ESAIM: COCV 32 (2026) 59
https://doi.org/10.1051/cocv/2026039

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

GLOBAL WELL-POSEDNESS OF FIRST-ORDER MEAN FIELD GAMES AND MASTER EQUATIONS WITH NONLINEAR DYNAMICS

Alain Bensoussan[1,*], Tak Kwong Wong[2], Sheung Chi Phillip Yam[3] and Hongwei Yuan[4]

Abstract. This paper addresses the generic first-order mean field game problem with nonlinear dynamics. A major contribution here is the provision of new crucial *a priori* estimates for the associated forward–backward ordinary differential equation (FBODE) system. In addition, we require monotonicity conditions intimately on the coefficient functions but not on the Hamiltonians to handle their nonseparable nature and nonlinear dynamics; as tackling Hamiltonians directly, it potentially dissolves much useful information. Moreover, we are able to deal with the nonseparable Hamiltonian under the hypothesis that the first-order derivative of the drift function in the measure variable is not too large relative to the convexity of the running cost function. Our approach involves demonstrating the local existence of a solution, providing new *a priori* estimates for the sensitivity of the backward equation with respect to the forward initial condition, and gluing these into a global solution. Furthermore, we establish the local and global existence as well as uniqueness of classical solutions for the mean field game and its master equation. To illustrate the effectiveness of our proposed general theory, we apply it to resolve two non-trivial non-LQ (non-linear-quadratic) examples with a nonseparable Hamiltonian and nonlinear dynamics, which remains unexplained in contemporary literature.

Mathematics Subject Classification. 49N80, 35Q89, 49J55, 60H30, 60H10.

Received August 26, 2024. Accepted May 10, 2026.

1. Introduction

The study of mean field games (MFG) and control problems has grown significantly over the past two decades, establishing itself as a prominent field of research in both mathematics and engineering. The topics involve the study of systems with a large number of interactive agents; in particular, as the number of agents approaches infinity, a mean field approximation becomes increasingly applicable. This methodology, first introduced by Huang–Malhamé–Caines [1] under the name "large population stochastic dynamic games" and Lasry–Lions

Keywords and phrases: Mean field game, Wasserstein metric space, forward–backward differential systems, decoupling field, Jacobian flows, master equations, non-linear-quadratic examples.

[1] International Center for Decision and Risk Analysis, Naveen Jindal School of Management, University of Texas at Dallas, 800 W Campbell Road, Richardson, TX 75080, USA.
[2] School of Mathematical Sciences, Shenzhen University, 518061 Shenzhen, PR China.
[3] Department of Statistics and Data Science, The Chinese University of Hong Kong, Ma Liu Shui, Hong Kong, PR China.
[4] Department of Mathematics, University of Macau, Avenida da Universidade, Taipa, Macau, PR China.

[*] Corresponding author: axb046100@utdallas.edu

[2] under another name "mean field games" independently, facilitates the analysis and comprehension of these complex systems composed of large agent populations. In the last two decades, the theory has been enriched with substantial advancements, fueled by both theoretical progress and practical applications. The pioneering monographs by Cardaliaguet, Delarue, Lasry and Lions [3], Carmona and Delarue [4], Gomes, Pimentel and Voskanyan [5], and Bensoussan, Frehse and Yam [6] have made remarkable strides in elucidating the fundamental problem formulation and optimization procedure. For a thorough understanding of the stochastic (Pontryagin) maximum principle, McKean–Vlasov dynamics, and forward–backward stochastic differential equations (FBSDEs), as well as their relationship to mean field games and control problems, interested readers may refer to the following key resources [1, 7–11, 13–24]. To delve deeper into the dynamic programming principle, Bellman equations, master equations and their connections with mean field games and control problems, the following references can provide invaluable insights [25–33].

One representative model of the mean field games problem is the linear quadratic setting, which has been investigated by Bensoussan–Sung–Yam–Yung [9], Carmona–Delarue–Lachapelle [34], Pham [21], Yong [35]. For the problem with a more complicated structure, there are three mainstream approaches in the literature. The first one is to solve the forward–backward Fokker–Planck (FP) and Hamilton–Jacobi–Bellman (HJB) system. Cardaliaguet [36] and Graber [37] studied the well-posedness of weak solutions of the first-order (in the absence of Brownian noise) FP–HJB system, with various structures of the Hamiltonian. Subsequently, Cardaliaguet–Graber [38] generalized their findings under more general conditions. For the second-order case (with only individual Brownian noise), when there is a nonlocal smoothing coupling, Lasry–Lions [2] showed the classical well-posedness of the FP–HJB system under the Lasry–Lions monotonicity. Porretta [39] showed the well-posedness of weak solutions for more general cost functionals. All these aforementioned articles are working on separable Hamiltonians and on a torus. Further, Ambrose [40] established the well-posedness of strong solutions to the FP–HJB system on a torus with a nonseparable Hamiltonian, provided that the initial and terminal data or the Hamiltonian is small.

The second approach has gained popularity of late, which is about solving an infinite-dimensional PDE, namely the *master equation*. This idea was first proposed by P.-L. Lions in the lecture [41] (see also Cardaliaguet [42], and the derivations in Bensoussan–Frehse–Yam [25, 43], Carmona–Delarue [4]), and then it gained significant attention from scholars. For the first-order master equation (without Brownian noise), a local (in time) classical solution on a torus was established by Gangbo-Święch [31] for a quadratic Hamiltonian; Mayorga [44] extended the work to a generic separable Hamiltonian on $\mathbb{R}^d$. For the second-order master equation (with individual Brownian noise), Ambrose-Mészáros [45] constructed local solutions on a torus with a nonseparable Hamiltonian locally depending on the measure argument. In Cardaliaguet–Cirant–Porretta [27], local solutions were constructed on $\mathbb{R}^d$ with a nonseparable Hamiltonian that also involves common noise *via* a novel splitting method. To the best of our knowledge, in order to extend the local theory of a generic problem to a global one, certain monotonicity assumptions are required on the Hamiltonian. The first type of monotonicity, known as *Lasry–Lions monotonicity* (LLM) as mentioned above, was introduced in Lasry–Lions [2]. Under this condition, Cardaliaguet–Delarue–Lasry–Lions [3] established the global well-posedness of the master equation on a torus with a generic separable Hamiltonian and a common noise using a PDE approach and Hölder estimates. Chassagneux–Crisan–Delarue [16] established, through a probabilistic approach with associated FBSDEs, the local existence and uniqueness of classical solutions for the master equation, and show that existence and uniqueness of such a classical solution may be extended to arbitrary large time intervals *via* an inductive argument, provided that a uniform (along the recursion) *a priori* estimate for the Lipschitz constant (in both the space variable and the measure argument) of the decoupling field can be obtained. They also identify two classes of examples, belonging to the framework of mean-field games, under which the assumed uniform *a priori* estimate can be shown to hold by combining either convexity or ellipticity with Lasry–Lions condition. In fact, our crucial *a priori* estimate (7.18) can directly derive the uniform *a priori* estimate for the Lipschitz constant of the decoupling field in the first order case. Additionally, the global well-posedness of different senses of weak solutions, with a separable Hamiltonian and the presence of a common noise, were established in Mou–Zhang [46] (on $\mathbb{R}^d$), Bertucci [47] (on a torus) and Cardaliaguet–Souganidis [48] (on $\mathbb{R}^d$, without individual noise, still, there is a common noise). Also, see the LLM for master equations in mean field type control problems in Bensoussan–Tai–Yam [11]. Finally, two examples in Cecchin–Pra–Fischer–Pelino [49] (two-state model) and Delarue–Tchuendom [50] (linear quadratic model) illustrated the non-unique existence of Nash equilibrium in the absence of LLM. A subsequently proposed monotonicity assumption is the *displacement monotonicity* (DM); see [30]. The concept of DM in the context of mean field games was initially introduced as "weak monotonicity" by Ahuja [51], who employed this concept to examine the well-posedness of mean field games with common noise using simple cost functions *via* the FBSDE approach. The DM was first used in Gangbo-Mészáros [29] for the global well-posedness of first-order master equation with separable Hamiltonians on $\mathbb{R}^d$, then in Gangbo-Mészáros–Mou–Zhang [30] for second-order master equations in the presence of common noise and nonseparable Hamiltonian with bounded derivatives up to fifth order. Also, see the role of DM for master equations in mean field type control problems in Bensoussan–Graber–Yam [8], Bensoussan–Tai–Yam [11] and also Bensoussan–Wong–Yam–Yuan [52]. One can read Section 2.1 in [10, 11], Remark 3.1 in [11], Section 5.3 in [52] and Graber-Mészáros [53] for more discussion about these two monotonicity assumptions. In addition, Bensoussan–Huang–Tang–Yam [54] introduce a so-called β-monotonicity that is commonly used in stochastic control theory and admits more interesting cases including but not limited to the displacement monotonicity, the small mean field effect condition or the Lasry–Lions monotonicity. Although some overlap may exist among these latter three conditions, their interrelationships are complex and not simply hierarchical; see [54] for detailed discussion. It is worth noting that Cecchin–Delarue [55] established the well-posedness of weak solutions to the master equation on a torus for a separable Hamiltonian, without imposing any monotonicity on the coefficients. Graber-Mészáros [53] propose two new monotonicity conditions, dichotomy with the Lasry–Lions monotonicity and displacement monotonicity conditions, that could serve as sufficient conditions for unconditional uniqueness of Nash equilibria in mean field games. In addition, they [56] apply their new monotonicity condition, through a newly derived nonlinear transport equation written on the space of probability measures, to establish both the uniqueness of Nash equilibria and the global in time well-posedness of master equations for a class of deterministic mean field games, where the interaction of the agents happens only at the terminal time, and, in the absence of monotonicity, they can employ the newly derived nonlinear transport equation in its conservative form to define weak entropy solutions for the master equation. Outside the monotonicity-based theory, Lions–Souganidis [57] investigates the homogenization of viscous backward–forward mean-field games systems in small viscosity limits and periodic environments, establishing first-order forward–backward limits, and demonstrating that the averaged system is of MFG-type in the nonlocal coupling case while the limit system is not necessarily of MFG-type in the local coupling case. Lasry–Lions–Seeger [58] presents dimension reduction techniques in deterministic mean field games by identifying conditions that enable the reduction of high-dimensional systems to lower-dimensional ones, validates these methods through consistency analysis with small-noise approximations, and offers computable frameworks for modeling complex interactions.

The third approach to resolving mean field games is to employ probabilistic tools. Carmona–Lacker [59] utilized this approach by first weakly solving the optimal control problem and then using the modified Kakutani's theorem to find the equilibrium measure. Carmona–Delarue–Lacker [60] found the weak equilibria of mean field games subject to common noise by applying the Kakutani–Fan–Glicksberg fixed-point theorem for set-valued functions, as well as a discretization procedure for the common noise. The weak equilibria can be shown to be strong under additional assumptions. We can also use the stochastic maximum principle to characterize the mean field games problem to the mean field (or McKean–Vlasov) forward–backward stochastic differential equations (FBSDEs). A major advantage of this approach is that the FBSDE is symbolically a finite-dimensional problem, as opposed to solving the master equation which is inherently infinite dimensional. The mean field SDEs were studied in Buckdahn–Li–Peng–Rainer [61]. The study of mean field BSDEs was first conducted by Buckdahn–Djehiche–Li–Peng [62] and Buckdahn–Li–Peng [63]. The well-posedness of the coupled mean field FBSDEs was first studied in Carmona–Delarue [64]. Later, Carmona–Delarue [15] relaxed the assumptions and established the well-posedness of FBSDEs under the framework of linear forward dynamics and convex coefficients of the backward equation. Bensoussan–Yam–Zhang [24] also addressed the well-posedness of FBSDEs with another set of generic assumptions on the coefficients. We also refer to the article [10, 11] which resolved the mean field games problem under the small mean field effect. In the article, the authors provided a brand new set of estimates for establishing the global well-posedness of the FBSDEs and hereto a new set of estimates for showing the regularity of the value function. The classical well-posedness of master equation was also provided in the paper. We also refer readers to [17, 19, 22, 23, 28, 32, 33] for the mean field type control problems.

To obtain classical solutions of first-order mean field games and their associated equations, the standard approach is to consider linear forward dynamics and impose certain convexity requirements on the cost functions with respect to the measure variable (for example, the Lasry–Lions monotonicity condition) and the control. This paper extends the aforementioned framework to incorporate nonlinear dynamics and provides tractable sufficient conditions for the global-in-time well-posedness of classical solutions by imposing assumptions on the drift and cost functions, rather than on the Hamiltonian (for example, the displacement monotonicity condition). This shift in focus is necessary to address the existence and uniqueness issues of the optimal control involved in defining the Hamiltonian within the context of nonlinear dynamics. The key innovation of our approach lies in the introduction of the novel cone condition (see (4.28)), a concept that has not been explored in prior works. In this paper, we analyze the problem with a generic dynamics in which the nonlinear drift function $f(x, \mu, a)$ depends not only on the control a, but also on the state variable x and mean-field measure μ. Contrary to the simpler cases considered in existing literature, such as $f(x, \mu, a) \equiv a$, our model assumes that the drift function can be at most linear-growth but quite arbitrarily nonlinear. In addition, we consider a running cost function $g(x, \mu, a)$ that can be nonseparable and of quadratic-growth in the entire space of $\mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x})$ in (x, μ), and thus the Hamiltonian $h(x, \mu, z) = \min_a (f(x, \mu, a) \cdot z + g(x, \mu, a))$ can also be nonseparable and of quadratic-growth in the entire space $\mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x}$ in (x, μ, z).

More precisely, we assume that $f(x, \mu, a)$ is Lipschitz continuous and grows at most linearly, while the running cost function $g(x, \mu, a)$ is convex in the control a and satisfies a monotonicity condition. This monotonicity condition can be motivated by the positive definiteness of the Schur complement of the Hessian matrix of the Lagrangian in the lifted version and was proposed in our previous paper [52]. Even though the calculus in the lifted variables provides useful insights, it creates additional technical difficulties in the mathematical analysis, and is not that effective in analyzing the problem with the generic drift rate; see Section 4 of [52] for the detailed discussions. The proposed monotonicity conditions (4.44), (4.46) and (7.1) are closely related to the Lasry–Lions monotonicity condition and displacement monotonicity condition; these conditions were first introduced in the works of [3, 4] and [29, 30], respectively, which we find to be interesting and important. For precise assumptions on the coefficient functions, refer to Sections 4.1 and 7.1, while Section 4.3 and Remark 7.2 provide more elaborate discussion on the relationship and comparison between our convexity assumptions and other monotonicity conditions. In particular, our Hypothesis (**h3**) in Section 7.1 highlights two important points: (i) excessively strong cross-interactions among the control a, state x, and its measure μ must be avoided to ensure long-term equilibrium strategies, as self-reinforcing interactions among these elements would destabilize the game dynamics over time; (ii) the drift rate f must neither exhibit substantial quadratic growth effects nor involve excessively strong couplings between x, μ and a, since such properties could induce a finite-time blowup of the state; see Remark 7.1 for details.

Under these assumptions, our main results include the global-in-time solvability of the generic mean field games, as well as the global-in-time existence, uniqueness, and classical regularity of the corresponding forward–backward ordinary differential equation (FBODE) system. Theorem 7.8 establishes the existence of a solution to the FBODE system, and Theorem 8.2 proves its uniqueness; while Theorems 7.9 and 8.4 address the respective issue for the master equation.

To demonstrate the usefulness of our newly proposed general theory, we provide a resolution of two non-trivial non-LQ examples with a nonseparable Hamiltonian in Section 9. In these examples, the drift function f exhibits a linear structure with controlled nonlinear perturbations that demonstrate sufficiently fast decay at infinity. The running cost g is strictly convex only in the state-control pair (x, a) while its dependence on μ is limited to mild quadratic perturbations without strict convexity; similarly, the terminal cost k is strictly convex only in x (but not in μ) with small quadratic perturbations. To the best of our knowledge, this configuration appears absent in the existing literature, thereby serving as an illustration that validates the broader applicability of our theoretical framework. Furthermore, it is worth noting that the explicit examples mentioned in Section 9 is not particularly exceptional. In fact, there exists a wide range of easily verifiable scenarios that also fall under the scope of our results. For the details, see a class of examples at the end of Section 4.1 and Remark 7.2.

To address the challenges arising from the nonlinear drift function in the generic mean field games (refer to (2.5)–(2.10)), we employ the approach established in our previous paper [52], despite facing different problems and equations in the current context. Firstly, unlike the simpler cases where the optimal control can be solved explicitly from the first-order condition, in our generic setting of the drift and running cost functions, the optimal control cannot be solved explicitly at all. To better understand its behavior, we introduce a cone condition (see (4.28)), which is a property (not an assumption) that our constructed solution of the FBODE system can be shown to fulfill (see (7.60)). This condition is incorporated with the assumption on the asymptotic behavior of second-order derivatives of the drift function (see (4.3)), which is generic enough and yet guarantees the solvability of the optimal control *via* the first-order condition (see (4.30)). To this end, we use a generalized version of the implicit function theorem (see the Appendix of [52]) that includes measure arguments. Secondly, we show that solving the generic mean field games (see (2.5)–(2.10)) can be reduced to solving a FBODE system (see (2.12)). Thirdly, we use fixed point arguments to prove the local-in-time existence, uniqueness, and regularity of solutions to the general FBODE system (6.1). This system is essentially the same as the FBODE system (2.12), except that it imposes a more general terminal data $p(x,\mu)$. The minimal lifespan (that can be guaranteed by the fixed point argument) of local solution depends mainly on the bound of the derivatives of the terminal data $p(x,\mu)$, as well as some other uniform-in-time constants related to the given drift and running cost functions; more discussions are given in Theorems 6.3 and 6.4. To obtain the global-in-time classical solution of the FBODE system (2.12), we paste together the local-in-time solutions. The key point is to obtain *a priori* uniform-in-time estimates on the derivatives of the backward equation $Z_s^{t,m}(x)$, since $Z_{t_i}^{t,m}(x)$, $i=1,2,...,N+1$ are the terminal data p on the respective sub-interval $[t_{i-1},t_i]$. We can obtain our new *a priori* estimates (see Thm. 7.5) under the monotonicity conditions (4.44), (4.46), and (7.1). Using both the local-in-time results and the new *a priori* estimates on the derivatives of the backward equation $Z_s^{t,m}(x)$, we can prove our global-in-time well-posedness claim.

The main contribution of this work is the provision of new *a priori* estimates (see Eq. (7.18)), whose proofs differ from those given in [52] as the associated forward–backward ordinary differential equation (FBODE) system is fundamentally different. In comparison to the proof of *a priori* estimates presented in [52], the main challenge in establishing new *a priori* estimates in this work lies in accurately tracking the term $\mathcal{T}_5$ in Step 4 of the proof of Theorem 7.5. To address this challenge, we introduce a novel monotonicity condition (7.1) that is aligned with the cone condition (4.28). This condition allows us to effectively monitor the nonlinear dynamics described by equation (2.5). These estimates are the key to guaranteeing the cone conditions and our global-in-time well-posedness.

Our approach has several direct consequences and advantages. Firstly, by using the push-forward operator $\#$, we can transform the study of μ_s in the Wasserstein metric space to X_s in the tangent space of Hilbert space L^2_m at the measure m. This allows us to avoid the difficulties caused by the lack of vector space structure in the Wasserstein metric space. Secondly, the optimal control $\widehat{\alpha}(x,\mu,z)$ is solved from the first-order condition $\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a g(x,\mu,\widehat{\alpha})=0$. However, in our generic setting, the drift function f can be nonlinear in control, and when the magnitude of variable z is large, we are not able to apply the implicit function theorem to guarantee the unique solvability of the optimal control from the first-order condition, even with the convexity in a of $g(x,\mu,a)$. Nonetheless, the proposed "cone condition", which our constructed solution of the FBODE system fulfills, suggests that, for any given (x,μ), only those z's in a bounded subset governed by (x,μ) are related to the solutions of FBODE. This allows us to apply the implicit function theorem in the related region while solving the FBODE. Thirdly, while deriving our new *a priori* estimates (7.18), we have to carefully deal with several terms (refer to Eq. (7.40)) that could be cancelled out in our previous paper [52] on mean field type control problems, but this is no longer possible in the current work. To manage these terms, we employ a monotonicity condition related to both the running cost function and drift function (see Eq. (7.1)). This condition is analogous to the displacement monotonicity condition proposed in [30], and we only require our monotonicity condition being valid within a cone subset. The monotonicity condition (7.1) can be directly deduced from the hypothesis (see (7.3)) that the first-order derivative of the drift function in the measure argument cannot be excessively large relative to the convexity of the running cost function. This requirement emerges only when the Hamiltonian is nonseparable; for more details, refer to Remark 7.2. Finally, the solution of the FBODE system fulfills the "cone condition" on the whole time interval $[0,T]$ and satisfies the *a priori* estimates (7.18) at the same time. Using our new *a priori* estimates (7.18) of the derivatives of the $Z_s^{t,m}(x)$, we can obtain a uniform lower bound on the minimal time mesh size of local solutions. This allows us to construct a global solution over the whole time interval $[0,T]$ for any fixed large time T by gluing these local solutions together.

The rest of this article is organized as follows. In Section 2, we introduce the Wasserstein space of measures, the push-forward operator and the formulation of first-order mean field game. Section 3 recalls various derivatives of functionals defined on the Wasserstein space. In Section 4, we introduce the assumptions used in our results, as well as some direct consequences and useful properties implied by these assumptions. Furthermore, we also compare these assumptions with other assumptions used in the literature. Section 5 introduces the HJ–FP system, master equation and forward–backward ordinary differential equation (FBODE) system, and explains why in order to solve the original first-order mean field game, it suffices to solve the FBODE system. In Section 6, we prove the existence and regularity of local-in-time solutions of the FBODE system, and then establish the global-in-time solution of the FBODE system (2.12) and master equation (2.11) in Section 7 by using some new *a priori* estimates. We also establish the uniqueness in Section 8. Finally, in Section 9, we provide two non-trivial examples of non-LQ mean field game with nonseparable Hamiltonian and nonlinear dynamics that satisfies our assumptions in Theorem 7.8.

2. Generic first-order mean field games

Let d and q be positive integers, and $\mathcal{P}_q(\mathbb{R}^d)$ be the space of all probability measures on $\mathbb{R}^d$, each of which has a finite q-th order moment, equipped with the Wasserstein metric $W_q(\mu,\nu)$ defined by:

$$W_q(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|\xi-\eta|^q d\pi(\xi,\eta)\right)^{\frac{1}{q}}, \tag{2.1}$$

where $\Pi(\mu,\nu)$ denotes the set of all joint probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ such that the marginals are the probability measures μ and ν, respectively. Also define $\|\mu\|_q = \left(\int_{\mathbb{R}^d}|x|^q\, d\mu(x)\right)^{1/q}$. Clearly, $\mathcal{P}_{q'}(\mathbb{R}^d)\subset\mathcal{P}_q(\mathbb{R}^d)$ for any integers $1\le q<q'<\infty$.

For any fixed $m\in\mathcal{P}_q(\mathbb{R}^d)$, we denote $L_m^{q,d_1,d_2} := L_m^q(\mathbb{R}^{d_1};\mathbb{R}^{d_2})$, the set of all measurable maps $\Phi:\mathbb{R}^{d_1}\to\mathbb{R}^{d_2}$ such that $\int_{\mathbb{R}^{d_1}}|\Phi(x)|^q dm(x)<\infty$. We equip L_m^{q,d_1,d_2} with a norm $\|\cdot\|_{L_m^{q,d_1,d_2}}$ or simply $\|\cdot\|_{L_m^{q,d}}$ when $d_1=d_2=d$:

$$\|X\|_{L_m^{q,d_1,d_2}} := \left(\int_{\mathbb{R}^{d_1}}|X(x)|^q dm(x)\right)^{\frac{1}{q}},\ \forall\ X\in L_m^{q,d_1,d_2}; \tag{2.2}$$

clearly, L_m^{q,d_1,d_2} is a Banach space and $L_m^{q',d_1,d_2}\subset L_m^{q,d_1,d_2}$ for any integers $1\le q<q'<\infty$. For $q=2$, we further equip L_m^{2,d_1,d_2} with an inner product:

$$\langle X,Y\rangle_{L_m^{2,d_1,d_2}} := \int_{\mathbb{R}^{d_1}}X(x)\cdot Y(x)dm(x),\ \forall\ X,Y\in L_m^{2,d_1,d_2}; \tag{2.3}$$

clearly, L_m^{2,d_1,d_2} is a Hilbert space. The space L_m^{2,d_1,d_2} can be heuristically interpreted as the tangent space associated with the point m in the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$, and we use subscript m to indicate this nature; one can refer to [65] for instance.

Definition 2.1. For any given probability measure $m\in\mathcal{P}_2(\mathbb{R}^d)$ and measurable map $X\in L_m^{2,d}$, define another probability measure $X\#m\in\mathcal{P}_2(\mathbb{R}^d)$ so that for every measurable function $\phi:\mathbb{R}^d\to\mathbb{R}$ such that

$\sup_{x\in\mathbb{R}^d} \frac{|\phi(x)|}{1+|x|^2} < \infty$,

$$\int_{\mathbb{R}^d} \phi(x)d(X\#m)(x) = \int_{\mathbb{R}^d} \phi(X(x))dm(x). \tag{2.4}$$

This new probability measure $X\#m$ is actually the push-forward measure of m under the map $X(\cdot)$.

In this article, we denote the state variable by $x \in \mathbb{R}^{d_x}$ and the control variable by $a \in \mathbb{R}^{d_a}$, also consider continuous coefficient and cost functions:

$$\begin{aligned}
&f(x,\mu,a) = (f_1,...,f_{d_x})(x,\mu,a) : \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_a} \to \mathbb{R}^{d_x};\\
&g(x,\mu,a) : \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_a} \to \mathbb{R};\\
&k(x,\mu) : \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \to \mathbb{R};
\end{aligned}$$

where these differentiable, in x and a, functions can be allowed to depend on time, but we here omit this dependence to avoid unnecessary technicalities; yet all the following discussions remain valid with time dependent coefficients and cost functions. Besides, f plays the role as a drift function, g is the running cost, and k stands for the terminal cost functional. The argument μ is the place for the mean field term of the state.

To describe the generic first-order mean field game, let us first introduce the notations for the dynamics and objective function as follows: given an arbitrary process of probability measures $\widetilde{m} := (\widetilde{m}_\tau)_{\tau\in[0,T]}$ with $\sup_{\tau\in[0,T]} \|\widetilde{m}_\tau\|_1 < \infty$ and any control process $\alpha := (\alpha_\tau)_{\tau\in[0,T]}$ with $\int_0^T |\alpha_\tau|^2 d\tau < \infty$, for any fixed $t \in [0,T]$ and $x \in \mathbb{R}^{d_x}$, consider the following dynamics of controlled state $X_s^{t,\widetilde{m},\alpha}(x)$:

$$X_s^{t,\widetilde{m},\alpha}(x) = \ x + \int_t^s f\left(X_\tau^{t,\widetilde{m},\alpha}(x), \widetilde{m}_\tau, \alpha_\tau\right) d\tau,\ \ s \in [t,T]. \tag{2.5}$$

The existence and uniqueness of solutions to (2.5) can be proven under standard Lipschitz and linear-growth conditions of f; moreover this solution satisfies $\sup_{s\in[t,T]} \left|X_s^{t,\widetilde{m},\alpha}(x)\right| \le C(1+|x|) < \infty$, where the constant C depends on T. Now, we define the objective function

$$j(t,x,\widetilde{m},\alpha) := \int_t^T g(X_s^{t,\widetilde{m},\alpha}(x), \widetilde{m}_s, \alpha_s)ds + k(X_T^{t,\widetilde{m},\alpha}(x), \widetilde{m}_T),\ \text{for}\ t \in [0,T]. \tag{2.6}$$

The optimal control process $\alpha^* := (\alpha_s^*)_{s\in[t,T]}$ is defined by

$$\alpha^*(t,x,\widetilde{m}) := \operatorname*{arg\,min}_{\alpha\in\mathbb{A}} j(t,x,\widetilde{m},\alpha), \tag{2.7}$$

provided the minimizer exists uniquely[1]. Here, the set of all admissible controls is specified as $\mathbb{A} := L^2([t,T];\mathbb{R}^{d_a})$.

The main purpose of this work is to solve the generic first order mean field game, which can be stated as follows: for any given initial data $m_0 \in \mathcal{P}_2(\mathbb{R}^{d_x})$ at the initial time $t = 0$, we aim at finding the Nash equilibrium probability measure $\widehat{m} := (\widehat{m}_s)_{s\in[0,T]}$ such that the corresponding optimal control process $\alpha^*(0,x,\widehat{m})$, obtained by minimizing the objective function j under the dynamics (2.5) with $\widetilde{m}_\tau := \widehat{m}_\tau$, will generate the same probability measure $\widehat{m}$ *via* (2.5) with $t = 0$, namely

$$\widehat{m}_s = X_s^{0,\widehat{m},*}\#m_0, \tag{2.8}$$

[1]The unique existence of the optimal control can be guaranteed in the framework that we consider, see Proposition 4.6 for more details.

where the solution map $X_s^{t,\widehat{m},*}$ is defined by solving

$$X_s^{t,\widehat{m},*}(x) = \ x + \int_t^s f\left(X_\tau^{t,\widehat{m},*}(x), \widehat{m}_\tau, \alpha_\tau^*(t,x,\widehat{m})\right) d\tau, \ \ s \in [t,T]. \tag{2.9}$$

Furthermore, we shall also solve for the corresponding value function

$$v(t,x) := j(t,x,\widehat{m},\alpha^*(t,x,\widehat{m})), \ \text{for any } (t,x) \in [0,T] \times \mathbb{R}^{d_x}. \tag{2.10}$$

Our work contributes to an important area of research in mean field games by presenting a new set of sufficient conditions for the global-in-time unique solvability of mean field games; indeed, we shall study the global-in-time well-posedness of the generic first-order mean field game (2.5)–(2.10) under Assumptions (**a1**)-(**a3**) and Hypotheses (**h1**)-(**h3**), to be specified in Sections 4.1 and 7.1. In fact, we shall study first-order mean field game (2.5)–(2.10) by analyzing the following first-order master equation (2.11) and its associated FBODE (2.12): for $(t,x,m) \in [0,T] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x})$,

$$\begin{cases} \partial_t V(t,x,m) + f\Big(x,m,\widehat{\alpha}\big(x,m,\partial_x V(t,x,m)\big)\Big) \cdot \partial_x V(t,x,m) + g\Big(x,m,\widehat{\alpha}\big(x,m,\partial_x V(t,x,m)\big)\Big) \\ \qquad\qquad\qquad + \int_{\mathbb{R}^{d_x}} \partial_m V(t,x,m)(y) \cdot f\Big(y,m,\widehat{\alpha}\big(y,m,\partial_x V(t,y,m)\big)\Big) dm(y) = 0, \\ V(T,x,m) = k(x,m), \end{cases} \tag{2.11}$$

where $\widehat{\alpha}(x,\mu,z)$ satisfies the first-order condition (5.4) and

$$\begin{cases} \dfrac{d}{ds} X_s^{t,m}(x) = f\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big), \ \ s \in [t,T], \\ \quad X_t^{t,m}(x) = x, \\ \dfrac{d}{ds} Z_s^{t,m}(x) = -\partial_x f\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big) \cdot Z_s^{t,m}(x) \\ \qquad\qquad - \partial_x g\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big), \ \ s \in [t,T], \\ \quad Z_T^{t,m}(x) = \partial_x k\big(X_T^{t,m}(x), X_T^{t,m}\#m\big); \end{cases} \tag{2.12}$$

The relationships between these components are further elaborated in Section 5.

3. Derivatives in Wasserstein metric space

We first list a few useful definitions related to the derivatives of functions in measure arguments.

Definition 3.1. (Linear Functional Derivative [4]). A function $f : \mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto f(\mu) \in \mathbb{R}$ is said to have a *linear functional derivative* if there exists a function

$$\frac{\delta f}{\delta \mu} : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$$

which is continuous with respect to the product topology of $\mathcal{P}_2(\mathbb{R}^d)$ and $\mathbb{R}^d$ so that the following holds:

(i) there is another function $c : \mathcal{P}_2(\mathbb{R}^d) \to [0,\infty)$ which is bounded on any bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$, such that

$$\left|\frac{\delta f}{\delta \mu}(\mu)(x)\right| \leq c(\mu)(1+|x|^2), \ \text{for any } \mu \in \mathcal{P}_2(\mathbb{R}^d) \text{ and } x \in \mathbb{R}^d; \tag{3.1}$$

(ii) for all ν_1, $\nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$, it also holds that

$$f(\nu_1) - f(\nu_2) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta f}{\delta \mu}(\theta\nu_1 + (1-\theta)\nu_2)(x) d(\nu_1 - \nu_2)(x) d\theta. \tag{3.2}$$

Note that $\frac{\delta f}{\delta \mu}$ defined here is unique only up to an additive constant, and the following normalization condition is taken in the rest of this article:

$$\int_{\mathbb{R}^d} \frac{\delta f}{\delta \mu}(\mu)(x) d\mu(x) = 0, \tag{3.3}$$

which in turn ensures the functional derivative of a constant function is always zero.

Let $\mathcal{H}$ and $\mathcal{H}'$ be two Hilbert spaces and $\mathcal{L}(\mathcal{H};\mathcal{H}')$ be the set of all bounded linear operators from $\mathcal{H}$ to $\mathcal{H}'$ equipped with the usual operator norm:

$$\|F\|_{\mathcal{L}(\mathcal{H};\mathcal{H}')} := \sup_{X \in \mathcal{H}, X \neq 0} \frac{\|F(X)\|_{\mathcal{H}'}}{\|X\|_{\mathcal{H}}}. \tag{3.4}$$

Definition 3.2. (*Fréchet differentiability*). A function $F : \mathcal{H} \to \mathcal{H}'$ is said to be *Fréchet differentiable* at X if there is a bounded linear operator $D_X F(X) \in \mathcal{L}(\mathcal{H};\mathcal{H}')$, such that for all $Y \in \mathcal{H}$,

$$\lim_{\|Y\|_{\mathcal{H}} \to 0} \frac{\|F(X+Y) - F(X) - D_X F(X)(Y)\|_{\mathcal{H}'}}{\|Y\|_{\mathcal{H}}} = 0.$$

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and all its square-integrable random $\mathbb{R}^d$-vectors, namely

$$\mathcal{H}^d := L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d),$$

which is equipped with an inner product:

$$\langle X, Y \rangle_{\mathcal{H}^d} := \int_\Omega X(\omega) \cdot Y(\omega) d\mathbb{P}(\omega), \text{ for all } X, Y \in \mathcal{H}^d. \tag{3.5}$$

Definition 3.3. (*L-differentiability* at a $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ [4, 43]). A function $f : \mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto f(\mu) \in \mathbb{R}$ is said to be *L-differentiable* at a $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists an $X_0 \in \mathcal{H}^d$ with the law μ_0, denoted by $\mathbb{L}_{X_0} = \mu_0$, such that its lifted version $F(X) := f(\mathbb{L}_X)$ is Fréchet differentiable at X_0; and we also call this Fréchet derivative $D_X F(X_0)(\cdot) \in \mathcal{L}(\mathcal{H}^d;\mathbb{R})$ as the *L-derivative* of $f(\mu)$ at μ_0, denote it by $\partial_\mu f(\mu_0)$.

Note that this *L-derivative* $\partial_\mu f(\mu_0)$ is uniquely defined, except $\mathbb{P}$-null set of points, in the sense that its definition is independent of the choice of X_0 corresponding to the same μ_0; also see Proposition 5.24 in [4] and [43] for details. A function f is said to be *L-differentiable* if it admits an *L-derivative* at every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ such that the associated derivative operator $\partial_\mu f(\cdot)(\cdot) : (\mu, x) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \mapsto \partial_\mu f(\mu)(x) \in \mathbb{R}^d$ is jointly measurable. Certainly, we also have

$$\partial_\mu f(\mathbb{L}_X)(X) = D_X F(X) \text{ a.s.}, \quad \forall X \in \mathcal{H}^d. \tag{3.6}$$

Next, we have the following properties for linear functional derivatives and L-derivatives.

Proposition 3.4. *(Prop. 5.51 in [4]) Suppose that $f : \mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto f(\mu) \in \mathbb{R}$ is L-differentiable on $\mathcal{P}_2(\mathbb{R}^d)$ and its Fréchet derivative $D_X F(\cdot)$ is uniformly Lipschitz in its argument. Also assume that its L-derivative*

$\partial_\mu f(\cdot)(\cdot) : (\mu, x) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \mapsto \partial_\mu f(\mu)(x) \in \mathbb{R}$ is jointly continuous with respect to the natural product topology. Then f has a linear functional derivative $\frac{\delta f}{\delta \mu}(\cdot)(\cdot)$ such that

$$\partial_\mu f(\mu)(x) = \partial_x \frac{\delta f}{\delta \mu}(\mu)(x), \text{ for any } \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{3.7}$$

Also see [43] for (3.7) for an alternative derivation.

Definition 3.5. A function $f : \mu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto f(\mu) \in \mathbb{R}$ is called regularly linear-functionally differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ if
(i) f has a linear functional derivative $\frac{\delta f}{\delta \mu}(\mu)(x)$ which, for each $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, is differentiable in x;
(ii) $\partial_x \frac{\delta f}{\delta \mu}(\cdot)(\cdot) : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ is jointly continuous with respect to the natural product topology such that

$$\left| \partial_x \frac{\delta f}{\delta \mu}(\mu)(x) \right| \leq c(\mu)(1 + |x|), \tag{3.8}$$

where $c(\mu)$ is another function in μ being bounded on any bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$.

To avoid ambiguity, we simply call f to be regularly differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. The derivative $\partial_x \frac{\delta f}{\delta \mu}$: $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ is sometimes called the *intrinsic derivative*, see [3] for example.

Proposition 3.6. *(Props. 5.44 and 5.48 in [4]) Suppose that $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is regularly differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Then, f is L-differentiable such that (3.7) holds. Furthermore, for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, and any sequence μ_n with $W_2(\mu_n, \mu) \to 0$ as $n \to \infty$, we also have*

$$\frac{f(\mu_n) - f(\mu) - \int_{\mathbb{R}^d} \frac{\delta f}{\delta \mu}(\mu)(x) d(\mu_n - \mu)(x)}{W_2(\mu_n, \mu)} \to 0. \tag{3.9}$$

Having laid down the basic principles, we can now shift our focus to the first-order mean field game described by equations (2.5)–(2.10). Within this context, the push-forward mapping $X(\cdot)$, that resides in L_m^{2,d_x} and is governed by (2.12), is deterministic. This deterministic nature is due to the absence of any additional randomness, with the only source of variability being carried over from the initial data.

4. Assumptions and preliminary results

4.1. Assumptions

We aim at solving the first-order mean field game (2.5)–(2.10) under the following assumptions.
(**a1**) The drift function $f : (x, \mu, a) \in \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_a} \mapsto \mathbb{R}^{d_x}$ is differentiable in $x \in \mathbb{R}^{d_x}$ and $a \in \mathbb{R}^{d_a}$; and $f(x, \mu, a)$, $\partial_x f(x, \mu, a)$ and $\partial_a f(x, \mu, a)$ are also regularly differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})$, as well as all of the following derivatives exist and are jointly continuous in their corresponding arguments, and they also satisfy the following estimates, respectively:
(i) For any $\xi = (\xi_1, ..., \xi_{d_x}) \in \mathbb{R}^{d_x}$,

$$\sum_{i=1}^{d_x} \sum_{j=1}^{d_x} \left(\xi_i \partial_a f_i(x, \mu, a) \right) \cdot \left(\xi_j \partial_a f_j(x, \mu, a) \right) \geq \lambda_f |\xi|^2, \tag{4.1}$$

where f_i is the i-th component of f.

(ii) Boundedness on first-order derivatives:

$$\begin{aligned}
&\sup_{(x,\mu,a)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}} \|\partial_a f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a};\mathbb{R}^{d_x})} \leq \Lambda_1;\\
&\sup_{(x,\mu,a,\widetilde{x})\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}\times\mathbb{R}^{d_x}} \|\partial_\mu f(x,\mu,a)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \leq \Lambda_2;\\
&\sup_{(x,\mu,a)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}} \|\partial_x f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \leq \Lambda_3;
\end{aligned} \tag{4.2}$$

also define $\Lambda_f := \max\{\Lambda_1,\Lambda_2,\Lambda_3\}$.
(iii) Boundedness on second-order derivatives:

$$\begin{aligned}
\sup_{(x,\mu,a,\widetilde{x})\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}\times\mathbb{R}^{d_x}} &\Big\{\|\partial_a\partial_a f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_a};\mathbb{R}^{d_x})}\\
&\vee \|\partial_a\partial_x f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_a};\mathbb{R}^{d_x})}\\
&\vee \|\partial_\mu\partial_a f(x,\mu,a)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\\
&\vee \|\partial_x\partial_x f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\\
&\vee \|\partial_\mu\partial_x f(x,\mu,a)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\Big\}\cdot(1+|x|+\|\mu\|_1) \leq \bar{l}_f.
\end{aligned} \tag{4.3}$$

(iv) The above second-order derivatives are jointly Lipschitz continuous in their corresponding arguments with Lipschitz constant decaying at the proximity of infinity: for any x, x', $\widetilde{x}$, $\widetilde{x}'\in\mathbb{R}^{d_x}$, a, $a'\in\mathbb{R}^{d_a}$, μ, $\mu'\in\mathcal{P}_2(\mathbb{R}^{d_x})$,

$$\begin{aligned}
a)\ &\big\|\partial_\mu\partial_x f(x,\mu,a)(\widetilde{x}) - \partial_\mu\partial_x f(x',\mu',a')(\widetilde{x}')\big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\\
\leq &\frac{\mathrm{Lip}_f}{1+\max\{|x|,|x'|\}+\max\{\|\mu\|_1,\|\mu'\|_1\}}\Big(|x-x'|+W_2(\mu,\mu')+|a-a'|+|\widetilde{x}-\widetilde{x}'|\Big),
\end{aligned} \tag{4.4}$$

$$\begin{aligned}
b)\ &\big\|\partial_a\partial_a f(x,\mu,a) - \partial_a\partial_a f(x',\mu',a')\big\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_a};\mathbb{R}^{d_x})}\\
\leq &\frac{\mathrm{Lip}_f}{1+\max\{|x|,|x'|\}+\max\{\|\mu\|_1,\|\mu'\|_1\}}\Big(|x-x'|+W_2(\mu,\mu')+|a-a'|\Big),
\end{aligned} \tag{4.5}$$

and so are for $\partial_\mu\partial_a f(x,\mu,a)(\widetilde{x})$, $\partial_a\partial_x f(x,\mu,a)$ and $\partial_x\partial_x f(x,\mu,a)$.

Here λ_f, Λ_1, Λ_2, Λ_3, $\bar{l}_f$ and Lip_f are some positive finite constants, and $\|\cdot\|_{\mathcal{L}(\mathcal{A};\mathcal{B})}$ is the operator norm of a linear mapping from $\mathcal{A}$ to $\mathcal{B}$.
(**a2**) The running cost function $g:(x,\mu,a)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}\mapsto\mathbb{R}$ is differentiable in $x\in\mathbb{R}^{d_x}$ and $a\in\mathbb{R}^{d_a}$; and $g(x,\mu,a)$, $\partial_x g(x,\mu,a)$ and $\partial_a g(x,\mu,a)$ are regularly differentiable in $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$, as well as all the second-order derivatives below are jointly Lipschitz continuous in their respective arguments, and they also satisfy the following estimates:
(i) For all $\eta\in\mathbb{R}^{d_a}$, any $x\in\mathbb{R}^{d_x}$, $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and $a\in\mathbb{R}^{d_a}$,

$$\eta^\top\partial_a\partial_a g(x,\mu,a)\eta \geq \lambda_g|\eta|^2. \tag{4.6}$$

(ii) For all $\xi\in\mathbb{R}^{d_x}$, any $x\in\mathbb{R}^{d_x}$, $a\in\mathbb{R}^{d_a}$, μ, $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and any X, $\widetilde{X}\in L^{2,d_x}_m$,

$$a)\ \xi^\top\partial_x\partial_x g(x,\mu,a)\xi \geq \lambda_g|\xi|^2 > 0, \tag{4.7}$$

$$b)\ \int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}} (\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widehat{x})\cdot\partial_\mu)g(x,\mu,a)\Big|_{x=X(\widetilde{x}),\mu=X\#m}\big(X(\widehat{x})\big)dm(\widehat{x})dm(\widetilde{x}) \geq -l_g\left\|\widetilde{X}\right\|^2_{L^{2,d_x}_m}. \tag{4.8}$$

(iii) We have the following bounds on the second-order derivatives of g:

$$a)\quad \sup_{(x,\mu,a,\widetilde{x})\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}\times\mathbb{R}^{d_x}} \Big\{\|\partial_a\partial_a g(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_a};\mathbb{R})} \vee \|\partial_\mu\partial_x g(x,\mu,a)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} \vee \|\partial_x\partial_x g(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})}\Big\} \le \Lambda_g, \tag{4.9}$$

$$b)\quad \sup_{(x,\mu,a,\widetilde{x})\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}\times\mathbb{R}^{d_x}} \Big\{\|\partial_x\partial_a g(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_x};\mathbb{R})} \vee \|\partial_\mu\partial_a g(x,\mu,a)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_x};\mathbb{R})}\Big\} \le \bar{l}_g. \tag{4.10}$$

Here λ_g, Λ_g, $\bar{l}_g$, λ_1 and λ_2 are some positive finite constants, and l_g is a constant which can be negative.
(a3) The terminal cost function $k:(x,\mu)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\mapsto\mathbb{R}$ is differentiable in $x\in\mathbb{R}^{d_x}$, and $k(x,\mu)$ and $\partial_x k(x,\mu)$ are regularly differentiable in $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$, as well as all the second-order derivatives below are jointly Lipschitz continuous in their respective arguments; also these derivatives satisfy the following estimates:
(i) For all $\xi\in\mathbb{R}^{d_x}$, any $x\in\mathbb{R}^{d_x}$, μ, $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and any X, $\widetilde{X}\in L^{2,d_x}_m$,

$$a)\ \xi^\top\partial_x\partial_x k(x,\mu)\xi \ge \lambda_k|\xi|^2 > 0, \tag{4.11}$$

$$b)\ \int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}} (\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widehat{x})\cdot\partial_\mu)k(x,\mu)\Big|_{x=X(\widetilde{x}),\mu=X\#m}\big(X(\widehat{x})\big)dm(\widehat{x})dm(\widetilde{x}) \ge -l_k\left\|\widetilde{X}\right\|^2_{L^{2,d_x}_m}. \tag{4.12}$$

(ii) We have the following bounds on the second-order derivatives of k:

$$\sup_{(x,\mu,\widetilde{x})\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})} \Big\{\|\partial_\mu\partial_x k(x,\mu)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} \vee \|\partial_x\partial_x k(x,\mu)\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})}\Big\} \le \Lambda_k. \tag{4.13}$$

Here λ_k and Λ_k are some positive finite constants, and l_k is a constant which can be negative.

Remark 4.1. Compared with the assumptions on coefficient functions f, g and k in our previous paper [52], we do not require the existence of $\partial_\mu\partial_\mu f(x,\mu,a)$, $\partial_\mu\partial_\mu g(x,\mu,a)$ and $\partial_\mu\partial_\mu k(x,\mu)$ and their bounds. This is due to the absence of the terms $\partial_\mu f(x,\mu,a)$ and $\partial_\mu g(x,\mu,a)$ on the right-hand side of the FBODE system (6.1) in comparison with (7.1) of [52]. In terms of problem setting, this is due to the fact that the mean-field term in the underlying mean-field game problem represents the probability measure of all other agents' states, which is an exogenous measure and remains unchanged when we perturb the control variable in the proof of Pontryagin's maximum principle.

To establish the existence of classical solutions, it is essential to impose sufficient regularity assumptions on the coefficient functions f, g, and k. While these regularity requirements may not represent the weakest possible constraints, they effectively circumvent overly technical considerations regarding smoothness while facilitating crucial *a priori* estimates. The decay rate assumptions $(\mathbf{a1})(\mathbf{iii})--(\mathbf{iv})$ imposed on the drift function f at infinity find strong justification in nonlinear dynamical systems theory. To illustrate, consider a prototypical example of a drift term exhibiting quadratic growth: the one-dimensional ordinary differential equation $\mathrm{d}X(t)/\mathrm{d}t = X(t)^2$. Solutions with positive initial data must blow-up in finite time. Therefore, to guarantee the global-in-time existence, it is natural to impose decay constraints on the second-order derivatives of f at infinity. Our theoretical framework permits f to be predominantly linear in nature, with nonlinear perturbations that are appropriately controlled – specifically requiring these perturbations to exhibit sufficiently rapid decay at infinity.

To elucidate our structural assumptions, consider the following class of coefficient functions representing nonlinear perturbations of the linear-quadratic setting:

$$f(x,\mu,a) := M_1 x + M_2 a + M_3\,\mathbb{E}(\mu) + \varepsilon_1 f_0\Big(x, \int_{\mathbb{R}^{d_x}} \phi_0(y)d\mu(y), a\Big), \tag{4.14}$$

$$\begin{aligned} g(x,\mu,a) := {} & a^\top M_4 a + x^\top M_5 x + x^\top M_6\,\mathbb{E}(\mu) + \mathbb{E}(\mu)^\top M_7\,\mathbb{E}(\mu) \\ & + \int_{\mathbb{R}^{d_x}} y^\top M_8 y\, d\mu(y) + \varepsilon_2 g_0\Big(x, \int_{\mathbb{R}^{d_x}} \phi_1(y)d\mu(y), a\Big), \end{aligned} \tag{4.15}$$

$$\begin{aligned} k(x,\mu) := {} & x^\top M_9 x + x^\top M_{10}\,\mathbb{E}(\mu) + \mathbb{E}(\mu)^\top M_{11}\,\mathbb{E}(\mu) \\ & + \int_{\mathbb{R}^{d_x}} y^\top M_{12} y\, d\mu(y) + \varepsilon_3 k_0\Big(x, \int_{\mathbb{R}^{d_x}} \phi_2(y)d\mu(y)\Big), \end{aligned} \tag{4.16}$$

where $\mathbb{E}(\mu) := \int_{\mathbb{R}^{d_x}} y\, d\mu(y) \in \mathbb{R}^{d_x}$, $f_0(x,w,a)$, $g_0(x,w,a)$ and $k_0(x,w)$ are functions of $x \in \mathbb{R}^{d_x}$, $w = (w^{(1)},\dots,w^{(d_w)}) \in \mathbb{R}^{d_w}$ and $a \in \mathbb{R}^{d_a}$, and $\phi_i(y) = (\phi_i^{(1)}(y),\dots,\phi_i^{(d_w)}(y)) \in \mathbb{R}^{d_w}$, $i = 0,1,2$, are functions of $y \in \mathbb{R}^{d_x}$, and M_j, $j = 1,\dots,10$ and ε_i, $i = 1,2,3$ are finite constant matrices. Without loss of generality, we impose symmetry requirements on M_4, M_5, M_7, M_8, M_9, M_{11} and M_{12}. Subsequently, we present verifiable sufficient conditions for Assumptions (**a1**)-(**a3**) applicable to coefficient functions f, g and k of type (4.14)-(4.16). These conditions prioritize practical verification over theoretical minimality.

Sufficient conditions for (**a1**): $M_2 M_2^\top$ is a $\mathbb{R}^{d_x} \times \mathbb{R}^{d_x}$ positive definite matrix, ε_1 is suitably small, $f_0 \in C^3(\mathbb{R}^{d_x} \times \mathbb{R}^{d_w} \times \mathbb{R}^{d_a}; \mathbb{R}^{d_x})$ with bounded first order derivatives, $\phi_0 \in C^{1,1}(\mathbb{R}^{d_x}; \mathbb{R}^{d_w})$ with bounded and uniformly Lipschitz continuous first order derivatives, and there exists a positive constant C such that, for all $x \in \mathbb{R}^{d_x}$, $\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})$, $a \in \mathbb{R}^{d_a}$ and $w = \int_{\mathbb{R}^{d_x}} \phi_0(y)d\mu(y)$,

$$\Big\|\nabla^2_{(x,w,a)} f_0(x,w,a)\Big\|_{l^\infty}, \Big\|\nabla^3_{(x,w,a)} f_0(x,w,a)\Big\|_{l^\infty} \le \frac{C}{1+|x|+\|\mu\|_1} \tag{4.17}$$

where $\|M\|_{l^\infty}$ is the supremum norm of M for those M composed of a finite number of elements. A sufficient condition for (4.17) is that, for all $x \in \mathbb{R}^{d_x}$, $w \in \mathbb{R}^{d_w}$, $a \in \mathbb{R}^{d_a}$, there exist $i \in \{1,\dots,d_w\}$ and $\beta \in \mathbb{R}$ such that

$$\begin{aligned} & \Big\|\nabla^2_{(x,w,a)} f_0(x,w,a)\Big\|_{l^\infty}, \Big\|\nabla^3_{(x,w,a)} f_0(x,w,a)\Big\|_{l^\infty} \le \frac{C}{1+|x|+|w^{(i)}|^\beta} \\ & \text{and } \left|\int_{\mathbb{R}^{d_x}} \phi_0^{(i)}(y)d\mu(y)\right|^\beta \ge \|\mu\|_1 \text{ for all } \mu \in \mathcal{P}_2(\mathbb{R}^{d_x}); \end{aligned} \tag{4.18}$$

for example, for any $p \ge 1$, if we choose $\beta := 1/p$ and $\phi_0^{(i)}(y) := |y|^p$, then $\Big(\int_{\mathbb{R}^{d_x}} \phi_0^{(i)}(y)d\mu(y)\Big)^{1/p} \ge \|\mu\|_1$ for all $\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})$.

Sufficient conditions for (**a2**): $g_0 \in C^{2,1}(\mathbb{R}^{d_x} \times \mathbb{R}^{d_w} \times \mathbb{R}^{d_a})$ with bounded second order derivatives; $\phi_1 \in C^{1,1}(\mathbb{R}^{d_x}; \mathbb{R}^{d_w})$ with bounded first order derivatives; ε_2 is suitably small; M_4 and M_5 are positive definite, to ensure the validity of (4.6) and (4.7) respectively; M_6 is positive definite, to ensure the validity of (4.8), since

$$\begin{aligned} & \int_{\mathbb{R}^{d_x}} \int_{\mathbb{R}^{d_x}} (\widetilde{X}(\widetilde{x}) \cdot \partial_x)(\widetilde{X}(\widehat{x}) \cdot \partial_\mu) g(x,\mu,a)\Big|_{x=X(\widetilde{x}),\mu=X\#m} \big(X(\widehat{x})\big) dm(\widehat{x}) dm(\widetilde{x}) \\ = {} & \left(\int_{\mathbb{R}^{d_x}} \widetilde{X}(y)dm(y)\right)^\top M_6 \left(\int_{\mathbb{R}^{d_x}} \widetilde{X}(y)dm(y)\right) + O(\varepsilon_2) \Big\|\widetilde{X}\Big\|^2_{L^{1,d_x}_m}; \end{aligned}$$

Sufficient conditions for (**a3**): $k_0 \in C^{2,1}(\mathbb{R}^{d_x} \times \mathbb{R}^{d_w})$ with bounded second order derivatives; $\phi_2 \in C^{1,1}(\mathbb{R}^{d_x}; \mathbb{R}^{d_w})$ with bounded first order derivatives; ε_3 is suitably small; M_9 and M_{10} are positive definite.

4.2. Properties of coefficient functions and the optimal control

Under Assumptions (**a1**)–(**a3**), we have the following useful properties which play a vital role in developing the general theory. All the proofs in this section are almost the same as that in Appendix of our previous paper [52], since these properties are only related to the coefficient functions themselves and the first-order condition, and are independent of the master equation or FBODE system (2.12). Therefore, we omit most details here.

Before delving into the properties of coefficient functions and controls, we first introduce a useful property regarding the Lipschitz continuity of functions defined on $\mathcal{P}_2(\mathbb{R}^{d_x})$. This property is used in deriving the properties of coefficient functions, controls, and solutions of the FBODE system (2.12).

Proposition 4.2. *For a function $h : \mu \in \mathcal{P}_2(\mathbb{R}^{d_x}) \mapsto h(\mu) \in \mathbb{R}^l$ (the dimension l can be any positive integer such as 1 or d_x) having a linear functional derivative $\frac{\delta}{\delta\mu} h(\mu)(\widetilde{x})$ which is also continuously differentiable in $\widetilde{x}$ and satisfies the linear growth of $\sup_{\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})} \left| \partial_{\widetilde{x}} \frac{\delta}{\delta\mu} h(\mu)(\widetilde{x}) \right| \leq l_1 + l_2 |\widetilde{x}|$ for some positive constants l_1 and l_2. Then we have*

$$|h(\mu) - h(\mu')| \leq l_1 W_1(\mu, \mu') + l_2 \left(\|\mu\|_2 + \|\mu'\|_2\right) W_2(\mu, \mu'). \tag{4.19}$$

Particularly, when $l_2 = 0$, we obtain

$$|h(\mu) - h(\mu')| \leq l_1 W_1(\mu, \mu'). \tag{4.20}$$

Proposition 4.3. *The drift function $f(x, \mu, a)$ is of linear-growth and Lipschitz continuous, that is, there exists a positive constant L_f, which can be taken as $L_f := \max\{\Lambda_f, |f(0, \delta_0, 0)|\}$[2], such that*

$$|f(x, \mu, a)| \leq L_f \left(1 + |x| + \|\mu\|_1 + |a|\right); \tag{4.21}$$

$$|f(x, \mu, a) - f(x', \mu', a')| \leq L_f \big(|x - x'| + W_1(\mu, \mu') + |a - a'|\big). \tag{4.22}$$

Proposition 4.4. *(Bounding first-order derivatives) We have the following estimates:*

$$|\partial_x g(x, \mu, a) - \partial_x g(x', \mu', a')| \leq \max\{\Lambda_g, \bar{l}_g\} \left(W_1(\mu, \mu') + |x - x'| + |a - a'|\right); \tag{4.23}$$

$$|\partial_a g(x, \mu, a) - \partial_a g(x', \mu', a')| \leq \max\{\Lambda_g, \bar{l}_g\} \left(W_1(\mu, \mu') + |x - x'| + |a - a'|\right); \tag{4.24}$$

$$\begin{aligned} &\sup_{x \in \mathbb{R}^{d_x},\, \mu \in \mathcal{P}_2(\mathbb{R}^{d_x}),\, a \in \mathbb{R}^{d_a}} \frac{|\partial_x g(x, \mu, a)|}{1 + |x| + \|\mu\|_1 + |a|} \vee \frac{|\partial_a g(x, \mu, a)|}{1 + |x| + \|\mu\|_1 + |a|} \\ &\leq L_g := \max\Big\{|\partial_x g(0, \delta_0, 0)|, |\partial_a g(0, \delta_0, 0)|, \Lambda_g, \bar{l}_g\Big\}; \end{aligned} \tag{4.25}$$

and

$$\sup_{x \in \mathbb{R}^{d_x},\, \mu \in \mathcal{P}_2(\mathbb{R}^{d_x})} \frac{|\partial_x k(x, \mu)|}{1 + |x| + \|\mu\|_1} \leq L_k := \max\Big\{|\partial_x k(0, \delta_0)|, \Lambda_k\Big\}. \tag{4.26}$$

[2] δ_0 is Dirac measure at 0.

Proposition 4.5. *For each fixed $(x,\mu,z)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$, we have* $\lim\limits_{|a|\to\infty} f(x,\mu,a)\cdot z+g(x,\mu,a)\to\infty$.

Define a constant

$$k_0:=4\max\{\Lambda_k,L_0^*\}>0, \tag{4.27}$$

where L_0^* will be defined later in (7.56) and it is a positive constant depending only on λ_f, Λ_1, Λ_2, Λ_3, λ_g, Λ_g, l_g, λ_k, Λ_k and l_k. In fact, we can show later in Remark 7.7 that $\Lambda_k<L_0^*$ and thus $k_0=4L_0^*$. We further define a cone set:

$$c_{k_0}:=\left\{(x,\mu,z)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}:|z|\le\frac{1}{2}k_0\left(1+|x|+\|\mu\|_1\right)\right\}, \tag{4.28}$$

for the constant $k_0>0$ as defined in (4.27); note that this cone needs not to be convex at all. In fact, Theorem 6.3 addresses the local existence of solution over small time interval, while Theorem 7.8 resolves the global existence of solution over an arbitrarily long time horizon. Both theorems ensure that $\left(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)\right)\in c_{k_0}$, where $\left(X_s^{t,m}(x),Z_s^{t,m}(x)\right)$ is the solution pair of the FBODE system (2.12). This property is crucial in ensuring the unique existence of the optimal control $\widehat{\alpha}(x,\mu,z)$ stated in Proposition 4.6. It is worth noting that this additional growth confinement, which will be stated in the proposition's hypothesis, is indeed justifiable.

Proposition 4.6. *Suppose further that*

$$\sup_{(x,\mu,a)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}}\|\partial_a\partial_a f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_a};\mathbb{R}^{d_x})}\cdot(1+|x|+\|\mu\|_1)\le\frac{1}{40\max\{\Lambda_k,L_0^*\}}\lambda_g. \tag{4.29}$$

Then, for each $(x,\mu,z)\in c_{k_0}$, the function $f(x,\mu,a)\cdot z+g(x,\mu,a)$ has a unique minimizer $\widehat{\alpha}(x,\mu,z)\in\mathbb{R}^{d_a}$, which satisfies the first-order condition

$$\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a g(x,\mu,\widehat{\alpha})=0, \tag{4.30}$$

and it is first-order differentiable in x, $z\in\mathbb{R}^{d_x}$ and L-differentiable in $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$ with jointly Lipschitz continuous first-order derivatives such that

$$\partial_x\widehat{\alpha}(x,\mu,z)=\ -\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha})\Big)^{-1}\left(\partial_x\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_x\partial_a g(x,\mu,\widehat{\alpha})\right), \tag{4.31}$$

$$\partial_z\widehat{\alpha}(x,\mu,z)=\ -\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha})\Big)^{-1}\partial_a f(x,\mu,\widehat{\alpha}), \tag{4.32}$$

$$\partial_\mu\widehat{\alpha}(x,\mu,z)(\widetilde{x})=\ -\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha})\Big)^{-1}\left(\partial_\mu\partial_a f(x,\mu,\widehat{\alpha})(\widetilde{x})\cdot z+\partial_\mu\partial_a g(x,\mu,\widehat{\alpha})(\widetilde{x})\right). \tag{4.33}$$

In addition, we have the following estimates:
(i) for an arbitrary $\xi\in\mathbb{R}^{d_x}$,

$$\Big(\Lambda_g+\frac{1}{20}\lambda_g\Big)|\xi|^2\ge\xi^\top\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha})\Big)\xi\ge\frac{19}{20}\lambda_g|\xi|^2; \tag{4.34}$$

(ii) for all $(x,\mu,z)\in c_{k_0}$,

$$\Big\|\partial_x\widehat{\alpha}(x,\mu,z)\Big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\vee\Big\|\partial_\mu\widehat{\alpha}(x,\mu,z)(\widetilde{x})\Big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\le\frac{20\big(\bar{l}_g+\frac{1}{2}k_0\cdot\bar{l}_f\big)}{19\lambda_g}, \tag{4.35}$$

$$\left\|\partial_z\widehat{\alpha}(x,\mu,z)\right\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\le\frac{20\Lambda_f}{19\lambda_g};\tag{4.36}$$

(iii) for any $(x,\mu,z)\in c_{k_0}$*, and any* $(x',\mu',z')\in c_{k_0}$ *such that* $\theta(x,\mu,z)+(1-\theta)(x',\mu',z')\in c_{k_0}$ *for all* $\theta\in[0,1]$*,*

$$|\widehat{\alpha}(x,\mu,z)-\widehat{\alpha}(x',\mu',z')|\le\ L_\alpha\big(|x-x'|+W_1(\mu,\mu')+|z-z'|\big),\tag{4.37}$$
$$|\widehat{\alpha}(x,\mu,z)|\le\ L_\alpha\cdot(1+|x|+\|\mu\|_1+|z|),\tag{4.38}$$

where

$$L_\alpha:=\max\Big\{\frac{20\Lambda_f}{19\lambda_g},\frac{20\big(\bar{l}_g+\frac{1}{2}k_0\cdot\bar{l}_f\big)}{19\lambda_g},\widehat{\alpha}(0,\delta_0,0)\Big\};\tag{4.39}$$

(iv) for any function $\Gamma:(x,\mu)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\mapsto\Gamma(x,\mu)\in\mathbb{R}^{d_x}$ *which is differentiable in* x *and is L-differentiable in* μ *with jointly continuous derivatives such that*

$$\left\|\partial_x\Gamma(x,\mu)\right\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\vee\left\|\partial_\mu\Gamma(x,\mu)(\widetilde{x})\right\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le L_\Gamma$$

for some positive constant $L_\Gamma\le\frac{1}{2}k_0$ *(recall (4.28)), then,* $|\Gamma(x,\mu)|\le L_\Gamma(1+|x|+\|\mu\|_1)$ *and*

$$|\widehat{\alpha}(x,\mu,\Gamma(x,\mu))-\widehat{\alpha}(x',\mu',\Gamma(x',\mu'))|\le L_\alpha(1+L_\Gamma)(|x-x'|+W_1(\mu,\mu')).\tag{4.40}$$

Proposition 4.7. *For the cone* c_{k_0} *defined in (4.28) and any* $\xi\in\mathbb{R}^{d_x}$*, and suppose also that (4.29) is valid, then we have the following inequality:*

$$\sup_{(x,\mu,z)\in c_{k_0}}\xi^\top\Big(\sum_{i=1}^{d_a}\partial_z\widehat{\alpha}_i(x,\mu,z)\otimes\partial_{a_i}f(x,\mu,\widehat{\alpha})\Big)\xi\le-\frac{\lambda_f^2}{\Lambda_g+\lambda_g/20}|\xi|^2,\tag{4.41}$$

where $\widehat{\alpha}(x,\mu,z)$ *is the unique minimizer solving the first-order condition (4.30) and we denote* $a\otimes b:=(a_ib_j)_{n\times m}$*, the tensor product of* a *and* b*, the* $n\times m$ *matrix for arbitrary vectors* $a\in\mathbb{R}^n$ *and* $b\in\mathbb{R}^m$*.*

Our key to handling nonlinear dynamics is the newly proposed "cone condition": while solving for the unique minimizer $\hat{\alpha}(x,\mu,z)$ of $f(x,\mu,a)\cdot z+g(x,\mu,a)$, we only have to restrict our attention to those (x,μ,z) belonging to the cone c_{k_0}, where the cone c_{k_0} was defined in (4.28); in other words, only the restriction of the Hamiltonian h to the cone c_{k_0} is relevant to the underlying first-order mean field game. We shall show in Theorems 6.3, 7.5 and 7.8 that this is justified because the solution pair $\big(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)\big)$ of the FBODE system (2.12) always satisfies the cone property

$$\big(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)\big)\in c_{k_0}.$$

This concept has not been introduced, studied or applied in the existing MFG literature. Its development is highly nontrivial because, in nonlinear settings, the second-order derivatives of Hamiltonian cannot be uniformly bounded over the entire space $(x,\mu,z)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$. This challenge persists even when the dynamics are reduced to a linear system with elementary nonlinear perturbations, and we shall illustrate this complicated issue by using a simple and explicit example at the end of this subsection.

Recently, Bansil-Mészáros [66] and Bansil-Mészáros-Mou [67] made some developments on MFG and master equations in the direction of displacement monotonicity, but these results still rely on the assumption that the

Hamiltonian $h : (x,\mu,z) \in \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x} \to \min_a\{f(x,\mu,a)\cdot z + g(x,\mu,a)\} \in \mathbb{R}$ has uniformly bounded second order derivatives. It means that, in particular, there exists a constant $C_h > 0$ such that

$$\begin{aligned}\big\|\partial_x\partial_x h(x,\mu,z)\big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} &= \Big\|\partial_x\partial_x f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z + \partial_x\widehat{\alpha}(x,\mu,z)^\top\partial_a\partial_x f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z \\ &\quad + \partial_x\widehat{\alpha}(x,\mu,z)^\top\partial_a\partial_x g(x,\mu,\widehat{\alpha}(x,\mu,z)) + \partial_x\partial_x g(x,\mu,\widehat{\alpha}(x,\mu,z))\Big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} \\ &\le C_h, \text{ for all } (x,\mu,z) \in \mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x},\end{aligned} \tag{4.42}$$

where $\widehat{\alpha}(x,\mu,z) \in \arg\min_a\{f(x,\mu,a)\cdot z + g(x,\mu,a)\}$.

In order to provide a simple and explicit illustration of the limitations of this uniform boundedness assumption on the second-order derivatives, let us consider a special kind of elementary perturbations of LQ problems as follows: suppose that f and g are in the form of (4.14)–(4.15) with the perturbations f_0 and g_0 independent of a, namely $f_0 := f_0(x,w)$ and $g_0 := g_0(x,w)$. Then the uniform boundedness assumption (4.42) becomes

$$\begin{aligned}\big\|\partial_x\partial_x h(x,\mu,z)\big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} &= \Big\|\varepsilon_1\partial_x\partial_x f_0\Big(x,\int_{\mathbb{R}^{d_x}}\phi_0(y)d\mu(y)\Big)\cdot z \\ &\quad + M_5 + \varepsilon_2\partial_x\partial_x g_0\Big(x,\int_{\mathbb{R}^{d_x}}\phi_1(y)d\mu(y)\Big)\Big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} \\ &\le C_h, \text{ for all } (x,\mu,z) \in \mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}.\end{aligned} \tag{4.43}$$

Since $(\mu,z) \in \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x}$ can be arbitrarily chosen, the inequality (4.43) implies that $\partial_x\partial_x f_0(x,w) = 0$ for all $x \in \mathbb{R}^{d_x}$ and $w \in \text{Range}(\phi_0)$. The same argument applies to $\big\|\partial_x\partial_\mu h(x,\mu,z)\big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})}$, $\big\|\partial_\mu\partial_\mu h(x,\mu,z)\big\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} \le C_h$, and hence $\partial_x\partial_w f_0(x,w) = \partial_w\partial_w f_0(x,w) = 0$ for all $x \in \mathbb{R}^{d_x}$ and $w \in \text{Range}(\phi_0)$ as well. Therefore, results based on the uniform boundedness assumption on the second-order derivatives of Hamiltonian only apply to linear $f_0(x,w)$ and are not applicable even to cases of linear dynamics with nonlinear perturbations, unless those nonlinearities occur exclusively for w outside Range(ϕ_0)—a region that is irrelevant to the underlying optimal control problem.

If one disregards the problem of existence and uniqueness of minimizers $\widehat{\alpha}(x,\mu,z)$ in the case of nonlinear dynamics, and assume that the Hamiltonian $h(x,\mu,z)$ is always well-defined in $\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$ and is twice continuously differentiable with uniformly bounded second order derivatives, then the examples for g and k provided in (4.15) and (4.16), respectively, are displacement semi-monotone.

4.3. Discussion on our assumptions

In this subsection, we aim to provide motivation behind and explanation of Assumptions (**a1**)-(**a3**) and the property (4.41) by comparing them with the coercivity and monotonicity conditions used in Cardaliaguet–Delarue–Lasry–Lions' monograph [3]. We shall also refer to Remark 7.2 to compare our monotonicity condition with the displacement monotonicity for nonseparable Hamiltonians introduced in Gangbo, Mészáros, Mou, and Zhang [30].

To simplify calculus and computations, we consider a special case where the drift function is simply $f(x,\mu,a) = a$ and the running cost function is $g(x,\mu,a) = \lambda a^2 + g_1(x,\mu)$, with a positive constant λ. Note that two more complicated examples are presented in Section 9, which is not in this special form. Under this assumption, we find that (4.1), (4.2), (4.3), and (4.10) are automatically satisfied with $\lambda_f = 1$, $\Lambda_f = 1$, and $\bar{l}_f = \bar{l}_g = 0$. Additionally, we can explicitly solve for the optimal control $\widehat{\alpha}$ using the first-order condition (4.30) as $\widehat{\alpha}(x,\mu,z) = -\frac{1}{2\lambda}z$. Therefore, the left-hand side of (4.41) equals to $-\frac{1}{2\lambda}|\xi|^2$. It is worth noting that the term $\Big(\sum_{i=1}^{d_a}\partial_z\widehat{\alpha}_i(x,\mu,z)\otimes\partial_{a_i}f(x,\mu,\widehat{\alpha})\Big)$ in (4.41) corresponds to the term $D^2_{zz}H(x,z)$ in (2.4) in [3]. Furthermore, there is no mean field term μ in $H(x,z)$ because the separable structure of the Hamiltonian was assumed in [3], however we can also treat the nonseparable case in the present work. Thus, (4.41) corresponds to the coercivity condition (2.4) in Cardaliaguet–Delarue–Lasry–Lions' monograph [3].

It is also worth noting that the running cost function $g(x,\mu,a)$ corresponds to the function $F(x,\mu)$ in (2.5) of [3]. However, since the separable structure of the running cost function $g(x,\mu,a)=\lambda a^2+g_1(x,\mu)$ was assumed in [3], there is no control variable a in $F(x,\mu)$. Note that, a sufficient condition to (4.8) is, for any $x\in\mathbb{R}^{d_x}$, $a\in\mathbb{R}^{d_a}$ and $\mu,\mu'\in\mathcal{P}_2(\mathbb{R}^{d_x})$,

$$\int_{\mathbb{R}^{d_x}}\Big(g(x,\mu',a)-g(x,\mu,a)\Big)d\big(\mu'-\mu\big)(x)\geq -l_g\left(\int_{\mathbb{R}} yd(\mu-\mu')(y)\right)^2 \text{ with } l_g<\lambda_g. \tag{4.44}$$

Indeed, for any $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and any $X,\ \widetilde{X}\in L^{2,d_x}_m$, by taking $\mu'=(X+\theta\widetilde{X})\#m$ with $\theta>0$, $\mu=X\#m$ and letting $\theta\to 0^+$ after dividing (4.44) by θ^2, the estimate (4.44) implies (4.8). However, condition (4.8) cannot imply (4.44) in general, since, when showing that (4.44) is a sufficient condition for (4.8), we have used the crucial fact $\left(\int_{\mathbb{R}} yd(\mu-\mu')(y)\right)^2=\left(\int_{\mathbb{R}}\theta\widetilde{X}(\widetilde{y})dm(\widetilde{y})\right)^2\leq\theta^2\left\|\widetilde{X}\right\|^2_{L^{1,d_x}_m}\leq\theta^2\left\|\widetilde{X}\right\|^2_{L^{2,d_x}_m}$ but the converse is not valid in general. Also note that the estimate (4.7) assumed in Assumption (**a2**) means that $g(x,\mu,a)$ is strictly convex in x. Moreover, (4.7) and (4.44) assumed in Assumption (**a2**) can be also replaced by

$$\begin{aligned}&\int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}(\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widetilde{x})\cdot\partial_x)g\big(x,\mu,a\big)\Big|_{x=X(\widetilde{x}),\mu=X\#m}dm(\widetilde{x})\\&+\int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}(\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widehat{x})\cdot\partial_\mu)g\big(x,\mu,a\big)\Big|_{x=X(\widetilde{x}),\mu=X\#m}\big(X(\widehat{x})\big)dm(\widehat{x})dm(\widetilde{x})\geq\lambda\left\|\widetilde{X}\right\|^2_{L^{2,d_x}_m},\end{aligned}\tag{4.45}$$

with $\lambda=\lambda_g>0$. The reasons of this replacement are as follows: i) (4.45) can imply (4.7) by taking $\widetilde{X}(\widetilde{x})=\epsilon^{-d_x}\widehat{X}(\epsilon^{-1}\widetilde{x})$ and then passing to the limit as $\epsilon\to 0^+$ after multiplying (4.45) by ϵ^{d_x} (also see Prop. B.6 in [29]); ii) while showing (7.41) in the proof of Theorem 7.5, we can directly use (4.45) instead of (4.8). On the other hand, (4.7) and (4.44) can imply (4.45) with $\lambda=\lambda_g-l_g$.

Similarly, the terminal cost function $k(x,\mu)$ corresponds to the function $G(x,\mu)$ in (2.5) of [3]. Likewise, a sufficient but not necessary condition to (4.12) is, for any $x\in\mathbb{R}^{d_x}$ and $\mu,\mu'\in\mathcal{P}_2(\mathbb{R}^{d_x})$,

$$\int_{\mathbb{R}^{d_x}}\Big(k(x,\mu')-k(x,\mu)\Big)d\big(\mu'-\mu\big)(x)\geq -l_k\left(\int_{\mathbb{R}} yd(\mu-\mu')(y)\right)^2 \text{ with } l_k<\lambda_k. \tag{4.46}$$

Indeed, for any $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and any $X,\ \widetilde{X}\in L^{2,d_x}_m$, by taking $\mu'=(X+\theta\widetilde{X})\#m$ with $\theta>0$, $\mu=X\#m$ and letting $\theta\to 0^+$ after dividing (4.46) by θ^2, the estimate (4.46) implies (4.12). Note again that the estimate (4.11) assumed in Assumption (**a3**) means that $k(x,\mu)$ is strictly convex in x. Again, (4.11) and (4.46) assumed in Assumption (**a3**) can also be replaced by

$$\begin{aligned}&\int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}(\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widetilde{x})\cdot\partial_x)k\big(x,\mu\big)\Big|_{x=X(\widetilde{x}),\mu=X\#m}dm(\widetilde{x})\\&+\int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}(\widetilde{X}(\widetilde{x})\cdot\partial_x)(\widetilde{X}(\widehat{x})\cdot\partial_\mu)k\big(x,\mu\big)\Big|_{x=X(\widetilde{x}),\mu=X\#m}\big(X(\widehat{x})\big)dm(\widehat{x})dm(\widetilde{x})\geq\lambda\left\|\widetilde{X}\right\|^2_{L^{2,d_x}_m}\end{aligned}\tag{4.47}$$

with $\lambda=\lambda_k>0$.

In summary, (4.41) has a close relationship with the coercivity condition (2.4) of the Hamiltonian in [3], while (4.44) and (4.46) have a close relationship with the monotonicity condition (2.5) in the same monograph. For the comparison with displacement monotonicity, one can refer to Remark 7.2.

5. Hamilton–Jacobi and Fokker–Planck system, master equation, forward–backward system and their relationships

It follows from the principle of dynamic programming that the value function $v(t,x)$ defined in (2.10) is expected to satisfy the Hamilton-Jacobi (HJ) equation: for $t\in[0,T]$, $x\in\mathbb{R}^{d_x}$,

$$\begin{cases}\partial_t v(t,x)+h\big(x,\widehat{m}_t,\partial_x v(t,x)\big)=0,\\ v(T,x)=k(x,\widehat{m}_T),\end{cases}\tag{5.1}$$

where the Hamiltonian $h:\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}\to\mathbb{R}$ is defined as

$$h(x,\mu,z):=\min_{a\in\mathbb{R}^{d_a}}\left(f(x,\mu,a)\cdot z+g(x,\mu,a)\right),\tag{5.2}$$

and, by (2.8), the density of the probability measure $\widehat{m}_t$ (if exists), denoted also by $\widehat{m}(t,x)$, satisfies the following Fokker-Planck (FP) equation: for $t\in[0,T]$, $x\in\mathbb{R}^{d_x}$,

$$\begin{cases}\partial_t\widehat{m}(t,x)+\operatorname{div}\Big(\partial_z h\big(x,\widehat{m}(t,x),\partial_x v(t,x)\big)\widehat{m}(t,x)\Big)=0,\\ \widehat{m}(0,x)=m_0(x).\end{cases}\tag{5.3}$$

Suppose that for any fixed $x\in\mathbb{R}^{d_x}$, $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and $z\in\mathbb{R}^{d_x}$ belonging to our interested domain[3], the function $f(x,\mu,a)\cdot z+g(x,\mu,a)$ has a unique minimizer $\widehat{\alpha}(x,\mu,z)\in\mathbb{R}^{d_a}$ and is differentiable in $a\in\mathbb{R}^{d_a}$. Then the unique minimizer $\widehat{\alpha}(x,\mu,z)$ must satisfy the usual first-order condition:

$$\partial_a f(x,\mu,\widehat{\alpha})\cdot z+\partial_a g(x,\mu,\widehat{\alpha})=0.\tag{5.4}$$

Substituting the optimal control $\widehat{\alpha}$ in (5.1)-(5.3) yields the following HJ–FP system: for $t\in[0,T]$, $x\in\mathbb{R}^{d_x}$,

$$\begin{cases}\partial_t v(t,x)+f\Big(x,\widehat{m}(t,x),\widehat{\alpha}\big(x,\widehat{m}(t,x),\partial_x v(t,x)\big)\Big)\cdot\partial_x v(t,x)+g\Big(x,\widehat{m}(t,x),\widehat{\alpha}\big(x,\widehat{m}(t,x),\partial_x v(t,x)\big)\Big)=0,\\ v(T,x)=k(x,\widehat{m}_T);\\ \partial_t\widehat{m}(t,x)+\operatorname{div}\bigg(f\Big(x,\widehat{m}(t,x),\widehat{\alpha}\big(x,\widehat{m}(t,x),\partial_x v(t,x)\big)\Big)\widehat{m}(t,x)\bigg)=0,\\ \widehat{m}(0,x)=m_0(x).\end{cases}\tag{5.5}$$

Define, for any $(t,x,m)\in[0,T]\times\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})$,

$$V(t,x,m):=j(t,x,\widehat{m},\alpha^*(t,x,\widehat{m})),\tag{5.6}$$

where the optimal control α^* is defined in (2.7), the Nash equilibrium $\widehat{m}:=(\widehat{m}_s)_{s\in[t,T]}$ is given by

$$\widehat{m}_s=X_s^{t,\widehat{m},*}\#m,\tag{5.7}$$

and the solution map $X_s^{t,\widehat{m},*}$ is defined by solving (2.9). Then, applying the principle of dynamic programming, one can show that $V(t,x,m)$ satisfies the first order master equation (2.11), also see [43] and [25]. Let

[3] More precisely, we shall show that only those (x,μ,z)'s belonging to the cone set c_{k_0} defined in (4.27) are related to the underlying mean field games, so Proposition 4.6 provides a solid and rigorous foundation for the discussion below.

$(v, \widehat{m}) := (v(t,x), \widehat{m}(t,x))$ be a smooth enough solution of the HJ–FP system (5.5) with the given initial condition $\widehat{m}(t_0, x) = \widehat{m}_{t_0}(x)$ at time $t = t_0$, then we can directly check that $V(t_0, x, \widehat{m}_{t_0}) := v(t_0, x)$ is a solution of the master equation (2.11). In other words, we can write the solution $V(t,x,m)$ of the master equation (2.11) in terms of the solution $(v, \widehat{m})$ of the HJ–FP system (5.5) by the characteristics method; also see section 1.2.4 in the monograph of Cardaliaguet, Delarue, Lasry and Lions [3].

Differentiating the master equation (2.11) with respect to x and using the first-order condition (5.4), we obtain

$$\begin{cases} \partial_t \partial_x V(t,x,m) + f\Big(x, m, \widehat{\alpha}\big(x, m, \partial_x V(t,x,m)\big)\Big) \cdot \partial_x \partial_x V(t,x,m) \\ +\partial_x f\Big(x, m, \widehat{\alpha}\big(x, m, \partial_x V(t,x,m)\big)\Big) \cdot \partial_x V(t,x,m) + \partial_x g\Big(x, m, \widehat{\alpha}\big(x, m, \partial_x V(t,x,m)\big)\Big) \\ + \int_{\mathbb{R}^{d_x}} \partial_m \partial_x V(t,x,m)(y) \cdot f\Big(y, m, \widehat{\alpha}\big(y, m, \partial_x V(t,y,m)\big)\Big) dm(y) = 0, \\ \partial_x V(T,x,m) = \partial_x k(x,m). \end{cases} \tag{5.8}$$

We can also write the solution $\partial_x V(t,x,m)$ of (5.8) in terms of the solution (X, Z) of a system of forward–backward ordinary differential equations (FBODE) defined in (2.12) by using the characteristics method and the push-forward operator $\#$ defined in Definition 2.1. Indeed, we can directly check that $\partial_x V(t,x,m) = Z_t^{t,m}(x)$ and $\partial_x V(s, X_s^{t,m}(x), X_s^{t,m}\#m) = Z_s^{t,m}(x)$. Once we solve the FBODE (2.12), then we can obtain the solution of master equation (2.11) by a direct integration:

$$V(t,x,m) := k\big(X_T^{t,m}(x), X_T^{t,m}\#m\big) + \int_t^T g\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big) ds. \tag{5.9}$$

In addition, we can obtain the optimal control process

$$\big(\alpha_t^*(0, x, X_\cdot^{0,m_0}\#m_0)\big)_{t\in[0,T]} = \Big(\widehat{\alpha}\big(X_t^{0,m_0}(x), X_t^{0,m_0}\#m_0, \partial_x V(t, X_t^{0,m_0}(x), X_t^{0,m_0}\#m_0)\big)\Big)_{t\in[0,T]}.$$

6. Local existence of forward–backward system (6.1)

For any fixed $0 \le t \le \widetilde{T} \le T$ and $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, consider the following forward–backward ordinary differential equation (FBODE) system, now with possibly a generic terminal cost functional p: for $s \in [t, \widetilde{T}]$,

$$\begin{cases} \dfrac{d}{ds} X_s^{t,m}(x) = f\left(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x))\right), \; X_t^{t,m}(x) = x, \\ \dfrac{d}{ds} Z_s^{t,m}(x) = -\partial_x f(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x))) \cdot Z_s^{t,m}(x) \\ \qquad -\partial_x g(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x))), \qquad Z_{\widetilde{T}}^{t,m}(x) = p(X_{\widetilde{T}}^{t,m}(x), X_{\widetilde{T}}^{t,m}\#m), \end{cases} \tag{6.1}$$

where the optimal control $\widehat{\alpha}(x, \mu, z)$ solves the first-order condition (5.4), and the terminal function p, which can be arbitrary now, satisfies the following assumptions:

(**P**) The terminal function $p : (x, \mu) \in \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \mapsto p(x,\mu) \in \mathbb{R}^{d_x}$ is first-order differentiable in $x \in \mathbb{R}^{d_x}$ and L-differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})$ with their first-order derivatives being Lipschitz continuous in their corresponding arguments and satisfying, for some finite positive constants L_p and $\overline{L}_p$,

$$\sup_{x\in\mathbb{R}^{d_x},\, \mu\in\mathcal{P}_2(\mathbb{R}^{d_x}),\, \widetilde{x}\in\mathbb{R}^{d_x}} \|\partial_x p(x,\mu)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \vee \|\partial_\mu p(x,\mu)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \le L_p, \tag{6.2}$$

$$|p(x,\mu)| \le \overline{L}_p \cdot (1 + |x| + \|\mu\|_1). \tag{6.3}$$

Assumption (**P**) echoes Assumption (**a3**)(ii) in Section 4.1 about the nature of the terminal data, and this structure will be preserved, upon enlarging the bound L_p gradually, while solving the FBODE over consecutive sub-intervals in a backward manner. In addition, (6.2) implies that, for any x, $x' \in \mathbb{R}^{d_x}$ and μ, $\mu' \in \mathcal{P}_2(\mathbb{R}^{d_x})$,

$$|p(x,\mu) - p(x',\mu')| \le L_p\left(|x-x'| + W_1(\mu,\mu')\right). \tag{6.4}$$

Then, clearly, we see that (6.3) is valid for any finite constant $\overline{L}_p \ge \max\{|p(0,\delta_0)|, L_p\}$. For the sake of convenience, we also define

$$L_B := L_f(1+L_\alpha+2L_pL_\alpha) \text{ and } \overline{L}_B := L_f(1+L_\alpha+2\overline{L}_pL_\alpha). \tag{6.5}$$

For local (in time) solutions of (2.12), for the last time interval, $p(x,\mu)$ is taken to be $\partial_x k(x,\mu)$ and thus, by (4.13), one can set $L_p = \Lambda_k$. In the proof for global (in time) existence Theorem 7.8, we shall glue (smoothly) consecutively together local (in time) solutions of (6.1) so as to obtain a global (in time) solution of (2.12) over $[0,T]$ of arbitrary length T; for a variety of terminal data p in different small time intervals $[t_i, t_{i+1}]$, $i = 1, 2, ..., N$, by using Theorem 7.5 to be proven later in Section 7, one can set $\overline{L}_p = \max\{\Lambda_k, L_0^*\} = L_0^*$ uniformly for various terminal data p, provided that L_0^* is to be defined in (7.56). In these cases, by (6.11), one has, for all $(s,x,m) \in [t,\widetilde{T}] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x})$, $(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)) \in c_{k_0}$ where the cone c_{k_0} was defined in (4.28) and the constant k_0 was defined in (4.27), namely $k_0 = 4\overline{L}_p$. Thus, when (4.29) is valid, by Proposition 4.6, the optimal control $\widehat{\alpha}\left(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\right)$ is well-defined and satisfies the first-order condition (4.30).

Define

$$\text{(a norm) } \|\gamma\|_{1,[t,T]} := \sup_{s\in[t,T],\, x\in\mathbb{R}^{d_x},\, \mu\in\mathcal{P}_2(\mathbb{R}^{d_x})} \frac{|\gamma(s,x,\mu)|}{1+|x|+\|\mu\|_1}, \tag{6.6}$$

$$\text{(a semi-norm) } \|\gamma\|_{2,[t,T]} := \sup_{s\in[t,T], x\in\mathbb{R}^{d_x},\, \mu\in\mathcal{P}_2(\mathbb{R}^{d_x}),\, \widetilde{x}\in\mathbb{R}^{d_x}} \|\partial_x\gamma(s,x,\mu)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \vee \|\partial_\mu\gamma(s,x,\mu)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}. \tag{6.7}$$

When there is no cause of ambiguity, we denote $\|\gamma\|_{i,[t,T]}$ by $\|\gamma\|_i$, for $i = 1, 2$, for the notational simplicity. Before stating our main results in this section, let us first recall the following concept.

Definition 6.1. (Decoupling field) We call a function $\gamma : [t,\widetilde{T}] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \to \mathbb{R}^{d_x}$ a decoupling field for the FBODE system (6.1) on $[t,\widetilde{T}]$ if γ solves the following system for all $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, $x \in \mathbb{R}^{d_x}$ and $s \in [t,\widetilde{T}]$:

$$\begin{aligned}\gamma(s,x,m) = p(X_{\widetilde{T}}^{s,m}(x), X_{\widetilde{T}}^{s,m}\#m) + \int_s^{\widetilde{T}} \bigg(&\partial_x f(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, \widehat{\alpha}(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, Z_\tau^{s,m}(x))) \cdot Z_\tau^{s,m}(x) \\ &+ \partial_x g(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, \widehat{\alpha}(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, Z_\tau^{s,m}(x)))\bigg) d\tau, \end{aligned} \tag{6.8}$$

where $Z_\tau^{s,m}(x) := \gamma\big(\tau, X_\tau^{s,m}(x), X_\tau^{s,m}\#m\big)$ and

$$X_\tau^{s,m}(x) = x + \int_s^\tau f\Big(X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m, \widehat{\alpha}\big(X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m, \gamma(\tau, X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m)\big)\Big) d\widetilde{\tau}, \tag{6.9}$$

and $\widehat{\alpha}(x,\mu,z)$ is the optimal control solved *via* the first-order condition (4.30).

Remark 6.2. In other words, the decoupling field $\gamma := \gamma(s,x,m)$ has the following property: using the given γ, one can first solve for $X_\tau^{s,m}(x)$ in (6.9), and then define $Z_\tau^{s,m}(x) := \gamma\big(\tau, X_\tau^{s,m}(x), X_\tau^{s,m}\#m\big)$. Then, it can be checked directly that the couple $\big(X_s^{t,m}(x), Z_s^{t,m}(x)\big)_{x\in\mathbb{R}^{d_x}, s\in[t,\widetilde{T}]}$ is a solution to the FBODE system (6.1).

By carefully constructing the iteration mappings of the decoupling field γ and its derivatives and utilizing Assumptions $(\mathbf{a1})$, $(\mathbf{a2})$ and $(\mathbf{P})$ on f, g, and p, we can prove the local-in-time existence, uniqueness, and regularity of solutions to the FBODE system (6.1) using the Banach fixed point argument. The proofs are similar to, but simpler[4] than, the proofs of Theorem 7.3 and 7.4 presented in Section 7.3 of our previous paper [52]. Therefore, we leave the technical details of these tedious but elementary proofs to the interested readers.

Theorem 6.3. *(Local-in-time existence) Assume that the drift function f and the running cost g satisfy Assumptions $(\mathbf{a1})$ and $(\mathbf{a2})$, respectively, the terminal function p satisfies Assumption $(\mathbf{P})$ with $L_p \leq \overline{L}_p := \max\{\Lambda_k, L_0^*\}$ and the inequality relation*[5]

$$\sup_{(x,\mu,a)\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_a}} \|\partial_a\partial_a f(x,\mu,a)\|_{\mathcal{L}(\mathbb{R}^{d_a}\times\mathbb{R}^{d_a};\mathbb{R}^{d_x})}\cdot(1+|x|+\|\mu\|_1) \leq \frac{1}{40\overline{L}_p}\lambda_g, \tag{6.10}$$

among f, g and p is valid, then there exists a constant $\epsilon_1 = \epsilon_1(\overline{L}_p; L_f, \Lambda_f, \bar{l}_f, L_g, \Lambda_g, \bar{l}_g, L_\alpha) > 0$, such that, for any $0 \leq t \leq \widetilde{T} \leq T$ with $\widetilde{T} - t \leq \epsilon_1$,
(a) *there exists a unique continuous decoupling field $\gamma : (s,x,m) \in [t,\widetilde{T}] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \mapsto \gamma(s,x,m) \in \mathbb{R}^{d_x}$ for the FBODE system (6.1) on $[t,\widetilde{T}]$, which also satisfies $|||\gamma|||_{1,[t,\widetilde{T}]} \leq 2\overline{L}_p$;*
(b) *moreover, for any $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, define $X_s^{t,m}(x)$ by solving (6.9) and $Z_s^{t,m}(x) := \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m)$, then the pair $(X_s^{t,m}(x), Z_s^{t,m}(x))_{x\in\mathbb{R}^{d_x}, s\in[t,\widetilde{T}]}$ is a solution of FBODE system (6.1) satisfying the following properties:*
(i) each of them is continuous in $(t,m,x) \in [t,\widetilde{T}] \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x}$ and is continuously differentiable in $s \in [t,\widetilde{T}]$;
(ii)

$$\left|Z_s^{t,m}(x)\right| \leq 2\overline{L}_p \cdot \left(1 + \left|X_s^{t,m}(x)\right| + \left\|X_s^{t,m}\right\|_{L_m^{1,d_x}}\right), \tag{6.11}$$

which means that $(X_s^{t,m}(x), Z_s^{t,m}(x))$ satisfies the cone property $(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)) \in c_{k_0}$ with the cone set c_{k_0} defined in (4.28);
(iii) in addition, $X_s^{t,m}(x)$ has the following flow property of $X_s^{t,m}(x)$, for any $0 \leq t \leq s \leq \tau \leq T$,

$$X_\tau^{t,m}(x) = X_\tau^{s,X_s^{t,m}\#m}\big(X_s^{t,m}(x)\big). \tag{6.12}$$

Theorem 6.4. *(Regularity of the local solution) Assume that the drift function f and the running cost g satisfy Assumptions $(\mathbf{a1})$ and $(\mathbf{a2})$ respectively, the terminal function p satisfies Assumption $(\mathbf{P})$ with $L_p \leq \overline{L}_p := \max\{\Lambda_k, L_0^*\}$ and the inequality relation (6.10) among f, g and p is valid, then there exists a constant $\epsilon_2 = \epsilon_2(\overline{L}_p; L_f, \Lambda_f, \bar{l}_f, L_g, \Lambda_g, \bar{l}_g, L_\alpha) > 0$, such that, for any $0 \leq t \leq \widetilde{T} \leq T$ with $\widetilde{T} - t \leq \epsilon_2$,*
(a) *the continuous decoupling field $\gamma(s,x,\mu)$ for the FBODE system (6.1) on $[t,\widetilde{T}]$ constructed in Theorem 6.3 has the following regularities and estimates:*
(i) $\gamma(s,x,\mu)$ is differentiable in $s \in [\widetilde{T}-\epsilon_2,\widetilde{T}]$ and $x \in \mathbb{R}^{d_x}$, and L-differentiable in $\mu \in \mathcal{P}_2(\mathbb{R}^{d_x})$, and its derivatives $\partial_s\gamma(s,x,\mu)$, $\partial_x\gamma(s,x,\mu)$ and $\partial_\mu\gamma(s,x,\mu)(\widetilde{x})$ are jointly continuous in their corresponding arguments $(s,x,\mu) \in [\widetilde{T}-\epsilon_2,\widetilde{T}] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x})$ and $(s,x,\mu,\widetilde{x}) \in [\widetilde{T}-\epsilon_2,\widetilde{T}] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x}) \times \mathbb{R}^{d_x}$ respectively;
(ii)

$$\sup_{s\in[\widetilde{T}-\epsilon_2,\widetilde{T}],\, x\in\mathbb{R}^{d_x},\, \mu\in\mathcal{P}_2(\mathbb{R}^{d_x})} \|\partial_x\gamma(s,x,\mu)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \leq 2L_p, \tag{6.13}$$

$$\sup_{s\in[\widetilde{T}-\epsilon_2,\widetilde{T}],\, x\in\mathbb{R}^{d_x},\, \mu\in\mathcal{P}_2(\mathbb{R}^{d_x}),\, \widetilde{x}\in\mathbb{R}^{d_x}} \|\partial_\mu\gamma(s,x,\mu)(\widetilde{x})\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})} \leq 2L_p. \tag{6.14}$$

[4]The simplification is due to the absence of the terms $\partial_\mu f(x,\mu,a)$ and $\partial_\mu g(x,\mu,a)$ on the right-hand side of the FBODE system (6.1), in comparison with (7.1) of [52].
[5]It echoes the condition (4.29).

(b) *Moreover, the solution pair* $\big(X_s^{t,m}(x), Z_s^{t,m}(x)=\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m)\big)_{x\in\mathbb{R}^{d_x},s\in[\widetilde{T}-\epsilon_2,\widetilde{T}]}$ *of FBODE system* (6.1) *constructed in Theorem 6.3 through* γ *has the following regularities:*
(i) $X_s^{t,m}(x)$ *is differentiable in* $x\in\mathbb{R}^{d_x}$ *and L-differentiable in* $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$*, and its derivatives* $\partial_x\big(X_s^{t,m}(x)\big)$ *and* $\partial_m\big(X_s^{t,m}(x)\big)(y)$ *are jointly continuous in their corresponding arguments* $(t,m,x)\in[\widetilde{T}-\epsilon_2,\widetilde{T}]\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$ *and* $(t,m,x,y)\in[\widetilde{T}-\epsilon_2,\widetilde{T}]\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}\times\mathbb{R}^{d_x}$ *respectively;* $X_s^{t,m}(x)$ *and these derivatives are also continuously differentiable in* $s\in[t,\widetilde{T}]$ *and satisfy the following estimates:*

$$\big\|\partial_x\big(X_s^{t,m}(x)\big)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le\exp\big(L_B(s-t)\big),\tag{6.15}$$

$$\big\|\partial_m\big(X_s^{t,m}(x)\big)(y)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le L_B(s-t)\Big(L_B(s-t)\exp\Big(2L_B(s-t)\Big)+1\Big)\exp\Big(2L_B(s-t)\Big);\tag{6.16}$$

(ii) $X_s^{t,m}(x)$ *is differentiable in* $t\in[\widetilde{T}-\epsilon_2,\widetilde{T}]$*; its derivative* $\partial_t\big(X_s^{t,m}(x)\big)$ *is jointly continuous in its arguments* $(t,m,x)\in[\widetilde{T}-\epsilon_2,\widetilde{T}]\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$*; this derivative is also continuously differentiable in* $s\in[t,\widetilde{T}]$ *and satisfies the following estimates:*

$$\big|\partial_t\big(X_s^{t,m}(x)\big)\big|\le\overline{L}_B\big(1+|x|+\|m\|_1\big)\Big(L_B(s-t)\exp\Big(2L_B(s-t)\Big)+1\Big)\exp\Big(L_B(s-t)\Big);\tag{6.17}$$

here L_B *and* $\overline{L}_B$ *are constants defined in* (6.5)*;*
(iii) $Z_s^{t,m}(x)=\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m)$ *is differentiable in* $t\in[\widetilde{T}-\epsilon_2,\widetilde{T}]$ *and* $x\in\mathbb{R}^{d_x}$*, and L-differentiable in* $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$*; and its derivatives* $\partial_t\big(Z_s^{t,m}(x)\big)$*,* $\partial_x\big(Z_s^{t,m}(x)\big)$ *and* $\partial_m\big(Z_s^{t,m}(x)\big)(y)$ *are jointly continuous in their corresponding arguments* $(t,m,x)\in[\widetilde{T}-\epsilon_2,\widetilde{T}]\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$ *and* $(t,m,x,y)\in[\widetilde{T}-\epsilon_2,\widetilde{T}]\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}\times\mathbb{R}^{d_x}$ *respectively;* $Z_s^{t,m}(x)$ *and these derivatives also continuously differentiable in* $s\in[t,\widetilde{T}]$ *and satisfy the following estimates:*

$$\big\|\partial_x\big(Z_s^{t,m}(x)\big)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le 2L_p\exp\big(L_B(s-t)\big);\tag{6.18}$$

$$\big\|\partial_m\big(Z_s^{t,m}(x)\big)(y)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le 4L_pL_B(s-t)\Big(L_B(s-t)\exp\Big(2L_B(s-t)\Big)+1\Big)\exp\Big(2L_B(s-t)\Big)+2L_p\exp\big(L_B(s-t)\big);\tag{6.19}$$

$$\big|\partial_t\big(Z_s^{t,m}(x)\big)\big|\le 4L_p\overline{L}_B\big(1+|x|+\|m\|_1\big)\Big(L_B(s-t)\exp\Big(2L_B(s-t)\Big)+1\Big)\exp\Big(L_B(s-t)\Big).\tag{6.20}$$

Remark 6.5. The assumption on the validity of (6.10) condition is used only for ensuring the unique solvability of the optimal control $\widehat{\alpha}(x,\mu,z)$, which serves a sufficient condition but not necessary one. However, this assumption simplifies the calculations involved, as shown in Proposition 4.6. If the optimal control can be found explicitly and is regular enough, the assumption (6.10) in Theorems 6.3 and 6.4 can even be removed.

7. Crucial estimate and global-in-time existence

7.1. Hypotheses and properties

To obtain a global-in-time solution of (2.12), we only need to establish the global existence of a decoupling field γ for the FBODE system (2.12) on the interval $[0,T]$, which also satisfies (6.8). This is because the pair $\big(X_s^{t,m}(x),Z_s^{t,m}(x)\big)$ defined by solving (6.9) and $Z_s^{t,m}(x):=\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m)$ is a solution of the FBODE system (6.1). Moreover, the FBODE system (6.1) reduces to the system (2.12) when $\widetilde{T}=T$ and $p(x,\mu)=\partial_x k(x,\mu)$. Therefore, establishing the global existence of a decoupling field γ yields a global-in-time solution to the FBODE system (2.12).

To extend the local-in-time solution of (6.8) to a global-in-time solution over the interval $[0,T]$, we need to derive several new *a priori* estimates, which play a crucial role of extending the existence lifespan and can be established under Assumptions (**a1**)–(**a3**) and the following additional **Hypotheses**:
(**h1**) $f(0,\delta_0,0)=0$, $\partial_x g(0,\delta_0,0)=\partial_a g(0,\delta_0,0)=0$, $\partial_\mu g(0,\delta_0,0)(0)=0$, $\partial_x k(0,\delta_0)=0$ and $\partial_\mu k(0,\delta_0)(0)=0$, where δ_0 is the Dirac measure with a point mass at 0.

(**h2**) We set the requirement: there are positive constants λ_1 and λ_2 such that, for any $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$ and $X,\ \widetilde{X},\ Z,\ \widetilde{Z} \in L_m^{2,d_x}$ satisfying $(X(x), X\#m, Z(x)) \in c_{k_0}$,

$$
\begin{aligned}
&(\lambda_g - l_g)\int_{\mathbb{R}^{d_x}} |\widetilde{X}(x)|^2 dm(x) - \int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}} \widetilde{Z}(x)^\top \partial_\mu \partial_z h(X(x), X\#m, Z(x))\big(X(\widetilde{x})\big)\widetilde{X}(\widetilde{x})dm(\widetilde{x})dm(x)\\
&-\int_{\mathbb{R}^{d_x}} \widetilde{Z}(x)^\top \partial_z\partial_z h(X(x), X\#m, Z(x))\widetilde{Z}(x)dm(x) \geq \lambda_1 \int_{\mathbb{R}^{d_x}} |\widetilde{X}(x)|^2 dm(x) + \lambda_2 \int_{\mathbb{R}^{d_x}} |\widetilde{Z}(x)|^2 dm(x), \quad (7.1)
\end{aligned}
$$

where $h(x,\mu,z)$ is the Hamiltonian defined in (5.2) and the cone set c_{k_0} was defined in (4.28).
(**h3**) Recall the constants defined in (**a1**)$-$(**a3**), we here specify $\bar{l}_f \leq \frac{1}{40\max\{\Lambda_k, L_0^*\}}\min\{\lambda_1,\lambda_g\}$ and $8\bar{l}_g \leq \min\{\lambda_1,\lambda_g\}$, where the positive constant L_0^* defined in (7.56) depends only on λ_f, Λ_1, Λ_2, Λ_3, λ_1, λ_2, λ_g, Λ_g, l_g, λ_k, Λ_k and l_k but not on $\bar{l}_g$ and $\bar{l}_f$.

Remark 7.1. (i) Hypotheses (**h1**)-(**h3**) are satisfied in many interesting cases including Linear-Quadratic setting (see [9]), also see two immediate non-LQ examples in Section 9. (**h1**) only serves for reducing the technicalities involved, it is not crucial and our methodology can even apply to the case without assuming it; we include here just for the sake of simplified illustration; indeed for the case with linear drift function f, we can even drop Hypothesis (**h1**) and set $\bar{l}_f = 0$, so that (**h3**) can be automatically partially satisfied.
(ii) (**h3**) tells us that exaggerated cross interactions among control, state and its measure should be avoided in order to get an equilibrium strategy for the game in the long run. Many practical examples in the literature also justifies the fact that self-reinforced interactions among them will jeopardize the long term stability of the game; this is a common understanding in the literature, and some authors of the present work also have obtained such a consensus in a private communication with Peter Caines [68]. The constants chosen in (**h3**) are just for convenience and we do not intend to optimize them.
(iii) By (7.1), we can derive that $\lambda_1 \leq \lambda_g - l_g$ by setting $\widetilde{Z}(x) = 0$, and thus $\lambda_1 \leq \lambda_g$ when l_g is non-negative. In addition, under Assumptions (**a1**)–(**a3**) and Hypothesis (**h3**), we have the results stated in Propositions 4.3–4.7 since $\bar{l}_f \leq \frac{1}{40\max\{\Lambda_k, L_0^*\}}\lambda_g$ implies (4.29).
(iv) The condition $\bar{l}_f \leq \frac{1}{40\max\{\Lambda_k, L_0^*\}}\min\{\lambda_1,\lambda_g\}$ in Hypothesis (**h3**) means that the rate of the second-order derivatives of drift function f is relatively small, it means that the drift rate cannot have a large quadratic-growth effect; even for the classical control problems, the drift function with quadratic-growth may sometimes lead to ill-posedness, and this limitation on modeling can also be observed in a lot of practical applications; see the monograph [69] for instance.
(v) We can also set the requirement in (**h2**) as follows: there are positive constants $\widetilde{\lambda}_1$ and λ_2 such that, for any $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$ and $X,\ \widetilde{X},\ Z,\ \widetilde{Z} \in L_m^{2,d_x}$ satisfying $(X(x), X\#m, Z(x)) \in c_{k_0}$,

$$
\begin{aligned}
&\left(\frac{1}{2}\lambda_g - l_g\right)\int_{\mathbb{R}^{d_x}} |\widetilde{X}(x)|^2 dm(x) - \int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}} \widetilde{Z}(x)^\top \partial_\mu \partial_z h(X(x), X\#m, Z(x))\big(X(\widetilde{x})\big)\widetilde{X}(\widetilde{x})dm(\widetilde{x})dm(x)\\
&-\int_{\mathbb{R}^{d_x}} \widetilde{Z}(x)^\top \partial_z\partial_z h(X(x), X\#m, Z(x))\widetilde{Z}(x)dm(x) \geq \widetilde{\lambda}_1 \int_{\mathbb{R}^{d_x}} |\widetilde{X}(x)|^2 dm(x) + \lambda_2 \int_{\mathbb{R}^{d_x}} |\widetilde{Z}(x)|^2 dm(x), \quad (7.2)
\end{aligned}
$$

which can imply (7.1) with $\lambda_1 = \widetilde{\lambda}_1 + \frac{1}{2}\lambda_g$. Thus, a sufficient condition for Hypothesis (**h3**) is $\bar{l}_f \leq \frac{1}{40\max\{\Lambda_k, L_0^*\}} \cdot \frac{1}{2}\lambda_g$ and $8\bar{l}_g \leq \frac{1}{2}\lambda_g$; in particular, $\widetilde{\lambda}_1$ can be chosen arbitrary close to 0.
(vi) The second condition of Hypothesis (**h3**) in the form $\bar{l}_g \leq C\lambda_g$ for some constant $C > 0$ is also implicitly involved in the displacement monotonicity condition (see Asm. 3.5 (ii) in [30]), even in the following simplest case: consider the simplified case that, $d_x = d_a = 1$, $f(x,\mu,a) = a$, and $g(x,\mu,a) = \frac{1}{2}\lambda_g x^2 + \frac{1}{2}\lambda_g a^2 + \bar{l}_g xa$ where the constants $\lambda_g > 0$ and $\bar{l}_g \geq 0$. Then $\widehat{\alpha}(x,\mu,z) = -\frac{1}{\lambda_g}(z + \bar{l}_g x)$ and $h(x,\mu,z) = -\frac{1}{2\lambda_g}(z+\bar{l}_g x)^2 + \frac{1}{2}\lambda_g x^2$. In

this case, since the Hamiltonian h is independent of μ, the displacement monotonicity condition is reduced to $\partial_x\partial_x h = -\frac{\bar{l}_g^2}{\lambda_g} + \lambda_g \geq 0$, that is, $\lambda_g \geq \bar{l}_g$.
(vii) The first condition of Hypothesis (**h3**) in the form $\bar{l}_f \leq C\min\{\lambda_1, \lambda_g\}$ for some constant $C > 0$ prevents the drift f in forward dynamics from exhibiting excessively strong interactions among x, μ, and a. This requirement is a natural sufficient condition for the global-in-time solvability of the forward dynamics, as it avoids f having some sort of quadratic growth that could potentially lead to a finite time blowup of the state. In practical applications, further structural information about f is usually available, and one may further relax this condition accordingly. However, for the sake of convenience, we impose this condition in this work to simplify our mathematical analysis for the generic case.

Remark 7.2. (i) A sufficient condition (on the coefficient functions f and g) for hypothesis (**h2**) is to assume that

$$\Big(\Lambda_2 + \frac{1}{4}\Lambda_1\Big)^2 - 4(\lambda_g - l_g)\cdot \frac{\lambda_f^2}{\Lambda_g + \lambda_g/20} < 0. \tag{7.3}$$

Under the condition (7.3), the corresponding λ_1 and λ_2 in (7.1) can be computed as follows. First, by Hypothesis (**h2**), (4.2), (4.35) and Proposition 4.7, we have, for any $(x,\mu,z) \in c_{k_0}$, where k_0 was defined in (4.27), and for any y, $\xi \in \mathbb{R}^{d_x}$,

$$\begin{aligned}
&\|\partial_\mu\partial_z h(x,\mu,z)(y)\|_{\mathcal{L}(\mathbb{R}^{d_x}\times\mathbb{R}^{d_x};\mathbb{R})} = \|\partial_\mu f(x,\mu,\widehat{\alpha}(x,\mu,z))(y) + \partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\partial_\mu\widehat{\alpha}(x,\mu,z)(y)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\\
&\leq \Lambda_2 + \Lambda_1\frac{20\big(\bar{l}_g + \frac{1}{2}k_0\cdot\bar{l}_f\big)}{19\lambda_g} < \Lambda_2 + \frac{1}{4}\Lambda_1 < \frac{5}{4}\Lambda_f, \qquad (7.4)\\
&\text{and } \; -\xi^\top\partial_z\partial_z h(x,\mu,z)\xi = -\xi^\top\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\partial_z\widehat{\alpha}(x,\mu,z)\xi \geq \frac{\lambda_f^2}{\Lambda_g + \lambda_g/20}|\xi|^2,
\end{aligned}$$

which imply that, for any $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, and for any $X,\ \widetilde{X},\ Z,\ \widetilde{Z} \in L_m^{2,d_x}$ so that $(X(x), X\#m, Z(x)) \in c_{k_0}$,

$$\begin{aligned}
&(\lambda_g - l_g)\int_{\mathbb{R}^{d_x}}|\widetilde{X}(x)|^2dm(x) - \int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}\widetilde{Z}(x)^\top\partial_\mu\partial_z h(X(x),X\#m,Z(x))\big(X(\widetilde{x})\big)\widetilde{X}(\widetilde{x})dm(\widetilde{x})dm(x)\\
&\quad - \int_{\mathbb{R}^{d_x}}\widetilde{Z}(x)^\top\partial_z\partial_z h(X(x),X\#m,Z(x))\widetilde{Z}(x)dm(x)\\
&\geq (\lambda_g - l_g)\int_{\mathbb{R}^{d_x}}|\widetilde{X}(\widetilde{x})|^2dm(\widetilde{x}) + \frac{\lambda_f^2}{\Lambda_g+\lambda_g/20}\int_{\mathbb{R}^{d_x}}|\widetilde{Z}(x)|^2dm(x)\\
&\quad - \Big(\Lambda_2 + \frac{1}{4}\Lambda_1\Big)\int_{\mathbb{R}^{d_x}}|\widetilde{X}(\widetilde{x})|dm(\widetilde{x})\int_{\mathbb{R}^{d_x}}|\widetilde{Z}(x)|dm(x)\\
&\geq \lambda_1\int_{\mathbb{R}^{d_x}}|\widetilde{X}(\widetilde{x})|^2dm(\widetilde{x}) + \lambda_2\int_{\mathbb{R}^{d_x}}|\widetilde{Z}(x)|^2dm(x),
\end{aligned}$$

for some positive constants λ_1 and λ_2 depending only on λ_f, Λ_1, Λ_2, λ_g, Λ_g and l_g since we have (7.3); for instance, we can set $\lambda_1 := (1-\theta)(\lambda_g - l_g)$ and $\lambda_2 := (1-\theta)\frac{\lambda_f^2}{\Lambda_g+\lambda_g/20}$ where $\theta = \Big(\Big(\Lambda_2 + \frac{1}{4}\Lambda_1\Big)^2\Big/\Big(4(\lambda_g - l_g)\cdot\frac{\lambda_f^2}{\Lambda_g+\lambda_g/20}\Big)\Big)^{1/2} \in [0,1)$.

In fact, by (7.4), the hypothesis (7.1) means that $\partial_\mu\widehat{\alpha}(x,\mu,z)(y)$ (and thus $\partial_\mu\partial_a f(x,\mu,a)(y)$ and $\partial_\mu\partial_a g(x,\mu,a)(y)$) should be suitably small, and the derivative $\partial_\mu f(x,\mu,\widehat{\alpha}(x,\mu,z))(y)$ should not be too large compared with the absolute value of $\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))(y)$ and the convexity of the running cost function $g(x,\mu,a)$ in (x,μ). Also note that, because of (7.4), a sufficient condition for the hypothesis (7.1) is $\frac{25}{16}\Lambda_f^2 - 4(\lambda_g - l_g)\cdot\frac{\lambda_f^2}{\Lambda_g+\lambda_g/20} < 0$, and readers can also find that the proposed non-LQ examples in Section 9 actually fulfills this sufficient condition.

(ii) If $\partial_\mu f(x,\mu,\widehat{\alpha}(x,\mu,z))(y) = \partial_\mu\partial_a f(x,\mu,a)(y) = \partial_\mu\partial_a g(x,\mu,a)(y) = 0$, then $\partial_\mu\partial_z h(x,\mu,z)(y) = 0$, which means that the Hamiltonian $h(x,\mu,z)$ is separable; in this case, (7.1) is automatically satisfied with the choices of $\lambda_1 = \lambda_g - l_g > 0$ and $\lambda_2 = \frac{\lambda_f^2}{\Lambda_g+\lambda_g/20}$. In addition, the displacement monotonicity condition for Hamiltonian, which is similar to (7.1) but not the same, is also assumed in Assumption 3.5 (ii) (that is (3.2)) in [30] to deal with nonseparable Hamiltonian. We here need to cope with nonlinear drift functions, both expressions $\partial_x\partial_x h(x,\mu,z)(y) = \partial_x\partial_x f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z + \partial_x\partial_x g(x,\mu,\widehat{\alpha}(x,\mu,z))$ and $\partial_\mu\partial_x h(x,\mu,z)(y) = \partial_\mu\partial_x f(x,\mu,\widehat{\alpha}(x,\mu,z))(y)\cdot z + \partial_\mu\partial_x g(x,\mu,\widehat{\alpha}(x,\mu,z))(y)$ allow z to take values without a bound, and to obtain more tractable conditions, one should impose assumptions to the coefficient functions f and g directly than to the Hamiltonian h only.

(iii) For the class of examples (4.14)–(4.16), sufficient conditions for (**h1**) are

$$\begin{aligned}&f_0(0,\phi_0(0),0) = \partial_x g_0(0,\phi_1(0),0) = \partial_a g_0(0,\phi_1(0),0) = \partial_\mu g_0(0,\phi_1(0),0)(0) = \partial_x k_0(0,\phi_2(0))\\&= \partial_\mu k_0(0,\phi_2(0))(0) = 0;\end{aligned}$$

sufficient conditions for (**h2**) are that $\|M_3\|_{l^\infty}^2 < 3.8\sigma_1(M_5)\sigma_1(M_2M_2^\top)/\max\{\|M_4\|_{l^\infty}, \|M_5\|_{l^\infty}, \|M_6\|_{l^\infty}\}$ where $\sigma_1(M)$ is the smallest eigenvalue of matrix M, and ε_1 and ε_2 are suitably small, by using (7.3) and (7.4); sufficient conditions for (**h3**) are that ε_1 and ε_2 are suitably small.

By using Hypothesis (**h1**), we have the following proposition which guides us with better linear-growth estimates for various coefficient functions and the equilibrium control.

Proposition 7.3. *Under Assumptions* (**a1**)–(**a3**) *and Hypotheses* (**h1**)–(**h3**), *we have* $\widehat{\alpha}(0,\delta_0,0) = 0$ *and the following linear-growth estimates hold:*

$$i)\ |f(x,\mu,a)| \le L_f\left(|x| + |a| + \|\mu\|_1\right); \tag{7.5}$$

$$ii)\ \sup_{x\in\mathbb{R}^{d_x},\,\mu\in\mathcal{P}_2(\mathbb{R}^{d_x}),\,a\in\mathbb{R}^{d_a},\,\widetilde{x}\in\mathbb{R}^{d_x}} \frac{|\partial_x g(x,\mu,a)|}{|x| + \|\mu\|_1 + |a|} \vee \frac{|\partial_a g(x,\mu,a)|}{|x| + \|\mu\|_1 + |a|} \le L_g; \tag{7.6}$$

$$iii)\ \sup_{x\in\mathbb{R}^{d_x},\,\mu\in\mathcal{P}_2(\mathbb{R}^{d_x}),\,\widetilde{x}\in\mathbb{R}^{d_x}} \frac{|\partial_x k(x,\mu)|}{|x| + \|\mu\|_1} \le L_k; \tag{7.7}$$

$$iv)\ |\widehat{\alpha}(x,\mu,z)| \le L_\alpha(|x| + \|\mu\|_1 + |z|), \tag{7.8}$$

where L_f, L_g, L_k *and* L_α *were defined in Proposition 4.3, 4.4 and 4.6, respectively.*

Proof. By using Hypothesis (**h1**) and the first-order condition (4.30), $\widehat{\alpha} = 0$ is a root to the first-order condition (4.30) at the point $(x,\mu,z) = (0,\delta_0,0)$, therefore in light of the uniqueness stated in Proposition 4.6, we know that the unique minimizer satisfies $\widehat{\alpha}(0,\delta_0,0) = 0$. By using Hypothesis (**h1**) together with the arguments leading to (4.21) (4.25), (4.26) and (4.38), we can similarly deduce (7.5)–(7.8); in a certain sense, the proofs for (7.5)–(7.8) are slightly simpler than those for (4.21) (4.25), (4.26) and (4.38). □

TABLE 1. Abbreviations for the derivatives of $\widehat{\alpha}$ and f, where $y,\widetilde{y},\widehat{y}\in\mathbb{R}^{d_x}$, $t\in[0,T]$, $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$, $s\in[t,T]$ and we also evaluate all of them at $(x,\mu,z,a,\widetilde{x},\widehat{x})=(X_s^{t,m}(y),X_s^{t,m}\#m,Z_s^{t,m}(y),\widehat{\alpha}_s^{t,m}(y),X_s^{t,m}(\widetilde{y}),X_s^{t,m}(\widehat{y}))\in\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}\times\mathbb{R}^{d_a}\times\mathbb{R}^{d_x}\times\mathbb{R}^{d_x}$.

$\big(\partial_x\widehat{\alpha}\big)_s^{t,m}(y)=\partial_x\widehat{\alpha}(x,\mu,z)$	$\big(\partial_\mu\widehat{\alpha}\big)_s^{t,m}(y,\widetilde{y})=\partial_\mu\widehat{\alpha}(x,\mu,z)(\widetilde{x})$	$\big(\partial_z\widehat{\alpha}\big)_s^{t,m}(y)=\partial_z\widehat{\alpha}(x,\mu,z)$
$\big(\partial_x f\big)_s^{t,m}(y)=\partial_x f(x,\mu,a)$	$\big(\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})=\partial_\mu f(x,\mu,a)(\widetilde{x})$	$\big(\partial_a f\big)_s^{t,m}(y)=\partial_a f(x,\mu,a)$
$\big(\partial_x\partial_x f\big)_s^{t,m}(y)=\partial_x\partial_x f(x,\mu,a)$	$\big(\partial_x\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})=\partial_x\partial_\mu f(x,\mu,a)(\widetilde{x})$	$\big(\partial_{\widetilde{x}}\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})=\partial_{\widetilde{x}}\partial_\mu f(x,\mu,a)(\widetilde{x})$
$\big(\partial_a\partial_a f\big)_s^{t,m}(y)=\partial_a\partial_a f(x,\mu,a)$	$\big(\partial_a\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})=\partial_a\partial_\mu f(x,\mu,a)(\widetilde{x})$	$\big(\partial_x\partial_a f\big)_s^{t,m}(y)=\partial_x\partial_a f(x,\mu,a)$
	$\big(\partial_\mu\partial_\mu f\big)_s^{t,m}(y,\widetilde{y},\widehat{y})=\partial_\mu\partial_\mu f(x,\mu,a)(\widetilde{x},\widehat{x})$	

7.2. Notations

For simplicity, we make use of the following notations by omitting some of their own arguments while only keeping the essential parameters and variables such as t, m, x and s:

$$\begin{cases}\widehat{\alpha}_s^{t,m}(x):=\widehat{\alpha}(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)),\ g_s^{t,m}(x):=g(X_s^{t,m}(x),X_s^{t,m}\#m,\widehat{\alpha}_s^{t,m}(x)),\\ f_s^{t,m}(x):=f(X_s^{t,m}(x),X_s^{t,m}\#m,\widehat{\alpha}_s^{t,m}(x)),\ k_s^{t,m}(x):=k(X_s^{t,m}(x),X_s^{t,m}\#m).\end{cases}\tag{7.9}$$

Likewise, we adopt the same style of notations for the derivatives of $\widehat{\alpha}$ and f, as shown in the following Table 1.

Using these notations and the chain rule, one can write:

$$\begin{aligned}\partial_t\big(\widehat{\alpha}_s^{t,m}(y)\big)&=\big(\partial_x\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_t\big(X_s^{t,m}(y)\big)+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu\widehat{\alpha}\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_t\big(X_s^{t,m}(\widetilde{y})\big)dm(\widetilde{y})+\big(\partial_z\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_t\big(Z_s^{t,m}(y)\big);\\ \partial_t\big(f_s^{t,m}(y)\big)&=\big(\partial_x f\big)_s^{t,m}(y)\cdot\partial_t\big(X_s^{t,m}(y)\big)+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_t\big(X_s^{t,m}(\widetilde{y})\big)dm(\widetilde{y})+\big(\partial_a f\big)_s^{t,m}(y)\cdot\partial_t\big(\widehat{\alpha}_s^{t,m}(y)\big);\\ \partial_t\big(g_s^{t,m}(y)\big)&=\big(\partial_x g\big)_s^{t,m}(y)\cdot\partial_t\big(X_s^{t,m}(y)\big)+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu g\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_t\big(X_s^{t,m}(\widetilde{y})\big)dm(\widetilde{y})+\big(\partial_a g\big)_s^{t,m}(y)\cdot\partial_t\big(\widehat{\alpha}_s^{t,m}(y)\big);\\ \partial_t\big(k_s^{t,m}(y)\big)&=\big(\partial_x k\big)_s^{t,m}(y)\cdot\partial_t\big(X_s^{t,m}(y)\big)+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu k\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_t\big(X_s^{t,m}(\widetilde{y})\big)dm(\widetilde{y});\end{aligned}\tag{7.10}$$

similarly, we also have a set of similar equations for $\partial_m\big(\widehat{\alpha}_s^{t,m}(y)\big)(\widetilde{y})$, $\partial_m\big(f_s^{t,m}(y)\big)(\widetilde{y})$, $\partial_m\big(g_s^{t,m}(y)\big)(\widetilde{y})$ and $\partial_m\big(k_s^{t,m}(y)\big)(\widetilde{y})$, and $\partial_x\big(\widehat{\alpha}_s^{t,m}(y)\big)$, $\partial_x\big(f_s^{t,m}(y)\big)$, $\partial_x\big(g_s^{t,m}(y)\big)$, $\partial_x\big(k_s^{t,m}(y)\big)$. For instance, $\partial_m\big(\widehat{\alpha}_s^{t,m}(y)\big)(\widetilde{y})$, $\partial_m\big(f_s^{t,m}(y)\big)(\widetilde{y})$, $\partial_y\big(\widehat{\alpha}_s^{t,m}(y)\big)$, $\partial_y\big(f_s^{t,m}(y)\big)$ can be written as, for any $y,\widetilde{y}\in\mathbb{R}^{d_x}$, $t\in[0,T]$, $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ and $s\in[t,T]$,

$$\begin{aligned}\partial_m\big(\widehat{\alpha}_s^{t,m}(y)\big)(\widetilde{y})&=\big(\partial_x\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_m\big(X_s^{t,m}(y)\big)(\widetilde{y})+\big(\partial_\mu\widehat{\alpha}\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_{\widetilde{y}}\big(X_s^{t,m}(\widetilde{y})\big)\\ &\quad+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu\widehat{\alpha}\big)_s^{t,m}(y,\widehat{y})\cdot\partial_m\big(X_s^{t,m}(\widehat{y})\big)(\widetilde{y})dm(\widehat{y})+\big(\partial_z\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_m\big(Z_s^{t,m}(y)\big)(\widetilde{y});\\ \partial_m\big(f_s^{t,m}(y)\big)(\widetilde{y})&=\big(\partial_x f\big)_s^{t,m}(y)\cdot\partial_m\big(X_s^{t,m}(y)\big)(\widetilde{y})+\big(\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})\cdot\partial_{\widetilde{y}}\big(X_s^{t,m}(\widetilde{y})\big)\\ &\quad+\int_{\mathbb{R}^{d_x}}\big(\partial_\mu f\big)_s^{t,m}(y,\widehat{y})\cdot\partial_m\big(X_s^{t,m}(\widehat{y})\big)(\widetilde{y})dm(\widehat{y})+\big(\partial_a f\big)_s^{t,m}(y)\cdot\partial_m\big(\widehat{\alpha}_s^{t,m}(y)\big)(\widetilde{y});\\ \partial_y\big(\widehat{\alpha}_s^{t,m}(y)\big)&=\big(\partial_x\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_y\big(X_s^{t,m}(y)\big)+\big(\partial_z\widehat{\alpha}\big)_s^{t,m}(y)\cdot\partial_y\big(Z_s^{t,m}(y)\big);\\ \partial_y\big(f_s^{t,m}(y)\big)&=\big(\partial_x f\big)_s^{t,m}(y)\cdot\partial_y\big(X_s^{t,m}(y)\big)+\big(\partial_a f\big)_s^{t,m}(y)\cdot\partial_y\big(\widehat{\alpha}_s^{t,m}(y)\big).\end{aligned}\tag{7.11}$$

For the sake of concise expression, in addition to the shorthand introduced in (7.9) and Table 1, we also adopt the following notations:

$$\begin{cases}\big(L_{xx}\big)_s^{t,m}(y):=\big(\partial_x\partial_x f\big)_s^{t,m}(y)\cdot Z_s^{t,m}(y)+\big(\partial_x\partial_x g\big)_s^{t,m}(y);\\ \big(L_{\mu x}\big)_s^{t,m}(y,\widetilde{y}):=\big(\partial_x\partial_\mu f\big)_s^{t,m}(y,\widetilde{y})\cdot Z_s^{t,m}(y)+\big(\partial_x\partial_\mu g\big)_s^{t,m}(y,\widetilde{y});\\ \big(L_{\mu\mu}\big)_s^{t,m}(y,\widetilde{y},\widehat{y}):=\big(\partial_\mu\partial_\mu f\big)_s^{t,m}(y,\widetilde{y},\widehat{y})\cdot Z_s^{t,m}(y)+\big(\partial_\mu\partial_\mu g\big)_s^{t,m}(y,\widetilde{y},\widehat{y});\end{cases}\tag{7.12}$$

$(L_{\mu\widetilde{x}})_s^{t,m}(y,\widetilde{y})$, $(L_{\mu a})_s^{t,m}(y,\widetilde{y})$, $(L_{xa})_s^{t,m}(y)$ and $(L_{aa})_s^{t,m}(y)$ are also defined similarly to the notations above; here, the right subscript of L is employed to denote the differentiation of f and g with respect to the variables specified at the right subscript.

7.3. Crucial a priori estimate

Now, we are ready to establish some *a priori* estimates. For any fixed $t \in [0,T]$ and $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, it follows from Theorems 6.3 and 6.4 that there is a (local-in-time) unique solution pair $\big(X_s^{t,m}(x), Z_s^{t,m}(x) := \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m)\big)$ to (2.12), each component function of which is continuously differentiable in $x \in \mathbb{R}^{d_x}$ and continuously L-differentiable in $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$ with the corresponding derivatives being continuously differentiable in $s \in [t,T]$. Then the couples $\big(\partial_m\big(X_s^{t,m}(x)\big)(y), \partial_m\big(Z_s^{t,m}(x)\big)(y)\big)$ and $\big(\partial_x\big(X_s^{t,m}(x)\big), \partial_x\big(Z_s^{t,m}(x)\big)\big)$ satisfy the linear FBODE systems, respectively:

$$\begin{cases} \dfrac{d}{ds}\partial_m\big(X_s^{t,m}(x)\big)(y) = \partial_m\big(f_s^{t,m}(x)\big)(y), \\ \quad \partial_m\big(X_t^{t,m}(x)\big)(y) = 0; \\ -\dfrac{d}{ds}\partial_m\big(Z_s^{t,m}(x)\big)(y) = (\partial_x f)_s^{t,m}(x)\cdot\partial_m\big(Z_s^{t,m}(x)\big)(y) + \partial_m\big(X_s^{t,m}(x)\big)(y)^\top (L_{xx})_s^{t,m}(x) \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^{d_x}} \partial_m\big(X_s^{t,m}(\widetilde{y})\big)(y)^\top (L_{x\mu})_s^{t,m}(x,\widetilde{y})dm(\widetilde{y}) \\ \qquad\qquad + \partial_y\big(X_s^{t,m}(y)\big)^\top (L_{x\mu})_s^{t,m}(x,y) + \partial_m\big(\widehat{\alpha}_s^{t,m}(x)\big)(y)^\top (L_{xa})_s^{t,m}(x), \\ \quad \partial_m\big(Z_T^{t,m}(x)\big)(y) = (\partial_x\partial_x k)_T^{t,m}(x)\cdot\partial_m\big(X_T^{t,m}(x)\big)(y) + (\partial_\mu\partial_x k)_T^{t,m}(x,y)\cdot\partial_y\big(X_T^{t,m}(y)\big) \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^{d_x}} (\partial_\mu\partial_x k)_T^{t,m}(x,\widetilde{x})\cdot\partial_m\big(X_T^{t,m}(\widetilde{x})\big)(y)dm(\widetilde{x}); \end{cases} \tag{7.13}$$

and

$$\begin{cases} \dfrac{d}{ds}\partial_x\big(X_s^{t,m}(x)\big) = \partial_x\big(f_s^{t,m}(x)\big), \\ \quad \partial_x\big(X_t^{t,m}(x)\big) = \mathcal{I}_{d_x\times d_x}; \\ -\dfrac{d}{ds}\partial_x\big(Z_s^{t,m}(x)\big) = \big(\partial_x f\big)_s^{t,m}(x)\cdot\partial_x\big(Z_s^{t,m}(x)\big) + \partial_x\big(X_s^{t,m}(x)\big)^\top (L_{xx})_s^{t,m}(x) \\ \qquad\qquad + \partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big)^\top (L_{xa})_s^{t,m}(x), \\ \quad \partial_x\big(Z_T^{t,m}(x)\big) = \big(\partial_x\partial_x k\big)_T^{t,m}(x)\cdot\partial_x\big(X_T^{t,m}(x)\big); \end{cases} \tag{7.14}$$

where $\mathcal{I}_{d_x\times d_x}$ is the $d_x \times d_x$ identity matrix.

Remark 7.4. Here, let us recall that the flow map $X_s^{t,m}(\cdot)$ actually maps the initial point x at the initial time t to another location at the later times s by using the forward dynamics, so as in (2.12), the initial condition of this flow map is $X_s^{t,m}(x) = x$, which is the identity map from $\mathbb{R}^{d_x}$ to $\mathbb{R}^{d_x}$. Thus, the x and m-derivatives of this flow map at the initial time t are as in (7.13) and (7.14).

For the sake of convenience, we denote

$$L^* := \|\!|\gamma|\!\|_{2,[t,T]} < \infty, \tag{7.15}$$

where $\|\!|\cdot|\!\|_2$ was defined in (6.7). Here, at least close to the terminal time T, the finiteness of L^* is guaranteed by the construction of the local-in-time solution, which was ensured at least for $t \in [0,T]$ with $T - t \le \epsilon_2$, where ϵ_2 was given in Theorem 6.4. Furthermore, the decoupling field γ in (7.15) was constructed in Theorem 6.3. Then, by using (4.20),

$$\big|Z_t^{t,m}(x) - Z_t^{t,m_0}(x_0)\big| = \big|\gamma(t,x,m) - \gamma(t,x_0,m_0)\big| \le\ L^*\Big(|x - x_0| + W_1(m,m_0)\Big).$$

By using Hypothesis (**h1**), one can easily check that, for any fixed $t\in[0,T]$, the pair $\Big(X_s^{t,\delta_0}(0)\equiv 0, Z_s^{t,\delta_0}(0)\equiv 0\Big)$, as a function couple in time s, solves the FBODE (2.12) with fixed $x=0$ and $m=\delta_0$. Therefore, we have $\big|Z_t^{t,m}(x)\big|=\big|\gamma(t,x,m)\big|\le L^*\Big(|x|+\|m\|_1\Big)$, which further implies that $\big\|Z_t^{t,m}\big\|_{L_m^{1,d_x}}=\big\|\gamma(t,\cdot,m)\big\|_{L_m^{1,d_x}}\le 2L^*\|m\|_1$ by taking an integration. Furthermore, we also have

$$\big|Z_s^{t,m}(x)\big|=\big|\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m)\big|\le L^*\Big(|X_s^{t,m}(x)|+\|X_s^{t,m}\|_{L_m^{1,d_x}}\Big),\tag{7.16}$$

$$\big\|Z_s^{t,m}\big\|_{L_m^{1,d_x}}\le 2L^*\|X_s^{t,m}\|_{L_m^{1,d_x}}.\tag{7.17}$$

Now, we have the following *a priori* estimates.

Theorem 7.5. *Under Assumptions* (**a1**)*–*(**a3**) *and Hypotheses* (**h1**)*–*(**h3**)*, let $\gamma(s,x,m)$ be the decoupling field for the FBODE system* (6.1) *on $[t,T]$ with terminal data $p(x,\mu)=\partial_x k(x,\mu)$, which is differentiable in $x\in\mathbb{R}^{d_x}$ and L-differentiable in $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$, and also suppose that $L^*\le 2\max\{L_0^*,\Lambda_k\}$, we then have the following estimate:*

$$L^*\le L_0^*,\tag{7.18}$$

where L^* is defined in (7.15) *and L_0^* to be defined in* (7.56)*.*

Remark 7.6. In particular, we shall show that the constant L_0^* defined in (7.56) is always larger than Λ_k; also see Remark 7.7 for more detailed justification. In the statement of Theorem 7.5, the use of $\max\{L_0^*,\Lambda_k\}$ (instead of L_0^*) is to strongly emphasize that, when the time t is close to the terminal time T, the quantity L^* must be less than or equal to $2\Lambda_k$. This fact is extremely important since the estimates of γ must strongly depend on the terminal cost function k.

Proof. Define the pair $\big(X_s^{t,m}(x),Z_s^{t,m}(x)\big)$ by the equation (6.9) and $Z_s^{t,m}(x):=\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m)$, then it is a solution pair of FBODE system (6.1) by (6.8). The differentiabilities of $X_s^{t,m}(x)$ and $Z_s^{t,m}(x)$ that were stated in Theorem 6.4 follow by the properties of $\gamma(s,x,\mu)$.

The idea of proof follows from that of Theorem 8.3 in our previous paper [52] and will also be progressed in five steps. In Step 1, we shall first provide some calculus identities and elementary estimates that are useful in deriving estimates for the solutions to the Jacobian flows. Step 2 provides the preliminary estimates on different derivatives of γ *via* the backward ODE. To control the upper bounds of these preliminary estimates, we shall derive our major estimates (*e.g.*, (7.39) and (7.44)) by integrating the inner products of forward and backward flows in Steps 3 and 4. Finally, in Step 5, we shall combine all the aforementioned estimates to obtain a uniform (in the terminal time T) bound on the derivatives of γ.

It follows from (4.27) that $\frac{1}{2}k_0=2\max\{L_0^*,\Lambda_k\}\ge L^*$, so using (4.28) and (7.16), we know that $\big(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)\big)\in c_{k_0}$, and hence the unique minimizer $\widehat{\alpha}_s^{t,m}(x):=\widehat{\alpha}\big(X_s^{t,m}(x),X_s^{t,m}\#m,Z_s^{t,m}(x)\big)$ always exists according to Proposition 4.6. Furthermore, Hypothesis (**h3**) implies that $\bar{l}_fL^*\le\frac{1}{2}k_0\bar{l}_f\le\frac{1}{20}\min\{\lambda_g,\lambda_1\}$ and $\bar{l}_g\le\frac{1}{8}\lambda_g$. Denote $\Lambda_h:=\frac{1}{2}k_0\bar{l}_f+\Lambda_g\le\frac{1}{20}\min\{\lambda_1,\lambda_g\}+\Lambda_g$ and $\bar{l}_h:=\frac{1}{2}k_0\bar{l}_f+\bar{l}_g\le\frac{1}{20}\min\{\lambda_1,\lambda_g\}+\frac{1}{8}\min\{\lambda_1,\lambda_g\}<\frac{1}{5}\min\{\lambda_1,\lambda_g\}$.

Step 1 (Calculus identities and some elementary estimates) We shall first introduce a few calculus identities that will be useful in Steps 3-4, and derive some elementary estimates that will be used in Steps 2-5. First of all, differentiating (4.30) with respect to x, μ and z, respectively, it yields that

$$\begin{aligned}&\partial_x\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z+\partial_x\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))\\&\qquad+\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))\Big)\cdot\partial_x\widehat{\alpha}(x,\mu,z)=0,\end{aligned}\tag{7.19}$$

$$\partial_\mu\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))(\widetilde{x})\cdot z+\partial_\mu\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))(\widetilde{x})$$

$$+\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))\Big)\cdot\partial_\mu\widehat{\alpha}(x,\mu,z)(\widetilde{x})=0, \tag{7.20}$$

$$\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))+\Big(\partial_a\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z+\partial_a\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))\Big)\cdot\partial_z\widehat{\alpha}(x,\mu,z)=0, \tag{7.21}$$

and hence, by considering (7.21) multiplied with $\partial_x\widehat{\alpha}(x,\mu,z)$ minus (7.19) multiplied with $\partial_z\widehat{\alpha}(x,\mu,z)$, and (7.21) multiplied with $\partial_\mu\widehat{\alpha}(x,\mu,z)$ minus (7.19) multiplied with $\partial_z\widehat{\alpha}(x,\mu,z)$, we obtain

$$\begin{aligned}&\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot\partial_x\widehat{\alpha}(x,\mu,z)\\&\qquad-\Big(\partial_x\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot z+\partial_x\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))\Big)\cdot\partial_z\widehat{\alpha}(x,\mu,z)=0,\end{aligned} \tag{7.22}$$

$$\begin{aligned}&\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))\cdot\partial_\mu\widehat{\alpha}(x,\mu,z)(\widetilde{x})\\&\qquad-\Big(\partial_\mu\partial_a f(x,\mu,\widehat{\alpha}(x,\mu,z))(\widetilde{x})\cdot z+\partial_\mu\partial_a g(x,\mu,\widehat{\alpha}(x,\mu,z))(\widetilde{x})\Big)\cdot\partial_z\widehat{\alpha}(x,\mu,z)=0.\end{aligned} \tag{7.23}$$

Next, we are going to derive some elementary estimates and understand how these estimates depend on L^* as follows. It follows from (4.3), (4.6) and (7.16) that for any $\xi\in\mathbb{R}^{d_a}$,

$$\xi^\top\Big((\partial_a\partial_a f)_s^{t,m}(x)\cdot Z_s^{t,m}(x)+(\partial_a\partial_a g)_s^{t,m}(x)\Big)\xi\geq\Big(\lambda_g-\bar{l}_f L^*\Big)|\xi|^2\geq\frac{19}{20}\lambda_g|\xi|^2>0. \tag{7.24}$$

By (4.2), (4.3), (4.10), (7.16), (7.19)–(7.21) and the first inequality of (7.24), we obtain

$$\Big\|(\partial_x\widehat{\alpha})_s^{t,m}(x)\Big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\vee\Big\|(\partial_\mu\widehat{\alpha})_s^{t,m}(x,\widetilde{x})\Big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\leq\frac{20\bar{l}_h}{19\lambda_g}, \tag{7.25}$$

$$\text{and }\Big\|(\partial_z\widehat{\alpha})_s^{t,m}(x)\Big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_a})}\leq\frac{20\Lambda_f}{19\lambda_g}. \tag{7.26}$$

Furthermore, applying (4.2), (7.25) and (7.26) to (7.11), one yields that

$$\begin{aligned}\Big|\partial_m\big(\widehat{\alpha}_s^{t,m}(x)\big)(y)\Big|\leq&\frac{20\bar{l}_h}{19\lambda_g}\Big(\Big|\partial_m\big(X_s^{t,m}(x)\big)(y)\Big|+\Big|\partial_y\big(X_s^{t,m}(y)\big)\Big|+\int_{\mathbb{R}^{d_x}}\Big|\partial_m\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})\Big)\\&+\frac{20\Lambda_f}{19\lambda_g}\Big|\partial_m\big(Z_s^{t,m}(x)\big)(y)\Big|;\end{aligned} \tag{7.27}$$

$$\Big|\partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big)\Big|\leq\frac{20\bar{l}_h}{19\lambda_g}\Big|\partial_x\big(X_s^{t,m}(x)\big)\Big|+\frac{20\Lambda_f}{19\lambda_g}\Big|\partial_x\big(Z_s^{t,m}(x)\big)\Big|; \tag{7.28}$$

$$\begin{aligned}\Big|\partial_m\big(f_s^{t,m}(x)\big)(y)\Big|\leq\Lambda_f\Big(&\Big|\partial_m\big(X_s^{t,m}(x)\big)(y)\Big|+\Big|\partial_y\big(X_s^{t,m}(y)\big)\Big|+\int_{\mathbb{R}^{d_x}}\Big|\partial_m\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})\\&+\Big|\partial_m\big(\widehat{\alpha}_s^{t,m}(x)\big)(y)\Big|\Big);\end{aligned} \tag{7.29}$$

$$\Big|\partial_x\big(f_s^{t,m}(x)\big)\Big|\leq\Lambda_f\Big(\Big|\partial_x\big(X_s^{t,m}(x)\big)\Big|+\Big|\partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big)\Big|\Big). \tag{7.30}$$

Step 2 (Preliminary estimates on derivatives of γ) For $i=1,2,3,...,d_x$, denote $\partial_{m_i}\gamma(t,x,m)(y):=\partial_{y_i}\frac{\delta}{\delta m}\gamma(t,x,m)(y)$. By using Assumptions (**a1**)–(**a3**) and Hypotheses (**h1**) and (**h3**), we can obtain, for

$i=1,2,3,...,d_x$,

$$
\begin{aligned}
&\Big|\partial_{m_i}\gamma(t,x,m)(y)\Big|^2 = \Big|\partial_{m_i}\big(Z_t^{t,m}(x)\big)(y)\Big|^2\\
=&\Big|\partial_{m_i}\big(Z_T^{t,m}(x)\big)(y)\Big|^2 - 2\int_t^T \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\cdot\frac{d}{ds}\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)ds\\
\le&\Lambda_k^2\bigg(\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})+\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|+\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|\bigg)^2\\
&+2\int_t^T\bigg(\Lambda_h\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})+\Lambda_h\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|\\
&\qquad+\Lambda_h\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|+\bar{l}_h\Big|\partial_{m_i}\big(\widehat{\alpha}_s^{t,m}(x)\big)(y)\Big|+\Lambda_f\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|\bigg)\cdot\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|ds\\
&\text{(by using (4.2), (4.3), (4.9), (4.10) and (4.13))}\\
\le&\Lambda_k^2\bigg(\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})+\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|+\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|\bigg)^2\\
&+2\int_t^T\bigg(\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|\\
&\qquad+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|+\Big(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h\Big)\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|\bigg)\cdot\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|ds\\
&\text{(by using (7.27))}\\
\le&3\Lambda_k^2\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y)\Big|^2dm(\widetilde{x})+3\Lambda_k^2\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2+3\Lambda_k^2\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
&+\int_t^T\Bigg(\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|^2dm(\widetilde{x})+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2\\
&\qquad+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2+\bigg(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+3\Lambda_h+\frac{60\bar{l}_h^2}{19\lambda_g}\bigg)\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2\Bigg)ds
\end{aligned}\tag{7.31}
$$

and, for $i=1,2,3,...,d_x$, we also consider

$$
\begin{aligned}
&\Big|\partial_{x_i}\gamma(t,x,m)\Big|^2=\Big|\partial_{x_i}\big(Z_t^{t,m}(x)\big)\Big|^2=\Big|\partial_{x_i}\big(Z_T^{t,m}(x)\big)\Big|^2-2\int_t^T\partial_{x_i}\big(Z_s^{t,m}(x)\big)\cdot\frac{d}{ds}\partial_{x_i}\big(Z_s^{t,m}(x)\big)ds\\
\le&\Lambda_k^2\Big|\partial_{x_i}\big(X_T^{t,m}(x)\big)\Big|^2+2\int_t^T\bigg(\Lambda_h\Big|\partial_{x_i}\big(X_s^{t,m}(x)\big)\Big|+\bar{l}_h\Big|\partial_{x_i}\big(\widehat{\alpha}_s^{t,m}(x)\big)\Big|+\Lambda_f\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|\bigg)\cdot\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|ds\\
&\text{(by using (4.2), (4.3), (4.9), (4.10) and (4.13))}\\
\le&\Lambda_k^2\Big|\partial_{x_i}\big(X_T^{t,m}(x)\big)\Big|^2+2\int_t^T\bigg(\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{x_i}\big(X_s^{t,m}(x)\big)\Big|+\Big(\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+\Lambda_f\Big)\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|\bigg)\cdot\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|ds\\
&\text{(by using (7.28))}\\
\le&\Lambda_k^2\Big|\partial_{x_i}\big(X_T^{t,m}(x)\big)\Big|^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\int_t^T\Big|\partial_{x_i}\big(X_s^{t,m}(x)\big)\Big|^2ds+\Big(\frac{40\Lambda_f}{19\lambda_g}\bar{l}_h+2\Lambda_f+\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\int_t^T\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|^2ds.
\end{aligned}\tag{7.32}
$$

Then, by integrating (7.31) with respect to x and and using Jensen's inequality to bound L^1-norm by the corresponding L^2-norm, we have, for $i = 1, 2, 3, ..., d_x$,

$$
\begin{aligned}
&\Big\|\partial_{m_i}\gamma(t,\cdot,m)(y)\Big\|^2_{L^{2,d_x}_m} = \Big\|\partial_{m_i}\big(Z^{t,m}_t(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\\
\leq & 6\Lambda_k^2\Big\|\partial_{m_i}\big(X^{t,m}_T(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m} + 3\Lambda_k^2\Big|\partial_{y_i}\big(X^{t,m}_T(y)\big)\Big|^2\\
&+\int_t^T\Bigg(\Big(2\Lambda_h+\frac{40\bar{l}_h^2}{19\lambda_g}\Big)\Big\|\partial_{m_i}\big(X^{t,m}_s(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X^{t,m}_s(y)\big)\Big|^2\\
&\qquad+\Big(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+3\Lambda_h+\frac{60\bar{l}_h^2}{19\lambda_g}\Big)\Big\|\partial_{m_i}\big(Z^{t,m}_s(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\Bigg)ds.
\end{aligned}
\tag{7.33}
$$

To bound the right hand sides of (7.31)-(7.33), we need the following major estimates.
Step 3 (Major estimate 1 of the following (7.39)) For $i = 1, 2, 3, ..., d_x$, by using (7.11), (7.14),

$$
\begin{aligned}
&\partial_{x_i}\big(Z^{t,m}_T(x)\big)\cdot\partial_{x_i}\big(X^{t,m}_T(x)\big)-\partial_{x_i}\big(Z^{t,m}_t(x)\big)\cdot\partial_{x_i}\big(X^{t,m}_t(x)\big)\\
=&\int_t^T\Big(\partial_{x_i}\big(Z^{t,m}_s(x)\big)\cdot\frac{d}{ds}\partial_{x_i}\big(X^{t,m}_s(x)\big)+\partial_{x_i}\big(X^{t,m}_s(x)\big)\cdot\frac{d}{ds}\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big)ds\\
=&\int_t^T\Bigg(\mathcal{T}_1+\partial_{x_i}\big(Z^{t,m}_s(x)\big)\cdot\Big(\big(\partial_a f\big)^{t,m}_s(x)\cdot\Big(\big(\partial_z\widehat{\alpha}\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big)\Big)\\
&\qquad-\mathcal{T}_2-\partial_{x_i}\big(X^{t,m}_s(x)\big)^\top\big(L_{xx}\big)^{t,m}_s(x)\partial_{x_i}\big(X^{t,m}_s(x)\big)\\
&\qquad-\Big(\big(\partial_x\widehat{\alpha}\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(X^{t,m}_s(x)\big)\Big)^\top\big(L_{xa}\big)^{t,m}_s(x)\partial_{x_i}\big(X^{t,m}_s(x)\big)\Bigg)ds,
\end{aligned}
$$

where $\mathcal{T}_1 := \partial_{x_i}\big(Z^{t,m}_s(x)\big)\cdot\Big(\big(\partial_x f\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(X^{t,m}_s(x)\big)+\big(\partial_a f\big)^{t,m}_s(x)\cdot\big(\partial_x\widehat{\alpha}\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(X^{t,m}_s(x)\big)\Big)$ and $\mathcal{T}_2 := \partial_{x_i}\big(X^{t,m}_s(x)\big)^\top\big(\partial_x f\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(Z^{t,m}_s(x)\big)+\Big(\big(\partial_z\widehat{\alpha}\big)^{t,m}_s(x)\cdot\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big)^\top\big(L_{xa}\big)^{t,m}_s(x)\partial_{x_i}\big(X^{t,m}_s(x)\big)$; by (7.22), $\mathcal{T}_1=\mathcal{T}_2$ and we can cancel these two terms. Thus, by using (4.3), (4.7), (4.10), (4.41), (7.16) and (7.25)), we have

$$
\begin{aligned}
&\partial_{x_i}\big(Z^{t,m}_T(x)\big)\cdot\partial_{x_i}\big(X^{t,m}_T(x)\big)-\partial_{x_i}\big(Z^{t,m}_t(x)\big)\cdot\partial_{x_i}\big(X^{t,m}_t(x)\big)\\
\leq&\int_t^T\Bigg(-\frac{\lambda_f^2}{\Lambda_h}\Big|\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big|^2-\Big(\lambda_g-\frac{1}{2}k_0\bar{l}_f-\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{x_i}\big(X^{t,m}_s(x)\big)\Big|^2\Bigg)ds,
\end{aligned}
$$

which implies, in light of (7.14) for the terminal condition on the left hand side, by using (4.11), we deduce

$$
\begin{aligned}
&\lambda_k\Big|\partial_{x_i}\big(X^{t,m}_T(x)\big)\Big|^2-\partial_{x_i}\big(Z^{t,m}_t(x)\big)_i\\
\leq&\int_t^T\Bigg(-\frac{\lambda_f^2}{\Lambda_h}\Big|\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big|^2-\Big(\lambda_g-\frac{1}{2}k_0\bar{l}_f-\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{x_i}\big(X^{t,m}_s(x)\big)\Big|^2\Bigg)ds, && (7.34)\\
\leq&\int_t^T\Big(-\lambda_z\Big|\partial_{x_i}\big(Z^{t,m}_s(x)\big)\Big|^2-\lambda_x\Big|\partial_{x_i}\big(X^{t,m}_s(x)\big)\Big|^2\Big)ds, && (7.35)
\end{aligned}
$$

where $\big(Z_t^{t,m}(x)\big)_i$ is the i-th component of $Z_t^{t,m}(x)$ and

$$\lambda_z := \frac{\lambda_f^2}{\Lambda_g + \lambda_g/20} \leq \frac{\lambda_f^2}{\Lambda_h}; \tag{7.36}$$

$$\lambda_x := \frac{4}{5}\lambda_g < \lambda_g - \frac{1}{20}\lambda_g - \frac{1}{20}\lambda_g \leq \lambda_g - \frac{1}{2}k_0\bar{l}_f - \frac{20\bar{l}_h^2}{19\lambda_g}. \tag{7.37}$$

Thus, by using (7.32) and (7.35), we can obtain

$$\begin{aligned}
&\Big|\partial_{x_i}\gamma(t,x,m)\Big|^2 = \Big|\partial_{x_i}\big(Z_t^{t,m}(x)\big)\Big|^2\\
\leq\ & \frac{\Lambda_k^2}{\lambda_k}\lambda_k\Big|\partial_{x_i}\big(X_T^{t,m}(x)\big)\Big|^2 + \frac{1}{\lambda_z}\left(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right)\lambda_z\int_t^T \Big|\partial_{x_i}\big(X_s^{t,m}(x)\big)\Big|^2 ds\\
&+\frac{1}{\lambda_x}\left(\frac{40\Lambda_f}{19\lambda_g}\bar{l}_h + 2\Lambda_f + \Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right)\lambda_x\int_t^T \Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|^2 ds\\
\leq\ & \max\left\{\frac{\Lambda_k^2}{\lambda_k}, \frac{1}{\lambda_z}\left(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right), \frac{1}{\lambda_x}\left(\frac{40\Lambda_f}{19\lambda_g}\bar{l}_h + 2\Lambda_f + \Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right)\right\}\Big|\partial_{x_i}\big(Z_t^{t,m}(x)\big)\Big|,
\end{aligned}$$

which implies the following *a priori* estimate, for any $i = 1, 2, 3, ..., d_x$, by dividing both sides of above by $\Big|\partial_{x_i}\big(Z_t^{t,m}(x)\big)\Big|$,

$$\begin{aligned}
&\Big|\partial_{x_i}\gamma(t,x,m)\Big| = \Big|\partial_{x_i}\big(Z_t^{t,m}(x)\big)\Big|\\
\leq\ & \max\left\{\frac{\Lambda_k^2}{\lambda_k}, \frac{1}{\lambda_z}\left(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right), \frac{1}{\lambda_x}\left(\frac{40\Lambda_f}{19\lambda_g}\bar{l}_h + 2\Lambda_f + \Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\right)\right\}\\
\leq\ & \max\left\{\frac{\Lambda_k^2}{\lambda_k}, \frac{1}{\lambda_z}\left(\Lambda_h + \frac{1}{20}\lambda_g\right), \frac{1}{\lambda_x}\left(\frac{5}{2}\Lambda_f + \Lambda_h + \frac{1}{20}\lambda_g\right)\right\} =: L_1^*,
\end{aligned} \tag{7.38}$$

since $\bar{l}_f L^* \leq \frac{1}{20}\lambda_g$, $\bar{l}_h < \frac{1}{5}\lambda_g$ and $\frac{20\bar{l}_h}{19\lambda_g} < \frac{1}{4}$. Therefore, we finally have

$$\lambda_k\Big|\partial_{x_i}\big(X_T^{t,m}(x)\big)\Big|^2 + \int_t^T\left(\lambda_z\Big|\partial_{x_i}\big(Z_s^{t,m}(x)\big)\Big|^2 + \lambda_x\Big|\partial_{x_i}\big(X_s^{t,m}(x)\big)\Big|^2\right)ds \leq L_1^*. \tag{7.39}$$

Step 4 (Major estimate 2 of the following (7.44)) For any $i = 1, 2, 3, ..., d_x$ and any $y \in \mathbb{R}^{d_x}$, integrating the inner product $\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\cdot\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)$ with respect to s over $[t,T]$ and using (7.13) yields

$$\begin{aligned}
&\partial_{m_i}\big(Z_T^{t,m}(x)\big)(y)\cdot\partial_{m_i}\big(X_T^{t,m}(x)\big)(y) - \partial_{m_i}\big(Z_t^{t,m}(x)\big)(y)\cdot\partial_{m_i}\big(X_t^{t,m}(x)\big)(y)\\
=&\int_t^T\left(\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\cdot\frac{d}{ds}\partial_{m_i}\big(X_s^{t,m}(x)\big)(y) + \partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\cdot\frac{d}{ds}\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\right)ds\\
=&\int_t^T\Bigg(\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\cdot\left(\big(\partial_\mu f\big)_s^{t,m}(x,y)\partial_{y_i}\big(X_s^{t,m}(y)\big) + \big(\partial_a f\big)_s^{t,m}(x)\cdot\left(\big(\partial_\mu\widehat{\alpha}\big)_s^{t,m}(x,y)\cdot\partial_{y_i}\big(X_s^{t,m}(y)\big)\right)\right)\\
&\qquad - \partial_{y_i}\big(X_s^{t,m}(y)\big)^\top\left(L_{x\mu}\right)_s^{t,m}(x,y)\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)
\end{aligned}$$

$$
\begin{aligned}
&- \bigg(\big(\partial_x \widehat{\alpha}\big)_s^{t,m}(x) \cdot \partial_{m_i}\big(X_s^{t,m}(x)\big)(y) + \big(\partial_\mu \widehat{\alpha}\big)_s^{t,m}(x,y) \cdot \partial_{y_i}\big(X_s^{t,m}(y)\big) \\
&\qquad\qquad + \int_{\mathbb{R}^{d_x}} \big(\partial_\mu \widehat{\alpha}\big)_s^{t,m}(x,\widehat{x}) \cdot \partial_{m_i}\big(X_s^{t,m}(\widehat{x})\big)(y) dm(\widehat{x}) \bigg) \cdot \left(L_{xa}\right)_s^{t,m}(x) \partial_{m_i}\big(X_s^{t,m}(x)\big)(y) \\
&+ \mathcal{T}_3 - \mathcal{T}_4 + \mathcal{T}_5 - \mathcal{T}_6 \Bigg) ds,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{T}_3 :=& \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) \cdot \Big(\big(\partial_x f\big)_s^{t,m}(x) \partial_{m_i}\big(X_s^{t,m}(x)\big)(y) + \big(\partial_a f\big)_s^{t,m}(x) \cdot \big(\partial_x \widehat{\alpha}\big)_s^{t,m}(x) \cdot \partial_{m_i}\big(X_s^{t,m}(x)\big)(y) \Big), \\
\mathcal{T}_4 :=& \partial_{m_i}\big(X_s^{t,m}(x)\big)(y)^\top \big(\partial_x f\big)_s^{t,m}(x) \cdot \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) + \Big(\big(\partial_z \widehat{\alpha}\big)_s^{t,m}(x) \cdot \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) \Big) \cdot \left(L_{xa}\right)_s^{t,m}(x) \partial_{m_i}\big(X_s^{t,m}(x)\big)(y), \\
\mathcal{T}_5 :=& \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) \cdot \Big(\int_{\mathbb{R}^{d_x}} \big(\partial_\mu f\big)_s^{t,m}(x,\widehat{x}) \partial_{m_i}\big(X_s^{t,m}(\widehat{x})\big)(y) dm(\widehat{x}) \\
&\qquad + \big(\partial_a f\big)_s^{t,m}(x) \cdot \int_{\mathbb{R}^{d_x}} \big(\partial_\mu \widehat{\alpha}\big)_s^{t,m}(x,\widehat{x}) \cdot \partial_{m_i}\big(X_s^{t,m}(\widehat{x})\big)(y) dm(\widehat{x}) \Big), \\
\mathcal{T}_6 :=& - \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) \cdot \Big(\big(\partial_a f\big)_s^{t,m}(x) \cdot \Big(\big(\partial_z \widehat{\alpha}\big)_s^{t,m}(x) \cdot \partial_{m_i}\big(Z_s^{t,m}(x)\big)(y) \Big) \Big) \\
&+ \partial_{m_i}\big(X_s^{t,m}(x)\big)(y)^\top \left(L_{xx}\right)_s^{t,m}(x) \partial_{m_i}\big(X_s^{t,m}(x)\big)(y) \\
&+ \int_{\mathbb{R}^{d_x}} \partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)^\top \left(L_{x\mu}\right)_s^{t,m}(x,\widetilde{x}) dm(\widetilde{x}) \partial_{m_i}\big(X_s^{t,m}(x)\big)(y),
\end{aligned}
$$

by using (7.11), (7.22), $\mathcal{T}_3 = \mathcal{T}_4$ and we can cancel these two terms. Special care must be exercised regarding the term $\mathcal{T}_5$ that exhibited cancellation through integration $\int_{\mathbb{R}^{d_x}} \mathcal{T}_i dm(x)$ when combined with adjoint term in our prior mean field control framework [cf. [52]]. However, this critical cancellation property fails to hold in the current mean field game (MFG) setting. Then, by using (4.2), (4.3), (4.9), (4.10), (7.16) and (7.25), we have

$$
\begin{aligned}
&\partial_{m_i}\big(Z_T^{t,m}(x)\big)(y) \cdot \partial_{m_i}\big(X_T^{t,m}(x)\big)(y) - \partial_{m_i}\big(Z_t^{t,m}(x)\big)(y) \cdot \partial_{m_i}\big(X_t^{t,m}(x)\big)(y) \\
\leq & \int_t^T \Bigg(\Lambda_f \Big(1 + \frac{20\bar{l}_h}{19\lambda_g}\Big) \Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big| \Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big| + \mathcal{T}_5 - \mathcal{T}_6 \\
&\qquad + \frac{20\bar{l}_h^2}{19\lambda_g} \Big(\int_{\mathbb{R}^{d_x}} \Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big| dm(\widetilde{x}) + \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \Big) \cdot \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \\
&\qquad + \Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big) \Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big| \cdot \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \Bigg) ds
\end{aligned}
$$

which implies, by using (4.11) and (4.13) on the left hand side,

$$
\begin{aligned}
&\lambda_k \Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2 - \Lambda_k \Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big| \Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big| \\
&+ \int_{\mathbb{R}^{d_x}} \partial_{m_i}\big(X_T^{t,m}(x)\big)(y)^\top \big(\partial_\mu \partial_x k\big)_T^{t,m}(x,\widetilde{x}) \partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) \\
\leq & \int_t^T \Bigg(\Lambda_f \Big(1 + \frac{20\bar{l}_h}{19\lambda_g}\Big) \Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big| \Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big| + \mathcal{T}_5 - \mathcal{T}_6 \\
&\qquad + \frac{20\bar{l}_h^2}{19\lambda_g} \Big(\int_{\mathbb{R}^{d_x}} \Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big| dm(\widetilde{x}) + \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \Big) \cdot \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \\
&\qquad + \Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big) \Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big| \cdot \Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big| \Bigg) ds. \qquad (7.40)
\end{aligned}
$$

Then, integrating (7.40) with respect to x, we have

$$\begin{aligned}
&\lambda_k\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}-\Lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|_{L^{1,d_x}_m}\\
&+\int_{\mathbb{R}^{d_x}}\int_{\mathbb{R}^{d_x}}\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)^\top\big(\partial_\mu\partial_x k\big)_T^{t,m}(x,\widetilde{x})\partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y)dm(\widetilde{x})dm(x)\\
\le\ &\int_t^T\Bigg(\Lambda_f\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)\Big\|\partial_{m_i}\big(Z_s^{t,m}(\cdot)\big)(y)\Big\|_{L^{1,d_x}_m}\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|+\int_{\mathbb{R}^{d_x}}\left(\mathcal{T}_5-\mathcal{T}_6\right)dm(x)\\
&\qquad+\frac{20\bar{l}_h^2}{19\lambda_g}\Big(\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{1,d_x}_m}+\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\Big)\\
&\qquad+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|\cdot\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|_{L^{1,d_x}_m}\Bigg)ds,
\end{aligned}$$

which implies, by using (4.7), (4.8), (4.12), (7.1) and using Jensen's inequality to replace L^1-norm by L^2-norm,

$$\begin{aligned}
&\Big(\lambda_k-l_k-\varepsilon_1\Big)\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}-\frac{1}{4\varepsilon_1}\Lambda_k^2\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
\le\ &\int_t^T\Bigg(\frac{1}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2-\Big(\lambda_2-\varepsilon_1\Big)\Big\|\partial_{m_i}\big(Z_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\\
&\qquad-\Big(\lambda_1-\frac{40\bar{l}_h^2}{19\lambda_g}-k_0\bar{l}_f-\varepsilon_1\Big)\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\Bigg)ds,
\end{aligned}\tag{7.41}$$

where, we set

$$\varepsilon_1:=\min\left\{\frac{1}{2}(\lambda_k-l_k),\frac{1}{2}\lambda_2,\frac{2}{5}\lambda_1\right\},\tag{7.42}$$

and thus we can define the following positive constants, under Hypothesis (**h3**),

$$\begin{aligned}
&\overline{\lambda}_k:=\frac{1}{2}(\lambda_k-l_k)=\lambda_k-l_k-\varepsilon_1;\ \overline{\lambda}_z:=\frac{1}{2}\lambda_2\le\lambda_2-\varepsilon_1;\\
&\overline{\lambda}_x:=\frac{2}{5}\lambda_1\le\lambda_1-\frac{1}{10}\lambda_1-\frac{1}{10}\lambda_1-\varepsilon_1\le\lambda_1-\frac{40\bar{l}_h^2}{19\lambda_g}-k_0\bar{l}_f-\varepsilon_1.
\end{aligned}\tag{7.43}$$

Therefore, we have

$$\begin{aligned}
&\overline{\lambda}_k\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+\int_t^T\Big(\overline{\lambda}_z\Big\|\partial_{m_i}\big(Z_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+\overline{\lambda}_x\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\Big)ds\\
\le\ &\frac{1}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\int_t^T\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2ds+\frac{1}{4\varepsilon_1}\Lambda_k^2\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2.
\end{aligned}\tag{7.44}$$

Step 5 (Consequences of major estimates obtained in Steps 3 and 4) Thus, by using (7.33), (7.39) and (7.44), we can obtain the following *a priori* estimate, for any $i=1,2,3,...,d_x$,

$$\begin{aligned}
&\Big\|\partial_{m_i}\gamma(t,\cdot,m)(y)\Big\|^2_{L^{2,d_x}_m}=\Big\|\partial_{m_i}\big(Z_t^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\\
\le\ &\frac{6\Lambda_k^2}{\overline{\lambda}_k}\overline{\lambda}_k\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+3\Lambda_k^2\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2
\end{aligned}$$

$$
\begin{aligned}
&+\int_t^T \Bigg(\Big(2\Lambda_h+\frac{40\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_x}\overline{\lambda}_x\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2\\
&\qquad+\Big(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+3\Lambda_h+\frac{60\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_z}\overline{\lambda}_z\Big\|\partial_{m_i}\big(Z_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}\Bigg)ds\\
&\le\Big(3+\frac{L_2^*}{4\varepsilon_1}\Big)\Lambda_k^2\frac{1}{\lambda_k}\lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
&\quad+\int_t^T\Bigg(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}+\frac{L_2^*}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\Bigg)\frac{1}{\lambda_x}\lambda_x\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2ds\\
&\le L_1^*L_3^*,
\end{aligned}
\tag{7.45}
$$

where L_1^* was defined in (7.38) and

$$
\begin{aligned}
L_2^*&:=\max\left\{\frac{6\Lambda_k^2}{\overline{\lambda}_k},\Big(2\Lambda_g+\frac{1}{10}\lambda_g\Big)\frac{1}{\overline{\lambda}_x},\Big(\frac{5}{4}\Lambda_f+3\Lambda_g+\frac{3}{10}\lambda_g\Big)\frac{1}{\overline{\lambda}_z}\right\}\\
&\ge\max\left\{\frac{6\Lambda_k^2}{\overline{\lambda}_k},\Big(2\Lambda_h+\frac{40\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_x},\Big(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+3\Lambda_h+\frac{60\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_z}\right\},
\end{aligned}
\tag{7.46}
$$

and

$$
\begin{aligned}
L_3^*&:=\max\left\{\Big(3+\frac{L_2^*}{4\varepsilon_1}\Big)\Lambda_k^2\frac{1}{\lambda_k},\Big(\Lambda_g+\frac{1}{10}\lambda_g+\frac{L_2^*}{4\varepsilon_1}\Big(\frac{25}{16}\Lambda_f^2+\Big(\Lambda_g+\frac{1}{10}\lambda_g\Big)^2\Big)\Big)\frac{1}{\lambda_x}\right\}\\
&\ge\max\left\{\Big(3+\frac{L_2^*}{4\varepsilon_1}\Big)\Lambda_k^2\frac{1}{\lambda_k},\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}+\frac{L_2^*}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\Big)\frac{1}{\lambda_x}\right\}.
\end{aligned}
\tag{7.47}
$$

In addition, using (7.39) and (7.44), we have

$$
\begin{aligned}
&\overline{\lambda_k}\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}+\overline{\lambda}_z\int_t^T\Big\|\partial_{m_i}\big(Z_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}ds+\overline{\lambda}_x\int_t^T\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L^{2,d_x}_m}ds\\
&\le\frac{1}{4\varepsilon_1}\Lambda_k^2\frac{1}{\lambda_k}\lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2+\frac{1}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\frac{1}{\lambda_x}\lambda_x\int_t^T\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2ds\\
&\le L_1^*L_4^*,
\end{aligned}
\tag{7.48}
$$

where

$$
\begin{aligned}
L_4^*&:=\max\left\{\frac{1}{4\varepsilon_1}\Lambda_k^2\frac{1}{\lambda_k},\frac{1}{4\varepsilon_1}\Big(\frac{25}{16}\Lambda_f^2+\Big(\Lambda_g+\frac{1}{10}\lambda_g\Big)^2\Big)\frac{1}{\lambda_x}\right\}\\
&\ge\max\left\{\frac{1}{4\varepsilon_1}\Lambda_k^2\frac{1}{\lambda_k},\frac{1}{4\varepsilon_1}\Big(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big)\frac{1}{\lambda_x}\right\}.
\end{aligned}
\tag{7.49}
$$

Now, we go back to (7.40). Using (4.2), (4.3), (4.9), (4.10), (4.13), (7.25) and (7.26), we obtain

$$
\begin{aligned}
&\lambda_k\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2-\Lambda_k\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_T^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|-\Lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|\\
&\le\int_t^T\Bigg(\Lambda_f\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)\cdot\Big(\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|+\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_s^{t,m}(\widehat{x})\big)(y)\Big|dm(\widehat{x})\Big)\\
&\quad-\lambda_z\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2-\lambda_x\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2\\
&\quad+\frac{20\bar{l}_h^2}{19\lambda_g}\Big(\int_{\mathbb{R}^{d_x}}\Big|\partial_{m_i}\big(X_s^{t,m}(\widetilde{x})\big)(y)\Big|dm(\widetilde{x})+\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|\Big)\cdot\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|\\
&\quad+\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|\Bigg)ds.
\end{aligned}
\tag{7.50}
$$

Applying Young's inequality to (7.50) yields

$$
\begin{aligned}
&(\lambda_k - 2\varepsilon_2)\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2 - \frac{1}{4\varepsilon_2}\Lambda_k^2\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}} - \frac{1}{4\varepsilon_2}\Lambda_k^2\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
\leq\ & \int_t^T \Bigg(\frac{1}{4\varepsilon_2}\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2\cdot\bigg(\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2 + \Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}}\bigg)\\
&\quad - (\lambda_z - 2\varepsilon_2)\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2 - \Big(\lambda_x - 2\varepsilon_2 - \frac{20\bar{l}_h^2}{19\lambda_g}\Big)\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2\\
&\quad + \frac{1}{4\varepsilon_2}\Big(\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}} + \frac{1}{4\varepsilon_2}\Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2\Bigg)ds,
\end{aligned}
$$

which implies, by using (7.39) and (7.48),

$$
\begin{aligned}
&(\lambda_k - 2\varepsilon_2)\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2 + (\lambda_z - 2\varepsilon_2)\int_t^T\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2 ds + \Big(\lambda_x - 2\varepsilon_2 - \frac{20\bar{l}_h^2}{19\lambda_g}\Big)\int_t^T\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2 ds\\
\leq\ & \frac{1}{4\varepsilon_2}\Lambda_k^2\frac{1}{\overline{\lambda}_k}\overline{\lambda}_k\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}} + \frac{1}{4\varepsilon_2}\Lambda_k^2\frac{1}{\lambda_k}\lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
&+\int_t^T\Bigg(\frac{1}{4\varepsilon_2}\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2\cdot\bigg(\frac{1}{\lambda_x}\lambda_x\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2 + \frac{1}{\overline{\lambda}_x}\overline{\lambda}_x\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}}\bigg)\\
&\quad + \frac{1}{4\varepsilon_2}\Big(\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\frac{1}{\overline{\lambda}_x}\overline{\lambda}_x\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{1,d_x}} + \frac{1}{4\varepsilon_2}\Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\frac{1}{\lambda_x}\lambda_x\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2\Bigg)ds\\
\leq\ & L_1^* L_4^*(1+L_5^*),
\end{aligned}
$$

where

$$
\begin{aligned}
L_5^* :=\ & \max\left\{\frac{1}{4\varepsilon_2}\Lambda_k^2\frac{1}{\overline{\lambda}_k}, \frac{1}{4\varepsilon_2\overline{\lambda}_x}\Big(\frac{25}{16}\Lambda_f^2 + \frac{1}{400}\lambda_g^2\Big)\right\}\\
\geq\ & \max\left\{\frac{1}{4\varepsilon_2}\Lambda_k^2\frac{1}{\overline{\lambda}_k}, \frac{1}{4\varepsilon_2\overline{\lambda}_x}\bigg(\Lambda_f^2\Big(1+\frac{20\bar{l}_h}{19\lambda_g}\Big)^2 + \Big(\frac{20\bar{l}_h^2}{19\lambda_g}\Big)^2\bigg)\right\}.
\end{aligned}
\tag{7.51}
$$

Here, we choose

$$
\varepsilon_2 := \min\Big\{\frac{1}{4}\lambda_k, \frac{1}{4}\lambda_z, \frac{1}{8}\lambda_g\Big\}, \tag{7.52}
$$

and thus, $\lambda_k - 2\varepsilon_2 = \frac{1}{2}\lambda_k > 0$, $\lambda_z - 2\varepsilon_2 = \frac{1}{2}\lambda_z > 0$ and $\lambda_x - 2\varepsilon_2 - \frac{20\bar{l}_h^2}{19\lambda_g} \geq \frac{1}{2}\lambda_g > 0$ are all positive, and we have

$$
\frac{1}{2}\lambda_k\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2 + \frac{1}{2}\lambda_z\int_t^T\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2 ds + \frac{1}{2}\lambda_g\int_t^T\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2 ds \leq L_1^*L_4^*(1+L_5^*). \tag{7.53}
$$

Thus, by using (7.31), (7.39), (7.48) and (7.53), we can obtain the following *a priori* estimate, for $i = 1, 2, 3, ..., d_x$,

$$
\begin{aligned}
&\Big|\partial_{m_i}\gamma(t,x,m)(y)\Big|^2 = \Big|\partial_{m_i}\big(Z_t^{t,m}(x)\big)(y)\Big|^2\\
\leq\ & 3\Lambda_k^2\frac{1}{\overline{\lambda}_k}\overline{\lambda}_k\Big\|\partial_{m_i}\big(X_T^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{2,d_x}} + 6\Lambda_k^2\frac{1}{\lambda_k}\cdot\frac{1}{2}\lambda_k\Big|\partial_{m_i}\big(X_T^{t,m}(x)\big)(y)\Big|^2 + 3\Lambda_k^2\frac{1}{\lambda_k}\lambda_k\Big|\partial_{y_i}\big(X_T^{t,m}(y)\big)\Big|^2\\
&+\int_t^T\Bigg(\Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_x}\overline{\lambda}_x\Big\|\partial_{m_i}\big(X_s^{t,m}(\cdot)\big)(y)\Big\|^2_{L_m^{2,d_x}} + 2\Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_g}\cdot\frac{1}{2}\lambda_g\Big|\partial_{m_i}\big(X_s^{t,m}(x)\big)(y)\Big|^2\\
&\quad + \Big(\Lambda_h + \frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_x}\lambda_x\Big|\partial_{y_i}\big(X_s^{t,m}(y)\big)\Big|^2 + 2\Big(\Lambda_f + \frac{20\Lambda_f}{19\lambda_g}\bar{l}_h + 3\Lambda_h + \frac{60\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_z}\cdot\frac{1}{2}\lambda_z\Big|\partial_{m_i}\big(Z_s^{t,m}(x)\big)(y)\Big|^2\Bigg)ds\\
\leq\ & \Big(L_4^*(2+L_5^*)+1\Big)L_1^*L_6^*,
\end{aligned}
\tag{7.54}
$$

where

$$
\begin{aligned}
L_6^* :=& \max\Big\{6\Lambda_k^2\frac{1}{\lambda_k}, 2\Big(\frac{5}{4}\Lambda_f+3\Lambda_g+\frac{3}{10}\lambda_g\Big)\frac{1}{\lambda_z}, 2\Big(\Lambda_g+\frac{1}{10}\lambda_g\Big)\frac{1}{\lambda_g},\\
&3\Lambda_k^2\frac{1}{\overline{\lambda}_k}, \Big(\Lambda_g+\frac{1}{10}\lambda_g\Big)\frac{1}{\overline{\lambda}_x}, 3\Lambda_k^2\frac{1}{\lambda_k}, \Big(\Lambda_g+\frac{1}{10}\lambda_g\Big)\frac{1}{\lambda_x}\Big\}\\
\geq& \max\Big\{6\Lambda_k^2\frac{1}{\lambda_k}, 2\Big(\Lambda_f+\frac{20\Lambda_f}{19\lambda_g}\bar{l}_h+3\Lambda_h+\frac{60\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_z}, 2\Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_g},\\
&3\Lambda_k^2\frac{1}{\overline{\lambda}_k}, \Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\overline{\lambda}_x}, 3\Lambda_k^2\frac{1}{\lambda_k}, \Big(\Lambda_h+\frac{20\bar{l}_h^2}{19\lambda_g}\Big)\frac{1}{\lambda_x}\Big\}.
\end{aligned}
\tag{7.55}
$$

In conclusion, by (7.38) and (7.54), we can obtain,

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}^{d_x}, m\in\mathcal{P}_2(\mathbb{R}^{d_x}), y\in\mathbb{R}^{d_x}} \|\partial_x\gamma(t,x,m)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\vee\|\partial_m\gamma(t,x,m)(y)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\\
&\leq d_x\cdot\max\Big\{L_1^*, \sqrt{\Big(L_4^*\big(2+L_5^*\big)+1\Big)L_1^*L_6^*}\Big\} =: L_0^*.
\end{aligned}
\tag{7.56}
$$

Since $\gamma(s,x,m)$ is also a decoupling field for the FBODE system (6.1) on $[t,T]$, for an arbitrary $s\in[t,T]$, $(X_\tau^{s,m}(x), Z_\tau^{s,m}(x)=\gamma(\tau, X_\tau^{s,m}(x), X_\tau^{s,m}\#m))_{x\in\mathbb{R}^{d_x},\tau\in[s,T]}$ is also a solution pair to (6.1) over the time horizon $[s,T]$, so the above estimates, for instance, (7.39), (7.44), (7.45), (7.48), (7.53), (7.54) and (7.56) are also valid for $\gamma(s,x,m)=\gamma(s,X_s^{s,m}(x),X_s^{s,m}\#m)=Z_s^{s,m}(x)$; in other words, we still have

$$
\sup_{x\in\mathbb{R}^{d_x}, m\in\mathcal{P}_2(\mathbb{R}^{d_x}), y\in\mathbb{R}^{d_x}} \|\partial_x\gamma(s,x,m)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\vee\|\partial_m\gamma(s,x,m)(y)\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\leq L_0^*.
$$

Therefore, $L^*=\|\!|\gamma|\!\|_{2,[t,T]}\leq L_0^*$. We emphasize here that the positive constants L_i^*, $i=0,1,\cdots,6$, which are defined in (7.38), (7.46), (7.47), (7.49), (7.51), (7.55) and (7.56), only depend on λ_f, Λ_1, Λ_2, Λ_3, λ_1, λ_2, λ_g, Λ_g, l_g, λ_k, Λ_k and l_k, but not L^*, $\bar{l}_f$, $\bar{l}_g$ and T. ☐

Remark 7.7. (i) It is worth noting that for our choice of L_0^*, we always have $L_0^*>\Lambda_k$. More precisely, it follows from (4.11), (4.13) and (7.43) that $\lambda_k\leq 2\Lambda_k$ and $\overline{\lambda}_k\leq\frac{1}{2}\Lambda_k$. Thus, by (7.38) and (7.55), we have

$$
L_1^*\geq\frac{\Lambda_k^2}{\lambda_k}\geq\frac{1}{2}\Lambda_k \text{ and } L_6^*\geq\frac{3\Lambda_k^2}{\overline{\lambda}_k}\geq 6\Lambda_k,
\tag{7.57}
$$

and hence,

$$
L_0^*\geq d_x\cdot\sqrt{\Big(L_4^*\big(2+L_5^*\big)+1\Big)L_1^*L_6^*}\geq\sqrt{L_1^*L_6^*}\geq\sqrt{3}\Lambda_k>\Lambda_k.
\tag{7.58}
$$

(ii) By Part (i) of Remark 7.2, if we directly assume (7.3), then L_0^* only depends on λ_f, Λ_1, Λ_2, Λ_3, λ_g, Λ_g, l_g, λ_k, Λ_k and l_k, since λ_1 and λ_2 can be chosen such that they depend only on λ_f, Λ_1, Λ_2, λ_g, Λ_g and l_g.

7.4. Global existence

With the *a priori* estimate (7.18) in hand, we are ready to establish the global solution existence in time of the FBODE system (2.12).

Theorem 7.8. *(Global in time solvability of the FBODE system* (2.12)*). Under Assumption* (**a1**)–(**a3**) *and Hypothesis* (**h1**)-(**h3**)*, for any* $T>0$*,*
(a) *there exists a unique global decoupling field* $\gamma(s,x,\mu)$ *for the FBODE system* (2.12) *on* $[0,T]$ *with terminal*

data $p(x,\mu)=\partial_x k(x,\mu)$ in the following class:
(i) it is differentiable in $s\in[0,T]$, $x\in\mathbb{R}^{d_x}$ and L-differentiable in $\mu\in\mathcal{P}_2(\mathbb{R}^{d_x})$, and its derivatives $\partial_s\gamma(s,x,\mu)$, $\partial_x\gamma(s,x,\mu)$ and $\partial_\mu\gamma(s,x,\mu)(\widetilde{x})$ are jointly continuous in their corresponding arguments $(s,x,\mu)\in[t,T]\times\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})$ and $(s,x,\mu,\widetilde{x})\in[t,T]\times\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$, respectively;
(ii) it also satisfies $\|\gamma\|_{2,[0,T]}\le L_0^*$ with L_0^* defined in (7.56).
***(b)** Moreover, for any $t\in[0,T]$ and $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$, there exists a solution pair $\big(X_s^{t,m}(x),Z_s^{t,m}(x)\big)_{x\in\mathbb{R}^{d_x},s\in[t,T]}$ of the FBODE system (2.12), which is defined by solving for $X_s^{t,m}(x)$ in (6.9) and*

$$Z_s^{t,m}(x):=\gamma(s,X_s^{t,m}(x),X_s^{t,m}\#m),$$

such that,
(i) both $X_s^{t,m}(x)$ and $Z_s^{t,m}(x)$ are continuously differentiable in $s\in[t,T]$;
(ii) for each $s\in[t,T]$ and $x\in\mathbb{R}^{d_x}$,

$$\big|Z_s^{t,m}(x)-Z_s^{t,m_0}(x_0)\big|\le\ L_0^*\Big(\big|X_s^{t,m}(x)-X_s^{t,m_0}(x_0)\big|+W_1(X_s^{t,m}\#m,X_s^{t,m_0}\#m_0)\Big),\tag{7.59}$$
$$\big|Z_s^{t,m}(x)\big|\le L_0^*\left(\big|X_s^{t,m}(x)\big|+\big\|X_s^{t,m}\big\|_{L_m^{1;d_x}}\right);\tag{7.60}$$

(iii) both $X_s^{t,m}(x)$ and $Z_s^{t,m}(x)$ are differentiable in $t\in[0,T]$, $x\in\mathbb{R}^{d_x}$ and are L-differentiable in $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$ with their corresponding derivatives $\Big(\partial_t\big(X_s^{t,m}(x)\big),\partial_t\big(Z_s^{t,m}(x)\big)\Big)$, $\Big(\partial_x\big(X_s^{t,m}(x)\big),\partial_x\big(Z_s^{t,m}(x)\big)\Big)$ and $\Big(\partial_m\big(X_s^{t,m}(x)\big)(y),\partial_m\big(Z_s^{t,m}(x)\big)(y)\Big)$ being continuous in their corresponding arguments, and these derivatives are also continuously differentiable in $s\in[t,T]$;
(iv) in addition, the following estimates hold:

$$\big\|\partial_x\big(X_s^{t,m}(x)\big)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le\exp\Big(L_B'(s-t)\Big);\tag{7.61}$$
$$\big\|\partial_m\big(X_s^{t,m}(x)\big)(y)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le L_M^{(s-t)};\tag{7.62}$$
$$\big|\partial_t\big(X_s^{t,m}(x)\big)\big|\le L_B'\big(1+|x|+\|m\|_1\big)\Big(L_B'(s-t)\exp\Big(2L_B'(s-t)\Big)+1\Big)\exp\Big(L_B'(s-t)\Big);\tag{7.63}$$
$$\big\|\partial_x\big(Z_s^{t,m}(x)\big)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le L_0^*\exp\Big(L_B'(s-t)\Big);\tag{7.64}$$
$$\big\|\partial_m\big(Z_s^{t,m}(x)\big)(y)\big\|_{\mathcal{L}(\mathbb{R}^{d_x};\mathbb{R}^{d_x})}\le 2L_0^*L_M^{(s-t)}+L_0^*\exp\Big(L_B'(s-t)\Big);\tag{7.65}$$
$$\big|\partial_t\big(Z_s^{t,m}(x)\big)\big|\le 2L_0^*L_B'\big(1+|x|+\|m\|_1\big)\Big(L_B'(s-t)\exp\Big(2L_B'(s-t)\Big)+1\Big)\exp\Big(L_B'(s-t)\Big),\tag{7.66}$$

where

$$L_B':=\Lambda_f(1+L_\alpha+L_0^*L_\alpha),\tag{7.67}$$
$$L_M^{(\tau)}:=L_B'\tau\Big(L_B'\tau\cdot\exp\Big(2L_B'\tau\Big)+1\Big)\exp\Big(2L_B'\tau\Big).\tag{7.68}$$

Proof. Define $\bar\epsilon_2=:\epsilon_2(L_0^*;L_f,\Lambda_f,\bar l_f,L_g,\Lambda_g,\bar l_g,L_\alpha)$ where ϵ_2 is given in Theorem 6.4. Recall that L_0^* was defined in (7.56). Partition the interval $[0,T]$ into sub-intervals with endpoints $0<t_1<...<t_{N-1}:=T-\bar\epsilon_2<t_N:=T-\frac{1}{2}\bar\epsilon_2<t_{N+1}=:T$ where $t_{N-i}:=T-\frac{i+1}{2}\bar\epsilon_2$, $i=0,1,...,N-1$ and N is the smallest integer such that $N\ge 2T/\bar\epsilon_2-1$. Note that the sub-interval $[0,t_1=T-\frac{N}{2}\bar\epsilon_2]$ has a length not longer than those other intervals $[t_1,t_2]$,..., and $[t_{N-1},t_N]$, everyone of which has a common length of size $\frac{1}{2}\bar\epsilon_2$. Now, we are ready to paste the local solution γ of (6.8) together to obtain a global one by using the idea depicted in Figure 1.

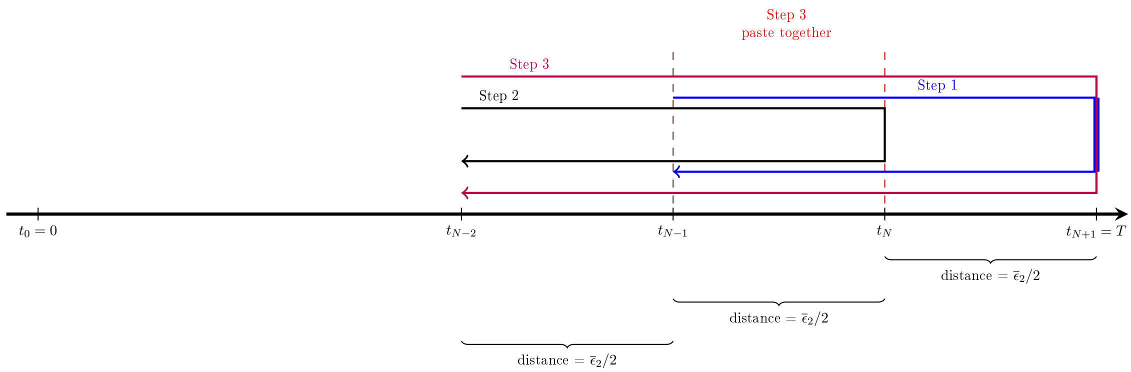

FIGURE 1. Algorithm for pasting the local-in-time solutions.

Step 1. Over the interval $[t_{N-1} = T - \bar{\epsilon}_2, T]$, take $p(x,\mu) = \partial_x k(x,\mu)$, and apply Theorems 6.3 and 6.4, then, the equation (6.8) has a unique solution $\gamma^{(N)}(s,x,\mu)$ over the time interval $[t_{N-1}, T]$ with the regularity (a) stated in Theorem 6.4 and satisfying $\left|\!\left|\!\left|\gamma^{(N)}\right|\!\right|\!\right|_{2,[t_{N-1},T]} \leq 2\max\{\Lambda_k, L_0^*\} = 2L_0^*$. Moreover, applying Theorems 7.5, we actually have a better refined estimate $\left|\!\left|\!\left|\gamma^{(N)}\right|\!\right|\!\right|_{2,[t_{N-1},T]} \leq L_0^*$.

Step 2. Over the interval $[t_{N-2} = t_N - \bar{\epsilon}_2, t_N]$, take $p(x,\mu) = \gamma^{(N)}(t_N,x,\mu)$, and by referring to (6.2) and (6.7), we have $L_p \leq L_0^*$, and then applying Theorems 6.3 and 6.4, the equation (6.8) has a unique solution $\gamma^{(N-1)}(s,x,\mu)$ over the time interval $[t_{N-2}, t_N]$ with the regularity (a) stated in Theorem 6.4 and satisfying $\left|\!\left|\!\left|\gamma^{(N-1)}\right|\!\right|\!\right|_{2,[t_{N-2},t_N]} \leq 2L_0^*$.

Step 3. By the local uniqueness of γ that was stated in Theorem 6.3, for $s \in [t_{N-1}, t_N]$,

$$\gamma^{(N-1)}(s,x,\mu) = \gamma^{(N)}(s,x,\mu),$$

since both of them satisfy the following equation: for any $(s,x,m) \in [t_{N-1}, t_N] \times \mathbb{R}^{d_x} \times \mathcal{P}_2(\mathbb{R}^{d_x})$,

$$\begin{aligned}\gamma(s,x,m) = \gamma^{(N)}(t_N, X_{t_N}^{s,m}(x), X_{t_N}^{s,m}\#m) &+ \int_s^{t_N} \Big(\partial_x f(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, \widehat{\alpha}(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, Z_\tau^{s,m}(x))) \cdot Z_\tau^{s,m}(x) \\ &+ \partial_x g(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, \widehat{\alpha}(X_\tau^{s,m}(x), X_\tau^{s,m}\#m, Z_\tau^{s,m}(x))) \Big) d\tau, \end{aligned} \tag{7.69}$$

where $Z_\tau^{s,m}(x) := \gamma\big(\tau, X_\tau^{s,m}(x), X_\tau^{s,m}\#m\big)$ and

$$X_\tau^{s,m}(x) = x + \int_s^\tau f\Big(X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m, \widehat{\alpha}\big(X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m, \gamma(\tau, X_{\widetilde{\tau}}^{s,m}(x), X_{\widetilde{\tau}}^{s,m}\#m)\big)\Big) d\widetilde{\tau}. \tag{7.70}$$

Therefore, for any $t \in [t_{N-1}, t_N]$, we paste together the local solutions $\gamma^{(N)}(s,x,\mu)$ and $\gamma^{(N-1)}(s,x,\mu)$ consecutively in order by extending $\gamma^{(N-1)}(s,x,\mu)$ as follows: for any $s \in [t_N, T]$, $\gamma^{(N-1)}(s,x,\mu) = \gamma^{(N)}(s,x,\mu)$. We still denote the unique extended solution over the time horizon up to the terminal time $[t_{N-2}, T]$ by $\gamma^{(N-1)}(s,x,\mu)$ without causing much ambiguity, which thus solves the equation (6.8) over the time horizon up to the terminal time $[t_{N-2}, T]$. In addition, $\gamma^{(N-1)}(s,x,\mu)$ on the time interval $[t_{N-2}, t_N]$ still has the regularity (a) stated in Theorem 6.4 and satisfies $\left|\!\left|\!\left|\gamma^{(N-1)}\right|\!\right|\!\right|_{2,[t_{N-2},T]} \leq 2L_0^*$. Moreover, applying Theorem 7.5 again, we also have the refined estimate $\left|\!\left|\!\left|\gamma^{(N-1)}\right|\!\right|\!\right|_{2,[t_{N-2},T]} \leq L_0^*$.

Step 4. Over the interval $[t_{N-3}, t_{N-1}]$, we repeat Step 2-Step 3. More precisely, due to the uniform (in time) estimate (7.18) in Theorem 7.5, the minimal lifespan $\overline{\epsilon}_2$ is uniform in all sub-intervals, so we can carry out the same pasting procedure for all the rest of the other intervals $[t_{N-4}, t_{N-2}]$,...,$[0, t_2]$ in a backward manner. As a result, we can obtain a unique global solution $\gamma(s, x, \mu)$ of the equation (6.8) on the whole time horizon $[0, T]$, so that it has the regularity (a) stated in Theorem 6.4, that is the regularity of γ stated in this theorem, and satisfies $|||\gamma|||_{2,[0,T]} \leq L_0^*$ as well.

Step 5. Finally, for any $t \in [0, T]$ and $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$, define the pair $\big(X_s^{t,m}(x), Z_s^{t,m}(x)\big)$ by (6.9) and $Z_s^{t,m}(x) := \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m)$ for all $s \in [t, T]$ and $x \in \mathbb{R}^{d_x}$, then it is a solution pair of the FBODE system (2.12), and (7.59) and (7.60) are valid. The regularity of $\big(X_s^{t,m}(x), Z_s^{t,m}(x)\big)$ stated in this theorem statement directly follows from that of γ stated in Theorem 7.8. In addition, by the same argument as that in the proof of (6.15)–(6.20), we have (7.61)–(7.63) and then (7.64)–(7.66) by using $|||\gamma|||_{2,[0,T]} \leq L_0^*$. □

Theorem 7.9. *(Global in time solvability of the master equation (2.11)). Under Assumption* **(a1)**–**(a3)** *and Hypothesis* **(h1)**-**(h3)***, then there exists a global-in-time classical solution $V(t, x, m)$, given by (5.9), of the master equation (2.11), and it is differentiable in $t \in [0, T]$, $x \in \mathbb{R}^{d_x}$, L-differentiable in $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$ with its corresponding derivatives $\partial_t V(t, x, m)$, $\partial_x V(t, x, m)$ and $\partial_m V(t, x, m)(y)$ being continuous in their respective arguments. In addition, $\partial_x V(t, x, m)$ is differentiable in $t \in [0, T]$, $x \in \mathbb{R}^{d_x}$ and L-differentiable in $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$ with its corresponding derivatives being continuous in their respective arguments, and $\partial_t V(t, x, m)$ and $\partial_m V(t, x, m)(y)$ are continuously differentiable in $x \in \mathbb{R}^{d_x}$.*

Proof. Under Assumption **(a1)**–**(a3)** and Hypothesis **(h1)**–**(h3)**, by Theorem 7.8, we then have a global solution pair $\big(X_s^{t,m}(x), Z_s^{t,m}(x)\big)_{x \in \mathbb{R}^{d_x}, s \in [t,T]}$ of the FBODE system (2.12) for any $t \in [0, T]$ and $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$. Define $V(t, x, m)$ by (5.9), then, the regularity of $V(t, x, m)$ directly follows from the regularity of $\big(X_s^{t,m}(x), Z_s^{t,m}(x)\big)$ and the coefficient functions f, g and k, and, by the flow property (6.12) of $X_s^{t,m}(x)$, we have

$$V(s, X_s^{t,m}(x), X_s^{t,m}\#m) = k\big(X_T^{t,m}(x), X_T^{t,m}\#m\big) + \int_s^T g\Big(X_\tau^{t,m}(x), X_\tau^{t,m}\#m, \widehat{\alpha}\big(X_\tau^{t,m}(x), X_\tau^{t,m}\#m, Z_\tau^{t,m}(x)\big)\Big) d\tau, \tag{7.71}$$

which implies, by taking derivative with respect to s on both sides,

$$\begin{aligned}
&\partial_s V(s, X_s^{t,m}(x), X_s^{t,m}\#m) + \partial_x V(s, X_s^{t,m}(x), X_s^{t,m}\#m) \cdot f\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big) \\
&\quad + \int_{\mathbb{R}^{d_x}} \partial_\mu V(s, X_s^{t,m}(x), X_s^{t,m}\#m)\big(X_s^{t,m}(\widetilde{x})\big) \cdot f\Big(X_s^{t,m}(\widetilde{x}), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(\widetilde{x}), X_s^{t,m}\#m, Z_s^{t,m}(\widetilde{x})\big)\Big) dm(\widetilde{x}) \\
&= -g\Big(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}\big(X_s^{t,m}(x), X_s^{t,m}\#m, Z_s^{t,m}(x)\big)\Big).
\end{aligned} \tag{7.72}$$

In addition, by taking the derivative of equation (5.9) with respect to x and using the notations in Section 7.2, we have

$$\begin{aligned}
&\partial_x V(t, x, m) = \partial_x \left(k_T^{t,m}(x) + \int_t^T g_s^{t,m}(x) ds \right) \\
&= (\partial_x k)_T^{t,m}(x) \cdot \partial_x\big(X_T^{t,m}(x)\big) \\
&\quad + \int_t^T \left(\partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big) \cdot (\partial_a g)_s^{t,m}(x) - \left(\frac{d}{ds} Z_s^{t,m}(x) + (\partial_x f)_s^{t,m}(x) \cdot Z_s^{t,m}(x) \right) \cdot \partial_x\big(X_s^{t,m}(x)\big) \right) ds,
\end{aligned} \tag{7.73}$$

where we use the second last equation of (2.12) to obtain the last term of the equality. By the forward equation (2.12) and a simple application of integration-by-parts, the following integral in (7.73) can be rewritten as

$$\begin{aligned}
&\int_t^T \frac{d}{ds} Z_s^{t,m}(x) \cdot \partial_x\big(X_s^{t,m}(x)\big) ds \\
&= Z_T^{t,m}(x) \cdot \partial_x\big(X_T^{t,m}(x)\big) - Z_t^{t,m}(x) \cdot \partial_x\big(X_t^{t,m}(x)\big) - \int_t^T Z_s^{t,m}(x) \cdot \frac{d}{ds} \partial_x\big(X_s^{t,m}(x)\big) ds \\
&= \big(\partial_x k\big)_T^{t,m}(x) \cdot \partial_x\big(X_T^{t,m}(x)\big) - Z_t^{t,m}(x) - \int_t^T Z_s^{t,m}(x) \cdot \partial_x\big(f_s^{t,m}(x)\big) ds
\end{aligned}$$

$$
\begin{aligned}
&= \big(\partial_x k\big)_T^{t,m}(x)\cdot\partial_x\big(X_T^{t,m}(x)\big) - Z_t^{t,m}(x)\\
&\quad -\int_t^T Z_s^{t,m}(x)\cdot\Big(\big(\partial_x f\big)_s^{t,m}(x)\partial_x\big(X_s^{t,m}(x)\big)+\big(\partial_a f\big)_s^{t,m}(x)\partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big)\Big)ds.
\end{aligned} \tag{7.74}
$$

Substituting (7.74) into (7.73) so as to cancel some similar terms, one has, by using the first-order condition (4.30) in the last equality,

$$
\partial_x V(t,x,m) = \int_t^T \partial_x\big(\widehat{\alpha}_s^{t,m}(x)\big)\cdot\Big(\big(\partial_a g\big)_s^{t,m}(x)+\big(\partial_a f\big)_s^{t,m}(x)\cdot Z_s^{t,m}(x)\Big)ds + Z_t^{t,m}(x) = Z_t^{t,m}(x). \tag{7.75}
$$

Then, the regularity of $\partial_x V(t,x,m)$ directly follows from the regularity of $Z_t^{t,m}(x)=\gamma(t,x,m)$. Set $s=t$ in (7.72) and then the resulting equation is exactly (2.11) since $X_t^{t,m}(x)=x$ and $X_t^{t,m}\#m=m$. In addition, the differentiability of $\partial_x V(t,x,m)=\gamma(t,x,m)$ follows from that of γ in Theorem 7.8 immediately.

Next, similar to the computations in deriving (7.73)-(7.75), we compute $\partial_t V$ as follows. Taking the derivative of the equation (5.9) with respect to t and using the notations in Section 7.2, we can obtain

$$
\begin{aligned}
&\partial_t V(t,x,m) = \partial_t\bigg(k_T^{t,m}(x)+\int_t^T g_s^{t,m}(x)ds\bigg)\\
&= \int_{\mathbb{R}^{d_x}}\big(\partial_\mu k\big)_T^{t,m}(x,\widetilde{x})\cdot\partial_t\big(X_T^{t,m}(\widetilde{x})\big)dm(\widetilde{x}) + \big(\partial_x k\big)_T^{t,m}(x)\cdot\partial_t\big(X_T^{t,m}(x)\big) - g_t^{t,m}(x)\\
&\quad+\int_t^T\bigg(\int_{\mathbb{R}^{d_x}}\big(\partial_\mu g\big)_s^{t,m}(x,\widetilde{x})\cdot\partial_t\big(X_s^{t,m}(\widetilde{x})\big)dm(\widetilde{x}) + \partial_t\big(\widehat{\alpha}_s^{t,m}(x)\big)\cdot\big(\partial_a g\big)_s^{t,m}(x)\\
&\qquad\qquad -\bigg(\frac{d}{ds}Z_s^{t,m}(x)+\big(\partial_x f\big)_s^{t,m}(x)\cdot Z_s^{t,m}(x)\bigg)\cdot\partial_t\big(X_s^{t,m}(x)\big)\bigg)ds,
\end{aligned} \tag{7.76}
$$

where we use the second last equation of (2.12) to obtain the last term of the equality. By the forward equation (2.12) and a simple application of integration-by-parts, the following integral in (7.76) can be rewritten as

$$
\begin{aligned}
&\int_t^T \frac{d}{ds}Z_s^{t,m}(x)\cdot\partial_t\big(X_s^{t,m}(x)\big)ds\\
&= Z_T^{t,m}(x)\cdot\partial_t\big(X_T^{t,m}(x)\big) - Z_t^{t,m}(x)\cdot\partial_t\big(X_s^{t,m}(x)\big)\big|_{s=t} - \int_t^T Z_s^{t,m}(x)\cdot\frac{d}{ds}\partial_t\big(X_s^{t,m}(x)\big)ds\\
&= \big(\partial_x k\big)_T^{t,m}(x)\cdot\partial_t\big(X_T^{t,m}(x)\big) + Z_t^{t,m}(x)\cdot f_t^{t,m}(x) - \int_t^T Z_s^{t,m}(x)\cdot\partial_t\big(f_s^{t,m}(x)\big)ds\\
&= \big(\partial_x k\big)_T^{t,m}(x)\cdot\partial_t\big(X_T^{t,m}(x)\big) + Z_t^{t,m}(x)\cdot f_t^{t,m}(x)\\
&\quad -\int_t^T Z_s^{t,m}(x)\cdot\bigg(\big(\partial_x f\big)_s^{t,m}(x)\cdot\partial_t\big(X_s^{t,m}(x)\big)+\partial_t\big(\widehat{\alpha}_s^{t,m}(x)\big)\cdot\big(\partial_a f\big)_s^{t,m}(x)\\
&\qquad\qquad +\int_{\mathbb{R}^{d_x}}\big(\partial_\mu f\big)_s^{t,m}(x,\widetilde{x})\cdot\partial_t\big(X_s^{t,m}(\widetilde{x})\big)dm(\widetilde{x})\bigg)ds,
\end{aligned} \tag{7.77}
$$

and thus, by substituting (7.77) into (7.76) so as to cancel some similar terms, one has, by using the first-order condition (4.30) in the last equality,

$$\begin{aligned}\partial_t V(t,x,m) = &\int_t^T \int_{\mathbb{R}^{d_x}} \Big(\big(\partial_\mu g\big)_s^{t,m}(x,\widetilde{x}) + \big(\partial_\mu f\big)_s^{t,m}(x,\widetilde{x}) \cdot Z_s^{t,m}(x) \Big) \cdot \partial_t \big(X_s^{t,m}(\widetilde{x})\big) dm(\widetilde{x}) ds \\ &+ \int_{\mathbb{R}^{d_x}} \big(\partial_\mu k\big)_T^{t,m}(x,\widetilde{x}) \cdot \partial_t \big(X_T^{t,m}(\widetilde{x})\big) dm(\widetilde{x}) - \Big(Z_t^{t,m}(x) \cdot f_t^{t,m}(x) + g_t^{t,m}(x) \Big), \end{aligned} \tag{7.78}$$

Finally, we compute $\partial_m V$ as follows. By taking the derivative of equation (5.9) with respect to m and using the notations in Section 7.2, we can obtain

$$\begin{aligned} &\partial_m V(t,x,m)(y) = \partial_m \Big(k_T^{t,m}(x) + \int_t^T g_s^{t,m}(x) ds \Big)(y) \\ &= \int_{\mathbb{R}^{d_x}} \big(\partial_\mu k\big)_T^{t,m}(x,\widetilde{x}) \cdot \partial_m \big(X_T^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) + \big(\partial_\mu k\big)_T^{t,m}(x,y) \cdot \partial_y \big(X_T^{t,m}(y)\big) + \big(\partial_x k\big)_T^{t,m}(x) \cdot \partial_m \big(X_T^{t,m}(x)\big)(y) \\ &\quad + \int_t^T \Big(\int_{\mathbb{R}^{d_x}} \big(\partial_\mu g\big)_s^{t,m}(x,\widetilde{x}) \cdot \partial_m \big(X_s^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) + \big(\partial_\mu g\big)_s^{t,m}(x,y) \cdot \partial_y \big(X_s^{t,m}(y)\big) \\ &\qquad + \partial_m \big(\widehat{\alpha}_s^{t,m}(x)\big)(y) \cdot \big(\partial_a g\big)_s^{t,m}(x) - \Big(\frac{d}{ds} Z_s^{t,m}(x) + \big(\partial_x f\big)_s^{t,m}(x) \cdot Z_s^{t,m}(x) \Big) \cdot \partial_m \big(X_s^{t,m}(x)\big)(y) \Big) ds, \end{aligned} \tag{7.79}$$

where we use the second last equation of (2.12) to obtain the last term of the equality. By the forward equation (2.12) and a simple application of integration-by-parts, the following integral in (7.79) can be rewritten as

$$\begin{aligned} &\int_t^T \frac{d}{ds} Z_s^{t,m}(x) \cdot \partial_m \big(X_s^{t,m}(x)\big)(y) ds \\ &= Z_T^{t,m}(x) \cdot \partial_m \big(X_T^{t,m}(x)\big)(y) - Z_t^{t,m}(x) \cdot \partial_m \big(X_t^{t,m}(x)\big)(y) - \int_t^T Z_s^{t,m}(x) \cdot \frac{d}{ds} \partial_m \big(X_s^{t,m}(x)\big)(y) ds \\ &= \big(\partial_x k\big)_T^{t,m}(x) \cdot \partial_m \big(X_T^{t,m}(x)\big)(y) - \int_t^T Z_s^{t,m}(x) \cdot \partial_m \big(f_s^{t,m}(x)\big)(y) ds \\ &= \big(\partial_x k\big)_T^{t,m}(x) \cdot \partial_m \big(X_T^{t,m}(x)\big)(y) \\ &\quad - \int_t^T Z_s^{t,m}(x) \cdot \Big(\big(\partial_x f\big)_s^{t,m}(x) \cdot \partial_m \big(X_s^{t,m}(x)\big)(y) + \int_{\mathbb{R}^{d_x}} \big(\partial_\mu f\big)_s^{t,m}(x,\widetilde{x}) \cdot \partial_m \big(X_s^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) \\ &\qquad + \big(\partial_\mu f\big)_s^{t,m}(x,y) \cdot \partial_y \big(X_s^{t,m}(y)\big) + \partial_m \big(\widehat{\alpha}_s^{t,m}(x)\big)(y) \cdot \big(\partial_a f\big)_s^{t,m}(x) \Big) ds, \end{aligned} \tag{7.80}$$

and thus, by substituting (7.80) into (7.79) so as to cancel some similar terms, one has, by using the first-order condition (4.30) in the last equality,

$$\begin{aligned} &\partial_m V(t,x,m)(y) \\ &= \int_{\mathbb{R}^{d_x}} \big(\partial_\mu k\big)_T^{t,m}(x,\widetilde{x}) \cdot \partial_m \big(X_T^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) + \big(\partial_\mu k\big)_T^{t,m}(x,y) \cdot \partial_y \big(X_T^{t,m}(y)\big) \\ &\quad + \int_t^T \Big(\int_{\mathbb{R}^{d_x}} \Big(\big(\partial_\mu f\big)_s^{t,m}(x,\widetilde{x}) \cdot Z_s^{t,m}(x) + \big(\partial_\mu g\big)_s^{t,m}(x,\widetilde{x}) \Big) \cdot \partial_m \big(X_s^{t,m}(\widetilde{x})\big)(y) dm(\widetilde{x}) \\ &\qquad + \Big(\big(\partial_\mu f\big)_s^{t,m}(x,y) \cdot Z_s^{t,m}(x) + \big(\partial_\mu g\big)_s^{t,m}(x,y) \Big) \cdot \partial_y \big(X_s^{t,m}(y)\big) \Big) ds. \end{aligned} \tag{7.81}$$

Therefore, the continuous differentiability of $\partial_t V(t,x,m)$ and $\partial_m V(t,x,m)(y)$ with respect to $x\in\mathbb{R}^{d_x}$ directly follows from the regularity of $Z_s^{t,m}(x)$ and the coefficient functions f, g and k. $\square$

8. Local and global uniqueness

Theorem 8.1. *(Local uniqueness of the solution to the FBODE system (6.1)) Assume that the drift function f and the running cost g satisfy Assumptions* (**a1**) *and* (**a2**) *respectively, the terminal function p satisfies Assumption* (**P**) *with $L_p \le \overline{L}_p := \max\{\Lambda_k, L_0^*\}$ and the inequality relation (6.10) among f, g and p is valid, then there exists a constant $\epsilon_3=\epsilon_3(\overline{L}_p; L_f,\Lambda_f,\bar{l}_f,L_g,\Lambda_g,\bar{l}_g,L_\alpha)>0$, such that, for any fixed $0\le t\le \widetilde{T}\le T$ with $\widetilde{T}-t\le\epsilon_3$ and $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$, if there exist two solution pairs $\big(X_s^{t,m}(x),Z_s^{t,m}(x)\big)_{x\in\mathbb{R}^{d_x},s\in[t,\widetilde{T}]}$ and $\big(\widetilde{X}_s^{t,m}(x),\widetilde{Z}_s^{t,m}(x)\big)_{x\in\mathbb{R}^{d_x},s\in[t,\widetilde{T}]}$ of FBODE system (6.1), each of them is continuously differentiable in $s\in[t,\widetilde{T}]$, then they must be equal for all $s\in[t,\widetilde{T}]$ and $x\in\mathbb{R}^{d_x}$.*

The proof of Theorem 8.1 is similar to, but simpler[6] than, the argument leading to Theorem 9.1 of our previous paper [52]. Therefore, we omit the details of the proof here.

Theorem 8.2. *(Global uniqueness of the solution to the FBODE system (2.12)) Under Assumptions* (**a1**)–(**a3**) *and Hypotheses* (**h1**)-(**h3**)*, for any fixed terminal time $T>0$ and initial distribution $m\in\mathcal{P}_2(\mathbb{R}^{d_x})$, if there exists a continuously differentiable, in $s\in[0,T]$, solution pair $\big(\widetilde{X}_s^{0,m}(x),\widetilde{Z}_s^{0,m}(x)\big)_{x\in\mathbb{R}^{d_x},s\in[0,T]}$ of FBODE system (2.12) subject to the initial distribution m at the initial time $t=0$, then it must be equal to the solution pair $\big(X_s^{0,m}(x),Z_s^{0,m}(x):=\gamma(s,X_s^{0,m}(x),X_s^{0,m}\#m)\big)_{x\in\mathbb{R}^{d_x},s\in[0,T]}$, where the forward dynamics $X_s^{0,m}(x)$ was obtained by solving (6.9) over $[0,T]$ and the decoupling field γ was constructed in Theorem 7.8.*

Proof. Under Assumptions (**a1**)–(**a3**) and Hypotheses (**h1**)–(**h3**), according to Theorem 7.8, we have a global decoupling field $\gamma(s,x,\mu)$ for the FBODE system (2.12) on $[0,T]$, which satisfies the equation (6.8). Let $\overline{L}_p:=\max\{\Lambda_k,L_0^*\}=L_0^*$ and $\epsilon_3=\epsilon_3(\overline{L}_p;L_f,\Lambda_f,\bar{l}_f,L_g,\Lambda_g,\bar{l}_g,L_\alpha)>0$ given in Theorem 8.1. Partition the interval $[0,T]$ into $0<t_1<...<t_{N-1}:=T-\epsilon_3<t_N=:T$, where $t_{N-i}=T-i\cdot\epsilon_3$, $i=0,1,...,N-1$ and N is the smallest integer such that $N\ge T/\epsilon_3$. Note that the interval $[0,t_1=T-(N-1)\epsilon_3]$ has a length not longer than those other intervals $[t_1,t_2]$,..., and $[t_{N-1},t_N]$, while every others has a common length of size ϵ_3.

Step 1. On the interval $[t_{N-1},t_N=T]$, define $y:=\widetilde{X}_{t_{N-1}}^{0,m}(x)$, $\widetilde{m}:=\widetilde{X}_{t_{N-1}}^{0,m}\#m$ and $p(x,\mu):=\partial_x k(x,\mu)$, then $\big(X_s^{t_{N-1},\widetilde{m}}(y),Z_s^{t_{N-1},\widetilde{m}}(y)\big):=\big(\widetilde{X}_s^{0,m}(x),\widetilde{Z}_s^{0,m}(x)\big)$ [7] is a solution pair of FBODE system (6.1) with the terminal data $Z_T^{t_{N-1},\widetilde{m}}(y)=p(X_T^{t_{N-1},\widetilde{m}}(y),X_T^{t_{N-1},\widetilde{m}}\#\widetilde{m})$ and the initial data $X_{t_{N-1}}^{t_{N-1},\widetilde{m}}(y)=y$. Note that, for any $0\le s\le\tau\le T$, $\widehat{x}\in\mathbb{R}^{d_x}$ and $\widehat{m}\in\mathcal{P}_2(\mathbb{R}^{d_x})$, the decoupling field γ solves (6.8) with (6.9):

$$
\begin{aligned}
\gamma(s,\widehat{x},\widehat{m}) = p(\widehat{X}_T^{s,\widehat{m}}(\widehat{x}),\widehat{X}_T^{s,\widehat{m}}\#\widehat{m}) \\
&+\int_s^T\Bigg(\partial_x f(\widehat{X}_\tau^{s,\widehat{m}}(\widehat{x}),\widehat{X}_\tau^{s,\widehat{m}}\#\widehat{m},\widehat{\alpha}(\widehat{X}_\tau^{s,\widehat{m}}(\widehat{x}),\widehat{X}_\tau^{s,\widehat{m}}\#\widehat{m},\widehat{Z}_\tau^{s,\widehat{m}}(\widehat{x})))\cdot\widehat{Z}_\tau^{s,\widehat{m}}(\widehat{x}) \\
&\qquad+\partial_x g(\widehat{X}_\tau^{s,\widehat{m}}(\widehat{x}),\widehat{X}_\tau^{s,\widehat{m}}\#\widehat{m},\widehat{\alpha}(\widehat{X}_\tau^{s,\widehat{m}}(\widehat{x}),\widehat{X}_\tau^{s,\widehat{m}}\#\widehat{m},\widehat{Z}_\tau^{s,\widehat{m}}(\widehat{x})))\Bigg)d\tau,
\end{aligned}
$$

[6]The simplification is due to the absence of the terms $\partial_\mu f(x,\mu,a)$ and $\partial_\mu g(x,\mu,a)$ on the right-hand side of the FBODE system (6.1), in comparison with (7.1) of [52].

[7]Since the initial measure at time t_N is $\widetilde{m}=\widetilde{X}_{t_{N-1}}^{0,m}\#m$ where $\widetilde{X}_{t_{N-1}}^{0,m}(\cdot)$ is a push-forward mapping, so the complement set S^c of the set $S:=\{y:y=\widetilde{X}_{t_{N-1}}^{0,m}(x),x\in\mathbb{R}^{d_x}\}=\widetilde{X}_{t_{N-1}}^{0,m}(\mathbb{R}^{d_x})$ is $\widetilde{m}$-measure zero, that is $\widetilde{m}(S^c)\le m(\emptyset)=0$, and thus we can ignore those elements $y\in S^c$ while solving the FBODE system (6.1) over $[t_{N-1},T]$.

where $\widehat{Z}_{\tau}^{s,\widehat{m}}(\widehat{x}) := \gamma\big(\tau, \widehat{X}_{\tau}^{s,\widehat{m}}(\widehat{x}), \widehat{X}_{\tau}^{s,\widehat{m}}\#\widehat{m}\big)$ and

$$\widehat{X}_{\tau}^{s,\widehat{m}}(\widehat{x}) = \widehat{x} + \int_s^{\tau} f\Big(\widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}(\widehat{x}), \widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}\#\widehat{m}, \widehat{\alpha}\big(\widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}(\widehat{x}), \widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}\#\widehat{m}, \gamma(\tau, \widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}(\widehat{x}), \widehat{X}_{\widetilde{\tau}}^{s,\widehat{m}}\#\widehat{m})\big)\Big)d\widetilde{\tau}.$$

Then, one can directly check that $\big(\widehat{X}_s^{t_{N-1},\widetilde{m}}(y), \widehat{Z}_s^{t_{N-1},\widetilde{m}}(y)\big)$ is also a solution pair of the FBODE system (6.1) on the interval $[t_{N-1}, t_N = T]$ with the same initial and terminal data as $\big(X_s^{t_{N-1},\widetilde{m}}(y), Z_s^{t_{N-1},\widetilde{m}}(y)\big)$ does. By Theorem 8.1, $\big(\widehat{X}_s^{t_{N-1},\widetilde{m}}(y), \widehat{Z}_s^{t_{N-1},\widetilde{m}}(y)\big) = \big(X_s^{t_{N-1},\widetilde{m}}(y), Z_s^{t_{N-1},\widetilde{m}}(y)\big) = \big(\widetilde{X}_s^{0,m}(x), \widetilde{Z}_s^{0,m}(x)\big)$ and thus $\widetilde{Z}_s^{0,m}(x) = \gamma(s, \widetilde{X}_s^{0,m}(x), \widetilde{X}_s^{0,m}\#m)$ on the interval $s \in [t_{N-1}, t_N = T]$.

Step 2. On the interval $[t_{N-2}, t_{N-1}]$, define $y := \widetilde{X}_{t_{N-2}}^{0,m}(x)$, $\widetilde{m} := \widetilde{X}_{t_{N-2}}^{0,m}\#m$ and $p(x,\mu) := \gamma(t_{N-1}, x, \mu)$, then $\big(X_s^{t_{N-2},\widetilde{m}}(y), Z_s^{t_{N-2},\widetilde{m}}(y)\big) := \big(\widetilde{X}_s^{0,m}(x), \widetilde{Z}_s^{0,m}(x)\big)$ is a solution pair of FBODE system (6.1) with terminal data $Z_{t_{N-1}}^{t_{N-2},\widetilde{m}}(y) = p(X_{t_{N-1}}^{t_{N-2},\widetilde{m}}(y), X_{t_{N-1}}^{t_{N-2},\widetilde{m}}\#\widetilde{m})$ and initial data $X_{t_{N-2}}^{t_{N-2},\widetilde{m}}(y) = y$. Same as Step 1, one can also directly check that $\big(\widehat{X}_s^{t_{N-2},\widetilde{m}}(y), \widehat{Z}_s^{t_{N-2},\widetilde{m}}(y)\big)$ is also a solution pair of FBODE system (6.1) on the interval $[t_{N-2}, t_{N-1}]$ with the same initial and terminal data as $\big(X_s^{t_{N-2},\widetilde{m}}(y), Z_s^{t_{N-2},\widetilde{m}}(y)\big)$ does. Since $\|\|\gamma\|\|_{2,[0,T]} \le L_0^*$, by Theorem 8.1 again, $\big(\widehat{X}_s^{t_{N-2},\widetilde{m}}(y), \widehat{Z}_s^{t_{N-2},\widetilde{m}}(y)\big) = \big(X_s^{t_{N-2},\widetilde{m}}(y), Z_s^{t_{N-2},\widetilde{m}}(y)\big) = \big(\widetilde{X}_s^{0,m}(x), \widetilde{Z}_s^{0,m}(x)\big)$, and thus $\widetilde{Z}_s^{0,m}(x) = \gamma(s, \widetilde{X}_s^{0,m}(x), \widetilde{X}_s^{0,m}\#m)$ on the interval $s \in [t_{N-2}, t_{N-1}]$. We can repeat the same procedure for all the rest of the other intervals $[t_{N-3}, t_{N-2}]$,...,$[0, t_1]$ in a backward manner since Theorem 7.8 can ensure $\|\|\gamma\|\|_{2,[0,T]} \le L_0^*$ independent of the current subinterval. Therefore, we can obtain $\widetilde{Z}_s^{0,m}(x) = \gamma(s, \widetilde{X}_s^{0,m}(x), \widetilde{X}_s^{0,m}\#m)$ on the whole interval $s \in [0,T]$. □

Remark 8.3. In the proof of the global existence of a continuously differentiable decoupling field γ (refer to (7.41)), we relied on the monotonicity conditions (4.46) and (7.1). Therefore, it is not unexpected that the global existence of such a decoupling field γ could lead to the global uniqueness of the solution to the FBODE system (2.12).

Theorem 8.4. *(Global uniqueness of the solution to the master equation (2.11)) Under Assumptions* (**a1**)–(**a3**) *and Hypotheses* (**h1**)-(**h3**)*, the master equation* (2.11) *has at most one solution $V(t,x,m)$ that has the regularity stated in Theorem 7.9.*

Proof. Suppose that there exist two solutions $V(t,x,m)$ and $\overline{V}(t,x,m)$ to the master equation (2.11), both of which have the regularity stated in Theorem 7.9. Define $\gamma(t,x,m) := \partial_x V(t,x,m)$, $\overline{\gamma}(t,x,m) := \partial_x \overline{V}(t,x,m)$, and

$$\begin{cases} X_s^{t,m}(x) := x + \int_t^s f\Big(X_\tau^{t,m}(x), X_\tau^{t,m}\#m, \widehat{\alpha}\big(X_\tau^{t,m}(x), X_\tau^{t,m}\#m, \gamma(\tau, X_\tau^{t,m}(x), X_\tau^{t,m}\#m)\big)\Big)d\tau, \\ \overline{X}_s^{t,m}(x) := x + \int_t^s f\Big(\overline{X}_\tau^{t,m}(x), \overline{X}_\tau^{t,m}\#m, \widehat{\alpha}\big(\overline{X}_\tau^{t,m}(x), \overline{X}_\tau^{t,m}\#m, \overline{\gamma}(\tau, \overline{X}_\tau^{t,m}(x), \overline{X}_\tau^{t,m}\#m)\big)\Big)d\tau, \end{cases} \tag{8.1}$$

$Z_s^{t,m}(x) := \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m)$ and $\overline{Z}_s^{t,m}(x) := \overline{\gamma}(s, \overline{X}_s^{t,m}(x), \overline{X}_s^{t,m}\#m)$. Then, by taking the derivative ∂_x to the master equation (2.11) and using the first-order condition (4.30) and $\gamma(t,x,m) = \partial_x V(t,x,m)$, we have

$$\begin{aligned} &\partial_t \gamma(t,x,m) + f(x,m,\widehat{\alpha}(x,m,\gamma(t,x,m))) \cdot \partial_x \gamma(t,x,m) + \int_{\mathbb{R}^{d_x}} f(y,m,\widehat{\alpha}(y,m,\gamma(t,y,m))) \cdot \partial_\mu \gamma(t,x,m)(y) dm(y) \\ &= -\partial_x f(x,m,\widehat{\alpha}(x,m,\gamma(t,x,m))) \cdot \gamma(t,x,m) - \partial_x g(x,m,\widehat{\alpha}(x,m,\gamma(t,x,m))), \end{aligned}$$

which implies that, by evaluating at $t = s$, $x = X_s^{t,m}(x)$ and $m = X_s^{t,m}\#m$,

$$\begin{aligned} &\partial_s \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m) + f(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m))) \cdot \partial_x \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m) \\ &+ \int_{\mathbb{R}^{d_x}} f(X_s^{t,m}(y), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(y), X_s^{t,m}\#m, \gamma(s, X_s^{t,m}(y), X_s^{t,m}\#m))) \cdot \partial_\mu \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m)(X_s^{t,m}(y)) dm(y) \\ &= -\partial_x f(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m))) \cdot \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m) \end{aligned}$$

$$- \partial_x g(X_s^{t,m}(x), X_s^{t,m}\#m, \widehat{\alpha}(X_s^{t,m}(x), X_s^{t,m}\#m, \gamma(s, X_s^{t,m}(x), X_s^{t,m}\#m))). \tag{8.2}$$

Applying the characteristics method to the derived equation (8.2), we can observe that the left-hand side of (8.2) is precisely equal to $\frac{d}{ds}Z_s^{t,m}(x)$. The corresponding characteristic equations satisfied by both $\left(X_s^{t,m}(x), Z_s^{t,m}(x)\right)$ and $\left(\overline{X}_s^{t,m}(x), \overline{Z}_s^{t,m}(x)\right)$ are exactly the same as the FBODE system (2.12). This implies that $\left(X_s^{t,m}(x), Z_s^{t,m}(x)\right) = \left(\overline{X}_s^{t,m}(x), \overline{Z}_s^{t,m}(x)\right)$ by Theorem 8.2. Therefore, we obtain $\partial_x V(t,x,m) = \gamma(t,x,m) = Z_t^{t,m}(x) = \overline{Z}_t^{t,m}(x) = \overline{\gamma}(t,x,m) = \partial_x \overline{V}(t,x,m)$ for all $t \in [0,T]$, $x \in \mathbb{R}^{d_x}$ and $m \in \mathcal{P}_2(\mathbb{R}^{d_x})$. Thus, by (8.1), $X_s^{t,m}(x) = \overline{X}_s^{t,m}(x)$. Next, using the master equation (2.11) again, we can conclude that $\partial_s\big(V(s, X_s^{t,m}(x), X_s^{t,m}(x)\#m) - \overline{V}(s, \overline{X}_s^{t,m}(x), \overline{X}_s^{t,m}\#m)\big) = 0$ for all $0 \le t \le s \le T$ and $V(T, X_T^{t,m}(x), X_T^{t,m}\#m) - \overline{V}(T, \overline{X}_T^{t,m}(x), \overline{X}_T^{t,m}\#m) = 0$ since both initial data are the same. Therefore, we have $V(s, X_s^{t,m}(x), X_s^{t,m}(x)\#m) = \overline{V}(s, \overline{X}_s^{t,m}(x), \overline{X}_s^{t,m}\#m)$ for all $0 \le t \le s \le T$, and thus $V(t,x,m) = \overline{V}(t,x,m)$ for all t by setting $s = t$. □

In a related study, Bertucci–Lasry–Lions [70] investigates the existence and uniqueness of Lipschitz solutions for mean field games (MFG) master equations across three settings: finite state spaces, Hilbert spaces, and probability measure spaces. It establishes that under Lipschitz continuity of initial data, a unique solution exists up to a maximal time, beyond which Lipschitz regularity fails, and demonstrates the propagation of higher-order regularity within this timeframe.

9. Non-LQ examples

We here provide two non-trivial non-LQ examples. The first example is used to illustrate that our theory can apply to a non-LQ case that other contemporary works cannot apply. The second example is used to illustrate the boundedness conditions on the second-order derivatives of Hamiltonian (defined on the whole space) is too restrictive.

In the first example, the drift function f, the running cost function g and the terminal cost function k are defined as follows; for simplicity, we just consider the process x_t living in $\mathbb{R}$. Let $\varepsilon_2 \in (0, 1/2]$, $\varepsilon_3 \in (0, 1/8]$, $\varepsilon_4 \in (0,1)$ be fixed constants. Define, for any x, $a \in \mathbb{R}$, $\mu \in \mathcal{P}_2(\mathbb{R})$,

$$f(x,\mu,a) := x + a + \int_{\mathbb{R}} y d\mu(y) + \varepsilon_1 x \exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y) d\mu(y)\Big)^2\Big), \tag{9.1}$$

$$g(x,\mu,a) := \frac{1}{2}a^2 + \frac{1}{2}x^2 - \varepsilon_2 x \int_{\mathbb{R}} y d\mu(y) + \varepsilon_3 a \int_{\mathbb{R}} y d\mu(y), \tag{9.2}$$

$$k(x,\mu) := \frac{1}{2}x^2 - \varepsilon_4 x \int_{\mathbb{R}} y d\mu(y), \tag{9.3}$$

where the constant $\varepsilon_1 \in (0,1)$ will be chosen sufficiently small below so that (**h2**) will be satisfied, and

$$\phi(y) := \begin{cases} |y|, & \text{for } |y| \ge 1; \\ -\frac{1}{8}y^4 + \frac{3}{4}y^2 + \frac{3}{8}, & \text{for } |y| < 1. \end{cases} \tag{9.4}$$

Remark 9.1. The drift function $f(x,\mu,a)$ in this example is both nonlinear and nonseparable across all variables. The running cost $g(x,\mu,a)$ demonstrates three distinct characteristics: (i) non-separability between the control a and measure μ, (ii) lack of strict convexity in μ[8], and (iii) quadratic growth in both x and a. Similarly, the terminal cost $k(x,\mu)$ shows quadratic growth in x and a while failing to be strictly convex in μ. Furthermore, the system exhibits weak cross-interactions among the control a, state x, and measure μ, with

[8] In our framework, the strict convexity of g and k with respect to μ plays no role in the existence theory; this is also consistent with observations made in the setting of displacement monotonicity.

the drift rate f satisfying two critical constraints: absence of significant quadratic growth effects and avoidance of overly strong couplings between x, μ and a. This actually echoes the philosophy discussed in Remark 7.1. To the best of our knowledge, this example does not fall within the scope of workable cases covered by the contemporary literature.

One can check that $\phi \in C^2(\mathbb{R})$ satisfies, $\phi(y) \geq |y|$ for all $y \in \mathbb{R}$,

$$\phi'(y) = \begin{cases} 1, & \text{for } y \geq 1; \\ -1, & \text{for } y \leq 1; \\ -\frac{1}{2}y^3 + \frac{3}{2}y, & \text{for } |y| < 1, \end{cases} \qquad \text{and} \qquad \phi''(y) = \begin{cases} 0, & \text{for } |y| \geq 1, \\ -\frac{3}{2}y^2 + \frac{3}{2}, & \text{for } |y| < 1. \end{cases} \tag{9.5}$$

In addition, we can check that $\|\mu\|_1 \leq \int_{\mathbb{R}} \phi(y)d\mu(y)$. It follows from routine calculations that

$$\begin{aligned}
(i)\ & \partial_x f(x,\mu,a) = 1 + \varepsilon_1(1-2x^2)\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big); \\
& \partial_\mu f(x,\mu,a)(\widetilde{x}) = 1 - 2\varepsilon_1 \phi'(\widetilde{x}) x \int_{\mathbb{R}} \phi(y)d\mu(y) \exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big); \\
& \partial_a f(x,\mu,a) = 1 - 2\varepsilon_1 ax \exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big); \\
& \partial_x\partial_x f(x,\mu,a) = -2x\Big(3-2x^2\Big)\varepsilon_1 \exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big); \\
& \partial_x\partial_\mu f(x,\mu,a)(\widetilde{x}) = -2\varepsilon_1\phi'(\widetilde{x})(1-2x^2)\int_{\mathbb{R}} \phi(y)d\mu(y)\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big) = \partial_\mu\partial_x f(x,\mu,a)(\widetilde{x}); \\
& \partial_x\partial_a f(x,\mu,a) = -2\varepsilon_1 a(1-2x^2)\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big) = \partial_a\partial_x f(x,\mu,a); \\
& \partial_a\partial_\mu f(x,\mu,a)(\widetilde{x}) = 4\varepsilon_1\phi'(\widetilde{x})xa\int_{\mathbb{R}} \phi(y)d\mu(y)\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big) = \partial_\mu\partial_a f(x,\mu,a)(\widetilde{x}); \\
& \partial_a\partial_a f(x,\mu,a) = -2\varepsilon_1(1-2a^2)x\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big). \\
(ii)\ & \partial_x g(x,\mu,a) = x - \varepsilon_2\int_{\mathbb{R}} y d\mu(y),\ \partial_a g(x,\mu,a) = a + \varepsilon_3\int_{\mathbb{R}} y d\mu(y); \\
& \partial_\mu g(x,\mu,a)(\widetilde{x}) = -\varepsilon_2 x + \varepsilon_3 a; \\
& \partial_a\partial_a g(x,\mu,a) = \partial_x\partial_x g(x,\mu,a) = 1,\ \partial_a\partial_\mu g(x,\mu,a)(\widetilde{x}) = \partial_\mu\partial_a g(x,\mu,a)(\widetilde{x}) = \varepsilon_3; \\
& \partial_x\partial_a g(x,\mu,a) = \partial_a\partial_x g(x,\mu,a) = 0,\ \partial_x\partial_\mu g(x,\mu,a)(\widetilde{x}) = \partial_\mu\partial_x g(x,\mu,a)(\widetilde{x}) = -\varepsilon_2. \\
(iii)\ & \partial_x k(x,\mu) = x - \varepsilon_4\int_{\mathbb{R}} y d\mu(y),\ \partial_\mu k(x,\mu)(\widetilde{x}) = -\varepsilon_4 x,\ \partial_{xx} k(x,\mu) = 1; \\
& \partial_x\partial_\mu k(x,\mu)(\widetilde{x}) = \partial_\mu\partial_x k(x,\mu)(\widetilde{x}) = -\varepsilon_4.
\end{aligned}$$

All of these functions are clearly jointly Lipschitz continuous in their corresponding arguments. Note also that $0 \leq x\exp\big(-x^2\big) \leq \frac{1}{\sqrt{2}}\exp\big(-\frac{1}{2}\big) \leq \frac{1}{2}$ and $0 \leq x^2\exp\big(-x^2\big) \leq \exp\big(-1\big) \leq \frac{1}{2}$. We emphasize here that both $g(x,\mu,a)$ and $k(x,\mu)$ do not satisfy Lasry–Lions monotonicity condition since $\partial_x\partial_\mu g(x,\mu,a)(\widetilde{x}) = -\varepsilon_2 < 0$ and $\partial_x\partial_\mu k(x,\mu)(\widetilde{x}) = -\varepsilon_4 < 0$.

Now, we shall check that whether the f, g and k defined in (9.1)–(9.3) do satisfy Assumptions (**a1**)–(**a3**) and Hypotheses (**h1**)-(**h3**) in the following:
(**a1**)(i) Since $|\partial_a f(x,\mu,a)|^2 \geq (1-\frac{1}{2}\varepsilon_1)^2$, then (4.1) is valid with $\lambda_f := 0.99$ for any $0 < \varepsilon_1 \leq 0.01$.
(ii) Since $|\partial_a f(x,\mu,a)| \leq 1 + \frac{1}{2}\varepsilon_1 =: \Lambda_1$, $|\partial_\mu f(x,\mu,a)(\widetilde{x})| \leq 1 + \frac{1}{2}\varepsilon_1 =: \Lambda_2$ and $|\partial_x f(x,\mu,a)| \leq 1 + 2\varepsilon_1 =: \Lambda_3$,

then (4.2) is valid with $\Lambda_f := 1.02$ for any $0 < \varepsilon_1 \leq 0.01$.
(iii) Using the decay nature of $\exp\Big(-x^2 - a^2 - \Big(\int_{\mathbb{R}} \phi(y)d\mu(y)\Big)^2\Big)$ and the fact that $\|\mu\|_1 \leq \int_{\mathbb{R}} \phi(y)d\mu(y)$, one can show that there exists a generic positive constant C_1 such that (4.3) is valid with $\bar{l}_f := C_1\varepsilon_1$.
(iv) The Lipschitz continuity can be verified in the same manner as in (**a1**)(iii).
(**a2**)(i) Since $\partial_a\partial_a g(x,\mu,a) = 1$, (4.6) is valid with $\lambda_g := 1$.
(ii) First of all, since $\partial_x\partial_x g(x,\mu,a) = 1$, (4.7) is valid with $\lambda_g = 1$; since $\partial_x\partial_\mu g(x,\mu,a)(\widetilde{x}) = \partial_\mu\partial_x g(x,\mu,a)(\widetilde{x}) = -\varepsilon_2$, (4.8) is valid with $l_g := \varepsilon_2 < \lambda_g$ for any $\varepsilon_2 < 1$; in addition, $\int_{\mathbb{R}} g(x,\mu,a) - g(x,\mu',a)d(\mu-\mu')(x) = -\varepsilon_2\left(\int_{\mathbb{R}} yd(\mu-\mu')(y)\right)^2$, which implies (4.44).
(iii) Using the explicit formulae for all second-order derivatives of g, one can check that (4.9) is valid with $\Lambda_g := 1$ for any $\varepsilon_2 \leq 1$ and (4.10) is valid with $\bar{l}_g := \varepsilon_3$.
(**a3**) (i) Since $\partial_x\partial_x k(x,\mu) = 1$, (4.11) is valid with $\lambda_k := 1$.
(ii) Since $\partial_x\partial_\mu k(x,\mu)(\widetilde{x}) = \partial_\mu\partial_x k(x,\mu)(\widetilde{x}) = -\varepsilon_4$, (4.12) is valid with $l_k := \varepsilon_4 < \lambda_k = 1$ for any $\varepsilon_4 < 1$; in addition, $\int_{\mathbb{R}} k(x,\mu) - k(x,\mu')d(\mu-\mu')(x) = -\varepsilon_4\left(\int_{\mathbb{R}} yd(\mu-\mu')(y)\right)^2$, which implies (4.46).
(iii) Using the explicit formulae for all second-order derivatives of k, one can check that (4.13) is valid with $\Lambda_k := 1$ for any $\varepsilon_4 \leq 1$.
(**h1**) Evaluating (9.1) and formulae for derivatives of g and k at $(x,\mu,a,\widetilde{x}) = (0,\delta_0,0,0)$ shows the fulfillment of (**h1**).
(**h2**) By some simple computations, we can show that $\frac{25}{16}\Lambda_f^2 < 1.63 < 3.733(1-\varepsilon_2) < 4(\lambda_g - l_g)\cdot\frac{\lambda_f^2}{\Lambda_g + \lambda_g/20}$ for any $\varepsilon_2 \leq \frac{1}{2} < 1 - \frac{1.63}{3.733}$, and then the validity of (**h2**) can be inferred from Remark 7.2.
(**h3**) It is worth noting that for the $\bar{l}_g$ and λ_g chosen above, we have $\bar{l}_g = \varepsilon_3 \leq \frac{1}{8} = \frac{1}{8}\lambda_g$. Furthermore, the constant L_0^* defined in (7.56) depends only on $\lambda_f = 0.99$, $\Lambda_f = 1.02$, $\lambda_g = 1$, $\Lambda_g = 1$, $\lambda_k = 1$ and $\Lambda_k = 1$ but not on $\bar{l}_f$ or ε_1, so we can always choose an ε_1 small enough, such that $\bar{l}_f = C_1\varepsilon_1 \leq \frac{1}{40\max\{\Lambda_k, L_0^*\}}\lambda_g = \frac{1}{40L_0^*}$. Therefore, (**h3**) is satisfied.

Now, we shall use a simpler model, with explicit computations, to illustrate the boundedness conditions on the second-order derivatives of Hamiltonian are too restrictive, so that even simple perturbations of LQ problem are not allowed. In this second example, we consider an even simpler drift function:

$$f(x,\mu,a) := x + a + \int_{\mathbb{R}} yd\mu(y) + \varepsilon_1 x\exp\Big(-x^2 - \Big(\int_{\mathbb{R}}\phi(y)d\mu(y)\Big)^2\Big), \tag{9.6}$$

subject to the running and terminal cost functions g and k in (9.2) and (9.3) with $\epsilon_2 = \epsilon_3 = \epsilon_4 = 0$. Due to the simpler structure of this example, it follows from the first-order condition that the optimal control $\widehat{\alpha}(x,\mu,z) = -z$, and hence, the Hamiltonian is

$$h(x,\mu,z) = \frac{1}{2}x^2 + xz - \frac{1}{2}z^2 + z\int_{\mathbb{R}} yd\mu(y) + \varepsilon_1 xz\exp\Big(-x^2 - \Big(\int_{\mathbb{R}}\phi(y)d\mu(y)\Big)^2\Big). \tag{9.7}$$

A direct differentiation yields

$$\partial_x\partial_x h(x,\mu,z) = 1 - 2x\Big(3 - 2x^2\Big)\varepsilon_1\exp\Big(-x^2 - \Big(\int_{\mathbb{R}}\phi(y)d\mu(y)\Big)^2\Big)\cdot z. \tag{9.8}$$

Clearly, $\partial_x\partial_x h(x,\mu,z)$ is not uniformly bounded in $\mathbb{R}^{d_x}\times\mathcal{P}_2(\mathbb{R}^{d_x})\times\mathbb{R}^{d_x}$, and thus those recent results in Bansil-Mészáros [66] and Bansil-Mészáros-Mou [67] cannot apply to this example. However, one can directly check that this example still satisfy all the Assumptions (**a1**)–(**a3**) and Hypotheses (**h1**)-(**h3**) used in this paper, provided that the constant ε_1 is small enough.

Acknowledgments

Alain Bensoussan is supported by the National Science Foundation under grant NSF-DMS-2204795. Tak Kwong Wong was partially supported by the HKU Seed Fund for Basic Research under the project code 201702159009, the Start-up Allowance for Croucher Award Recipients, and Hong Kong General Research Fund (GRF) grants with project numbers 17306420, 17302521, and 17315322. Phillip Yam acknowledges the financial supports from HKGRF-14301321 with the project title "General Theory for Infinite Dimensional Stochastic Control: Mean Field and Some Classical Problems", and HKGRF-14300123 with the project title "Well-posedness of Some Poisson-driven Mean Field Learning Models and their Applications". The work described in this article was supported by a grant from the Germany/Hong Kong Joint Research Scheme sponsored by the Research Grants Council of Hong Kong and the German Academic Exchange Service of Germany (Reference No. G-CUHK411/23). He also thanks The University of Texas at Dallas for the kind invitation to be a Visiting Professor in Naveen Jindal School of Management during his sabbatical leave. Hongwei Yuan thanks the Department of Statistics and Data Science of The Chinese University of Hong Kong and the Department of Mathematics of University of Macau for the financial support. The authors are grateful to an anonymous referee for encouraging us to make a serious comparison with some contemporary literature; this discussion has further highlighted the novelty of our results.

Data availability statement

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

References

[1] M. Huang, R.P. Malhamé and P.E Caines, Large population stochastic dynamic games: closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6** (2006) 221–252.

[2] P.-L. Lasry and J.-M. Lions, Mean field games. *Jpn. J. Math.* **2** (2007) 229–260.

[3] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games (AMS-201)*, vol. 201. Princeton University Press (2019).

[4] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I–II.* Springer (2018).

[5] D.A. Gomes, E.A. Pimentel and V. Voskanyan, *Regularity Theory for Mean-Field Game Systems.* Springer (2016).

[6] A. Bensoussan, J. Frehse and S.C.P. Yam, *Mean Field Games and Mean Field Type Control Theory*, vol. 101. Springer (2013).

[7] D. Andersson and B. Djehiche, A maximum principle for SDEs of mean-field type. *Appl. Math. Optim.* **63** (2011) 341–356.

[8] A. Bensoussan, P.J. Graber and S.C.P. Yam, Control on Hilbert spaces and application to some mean field type control problems. *Ann. Appl. Probab.* **34** (2024) 4085–4136.

[9] A. Bensoussan, K.C.J. Sung, S.C.P. Yam and S.P. Yung, Linear-quadratic mean field games. *J. Optim. Theory Appl.* **169** (2016) 496–529.

[10] A. Bensoussan, H.M. Tai, T.K. Wong and S.C.P. Yam, A control theoretical Approach to Mean Field Games: Part I–Global Equilibrium Solution. *Appl. Math. Optim.* **92** (2025) 44.

[11] A. Bensoussan, H.M. Tai, T.K. Wong, and S.C.P. Yam, A Control Theoretical Approach to Mean Field Games. Part II: Global Well-Posedness of Master Equations. *Appl. Math. Optim.* **93** (2026) 45.

[12] A. Bensoussan, H.M. Tai and S.C.P. Yam, Mean field type control problems, some Hilbert-space-valued fbsdes, and related equations. *ESAIM Control Optim. Calc. Var.* **31** (2025) 33.

[13] R. Buckdahn, B. Djehiche and J. Li, A general stochastic maximum principle for SDEs of mean-field type. *Appl. Math. Optim.* **64** (2011) 197–216.

[14] R. Buckdahn, J. Li and J. Ma, A mean-field stochastic control problem with partial observations. *Ann. Appl. Probab.* **27** (2017) 3201–3245.

[15] R. Carmona and F. Delarue, Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *Ann. Probab.* **43** (2015) 2647–2700.

[16] J.-F. Chassagneux, D. Crisan and F. Delarue, A probabilistic approach to classical solutions of the master equation for large population equilibria. *Mem. Am. Math. Soc.* **280** (2022).

[17] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham and M. Rosestolato, Optimal control of path-dependent McKean–Vlasov SDEs in infinite-dimension. *Ann. Appl. Probab.* **33** (2023) 2863–2918.

[18] A. Cosso and H. Pham, Zero-sum stochastic differential games of generalized McKean–Vlasov type. *J. Math. Pures Appl.* **129** (2019) 180–212.

[19] M.F. Djete, D. Possamaï and X. Tan, Mckean–vlasov optimal control: the dynamic programming principle. *Ann. Probab.* **50** (2022) 791–833.

[20] E. Pardoux and S. Peng, Backward stochastic differential equations and quasilinear parabolic partial differential equations, in *Stochastic Partial Differential Equations and Their Applications.* Springer (2005) 200–217.

[21] H. Pham, Linear quadratic optimal control of conditional McKean–Vlasov equation with random coefficients and applications. *Probab. Uncertainty Quant. Risk* **1** (2016) 1–26.

[22] H. Pham and X. Wei, Dynamic programming for optimal control of stochastic McKean–Vlasov dynamics. *SIAM J. Control Optim.* **55** (2017) 1069–1101.

[23] H. Pham and X. Wei, Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control Optim. Calc. Var.* **24** (2018) 437–461.

[24] A. Bensoussan, S.C.P. Yam and Z. Zhang, Well-posedness of mean-field type forward–backward stochastic differential equations. *Stoch. Processes Appl.* **125** (2015) 3327–3354.

[25] A. Bensoussan, J. Frehse and S.C.P. Yam, The master equation in mean field theory. *J. Math. Pures Appl.* **103** (2015) 1441–1474.

[26] A. Bensoussan and S.C.P. Yam, Control problem on space of random variables and master equation. *ESAIM Control Optim. Calc. Var.* **25** (2019) 10.

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

[28] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham and M. Rosestolato, Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions. *Trans. Am. Math. Soc.* (2023).

[29] W. Gangbo and A. R. Mészáros, Global well-posedness of master equations for deterministic displacement convex potential mean field games. *Commun. Pure Appl. Math.* **75** (2022) 2685–2801.

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

[31] W. Gangbo and A. Święch, Existence of a solution to an equation arising from the theory of mean field games. *J. Differ. Equ.* **259** (2015) 6573–6643.

[32] M. Laurière and O. Pironneau, Dynamic programming for mean-field type control. *Comptes Rendus Math.* **352** (2014) 707–713.

[33] J. Li, Stochastic maximum principle in the mean-field controls. *Automatica* **48** (2012) 366–373.

[34] R. Carmona, F. Delarue and A. Lachapelle, Control of McKean–Vlasov dynamics versus mean field games. *Math. Financial Econ.* **7** (2013) 131–166.

[35] J. Yong, Linear-quadratic optimal control problems for mean-field stochastic differential equations. *SIAM J. Control Optim.* **51** (2013) 2809–2838.

[36] P. Cardaliaguet, Weak solutions for first order mean field games with local coupling. *Analysis and Geometry in Control Theory and its Applications* (2015) 111–158.

[37] P.J. Graber, Optimal control of first-order Hamilton–Jacobi equations with linearly bounded Hamiltonian. *Appl. Math. Optim.* **70** (2014) 185–224.

[38] P. Cardaliaguet and P.J. Graber, Mean field games systems of first order. *ESAIM Control Optim. Calc. Var.* **21** (2015) 690–722.

[39] A. Porretta, Weak solutions to Fokker–Planck equations and mean field games. *Arch. Rational Mech. Anal.* **216** (2015) 1–62.

[40] D. Ambrose, Strong solutions for time-dependent mean field games with non-separable Hamiltonians. *J. Math. Pures Appl.* **113** (2018) 141–154.

[41] P.-L. Lions, Cours au Collège de France.

[42] P. Cardaliaguet, Notes on mean-field games (from P.-L. Lions' lectures at Collège de France) (2013).

[43] A. Bensoussan, J. Frehse and S.C.P. Yam, On the interpretation of the master equation. *Stoch. Processes Appl.* **127** (2017) 2093–2137.

[44] S. Mayorga, Short time solution to the master equation of a first order mean field game. *J. Differ. Equ.* **268** (2020) 6251–6318.

[45] D. Ambrose and A. Mészáros, Well-posedness of mean field games master equations involving non-separable local Hamiltonians. *Trans. Am. Math. Soc.* **376** (2023) 2481–2523.

[46] C. Mou and J. Zhang, Wellposedness of second order master equations for mean field games with nonsmooth data. *Mem. Am. Math. Soc.* (2021).

[47] C. Bertucci, Monotone solutions for mean field games master equations: continuous state space and common noise. *Commun. Part. Differ. Equ.* **48** (2023) 1245–1285.

[48] P. Cardaliaguet and P. Souganidis, Monotone Solutions of the Master Equation for Mean Field Games with Idiosyncratic Noise. *SIAM J. Math. Anal.* **54** (2022). DOI: https://doi.org/10.1137/21M1450008

[49] A. Cecchin, D.P. Paolo, F. Markus and P. Guglielmo, On the convergence problem in mean field games: a two state model without uniqueness. *SIAM J. Control Optim.* **57** (2019) 2443–2466.

[50] F. Delarue and R.F. Tchuendom, Selection of equilibria in a linear quadratic mean-field game. *Stoch. Processes Appl.* **130** (2020) 1000–1040.

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

[52] A. Bensoussan, T.K. Wong, S.C.P. Yam and H. Yuan, A theory of first order mean field type control problems and their equations. *J. Eur. Math. Soc.* (2026), published online first. DOI: https://doi.org/10.4171/JEMS/1781

[53] P.J. Graber and A.R. Mészáros, On monotonicity conditions for mean field games. *J. Funct. Anal.* **285** (2023) 110095.

[54] A. Bensoussan, Z. Huang, S. Tang, and S.C.P. Yam, On mean field monotonicity conditions from a control theoretical perspective. *Rend. Lincei Mat. Appl.* **36** (2025) 89–137.

[55] A. Cecchin and F. Delarue Weak solutions to the Master Equation of Potential Mean Field Games. *Mem. Am. Math. Soc.* (2025). DOI: https://doi.org/10.1090/memo/1600

[56] P.J. Graber and A. Mészáros, On some mean field games and master equations through the lens of conservation laws. *Math. Ann.* **390** (2024) 4497–4533.

[57] P.-L. Lions and P.E. Souganidis, Homogenization of the backward–forward mean-field games systems in periodic environments. *Rend. Lincei* **31** (2021) 733–755.

[58] J.-M. Lasry, P.-L. Lions and B. Seeger, Dimension reduction techniques in deterministic mean field games. *Commun. Part. Differ. Equ.* **47** (2022) 701–723.

[59] R. Carmona and D. Lacker, A probabilistic weak formulation of mean field games and applications. *The Ann. Appl. Probab.* **25** (2018) 1189–1231.

[60] R. Carmona, F. Delarue and D. Lacker, Mean field games with common noise. *Ann. Probab.* **44** (2016) 3740–3803.

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

[62] R. Buckdahn, B. Djehiche, J. Li and S. Peng, Mean-field backward stochastic differential equations: a limit approach. *Ann. Probab.* **37** (2009) 1524–1565.

[63] R. Buckdahn, J. Li and S. Peng, Mean-field backward stochastic differential equations and related partial differential equations. *Stoch. Processes Appl.* **119** (2009) 3133–3154.

[64] R. Carmona and F. Delarue. Mean field forward–backward stochastic differential equations. *Electron. Commun. Probab.* **18** (2013) 1–15.

[65] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media (2008).

[66] M. Bansil and A. Mészáros, Hidden monotonicity and canonical transformations for mean field games and master equations, in *Forum of Mathematics, Sigma*, vol. 13. Cambridge University Press (2025) e182.

[67] M. Bansil, A. Mészáros and C. Mou, Global well-posedness of displacement monotone degenerate mean field games master equations. *SIAM J. Control Optim.* **63** (2025) 993–1021.

[68] P. Caines, Mean field game control theory: Introduction and recent results. https://www.cityu.edu.hk/ma/news-events/mean-field-game-control-theory-introduction-and-recent-results (2015).

[69] A. Bensoussan, *Estimation and Control of Dynamical Systems*, vol. 48. Springer (2018).

[70] C. Bertucci, J.-M. Lasry and P.-L. Lions, On Lipschitz solutions of mean field games master equations. *J. Funct. Anal.* **287** (2024) 110486.