\langle \partial_{x\mu} H(\xi, \mathcal{L}_\xi, \widetilde{\xi}, \varphi(\xi))\widetilde{\eta} + \partial_{xx} H(\xi, \mathcal{L}_\xi, \varphi(\xi))\eta, \eta\rangle \\ &+ \frac{1}{4}|(\partial_{pp} H(\xi, \mathcal{L}_\xi, \varphi(\xi)))^{-\frac{1}{2}} \widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{p\mu} H(\xi, \mathcal{L}_\xi, \widetilde{\xi}, \varphi(\xi))\widetilde{\eta}]|^2\Big] \leq 0. \end{aligned} \tag{3.8}$$

Remark 3.7. (i) The above definition of displacement monotonicity for non-separable Hamiltonians is not really used in the rest of the paper except for the comparison with the new notion of anti-monotonicity introduced below. We refer to [31, Proposition 3.7] for another equivalent definition of the above one.

(ii) The function $\varphi(\xi)$ in (3.8) is chosen to be $\partial_x V(t, \xi, \mathcal{L}_\xi)$ in the proof of the propagation of the displacement monotonicity (3.6) along $V(t, \cdot)$ in [31]. Since $\partial_x V$ is not known priorly, the displacement monotonicity (3.8) is made for any desirable function φ.

(iii) When H is non-separable, it still remains a challenge to find appropriate conditions on H so that the Lasry–Lions monotonicity (3.5) could propagate along the solution $V(t, \cdot)$.

Finally, we introduce the anti-monotonicity condition, which is the main structural condition in this paper and serves as an alternative sufficient condition for the global well-posedness of the master equation. Denote

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

Definition 3.8. Let $U \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$ and $\vec{\lambda} \in D_4$. We say U is $\vec{\lambda}$-anti-monotone if

$$\begin{aligned}(\text{AntiMon})_{\xi}^{\vec{\lambda}} U(\eta,\eta) := \widetilde{\mathbb{E}}\big[&\lambda_0 \langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \lambda_1 \langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle \\ &+ |\partial_{xx} U(\xi, \mathcal{L}_\xi)\eta|^2 + \lambda_2 |\widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}]|^2 \\ &- \lambda_3 |\eta|^2\big] \leq 0 \quad \forall \xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1).\end{aligned} \tag{3.10}$$

Remark 3.9. (i) The main feature of (3.10) is that the direction of the inequality is opposite to those in Definition 3.4. In particular, (3.10) implies the Lasry–Lions anti-monotonicity,

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] \leq 0, \tag{3.11}$$

for the case that $\lambda_0 = \lambda_3 = 0$ and $\lambda_1 = \lambda_2 = 1$. In fact, in this case condition (3.10) is stronger than (3.11), and we interpret it as U is sufficiently Lasry–Lions anti-monotone,

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] \leq -\widetilde{\mathbb{E}}[|\partial_{xx} U(\xi, \mathcal{L}_\xi)\eta|^2 + |\widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}]|^2] \leq 0.$$

Similarly, in the case $\lambda_0 = \lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 0$, we see that (3.10) implies U is sufficiently displacement anti-monotone,

$$\begin{aligned}&\widetilde{\mathbb{E}}[\langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] \\ &\quad \leq -\widetilde{\mathbb{E}}[|\partial_{xx} U(\xi, \mathcal{L}_\xi)\eta|^2 + |\widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}]|^2] \leq 0.\end{aligned} \tag{3.12}$$

Note that the concavity of U in x could help in (3.12), while in (3.6) its convexity is helpful.

(ii) Inequality (3.10) implies the displacement semi-anti-monotonicity, i.e.,

$$\widetilde{\mathbb{E}}[\langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] \leq \lambda_3 \widetilde{\mathbb{E}}[|\eta|^2] \tag{3.13}$$

if $\vec{\lambda} \in D_4$, $\lambda_0 = \lambda_1 = 1$ and $\lambda_3 \geq 0$. Note that condition (3.13) is weaker than (3.12) for the case. We recall that in the literature a function $u : \mathbb{R}^d \to \mathbb{R}$ is said to be semi-concave, or λ-concave, if $\partial_{xx} u \leq \lambda I_d$ for some constant $\lambda > 0$, where I_d is the $d \times d$ identity matrix. We follow the same spirit to call U $\vec{\lambda}$-anti-monotone if U satisfies (3.10).

We next provide an example which is $\vec{\lambda}$-anti-monotone.

Example 3.10. Let $d = 1$ and consider the function for some constants a_0, a_1,

$$U(x, \mu) = \frac{a_0}{2}|x|^2 + a_1 x \int_{\mathbb{R}} y \mu(dy), \quad (x, \mu) \in \mathbb{R} \times \mathcal{P}_2.$$

It is clear that $\partial_{xx} U = a_0$ and $\partial_{x\mu} U = a_1$.

(i) For any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_T^1)$, we have

$$\widetilde{\mathbb{E}}[\langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] = a_1 |\mathbb{E}[\eta]|^2.$$

So U is Lasry–Lions monotone if $a_1 \geq 0$, and Lasry–Lions anti-monotone if $a_1 \leq 0$.

(ii) Similarly, we have

$$\widetilde{\mathbb{E}}[\langle \partial_{xx} U(\xi, \mathcal{L}_\xi)\eta, \eta\rangle + \langle \partial_{x\mu} U(\xi, \mathcal{L}_\xi, \widetilde{\xi})\widetilde{\eta}, \eta\rangle] = a_0 \mathbb{E}[|\eta|^2] + a_1 |\mathbb{E}[\eta]|^2.$$

Then one can easily check that U is displacement monotone if $a_0 \geq 0$, $a_1 \geq -a_0$, and displacement anti-monotone if $a_0 \leq 0$, $a_1 \leq -a_0$.

(iii) For any $\vec{\lambda} \in D_4$, we have

$$(\mathrm{AntiMon})_\xi^{\vec{\lambda}} U(\eta, \eta) := [\lambda_0 a_0 + |a_0|^2 - \lambda_3]\mathbb{E}[|\eta|^2] + [\lambda_1 a_1 + \lambda_2 |a_1|^2] |\mathbb{E}[\eta]|^2.$$

Then U is $\vec{\lambda}$-anti-monotone if

$$\lambda_0 a_0 + |a_0|^2 - \lambda_3 \leq 0, \quad \lambda_0 a_0 + |a_0|^2 - \lambda_3 \leq -[\lambda_1 a_1 + \lambda_2 |a_1|^2],$$

which is equivalent to

$$\lambda_3 \geq \max(\lambda_0 a_0 + |a_0|^2, \lambda_0 a_0 + |a_0|^2 + \lambda_1 a_1 + \lambda_2 |a_1|^2).$$

In particular, if we set $\lambda_0 = \lambda_1 = \lambda_2 = 1$, $\lambda_3 = 0$, and $-1 \leq a_0, a_1 \leq 0$, we see that U is $\vec{\lambda}$-anti-monotone.

Remark 3.11. Let $U \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2)$ and $\vec{\lambda} \in D_4$.

(i) When $\lambda_0 = 0$, then for all $\xi_1, \xi_2 \in \mathbb{L}^2(\mathcal{F}_T^1)$, (3.10) is equivalent to the following integral form:

$$\begin{aligned}
&\lambda_1 \mathbb{E}[U(\xi_1, \mathcal{L}_{\xi_1}) + U(\xi_2, \mathcal{L}_{\xi_2}) - U(\xi_1, \mathcal{L}_{\xi_2}) - U(\xi_2, \mathcal{L}_{\xi_1})] \\
&\quad + \mathbb{E}[|\partial_x U(\xi_1, \mathcal{L}_{\xi_2}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2 + \lambda_2 |\partial_x U(\xi_2, \mathcal{L}_{\xi_1}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2] \\
&\quad \leq \lambda_3 \mathbb{E}[|\xi_1 - \xi_2|^2] + o(\mathbb{E}[|\xi_1 - \xi_2|^2]).
\end{aligned}$$

Here $o(\varepsilon)$ means it vanishes faster than ε as $\varepsilon \to 0$.

(ii) When $\lambda_0 = \lambda_1$, then for any $\xi_1, \xi_2 \in \mathbb{L}^2(\mathcal{F}_T^1)$, (3.10) is equivalent to the following form:

$$\begin{aligned}
&\lambda_0 \mathbb{E}[\langle \partial_x U(\xi_1, \mathcal{L}_{\xi_1}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2}), \xi_1 - \xi_2\rangle] \\
&\quad + \mathbb{E}[|\partial_x U(\xi_1, \mathcal{L}_{\xi_2}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2 + \lambda_2 |\partial_x U(\xi_2, \mathcal{L}_{\xi_1}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2] \\
&\quad \leq C \mathbb{E}[|\xi_1 - \xi_2|^2] + o(\mathbb{E}[|\xi_1 - \xi_2|^2]).
\end{aligned}$$

(iii) In general, for any $\xi_1, \xi_2 \in \mathbb{L}^2(\mathcal{F}_T^1)$, (3.10) is equivalent to the following form:

$$\begin{aligned}
&\mathbb{E}\big[\lambda_0 \langle \partial_x U(\xi_1, \mathcal{L}_{\xi_2}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2}), \xi_1 - \xi_2\rangle \\
&\quad + \lambda_1 \langle \partial_x U(\xi_2, \mathcal{L}_{\xi_1}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2}), \xi_1 - \xi_2\rangle\big] \\
&\quad + \mathbb{E}[|\partial_x U(\xi_1, \mathcal{L}_{\xi_2}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2 + \lambda_2 |\partial_x U(\xi_2, \mathcal{L}_{\xi_1}) - \partial_x U(\xi_2, \mathcal{L}_{\xi_2})|^2] \\
&\quad \leq \lambda_3 \mathbb{E}[|\xi_1 - \xi_2|^2] + o(\mathbb{E}[|\xi_1 - \xi_2|^2]).
\end{aligned}$$

Assumption 3.12. (i) G satisfies Assumption 3.2 (i) and is $\vec{\lambda}$-anti-monotone for some $\vec{\lambda} \in D_4$;

(ii) H satisfies Assumption 3.1 (i) and there exist constants $\underline{L}^H_{xp} > 0$, $\underline{L}^H_{xx} > 0$, $\overline{\gamma} > \underline{\gamma} > 0$ such that

$$\underline{\kappa}(\partial_{xp} H) \geq \underline{L}^H_{xp}, \quad \underline{\kappa}(\partial_{xx} H) \geq \underline{L}^H_{xx}, \tag{3.14}$$

and

$$\underline{\gamma}\underline{L}^H_{xp} \leq \underline{L}^H_{xx} \leq \overline{L}^H_{xx} \leq \overline{\gamma}\underline{L}^H_{xp}, \quad \overline{L}^H_{xp} \leq \overline{\gamma}\underline{L}^H_{xp}. \tag{3.15}$$

Note that no structural conditions are required for $\partial_{x\mu} H$ here, and $\partial_{pp} H$ can be degenerate.

4. Propagation of anti-monotonicity

In this section, we show that any classical solution V to the master equation (1.1) could propagate the anti-monotonicity under appropriate conditions.

Theorem 4.1. *Let Assumption 3.12 hold and let V be a classical solution of the master equation (1.1) such that*

$$\partial_{xx} V(t,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2), \quad \partial_{x\mu} V(t,\cdot,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d),$$

and all the second- and higher-order derivatives of V involved above are also continuous in the time variable and are uniformly bounded. Assume further that there exist a constant $L^V_{xx} > 0$ such that

$$|\partial_{xx} V| \leq L^V_{xx}, \tag{4.1}$$

and a constant

$$\lambda_0 > \frac{\overline{\gamma}^2[1 + L^V_{xx}]^2 - 8\lambda_3}{4\underline{\gamma}} \quad \text{such that } \theta_1 := \frac{\overline{\gamma}[1 + L^V_{xx}]}{\sqrt{4(\underline{\gamma}\lambda_0 + 2\lambda_3)}} < 1. \tag{4.2}$$

Introduce the following symmetric matrices, which depend only on $\underline{\gamma}$, $\overline{\gamma}$, $\vec{\lambda}$, and L^V_{xx}:

$$A_1 := \begin{bmatrix} 4[1-\theta_1] & 0 & 0 \\ 0 & 2\lambda_2 & 0 \\ 0 & 0 & [1-\theta_1][\lambda_0\underline{\gamma} + 2\lambda_3] \end{bmatrix}, \tag{4.3}$$

$$A_2 := \begin{bmatrix} \lambda_0 & \lambda_0 & |\lambda_0 - \frac{1}{2}\lambda_1| + \lambda_3 \\ \lambda_0 & |\lambda_1| & \frac{1}{2}|\lambda_1| + \lambda_2 + \lambda_3 \\ |\lambda_0 - \frac{1}{2}\lambda_1| + \lambda_3 & \frac{1}{2}|\lambda_1| + \lambda_2 + \lambda_3 & |\lambda_1| + 2\lambda_3 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 \\ 1 & \lambda_2 & \lambda_2 \\ 1 & \lambda_2 & 0 \end{bmatrix} L^V_{xx}.$$

Then, whenever

$$\underline{L}^H_{xp} \geq \underline{\kappa}(A_1^{-1}A_2) L^H_2, \tag{4.4}$$

$V(t,\cdot)$ is $\vec{\lambda}$-anti-monotone in the sense of (3.10) for all $t \in [0,T]$.

Proof. Without loss of generality, we shall prove the theorem only for $t_0 = 0$.

Fix $\xi \in \mathbb{L}^2(\mathscr{F}_0)$ and $\eta \in \mathbb{L}^2(\mathscr{F}_0)$. Given the desired regularity of V and H, the following system of McKean–Vlasov SDEs has a unique solution $(X, \delta X)$:

$$
\begin{aligned}
X_t &= \xi - \int_0^t \partial_p H(X_s, \mu_s, \partial_x V(s, X_s, \mu_s))ds + B_t + \beta B_t^0, \quad \mu_t := \mathscr{L}_{X_t|\mathscr{F}_t^0}, \\
\delta X_t &= \eta - \int_0^t \big[\partial_{px} H(X_s, \mu_s, \partial_x V(s, X_s, \mu_s))\delta X_s \\
&\qquad + \tilde{\mathbb{E}}_{\mathscr{F}_s}[\partial_{p\mu} H(X_s, \mu_s, \partial_x V(s, X_s, \mu_s), \tilde{X}_s)\delta \tilde{X}_s] \\
&\qquad + \partial_{pp} H(X_s, \mu_s, \partial_x V(s, X_s, \mu_s))[\Upsilon_s + \bar{\Upsilon}_s]\big]ds,
\end{aligned}
\tag{4.5}
$$

where

$$
\begin{aligned}
\Upsilon_t &:= \tilde{\mathbb{E}}_{\mathscr{F}_t}[\partial_{x\mu} V(t, X_t, \mu_t, \tilde{X}_t)\delta \tilde{X}_t], \\
\bar{\Upsilon}_t &:= \partial_{xx} V(t, X_t, \mu_t)\delta X_t.
\end{aligned}
$$

In the sequel, for simplicity of notation, we omit the variables (t, μ_t) as well as the dependence on $\partial_x V$, and denote

$$
\begin{aligned}
H_p(X_t) &:= \partial_p H(X_t, \mu_t, \partial_x V(t, X_t, \mu_t)), \\
H_{p\mu}(X_t, \tilde{X}_t) &:= \partial_{p\mu} H(X_t, \mu_t, \tilde{X}_t, \partial_x V(t, X_t, \mu_t)),
\end{aligned}
$$

and similarly for H_{xp}, H_{pp}, $H_{x\mu}$, $\partial_{xx} V$, $\partial_{x\mu} V$, etc. We remark that, $(\tilde{X}_t, \delta \tilde{X}_t)$ is a conditionally independent copy of $(X_t, \delta X_t)$ and μ_t is $\mathscr{F}_t^0$-measurable.

Recall (4.5) and introduce

$$
\begin{aligned}
I_t &:= \mathbb{E}[\langle \Upsilon_t, \delta X_t\rangle], \quad \bar{I}_t := \mathbb{E}[\langle \bar{\Upsilon}_t, \delta X_t\rangle], \\
\Gamma_t &:= (\text{AntiMon})_{X_t}^{\vec{\lambda}} V(t, \cdot)(\delta X_t, \delta X_t) \\
&= \lambda_0 \bar{I}_t + \lambda_1 I_t + \mathbb{E}[|\bar{\Upsilon}_t|^2 + \lambda_2 |\Upsilon_t|^2 - \lambda_3 |\delta X_t|^2].
\end{aligned}
$$

By the calculation in [31, Theorem 4.1], we have

$$
\begin{aligned}
\frac{d}{dt} I(t) &= \mathbb{E}\big[-\langle H_{pp}(X_t)\Upsilon_t, \Upsilon_t\rangle - \langle \tilde{\mathbb{E}}_{\mathscr{F}_t}[H_{p\mu}(X_t, \tilde{X}_t)\delta \tilde{X}_t], \Upsilon_t - \bar{\Upsilon}_t\rangle \\
&\qquad + \langle \tilde{\mathbb{E}}_{\mathscr{F}_t}[H_{x\mu}(X_t, \tilde{X}_t)\delta \tilde{X}_t], \delta X_t\rangle\big], \\
\frac{d}{dt} \bar{I}(t) &= \mathbb{E}\big[-\langle H_{pp}(X_t)\bar{\Upsilon}_t, \bar{\Upsilon}_t\rangle - 2\langle H_{pp}(X_t)\bar{\Upsilon}_t, \Upsilon_t\rangle \\
&\qquad - 2\langle \bar{\Upsilon}_t, \tilde{\mathbb{E}}_{\mathscr{F}_t}[H_{p\mu}(X_t, \tilde{X}_t)\delta \tilde{X}_t]\rangle + \langle H_{xx}(X_t)\delta X_t, \delta X_t\rangle\big],
\end{aligned}
\tag{4.6}
$$

and, by the calculation in [31, Theorem 5.1], we have

$$
\begin{aligned}
d\Upsilon_t &= (dB_t)^\top K_1(t) + \beta(dB_t^0)^\top K_2(t) + [K_3(t)\Upsilon_t + K_4(t)]dt, \\
d\bar{\Upsilon}_t &= (dB_t)^\top \bar{K}_1(t) + \beta(dB_t^0)^\top \bar{K}_2(t) \\
&\qquad + [2H_{xp}(X_t)\bar{\Upsilon}_t - \partial_{xx} V(X_t) H_{pp}(X_t)\Upsilon_t + \bar{K}_3(t)]dt,
\end{aligned}
\tag{4.7}
$$

where ($K_5(t)$ and $K_6(t)$ in [31] turn to $K_3(t)$ and $K_4(t)$, respectively, here)

$$\begin{aligned}
K_1(t) &:= \widetilde{\mathbb{E}}_{\mathcal{F}_t}[\partial_{xx\mu}V(X_t,\widetilde{X}_t)\delta\widetilde{X}_t],\\
K_2(t) &:= K_1(t)+\bar{\widetilde{\mathbb{E}}}_{\mathcal{F}_t}[[(\partial_{\mu x\mu}V)(X_t,\bar{X}_t,\widetilde{X}_t)+\partial_{\tilde{x}x\mu}V(X_t,\widetilde{X}_t)]\delta\widetilde{X}_t],\\
K_3(t) &:= H_{xp}(X_t)+\partial_{xx}V(X_t)H_{pp}(X_t),\\
K_4(t) &:= \widetilde{\mathbb{E}}_{\mathcal{F}_t}[[H_{x\mu}(X_t,\widetilde{X}_t)+\partial_{xx}V(X_t)H_{p\mu}(X_t,\widetilde{X}_t)]\delta\widetilde{X}_t],\\
\bar{K}_1(t) &:= \partial_{xxx}V(X_t)\delta X_t,\\
\bar{K}_2(t) &:= \bar{K}_1(t)+\widetilde{\mathbb{E}}_{\mathcal{F}_t}[(\partial_{\mu xx}V)(X_t,\widetilde{X}_t)\delta\widetilde{X}_t],\\
\bar{K}_3(t) &:= [H_{xx}(X_t)-\partial_{xx}V(X_t)H_{px}(X_t)]\delta X_t-\partial_{xx}V(X_t)\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t].
\end{aligned}$$

In particular, this implies that

$$\begin{aligned}
&\frac{d}{dt}\mathbb{E}[|\Upsilon_t|^2]\geq 2\mathbb{E}[\langle\Upsilon_t,K_3(t)\Upsilon_t+K_4(t)\rangle],\\
&\frac{d}{dt}\mathbb{E}[|\bar{\Upsilon}_t|^2]\geq 2\mathbb{E}[\langle\bar{\Upsilon}_t,2H_{xp}(X_t)\bar{\Upsilon}_t-\partial_{xx}V(X_t)H_{pp}(X_t)\Upsilon_t+\bar{K}_3(t)\rangle].
\end{aligned}\tag{4.8}$$

Moreover, by (4.5) we have

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}[|\delta X_t|^2]=-2\mathbb{E}[\langle &H_{px}(X_t)\delta X_t+\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]\\
&+H_{pp}(X_t)[\Upsilon_t+\bar{\Upsilon}_t],\delta X_t\rangle].
\end{aligned}\tag{4.9}$$

Thus, by (4.6), (4.8), and (4.9), we have

$$\begin{aligned}
\frac{d}{dt}\Gamma_t \geq{}& \lambda_0\mathbb{E}\big[-\langle H_{pp}(X_t)\bar{\Upsilon}_t,\bar{\Upsilon}_t\rangle-2\langle H_{pp}(X_t)\bar{\Upsilon}_t,\Upsilon_t\rangle\\
&\qquad-2\langle\bar{\Upsilon}_t,\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]\rangle+\langle H_{xx}(X_t)\delta X_t,\delta X_t\rangle\big]\\
&+\lambda_1\mathbb{E}\big[-\langle H_{pp}(X_t)\Upsilon_t,\Upsilon_t\rangle-\langle\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t],\Upsilon_t-\bar{\Upsilon}_t\rangle\\
&\qquad+\langle\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{x\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t],\delta X_t\rangle\big]\\
&+2\mathbb{E}\big[\langle\bar{\Upsilon}_t,[2H_{xp}(X_t)\bar{\Upsilon}_t-\partial_{xx}V(X_t)H_{pp}(X_t)\Upsilon_t+\bar{K}_3(t)]\rangle\\
&\qquad+\lambda_2\langle\Upsilon_t,[K_3(t)\Upsilon_t+K_4(t)]\rangle\big]\\
&+2\lambda_3\mathbb{E}[\langle H_{px}(X_t)\delta X_t+\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]+H_{pp}(X_t)[\Upsilon_t+\bar{\Upsilon}_t],\delta X_t\rangle]\\
={}&\mathbb{E}\big[\langle[-\lambda_0H_{pp}(X_t)+4H_{xp}(X_t)]\bar{\Upsilon}_t,\bar{\Upsilon}_t\rangle\\
&\quad+\langle[-\lambda_1H_{pp}(X_t)+2\lambda_2K_3(t)]\Upsilon_t,\Upsilon_t\rangle\\
&\quad+\langle[\lambda_0H_{xx}(X_t)+2\lambda_3H_{px}(X_t)]\delta X_t,\delta X_t\rangle\\
&\quad+\langle\lambda_1\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{x\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]+2\lambda_3\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t],\delta X_t\rangle\\
&\quad-\langle2[\lambda_0H_{pp}(X_t)+\partial_{xx}V(X_t)H_{pp}(X_t)]\Upsilon_t,\bar{\Upsilon}_t\rangle\\
&\quad+\langle[-2\lambda_0+\lambda_1]\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]+2\bar{K}_3(t)+2\lambda_3H_{pp}(X_t)\delta X_t,\bar{\Upsilon}_t\rangle\\
&\quad+\langle-\lambda_1\widetilde{\mathbb{E}}_{\mathcal{F}_t}[H_{p\mu}(X_t,\widetilde{X}_t)\delta\widetilde{X}_t]+2\lambda_2K_4(t)+2\lambda_3H_{pp}(X_t)\delta X_t,\Upsilon_t\rangle\big].
\end{aligned}$$

Next, by Assumptions 3.1 (i) and 3.12 (ii), and (3.15) we have

$$\begin{aligned}\frac{d}{dt}\Gamma_t \geq\ & [4\underline{L}^H_{xp} - \lambda_0 L^H_2]\mathbb{E}[|\overline{\Upsilon}_t|^2] + [2\lambda_2\underline{L}^H_{xp} - [|\lambda_1| + \lambda_2 L^V_{xx}]L^H_2]\mathbb{E}[|\Upsilon_t|^2]\\ & + [\lambda_0\underline{L}^H_{xx} + 2\lambda_3\underline{L}^H_{xp} - [|\lambda_1| + 2\lambda_3]L^H_2]\mathbb{E}[|\delta X_t|^2]\\ & - 2L^H_2[\lambda_0 + L^V_{xx}]\mathbb{E}[|\Upsilon_t||\overline{\Upsilon}_t|]\\ & - [|\lambda_1 - 2\lambda_0|L^H_2 + 2\overline{\gamma}[1 + L^V_{xx}]\underline{L}^H_{xp} + 2L^V_{xx}L^H_2 + 2\lambda_3 L^H_2]\\ & \times (\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}(\mathbb{E}[|\overline{\Upsilon}_t|^2])^{\frac{1}{2}}\\ & - L^H_2[|\lambda_1| + 2\lambda_2[1 + L^V_{xx}] + 2\lambda_3](\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}(\mathbb{E}[|\Upsilon_t|^2])^{\frac{1}{2}}.\end{aligned}$$

Note that, recalling the θ_1 in (4.2),

$$4\theta_1\mathbb{E}[|\overline{\Upsilon}_t|^2] + 2\overline{\gamma}[1 + L^V_{xx}](\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}(\mathbb{E}[|\overline{\Upsilon}_t|^2])^{\frac{1}{2}} + \theta_1[\lambda_0\underline{\gamma} + 2\lambda_3]\mathbb{E}[|\delta X_t|^2] \geq 0.$$

Then, recalling (4.3) and denoting $a := [(\mathbb{E}[|\overline{\Upsilon}_t|^2])^{\frac{1}{2}}, (\mathbb{E}[|\Upsilon_t|^2])^{\frac{1}{2}}, (\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}]$,

$$\begin{aligned}\frac{d}{dt}\Gamma_t \geq\ & [4[1-\theta_1]\underline{L}^H_{xp} - \lambda_0 L^H_2]\mathbb{E}[|\overline{\Upsilon}_t|^2] + [2\lambda_2\underline{L}^H_{xp} - [|\lambda_1| + \lambda_2 L^V_{xx}]L^H_2]\mathbb{E}[|\Upsilon_t|^2]\\ & + [[1-\theta_1][\lambda_0\underline{\gamma} + 2\lambda_3]\underline{L}^H_{xp} - [|\lambda_1| + 2\lambda_3]L^H_2]\mathbb{E}[|\delta X_t|^2]\\ & - 2L^H_2[\lambda_0 + L^V_{xx}]\mathbb{E}[|\Upsilon_t||\overline{\Upsilon}_t|]\\ & - L^H_2[|\lambda_1 - 2\lambda_0| + 2L^V_{xx} + 2\lambda_3](\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}(\mathbb{E}[|\overline{\Upsilon}_t|^2])^{\frac{1}{2}}\\ & - L^H_2[|\lambda_1| + 2\lambda_2[1 + L^V_{xx}] + 2\lambda_3](\mathbb{E}[|\delta X_t|^2])^{\frac{1}{2}}(\mathbb{E}[|\Upsilon_t|^2])^{\frac{1}{2}}\\ =\ & a[A_1\underline{L}^H_{xp} - A_2 L^H_2]a^\top \geq 0,\end{aligned}$$

where the last inequality thanks to (4.4) and the fact that $A_1 \geq 0$. Thus

$$(\text{AntiMon})^{\vec{\lambda}}_{\xi}V(0,\eta,\eta) = \Gamma_0 \leq \Gamma_T = (\text{AntiMon})^{\vec{\lambda}}_{X_T}G(\delta X_T, \delta X_T) \leq 0.$$

That is, $V(0,\cdot,\cdot)$ is $\vec{\lambda}$-anti-monotone. ■

5. The Lipschitz continuity

We first show that the anti-monotonicity of V implies the uniformly Lipschitz continuity of $\partial_x V$ in μ under W_2. Unlike in [31], since we do not require the first-order derivatives of G, H to be bounded, here we do not expect the Lipschitz continuity of V itself.

Theorem 5.1. *Let Assumptions* 3.1 (i), 3.2 (i) *hold and V be a classical solution of the master equation* (1.1) *such that*

$$\partial_{xx}V(t,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2), \quad \partial_{x\mu}V(t,\cdot,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d),$$

and all the second- and higher-order derivatives of V involved above are also continuous in the time variable and are uniformly bounded. Assume further that $V(t,\cdot,\cdot)$ is

$\vec{\lambda}$-anti-monotone in the sense of (3.10) *for all $t \in [0,T]$. Then $\partial_x V$ is uniformly Lipschitz continuous in μ under W_2, with a Lipschitz constant C_2^μ depending only on $\vec{\lambda}$, the parameters in* (3.3) *and* (3.4)*, and L_{xx}^V.*

Proof. In this proof, $C > 0$ denotes a generic constant depending only on quantities mentioned in the statement of the theorem. As in the proof of Theorem 4.1, without loss of generality we show the theorem only for $t_0 = 0$. First, by (3.10) we have, for any $\xi, \eta \in \mathbb{L}^2(\mathcal{F}_t^1)$,

$$\mathbb{E}[|\widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{x\mu} V(t,\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta}]|^2] \le C|\widetilde{\mathbb{E}}[\langle \partial_{x\mu} V(t,\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta},\eta\rangle]| + C\mathbb{E}[|\eta|^2]. \quad (5.1)$$

Next, applying Hölder's inequality to (5.1) we have

$$\mathbb{E}[|\widetilde{\mathbb{E}}_{\mathcal{F}_T^1}[\partial_{x\mu} V(t,\xi,\mathcal{L}_\xi,\tilde{\xi})\tilde{\eta}]|^2] \le C\mathbb{E}[|\eta|^2]. \quad (5.2)$$

From now on, we fix $\xi \in \mathbb{L}^2(\mathcal{F}_0)$ and $\eta \in \mathbb{L}^2(\mathcal{F}_0)$ and continue to use the notation as in the proof of Theorem 4.1. In particular, $X, \delta X, \mu_t, \Upsilon, \bar{\Upsilon}$ are defined by (4.5). Applying (5.2) by replacing $\mathbb{E}$ by $\mathbb{E}_{\mathcal{F}_t^0}$ and noting that X_t is $\mathcal{F}_t$-measurable, we have

$$\begin{aligned}\mathbb{E}[|\Upsilon_t|^2] &= \mathbb{E}[\mathbb{E}_{\mathcal{F}_t^0}[|\widetilde{\mathbb{E}}_{\mathcal{F}_T}[\partial_{x\mu} V(t,X_t,\mu_t,\tilde{X}_t)\delta\tilde{X}_t]|^2]] \\ &\le C\mathbb{E}[\mathbb{E}_{\mathcal{F}_t^0}[|\delta X_t|^2]] \le C\mathbb{E}[|\delta X_t|^2]. \end{aligned} \quad (5.3)$$

Using Hölder's inequality on (4.5) and noting in particular

$$|\bar{\Upsilon}_t| \le L_{xx}^V |\delta X_t|,$$

we obtain

$$|\delta X_t|^2 \le 2|\eta|^2 + C\int_0^t [|\delta X_s|^2 + |\widetilde{\mathbb{E}}_{\mathcal{F}_s}[|\delta\tilde{X}_s|]|^2 + |\Upsilon_s|^2]ds. \quad (5.4)$$

Taking expectation on (5.4) and using (5.3), we derive

$$\mathbb{E}[|\delta X_t|^2] \le 2\mathbb{E}[|\eta|^2] + C\int_0^t \mathbb{E}[|\delta X_s|^2]ds.$$

Then it follows from Grönwall's inequality that

$$\sup_{t\in[0,T]} \mathbb{E}[|\delta X_t|^2] \le C\mathbb{E}[|\eta|^2]. \quad (5.5)$$

Next, by (4.7), we have

$$\Upsilon_t = \Upsilon_T - \int_t^T [K_3(s)\Upsilon_s + K_4(s)]ds - \int_t^T (dB_s)^\top K_1(s) - \beta\int_t^T (dB_s^0)^\top K_2(s).$$

Taking conditional expectation $\widetilde{\mathbb{E}}_{\mathcal{F}_t}$, we have

$$\Upsilon_t = \widetilde{\mathbb{E}}_{\mathcal{F}_t}[\partial_{x\mu} G(X_T,\mu_T,\tilde{X}_T)\delta\tilde{X}_T] - \int_t^T \widetilde{\mathbb{E}}_{\mathcal{F}_t}[K_3(s)\Upsilon_s + K_4(s)]ds. \quad (5.6)$$

Then by (5.6) and the required regularity of G, H and V, we have

$$|\Upsilon_t|^2 \le C\widetilde{\mathbb{E}}_{\mathcal{F}_t}[|\delta\widetilde{X}_T|^2] + C\int_t^T \widetilde{\mathbb{E}}_{\mathcal{F}_t}[|\Upsilon_s|^2 + |\delta\widetilde{X}_s|^2]ds.$$

Now taking conditional expectation $\widetilde{\mathbb{E}}_{\mathcal{F}_0}$, we get

$$\widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\Upsilon_t|^2] \le C\widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\delta\widetilde{X}_T|^2] + C\int_t^T \widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\Upsilon_s|^2 + |\delta\widetilde{X}_s|^2]ds.$$

Thus, by Grönwall's inequality we have

$$|\Upsilon_0|^2 = \widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\Upsilon_0|^2] \le C\widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\delta\widetilde{X}_T|^2] + C\int_0^T \widetilde{\mathbb{E}}_{\mathcal{F}_0}[|\delta\widetilde{X}_s|^2]ds. \tag{5.7}$$

Note that, recalling the setting in Section 2, $\delta\widetilde{X}_t$ is measurable with respect to $\mathcal{F}_t^0 \vee \widetilde{\mathcal{F}}_t^1$, which is independent of $\mathcal{F}_0$ under $\widetilde{\mathbb{P}}$. Then the conditional expectation in the right-hand side of (5.7) is actually an expectation. Plug (5.5) into (5.7), we obtain

$$|\widetilde{\mathbb{E}}_{\mathcal{F}_0}[\partial_{x\mu}V(0,\xi,\mu_0,\widetilde{\xi})\widetilde{\eta}]|^2 = |\Upsilon_0|^2 \le C\mathbb{E}[|\eta|^2].$$

This implies

$$|\widetilde{\mathbb{E}}[\partial_{x\mu}V(0,x,\mu_0,\widetilde{\xi})\widetilde{\eta}]| \le C(\mathbb{E}|\eta|^2)^{\frac{1}{2}}, \quad \mu_0 - \text{a.e. } x. \tag{5.8}$$

Since $\partial_\mu V$ is continuous, then (5.8) actually holds for all x. In particular, this implies that there exists a constant $C_2^{\mu_0} > 0$ such that

$$\begin{aligned}|\partial_x V(0,x,\mathcal{L}_{\xi+\eta}) - \partial_x V(0,x,\mathcal{L}_\xi)| &= \left|\int_0^1 \mathbb{E}[\partial_{x\mu}V(0,x,\mathcal{L}_{\xi+\theta\eta},\xi+\theta\eta)\eta]d\theta\right| \\ &\le C_2^{\mu_0}(\mathbb{E}[|\eta|^2])^{\frac{1}{2}}.\end{aligned}$$

Now, taking random variables ξ, η such that

$$W_2^2(\mathcal{L}_{\xi+\eta},\mathcal{L}_\xi) = \mathbb{E}[|\eta|^2],$$

the above inequality exactly means that $\partial_x V(0,x,\cdot)$ is uniformly Lipschitz continuous in μ_0 under W_2 with uniform Lipschitz constant $C_2^{\mu_0}$. ■

We emphasize that the above Lipschitz continuity is under W_2, while the global well-posedness of the master equation requires the W_1-Lipschitz continuity. As in [31], we shall derive the desired W_1-Lipschitz continuity from the W_2-Lipschitz continuity by utilizing the pointwise representation for the Wasserstein derivative developed in [43]. Note again that in Theorem 5.1 we only have the Lipschitz continuity for $\partial_x V$, but not for V, so at below we shall also consider

$$\vec{U}(t,x,\mu) := \partial_x V(t,x,\mu),$$

which formally should satisfy the following vectorial master equation on $[0, T) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, with terminal condition $\vec{U}(T, x, \mu) = \partial_x G(x, \mu)$:

$$-\partial_t \vec{U} - \frac{\widehat{\beta}^2}{2}\mathrm{tr}(\partial_{xx}\vec{U}) + \partial_x H(x, \mu, \vec{U}) + \partial_p H(x, \mu, \vec{U}) \cdot \partial_x \vec{U} - \mathcal{N}\vec{U} = 0, \quad (5.9)$$

where

$$\mathcal{N}\vec{U}(t, x, \mu) := \mathrm{tr}\Big(\bar{\widetilde{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2}\partial_{\tilde{x}}\partial_\mu \vec{U}(t, x, \mu, \tilde{\xi}) - \partial_\mu \vec{U}(t, x, \mu, \tilde{\xi})(\partial_p H)^\top(\tilde{\xi}, \mu, \vec{U}(t, \tilde{\xi}, \mu)) + \beta^2 \partial_x \partial_\mu \vec{U}(t, x, \mu, \tilde{\xi}) + \frac{\beta^2}{2}\partial_{\mu\mu}\vec{U}(t, x, \mu, \bar{\xi}, \tilde{\xi})\Big]\Big).$$

To be precise, fix t_0, ξ, we first consider the following McKean–Vlasov SDE on $[t_0, T]$:

$$\begin{aligned} X_t^\xi &= \xi - \int_{t_0}^t \partial_p H(X_s^\xi, \rho_s, \nabla Y_s^\xi) ds + B_t^{t_0} + \beta B_t^{0,t_0}, \quad \rho_t := \rho_t^\xi := \mathcal{L}_{X_t^\xi | \mathcal{F}_t^0}, \\ \nabla Y_t^\xi &= \partial_x G(X_T^\xi, \rho_T) - \int_t^T \partial_x H(X_s^\xi, \rho_s, \nabla Y_s^\xi) ds - \int_t^T \nabla Z_s^\xi \cdot dB_s \\ &\quad - \int_t^T \nabla Z_s^{0,\xi} \cdot dB_s^0. \end{aligned} \quad (5.10)$$

Next, given ρ as above, for fixed $x \in \mathbb{R}^d$ and letting $(e_1, \ldots, e_d)$ denote the natural basis of $\mathbb{R}^d$, we introduce a series of FBSDEs, possibly McKean–Vlasov type,

$$\begin{cases} X_t^{\xi,x} = x - \int_{t_0}^t \partial_p H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) ds + B_t^{t_0} + \beta B_t^{0,t_0}, \\ \nabla Y_t^{\xi,x} = \partial_x G(X_T^{\xi,x}, \rho_T) - \int_t^T \partial_x H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) ds \\ \qquad - \int_t^T \nabla Z_s^{\xi,x} \cdot dB_s - \int_t^T \nabla Z_s^{0,\xi,x} \cdot dB_s^0; \end{cases} \quad (5.11)$$

$$\begin{cases} \nabla_k X_t^{\xi,x} = e_k - \int_{t_0}^t \big[(\nabla_k X_s^{\xi,x})^\top \partial_{xp} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) \\ \qquad + (\nabla_k^2 Y_s^{\xi,x})^\top \partial_{pp} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x})\big] ds, \\ \nabla_k^2 Y_t^{\xi,x} = (\nabla_k X_T^{\xi,x})^\top \partial_{xx} G(X_T^{\xi,x}, \rho_T) \\ \qquad - \int_t^T \big[(\nabla_k X_s^{\xi,x})^\top \partial_{xx} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) \\ \qquad + (\nabla_k^2 Y_s^{\xi,x})^\top \partial_{px} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x})\big] ds \\ \qquad - \int_t^T \nabla_k^2 Z_s^{\xi,x} \cdot dB_s^{t_0} - \int_t^T \nabla_k^2 Z_s^{0,\xi,x} \cdot dB_s^{0,t_0}; \end{cases} \quad (5.12)$$

$$
\begin{cases}
\nabla_k \mathcal{X}_t^{\xi,x} = -\int_{t_0}^t \big[(\nabla_k \mathcal{X}_s^{\xi,x})^\top \partial_{xp} H(X_s^\xi, \rho_s, \nabla Y_s^\xi) \\
\qquad + (\nabla_k^2 \mathcal{Y}_s^{\xi,x})^\top \partial_{pp} H(X_s^\xi, \rho_s, \nabla Y_s^\xi) \\
\qquad + \widetilde{\mathbb{E}}_{\mathcal{F}_s}[(\nabla_k \widetilde{X}_s^{\xi,x})^\top (\partial_{\mu p} H)(X_s^\xi, \rho_s, \widetilde{X}_s^{\xi,x}, \nabla Y_s^\xi) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_s^{\xi,x})^\top \partial_{\mu p} H(X_s^\xi, \rho_s, \widetilde{X}_s^\xi, \nabla Y_s^\xi)]\big] ds, \\
\nabla_k^2 \mathcal{Y}_t^{\xi,x} = \widetilde{\mathbb{E}}_{\mathcal{F}_T}\big[(\nabla_k \widetilde{X}_T^{\xi,x})^\top \partial_{\mu x} G(X_T^\xi, \rho_T, \widetilde{X}_T^{\xi,x}) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_T^{\xi,x})^\top \partial_{\mu x} G(X_T^\xi, \rho_T, \widetilde{X}_T^\xi)\big] \\
\qquad + (\nabla_k \mathcal{X}_T^{\xi,x})^\top \partial_{xx} G(X_T^\xi, \rho_T) \\
\qquad - \int_t^T \nabla_k^2 \mathcal{Z}_s^{\xi,x} \cdot dB_s^{t_0} - \int_t^T \nabla_k^2 \mathcal{Z}_s^{0,\xi,x} \cdot dB_s^{0,t_0} \\
\qquad - \int_t^T \big[(\nabla_k \mathcal{X}_s^{\xi,x})^\top \partial_{xx} H(X_s^\xi, \rho_s, \nabla Y_s^\xi) \\
\qquad + (\nabla_k^2 \mathcal{Y}_s^{\xi,x})^\top \partial_{px} H(X_s^\xi, \rho_s, \nabla Y_s^\xi) \\
\qquad + \widetilde{\mathbb{E}}_{\mathcal{F}_s}[(\nabla_k \widetilde{X}_s^{\xi,x})^\top \partial_{\mu x} H(X_s^\xi, \rho_s, \widetilde{X}_s^{\xi,x}, \nabla Y_s^\xi) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_s^{\xi,x})^\top \partial_{\mu x} H(X_s^\xi, \rho_s, \widetilde{X}_s^\xi, \nabla Y_s^\xi)]\big] ds;
\end{cases}
\tag{5.13}
$$

$$
\begin{cases}
\nabla_{\mu_k} X_t^{x,\xi,\tilde{x}} = -\int_{t_0}^t \big[\widetilde{\mathbb{E}}_{\mathcal{F}_s}\big[(\nabla_k \widetilde{X}_s^{\xi,\tilde{x}})^\top \partial_{\mu p} H(X_s^{\xi,x}, \rho_s, \widetilde{X}_s^{\xi,\tilde{x}}, \nabla Y_s^{\xi,x}) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_s^{\xi,\tilde{x}})^\top \partial_{\mu p} H(X_s^{\xi,x}, \rho_s, \widetilde{X}_s^\xi, \nabla Y_s^{\xi,x})\big] \\
\qquad + (\nabla_{\mu_k} X_s^{x,\xi,\tilde{x}})^\top \partial_{xp} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) \\
\qquad + (\nabla_{\mu_k}^2 Y_s^{x,\xi,\tilde{x}})^\top \partial_{pp} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x})\big] ds, \\
\nabla_{\mu_k}^2 Y_t^{x,\xi,\tilde{x}} = \widetilde{\mathbb{E}}_{\mathcal{F}_T}\big[(\nabla_k \widetilde{X}_T^{\xi,\tilde{x}})^\top \partial_{\mu x} G(X_T^{\xi,x}, \rho_T, \widetilde{X}_T^{\xi,\tilde{x}}) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_T^{\xi,\tilde{x}})^\top \partial_{\mu x} G(X_T^{\xi,x}, \rho_T, \widetilde{X}_T^\xi)\big] \\
\qquad + (\nabla_{\mu_k} X_T^{x,\xi,\tilde{x}})^\top \partial_{xx} G(X_T^{\xi,x}, \rho_T) \\
\qquad - \int_t^T \nabla_{\mu_k}^2 Z_s^{x,\xi,\tilde{x}} \cdot dB_s - \int_t^T \nabla_{\mu_k}^2 Z_s^{0,x,\xi,\tilde{x}} \cdot dB_s^0 \\
\qquad - \int_t^T \big[(\nabla_{\mu_k} X_s^{x,\xi,\tilde{x}})^\top \partial_{xx} H(X_s^{\xi,x}, \rho_s, \nabla Y_s^{\xi,x}) \\
\qquad + (\nabla_{\mu_k}^2 Y_s^{x,\xi,\tilde{x}})^\top \partial_{px} H(X_s^x, \rho_s, \nabla Y_s^{\xi,x}) \\
\qquad + \widetilde{\mathbb{E}}_{\mathcal{F}_s}\big[(\nabla_k \widetilde{X}_s^{\xi,\tilde{x}})^\top \partial_{\mu x} H(X_s^{\xi,x}, \rho_s, \widetilde{X}_s^{\xi,\tilde{x}}, \nabla Y_s^{\xi,x}) \\
\qquad + (\nabla_k \widetilde{\mathcal{X}}_s^{\xi,\tilde{x}})^\top \partial_{\mu x} H(X_s^{\xi,x}, \rho_s, \widetilde{X}_s^\xi, \nabla Y_s^{\xi,x})\big]\big] ds.
\end{cases}
\tag{5.14}
$$

The following local (in time) result provides the crucial W_1-Lipschitz continuity of $\vec{U}$.

Proposition 5.2. *Let Assumptions 3.1* (i) *and 3.2* (i) *hold. Recall the constants $\overline{L}^H_{xx}$, $\overline{L}^H_{xp}$, L^H_2 in (3.3), L^G_2, $\overline{L}^G_{xx}$ in (3.4), and $\widetilde{L}^G_2$ in Remark 3.3. Then there exists $\delta > 0$, depending only on d, $\overline{L}^H_{xx}$, $\overline{L}^H_{xp}$, L^H_2, $\overline{L}^G_{xx}$, $\widetilde{L}^G_2$, such that whenever $T - t_0 \le \delta$, the following hold.*

(i) *The McKean–Vlasov FBSDEs (5.10), (5.11), (5.12), (5.13), and (5.14) are well posed on $[t_0, T]$ for any $\mu \in \mathcal{P}_2$ and $\xi \in \mathbb{L}^2(\mathcal{F}_{t_0}, \mu)$.*

(ii) *Define $\vec{U}(t_0, x, \mu) := \nabla Y^{x,\xi}_{t_0}$. Then we have the pointwise representation*

$$\partial_{\mu_k} \vec{U}(t_0, x, \mu, \tilde{x}) = \nabla^2_{\mu_k} Y^{x,\xi,\tilde{x}}_{t_0}. \tag{5.15}$$

Moreover, there exists a constant $C^\mu_1 > 0$, depending only on d, $\overline{L}^H_{xx}$, $\overline{L}^H_{xp}$, L^H_2, L^G_2, $\overline{L}^G_{xx}$ such that

$$|\partial_\mu \vec{U}(0, x, \mu, \tilde{x})| \le C^\mu_1. \tag{5.16}$$

(iii) *Assume further that Assumptions 3.1* (ii) *and 3.2* (ii) *hold true. Then the vectorial master equation (5.9) has a unique classical solution $\vec{U}$. Moreover,*

$$\vec{U}(t, \cdot, \cdot),\ \ \partial_x \vec{U}(t, \cdot, \cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2),\ \ \partial_\mu \vec{U}(t, \cdot, \cdot, \cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d),$$

and all their derivatives in the state and probability measure variables are continuous in the time variable and are uniformly bounded.

(iv) *The following decoupled McKean–Vlasov FBSDE*

$$\begin{aligned} X^x_t &= x + B^{t_0}_t + \beta B^{0,t_0}_t, \\ Y^{x,\xi}_t &= G(X^x_T, \rho_T) - \int_t^T H(X^x_s, \rho_s, \vec{U}(s, X^x_s, \rho_s)) ds \\ &\quad - \int_t^T Z^{x,\xi}_s \cdot dB_s - \int_t^T Z^{0,x,\xi}_s \cdot dB^0_s \end{aligned} \tag{5.17}$$

is well posed on $[t_0, T]$ for any $x \in \mathbb{R}^d$. Define $V(t_0, x, \mu) := Y^{x,\xi}_{t_0}$. Then V is the unique classical solution of the master equation (1.1) and $\partial_x V = \vec{U}$ on $[0, T] \times \mathbb{R}^d \times \mathcal{P}_2$.

We emphasize that at above C^μ_1 depends on L^G_2 in (3.4), but δ depends only on $\widetilde{L}^G_2$ in Remark 3.3, not on L^G_2.

Proof. The proof of (i)–(iii) is very lengthy, but essentially identical to that of [31, Proposition 6.2], except that [31] considers both $\partial_\mu V$ and $\partial_{x\mu} V = \partial_\mu \vec{U}$. So we omit it here.

(iv) By the smoothness of $\vec{U}$ obtained in (iii), clearly the V defined in (iv) is smooth and $Y^{x,\xi}_t = V(t, X^x_t, \rho_t)$. By applying Itô's formula (2.2), we see that V satisfies the PDE

$$\begin{aligned} &- \partial_t V - \frac{\widehat{\beta}^2}{2} \mathrm{tr}(\partial_{xx} V) + H(x, \mu, \vec{U}) - \mathrm{tr}\Big(\bar{\tilde{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2} \partial_{\tilde{x}} \partial_\mu V(t, x, \mu, \tilde{\xi}) \\ &\quad + \frac{\beta^2}{2} \partial_{\mu\mu} V(t, x, \mu, \bar{\xi}, \tilde{\xi}) - \partial_\mu V(t, x, \mu, \tilde{\xi}) (\partial_p H)^\top (\tilde{\xi}, \mu, \vec{U}(t, \tilde{\xi}, \mu)) \\ &\quad + \beta^2 \partial_x \partial_\mu V(t, x, \mu, \tilde{\xi})\Big]\Big) = 0. \end{aligned} \tag{5.18}$$

Differentiate it with respect to x, we obtain the PDE for $\vec{U}' := \partial_x V$:

$$\begin{aligned}
&-\partial_t \vec{U}' - \frac{\widehat{\beta}^2}{2}\operatorname{tr}(\partial_{xx}\vec{U}') + \partial_x H(x,\mu,\vec{U}) + \partial_p H(x,\mu,\vec{U})\cdot\partial_x\vec{U} \\
&\quad - \operatorname{tr}\Big(\bar{\widetilde{\mathbb{E}}}\Big[\frac{\widehat{\beta}^2}{2}\partial_{\tilde{x}}\partial_\mu \vec{U}'(t,x,\mu,\tilde{\xi}) + \frac{\beta^2}{2}\partial_{\mu\mu}\vec{U}'(t,x,\mu,\bar{\xi},\tilde{\xi}) \\
&\qquad - \partial_\mu \vec{U}'(t,x,\mu,\tilde{\xi})(\partial_p H)^\top(\tilde{\xi},\mu,\vec{U}(t,\tilde{\xi},\mu)) \\
&\qquad + \beta^2\partial_x\partial_\mu \vec{U}'(t,x,\mu,\tilde{\xi})\Big]\Big) = 0. \qquad (5.19)
\end{aligned}$$

Compare this with (5.9), we see that $\vec{U}$ also satisfies (5.19). Thus, by the uniqueness we have

$$\vec{U} = \vec{U}' = \partial_x V.$$

Plug this into (5.18), we verify that V satisfies (1.1). ■

6. Uniform estimates of $\partial_{xx} V$

We note that all the above results rely on the bound L^V_{xx} of $\partial_{xx}V$ in (4.1). In particular, in Theorem 4.1 the $\underline{L}^H_{xp}$ depends on L^V_{xx}. Then it is crucial to obtain an a priori uniform estimate of L^V_{xx} which is independent of $\underline{L}^H_{xp}$. Recall (2.5), we have $\partial_{xx}V = \partial_{xx}u$, so it suffices to establish the a priori estimate for the solution u to the backward SPDE in (2.4), for an arbitrarily given ρ (not necessarily satisfying the forward SPDE in (2.4)).

For this purpose, we consider a special form of H.

Assumption 6.1. The Hamiltonian H takes the following form:

$$H(x,\mu,p) = \langle A_0 x, p\rangle + H_0(x,\mu,p), \qquad (6.1)$$

where $A_0 \in \mathbb{R}^{d\times d}$ is a constant matrix and $H_0 : \mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d \to \mathbb{R}$ is a function satisfying

(i) $H_0 \in \mathcal{C}^2(\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d)$ and there exist constants $\underline{L}^{H_0}_{xx}, \overline{L}^{H_0}_{xx}, L^{H_0}_2 > 0$ such that

$$\underline{\kappa}(\partial_{xx}H_0) \ge \underline{L}^{H_0}_{xx}, \quad |\partial_{xx}H_0| \le \overline{L}^{H_0}_{xx}, \qquad (6.2)$$

$$|\partial_{xp}H_0|, |\partial_{pp}H_0|, |\partial_{x\mu}H_0|, |\partial_{p\mu}H_0| \le L^{H_0}_2 \quad \text{on } \mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d. \qquad (6.3)$$

(ii) H_0 satisfies Assumption 3.1 (ii).

Given A_0, consider its Jordan decomposition

$$A_0 = Q_0 J_0 Q_0^{-1}, \qquad (6.4)$$

where $J_0 \in \mathbb{C}^{d\times d}$ is the Jordan normal form of A_0 and $Q_0 \in \mathbb{C}^{d\times d}$ is invertible. Let $\bar{Q}_0$ denote the conjugate of Q_0 and thus $Q_0\bar{Q}_0^\top$ is positive definite. The following estimate will be crucial.

Lemma 6.2. *Recall* (3.1). *For any* $t \geq 0$, *we have*

$$|e^{-A_0 t}| \leq \sqrt{L^{A_0}} e^{[1-\underline{\kappa}'(A_0)]t}, \quad \textit{where } L^{A_0} := \inf_{Q_0} \frac{\overline{\kappa}(Q_0 \bar{Q}_0^\top)}{\underline{\kappa}(Q_0 \bar{Q}_0^\top)}. \tag{6.5}$$

Here the infimum is over all Q_0 *satisfying* (6.4).

Proof. Fix J_0, Q_0 as in (6.4). It is obvious that $e^{-A_0 t} = Q_0 e^{-J_0 t} Q_0^{-1}$. We claim that

$$|\langle e^{-J_0 t} x, y\rangle| \leq e^{[1-\underline{\kappa}'(A_0)]t} |x||y| \quad \forall x, y \in \mathbb{C}^d. \tag{6.6}$$

Then, for any $x, y \in \mathbb{R}^d$ with $|x| = |y| = 1$, we have

$$\begin{aligned}
|\langle e^{-A_0 t} x, y\rangle| &= |\langle e^{-J_0 t} Q_0^{-1} x, Q_0 \top y\rangle| \leq e^{[1-\underline{\kappa}'(A_0)]t} |Q_0^{-1} x||Q_0 \top y| \\
&\leq e^{[1-\underline{\kappa}'(A_0)]t} \sqrt{\overline{\kappa}(Q_0^{-1}(\bar{Q}_0^\top)^{-1})} \sqrt{\overline{\kappa}(Q_0 \bar{Q}_0^\top)} \\
&= e^{[1-\underline{\kappa}'(A_0)]t} \sqrt{\frac{\overline{\kappa}(Q_0 \bar{Q}_0^\top)}{\underline{\kappa}(Q_0 \bar{Q}_0^\top)}}.
\end{aligned}$$

Since Q_0 is arbitrary, this implies (6.5) immediately.

To see (6.6), assume the Jordan normal form $J_0 = \text{diag}(J_1, \ldots, J_k)$. Here $d_1 + \cdots + d_k = d$; $J_i = \lambda_i I_{d_i} + U_{d_i} \in \mathbb{R}^{d_i \times d_i}$, $i = 1, \ldots, k$; $\lambda_1, \ldots, \lambda_k$ are all the eigenvalues of A_0; and U_{d_i} is the matrix whose $(j, j+1)$-component is 1, $j = 1, \ldots, d_i - 1$, and all other components are 0. It is straightforward to see that

$$e^{-J_0 t} = \text{diag}(e^{-J_1 t}, \ldots, e^{-J_k t}).$$

Note that, for each i, since I_{d_i} and U_{d_i} can commute, and $U_{d_i}^{d_i} = 0$,

$$e^{-J_i t} = e^{-\lambda_i t} e^{-U_{d_i} t} = e^{-\lambda_i t} \sum_{n=0}^{d_i - 1} \frac{(-t)^n}{n!} U_{d_i}^n.$$

For any $x^{(i)}, y^{(i)} \in \mathbb{C}^{d_i}$, it is clear that

$$|\langle U_{d_i}^n x^{(i)}, y^{(i)}\rangle| \leq \frac{1}{2}[|x^{(i)}|^2 + |y^{(i)}|^2].$$

Then, for $x = (x^{(1)}, \ldots, x^{(k)})$, $y = (y^{(1)}, \ldots, y^{(k)}) \in \mathbb{C}^d$ with $|x| = |y| = 1$, we have

$$\begin{aligned}
|\langle e^{-J_0 t} x, y\rangle| &= \left| \sum_{i=1}^k \langle e^{-J_i t} x^{(i)}, y^{(i)}\rangle \right| \leq \sum_{i=1}^k |e^{-\lambda_i t}| \sum_{n=0}^{d_i - 1} \frac{t^n}{n!} |\langle U_{d_i}^n x^{(i)}, y^{(i)}\rangle| \\
&\leq e^{-\underline{\kappa}'(A_0)t} \sum_{i=1}^k \sum_{n=0}^{d_i - 1} \frac{t^n}{n!} \frac{1}{2}[|x^{(i)}|^2 + |y^{(i)}|^2] \leq e^{-\underline{\kappa}'(A_0)t} \sum_{n=0}^{d-1} \frac{t^n}{n!}.
\end{aligned}$$

This implies (6.6) immediately. ∎

Remark 6.3. (i) Form (6.1) is assumed for estimate (6.5) and for the property

$$de^{-A_0 t} = -e^{-A_0 t} A_0 dt = -A_0 e^{-A_0 t} dt, \tag{6.7}$$

required in the proof of Theorem 6.4 below. In general, $e^{-\int_0^t \partial_{xp} H ds}$ does not enjoy these properties. When $d = 1$, however, $e^{-\int_0^t \partial_{xp} H ds}$ obviously satisfies similar properties and thus we do not need the special form (6.1). Moreover, we remark that any alternative structures which could ensure a uniform a priori bound for $\partial_{xx} u$ can serve our purpose.

(ii) It is clear that, under (6.1), (6.2), and (6.3), we may set

$$\begin{aligned} &\underline{L}^H_{xp} := \underline{\kappa}(A_0) - L^{H_0}_2, \quad \overline{L}^H_{xp} := |A_0| + L^{H_0}_2;\\ &\underline{L}^H_{xx} := \underline{L}^{H_0}_{xx}, \quad \overline{L}^H_{xx} := \overline{L}^{H_0}_{xx}, \quad L^H_2 := L^{H_0}_2. \end{aligned} \tag{6.8}$$

Then (3.3) and (3.14) hold true. We shall remark though that the term $\underline{\kappa}(A_0)$ and the condition $\underline{\kappa}(\partial_{xx} H_0) \geq \underline{L}^{H_0}_{xx}$ are not used in Theorem 6.4 below.

(iii) When A_0 is symmetric, one can easily see that $L^{A_0} = 1$, and in this case (6.5) can be improved: $|e^{-A_0 t}| \leq e^{-\underline{\kappa}'(A_0) t}$.

Then we have the following uniform a priori estimate.

Theorem 6.4. *Let Assumptions 3.2 (i) and 6.1 hold and let ρ: $[0, T] \times \Omega \to \mathscr{P}_2$ be $\mathbb{F}^0$-progressively measurable with*

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\int_{\mathbb{R}^d} |x|^2 \rho_t(dx)\Big] < +\infty.$$

Assume (u, v) is a classical solution to the backward SPDE in (2.4) for the given ρ here (ρ is not necessarily a solution to the forward SPDE in (2.4)) such that $\partial_{xx} u$ is bounded and, for some fixed constant $\overline{L}^A \geq 1$,

$$L^{A_0} \leq \overline{L}^A, \quad \underline{\kappa}'(A_0) \geq \theta_2 := \max\Big\{\theta_3, \frac{\overline{L}^{H_0}_{xx}}{2\overline{L}^G_{xx}} + 1\Big\}, \tag{6.9}$$

where

$$\theta_3 := 1 + L^{H_0}_2 \overline{L}^A [1 + \overline{L}^G_{xx} \overline{L}^A + \sqrt{(1 + \overline{L}^G_{xx} \overline{L}^A)^2 - 1}].$$

Then the following estimate holds: for any $\theta \geq \theta_3$,

$$|\partial_{xx} u(t, x)| \leq L^u_{xx}(\theta_3) \quad \forall (t, x), \tag{6.10}$$

where

$$L^u_{xx}(\theta) := \frac{\theta - 1 - L^{H_0}_2 \overline{L}^A - \sqrt{(\theta - 1 - L^{H_0}_2 \overline{L}^A)^2 - 2L^{H_0}_2 \overline{L}^G_{xx} (\overline{L}^A)^2 [\theta - 1]}}{L^{H_0}_2 \overline{L}^A}.$$

We note that (6.9) implies $L^u_{xx}(\theta)$ is well defined for $\theta \geq \theta_3$, and we emphasize that the bound $L^u_{xx}(\theta_3)$ depends only on $L^{H_0}_2$, $\overline{L}^G_{xx}$ and $\overline{L}^A$, in particular not on T, $\underline{\kappa}'(A_0)$, or $\overline{L}^{H_0}_{xx}$.

Proof. Fix (t_0, x). First, under our conditions it is clear that the following FBSDE on $[t_0, T]$ has a unique solution $(X^x, \nabla Y^x, \nabla Z^x, \nabla Z^{0,x})$:

$$\begin{aligned} X_t^x &= x - \int_{t_0}^t \partial_p H(X_s^x, \rho_s, \nabla Y_s^x)ds + B_t^{t_0} + \beta B_t^{0,t_0}, \\ \nabla Y_t^x &= \partial_x G(X_T^x, \rho_T) - \int_t^T \partial_x H(X_s^x, \rho_s, \nabla Y_s^x)ds - \int_t^T \nabla Z_s^x \cdot dB_s \\ &\quad - \int_t^T \nabla Z_s^{0,x} \cdot dB_s^0. \end{aligned} \tag{6.11}$$

In particular, $\partial_x u$ serves as the decoupling field:

$$\nabla Y_t^x = \partial_x u(t, X_t^x), \quad t \in [t_0, T]. \tag{6.12}$$

Next, denote $L_0 := L_{xx}^u(\underline{\kappa}'(A_0))$, and consider the following BSDE on $[t_0, T]$:

$$\begin{aligned} \nabla^2 Y_t^x &= \partial_{xx} G(X_T^x, \rho_T) - \int_t^T \nabla^2 Z_s^x \cdot dB_s - \int_t^T \nabla^2 Z_s^{0,x} \cdot dB_s^0 \\ &\quad - \int_t^T \big[\nabla^2 Y_s^x [A_0^\top + \partial_{px} H_0(X_s^x, \rho_s, \nabla Y_s^x)] \\ &\qquad + [A_0 + \partial_{xp} H_0(X_s^x, \rho_s, \nabla Y_s^x)]\nabla^2 Y_s^x + \partial_{xx} H_0(X_s^x, \rho_s, \nabla Y_s^x) \\ &\qquad + [\nabla^2 Y_s^x \wedge L_0]\partial_{pp} H_0(X_s^x, \rho_s, \nabla Y_s^x)[\nabla^2 Y_s^x \wedge L_0]\big]ds. \end{aligned} \tag{6.13}$$

Here $A \wedge L_0 := [(-L_0) \vee a_{ij} \wedge L_0]_{i,j}$ is the truncated matrix. The above BSDE has a Lipschitz continuous driver and thus is well posed. Recalling (6.7) and applying Itô's formula, we have

$$\begin{aligned} e^{-A_0 t}\nabla^2 Y_t^x e^{-A_0^\top t} &= e^{-A_0 T}\partial_{xx} G(X_T^x, \rho_T)e^{-A_0^\top T} \\ &\quad - \int_t^T e^{-A_0 s}[\nabla^2 Z_s^x \cdot dB_s + \nabla^2 Z_s^{0,x} \cdot dB_s^0]e^{-A_0 s} \\ &\quad - \int_t^T e^{-A_0 s}\big[\nabla^2 Y_s^x \partial_{px} H_0(X_s^x, \rho_s, \nabla Y_s^x) \\ &\qquad + \partial_{xp} H_0(X_s^x, \rho_s, \nabla Y_s^x)\nabla^2 Y_s^x + \partial_{xx} H_0(X_s^x, \rho_s, \nabla Y_s^x) \\ &\qquad + [\nabla^2 Y_s^x \wedge L_0]\partial_{pp} H_0(X_s^x, \rho_s, \nabla Y_s^x)[\nabla^2 Y_s^x \wedge L_0]\big]e^{-A_0^\top s}ds. \end{aligned}$$

Taking conditional expectation $\mathbb{E}_{\mathcal{F}_t}$ on both sides, we obtain

$$\begin{aligned} \nabla^2 Y_t^x &= e^{A_0(t-T)}\mathbb{E}_{\mathcal{F}_t}[\partial_{xx} G(X_T^x, \rho_T)]e^{A_0^\top (t-T)} \\ &\quad - \int_t^T e^{A_0(t-s)}\mathbb{E}_{\mathcal{F}_t}\big[\nabla^2 Y_s^x \partial_{px} H_0(X_s^x, \rho_s, \nabla Y_s^x) + \partial_{xp} H_0(X_s^x, \rho_s, \nabla Y_s^x)\nabla^2 Y_s^x \\ &\qquad + \partial_{xx} H_0(X_s^x, \rho_s, \nabla Y_s^x) \\ &\qquad + [\nabla^2 Y_s^x \wedge L_0]\partial_{pp} H_0(X_s^x, \rho_s, \nabla Y_s^x)[\nabla^2 Y_s^x \wedge L_0]\big]e^{A_0^\top (t-s)}ds. \end{aligned}$$

Recalling (3.2) and applying Lemma 6.2, we have

$$|\nabla^2 Y_t^x| \le e^{2[1-\underline{\kappa}'(A_0)](T-t)} \overline{L}_{xx}^G \overline{L}^A + \frac{\overline{L}_{xx}^{H_0} \overline{L}^A}{2[\underline{\kappa}'(A_0)-1]}[1-e^{2[1-\underline{\kappa}'(A_0)](T-t)}] + L_2^{H_0} \overline{L}^A [2+L_0] \int_t^T e^{2[1-\underline{\kappa}'(A_0)](s-t)} \mathbb{E}_{\mathscr{F}_t}[|\nabla^2 Y_s^x|] ds.$$

Taking the conditional expectation $\mathbb{E}_{\mathscr{F}_{t_0}}$ and noting that

$$\underline{\kappa}'(A_0) \ge \theta_2 \ge \frac{\overline{L}_{xx}^{H_0}}{2\overline{L}_{xx}^G} + 1,$$

we derive

$$\begin{aligned}\mathbb{E}_{\mathscr{F}_{t_0}}[|\nabla^2 Y_t^x|] &\le e^{2[1-\underline{\kappa}'(A_0)](T-t)} \overline{L}_{xx}^G \overline{L}^A + \overline{L}_{xx}^G \overline{L}^A [1-e^{2[1-\underline{\kappa}'(A_0)](T-t)}] \\ &\quad + L_2^{H_0} \overline{L}^A [2+L_0] \int_t^T e^{2[1-\underline{\kappa}'(A_0)](s-t)} \mathbb{E}_{\mathscr{F}_{t_0}}[|\nabla^2 Y_s^x|] ds \\ &\le \overline{L}_{xx}^G \overline{L}^A + L_2^{H_0} \overline{L}^A [2+L_0] \int_t^T e^{2[1-\underline{\kappa}'(A_0)](s-t)} \mathbb{E}_{\mathscr{F}_{t_0}}[|\nabla^2 Y_s^x|] ds.\end{aligned}$$

Then by Grönwall's inequality, we have

$$\begin{aligned}\mathbb{E}_{\mathscr{F}_{t_0}}[|\nabla^2 Y_t^x|] \le \overline{L}_{xx}^G \overline{L}^A &+ \frac{\overline{L}_{xx}^G L_2^{H_0} |\overline{L}^A|^2 [2+L_0]}{2[\underline{\kappa}'(A_0)-1] - L_2^{H_0} \overline{L}^A [2+L_0]} \\ &\times [1 - e^{-[2[\underline{\kappa}'(A_0)-1] - L_2^{H_0} \overline{L}^A [2+L_0]][T-t]}]. \end{aligned} \tag{6.14}$$

Recall (6.10), one can check straightforwardly that for any $\theta \ge \theta_3$,

$$\begin{aligned}\frac{d}{d\theta} L_{xx}^u(\theta) = \frac{1}{L_2^{H_0}\overline{L}^A}\Big[1 - \frac{(\theta - 1 - L_2^{H_0}\overline{L}^A) - L_2^{H_0}\overline{L}_{xx}^G (\overline{L}^A)^2}{\sqrt{(\theta-1-L_2^{H_0}\overline{L}^A)^2 - 2L_2^{H_0}\overline{L}_{xx}^G(\overline{L}^A)^2[\theta-1]}}\Big] \\ < 0.\end{aligned} \tag{6.15}$$

Then, since $\underline{\kappa}'(A_0) \ge \theta_2 \ge \theta_3$ and $L_0 = L_{xx}^u(\underline{\kappa}'(A_0))$, by (6.9) and (6.10) we have

$$2[\underline{\kappa}'(A_0)-1] - L_2^{H_0}\overline{L}^A[2+L_0] \ge 2[\theta_3 - 1] - L_2^{H_0}\overline{L}^A[2+L_{xx}^u(\theta_3)] \ge 0.$$

Thus (6.14) implies

$$\begin{aligned}\mathbb{E}_{\mathscr{F}_{t_0}}[|\nabla^2 Y_t^x|] &\le \overline{L}_{xx}^G \overline{L}^A + \frac{\overline{L}_{xx}^G L_2^{H_0}|\overline{L}^A|^2[2+L_0]}{2[\underline{\kappa}'(A_0)-1] - L_2^{H_0}\overline{L}^A[2+L_0]} \\ &= \frac{2\overline{L}_{xx}^G \overline{L}^A [\underline{\kappa}'(A_0)-1]}{2[\underline{\kappa}'(A_0)-1] - L_2^{H_0}\overline{L}^A[2+L_0]} \\ &= L_0,\end{aligned}$$

where the last equality is due to the straightforward calculation. In particular, by setting $t = t_0$, we have $|\nabla^2 Y_{t_0}^x| \le L_0$. Similarly, we can show $|\nabla^2 Y_t^x| \le L_0$ for all $t \in [t_0, T]$. Then $\nabla^2 Y_s^x \wedge L_0 = \nabla^2 Y_s^x$ and thus (6.13) becomes

$$\begin{aligned}\nabla^2 Y_t^x = \partial_{xx} G(X_T^x, \rho_T) - \int_t^T \nabla^2 Z_s^x \cdot dB_s - \int_t^T \nabla^2 Z_s^{0,x} \cdot dB_s^0 \\ - \int_t^T \big[\nabla^2 Y_s^x [A_0^\top + \partial_{px} H_0(X_s^x, \rho_s, \nabla Y_s^x)] \\ + [A_0 + \partial_{xp} H_0(X_s^x, \rho_s, \nabla Y_s^x)]\nabla^2 Y_s^x + \partial_{xx} H_0(X_s^x, \rho_s, \nabla Y_s^x) \\ + \nabla^2 Y_s^x \partial_{pp} H_0(X_s^x, \rho_s, \nabla Y_s^x)\nabla^2 Y_s^x\big] ds. \end{aligned} \tag{6.16}$$

By considering the equation for $\partial_{xx} u$ derived from the BSPDE in (2.4), one can readily see from (6.11), (6.12), and (6.16) that $\nabla^2 Y_t^x = \partial_{xx} u(t, X_t^x)$. In particular, $|\partial_{xx} u(t_0, x)| = |\nabla^2 Y_{t_0}^x| \le L_0$. Since (t_0, x) is arbitrary, we have $|\partial_{xx} u(t, x)| \le L_0 = L_{xx}^u(\underline{\kappa}'(A_0))$ for all (t, x). This, together with (6.15), implies (6.10). ∎

7. Global well-posedness

In this section, we establish the global well-posedness of the master equation. We shall first construct the global well-posedness of the vectorial master equation (5.9). Following the idea from [21,24,31,43], the key is to extend a local classical solution to a global one through an a priori uniform Lipschitz continuity estimate of the solution in μ. We note that Theorem 6.4 implies the uniform a priori bound of $\partial_{xx} V$. Then, by applying Theorems 4.1 and 5.1, we obtain the uniform a priori Lipschitz continuity of $\vec{U} = \partial_x V$ with respect to μ under W_2. Moreover, by Proposition 5.2 we derive the desired uniform a priori Lipschitz continuity of $\vec{U}$ with respect to μ under W_1.

We now present the main well-posedness result. Note that the dependence on the parameters is quite subtle, so we will introduce them carefully. Following the order of the assumptions below, one can easily construct a class of G and H satisfying all of them, see, e.g., Example 7.2. In particular, in light of Remark 6.3 (iii), we may set $\overline{L}^A = 1$ and consider symmetric A_0.

Theorem 7.1. *Let Assumption 3.2 with $\overline{L}_{xx}^G$, L_2^G and Assumption 3.12 (i) with $\vec{\lambda} \in D_4$ hold true, and H takes the form (6.1) such that Assumption 6.1 (ii) holds and there exists $L_2^{H_0}$ satisfying the requirements in (6.3). Fix an arbitrary $\overline{L}^A \ge 1$ and set θ_3 as in (6.9) and $L_{xx}^V := L_{xx}^u(\theta_3)$ as in (6.10). Assume further the following hold true:*

(i) *There exist $0 < \underline{\gamma} < \overline{\gamma}$ such that $\underline{\gamma} \le \overline{L}_{xx}^G$, $\overline{\gamma} > 1$, and (4.2) holds true.*

(ii) *Set A_1, A_2 as in (4.3). The matrix A_0 satisfies*

$$\begin{aligned} L^{A_0} \le \overline{L}^A, \quad \underline{\kappa}(A_0) \ge [1 + \underline{\kappa}(A_1^{-1} A_2)] L_2^{H_0}, \\ \underline{\kappa}'(A_0) \ge \theta_3, \quad |A_0| + L_2^{H_0} \le \overline{\gamma}[\underline{\kappa}(A_0) - L_2^{H_0}]. \end{aligned} \tag{7.1}$$

(iii) *There exist* $0 < \underline{L}_{xx}^{H_0} \le \overline{L}_{xx}^{H_0}$ *satisfying* (6.2) *and*

$$\underline{\gamma}[\underline{\kappa}(A_0) - L_2^{H_0}] \le \underline{L}_{xx}^{H_0} \le \overline{L}_{xx}^{H_0} \le [\overline{\gamma}[\underline{\kappa}(A_0) - L_2^{H_0}]] \wedge [2\overline{L}_{xx}^{G}[\underline{\kappa}'(A_0) - 1]]. \tag{7.2}$$

Then the master equation (1.1) *on* $[0, T]$ *admits a unique classical solution* V *with bounded* $\partial_x V$, $\partial_{xx} V$ *and* $\partial_{x\mu} V$.

Furthermore, the McKean–Vlasov FBSDEs (5.10)*,* (5.11)*,* (5.12)*,* (5.13)*,* (5.14) *and* (5.17) *are also well posed on* $[0, T]$ *and the representation formula* (5.15) *remains true on* $[0, T]$.

Proof. The uniqueness as well as the well-posedness of the involved FBSDEs and the representation formula (5.15) follow exactly the same arguments as in [31, Theorem 6.3]. Thus we shall only prove the existence.

Set $\underline{L}_{xp}^H$, $\overline{L}_{xp}^H$, $\underline{L}_{xx}^H$, $\overline{L}_{xx}^H$, L_2^H as in (6.8). Then clearly Assumptions 3.1 and 3.12 hold true. By (7.1) and (7.2), we see that (6.9) holds true and thus we have the a priori estimate (6.10). Moreover, by (7.1) we have $\underline{L}_{xp}^H \ge \underline{\kappa}(A_1^{-1} A_2) L_2^H$, and thus the result of Theorem 4.1 holds true.

We now let C_2^μ be the a priori (global) uniform Lipschitz estimate of $\partial_x V$ with respect to μ under W_2, as established by Theorems 4.1 and 5.1. Let $\delta > 0$ be the constant in Proposition 5.2, but with $\overline{L}_{xx}^G$ replaced with L_{xx}^V and L_2^G replaced with C_2^μ. Let $0 = T_0 < \cdots < T_n = T$ be a partition such that $T_{i+1} - T_i \le \frac{\delta}{2}$, $i = 0, \ldots, n-1$.

First, since $T_n - T_{n-2} \le \delta$, by Proposition 5.2 the master equation (1.1) on $[T_{n-2}, T_n]$ with terminal condition G has a unique classical solution V. For each $t \in [T_{n-2}, T_n]$, applying Theorem 6.4 we have $|\partial_{xx} V(T_{n-1}, \cdot, \cdot)| \le L_{xx}^V$. Note that by Proposition 5.2 (iii) and (iv) $V(t, \cdot, \cdot)$ has further regularities, this enables us to apply Theorems 4.1 and 5.1 and obtain that $\partial_x V(t, \cdot, \cdot)$ is uniform Lipschitz continuous in μ under W_2 with Lipschitz constant C_2^μ. Moreover, by Proposition 5.2 (ii) $\partial_x V(T_{n-1}, \cdot, \cdot)$ is also uniformly Lipschitz continuous in μ under W_1.

We next consider the master equation (1.1) on $[T_{n-3}, T_{n-1}]$ with terminal condition $V(T_{n-1}, \cdot, \cdot)$. We emphasize that $\partial_x V(T_{n-1}, \cdot, \cdot)$ has the above uniform regularity with the same constants L_{xx}^V, C_2^μ, then we may apply Proposition 5.2 with the same δ and obtain a classical solution V on $[T_{n-3}, T_{n-1}]$ with the additional regularities specified in Proposition 5.2 (iii) and (iv). Clearly, this extends the classical solution of the master equation to $[T_{n-3}, T_n]$. We emphasize again that, while the bound of $\partial_{x\mu} V(t, \cdot)$ may become larger for $t \in [T_{n-3}, T_{n-2}]$ because the C_1^μ in (5.16) now depends on $\|\partial_{x\mu} V(T_{n-1}, \cdot)\|_{L^\infty}$ instead of $\|\partial_{x\mu} V(T_n, \cdot)\|_{L^\infty}$, by the global a priori estimates in Theorems 4.1 and 5.1 we see that $\partial_x V(t, \cdot)$ corresponds to the same L_{xx}^V and C_2^μ for all $t \in [T_{n-3}, T_n]$. This enables us to consider the master equation (1.1) on $[T_{n-4}, T_{n-2}]$ with terminal condition $V(T_{n-2}, \cdot, \cdot)$, and then we obtain a classical solution on $[T_{n-4}, T_n]$ with the desired uniform estimates and additional regularities.

Repeat the arguments backwardly in time, we may construct a classical solution V for the original master equation (1.1) on $[0, T]$ with terminal condition G. Moreover, since this procedure is repeated only n times, by applying (5.16) repeatedly we see that (5.16) indeed holds true on $[0, T]$. ■

We conclude the paper by providing an example which satisfies all the assumptions in Theorem 7.1. We emphasize that there is no smallness assumption imposed here.

Example 7.2. For simplicity, let $d = 1$. Fix positive constants $0 < \underline{\alpha} \le \overline{\alpha}$ and $0 < \underline{\gamma} < \overline{\gamma}$ with $\overline{\gamma} > 1$, and fix $(\lambda_1, \lambda_2, \lambda_3)$ satisfying the requirements in (3.9). Set $\overline{L}^A := 1$ and let M_0 be a large number which will be specified later. Assume

(i) G satisfies Assumption 3.2 with

$$-\overline{\alpha} M_0 \le \partial_{xx} G(x, \mu) \le -\underline{\alpha} M_0 \quad \text{on } \mathbb{R} \times \mathscr{P}_2(\mathbb{R}); \tag{7.3}$$

(ii) H satisfies Assumption 6.1 with $A_0 := M_0^3 > L_2^{H_0}$ in (6.1), and

$$\underline{\gamma}[A_0 - L_2^{H_0}] \le \partial_{xx} H_0(x, \mu, p) \le \overline{\gamma}[A_0 - L_2^{H_0}] \quad \text{on } \mathbb{R} \times \mathscr{P}_2(\mathbb{R}) \times \mathbb{R}. \tag{7.4}$$

Then, for M_0 large enough, which may depend on $\underline{\alpha}$, $\overline{\alpha}$, $\underline{\gamma}$, $\overline{\gamma}$, $(\lambda_1, \lambda_2, \lambda_3)$, and L_2^G, $L_2^{H_0}$, one can choose appropriate λ_0 such that all the conditions in Theorem 7.1 hold true.

Proof. We first emphasize that (7.3) and (7.4) involve only $\partial_{xx} G$ and $\partial_{xx} H_0$. Note that the parameters L_2^G, $L_2^{H_0}$, which M_0 will depend on, do not involve these derivatives. So it is rather easy to construct G and H_0 satisfying both Assumptions 3.2 and 6.1, and (7.3), (7.4) with arbitrarily large M_0. Moreover, recall (3.4) and (6.2), by (7.3) and (7.4) it is clear that

$$\overline{L}_{xx}^G = \overline{\alpha} M_0, \quad \underline{L}_{xx}^{H_0} = \underline{\gamma}[A_0 - L_2^{H_0}], \quad \overline{L}_{xx}^{H_0} = \overline{\gamma}[A_0 - L_2^{H_0}]. \tag{7.5}$$

Then the θ_3 in (6.9) and $L_{xx}^u(\theta_3)$ in (6.10) become, by recalling $\overline{L}^A = 1$,

$$\theta_3 := 1 + L_2^{H_0}[1 + \overline{\alpha} M_0 + \sqrt{(1 + \overline{\alpha} M_0)^2 - 1}],$$

$$L_{xx}^V = L_{xx}^u(\theta_3) := \frac{2\overline{\alpha} M_0(\theta_3 - 1)}{\theta_3 - 1 - L_2^{H_0} + \sqrt{(\theta_3 - 1 - L_2^{H_0})^2 - 2L_2^{H_0}\overline{\alpha} M_0(\theta_3 - 1)}}.$$

We now show that the following λ_0 satisfies all the requirements

$$\lambda_0 = \frac{\overline{\gamma}^2[1 + L_{xx}^V]^2 - 8\lambda_3}{4\underline{\gamma}} + 1.$$

First, by the choice of λ_0, it is obvious that $\lambda_0 > \frac{\overline{\gamma}^2[1+L_{xx}^V]^2-8\lambda_3}{4\underline{\gamma}}$, which verifies (4.2).

Next, let $O(M)$ denote a generic positive function of M such that $\frac{O(M)}{M}$ is bounded both from above and away from 0. Then we see that

$$\theta_3 = O(M_0), \quad L_{xx}^V = O(M_0), \quad \lambda_0 = O(M_0^2). \tag{7.6}$$

By (3.10), we have

$$(\text{AntiMon})_{\xi}^{\vec{\lambda}} U(\eta,\eta) \le [-\lambda_0 \underline{\alpha} M_0 + \overline{\alpha}^2 M_0^2 - \lambda_3]\mathbb{E}[|\eta|^2] + [|\lambda_1| L_2^G + \lambda_2 |L_2^G|^2]|\mathbb{E}[\eta]|^2.$$

Since $\lambda_0 M_0 = O(M_0^3)$, it is clear that G is $\vec{\lambda}$-anti-monotone when M_0 is large enough.

Moreover, since $d = 1$, we have $\underline{\kappa}(A_0) = \overline{\kappa}(A_0) = \underline{\kappa}'(A_0) = A_0$ and $L^{A_0} = 1 \le \overline{L}^A = 1$. Recall (4.2) and (4.3). When M_0 is large, it is clear that $1 - \theta_1$ is uniformly away from 0 and then it follows from (7.6) that $\underline{\kappa}(A_1^{-1} A_2) = O(M_0^2)$. Thus, since $A_0 = M_0^3$ and $\overline{\gamma} > 1$, for M_0 sufficiently large we have the following inequalities which verify (7.1):

$$A_0 \ge 1 + \underline{\kappa}(A_1^{-1} A_2) L_2^{H_0}, \quad A_0 \ge \theta_3 \quad \text{and} \quad A_0 + L_2^{H_0} \le \overline{\gamma}[A_0 - L_2^{H_0}].$$

Finally, since $\overline{L}_{xx}^G = \overline{\alpha} M_0$, it is clear that $2\overline{L}_{xx}^G[A_0 - 1] \ge \overline{\gamma}[A_0 - L_2^{H_0}]$ for M_0 large enough. Then (7.4) implies (7.2). ∎

Funding. CM was partially supported by CityU Start-up Grant 7200684, Hong Kong RGC Grant ECS 9048215 and Hong Kong RGC Grant GRF 9043379. JZ was supported in part by NSF grants DMS-1908665 and DMS-2205972.

References

[1] Ahuja, S.: Wellposedness of mean field games with common noise under a weak monotonicity condition. SIAM J. Control Optim. **54**, 30–48 (2016) Zbl 1327.93403 MR 3439756

[2] Ahuja, S., Ren, W., Yang, T.-W.: Forward-backward stochastic differential equations with monotone functionals and mean field games with common noise. Stochastic Process. Appl. **129**, 3859–3892 (2019) Zbl 1479.60110 MR 3997664

[3] Bayraktar, E., Cecchin, A., Cohen, A., Delarue, F.: Finite state mean field games with Wright–Fisher common noise. J. Math. Pures Appl. (9) **147**, 98–162 (2021) Zbl 1459.91012 MR 4213680

[4] Bayraktar, E., Cecchin, A., Cohen, A., Delarue, F.: Finite state mean field games with Wright–Fisher common noise as limits of N-player weighted games. Math. Oper. Res. **47**, 2840–2890 (2022) Zbl 1505.91058 MR 4515487

[5] Bayraktar, E., Cohen, A.: Analysis of a finite state many player game using its master equation. SIAM J. Control Optim. **56**, 3538–3568 (2018) Zbl 1416.91013 MR 3860894

[6] Bayraktar, E., Zhang, X.: On non-uniqueness in mean field games. Proc. Amer. Math. Soc. **148**, 4091–4106 (2020) Zbl 1444.91025 MR 4127851

[7] Bensoussan, A., Frehse, J., Yam, P.: Mean field games and mean field type control theory. SpringerBriefs Math., Springer, New York (2013) Zbl 1287.93002 MR 3134900

[8] Bensoussan, A., Graber, P. J., Yam, S. C. P.: Stochastic control on space of random variables. arXiv:1903.12602 (2019)

[9] Bensoussan, A., Graber, P. J., Yam, S. C. P.: Control on Hilbert spaces and application to some mean field type control problems. arXiv:2005.10770 (2023)

[10] Bensoussan, A., Yam, S. C. P.: Control problem on space of random variables and master equation. ESAIM Control Optim. Calc. Var. **25**, article no. 10 (2019) Zbl 1450.35305 MR 3943358

[11] Bertucci, C.: Monotone solutions for mean field games master equations: finite state space and optimal stopping. J. Éc. Polytech. Math. **8**, 1099–1132 (2021) Zbl 1472.35394 MR 4275225

[12] Bertucci, C.: Monotone solutions for mean field games master equations: continuous state space and common noise. Comm. Partial Differential Equations **48**, 1245–1285 (2023) Zbl 07793273 MR 4687277

[13] Bertucci, C., Lasry, J.-M., Lions, P.-L.: Some remarks on mean field games. Comm. Partial Differential Equations **44**, 205–227 (2019) Zbl 1411.91100 MR 3941633

[14] Buckdahn, R., Li, J., Peng, S., Rainer, C.: Mean-field stochastic differential equations and associated PDEs. Ann. Probab. **45**, 824–878 (2017) Zbl 1402.60070 MR 3630288

[15] Cardaliaguet, P.: A short course on mean field games (Lectures by P.-L. Lions, Collège de France). https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20200303.pdf, visited on 16 April 2024 (2021)

[16] Cardaliaguet, P., Cirant, M., Porretta, A.: Splitting methods and short time existence for the master equations in mean field games. J. Eur. Math. Soc. (JEMS) **25**, 1823–1918 (2023) Zbl 1522.35511 MR 4592861

[17] Cardaliaguet, P., Delarue, F., Lasry, J.-M., Lions, P.-L.: The master equation and the convergence problem in mean field games. Ann. of Math. Stud. 201, Princeton University Press, Princeton, NJ (2019) Zbl 1430.91002 MR 3967062

[18] Cardaliaguet, P., Porretta, A.: An introduction to mean field game theory. In: Mean field games, Lecture Notes in Math. 2281, Springer, Cham, 1–158 (2020) Zbl 1457.91058 MR 4214774

[19] Cardaliaguet, P., Souganidis, P.: Monotone solutions of the master equation for mean field games with idiosyncratic noise. SIAM J. Math. Anal. **54**, 4198–4237 (2022) Zbl 1497.35471 MR 4451309

[20] Carmona, R., Delarue, F.: Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games. Probab. Theory Stoch. Model. 83, Springer, Cham (2018) Zbl 1422.91014 MR 3752669

[21] Carmona, R., Delarue, F.: Probabilistic theory of mean field games with applications. II. Mean field games with common noise and master equations. Probab. Theory Stoch. Model. 84, Springer, Cham (2018) Zbl 1422.91015 MR 3753660

[22] Cecchin, A., Dai Pra, P., Fischer, M., Pelino, G.: On the convergence problem in mean field games: a two state model without uniqueness. SIAM J. Control Optim. **57**, 2443–2466 (2019) Zbl 1426.91025 MR 3981375

[23] Cecchin, A., Delarue, F.: Selection by vanishing common noise for potential finite state mean field games. Comm. Partial Differential Equations **47**, 89–168 (2022) Zbl 1490.91018 MR 4373880

[24] Chassagneux, J.-F., Crisan, D., Delarue, F.: A probabilistic approach to classical solutions of the master equation for large population equilibria. Mem. Amer. Math. Soc. **280**, v+123 pp. (2022) Zbl 1520.91003 MR 4493576

[25] Delarue, F., Foguen Tchuendom, R.: Selection of equilibria in a linear quadratic mean-field game. Stochastic Process. Appl. **130**, 1000–1040 (2020) Zbl 1471.91029 MR 4046528

[26] Delarue, F., Lacker, D., Ramanan, K.: From the master equation to mean field game limit theory: a central limit theorem. Electron. J. Probab. **24**, article no. 51 (2019) Zbl 1508.60032 MR 3954791

[27] Delarue, F., Lacker, D., Ramanan, K.: From the master equation to mean field game limit theory: Large deviations and concentration of measure. Ann. Probab. **48**, 211–263 (2020) Zbl 1445.60025 MR 4079435

[28] Djete, M. F.: Large population games with interactions through controls and common noise: convergence results and equivalence between open-loop and closed-loop controls. ESAIM Control Optim. Calc. Var. **29**, article no. 39 (2023) Zbl 07734411 MR 4598796

[29] Foguen Tchuendom, R.: Uniqueness for linear-quadratic mean field games with common noise. Dyn. Games Appl. **8**, 199–210 (2018) Zbl 1390.91061 MR 3764697

[30] Gangbo, W., Mészáros, A. R.: Global well-posedness of master equations for deterministic displacement convex potential mean field games. Comm. Pure Appl. Math. **75**, 2685–2801 (2022) Zbl 1510.35342 MR 4509653

[31] Gangbo, W., Mészáros, A. R., Mou, C., Zhang, J.: Mean field games master equations with nonseparable Hamiltonians and displacement monotonicity. Ann. Probab. **50**, 2178–2217 (2022) Zbl 1501.35403 MR 4499277

[32] Gangbo, W., Święch, A.: Existence of a solution to an equation arising from the theory of mean field games. J. Differential Equations **259**, 6573–6643 (2015) Zbl 1359.35221 MR 3397332

[33] Huang, M., Malhamé, R. P., Caines, P. E.: Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. Commun. Inf. Syst. **6**, 221–251 (2006) Zbl 1136.91349 MR 2346927

[34] Iseri, M., Zhang, J.: Set values for mean field games. arXiv:2107.01661 (2024)

[35] Lacker, D.: A general characterization of the mean field limit for stochastic differential games. Probab. Theory Related Fields **165**, 581–648 (2016) Zbl 1344.60065 MR 3520014

[36] Lacker, D.: Limit theory for controlled McKean–Vlasov dynamics. SIAM J. Control Optim. **55**, 1641–1672 (2017) Zbl 1362.93167 MR 3654119

[37] Lacker, D.: On a strong form of propagation of chaos for McKean–Vlasov equations. Electron. Commun. Probab. **23**, article no. 45 (2018) Zbl 1396.65013 MR 3841406

[38] Lacker, D.: On the convergence of closed-loop Nash equilibria to the mean field game limit. Ann. Appl. Probab. **30**, 1693–1761 (2020) Zbl 1470.91036 MR 4133381

[39] Lacker, D., Le Flem, L.: Closed-loop convergence for mean field games with common noise. Ann. Appl. Probab. **33**, 2681–2733 (2023) Zbl 1519.49026 MR 4612653

[40] Lasry, J.-M., Lions, P.-L.: Mean field games. Jpn. J. Math. **2**, 229–260 (2007) Zbl 1156.91321 MR 2295621

[41] Lions, P.-L.: Théorie des jeux de champ moyen et applications. Collège de France, https://www.youtube.com/watch?v=nXl0Tk_8kQ8, visited on 16 April 2024 (2007)

[42] Mayorga, S.: Short time solution to the master equation of a first order mean field game. J. Differential Equations **268**, 6251–6318 (2020) Zbl 1447.35331 MR 4069006

[43] Mou, C., Zhang, J.: Wellposedness of second order master equations for mean field games with nonsmooth data. arXiv:1903.09907 (2022)

[44] Nutz, M., San Martin, J., Tan, X.: Convergence to the mean field game limit: A case study. Ann. Appl. Probab. **30**, 259–286 (2020) Zbl 1437.91058 MR 4068311