Splitting methods and short time existence for the master equations in mean field games

Pierre Cardaliaguet, Marco Cirant, Alessio Porretta

Splitting methods and short time existence for the master equations in mean field games

Pierre Cardaliaguet* Marco Cirant† Alessio Porretta‡

January 24, 2020

Abstract

We develop a splitting method to prove the well-posedness, in short time, of solutions for two master equations in mean field game (MFG) theory: the second order master equation, describing MFGs with a common noise, and the system of master equations associated with MFGs with a major player. Both problems are infinite dimensional equations stated in the space of probability measures. Our new approach simplifies, shortens and generalizes previous existence results for second order master equations and provides the first existence result for systems associated with MFG problems with a major player.

Contents

*Université Paris-Dauphine, PSL Research University, CNRS, Ceremade, 75016 Paris, France. cardaliaguet@ceremade.dauphine.fr

†Università di Parma, Parco Area delle Scienze 53/a, 43124 Parma, Italy. marcoalessandro.cirant@unipr.it

‡Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy. porretta@mat.uniroma2.it

1 Introduction

The paper is dedicated to a construction of a solution of the so-called "master equations" in mean field game theory (MFG). These equations have been introduced by Lasry and Lions and discussed by Lions in [27]. Let us recall that mean field games describe the behavior of infinitely many agents in interaction. We consider here two problems: the master equation with common noise and the master equation with a major player. We present a general approach valid for both problems.

Let us first discuss the master equation with common noise. In this problem, the agents are subject to a common source of randomness. The master equation is then a second order equation in the space of measures and reads as follows:

$$\begin{cases} -\partial_t U(t,x,m) - \mathrm{Tr}((a(t,x)+a^0(t,x))D^2_{xx}U(t,x,m)) + H(x, D_xU(t,x,m), m) \\ \quad - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}((a(t,y)+a^0(t,y))D^2_{ym}U(t,x,m,y))\ dm(y) \\ \quad + \displaystyle\int_{\mathbb{R}^d} D_mU(t,x,m,y)\cdot H_p(y, D_xU(t,y,m), m)\ dm(y) \\ \quad -2\displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[\sigma^0(t,y)(\sigma^0(t,x))^T D^2_{xm}U(t,x,m,y)\right] m(dy) \\ \quad - \displaystyle\int_{\mathbb{R}^{2d}} \mathrm{Tr}[\sigma^0(t,y)(\sigma^0(t,y'))^T D^2_{mm}U(t,x,m,y,y')]m(dy)m(dy') = 0 \\ \qquad\qquad \text{in } (0,T)\times\mathbb{R}^d\times\mathcal{P}_2 \\ U(T,x,m) = G(x,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2 \end{cases} \tag{1}$$

In the above equation, the unknown $U = U(t,x,m)$ is scalar valued and depends on the time variable $t \in [0,T]$, the space variable $x \in \mathbb{R}^d$ and the distribution of the agents $m \in \mathcal{P}_2$ ($\mathcal{P}_2$ is the space of Borel probability measures with finite second order moment); the derivatives D_mU and $D^2_{mm}U$ refer to the derivative with respect to the probability measure (see subsection 2.2); the maps $H = H(x,p,m)$ and $G = G(x,m)$ reflect the running and terminal costs of the agents. The matrix valued function $a = a(t,x)$ is the volatility term corresponding to idiosyncratic noise of the small players while $a^0 = a^0(t,x) = \sigma^0(\sigma^0)^T(t,x)$ is the volatility corresponding to the common noise.

As explained by Lions [27], the master equation can be understood as a non-linear transport equation in the space of probability measures. When $a^0 = 0$ (i.e., in the so-called first order master equation), the characteristics of this transport equation are given by the MFG system: if we fix an initial time t_0 and

an initial probability measure m_0 on $\mathbb{R}^d$, and if the pair (u, m) is a solution of the MFG system

$$\left\{\begin{array}{ll} (i) & -\partial_t u - \mathrm{Tr}(a(t,x)D^2u) + H(x, Du, m(t)) = 0 \qquad \text{in } (t_0,T)\times\mathbb{R}^d \\ (ii) & \partial_t m - \sum_{i,j} D_{ij}(a_{i,j}m) - \mathrm{div}(mH_p(x, Du, m(t)) = 0 \qquad \text{in } (t_0,T)\times\mathbb{R}^d \\ (iii) & m(t_0) = m_0, \qquad u(T,x) = G(x, m(T)) \qquad \text{in } \mathbb{R}^d \end{array}\right. \tag{2}$$

then we expect the equality

$$U(t, x, m(t)) = u(t, x) \qquad \forall t \in [t_0, T]. \tag{3}$$

The interpretation of the MFG system (2) is the following: the map u is the value function of a typical small agent (anticipating the evolution of the population density $(m(t))$ and accordingly solves the Hamilton-Jacobi equation (2)-(i). When this agent plays in an optimal way, the drift in the dynamic of its state is given by the term $-H_p(x, Du, m(t))$. By a mean field argument (assuming that the noise of the agents are independent), the resulting evolution of the population density $\tilde{m}$ satisfies the Kolmogorov equation

$$\left\{\begin{array}{l} \partial_t\tilde{m} - \sum_{i,j} D_{ij}(a_{i,j}\tilde{m}) - \mathrm{div}(\tilde{m}H_p(x, Du, m(t)) = 0 \qquad \text{in } (t_0,T)\times\mathbb{R}^d \\ m(t_0) = m_0 \qquad \text{in } \mathbb{R}^d \end{array}\right.$$

In an equilibrium configuration, i.e., when agents anticipate correctly the evolving measure, one has $\tilde{m} = m$ and therefore the population density m solves (2)-(ii).

The existence/uniqueness of the solution for the MFG system is rather well understood: it relies on Schauder estimates, fixed point methods and monotonicity arguments (see, in particular, [24, 25]). From the well-posedness of the MFG system, one can derive the existence of a solution to the first order master equation "quite easily": one just needs to define the map U by (3) with $t = t_0$ and check that the map U thus defined is a classical solution to the first order master equation. This is the path followed in [17, 28] (when there is no diffusion at all: $a = a^0 \equiv 0$) and in [16] (when $a > 0$ is constant and $a^0 = 0$). See also [9] for a similar result (in the torus) using PDE linearization techniques.

When $a^0 \not\equiv 0$ (i.e., for the second order master equation, or master equation with a common noise), the characteristics are now given by the system of SPDEs (called "stochastic MFG system"):

$$\left\{\begin{array}{l} du(t,x) = \big[-\mathrm{Tr}((a+a^0)(t,x)D^2u(t,x)) + H(x, Du(t,x), m(t)) \\ \qquad\qquad -\sqrt{2}\mathrm{Tr}(\sigma^0(t,x)Dv(t,x))\big]dt + v(t,x)\cdot dW_t \qquad \text{in } (0,T)\times\mathbb{R}^d, \\ dm(t,x) = \big[\sum_{i,j} D_{ij}(((a_{ij}) + a^0_{ij})(t,x)m(t,x)) + \mathrm{div}\big(m(,x)D_pH(x, Du(t,x), m(t))\big)\big]dt \\ \qquad\qquad -\mathrm{div}(m(t,x)\sqrt{2}\sigma^0(t,x)dW_t\big), \qquad \text{in } (0,T)\times\mathbb{R}^d, \\ u(T,x) = G(x, m(T)),\ m(0) = m_0, \qquad \text{in } \mathbb{R}^d \end{array}\right. \tag{4}$$

In the above system, (W_t) is the common noise (here a Brownian motion) and the unknown is the triplet (u, m, v), where the new variable v (a random vector field in $\mathbb{R}^d$) ensures the solution u of the backward Hamilton-Jacobi (HJ) equation to be adapted to the filtration generated by the common noise (W_t). The analysis of this system is much more involved than the deterministic one: Schauder estimates are no longer available and the usual fixed point methods based on compactness arguments can no longer be applied. One has to replace them by continuation methods, which are much heavier to handle, see [9]. Besides the PDE approach we just mentioned, MFG with common noise can also be handled through a probabilistic formulation: see the pioneering result [12], as well as [2, 21] and the monograph [11]. Once the analysis of the stochastic MFG system has been performed, one can proceed with the construction of the second order master equation as in the first order case, defining the map U by (3) for $t = t_0$, where u is now the $u-$component of the solution of the stochastic MFG system ($u(t_0, \cdot)$ turns out to be deterministic). However, here again, the verification that the map U defined so far is smooth enough to satisfy (1) requires a lot of work: see [9] and [11].

Let us finally recall another approach, suggested by P.-L. Lions in the seminar [29]: it consists in writing the equation for the quantity D_mU as an hyperbolic equation in a Hilbert space of random variables. The construction requires, however, convexity conditions on the system with respect to the space variable (but no uniform ellipticity for the matrix a).

We now discuss the second equation considered in this paper: the master equation corresponding to MFG models with a major player. It reads as follows:

$$\left\{\begin{array}{ll}
(i) & -\partial_t U^0 - \Delta_{x_0} U^0 + H^0(x_0, D_{x_0} U^0, m) - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U^0(t, x_0, m, y) dm(y) \\
& \quad + \displaystyle\int_{\mathbb{R}^d} D_m U^0(t, x_0, m, y) \cdot H_p(x_0, y, D_x U(t, x_0, y, m), m) dm(y) = 0 \\
& \qquad\qquad \text{in } (0,T) \times \mathbb{R}^{d_0} \times \mathcal{P}_2, \\
(ii) & -\partial_t U - \Delta_x U - \Delta_{x_0} U + H(x_0, x, D_x U, m) - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U(t, x_0, x, m, y) dm(y) \\
& \quad + D_{x_0} U \cdot H_p^0(x_0, D_{x_0} U^0(t, x_0, m), m) \\
& \quad + \displaystyle\int_{\mathbb{R}^d} D_m U(t, x_0, x, m, y) \cdot H_p(x_0, y, D_x U(t, x_0, y, m), m) dm(y) = 0 \\
& \qquad\qquad \text{in } (0,T) \times \mathbb{R}^{d_0} \times \mathbb{R}^d \times \mathcal{P}_2, \\
(iii) & U^0(T, x_0, m) = G^0(x_0, m), \qquad \text{in } \mathbb{R}^{d_0} \times \mathcal{P}_2, \\
(iv) & U(T, x_0, x, m) = G(x_0, x, m) \qquad \text{in } \mathbb{R}^{d_0} \times \mathbb{R}^d \times \mathcal{P}_2.
\end{array}\right. \tag{5}$$

In the above system, $U^0 = U^0(t, x_0, m)$ corresponds to the payoff at equilibrium for a major player interacting with a crowd in which each agent has at equilibrium a payoff given by $U = U(t, x_0, x, m)$. Here m is the distribution law of the agents. Notice that each agent is influenced by the major player whereas the latter is only influenced by the distribution of the whole population. Mean field games with a major player have been first discussed by Huang in [20] and several notions of equilibria, in different contexts, have been proposed in the literature since then: see [5, 6, 7, 8, 11, 13, 14, 15, 26]. The above system has been introduced by Lasry and Lions in [26]. In the companion paper [10], we explain how the above master equation is related to the approach by Carmona and al. [13, 14, 15]. Concerning the existence of a solution, [15] shows the existence of an equilibrium in short time for the case of a finite state space, [26] proves the existence of a solution to the master equation still in the finite state space framework and notes that the Hilbertian techniques described in [29] could be adapted to the master equation with a major player (5).

The purpose of this paper is to introduce a different path towards the construction of a solution to the second order master equation and to the master equation with a major player, using as a building block the construction of a solution to the first order master equation. For the second order master equation, we justify this point of view by the fact that the deterministic MFG system and the first order master equation are much easier to manipulate than the stochastic MFG system. Our approach allows for instance to build solutions of the second order master equation (in short time) under more general assumptions than in [9, 11]. For the MFG problem with a major player, we prove for the first time the (short time) well-posedness of the associated system of master equations in continuous space.

Let us first explain our ideas for the master equation with common noise (1). In contrast to previous works, we do not use directly the representation formula (3) (for $t = t_0$) for the solution of the second order master equation. Instead, we somehow decompose the second order master equation as the superposition of the first order master equation:

$$\left\{\begin{array}{l}
-\partial_t U - \mathrm{Tr}(a(t,x) D^2_{xx} U) + H(x, D_x U, m) - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}(a(t,y) D^2_{ym} U)\ dm(y) \\
\qquad + \displaystyle\int_{\mathbb{R}^d} D_m U \cdot H_p(y, D_x U, m)\ dm(y) = 0 \qquad \text{in } (0,T) \times \mathbb{R}^d \times \mathcal{P}_2 \\
U(T, x, m) = G(x, m) \qquad \text{in } \mathbb{R}^d \times \mathcal{P}_2
\end{array}\right. \tag{6}$$

and of a linear second order master equation:

$$\begin{cases} -\partial_t U - \mathrm{Tr}\left[\sigma^0(\sigma^0)^T(t,x)D^2_{xx}U\right] - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[\sigma^0(\sigma^0)^T(t,y)D^2_{ym}U\right] m(dy) \\ \qquad -2\displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[\sigma^0(t,y)(\sigma^0(t,x))^T D^2_{xm}U\right] m(dy) \\ \qquad -\displaystyle\int_{\mathbb{R}^{2d}} \mathrm{Tr}[\sigma^0(t,y)(\sigma^0(t,y'))^T D^2_{mm}U]m(dy)m(dy') = 0 \\ \qquad\qquad\qquad\qquad \text{in } (0,T)\times\mathbb{R}^d\times\mathcal{P}_2 \\ U(T,x,m) = G(x,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2 \end{cases} \tag{7}$$

The solution to this linear second order master equation is just given by a Feynman-Kac formula, and thus it is very easy to handle. Then we use Trotter-Kato formula, alternating the two equations in short time intervals to build in the limit a solution of the full equation (1). If the technique is quite transparent, its actual implementation requires some care. Indeed, one has to check that, at each step of the process, the regularity of the solution does not deteriorate too much, meaning at least in a linear way in time. The aim of Section 6 is precisely to quantify this deterioration for the solution U of the first order master equation (6), as well as for its derivatives in the measure variable. As the solution of (6) is built by using the representation formula (3) (where $t = t_0$) presented above, one has first to do the analysis on the MFG system (2) and this is the aim of Section 5. Note that we are able to control the regularity of the linear second order equation (7) only when the matrix a^0 is constant. Hence we only prove the short time existence of a solution to (1) in that case.

For the problem with a major player, we argue in a similar way: we view equation (5) as the superposition of two systems: the first one is a first order system of master equations (for a fixed x_0):

$$\begin{cases} (i) \quad -\partial_t U^0 - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U^0(t,x_0,m,y)dm(y) \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d} D_m U^0(t,x_0,m,y)\cdot H_p(x_0,y,D_xU(t,x_0,y,m),m)dm(y) = 0 \\ (ii) \quad -\partial_t U - \Delta_x U + H(x_0,x,D_xU,m) - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U(t,x_0,x,m,y)dm(y) \\ \qquad\qquad + \displaystyle\int_{\mathbb{R}^d} D_m U(t,x_0,x,m,y)\cdot H_p(x_0,y,D_xU(t,x_0,y,m),m)dm(y) = 0 \end{cases}$$

It turns out that this system can be handled by the method of characteristics. As for the second one, it is a simple system of HJ equations (for fixed x, m):

$$\begin{cases} (i) \quad -\partial_t U^0 - \Delta_{x_0} U^0 + H^0(x_0, D_{x_0}U^0, m) = 0 \\ (ii) \quad -\partial_t U - \Delta_{x_0} U + D_{x_0}U\cdot H^0_p(x_0, D_{x_0}U^0(t,x_0,m),m) = 0\,. \end{cases}$$

The idea of splitting time is not completely new in the framework of mean field games. For instance, the construction, given in [12], of (weak) equilibria for MFG problems with common noise relies on a time splitting. The main difference is that it is done at the level of the MFG equilibrium, while we do the construction at the (stronger) level of the master equation. One consequence is that, with our approach, the construction of a solution to the stochastic MFG system (in short time, though) is straightforward once the solution of the master equation is built, while deriving a solution of the master equation from the stochastic MFG system is much trickier. Let us also quote the paper in preparation [1] in which the authors use a splitting technique similar to the one described above to compute numerically the solution of MFGs with a major player.

Let us finally point out that, in this paper, we do not address at all the problem of the existence of a solution on a large time interval. For the first and second order master equation, this question is related to the Lasry-Lions monotonicity condition [24, 25]. The existence of a solution on a large time interval can be obtained under this condition either by the Hilbertian approach, as explained in [29], or by a continuation method, as in [16] and [11] or even directly by using the long time existence of a solution for the MFG system, as in [9]. Let us recall that, when the monotonicity condition is not fulfilled, the solution to the second order master equation is expected to develop shocks (i.e., discontinuities) in finite

time. Note also that a structure condition similar to the monotonicity condition is not known for MFGs with a major player.

The paper is organized in the following way. In Section 2 we fix the notation, we recall the definition of derivatives in the space of measures and state our main assumptions. The main existence results for the second order master equation (equation (1)) and for the system of master equations for MFG with a major player (system (5)) are collected in Sections 3 and 4 respectively. Both sections require estimates on the first order master equations. As first order master equations are built by the method of characteristics involving the solutions of classical MFG systems (2), Section 5 first provides estimates for these systems. Then Section 6 is devoted to the first order master equations. We complete the paper by appendices in which we prove short-time estimates for the standard Hamilton-Jacobi equations (Section A) and we discuss several facts on maps defined on the space of measures (differentiability, interpolation and Ascoli Theorem, Section B).

Acknowledgement. The first author was partially supported by the ANR (Agence Nationale de la Recherche) project ANR-16-CE40-0015-01 and by the AFOSR grant FA9550-18-1-0494. The second author was partially supported by the Fondazione CaRiPaRo Project "Nonlinear Partial Differential Equations: Asymptotic Problems and Mean-Field Games" and the Programme "FIL-Quota Incentivante" of University of Parma, co-sponsored by Fondazione Cariparma. The third author was partially supported by Indam (Gnampa national project 2018) and by FSMP (Foundation Sciences Mathématiques de Paris).

2 Notation and assumptions

2.1 Notation

Throughout the paper, we work in the euclidean space $\mathbb{R}^d$ (with $d \in \mathbb{N}$, $d \geqslant 1$), endowed with the scalar product $(x, y) \to x \cdot y$ and the distance $|\cdot|$. Given $T > 0$ and a map $\phi : (0, T) \times \mathbb{R}^d \to \mathbb{R}$, we denote by $\partial_t \phi$ the derivative of ϕ with respect to the time-variable, by $\partial_{x_i} \phi$ its partial derivative with respect to the $i-$th space variable ($i = 1, \dots, d$) and by $D\phi$ the gradient with respect to the space variable.

For $n \in \mathbb{N}$, we denote by C_b^n the set of maps $\phi : \mathbb{R}^d \to \mathbb{R}$ which are $n-$times differentiable with continuous and bounded derivatives: in particular, C_b^0 is the set of continuous and bounded maps. Given $\phi \in C_b^n$ and a multi-index $k = (k_1, \dots, k_d) \in \mathbb{N}^d$, with length $|k| := \sum_{i=1}^d k_i \leqslant n$, we denote by $\partial^k \phi = \frac{\partial^{k_1}}{\partial x_1^{k_1}} \dots \frac{\partial^{k_d}}{\partial x_d^{k_d}} \phi$ (or briefly ϕ_k) the $k-$th derivative of ϕ. We also denote by $D^n \phi$ ($n \in \mathbb{N}$, $n \geqslant 1$) the vector $(\partial^k \phi)_{|k|=n}$. The norm of ϕ in C_b^n is

$$\|\phi\|_n := \sum_{r=0}^n \sup_x \left(\sum_{|\alpha|=r} |\partial^\alpha \phi(x)|^2 \right)^{1/2} = \sum_{r=0}^n \|D^r \phi\|_\infty.$$

For $n = 0$, we use indifferently the notation $\|\phi\|_0$ or $\|\phi\|_\infty$.

For $(n_1, \dots, n_k) \in \mathbb{N}^k$ ($k \in \mathbb{N}$, $k \geqslant 2$), we denote by $C_b^{n_1, \dots, n_k}$ the space of functions $\phi : \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_k} \to \mathbb{R}$ ($d_i \geqslant 1$) having continuous and bounded derivatives $D_{x_1}^{l_1} \cdots D_{x_k}^{l_k} \phi$ for all $l_1 \leqslant n_1, \dots, l_k \leqslant n_k$, endowed with the norm

$$\|\phi\|_{n_1, \dots, n_k} = \|\phi(\cdot_{x_1}, \dots, \cdot_{x_k})\|_{n_1, \dots, n_k} := \sum_{l_1 \leqslant n_1, \dots, l_k \leqslant n_k} \|D_{x_1}^{l_1} \cdots D_{x_k}^{l_k} \phi\|_\infty,$$

where now $(x_1, \dots, x_k)$ stands for a generic element $\mathbb{R}^{d_1} \times \mathbb{R}^{d_k}$.

We denote by C^{-n} the dual space of C_b^n, endowed with the usual norm

$$\|\rho\|_{-n} := \sup_{\|\phi\|_n \leqslant 1} \rho(\phi) \qquad \forall \rho \in C^{-n}.$$

Finally, when a map $\phi = \phi(t, x)$ depends also on time t belonging to an interval I, we often write $\sup_{t \in I} \|\phi(t)\|_n$ for $\sup_{t \in I} \|\phi(t, \cdot)\|_n$. We use a corresponding notation for a map $\rho \in C^0([0, T], C^{-k})$.

Throughout the paper, $\mathcal{P}$ stands for the set of Borel probability measures on $\mathbb{R}^d$ and for $k \geqslant 1$, $\mathcal{P}_k$ stands for the set of measures in $\mathcal{P}$ with finite moment of order k: namely,

$$M_k(m) := \left(\int_{\mathbb{R}^d} |x|^k m(dx) \right)^{1/k} < +\infty \qquad \text{if } m \in \mathcal{P}_k.$$

The set $\mathcal{P}_k$ is endowed with the distance (see for instance [4, 30, 31])

$$\mathbf{d}_k(m, m') = \inf_{\pi} \left(\int_{\mathbb{R}^d} |x - y|^k \pi(dx, dy) \right)^{1/k}, \qquad \forall m, m' \in \mathcal{P}_k,$$

where the infimum is taken over the couplings π between m and m', i.e., over the Borel probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal m and second marginal m'. Note that $\mathcal{P}_2 \subset \mathcal{P}_1$ and $\mathbf{d}_1 \leqslant \mathbf{d}_2$ by Cauchy-Schwarz inequality. We will often use the fact that, if $\phi : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous with a Lipschitz constant $L \geqslant 0$, then

$$\left| \int_{\mathbb{R}^d} \phi(x)(m - m')(dx) \right| \leqslant L \mathbf{d}_1(m, m'), \qquad \forall m, m' \in \mathcal{P}_1.$$

Moreover, $\mathbf{d}_1(m, m')$ is the smallest constant for which the above inequality holds for any $L-$Lipschitz continuous map ϕ (see for instance [30, 31]). Given $m \in \mathcal{P}$ and $\phi \in C_b^0$, the image $\phi \sharp m$ of m by ϕ is the element of $\mathcal{P}$ defined by

$$\int_{\mathbb{R}^d} f(x)\ m\sharp\phi(dx) = \int_{\mathbb{R}^d} f(\phi(x)) m(dx) \qquad \forall f \in C_b^0.$$

2.2 Derivatives in the space of measures

We now define the derivative in the space $\mathcal{P}_2$. For this, we follow mostly the definition and notations introduced in [9] (in a slightly different context) and which are reminiscent of earlier approaches: see [3, 4] and the references in [11]. We say that a map $U : \mathcal{P}_2 \to \mathbb{R}$ is C^1 if there exists a *continuous and bounded* map $\frac{\delta U}{\delta m} : \mathcal{P}_2 \times \mathbb{R}^d \to \mathbb{R}$ such that

$$U(m') - U(m) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}((1-s)m + sm', y)(m' - m)(dy)ds \qquad \forall m, m' \in \mathcal{P}_2. \tag{8}$$

Note that the restriction on $\frac{\delta U}{\delta m}$ to be continuous on the entire space $\mathbb{R}^d$ and globally bounded is restrictive: it will however simplify our forthcoming construction. The map $\frac{\delta U}{\delta m}$ is defined only up to an additive constant that we fix with the convention

$$\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m, y) m(dy) = 0 \qquad \forall m \in \mathcal{P}_2. \tag{9}$$

We say that the map U is continuously $L-$differentiable (in short: $L-C^1$) if U is C^1 and if $y \to \frac{\delta U}{\delta m}(m, y)$ is everywhere differentiable with a continuous and globally bounded derivative on $\mathcal{P}_2 \times \mathbb{R}^d$. We denote by

$$D_m U(m, y) := D_y \frac{\delta U}{\delta m}(m, y)$$

this $L-$derivative. In view of the discussion in [9], $D_m U$ coincides with the Lions derivative as introduced in [27] and discussed in [11]. In particular, it estimates the Lipschitz regularity of U in $\mathcal{P}_2$ (Remark 5.27 in [11]):

$$|U(m) - U(m')| \leqslant \mathbf{d}_2(m, m') \sup_{\mu \in \mathcal{P}_2} \left(\int_{\mathbb{R}^d} |D_m U(\mu, y)|^2 \mu(dy) \right)^{1/2} \qquad \forall m, m' \in \mathcal{P}_2. \tag{10}$$

Of course one can also estimate the Lipschitz regularity of U through the $\mathbf{d}_1$ norm, as

$$|U(m) - U(m')| \leqslant \mathbf{d}_1(m, m') \sup_{\mu \in \mathcal{P}_2} \|D_m U(\mu, \cdot)\|_\infty \leqslant \mathbf{d}_2(m, m') \sup_{\mu \in \mathcal{P}_2} \|D_m U(\mu, \cdot)\|_\infty . \tag{11}$$

Note that, with our boundedness convention, if U is continuously $L-$differentiable, then U is automatically globally Lipschitz continuous.

When U is smooth enough, we often see the map $\frac{\delta U}{\delta m}$ as a linear map on C^{-k} by

$$\frac{\delta U}{\delta m}(m)(\rho) = \langle \rho, \frac{\delta U}{\delta m}(m,\cdot)\rangle_{C^{-k},C^k} \qquad \forall \rho \in C^{-k}.$$

We say that U is C^2 if $\frac{\delta U}{\delta m}$ is C^1 in m with a continuous and bounded derivative: namely $\frac{\delta^2 U}{\delta m^2} = \frac{\delta}{\delta m}(\frac{\delta U}{\delta m}) : \mathcal{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is continuous in all variables and bounded. We say that U is twice $L-$differentiable if the map $D_m U$ is $L-$differentiable with respect to m with a second order derivative $D^2_{mm}U = D^2_{mm}U(m,y,y')$ which is continuous and bounded on $\mathcal{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d$ with values in $\mathbb{R}^{d\times d}$.

When a map $U : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$ is of class C^n_b with respect to the space variable, uniformly with respect to the measure variable, we often set

$$\|U\|_n := \sup_{m\in\mathcal{P}_2} \|U(\cdot,m)\|_n. \tag{12}$$

We use similar notation for a map U depending on several space variables and on a measure.

When a map $U : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$ is Lipschitz continuous with respect to m, uniformly with respect to the space variable in some C^n norm, we define $\mathrm{Lip}_n(U)$ as the smallest constant C such that

$$\|U(\cdot,m_1) - U(\cdot,m_2)\|_n \leqslant C\mathbf{d}_2(m_1,m_2) \qquad \forall m,m' \in \mathcal{P}_2.$$

Namely:

$$\mathrm{Lip}_n(U) := \sup_{m_1\neq m_2} \frac{\|U(\cdot,m_1) - U(\cdot,m_2)\|_n}{\mathbf{d}_2(m_1,m_2)}.$$

More generally, if $U : (\mathbb{R}^d)^k \times \mathcal{P}_2 \to \mathbb{R}$ (for $k \in \mathbb{N}$, $k \geqslant 1$) is Lipschitz continuous in the measure variable in some $C^{n_1,\dots,n_k}_b$ norm (where $n_i \in \mathbb{N}$ for $i = 1,\dots,k$), then we set

$$\mathrm{Lip}_{n_1,\dots,n_k}(U) := \sup_{m_1\neq m_2} \frac{\|U(\cdot_{x_1},\dots,\cdot_{x_k},m_1) - U(\cdot_{x_1},\dots,\cdot_{x_k},m_2)\|_{n_1,\dots,n_k}}{\mathbf{d}_2(m_1,m_2)}.$$

We will typically use this notation for the derivatives of a map $U : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$: indeed we will often have to estimate quantities of the form

$$\mathrm{Lip}_{n_1,n_2}(D_m U) := \sup_{m_1\neq m_2} \frac{\|D_m U(\cdot_x,m_1,\cdot_y) - D_m U(\cdot_x,m_2,\cdot_y)\|_{n_1,n_2}}{\mathbf{d}_2(m_1,m_2)}$$

and

$$\mathrm{Lip}_{n_1,n_2,n_3}(D^2_{mm}U) := \sup_{m_1\neq m_2} \frac{\|D^2_{mm}U(\cdot_x,m_1,\cdot_y,\cdot_{y'}) - D^2_{mm}U(\cdot_x,m_2,\cdot_y,\cdot_{y'})\|_{n_1,n_2,n_3}}{\mathbf{d}_2(m_1,m_2)}.$$

Concerning the Lipschitz continuity with respect to one of the entries x_i, we will use the following notation:

$$\mathrm{Lip}^{x_i}_{n_1,\dots,n_{i-1},n_{i+1},\dots,n_k}(U) :=$$
$$\sup_{m,x_i^1\neq x_i^2} \frac{\|U(\cdot_{x_1},\dots,\cdot_{x_{i-1}},x_i^1,\cdot_{x_{i+1}},\dots,\cdot_{x_k},m) - U(\cdot_{x_1},\dots,\cdot_{x_{i-1}},x_i^2,\cdot_{x_{i+1}},\dots,\cdot_{x_k},m)\|_{n_1,\dots,n_{i-1},n_{i+1},\dots,n_k}}{|x_i^1 - x_i^2|}.$$

Further norms: In order to estimate the $y-$dependence of a derivative with respect to the measure of a map $U = U(x,m)$, we systematically proceed by duality method, testing this derivative against distributions. This yields to the following norms, for $n,k \in \mathbb{N}$ (note the the subtle difference in notation between $\|\cdot\|_{n,k}$ and $\|\cdot\|_{n;k}$):

$$\left\|\frac{\delta U}{\delta m}\right\|_{n;k} := \sup_{m\in\mathcal{P}_2} \sum_{r=0}^n \sup_{\substack{x\in\mathbb{R}^d,\rho\in C^0_c\\ \|\rho\|_{-k}=1}} \left(\sum_{|\alpha|=r} \left|\partial_x^\alpha \frac{\delta U}{\delta m}(x,m)(\rho)\right|^2\right)^{1/2} = \sup_{m\in\mathcal{P}_2} \sum_{r=0}^n \sup_{\substack{x\in\mathbb{R}^d,\rho\in C^0_c,\\ \|\rho\|_{-k}=1}} \left|D_x^r \frac{\delta U}{\delta m}(x,m)(\rho)\right|,$$

$$\left\|\frac{\delta^2 U}{\delta m^2}\right\|_{n;k,k'} := \sup_{m\in\mathcal{P}_2} \sum_{r=1}^n \sup_{\substack{x\in\mathbb{R}^d,\rho,\rho'\in C^0_c,\\ \|\rho\|_{-k}=\|\rho'\|_{-k'}=1}} \left|D_x^r \frac{\delta^2 U}{\delta m^2}(x,m)(\rho,\rho')\right|.$$

For maps $U = U(x_1, x_2, m)$ depending on two (or more) space variables, we use the transparent notation $\|\cdot\|_{n_1,n_2;k}$ (and, if $n_1 = 0$ (say), we simply set $\|\cdot\|_{n_2,k} = \|\cdot\|_{0,n_2;k}$). Finally, we use similar notation for the Lipschitz norms, setting, for instance for a map $U = U(x, m)$,

$$\mathrm{Lip}_{n;k,k'}\Big(\frac{\delta^2 U}{\delta m^2}\Big) := \sup_{m_1\neq m_2} \mathbf{d}_2(m_1,m_2)^{-1} \sum_{r=0}^{n} \sup_{\substack{x\in\mathbb{R}^d,\rho,\rho'\in C_c^0\\ \|\rho\|_{-k}=\|\rho'\|_{-k'}=1}} \Big| D_x^r \frac{\delta^2 U}{\delta m^2}(x,m_2)(\rho,\rho') - D_x^r \frac{\delta^2 U}{\delta m^2}(x,m_1)(\rho,\rho')\Big|.$$

Some comment about the norms we just introduced are now in order. We discuss the norm $\|\cdot\|_{n;k}$ to fix the ideas. With these notations, we have, if $U = U(x, m)$ is smooth enough,

$$\|\frac{\delta U}{\delta m}(\cdot, m)(\rho)\|_n \leqslant \|\frac{\delta U}{\delta m}\|_{n;k}\|\rho\|_{-k},$$

for every fixed $m \in \mathcal{P}_2$. Inequalities of this type are used throughout the text. Next we note that the norms $\sup_{m\in\mathcal{P}_2}\|\cdot\|_{n,k}$ and $\|\cdot\|_{n;k}$ are equivalent if we know a priori that $\delta U/\delta m = \delta U/\delta m(x, m, y)$ is in $C_b^{n,k}$. In general we do not have this information, but only know that $\delta U/\delta m$ is (at least) continuous. In this case, we use the following remark:

Lemma 2.1. *Let $k \in \mathbb{N}$ with $k \geqslant 1$ and $u \in C^0$ be such that*

$$\theta := \sup_{\rho\in C_c^0,\ \|\rho\|_{-k}=1} \int_{\mathbb{R}^d} u(y)\rho(y)dy < +\infty. \tag{13}$$

Then $u \in C_b^{k-1}$ with $\|u\|_{k-1} \leqslant C_k\theta$ (where C_k depends on d and k) and, for any $\beta \in \mathbb{N}^d$ with $|\beta| = k-1$, $\partial^\beta u$ is $\theta-$Lipschitz continuous.

Remark 2.2. *In particular, if $\frac{\delta U}{\delta m} \in C_b^{n,0}$ and $\|\frac{\delta U}{\delta m}\|_{n;k}$ is finite for some $n, k \in \mathbb{N}$ with $k \geqslant 1$, then $\frac{\delta U}{\delta m} \in C_b^{n,k-1}$ and*

$$\|\delta U/\delta m\|_{n,k-1} \leqslant C_{n,k}\|\delta U/\delta m\|_{n;k}$$

for some constant $C_{n,k}$ depending in addition on dimension only. Moreover, the derivatives of the form $\partial_x^\alpha\partial_y^\beta\frac{\delta U}{\delta m}$ for $|\alpha| \leqslant n$ and $\beta \leqslant k-1$ are Lipschitz continuous with respect to y and thus—by (11)—also with respect to m, with a Lipschitz constant bounded by $\|\frac{\delta U}{\delta m}\|_{n;k}$.

Proof. For $k = 1$, we have, approximating Dirac masses by continuous maps with compact support: for any $x, y \in \mathbb{R}^d$,

$$|u(x)| \leqslant \theta\|\delta_x\|_{-1} = \theta \qquad \text{and} \qquad |u(x) - u(y)| \leqslant \theta\|\delta_x - \delta_y\|_{-1} = \theta|x - y|.$$

This proves the claim for $k = 1$. Let now assume that (13) holds for $k = 2$. Then u can be extended to an element T in $(C^{-2})'$ with norm $\|T\| \leqslant \theta$, such that $T(\rho) = \int u d\rho$ for any Radon measure ρ. As, for any $v \in \mathbb{R}^d$,

$$\lim_{h\to 0, v'\to v} h^{-1}(\delta_{x+hv'} - \delta_x) = -\partial_v\delta_x \qquad \text{in } C^{-2},$$

we infer that

$$\lim_{h\to 0, v'\to v} h^{-1}(u(x + hv') - u(x)) = -T(\partial_v\delta_x).$$

The map $(x, v) \to \partial_v\delta_x$ being continuous in C^{-2} with $\|\partial_v\delta_x\|_{-2} \leqslant |v|$, u is in C^1 with $\|Du\| \leqslant \theta$. Then, arguing as for $k = 1$, one can easily check that Du is $\theta-$Lipschitz continuous. So the result also holds for $k = 2$. The proof can be completed in the same way for any k by induction. □

2.3 Assumptions on the data

We state here the assumptions needed on a, H and G for the existence of a classical solution to the second order master equation (1) and to the master equation (5) for the MFG problem with a major player. These assumptions are in force throughout the paper. Note that they are common to both problems (1) and (5) since both require the same kind of estimates on the first order master equation (see Section 6).

We assume that the map $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ can be written as $a = \sigma\sigma^T$ where $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ ($M \in \mathbb{N}, \quad M \geqslant 1$) is bounded in C_b^n with respect to the space variable, uniformly with respect to the time variable, for some $n \geqslant 4$. We also assume that the following uniform ellipticity condition holds:

$$a(t,x) \geqslant C_0^{-1} I_d, \qquad \|Da\|_\infty \leqslant C_0\,, \tag{14}$$

for some $C_0 > 0$.

We assume that the map $H : \mathbb{R}^{d_0} \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$ satisfies the growth condition

$$\sup_{x_0\in\mathbb{R}^{d_0},\ x\in\mathbb{R}^d,\ m\in\mathcal{P}_2} |D_x H(x_0,x,p,m)| \leqslant C_0(1+|p|^\gamma), \quad \forall p \in \mathbb{R}^d, \tag{15}$$

for some $\gamma > 1$. We also suppose that, for any $R > 0$, the quantities

$$\|H(\cdot_{x_0},\cdot_x,\cdot_p,m)\|_{3,n,n+1},\ \left\|\frac{\delta H}{\delta m}(\cdot_{x_0},\cdot_x,\cdot_p,m,\cdot_y)\right\|_{2,n-1,n,k},\ \left\|\frac{\delta^2 H}{\delta m^2}(\cdot_{x_0},\cdot_x,\cdot_p,m,\cdot_y,\cdot_{y'})\right\|_{1,n-2,n-1,k-1,k-1},$$

and $\mathrm{Lip}_{1,n-3,n-2,k-1,k-1}(\frac{\delta^2 H}{\delta m^2})$ are bounded for $|p| \leqslant R$, $m \in \mathcal{P}_2$ and $x_0 \in \mathbb{R}^{d_0}$, for any $k \in \{2,\dots,n-1\}$. Note that we could also allow for a time dependence for H without changing at all the arguments: we will not do so to simplify a little the notation. For the second order master equation, the Hamiltonian H actually does not depend on x_0, but this dependence is important to handle the MFG problem with a major player.

As for the initial condition $G : \mathbb{R}^{d_0} \times \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$, we assume that G is of class C^2 with respect to all variables and that the quantities

$$\|G(\cdot_{x_0},\cdot_x,m)\|_{3,n},\ \left\|\frac{\delta G}{\delta m}(\cdot_{x_0},\cdot_x,m,\cdot_y)\right\|_{2,n-1,k},\ \left\|\frac{\delta^2 G}{\delta m^2}(\cdot_{x_0},\cdot_x,m,\cdot_y,\cdot_{y'})\right\|_{1,n-2,k-1,k-1},$$
$$\mathrm{Lip}_{1,n-3,k-2,k-2}(\frac{\delta^2 G}{\delta m^2})(\cdot_{x_0},\cdot_x,m,\cdot_y,\cdot_{y'}),$$

are bounded uniformly with respect to $m \in \mathcal{P}_2$. Here again, for the second order master equation, the terminal condition G does not depend on x_0, but this dependence is needed in the MFG problem with a major player.

Additional assumptions for the MFG problem with a major player. This problem involves in addition a Hamiltonian $H^0 : \mathbb{R}^{d_0} \times \mathbb{R}^{d_0} \times \mathcal{P}_2 \to \mathbb{R}$ and a terminal condition $G^0 : \mathbb{R}^{d_0} \times \mathcal{P}_2 \to \mathbb{R}$. We assume that the map H^0 satisfies the growth property

$$\sup_{x_0\in\mathbb{R}^{d_0},\ m\in\mathcal{P}_2} |D_{x_0,p} H^0(x_0,p,m)| + |D^2_{x_0,p} H^0(x_0,p,m)| \leqslant C_0(|p|^\gamma + 1), \tag{16}$$

for some $\gamma > 1$. We also suppose that, for any $R > 0$, the quantities

$$\|H^0(\cdot_{x_0},\cdot_p,m)\|_{3,4},\ \left\|\frac{\delta H^0}{\delta m}(\cdot_{x_0},\cdot_p,m,\cdot_y)\right\|_{2,3,k},\ \left\|\frac{\delta^2 H^0}{\delta m^2}(\cdot_{x_0},\cdot_p,m,\cdot_y,\cdot_{y'})\right\|_{1,2,k-1,k-1},$$

and $\mathrm{Lip}_{0,1,k-2,k-2}(\frac{\delta^2 H}{\delta m^2})$ are bounded for $|p| \leqslant R$, $m \in \mathcal{P}_2$ and $x_0 \in \mathbb{R}^{d_0}$, for any $k \in \{2,\dots,n-1\}$.

The initial condition $G^0 : \mathbb{R}^{d_0} \times \mathcal{P}_2 \to \mathbb{R}$ is assumed to be of class C^2 with respect to the measure variable and the quantities

$$\|G^0(\cdot,m)\|_3,\ \left\|\frac{\delta G^0}{\delta m}(\cdot,m,\cdot)\right\|_{2,k},\ \left\|\frac{\delta^2 G^0}{\delta m^2}(\cdot_{x_0},m,\cdot_y,\cdot_{y'})\right\|_{1,k-1,k-1},$$
$$\mathrm{Lip}_{0,k-2,k-2}(\frac{\delta^2 G^0}{\delta m^2})(\cdot_{x_0},m,\cdot_y,\cdot_{y'}),$$

are supposed to be bounded uniformly with respect to $m \in \mathcal{P}_2$.

Throughout the proofs, we assume that the time horizon T is small, say $T \leqslant 1$. We denote by C and C_M a constant which might change from line to line and which depends only on the data of the problem, i.e., on a, H and H^0—the dependence in G and G^0 being always explicitly written—and, for C_M, on the additional real number M. In some proofs, when there is no ambiguity, we drop the M dependence of C_M to simplify the expressions.

3 The second order master equation

The aim of the section is to show the short-time existence of the second order master equation:

$$\begin{cases} -\partial_t U(t,x,m) - \mathrm{Tr}((a(t,x)+a^0)D^2_{xx}U(t,x,m)) + H(x, D_xU(t,x,m), m) \\ \quad - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}((a(t,y)+a^0)D^2_{ym}U(t,x,m,y))\, dm(y) \\ \quad + \displaystyle\int_{\mathbb{R}^d} D_mU(t,x,m,y)\cdot H_p(y, D_xU(t,y,m), m)\, dm(y) \\ \quad -2\displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{xm}U(t,x,m,y)\right] m(dy) \\ \quad - \displaystyle\int_{\mathbb{R}^{2d}} \mathrm{Tr}[a^0 D^2_{mm}U(t,x,m,y,y')]m(dy)m(dy') = 0 \\ \qquad\qquad \text{in } (0,T)\times\mathbb{R}^d\times\mathcal{P}_2\,, \\ U(T,x,m) = G(x,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2\,, \end{cases} \tag{17}$$

where a^0 is a symmetric positive definite $d\times d$ matrix (independent of time and space).

Definition 3.1. *We say that $U : [0,T]\times\mathbb{R}^d\times\mathcal{P}_2\to\mathbb{R}$ is a classical solution of* (17) *if U and its derivatives involved in* (17) *exist, are continuous in all variables and are bounded, and if* (17) *holds.*

Our main result is the following short time existence Theorem:

Theorem 3.2. *Under the assumptions of Subsection 2.3, there exists a time $T>0$ such that the second order master equation* (17) *has a classical solution U on $[0,T]$.*

We shall not prove here the uniqueness of the solution to (17), which holds under our assumptions: this point has been often discussed in the literature (see [9, 11] for instance). The reader may notice that we cannot handle a second order master equation with a space dependent matrix $a^0 = a^0(t,x)$. The reason is that we do not know how to extend the estimate in Proposition 3.3 to the space dependent case.

The proof of Theorem 3.2 is given at the end of the section, after a few preliminary steps. The key idea is to use a Trotter-Kato scheme alternating the first order master equation as in (6) and a linear second order master equation. The analysis of the first order master equation, being quite technical, is postponed to Section 6 below. We now discuss the linear second order master equation.

3.1 The linear second order master equation

In this section we consider the (forward) second order linear master equation

$$\begin{cases} \partial_t U(t,x,m) - \mathrm{Tr}\left[a^0 D^2_{xx}U(t,x,m)\right] - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{ym}U(t,x,m,y)\right] m(dy) \\ \quad -2\displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{xm}U(t,x,m,y)\right] m(dy) \\ \quad - \displaystyle\int_{\mathbb{R}^{2d}} \mathrm{Tr}[a^0 D^2_{mm}U(t,x,m,y,y')]m(dy)m(dy') = 0 \\ \qquad\qquad \text{in } (0,T)\times\mathbb{R}^d\times\mathcal{P}_2\,, \\ U(0,x,m) = G(x,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2\,. \end{cases} \tag{18}$$

Let Γ be the fundamental solution of the equation associated with a^0:

$$\begin{cases} \partial_t\Gamma(t,x) - \mathrm{Tr}\left[a^0 D^2_{xx}\Gamma(t,x)\right] = 0 \qquad \text{in } (0,+\infty)\times\mathbb{R}^d, \\ \Gamma(0,x) = \delta_0(x) \qquad \text{in } \mathbb{R}^d, \end{cases}$$

and, given a map $G:\mathbb{R}^d\times\mathcal{P}_2\to\mathbb{R}$ of class C^2 in (x,m), let us set

$$U(t,x,m) = \int_{\mathbb{R}^d} G(\xi, (id - x + \xi)\sharp m)\Gamma(t,x-\xi)d\xi \qquad \forall (t,x,m)\in[0,T]\times\mathbb{R}^d\times\mathcal{P}_2.$$

Proposition 3.3. *The map U is a classical solution to the second order equation* (18)*. Moreover, there exists a constant $C > 0$ (depending only on n, k and a^0), such that*

$$\sup_{t\in[0,T]} \|U(t)\|_n \leqslant (1+CT) \sup_{m\in\mathcal{P}_2} \|G\|_n$$

and, for $k \in \{2,\dots,n-1\}$,

$$\sup_{t\in[0,T]} \left\| \frac{\delta U}{\delta m}(t) \right\|_{n-1;k} \leqslant (1+CT) \left\| \frac{\delta G}{\delta m} \right\|_{n-1;k},$$

$$\sup_{t\in[0,T]} \left\| \frac{\delta^2 U}{\delta m^2}(t) \right\|_{n-2;k-1,k-1} \leqslant (1+CT) \left\| \frac{\delta^2 G}{\delta m^2} \right\|_{n-2;k-1,k-1}$$

and

$$\sup_{t\in[0,T]} \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 U}{\delta m^2}(t)) \leqslant (1+CT)\mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 G}{\delta m^2}).$$

Remark 3.4. *If we assume that, for some constant M,*

$$\|G\|_n + \left\| \frac{\delta G}{\delta m} \right\|_{n-1;k} + \left\| \frac{\delta^2 G}{\delta m^2} \right\|_{n-2;k-1,k-1} + \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 G}{\delta m^2}) \leqslant M,$$

then the above estimates can be rewritten in the form:

$$\begin{aligned}\sup_{t\in[0,T]} &(\|U(t)\|_n + \left\| \frac{\delta U}{\delta m}(t) \right\|_{n-1;k} + \left\| \frac{\delta^2 U}{\delta m^2}(t) \right\|_{n-2;k-1,k-1} + \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 U}{\delta m^2}(t))) \\ &\leqslant \|G\|_n + \left\| \frac{\delta G}{\delta m} \right\|_{n-1;k} + \left\| \frac{\delta^2 G}{\delta m^2} \right\|_{n-2;k-1,k-1} + \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 G}{\delta m^2}) + C_M T,\end{aligned}$$

for some constant C_M depending on n, k, a^0 and M.

In order to prove this Proposition, we need two Lemmas, the proof of which are easy and left to the reader.

Lemma 3.5. *Let $U : \mathcal{P}_2 \to \mathbb{R}$ be $L-C^1$ and let $\phi : \mathbb{R}^d \to \mathbb{R}^d$ be of class C^1 with bounded derivative. Let us set $V(m) = U(\phi\sharp m)$. Then V is $L-C^1$ with*

$$D_m V(m,y) = (D\phi(y))^T D_m U(\phi\sharp m, \phi(y)).$$

Lemma 3.6. *Let $U : \mathcal{P}_2 \to \mathbb{R}$ be $L-C^1$ and let $V(x,m) = U((id+x)\sharp m)$. Then V is of class C^1 with*

$$D_x V(x,m) = \int_{\mathbb{R}^d} D_m U((id+x)\sharp m, x+y) dm(y).$$

Proof of Proposition 3.3. Let us first note that

$$U(t,x,m) = \int_{\mathbb{R}^d} G(\xi, (id-x+\xi)\sharp m)\Gamma(t,x-\xi)d\xi = \int_{\mathbb{R}^d} G(x-z,(id-z)\sharp m)\Gamma(t,z)dz.$$

In particular, U is C^1 in t, C^2 in x and has second order derivatives which are C^2 in the space variables with, in view of Lemma 3.5 and Lemma 3.6,

$$D_x U(t,x,m) = \int_{\mathbb{R}^d} D_x G(x-y,(id-y)\sharp m)\Gamma(t,y)dy,$$

$$D_x^2 U(t,x,m) = \int_{\mathbb{R}^d} D_x^2 G(x-y,(id-y)\sharp m)\Gamma(t,y)dy,$$

$$D_m U(t,x,m,y) = \int_{\mathbb{R}^d} D_m G(x-z,(id-z)\sharp m, y-z)\Gamma(t,z)dz,$$

and

$$D_m^2 U(t,x,m,y,y') = \int_{\mathbb{R}^d} D_m^2 G(x-z,(id-z)\sharp m, y-z, y'-z)\Gamma(t,z)dz.$$

This easily implies the estimates on U and its derivatives.

On the other hand, since $(id-w)\sharp[(id-z)\sharp m] = (id-z-w)\sharp m$, we have, for any $t \in (0,T)$ and $h \in (0, T-t)$,

$$\begin{aligned}
&\int_{\mathbb{R}^d} U(t,x-z,(id-z)\sharp m)\Gamma(h,z)dz \\
&\quad = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} G(x-z-w,(id-z-w)\sharp m)\Gamma(h,z)\Gamma(t,w)dwdz \\
&\quad = \int_{\mathbb{R}^d} G(x-u,(id-u)\sharp m)\left(\int_{\mathbb{R}^d}\Gamma(h,u-w)\Gamma(t,w)dw\right)du \\
&\quad = \int_{\mathbb{R}^d} G(x-u,(id-u)\sharp m)\Gamma(t+h,u)du = U(t+h,x,m).
\end{aligned}$$

So, taking the derivative with respect to $h > 0$ in the above expression:

$$\partial_t U(t+h,x,m) = \int_{\mathbb{R}^d} U(t,x-z,(id-z)\sharp m)\partial_t\Gamma(h,z)dz.$$

Integrating by parts and using Lemma 3.5 and Lemma 3.6:

$$\begin{aligned}
\partial_t U(t+h,x,m) &= \int_{\mathbb{R}^d} U(t,x-z,(id-z)\sharp m)\Big(\mathrm{Tr}\left[a^0 D_{zz}^2\Gamma(h,z)\right]\Big)dz \\
&= \int_{\mathbb{R}^d}\Big(\mathrm{Tr}\left[a^0 D_{xx}^2 U(t,x-z,(id-z)\sharp m)\right] + 2\int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{xm}^2 U(t,x-z,(id-z)\sharp m, y-z)\right]m(dy) \\
&\quad + \int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{ym}^2 U(t,x-z,(id-z)\sharp m, y-z)\right]m(dy) \\
&\quad + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{mm}^2 U(t,x-z,(id-z)\sharp m, y-z, y'-z)\right]m(dy)m(dy')\Big)\Gamma(h,z))dz.
\end{aligned}$$

Letting $h \to 0$ we obtain

$$\begin{aligned}
\partial_t U(t,x,m) &= \mathrm{Tr}\left[a^0 D_{xx}^2 U(t,x,m)\right] + 2\int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{xm}^2 U(t,x,m,y)\right]m(dy) \\
&\quad + \int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{ym}^2 U(t,x,m,y)\right]m(dy) + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathrm{Tr}\left[a^0 D_{mm}^2 U(t,x,m,y,y')\right]m(dy)m(dy').
\end{aligned}$$

So U is a solution to (18). □

3.2 Existence of a solution

3.2.1 Definition of the semi-discrete scheme

Let us fix some horizon $T > 0$ (small) and a step-size $\tau := T/(2N)$ (where $N \in \mathbb{N}$, $N \geqslant 1$). We set $t_k = kT/(2N)$, $k \in \{0, 2N\}$. We define by backward induction a continuous map $U^N = U^N(t,x,m)$, with $U^N : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}$, as follows: we require that

(i) U^N satisfies the terminal condition

$$U^N(T,x,m) = G(x,m) \qquad \forall (x,m) \in \mathbb{R}^d \times \mathcal{P}_2,$$

(ii) U^N solves the backward linear second order master equation

$$\begin{aligned}-\partial_t U^N - 2\mathrm{Tr}\left[a^0 D^2_{xx} U^N\right] - 2\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{ym} U^N\right] m(dy) \\ -4\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{xm} U^N\right] m(dy) - 2\int_{\mathbb{R}^{2d}} \mathrm{Tr}[a^0 D^2_{mm} U^N] m(dy) m(dy') = 0\end{aligned} \tag{19}$$

on time intervals of the form (t_{2j+1}, t_{2j+2}), for $j = 0, \dots, N-1$,

(iii) U^N solves the first order master equation

$$\begin{aligned}-\partial_t U^N - 2\mathrm{Tr}(aD^2_{xx} U^N) + 2H(x, D_x U^N, m) - 2\int_{\mathbb{R}^d} \mathrm{Tr}(aD^2_{ym} U^N)\, dm(y) \\ +2\int_{\mathbb{R}^d} D_m U^N \cdot H_p(y, D_x U^N, m)\, dm(y) = 0\end{aligned} \tag{20}$$

on time intervals of the form (t_{2j}, t_{2j+1}), for $j = 0, \dots, N-1$.

Our aim is to show that, if the time horizon is short enough, U^N converges to a solution of the second order master equation as $N \to +\infty$.

3.2.2 Estimates of U^N

For $n \geqslant 4$ and $k \in \{3, \dots, n-1\}$, let

$$M := \|G\|_n + \left\|\frac{\delta G}{\delta m}\right\|_{n-1;k} + \left\|\frac{\delta^2 G}{\delta m^2}\right\|_{n-2;k-1,k-1} + \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 G}{\delta m^2}) + 1. \tag{21}$$

Lemma 3.7. *There exists $T_M > 0$ such that, for any $T \in (0, T_M]$ and $N \geqslant 1$, we have*

$$\sup_{t\in[0,T]} (\|U^N(t)\|_n + \|\frac{\delta U^N}{\delta m}(t)\|_{n-1;k} + \left\|\frac{\delta^2 U^N}{\delta m^2}(t)\right\|_{n-2;k-1,k-1} + \mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2 U^N}{\delta m^2}(t)\right)) \leqslant M.$$

Moreover:

- *The maps U^N, $D_x U^N$, $D^2_{xx} U^N$ are globally Lipschitz continuous in (t, x, m), uniformly with respect to N.*
- *The maps $D_m U$, $D_m D_x U^N$, $D_y D_m U^N$ are Holder continuous in (t, x, m, y), uniformly with respect to N, in any set of the form*

$$\{(t, x, m, y) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d,\ M_2(m) \leqslant R,\ |y| \leqslant R\}, \tag{22}$$

  (where $M_2(m) = (\int_{\mathbb{R}^d} |y|^2 m(dy))^{1/2}$).
- *The map $D^2_m U^N$ is Holder continuous in (t, x, m, y, y'), uniformly with respect to N, in any set of the form*

$$\{(t, x, m, y, y') \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d,\ M_2(m) \leqslant R,\ |y|, |y'| \leqslant R\}. \tag{23}$$

Proof. In order to prove the estimate, we use Proposition 3.3 as well as Propositions 6.1, 6.6, 6.8 (in Section 6 below). Let T_M be the smallest positive constant associated with these Propositions. Let also C_M be the largest constant in Propositions 3.3, 6.1, 6.6 and 6.8. We assume without loss of generality that $T_M < 1/(2C_M)$ and we fix $T \in (0, T_M]$.

We define the sequence $(\theta_k)_{k=0,\dots,2N}$ by

$$\theta_{2j} = M - 1 + C_M \frac{T}{N}(N - j), \qquad j = 0, \dots, N.$$

As $T_M \leqslant 1/(2C_M)$, we have $\theta_{2j} \leqslant M$ for any $T \in (0, T_M]$ and $N \geqslant 1$.

Now, using Propositions 6.1, 6.6, 6.8 and 3.3 one checks by backward induction that

$$\sup_{t\in[t_{2j},t_{2j+2}]}\left\{\|U^N(t)\|_n+\|\frac{\delta U^N}{\delta m}(t)\|_{n-1;k}+\left\|\frac{\delta^2U^N}{\delta m^2}(t)\right\|_{n-2;k-1,k-1}\right.$$
$$\left.+\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2U^N}{\delta m^2}(t)\right)\right\}\leqslant\theta_{2j}\leqslant M\qquad\forall j=0,\dots,N-1. \tag{24}$$

Indeed, assume that this is true for $j+1$; Proposition 3.3 (see also Remark 3.4), applied in the interval $[t_{2j+1},t_{2j+2}]$ and with the terminal condition $U^N(t_{2j+2},\cdot,\cdot)$ which satisfies (24) by assumption, implies that

$$\sup_{t\in[t_{2j+1},t_{2j+2}]}\left\{\|U^N(t)\|_n+\|\frac{\delta U^N}{\delta m}(t)\|_{n-1;k}+\left\|\frac{\delta^2U^N}{\delta m^2}(t)\right\|_{n-2;k-1,k-1}\right.$$
$$\left.+\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2U^N}{\delta m^2}(t)\right)\right\}\leqslant\theta_{2j+2}+\frac{C_MT}{2N}\,.$$

Then using Propositions 6.1, 6.6, 6.8 for the interval $[t_{2j},t_{2j+1}]$ and the terminal condition $U^N(t_{2j+1},\cdot,\cdot)$ for which (24) now holds, one gets:

$$\sup_{t\in[t_{2j},t_{2j+1}]}\left\{\|U^N(t)\|_n+\|\frac{\delta U^N}{\delta m}(t)\|_{n-1;k}+\left\|\frac{\delta^2U^N}{\delta m^2}(t)\right\|_{n-2;k-1,k-1}\right.$$
$$\left.+\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2U^N}{\delta m^2}(t)\right)\right\}\leqslant\theta_{2j+2}+\frac{C_MT}{N}=\theta_{2j}\,,$$

so (24) holds for j. Since the first step ($j=N-1$) can be proved similarly using the very definition of M in (21), we can conclude that (24) holds for every $j=0,\dots,N-1$.

We now prove the second part of the Lemma. As U^N solves (19) on the time intervals (t_{2j+1},t_{2j+2}) and (20) on time intervals (t_{2j},t_{2j+1}), we obtain directly, by the space estimates proved above:

$$\sup_{t,m}\|\partial_tU(t,\cdot,m)\|_{n-2}\leqslant C_M, \tag{25}$$

where C_M does not depend on N.

Let now $l\in\mathbb{N}^d$ with $|l|\leqslant 2$. By (25) and the fact that $\|U^N\|_n$ is bounded for $n>|l|$, D^lU^N is uniformly Lipschitz continuous in t and x. Moreover, since $\|\delta U^N/\delta m\|_{n-1;k}$ is bounded (for $k\geqslant 1$), D^lU^N is uniformly Lipschitz continuous in m as well by Remark 2.2 since $|l|\leqslant n-1$.

Next we prove the uniform continuity of $D_x^lD_y^rD_mU^N$ for $|l|,|r|\leqslant 1$. First we recall that $\|\delta U^N/\delta m\|_{n-1;k}$ is bounded, so that $\|D_mU^N\|_{n-1;k-1}$ is bounded, with $n-1\geqslant 2$ and $k-1\geqslant 2$. Therefore $D_x^lD_y^rD_mU^N$ is uniformly Lipschitz continuous in (x,y) (for y, this is Remark 2.2). Second, recall that $\|\delta^2U^N/\delta m^2\|_{n-2;k-1,k-1}$ is bounded, so that $\|\frac{\delta}{\delta m}D_mU^N\|_{n-2;k-2,k-1}$ is bounded as well, with $n\geqslant 3$ and $k\geqslant 3$: therefore $D_x^lD_y^rD_mU^N$ is uniformly Lipschitz continuous in m. As we have already proved that U^N is uniformly Lipschitz continuous in t, we can deduce from Lemma B.4 below applied to U^N that D_mU^N is also Holder continuous in time in any set of the form (22).

Finally we consider $D^2_{mm}U^N=D^2_{mm}U^N(t,x,m,y,y')$. Since $\|\delta^2U^N/\delta m\|_{n-2;k-1,k-1}$ and $\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2U^N}{\delta m^2}\right)$ are bounded, with $n\geqslant 4$ and $k\geqslant 3$, $D^2_{mm}U^N$ is uniformly Lipschitz continuous in (x,m,y,y'). Applying Lemma B.4 to the map D_mU^N, which is Holder continuous in time in sets of the form (22) (as we have seen above) and such that $D^2_{mm}U$ is uniformly Lipschitz in (m,y,y'), we deduce that $D^2_{mm}U^N$ is also Holder continuous in time, uniformly in N, in sets of the form (23). So we conclude that $D^2_{mm}U^N$ is uniformly Holder continuous in all variables. □

3.2.3 Proof of Theorem 3.2

Proof of Theorem 3.2. In view of Lemma 3.7, the maps U^N, D_xU^N, $D^2_{xx}U^N$, D_mU^N, $D_mD_xU^N$, $D_yD_mU^N$ and $D^2_mU^N$ are locally Holder continuous in all variables, uniformly with respect to N. So, by a version of Arzela-Ascoli Theorem (see Lemma B.5 below), there is a subsequence denoted in the same way such

that U^N, D_xU^N, $D^2_{xx}U^N$, D_mU^N, $D_mD_xU^N$, $D_yD_mU^N$ and $D^2_mU^N$ converge pointwisely in m and locally uniformly in time-space to some maps U, D_xU, $D^2_{xx}U$, V, D_xV, D_yV and W. Moreover, using the integral formula (8), is easy to check that $V = D_mU$ and $W = D^2_mU$.

By the equation satisfied by U^N, we have, for any $0 \leqslant s < t \leqslant T$,

$$
\begin{aligned}
&U^N(t,x,m) - U^N(s,x,m) \\
&\quad = -\sum_{k=0}^{N-1} \int_{t_{2k+1}}^{t_{2k+2}} 2\Big\{ \mathrm{Tr}\left[a^0 D^2_{xx} U^N\right] + \int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{ym} U^N\right] m(dy) \\
&\qquad + 2\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{xm} U^N\right] m(dy) + \int_{\mathbb{R}^{2d}} \mathrm{Tr}\left[a^0 D^2_{mm} U^N\right] m(dy)m(dy') \Big\} \mathbf{1}_{[s,t]}(\tau) d\tau \\
&\qquad - \sum_{k=0}^{N-1} \int_{t_{2k}}^{t_{2k+1}} 2\Big\{ \mathrm{Tr}(aD^2_{xx}U^N) - H(x, D_xU^N, m) \\
&\qquad\quad + \int_{\mathbb{R}^d} \mathrm{Tr}(aD^2_{ym}U^N)\ m(dy) - \int_{\mathbb{R}^d} D_mU^N \cdot H_p(y, D_xU^N, m)\ m(dy) \Big\} \mathbf{1}_{[s,t]}(\tau) d\tau.
\end{aligned}
$$

Since, as N tends to infinity, the maps

$$
t \to \sum_{k=0}^{N-1} \mathbf{1}_{[t_{2k+1}, t_{2k+2}]}(t) \qquad \text{and} \qquad t \to \sum_{k=0}^{N-1} \mathbf{1}_{[t_{2k}, t_{2k+1}]}(t)
$$

weakly converge to the constant $1/2$ and since the space integrals in the above equation converge pointwisely to the corresponding quantities for the limit U, we obtain by the dominated convergence Theorem:

$$
\begin{aligned}
&U(t,x,m) - U(s,x,m) \\
&\quad = -\int_s^t \Big(\mathrm{Tr}\left[a^0 D^2_{xx} U\right] + \int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{ym} U\right] dm \\
&\qquad + 2\int_{\mathbb{R}^d} \mathrm{Tr}\left[a^0 D^2_{xm} U\right] dm + \int_{\mathbb{R}^{2d}} \mathrm{Tr}\left[a^0 D^2_{mm} U\right] dm \otimes dm \\
&\qquad + \mathrm{Tr}(aD^2_{xx}U) - H(x, D_xU, m) \\
&\qquad + \int_{\mathbb{R}^d} \mathrm{Tr}(aD^2_{ym}U)\ dm - \int_{\mathbb{R}^d} D_mU \cdot H_p(y, D_xU, m)\ dm \Big) d\tau,
\end{aligned}
$$

so that U is a classical solution to (17). □

3.3 Existence of the solution to the stochastic MFG system

An easy consequence of the existence of a solution to the master equation is the well-posedness of the stochastic MFG system:

$$
\left\{
\begin{array}{ll}
du(t,x) = \big[-\mathrm{Tr}((a+a^0)(t,x)D^2u(t,x)) + H(x, Du(t,x), m(t)) & \\
\qquad\qquad -\sqrt{2}\mathrm{Tr}(\sigma^0 Dv(t,x)) \big] dt + v(t,x)\cdot dW_t & \text{in } (0,T)\times\mathbb{R}^d, \\
dm(t,x) = \big[\sum_{i,j} D_{ij}(((a_{ij}) + a^0_{ij})(t,x)m(t,x)) + \mathrm{div}\big(m(,x)H_p(x, Du(t,x), m(t))\big) \big] dt & \\
\qquad\qquad -\mathrm{div}(m(t,x)\sqrt{2}\sigma^0 dW_t\big), & \text{in } (0,T)\times\mathbb{R}^d, \\
u(T,x) = G(x, m(T)),\ m(0) = m_0, & \text{in } \mathbb{R}^d
\end{array}
\right. \tag{26}
$$

We say that (u, m, v) is a classical solution to (26) if u, m and v are random with values in $C^0([0,T], C^2_b)$, $C^0([0,T], \mathcal{P}_2)$ and $C^0([0,T], C^1_b(\mathbb{R}^d, \mathbb{R}^d))$ respectively and adapted to the filtration generated by W and if the backward HJ equation is satisfied in a classical sense:

$$
\begin{aligned}
u(t,x) = G(x, m(T)) - \int_t^T &\big(-\mathrm{Tr}((a+a^0)(s,x)D^2u(s,x)) + H(x, Du(s,x), m(s)) \\
&- \sqrt{2}\mathrm{Tr}(\sigma^0 Dv(s,x)) \big) ds - \int_t^T v(s,x)\cdot dW_s
\end{aligned}
$$

while the Fokker-Planck equation is satisfied in the sense of distributions: for any $\phi \in C_c^\infty([0,T)\times\mathbb{R}^d)$,

$$\begin{aligned}0 = &\int_{\mathbb{R}^d}\phi(0,x)m_0(dx)\\ &+\int_0^T\int_{\mathbb{R}^d}\Big(\mathrm{Tr}((a+a^0)(s,x)D^2\phi(s,x) - D\phi(s,x)\cdot H_p(x,Du(s,x),m(s))\Big)m(s,dx)ds\\ &+\sqrt{2}\int_0^T\int_{\mathbb{R}^d}((\sigma^0)^T D\phi(s,x)m(s,dx)\cdot dW_s.\end{aligned}$$

Theorem 3.8. *Under the assumptions of Theorem 3.2, there exists a time $T>0$ for which the stochastic MFG system* (26) *has a classical solution (u,m,v) in $[0,T]$. Moreover,*

$$v(t,x) = \sqrt{2}\int_{\mathbb{R}^d}(\sigma^0)^T D_m U(t,x,m(t),y)m(t,dy), \tag{27}$$

where U is the solution to the second order master equation (17)*.*

Proof. Let m be the solution to the stochastic McKean-Vlasov equation:

$$\begin{cases} dm(t,x) = \big[\sum_{i,j} D_{ij}((a_{i,j}+a^0_{i,j})(t,x)m(t,x)) + \mathrm{div}\big(m(t,x)H_p(x,DU(t,x,m(t)),m(t))\big)\big]dt \\ \qquad\qquad -\mathrm{div}(m(t,x)\sqrt{2}\sigma^0 dW_t\big), \qquad\qquad \text{in } (0,T)\times\mathbb{R}^d, \\ m(0,dx) = m_0, \qquad \text{in } \mathbb{R}^d \end{cases} \tag{28}$$

Existence of a solution for this system can be obtained, for instance, as the mean field limit of the SDE

$$\begin{cases} dX_s^{N,i} = -H_p(X_s^{N,i}, D_xU(t,X_s^{N,i},m^N_{\mathbf{X}_s^N}), m^N_{\mathbf{X}_s^N})ds + \sqrt{2}\sigma(s,X_s^{N,i})dB_s^i + \sqrt{2}\sigma^0(s,X_s^{N,i})dW_s \\ X_0^{N,i} = \bar{X}_0^{N,i} \end{cases}$$

where $\bar{X}_0^{N,i}$ is a family of i.i.d. r.v. of law m_0 and where $m^N_{\mathbf{X}_s^N} = \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{X_s^{N,i}}$. Indeed, one can show that the family of laws of $(m^N_{\mathbf{X}_s^N})$ is tight in $C^0([0,T],\mathcal{P}_2)$ and that its limit is a solution to (28). Uniqueness for (28) comes from the regularity of U and Gronwall's Lemma.

Then one can use the Itô's formula in [9, Theorem A.1] (see also [11, Theorem 11.13]) to derive that $u(t,x) := U(t,x,m(t))$ solves the backward stochastic HJ equation

$$\begin{cases} du(t,x) = \big[-\mathrm{Tr}((a+a^0)(t,x)D^2u(t,x)) + H(x,Du(t,x),m(t)) \\ \qquad\qquad -\sqrt{2}\mathrm{Tr}(\sigma^0 Dv(t,x))\big]dt + v(t,x)\cdot dW_t \qquad\qquad \text{in } (0,T)\times\mathbb{R}^d, \\ u(T,x) = G(x,m(T)) \qquad \text{in } \mathbb{R}^d \end{cases}$$

where v is given by (27). Note that, by the regularity of U, u and v have the required regularity. □

4 The master equation for MFGs with a major player

In this section we investigate the well-posedness of the master equation associated with the MFG problem with a major player. The unknown (U^0, U) solves the system of master equations:

$$
\left\{
\begin{array}{ll}
(i) & -\partial_t U^0(t,x_0,m) - \Delta_{x_0} U^0(t,x_0,m) + H^0(x_0, D_{x_0}U^0(t,x_0,m),m) \\
& \quad - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U^0(t,x_0,m,y) dm(y) \\
& \quad + \displaystyle\int_{\mathbb{R}^d} D_m U^0(t,x_0,m,y) \cdot H_p(x_0,y,D_x U(t,x_0,y,m),m) dm(y) = 0 \\
& \qquad\qquad\qquad\qquad\qquad\qquad \text{in } (0,T)\times\mathbb{R}^{d_0}\times\mathcal{P}_2, \\
(ii) & -\partial_t U(t,x_0,x,m) - \Delta_x U(t,x_0,x,m) - \Delta_{x_0} U(t,x_0,x,m) \\
& \quad + H(x_0,x,D_x U(t,x_0,x,m),m) \\
& \quad - \displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U(t,x_0,x,m,y) dm(y) + D_{x_0}U \cdot H_p^0(x_0, D_{x_0}U^0(t,x_0,m),m) \\
& \quad + \displaystyle\int_{\mathbb{R}^d} D_m U(t,x_0,x,m,y) \cdot H_p(x_0,y,D_x U(t,x_0,y,m),m) dm(y) = 0 \\
& \qquad\qquad\qquad\qquad\qquad\qquad \text{in } (0,T)\times\mathbb{R}^{d_0}\times\mathbb{R}^d\times\mathcal{P}_2, \\
(iii) & U^0(T,x_0,m) = G^0(x_0,m), \qquad \text{in } \mathbb{R}^{d_0}\times\mathcal{P}_2, \\
(iv) & U(T,x_0,x,m) = G(x_0,x,m) \qquad \text{in } \mathbb{R}^{d_0}\times\mathbb{R}^d\times\mathcal{P}_2.
\end{array}
\right. \tag{29}
$$

Definition 4.1. *Let $U^0 : [0,T]\times\mathbb{R}^{d_0}\times\mathcal{P}_2 \to \mathbb{R}$ and $U : [0,T]\times\mathbb{R}^{d_0}\times\mathbb{R}^d\times\mathcal{P}_2 \to \mathbb{R}$. We say that (U^0, U) is a classical solution of (29) if U^0 and U and their derivatives involved in (29) exist, are continuous in all variables and are bounded, and if (29) holds.*

Throughout this part, assumptions in Subsection 2.3 are in force. Our main result is the following:

Theorem 4.2. *Under the assumptions of Subsection 2.3, there exists a time $T > 0$ and a classical solution (U^0, U) to (29) on the time interval $[0,T]$, which is, in addition, such that $D_{x_0}U^0$ and $D_{x_0,x}U$ are uniformly Lipschitz continuous in the space and measure variables.*

The result can be easily extended to non constant diffusions. We work here with a constant diffusion to simplify the notation.

The idea of the proof follows a similar splitting method as we did in Section 3, by dividing the time interval $[0,T]$ into $[t_{2k}, t_{2k+1})$ and $[t_{2k+1}, t_{2k+2})$, where $t_k = kT/(2N)$, $k \in \{0, 2N\}$. This time we alternate the two following problems: in $[t_{2k+1}, t_{2k+2})$ we solve, for a fixed $x_0 \in \mathbb{R}^{d_0}$, the first order system of master equations in $\mathbb{R}^d \times \mathcal{P}_2$:

$$
\left\{
\begin{array}{ll}
(i) & -\partial_t U^0 - 2\displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U^0(t,x_0,m,y) dm(y) \\
& \quad + 2\displaystyle\int_{\mathbb{R}^d} D_m U^0(t,x_0,m,y) \cdot H_p(x_0,y,D_x U(t,x_0,y,m),m) dm(y) = 0\,, \\
(ii) & -\partial_t U - 2\Delta_x U + 2H(x_0,x,D_x U,m) - 2\displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U(t,x_0,x,m,y) dm(y) \\
& \quad + 2\displaystyle\int_{\mathbb{R}^d} D_m U(t,x_0,x,m,y) \cdot H_p(x_0,y,D_x U(t,x_0,y,m),m) dm(y) = 0\,,
\end{array}
\right. \tag{30}
$$

while on $[t_{2k}, t_{2k+1})$ we solve for a fixed $(x,m) \in \mathbb{R}^d \times \mathcal{P}_2$ the system of HJ equations in $\mathbb{R}^{d_0}$:

$$
\left\{
\begin{array}{ll}
(i) & -\partial_t U^0 - 2\Delta_{x_0} U^0 + 2H^0(x_0, D_{x_0}U^0, m) = 0\,, \\
(ii) & -\partial_t U - 2\Delta_{x_0} U + 2D_{x_0}U \cdot H_p^0(x_0, D_{x_0}U^0(t,x_0,m),m) = 0\,.
\end{array}
\right. \tag{31}
$$

The (technical) analysis of System (30) is postponed to Section 6. We now concentrate on System (31). In order to write the estimate, we need to treat the pair of maps (U^0, U) simultaneously: this requires specific notation that we discuss first.

4.1 Analysis of the simple system of HJ equations

In this section we consider the system

$$\left\{\begin{array}{ll} (i) & -\partial_t U^0(t,x_0;m) - \Delta_{x_0} U^0(t,x_0;m) + H^0(x_0, D_{x_0}U^0(t,x_0;m),m) = 0 \text{ in } (0,T)\times\mathbb{R}^{d_0} \\ (ii) & -\partial_t U(t,x_0;x,m) - \Delta_{x_0} U(t,x_0;x,m) \\ & \qquad + D_{x_0}U(t,x_0;x,m)\cdot H_p^0(x_0, D_{x_0}U^0(t,x_0,m),m) = 0 \qquad \text{in } (0,T)\times\mathbb{R}^{d_0} \\ (iii) & U^0(T,x_0;m) = G^0(x_0,m) \text{ in } \mathbb{R}^{d_0},\ U(T,x_0;x,m) = G(x_0,x,m) \text{ in } \mathbb{R}^{d_0}, \end{array}\right. \tag{32}$$

where $(x,m)\in\mathbb{R}^d\times\mathcal{P}_2$ are fixed. The main part of this subsection consists in proving estimates on the solution (U^0,U) to (32).

4.1.1 Notation for the norms

In this section, we are dealing with pairs of maps $(V^0,V) = (V^0(x_0,m), V(x_0,x,m))$ which might also depend on time t, not indicated here. The way we compute the norms is crucial in order to match all the estimates. We use the following norms:

$$\big\|(V^0,V)\big\|_n := \sup_{m\in\mathcal{P}_2}\sum_{r=0}^n \sup_{x_0\in\mathbb{R}^{d_0},x\in\mathbb{R}^d} \big(|V^0(x_0,m)|^2 + |D_x^r V(x_0,x,m)|^2\big)^{1/2},$$

$$\left\|\frac{\delta(V^0,V)}{\delta m}\right\|_{n;k} := \sup_{m\in\mathcal{P}_2}\sum_{r=0}^n \sup_{\substack{x_0\in\mathbb{R}^{d_0},x\in\mathbb{R}^d,\\ \rho\in C_b^0,\|\rho\|_{-k}=1}} \left(\left|\frac{\delta V^0}{\delta m}(x_0,m)(\rho)\right|^2 + \left|D_x^r\frac{\delta V}{\delta m}(x_0,x,m)(\rho)\right|^2\right)^{1/2},$$

$$\left\|\frac{\delta^2(V^0,V)}{\delta m^2}\right\|_{n;k,k} := \sup_{m\in\mathcal{P}_2}\sum_{r=0}^n \sup_{\substack{x_0\in\mathbb{R}^{d_0},x\in\mathbb{R}^d,\\ \rho,\rho'\in C_b^0,\|\rho\|_{-k}=\|\rho'\|_{-k}=1}} \left(\left|\frac{\delta^2 V^0}{\delta m^2}(x_0,m)(\rho,\rho')\right|^2 + \left|D_x^r\frac{\delta^2 V}{\delta m^2}(x_0,x,m)(\rho,\rho')\right|^2\right)^{1/2}$$

and

$$\begin{aligned}\mathrm{Lip}_{n;k,k}\Big(\frac{\delta^2(V^0,V)}{\delta m^2}\Big) &:= \sup_{m_1\neq m_2}\mathbf{d}_2(m_1,m_2)^{-1}\left\|\frac{\delta^2}{\delta m^2}\big(V^0(m_2)-V^0(m_1),V(m_2)-V(m_1)\big)\right\|_{n;k,k} \\ &= \sup_{m_1\neq m_2}\mathbf{d}_2(m_1,m_2)^{-1}\sum_{r=0}^n \sup_{\substack{x_0\in\mathbb{R}^{d_0},\,x\in\mathbb{R}^d,\\ \rho,\rho'\in C_b^0,\,\|\rho\|_{-k}=\|\rho'\|_{-k}=1}} \\ &\left(\left|\frac{\delta^2 V^0}{\delta m^2}(x_0,m_2)(\rho,\rho') - \frac{\delta^2 V^0}{\delta m^2}(x_0,m_1)(\rho,\rho')\right|^2 + \left|D_x^r\frac{\delta^2 V}{\delta m^2}(x_0,x,m_2)(\rho,\rho') - D_x^r\frac{\delta^2 V}{\delta m^2}(x_0,x,m_1)(\rho,\rho')\right|^2\right)^{1/2}.\end{aligned}$$

We define in a similar way the quantities $\mathrm{Lip}_n^{x_0}(D_{x_0}^2V^0, D_{x_0}^2V)$, $\mathrm{Lip}_{n;k}(\frac{\delta V_{x_0}^0}{\delta m},\frac{\delta V_{x_0}}{\delta m})$ and $\mathrm{Lip}_n(D_{x_0}^2V^0, D_{x_0}^2U)$.

Note that arguing as in Remark 2.2, a control on $\left\|\frac{\delta(V^0,V)}{\delta m}\right\|_{n;k}$ yields a control on $\left\|\frac{\delta V^0}{\delta m}\right\|_{n,k-1}$ and $\left\|\frac{\delta V}{\delta m}\right\|_{n,k-1}$, and similarly for $\left\|\frac{\delta^2(V^0,V)}{\delta m^2}\right\|_{n;k,k}$, $\mathrm{Lip}_{n;k,k}\Big(\frac{\delta^2(V^0,V)}{\delta m^2}\Big)$, ...

4.1.2 Basic regularity of (U^0,U)

We recall that H^0 satisfies the assumptions of Section 2.3, in particular condition (16) is in force.

Proposition 4.3. *Fix $M>0$ and $n\geqslant 3$. There are constants $K_M, T_M>0$, depending on M, C_0 and γ, and a constant $C_M>0$ depending on*

$$\sup_{|p|\leqslant K_M}\sup_{m\in\mathcal{P}_2}\sum_{k=0}^3 \|D_{(x_0,p)}^k H^0(\cdot,p,m)\|_\infty + \sum_{k=0}^3 \|D_{(x_0,p)}^k H_p^0(\cdot,p,m)\|_\infty,$$

such that, if $\sum_{k=0}^{2}\big(\|D^kG^0\|_\infty+\|D^k_{x_0}G\|_{0,n-k}\big)+\Big(\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}G^0,D^2_{x_0}G)\Big)\leqslant M$*, then, for any* $T\in(0,T_M)$*, we have*

$$\begin{aligned}\sup_t\Big(\Big\|(U^0,U)(t)\Big\|_n+\Big\|D_{x_0}(U^0,U)(t)\Big\|_{n-1}+\Big\|D^2_{x_0}(U^0,U)(t)\Big\|_{n-2}+(\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}(U^0,U)(t)))\Big)\\ \leqslant\Big\|(G^0,G)\Big\|_n+\Big\|D_{x_0}(G^0,G)\Big\|_{n-1}+\Big\|D^2_{x_0}(G^0,G))\Big\|_{n-2}+(\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}(G^0,G)))+C_MT.\end{aligned}$$

Proof. To estimate $\|(U^0,U)\|_n$ it suffices to apply successively Proposition A.8 with $r=0$ and $l\leqslant n$, and to sum over l. The argument to estimate first and higher order derivatives with respect to x_0 is identical: apply successively Proposition A.8 with $r=1$ and $l\leqslant n-1$ (for $\|(D_{x_0}U^0,D_{x_0}U)\|_{n-1}$), with $r=2$ and $l\leqslant n-2$ (for $\|(D^2_{x_0}U^0,D^2_{x_0}U)\|_{n-2}$) and finally with $r=3$ and $l\leqslant n-3$ (for the Lipschitz bound in x_0 of $D^2_{x_0}(U^0,U)$). □

4.1.3 First order differentiability in m

Proposition 4.4. *Under the assumptions of Proposition 4.3, the pair* (U^0,U) *is of class* C^1 *with respect to* m*, as well as its derivatives with respect to* x *appearing below, and, for any fixed* $(x,m,\rho)\in\mathbb{R}^d\times\mathcal{P}_2\times C^{-k}$ *the derivative*

$$(v^0,v)=\Big(\frac{\delta U^0}{\delta m}(t,x_0;m)(\rho),\ \frac{\delta U}{\delta m}(t,x_0;x,m)(\rho)\Big)$$

solves,

$$\begin{cases}(i)\quad -\partial_tv^0-\Delta_{x_0}v^0+\dfrac{\delta H^0}{\delta m}(x_0,D_{x_0}U^0,m)(\rho)\\ \qquad\qquad +H^0_p(x_0,D_{x_0}U^0,m)\cdot D_{x_0}v^0=0\qquad\text{in }(0,T)\times\mathbb{R}^{d_0},\\ (ii)\quad -\partial_tv-\Delta_{x_0}v+D_{x_0}v\cdot H^0_p(x_0,D_{x_0}U^0,m)\\ \qquad +D_{x_0}U\cdot\left(\dfrac{\delta H^0_p}{\delta m}(x_0,D_{x_0}U^0,m)(\rho)+H^0_{pp}(x_0,D_{x_0}U^0,m)D_{x_0}v^0\right)=0\text{ in }(0,T)\times\mathbb{R}^{d_0},\\ (iii)\quad v^0(T,x_0;m)=\dfrac{\delta G^0}{\delta m}(x_0,m)(\rho),\ v(T,x_0,x;m)=\dfrac{\delta G}{\delta m}(x_0,x,m)(\rho)\qquad\text{in }\mathbb{R}^{d_0}.\end{cases}\tag{33}$$

Suppose in addition that, for $k\geqslant 2$*,*

$$\sup_{x_0,m}\Big(\|\frac{\delta G^0}{\delta m}\|_k+\|\frac{\delta G}{\delta m}\|_{n-1;k}+\|\frac{\delta G^0_{x_0}}{\delta m}\|_{k-1}+\|\frac{\delta G_{x_0}}{\delta m}\|_{n-2;k-1}+\big(\mathrm{Lip}^{x_0}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m},\frac{\delta G_{x_0}}{\delta m})\big)\Big)\leqslant M.$$

Then there exists $T_M,C_M>0$ *such that, for any* $T\in(0,T_M)$*, we have*

$$\begin{aligned}\sup_t\Big(\Big\|\frac{\delta(U^0,U)}{\delta m}(t)\Big\|_{n-1;k}+\Big\|\frac{\delta(U^0_{x_0},U_{x_0})}{\delta m}(t)\Big\|_{n-2;k-1}+(\mathrm{Lip}^{x_0}_{n-3;k-2}(\frac{\delta U^0_{x_0}}{\delta m},\frac{\delta U_{x_0}}{\delta m})(t))\Big)\\ \leqslant\Big(\Big\|\frac{\delta(G^0,G)}{\delta m}\Big\|_{n-1;k}+\Big\|\frac{\delta(G^0_{x_0},G_{x_0})}{\delta m}\Big\|_{n-2;k-1}+\mathrm{Lip}^{x_0}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m},\frac{\delta G_{x_0}}{\delta m})\Big)+C_MT,\end{aligned}$$

where C_M *depends on* M*,* r*,* n*,* k *and on the regularity of* H^0*.*

Proof. In order to show that U^0 is C^1 with respect to m, let us define

$$\hat U^0(t,x_0;s,m,y):=U^0(t,x_0,(1-s)m+s\delta_y).$$

Then, as $\hat H^0:=H^0(x_0,p,(1-s)m+s\delta_y)$ and $\hat g^0:=G^0(x_0,(1-s)m+s\delta_y)$ are of class C^1 with respect to the parameter $s\in[0,1]$, the map $\hat U^0$ is C^1 in s and its derivative $\hat v^0(t,x_0;m,y):=(d\hat U^0/ds)(t,x_0;0,m,y)$ solves the linearized equation

$$\begin{cases}-\partial_t\hat v^0-\Delta_{x_0}\hat v^0+\dfrac{\delta H^0}{\delta m}(x_0,D_{x_0}U^0,m,y)+H^0_p(x_0,D_{x_0}U^0,m)\cdot D_{x_0}\hat v^0=0\text{ in }(0,T)\times\mathbb{R}^{d_0},\\ \hat v^0(T,x_0)=\dfrac{\delta G^0}{\delta m}(x_0,m,y)\qquad\text{in }\mathbb{R}^{d_0}.\end{cases}$$

By uniqueness and parabolic regularity, the solution to this equation depends continuously of the parameters (m, y). Hence Lemma B.1 states that U^0 is C^1 in m with $\delta U^0/\delta m(t, x_0, m, y) = \hat{v}^0(t, x_0; m, y)$.

Next we consider the linear equation satisfied by U. By our previous discussion on U^0, the vector field

$$(t, x_0) \to H_p^0(x_0, D_{x_0}U^0(t, x_0, m), m)$$

is C^1 with respect to m. For $(s, m, y) \in [0, 1] \times \mathcal{P}_2 \times \mathbb{R}^{d_0}$, the map $\hat{U}(t, x_0; s, x, m, y) := U(t, x_0, x, (1 - s)m + s\delta_y)$ solves a linear equation in which the vector field

$$\hat{V}(t, x_0; s, m, y) := H_p^0(x_0, D_{x_0}U^0(t, x_0; (1 - s)m + s\delta_y), (1 - s)m + s\delta_y)$$

and the terminal condition $\hat{g}(x_0; x, s, m, y) := G(x_0, x, (1 - s)m + s\delta_y)$ are C^1 in s. Then $\hat{U}$ is C^1 in s and its derivative $\hat{v}(t, x_0; x, m, y) := (d/ds)\hat{U}(t, x_0; 0, x, m, y)$ solves the linear equation

$$\begin{cases} -\partial_t \hat{v} - \Delta_{x_0}\hat{v} + D_{x_0}\hat{v} \cdot H_p^0(x_0, D_{x_0}U^0, m) \\ \qquad + D_{x_0}U \cdot \left(\dfrac{\delta H_p}{\delta m}(x_0, D_{x_0}U^0, m, y) + H_{pp}(x_0, D_{x_0}U^0, m) D_{x_0}\hat{v}^0 \right) = 0 \qquad \text{in } (0, T) \times \mathbb{R}^{d_0}, \\ \hat{v}(T, x_0; x, m, y) = \dfrac{\delta G}{\delta m}(x_0, x, m, y) \qquad \text{in } \mathbb{R}^{d_0}. \end{cases}$$

As the solution to this equation depends continuously of the parameters (m, y), Lemma B.1 states that U is C^1 in m with $\delta U/\delta m(t, x_0, x, m, y) = \hat{v}(t, x_0; x, m, y)$. This proves that the derivative $(\hat{v}^0, \hat{v}) = (\delta U^0/\delta m, \delta U/\delta m)(t, x_0, x, m, y)$ solves (33) with $\rho = \delta_y$.

Hence, for any $\rho \in C_b^0$, the pair $(v^0, v) = (\frac{\delta U^0}{\delta m}(t, x_0; m)(\rho), \frac{\delta U}{\delta m}(t, x_0; x, m)(\rho))$ solves a linear system of the form (113) in which the drifts

$$V^0(t, x^0; m) := H_p^0(x_0, D_{x_0}U^0(t, x_0, m), m)$$

and

$$V(t, x^0; x, m) := H_{pp}^0(x_0, D_{x_0}U^0(t, x_0, m), m) D_{x_0}U(t, x_0; x)$$

are bounded in class C_b^1 and $C_b^{0,n-1} \cap C_b^{1,n-2}$ respectively, while the source terms

$$f^0(t, x^0; m) := \frac{\delta H^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho)$$

and

$$f(t, x_0; x, m) := D_{x_0}U(t, x_0; x) \cdot \frac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho)$$

are in C_b^1 and $C_b^{0,n-1} \cap C_b^{1,n-2}$ respectively, thanks to Proposition 4.3. We then use Proposition A.9 successively to obtain the estimates: first with $r = 0$ and $l \leqslant n - 1$, we get

$$\Big(|\frac{\delta U^0}{\delta m}(t, x_0; m)(\rho)|^2 + |D_x^l \frac{\delta U}{\delta m}(t, x_0; x, m)(\rho)|^2 \Big)^{1/2} \leqslant \\ (1 + CT) \sup_{x_0, x} \Big(|\frac{\delta G^0}{\delta m}(x_0; m)(\rho)|^2 + |D_x^l \frac{\delta G}{\delta m}(x_0; x, m)(\rho)|^2 \Big)^{1/2} + CT.$$

Then by taking the supremum over $\|\rho\|_{-k} = 1, x_0, x$ and summing over $l \leqslant n - 1$ we find the estimate for $\|\frac{\delta(U^0, U)}{\delta m}\|_{n-1;k}$. An analogous application of Proposition A.9 with $r = 1$ and $l \leqslant n - 2$ provides the bound for $\|\frac{\delta(U^0_{x_0}, U_{x_0})}{\delta m}\|_{n-2;k-1}$, while the Lipschitz estimate in x_0 for $(\frac{\delta U^0_{x_0}}{\delta m}, \frac{\delta U_{x_0}}{\delta m})$ is obtained similarly with $r = 2$, and $l \leqslant n - 3$. □

4.1.4 Second order differentiability with respect to m

Proposition 4.5. *Under the assumptions of Proposition 4.4, $k \geqslant 3$, the pair (U^0, U) (and its derivatives with respect to x) is of class C^2 with respect to m and, for any fixed $(x, m, \rho, \rho') \in \mathbb{R}^d \times \mathcal{P}_2 \times C^{-(k-1)} \times C^{-(k-1)}$ the derivative $(w^0, w) = (\delta^2 U^0/\delta m^2(t, x_0; m)(\rho, \rho'), \delta^2 U/\delta m^2(t, x_0; x, m)(\rho, \rho'))$ solves*

$$
\begin{cases}
(i) & -\partial_t w^0 - \Delta_{x_0} w^0 + H_p^0(x_0, D_{x_0}U^0, m)\cdot D_{x_0} w^0 + \dfrac{\delta^2 H^0}{\delta m^2}(x_0, D_{x_0}U^0, m)(\rho, \rho') \\
& \quad + H_{pp}^0(x_0, D_{x_0}U^0, m) D_{x_0} v^0 \cdot D_{x_0}(v')^0 + \dfrac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho)\cdot D_{x_0}(v')^0 \\
& \quad + \dfrac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho')\cdot D_{x_0} v^0 = 0 \qquad \text{in } (0,T)\times\mathbb{R}^{d_0} \\
(ii) & -\partial_t w - \Delta_{x_0} w + H_p^0(x_0, D_{x_0}U^0, m)\cdot D_{x_0} w \\
& + D_{x_0} v \cdot \left(\dfrac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho') + H_{pp}^0(x_0, D_{x_0}U^0, m) D_{x_0}(v')^0 \right) \\
& + D_{x_0} v' \cdot \left(\dfrac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho) + H_{pp}^0(x_0, D_{x_0}U^0, m) D_{x_0} v^0 \right) \\
& + D_{x_0} U \cdot \Big(\dfrac{\delta H_{pp}^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho) D_{x_0}(v')^0 + \dfrac{\delta^2 H_p^0}{\delta m^2}(x_0, D_{x_0}U^0, m)(\rho, \rho') \\
& \quad + H_{ppp}^0(x_0, D_{x_0}U^0, m) D_{x_0} v^0 D_{x_0}(v')^0 + \dfrac{\delta H_{pp}^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho') D_{x_0} v^0 \\
& \quad + H_{pp}^0(x_0, D_{x_0}U^0, m) D_{x_0} w^0 \Big) = 0 \text{ in } (0,T)\times\mathbb{R}^{d_0} \\
(iii) & w^0(T, x_0; m) = \dfrac{\delta^2 G^0}{\delta m^2}(x_0, m)(\rho, \rho'),\ w(T, x_0; x, m) = \dfrac{\delta^2 G}{\delta m^2}(x_0, x, m)(\rho, \rho') \text{ in } \mathbb{R}^{d_0},
\end{cases}
\tag{34}
$$

where $(v^0, v), ((v')^0, v')$ are the solutions to (33) associated with ρ and ρ' respectively. Moreover, if

$$
\left\| \frac{\delta^2 (G^0, G)}{\delta m^2} \right\|_{n-2;k-1,k-1} + \mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) \leqslant M,
$$

then there exists $T_M, C_M > 0$ such that, for any $T \in (0, T_M)$,

$$
\begin{aligned}
&\sup_t \Big(\left\| \frac{\delta^2 (U^0, U)}{\delta m^2}(t) \right\|_{n-2;k-1,k-1} + \big(\mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\big(\frac{\delta^2 U^0}{\delta m^2}, \frac{\delta^2 U}{\delta m^2}\big)(t)\big) \Big) \\
&\quad \leqslant \Big(\left\| \frac{\delta^2 (G^0, G)}{\delta m^2} \right\|_{n-2;k-1,k-1} + \mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) \Big) + C_M T.
\end{aligned}
$$

Proof. The differentiability of $\delta U^0/\delta m$ and of $\delta U/\delta m$ and the representation formula (34) can be established as for U^0 and U in Proposition 4.4. To prove the estimate, we use Proposition A.9 with

$$
V^0(t, x^0; m) := H_p^0(x_0, D_{x_0}U^0(t, x_0, m), m)
$$

and

$$
V(t, x^0 x; m) := H_{pp}^0(x_0, D_{x_0}U^0(t, x_0, m), m) D_{x_0} U(t, x_0, x),
$$

which are bounded in class C_b^1 and $C_b^{0,n-1} \cap C_b^{1,n-2}$ respectively, while the source terms

$$
\begin{aligned}
f^0(t, x^0; m) :=& \frac{\delta^2 H^0}{\delta m^2}(x_0, D_{x_0}U^0, m)(\rho, \rho') + H_{pp}^0(x_0, D_{x_0}U^0, m) D_{x_0} v^0 \cdot D_{x_0}(v')^0 \\
&+ \frac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho)\cdot D_{x_0}(v')^0 + \frac{\delta H_p^0}{\delta m}(x_0, D_{x_0}U^0, m)(\rho')\cdot D_{x_0} v^0
\end{aligned}
$$

and

$$
\begin{aligned}
f(t,x_0,x;m) := D_{x_0}v\cdot\left(\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^0,m)(\rho')+H_{pp}^0(x_0,D_{x_0}U^0,m)D_{x_0}(v')^0\right)\\
+D_{x_0}v'\cdot\left(\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^0,m)(\rho)+H_{pp}^0(x_0,D_{x_0}U^0,m)D_{x_0}v^0\right)\\
+D_{x_0}U\cdot\Big(\frac{\delta H_{pp}^0}{\delta m}(x_0,D_{x_0}U^0,m)(\rho)D_{x_0}(v')^0+\frac{\delta^2 H_p^0}{\delta m^2}(x_0,D_{x_0}U^0,m)(\rho,\rho')\\
+H_{ppp}^0(x_0,D_{x_0}U^0,m)D_{x_0}v^0D_{x_0}(v')^0+\frac{\delta H_{pp}^0}{\delta m}(x_0,D_{x_0}U^0,m)(\rho')D_{x_0}v^0\Big)
\end{aligned}
$$

are in C_b^0 and $C_b^{0,n-2}$ respectively, thanks to Propositions 4.3 and 4.4. By Proposition A.9, with $r=0$ and $n-2$ we obtain the estimates for $\|\frac{\delta^2(U^0,U)}{\delta m^2}\|_{n-2;k-1,k-1}$. The Lipschitz bound in x_0 of $(\frac{\delta^2U^0}{\delta m^2},\frac{\delta^2U}{\delta m^2})$ follows analogously. □

4.1.5 Lipschitz regularity of the second order derivatives

We finally address the Lipschitz regularity of second order derivatives of U^0 and U with respect to m and x_0.

Proposition 4.6. *Under the assumptions of Proposition 4.5 and if, in addition,* $\mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2G^0}{\delta m^2},\frac{\delta^2G}{\delta m^2})\leqslant M$, *then we have,*

$$
\sup_t(\mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2U^0}{\delta m^2},\frac{\delta^2U}{\delta m^2})(t))\leqslant \mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2G^0}{\delta m^2},\frac{\delta^2G}{\delta m^2})+C_MT,
$$

where the constant C_M depend on the regularity of H and H^0 and on M.

Proof. Let $(x,\rho,\rho')\in\mathbb{R}^d\times C^{-(k-2)}\times C^{-(k-2)}$, $m^1,m^2\in\mathcal{P}_2$, $(U^{0,1},U^1)$ be the solution to (32) associated with (x,m^1) and $(U^{0,2},U^2)$ be the solution associated with (x,m^2). We denote by $(v^{0,1},v^1)$, $((v')^{0,1},(v')^1)$ (resp. $(v^{0,2},v^2)$, $((v')^{0,2},(v')^2)$) the corresponding solutions to the first order linearized system (33) associated with ρ and ρ', and by $(w^{0,1},w^1)$ (resp. $(w^{0,2},w^2)$) the corresponding solution of the second order linearized system (34). We want to estimate the difference $(z^0,z):=(w^{0,2}-w^{0,1},w^2-w^1)$. We have

$$
\begin{cases}
-\partial_t z^0-\Delta_{x_0}z^0+H_p^0(x_0,D_{x_0}U^{0,1}(t,x_0,m^1),m^1)\cdot D_{x_0}z^0+f^0=0,\\
-\partial_t z-\Delta_{x_0}z+D_{x_0}z\cdot H_p^0(x_0,D_{x_0}U^{0,1},m^1)-H_{pp}^0(x_0,D_{x_0}U^{0,1},m)D_{x_0}U^1\cdot D_{x_0}z^0+f=0\\
z^0(T)=\dfrac{\delta^2G^0}{\delta m^2}(x_0,m^2)(\rho,\rho')-\dfrac{\delta^2G^0}{\delta m^2}(x_0,m^1)(\rho,\rho'),\\
z(T)=\dfrac{\delta^2G}{\delta m^2}(x_0,x,m^2)(\rho,\rho')-\dfrac{\delta^2G}{\delta m^2}(x_0,x,m^1)(\rho,\rho'),
\end{cases}
$$

where

$$
\begin{aligned}
f^0:=&(H_p^0(x_0,D_{x_0}U^{0,2},m^2)-H_p^0(x_0,D_{x_0}U^{0,1},m^1))\cdot D_{x_0}w^{0,2}\\
&+\frac{\delta^2H^0}{\delta m^2}(x_0,D_{x_0}U^{0,2},m^2)(\rho,\rho')-\frac{\delta^2H^0}{\delta m^2}(x_0,D_{x_0}U^{0,1},m^1)(\rho,\rho')\\
&+H_{pp}^0(x_0,D_{x_0}U^{0,2},m^2)D_{x_0}v^{0,2}\cdot D_{x_0}(v')^{0,2}-H_{pp}^0(x_0,D_{x_0}U^{0,1},m^1)D_{x_0}v^{0,1}\cdot D_{x_0}(v')^{0,1}\\
&+\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^{0,2},m^2)(\rho)\cdot D_{x_0}(v')^{0,2}-\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^{0,1},m^1)(\rho)\cdot D_{x_0}(v')^{0,1}\\
&+\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^{0,2},m^2)(\rho')\cdot D_{x_0}v^{0,2}-\frac{\delta H_p^0}{\delta m}(x_0,D_{x_0}U^{0,1},m^1)(\rho')\cdot D_{x_0}v^{0,1}
\end{aligned}
$$

and

$$
\begin{aligned}
f :=& D_{x_0}w^2 \cdot \big(H^0_p(x_0, D_{x_0}U^{0,2}, m^2) - H^0_p(x_0, D_{x_0}U^{0,1}, m^1)\big) \\
&+ D_{x_0}v^2 \cdot \left(\frac{\delta H^0_p}{\delta m}(x_0, D_{x_0}U^{0,2}, m^2)(\rho') + H^0_{pp}(x_0, D_{x_0}U^{0,2}, m^2)D_{x_0}(v')^{0,2}\right) \\
&- D_{x_0}v^1 \cdot \left(\frac{\delta H^0_p}{\delta m}(x_0, D_{x_0}U^{0,1}, m^1)(\rho') + H^0_{pp}(x_0, D_{x_0}U^{0,1}, m^1)D_{x_0}(v')^{0,1}\right) \\
&+ D_{x_0}(v')^2 \cdot \left(\frac{\delta H^0_p}{\delta m}(x_0, D_{x_0}U^{0,2}, m^2)(\rho) + H^0_{pp}(x_0, D_{x_0}U^{0,2}, m^2)D_{x_0}v^{0,2}\right) \\
&- D_{x_0}(v')^1 \cdot \left(\frac{\delta H^0_p}{\delta m}(x_0, D_{x_0}U^{0,1}, m^1)(\rho) + H^0_{pp}(x_0, D_{x_0}U^{0,1}, m^1)D_{x_0}v^{0,1}\right) \\
&+ D_{x_0}U^2 \cdot \Big(\frac{\delta H^0_{pp}}{\delta m}(x_0, D_{x_0}U^{0,2}, m^2)(\rho)D_{x_0}(v')^{0,2} + \frac{\delta^2 H^0_p}{\delta m^2}(x_0, D_{x_0}U^{0,2}, m^2)(\rho, \rho') \\
&\qquad + H^0_{ppp}(x_0, D_{x_0}U^{0,2}, m^2)D_{x_0}v^{0,2}D_{x_0}(v')^{0,2} + \frac{\delta H^0_{pp}}{\delta m}(x_0, D_{x_0}U^{0,2}, m^2)(\rho')D_{x_0}v^{0,2}\Big) \\
&+ \Big(H^0_{pp}(x_0, D_{x_0}U^{0,2}, m)D_{x_0}U^2 - H^0_{pp}(x_0, D_{x_0}U^{0,1}, m)D_{x_0}U^1\Big) \cdot D_{x_0}w^{0,2} \\
&- D_{x_0}U^1 \cdot \Big(\frac{\delta H^0_{pp}}{\delta m}(x_0, D_{x_0}U^{0,1}, m^1)(\rho)D_{x_0}(v')^{0,1} + \frac{\delta^2 H^0_p}{\delta m^2}(x_0, D_{x_0}U^{0,1}, m^1)(\rho, \rho') \\
&\qquad + H^0_{ppp}(x_0, D_{x_0}U^{0,1}, m^1)D_{x_0}v^{0,1}D_{x_0}(v')^{0,1} + \frac{\delta H^0_{pp}}{\delta m}(x_0, D_{x_0}U^{0,1}, m^1)(\rho')D_{x_0}v^{0,1}\Big).
\end{aligned}
$$

Proposition 4.4 (for the representation of the $(v^{0,i}, v^i)$) and Proposition 4.5 (for their Lipschitz regularity in m and in x_0) imply in particular that

$$\sup_t \big(\|D_{x_0}(v^{0,2} - v^{0,1})\|_\infty + \|D_{x_0}(v^2 - v^1)\|_{0,n-3}\big) \leqslant C\mathbf{d}_2(m_1, m_2)$$

and hence we have, using also Proposition 4.5,

$$\sup_t \big(\|f^0\|_\infty + \|f\|_{0,n-3}\big) \leqslant C\mathbf{d}_2(m^1, m^2).$$

Using Proposition A.9 (with $r = 0$), we obtain, for any $l \leqslant n-3$,

$$
\begin{aligned}
&\sup_{t,x_0,x} (|z^0(t,x_0)|^2 + |D^l_x z(t,x_0,x)|^2)^{1/2} \\
&\leqslant (1 + CT)\sup_{x_0,x}\Big(\left|\frac{\delta^2 G^0}{\delta m^2}(x_0, m^2)(\rho,\rho') - \frac{\delta^2 G^0}{\delta m^2}(x_0, m^1)(\rho,\rho')\right|^2 \\
&\qquad + \left|D^l_x\frac{\delta^2 G}{\delta m^2}(x_0, x, m^2)(\rho,\rho') - D^l_x\frac{\delta^2 G}{\delta m^2}(x_0, x, m^1)(\rho,\rho')\right|^2\Big)^{1/2} + CT\mathbf{d}_2(m^1, m^2),
\end{aligned}
$$

which gives the claim. □

We complete this section by stating similar estimates on the Lipschitz regularity of the other second order derivatives:

Proposition 4.7. *Under the assumptions of Proposition 4.5 and if, in addition,* $\mathrm{Lip}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}) + \mathrm{Lip}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) \leqslant M$*, then we have,*

$$\sup_t(\mathrm{Lip}_{n-3;k-2}(\frac{\delta U^0_{x_0}}{\delta m}, \frac{\delta U_{x_0}}{\delta m})(t)) \leqslant \mathrm{Lip}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}) + C_M T$$

and

$$\sup_t(\mathrm{Lip}_{n-3}(D^2_{x_0}U^0, D^2_{x_0}U)(t)) \leqslant \mathrm{Lip}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) + C_M T,$$

where the constant C_M depend on the regularity of H and H^0 and on M.

As the proof is completely similar to the proof of Proposition 4.6, we omit it.

4.2 Existence of a solution

4.2.1 Definition of the semi-discrete scheme

Let us fix some horizon $T > 0$ (small) and a step-size $\tau := T/(2N)$ (where $N \in \mathbb{N}$, $N \geqslant 1$). We set $t_k = kT/(2N)$, $k \in \{0, 2N\}$. We define by backward induction the continuous maps $U^{0,N} = U^{0,N}(t, x_0, m)$ and $U^N = U^N(t, x_0, x, m)$ as follows: we require that

(i) $(U^{0,N}, U^N)$ satisfies the terminal condition:

$$U^{0,N}(T, x_0, m) = G^0(x_0, m),\ U^N(T, x_0, x, m) = G(x_0, x, m) \qquad \forall (x_0, x, m) \in \mathbb{R}^d \times \mathbb{R}^{d_0} \times \mathcal{P}_2,$$

(ii) for $x_0 \in \mathbb{R}^{d_0}$ fixed, $(U^{0,N}, U^N)$ solves the backward system of first order master equations:

$$\left\{\begin{array}{ll} (i) & -\partial_t U^0 - 2\displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U^0(t, x_0, m, y) dm(y) \\ & \quad +2\displaystyle\int_{\mathbb{R}^d} D_m U^0(t, x_0, m, y) \cdot H_p(x_0, y, D_x U(t, x_0, y, m), m) dm(y) = 0 \\ (ii) & -\partial_t U - 2\Delta_x U + 2H(x_0, x, D_x U, m) - 2\displaystyle\int_{\mathbb{R}^d} \mathrm{div}_y D_m U(t, x_0, x, m, y) dm(y) \\ & \quad +2\displaystyle\int_{\mathbb{R}^d} D_m U(t, x_0, x, m, y) \cdot H_p(x_0, y, D_x U(t, x_0, y, m), m) dm(y) = 0 \end{array}\right. \tag{35}$$

on time intervals of the form (t_{2j+1}, t_{2j+2}), for $j = 0, \dots, N-1$,

(iii) for $(x, m) \in \mathbb{R}^d \times \mathcal{P}_2$ fixed, $(U^{0,N}, U^N)$ solves the backward system of HJ equations:

$$\left\{\begin{array}{ll} (i) & -\partial_t U^0 - 2\Delta_{x_0} U^0 + 2H^0(x_0, D_{x_0} U^0, m) = 0 \\ (ii) & -\partial_t U - 2\Delta_{x_0} U + 2D_{x_0} U \cdot H^0_p(x_0, D_{x_0} U^0(t, x_0, m), m) = 0 \end{array}\right. \tag{36}$$

on time intervals of the form (t_{2j}, t_{2j+1}), for $j = 0, \dots, N-1$.

Our aim is to show that, if the time horizon is short enough, $(U^{0,N}, U^N)$ converges to a solution of the master equation for MFGs with a major player as $N \to +\infty$.

4.2.2 Proof of the existence of a solution

For $n \geqslant 4$ and $k \in \{3, \dots, n-1\}$, let

$$\begin{aligned} M := 1 &+ \left\|(G^0, G)\right\|_n + \left\|D_{x_0}(G^0, G)\right\|_{n-1} + \left\|D^2_{x_0}(G^0, G)\right\|_{n-2} + (\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G)) \\ &+ \left\|\frac{\delta(G^0, G)}{\delta m}\right\|_{n-1;k} + \left\|\frac{\delta(G^0_{x_0}, G_{x_0})}{\delta m}\right\|_{n-2;k-1} + (\mathrm{Lip}^{x_0}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m})) \\ &+ \left\|\frac{\delta^2(G^0, G)}{\delta m^2}\right\|_{n-2;k-1,k-1} + (\mathrm{Lip}^{x_0}_{n-3;k-2,k-2}(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2})) \\ &+ (\mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2})) + (\mathrm{Lip}_{n-3;k-2}(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m})) + (\mathrm{Lip}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G)). \end{aligned}$$

Lemma 4.8. *There exists $T_M > 0$, depending on the regularity of H^0, H and on M, such that, for any $T \in (0, T_M]$ and $N \geqslant 1$, we have, for any $t \in [0,T]$,*

$$\begin{aligned}
&\Big\|(U^0,U)(t)\Big\|_n + \Big\|D_{x_0}(U^0,U)(t)\Big\|_{n-1} + \Big\|D^2_{x_0}(U^0,U)(t)\Big\|_{n-2} + (\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}U^0, D^2_{x_0}U)(t)) \\
&+ \Big\|\frac{\delta(U^0,U)}{\delta m}(t)\Big\|_{n-1;k} + \Big\|\frac{\delta(U^0_{x_0},U_{x_0})}{\delta m}(t)\Big\|_{n-2;k-1} + (\mathrm{Lip}^{x_0}_{n-3;k-2}(\frac{\delta U^0_{x_0}}{\delta m}, \frac{\delta U_{x_0}}{\delta m})(t)) \qquad (37)\\
&+ \Big\|\frac{\delta^2(U^0,U)}{\delta m^2}(t)\Big\|_{n-2;k-1,k-1} + (\mathrm{Lip}^{x_0}_{n-3;k-2,k-2}(\frac{\delta^2U^0}{\delta m^2}, \frac{\delta^2U}{\delta m^2})(t)) \\
&+ (\mathrm{Lip}_{n-3;k-2,k-2}(\frac{\delta^2U^0}{\delta m^2}, \frac{\delta^2U}{\delta m^2})(t)) + (\mathrm{Lip}_{n-3;k-2}(\frac{\delta U^0_{x_0}}{\delta m}, \frac{\delta U_{x_0}}{\delta m})(t)) + (\mathrm{Lip}_{n-3}(D^2_{x_0}U^0, D^2_{x_0}U)(t)) \ \leqslant\ M.
\end{aligned}$$

Moreover:

- *The maps $U^{0,N}$ and U^N are globally Lipschitz continuous in all variables and their first and second space derivatives are globally Holder continuous in all variables, uniformly with respect to N.*
- *The maps $D_mU^{0,N}$ and D_mU^N are Holder continuous in (t,x_0,m,y) and (t,x_0,x,m,y) respectively, uniformly with respect to N, in any set of the form*

$$\begin{aligned}
&\{(t,x_0,m,y) \in [0,T]\times\mathbb{R}^{d_0}\times\mathcal{P}_2\times\mathbb{R}^d,\ M_2(m)\leqslant R,\ |y|\leqslant R\} \\
&\text{and } \{(t,x_0,x,m,y) \in [0,T]\times\mathbb{R}^{d_0}\times\mathbb{R}^d\times\mathcal{P}_2\times\mathbb{R}^d,\ M_2(m)\leqslant R,\ |y|\leqslant R\} \qquad (38)
\end{aligned}$$

respectively (where $M_2(m) = (\int_{\mathbb{R}^d}|y|^2m(dy))^{1/2}$).

Proof. We only sketch the proof, since it is exactly the same as for the second order master equation (see Lemma 3.7). The proof of (37) can be established by collecting the estimates in Propositions 6.5, 6.7 and 6.9 in Section 6 below, which provide the bounds on intervals of the form (t_{2j+1}, t_{2j+2}), and, for the intervals of the form (t_{2j}, t_{2j+1}), by Proposition 4.3, 4.4, 4.5 and 4.6.

The Lipschitz regularity in space of $U^{0,N}$ and U^N and of their first and second order space derivatives follows immediately from (37). As $D_mU^{0,N}$ and D_mU^N are bounded according to (37), $U^{0,N}$ and U^N and their first and second order space derivatives are also Lipschitz continuous in m. Finally, since $U^{0,N}$ and U^N satisfy the equations (35) and (36), the bounds in (37) show that $\partial_tU^{0,N}$ and ∂_tU^N are bounded and therefore that $U^{0,N}$ and U^N are also Lipschitz continuous in time. The global Holder regularity of the first and second space derivatives of $U^{0,N}$ and U^N then follows by interpolation (Lemma B.2).

The Lipschitz regularity in space and in measure of $D_mU^{0,N}$ and D_mU^N is a consequence of (37) while the Holder regularity in time in sets of the form (38) comes from interpolation (Lemma B.4). □

Proof of Theorem 4.2. It relies exactly on the same argument as in the proof of Theorem 3.2 and we omit it. □

4.3 Uniqueness of the solution

We finally address the uniqueness of the solution of the master equation for MFGs with a major player:

Theorem 4.9. *Let $(U^{0,1},U^1)$ and $(U^{0,2},U^2)$ be two classical solutions to (29) defined on the time interval $[0,T]$ and such that $D_{x_0}U^{0,1}$ and $D_{x_0,x}U^1$ are uniformly Lipschitz continuous in the space and measure variables. Then $(U^{0,1},U^1) = (U^{0,2},U^2)$.*

Proof. Let $(t_0,\bar{x}_0,\bar{m}_0) \in [0,T)\times\mathbb{R}^{d_0}\times\mathcal{P}_2$ be an initial condition, Z be a random variable with law $\bar{m}_0$ and let (X^0_t, m_t, X_t) be the solution to

$$\begin{cases}
dX^0_t = -H^0_p(X^0_t, D_{x_0}U^{0,1}(t,X^0_t,m_t), m_t)dt + \sqrt{2}dW^0_t \text{ in } (0,T) \\
dm_t = \big(\Delta m_t + \mathrm{div}(m_tH_p(X^0_t,x,D_xU^1(t,X^0_t,x,m_t),m_t))\big)\,dt \text{ in } (0,T)\times\mathbb{R}^d \\
dX_t = -H_p(X^0_t,X_t,D_xU^1(t,X^0_t,X_t,m_t)dt + \sqrt{2}dW_t \text{ in } (0,T) \\
X^0_{t_0} = \bar{x}_0,\ m_{t_0} = \bar{m}_0,\ X_{t_0} = Z
\end{cases}$$

where (W_t^0) and (W_t) are Brownian motions, (W_t^0), (W_t) and Z being independent. As $D_xU^{0,1}$ and D_xU^1 are globally Lipschitz continuous, the above system has a unique solution. Note that m_t is the conditional law of X_t given $(W_s^0)_{s\leqslant t}$.

We compute the variation of $U^{0,1}$ along (t, X_t^0, m_t):

$$\begin{aligned}
dU^{0,1}(t,X_t^0,m_t) =& \Big(\partial_t U^{0,1} + \Delta_{x_0}U^{0,1} - H_p^0(X_t^0, D_{x_0}U^{0,1}, m_t)\cdot D_{x_0}U^{0,1} \\
&- \int_{\mathbb{R}^d} D_m U^{0,1}\cdot H_p(X_t^0, y, D_xU^1(t,X_t^0,y,m_t), m_t)m_t(dy) + \int_{\mathbb{R}^d}\mathrm{div}_y D_m U^{0,1} m_t(dy)\Big)dt \\
&+ \sqrt{2}D_{x_0}U^{0,1}\cdot dW_t^0,
\end{aligned}$$

where, unless specified otherwise, $U^{0,1}$ and its space derivatives are computed at (t, X_t^0, m_t) while $D_mU^{0,1}$ and its space derivatives are computed at (t, X_t^0, m_t, y). In view of the equation satisfied by $U^{0,1}$, we find

$$\begin{aligned}
dU^{0,1}(t,X_t^0,m_t) =& \Big(H^0(X_t^0, D_{x_0}U^{0,1}, m_t) - H_p^0(X_t^0, D_{x_0}U^{0,1}, m_t)\cdot D_{x_0}U^{0,1}\Big)dt \\
&+ \sqrt{2}D_{x_0}U^{0,1}\cdot dW_t^0\,.
\end{aligned}$$

We proceed in the same way for $U^{0,2}$ and obtain, in view of the equation satisfied by $U^{0,2}$:

$$\begin{aligned}
dU^{0,2}(t,X_t^0,m_t) =& \Big(H^0(X_t^0, D_{x_0}U^{0,2}, m_t) - H_p^0(X_t^0, D_{x_0}U^{0,1}, m_t)\cdot D_{x_0}U^{0,2} \\
&+ \int_{\mathbb{R}^d} D_m U^{0,2}\cdot (H_p(X_t^0, y, D_xU^2(t,X_t^0,y,m_t), m_t) - H_p(X_t^0, y, D_xU^1(t,X_t^0,y,m_t)), m_t)m_t(dy)\Big)dt \\
&+ \sqrt{2}D_{x_0}U^{0,2}\cdot dW_t^0,
\end{aligned}$$

where, unless specified otherwise, $U^{0,2}$ and its space derivatives are computed at (t, X_t^0, m_t) while $D_mU^{0,2}$ and its space derivatives are computed at (t, X_t^0, m_t, y). Therefore

$$\begin{aligned}
d(U^{0,2}-U^{0,1})^2 =& 2(U^{0,2}-U^{0,1})\Big(H^0(X_t^0, D_{x_0}U^{0,2}, m_t) - H^0(X_t^0, D_{x_0}U^{0,1}, m_t) \\
&- H_p^0(X_t^0, D_{x_0}U^{0,1}, m_t)\cdot(D_{x_0}U^{0,2} - D_{x_0}U^{0,1}) \\
&+ \int_{\mathbb{R}^d} D_m U^{0,2}\cdot (H_p(X_t^0, y, D_xU^2(t,X_t^0,y,m_t), m_t) - H_p(X_t^0, y, D_xU^1(t,X_t^0,y,m_t), m_t))m_t(dy)\Big)dt \\
&+ 2(D_{x_0}U^{0,2} - D_{x_0}U^{0,1})^2 dt + 2\sqrt{2}(U^{0,2}-U^{0,1})(D_{x_0}U^{0,2} - D_{x_0}U^{0,1})\cdot dW_t^0.
\end{aligned}$$

Let us set $U_t^{0,i} = U^{0,i}(t, X_t^0, m_t)$ (for $i = 1, 2$). We integrate in time between $s \in [t_0, T]$ and T, take expectation and use the fact that $U_T^{0,1} = U_T^{0,2} = G^0(X_T^0, m_T)$:

$$\begin{aligned}
0 =& \mathbb{E}\Big[(U_s^{0,2}-U_s^{0,1})^2 + \int_s^T 2(U_t^{0,2}-U_t^{0,1})\Big(H^0(X_t^0, D_{x_0}U^{0,2}, m_t) - H^0(X_t^0, D_{x_0}U^{0,1}, m_t) \\
&- H_p^0(X_t^0, D_{x_0}U^{0,1}, m_t)\cdot(D_{x_0}U^{0,2} - D_{x_0}U^{0,1}) \\
&+ \int_{\mathbb{R}^d} D_m U^{0,2}\cdot (H_p(X_t^0, y, D_xU^2(t,X_t^0,y,m_t), m_t) - H_p(X_t^0, y, D_xU^1(t,X_t^0,y,m_t), m_t))m_t(dy)\Big)dt \\
&+ 2\int_s^T |D_{x_0}U^{0,2} - D_{x_0}U^{0,1}|^2 dt\Big].
\end{aligned}$$

Thanks to the regularity of the solutions, we have by Cauchy-Schwarz inequality and for any $\epsilon > 0$:

$$\begin{aligned}
0 \geqslant& \mathbb{E}\Big[(U_s^{0,2}-U_s^{0,1})^2 - \int_s^T \Big(C_\epsilon(U_t^{0,2}-U_t^{0,1})^2 + \epsilon|D_{x_0}(U^{0,2}-U^{0,1})|^2 \\
&+ \epsilon\int_{\mathbb{R}^d}|D_x(U^2-U^1)(t,X_t^0,y,m_t)|^2 m_t(dy)\Big)dt + 2\int_s^T |D_{x_0}(U^{0,2}-U^{0,1})|^2\,dt\Big].
\end{aligned}$$

So, for ϵ small enough, we obtain

$$0 \geqslant \mathbb{E}\Big[(U_s^{0,2} - U_s^{0,1})^2 - \int_s^T \Big(C_\epsilon (U_t^{0,2} - U_t^{0,1})^2 + \epsilon \int_{\mathbb{R}^d} |D_x(U^2 - U^1)(t, X_t^0, y, m_t)|^2 m_t(dy)\Big) dt + \int_s^T |D_{x_0}(U^{0,2} - U^{0,1})|^2 dt \Big].$$

We argue in the same way for $U_t^i := U^i(t, X_t^0, X_t, m_t)$ $(i = 1, 2)$ and find:

$$0 \geqslant \mathbb{E}\Big[(U_s^2 - U_s^1)^2 - \int_s^T \Big(C_\epsilon (U_t^2 - U_t^1)^2 + \epsilon |D_{x_0}(U^{0,2} - U^{0,1})|^2 + \epsilon \int_{\mathbb{R}^d} |D_x(U^2 - U^1)(t, X_t^0, y, m_t)|^2 m_t(dy)\Big) dt + \int_s^T |D_{x_0}(U^2 - U^1)|^2 + |D_x(U^2 - U^1)|^2 \, dt \Big].$$

We add the last two inequalities to obtain:

$$\begin{aligned} 0 \geqslant \mathbb{E}\Big[&(U_s^{0,2} - U_s^{0,1})^2 + (U_s^2 - U_s^1)^2 - \int_s^T \Big(C_\epsilon ((U_t^{0,2} - U_t^{0,1})^2 + (U_t^2 - U_t^1)^2) \\ &+ \epsilon |D_{x_0}(U^{0,2} - U^{0,1})|^2 + 2\epsilon \int_{\mathbb{R}^d} |D_x(U^2 - U^1)(t, X_t^0, y, m_t)|^2 m_t(dy)\Big) dt \\ &+ \int_s^T \Big(|D_{x_0}(U^{0,2} - U^{0,1})|^2 + |D_{x_0}(U^2 - U^1)|^2 + |D_x(U^2 - U^1)|^2\Big) \, dt \Big]. \end{aligned} \tag{39}$$

Note that, as m_t is the conditional law of X_t given $(W_u^0)_{u \leqslant t}$, we have

$$\begin{aligned} \mathbb{E}\Big[|D_x(U^2 - U^1)(t, X_t^0, X_t, m_t)|^2\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[|D_x(U^2 - U^1)(t, X_t^0, X_t, m_t)|^2 \mid (W_u^0)_{u \leqslant t}\Big]\Big] \\ &= \mathbb{E}\Big[\int_{\mathbb{R}^d} |D_x(U^2 - U^1)(t, X_t^0, y, m_t)|^2 m_t(dy)\Big] \end{aligned}$$

since X_t^0 and X_t are adapted to $(W_u^0)_{u \leqslant t}$. Plugging this relation into (39) we find therefore, for $\epsilon > 0$ small enough,

$$\begin{aligned} 0 \geqslant \mathbb{E}\Big[&(U_s^{0,2} - U_s^{0,1})^2 + (U_s^2 - U_s^1)^2 - \int_s^T \Big(C_\epsilon ((U_t^{0,2} - U_t^{0,1})^2 + (U_t^2 - U_t^1)^2) dt \\ &+ (1/2) \int_s^T |D_{x_0}(U^{0,2} - U^{0,1})|^2 + |D_{x_0}(U^2 - U^1)|^2 + |D_x(U^2 - U^1)|^2 \, dt \Big]. \end{aligned}$$

We conclude by Gronwall's inequality that, for any $t \in [t_0, T]$,

$$\mathbb{E}\Big[(U^{0,2}(t, X_t^0, m_t) - U^{0,1}(t, X_t^0, m_t))^2 + (U^2(t, X_t^0, X_t, m_t) - U^1(t, X_t^0, X_t m_t))^2\Big] = 0.$$

For $t = t_0$, we have therefore $U^{0,2}(t_0, \bar{x}_0, \bar{m}_0) = U^{0,1}(t_0, \bar{x}_0, \bar{m}_0)$ and

$$U^1(t_0, \bar{x}_0, Z, \bar{m}_0) = U^2(t_0, \bar{x}_0, Z, \bar{m}_0) \;\; a.s.$$

If $\bar{m}_0$ has a positive density, the fact that the law of Z is $\bar{m}_0$ easily implies the equality of U^1 and U^2 at any point $(t_0, \bar{x}_0, x, \bar{m}_0)$ for $x \in \mathbb{R}^d$. We conclude by density of such laws and by continuity of the U^i's. □

5 Estimates on the MFG system

We are now left to prove the estimates on the first order master equations considered in the two previous sections. As the solutions of these equations are built by a method of characteristics, where the

characteristics are the solutions of the MFG system, we first need to discuss the well-posedness and the regularity properties of this system:

$$\begin{cases} (i) & -\partial_t u(t,x) - \mathrm{Tr}(a(t,x)D^2u(t,x)) + H(x_0,x,Du(t,x),m(t)) = 0 \qquad \text{in } (t_0,T)\times\mathbb{R}^d, \\ (ii) & \partial_t m(t,x) - \sum_{i,j} D_{ij}(a_{i,j}(t,x)m(t,x)) - \mathrm{div}(m(t,x)H_p(x_0,x,Du(t,x),m(t))) = 0 \\ & \qquad\qquad \text{in } (t_0,T)\times\mathbb{R}^d, \\ (iii) & m(t_0) = m_0, \qquad u(T,x) = G(x_0,x,m(T)) \qquad \text{in } \mathbb{R}^d. \end{cases} \tag{40}$$

Here $x_0 \in \mathbb{R}^{d_0}$ is treated as a fixed parameter. We also present similar results for the corresponding linearized systems. These estimates are motivated by the construction and the regularity of the first order master equation in the next section.

Let us first explain the notion of solution to (40). Fix $(t_0, m_0) \in [0,T]\times\mathcal{P}_2$ and $x_0 \in \mathbb{R}^{d_0}$. We say that (u,m) is a solution to (40) if $u \in C^0([t_0,T],C_b^2)$ satisfies

$$u(t,x) = G(x_0,x,m(T)) + \int_t^T (\mathrm{Tr}(a(s,x)D^2u(s,x)) - H(x_0,x,Du(s,x),m(s)))ds \qquad \forall t\in[t_0,T]$$

and if $m \in C^0([t_0,T],\mathcal{P}_2)$ solves the Fokker-Planck equation in the sense of distributions: for any $\phi \in C_c^\infty([0,T)\times\mathbb{R}^d)$,

$$\begin{aligned} 0 = &\int_{\mathbb{R}^d} \phi(0,x)m_0(dx) \\ &+ \int_0^T \int_{\mathbb{R}^d} (\mathrm{Tr}(a(s,x)D^2\phi(s,x)) - D\phi(s,x)\cdot H_p(x_0,x,Du(s,x),m(s)))m(s,dx)ds\,. \end{aligned}$$

The assumptions on a, H and G given in Subsection 2.3 are in force throughout the section.

5.1 Well-posedness and regularity of the MFG system

We discuss here the well-posedness of the MFG system (40) and provide several estimates. Let us start with the Hamilton-Jacobi (HJ) equation (general estimates on this equation are given in the Appendix A).

Proposition 5.1. *For any $M>0$, there exist $T_M>0$, $L_M>0$, depending on C_0 and γ given in assumptions (14) and (15), such that, if $\sup_{x_0,m}\|G(x_0,\cdot,m)\|_1 \leqslant M$, then, for any $T\in(0,T_M)$ and any $m\in C^0([0,T],\mathcal{P}_2)$, the solution u to the HJ equation*

$$\begin{cases} -\partial_t u(t,x) - \mathrm{Tr}(a(t,x)D^2u(t,x)) + H(x_0,x,Du(t,x),m(t)) = 0 \qquad \text{in } (t_0,T)\times\mathbb{R}^d \\ u(T,x) = G(x_0,x,m(T)) \qquad \text{in } \mathbb{R}^d \end{cases} \tag{41}$$

satisfies

$$\sup_{t\in[t_0,T]}\|u\|_1 \leqslant \sup_{x_0,m}\|G(x_0,\cdot,m)\|_1 + L_M T\,.$$

Henceforth, we set $K_M := \sup_{x_0,m}\|G(x_0,\cdot,m)\|_1 + L_M T_M$.

If, in addition, $\sup_{x_0,m}\|G(x_0,\cdot,m)\|_n \leqslant M$, then there exists $C_M>0$, depending on n, C_0, γ and

$$\sup_{t\in[0,T_M]}\|a(t)\|_n + \sup_{|p|\leqslant K_M, x_0\in\mathbb{R}^{d_0}, m\in\mathcal{P}_2}\sum_{k=0}^n \|D^k_{(x,p)}H(x_0,\cdot,p,m)\|_\infty,$$

such that u satisfies, for any $T\in(0,T_M)$, $x_0\in\mathbb{R}^{d_0}$ and $r\leqslant n$,

$$\sup_{t\in[t_0,T],x\in\mathbb{R}^d}|D_x^r u(t,x)| \leqslant \sup_{x\in\mathbb{R}^d}|D_x^r G(x_0,x,m(T))| + C_M T.$$

Therefore, for any $x_0\in\mathbb{R}^{d_0}$,

$$\sup_{t\in[t_0,T]}\|u(t)\|_n \leqslant \sup_m \|G(x_0,\cdot,m)\|_n + C_M T. \tag{42}$$

Proof. Use Propositions A.1 and A.6. □

Next we discuss the dependence of the solution u of (41) with respect to $(m(t))_{t\in[t_0,T]}$ and $x_0 \in \mathbb{R}^{d_0}$. We stress that, hereafter, we currently use the preliminary gradient estimate $\sup_{t\in[t_0,T_M]} \|u(t)\|_1 \leqslant K_M$ which is obtained as a first step in Proposition 5.1. In particular, the Hamiltonian $H(x_0,x,p,m)$ will be systematically estimated for $|p| \leqslant K_M$.

Proposition 5.2. *If the assumptions of Proposition 5.1 are satisfied so that* (42) *holds true, then there exists $T_M > 0$ such that, for $T \in (0,T_M)$ and any $t_0 \in [0,T]$, for any $m^1, m^2 \in C^0([0,T],\mathcal{P}_2)$ and any $x_0^1, x_0^2 \in \mathbb{R}^{d_0}$, if u^1 and u^2 are the corresponding solutions to the HJ equation* (41)*, then we have, for $n \geqslant 2$,*

$$\begin{aligned}\sup_{t\in[t_0,T]} \|u^1(t)-u^2(t)\|_{n-1} &\leqslant C_M\, T\big(\sup_{t\in[t_0,T]} \mathbf{d}_2(m^1(t),m^2(t)) + |x_0^1-x_0^2|\big)\\ &\quad + (1+C_MT)\big\{[\mathrm{Lip}_{0,n-1}(G)]\mathbf{d}_2(m^1(T),m^2(T)) + [\mathrm{Lip}_{n-1}^{x_0}(G)]|x_0^1-x_0^2|\big\}\end{aligned}$$

where C_M depends on the same quantities as in Proposition 5.1 as well as on $\mathrm{Lip}_{n-1,n}(H(x_0,\cdot,\cdot,m))$, $\mathrm{Lip}_{n-1,n}^{x_0}(H(x_0,\cdot,\cdot,m))$ (for $x \in \mathbb{R}^d$ and $|p| \leqslant K_M$) and $\sup_{x_0,m} \|G(x_0,\cdot,m)\|_n$.

Proof. The map $v := u^1 - u^2$ satisfies

$$\begin{cases} -\partial_t v - \mathrm{Tr}(a(t,x)D^2v) + V(t,x)\cdot Dv + f(t,x) = 0\\ v(T,x) = G(x_0^1,x,m^1(T)) - G(x_0^2,x,m^2(T))\end{cases}$$

where $V(t,x) = \displaystyle\int_0^1 H_p(x,x_0^2,sDu^1(t,x)+(1-s)Du^2(t,x),m^2(t))ds$ and

$$f(t,x) := H(x_0^1,x,Du^1(t,x),m^1(t)) - H(x_0^2,x,Du^1(t,x),m^2(t)).$$

By Proposition A.7 (applied with $k=1$ and $n-1$), we have

$$\begin{aligned}\sup_{t\in[0,T]} \|u^1(t)-u^2(t)\|_{n-1} &\leqslant (1+CT)\|G(x_0^1,\cdot,m^1(T)) - G(x_0^2,\cdot,m^2(T))\|_{n-1} + CT\sup_{t\in[t_0,T]}\|f(t)\|_{n-1}\\ &\leqslant (1+CT)\big\{[\mathrm{Lip}_{0,n-1}(G)]\mathbf{d}_2(m^1(T),m^2(T)) + [\mathrm{Lip}_{n-1}^{x_0}(G)]|x_0^1-x_0^2|\big\}\\ &\quad + CT\big(\sup_{t\in[t_0,T]} \mathbf{d}_2(m^1(t),m^2(t)) + |x_0^1-x_0^2|\big),\end{aligned}$$

where the constant C depends on H and on $\sup_{t\in[0,T]}\|V(t)\|_{n-1}$, hence on $\sup_{t\in[0,T]}[\|u^1(t)\|_n, \sup_{t\in[0,T]}\|u^2(t)\|_n]$, which are estimated thanks to Proposition 5.1. □

In our next step, we consider the solution to the Fokker-Planck equation

$$\begin{cases} \partial_t\tilde m(t,x) - \displaystyle\sum_{i,j} D_{ij}(a_{i,j}(t,x)\tilde m(t,x)) - \mathrm{div}(\tilde m(t,x)H_p(x_0,x,Du(t,x),m(t))) = 0 \text{ in } (t_0,T)\times\mathbb{R}^d\\ \tilde m(t_0) = m_0 \qquad \text{in } \mathbb{R}^d\end{cases} \tag{43}$$

where $(m(t))_{t\in[t_0,T]}$ is given and u satisfies (41). Let us recall that, under the assumptions of Proposition 5.1, there exists a unique weak solution $\tilde m \in C^0([t_0,T],\mathcal{P}_2)$ to (43).

Proposition 5.3. *Assume that*

$$\|D_xG\|_\infty \leqslant M\,, \quad \|D_{xx}^2G\|_\infty \leqslant M\,, \quad \mathrm{Lip}_{0,1}(G) + \mathrm{Lip}_1^{x_0}(G) \leqslant M\,. \tag{44}$$

Then there exists a constant $C_M > 0$, only depending on M, $\|a\|_2$ and the regularity of H, such that, for any $m^1, m^2 \in C^0([0,T],\mathcal{P}_2)$, $x_0^1, x_0^2 \in \mathbb{R}^{d_0}$ and $m_0^1, m_0^2 \in \mathcal{P}_2$, if u^1 and u^2 are the corresponding solutions to the HJ equation (41) *with $x_0 = x_0^i$ and if $\tilde m_1, \tilde m_2$ are the corresponding solutions to* (43) *starting from m_0^1 and m_0^2 respectively, then*

$$\sup_{t\in[t_0,T]} \mathbf{d}_2^2(\tilde m^1(t),\tilde m^2(t)) \leqslant (1+C_MT)\mathbf{d}_2^2(m_0^1,m_0^2) + C_MT\left(\sup_{t\in[t_0,T]} \mathbf{d}_2^2(m^1(t),m^2(t)) + |x_0^1-x_0^2|^2\right).$$

Proof. We can represent $\tilde{m}^i(t)$ as the law of X^i_t where $\mathbb{E}[|X^1_0 - X^2_0|^2] = \mathbf{d}_2^2(m^1_0, m^2_0)$ and X^i solves

$$X^i_t = X^i_0 - \int_0^t H_p(x^i_0, X^i_s, Du^i(s, X^i_s), m^i(s))ds + \sqrt{2}\int_0^t \sigma(s, X^i_s)dB_s,$$

so that

$$\begin{aligned}
\mathbb{E}\left[|X^1_t - X^2_t|^2\right] &\leqslant \mathbb{E}\left[|X^1_0 - X^2_0|^2\right] \\
&+ 2\mathbb{E}\left[\int_0^t (X^1_s - X^2_s)\cdot\left(H_p(x^1_0, X^1_s, Du^1, m^1(t)) - H_p(x^2_0, X^2_s, Du^2, m^2(t))\right)ds\right] \\
&+ \mathbb{E}\left[\int_0^t \mathrm{Tr}\left((\sigma(s, X^1_s) - \sigma(s, X^2_s))(\sigma(s, X^1_s) - \sigma(s, X^2_s))^*\right)ds|\right] \leqslant \mathbb{E}\left[|X^1_0 - X^2_0|^2\right] \\
&\quad + C_M\mathbb{E}\left[\int_0^t (|X^1_s - X^2_s|^2 + |D(u^1 - u^2)(s, X^1_s)|^2 + \mathbf{d}_2^2(m^1(s), m^2(s)) + |x^1_0 - x^2_0|^2)ds\right]
\end{aligned}$$

where C_M depends on the Lipschitz regularity of H_p in $\mathbb{R}^{d_0}\times\mathbb{R}^d\times B(K_M)\times\mathcal{P}_2$ (where K_M is defined in Proposition 5.1), on $\sup_t \|u^1(t)\|_2$, and on the Lipschitz regularity of σ. We infer from Gronwall's Lemma that

$$\begin{aligned}
\mathbb{E}\left[|X^1_t - X^2_t|^2\right] &\leqslant (1 + C_M T)\mathbb{E}\left[|X^1_0 - X^2_0|^2\right] \\
&\quad + C_M T\left(\sup_t \|D(u^1 - u^2)(t)\|^2_\infty + \sup_{t\in[t_0,T]} \mathbf{d}_2^2(m^1(t), m^2(t)) + |x^1_0 - x^2_0|^2\right).
\end{aligned}$$

As $\mathbb{E}[|X^1_0 - X^2_0|^2] = \mathbf{d}_2^2(m^1_0, m^2_0)$ and $\mathbf{d}_2^2(\tilde{m}^1(t), \tilde{m}^2(t)) \leqslant \mathbb{E}\left[|X^1_t - X^2_t|^2\right]$, we obtain:

$$\begin{aligned}
\sup_{t\in[t_0,T]} \mathbf{d}_2^2(\tilde{m}^1(t), \tilde{m}^2(t)) &\leqslant (1 + C_M T)\mathbf{d}_2^2(m^1_0, m^2_0) \\
&\quad + C_M T\left(\sup_t \|D(u^1 - u^2)(t)\|^2_\infty + \sup_{t\in[t_0,T]} \mathbf{d}_2^2(m^1(t), m^2(t)) + |x^1_0 - x^2_0|^2\right).
\end{aligned}$$

We estimate the term $\sup_t \|D(u^1 - u^2)(t)\|^2_\infty$ by Proposition 5.2 (with $n = 2$): since $\mathrm{Lip}_{0,1}(G)$ and $\mathrm{Lip}_1^{x_0}(G)$ are estimated by (44), we deduce, for some (possibly different) constant C_M:

$$\sup_{t\in[t_0,T]} \mathbf{d}_2^2(\tilde{m}^1(t), \tilde{m}^2(t)) \leqslant (1 + C_M T)\mathbf{d}_2^2(m^1_0, m^2_0) + C_M T\left(\sup_{t\in[t_0,T]} \mathbf{d}_2^2(m^1(t), m^2(t)) + |x^1_0 - x^2_0|^2\right).$$

□

Collecting the estimates in Propositions 5.1, 5.2 and 5.3 yields the well-posedness of the MFG system and estimates on the solution:

Proposition 5.4. *Fix $M > 0$ and assume that (44) holds true and that $\|G\|_n \leqslant M$ holds. Then there exists $T_M, C_M > 0$, depending on M, n, C_0, γ and*

$$\sup_{t\in[0,T_M]} \|a(t)\|_n + \sup_{|p|\leqslant K_M, x_0\in\mathbb{R}^{d_0}, m\in\mathcal{P}_2} \sum_{k=0}^n \|D^k_{(x,p)} H(x_0, \cdot, p, m)\|_\infty,$$

(where K_M is given in Proposition 5.1) such that, for any $T \in (0, T_M)$, for any $(t_0, m_0) \in [0, T]\times\mathcal{P}_2$, there exists a unique solution to the MFG system (40). This solution satisfies

$$\sup_{t\in[t_0,T]} \|u(t)\|_n \leqslant \|G(x_0, \cdot, m(T))\|_n + C_M T.$$

Moreover, if (t_0, m^1_0) and (t_0, m^2_0) are two initial conditions in $[0, T]\times\mathcal{P}_2$ and $x^1_0, x^2_0 \in \mathbb{R}^{d_0}$, and if (u^1, m^1) and (u^2, m^2) are the corresponding solutions to the MFG system (40) with $x_0 = x^1_0$ and $x_0 = x^2_0$ respectively, then

$$\sup_{t\in[t_0,T]} \mathbf{d}_2(m^1(t), m^2(t)) \leqslant (1 + C_M T)\mathbf{d}_2(m^1_0, m^2_0) + C_M T|x^1_0 - x^2_0|,$$

and

$$\sup_{t\in[t_0,T]}\|u^1(t)-u^2(t)\|_{n-1}\leqslant C_MT\big(\mathbf{d}_2(m_0^1,m_0^2)+|x_0^1-x_0^2|\big)$$
$$+(1+C_MT)\big\{[\mathrm{Lip}_{0,n-1}(G)](\mathbf{d}_2(m_0^1,m_0^2)+|x_0^1-x_0^2|)+[\mathrm{Lip}_{n-1}^{x_0}(G)]|x_0^1-x_0^2|\big\}.$$

Proof. The existence and uniqueness result come from a standard fixed point argument on $C^0([t_0,T],\mathcal{P}_2)$ for T small enough (say $T\leqslant T_M$ where $C_MT_M\leqslant 1/2$, C_M being given by the previous Propositions). For the stability with respect to the initial condition, one first uses the estimate in Proposition 5.3 with $\tilde m^i=m^i$:

$$\sup_{t\in[t_0,T]}\mathbf{d}_2^2(m^1(t),m^2(t))\leqslant(1+C_MT)\mathbf{d}_2^2(m_0^1,m_0^2)+C_MT\left(\sup_{t\in[t_0,T]}\mathbf{d}_2^2(m^1(t),m^2(t))+|x_0^1-x_0^2|^2\right).$$

Thus, as $C_MT\leqslant 1/2$, one obtains

$$\sup_{t\in[t_0,T]}\mathbf{d}_2(m^1(t),m^2(t))\leqslant(1+C_MT)\mathbf{d}_2(m_0^1,m_0^2)+C_MT|x_0^1-x_0^2|,$$

modifying C_M if necessary. Plugging this estimate into the estimate for the u^i in Proposition 5.2 gives the result. □

5.2 The first order linearized system

Next we consider the linearized system

$$\begin{cases}
(i)\quad -\partial_t v-\mathrm{Tr}(a(t,x)D^2v)+H_p(x_0,x,Du,m(t))\cdot Dv+\dfrac{\delta H}{\delta m}(x_0,x,Du,m(t))(\rho(t))=R_1(t,x)\\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{in } (t_0,T)\times\mathbb{R}^d\\
(ii)\quad \partial_t\rho-\displaystyle\sum_{i,j}D_{ij}(a_{i,j}\rho)-\mathrm{div}(\rho H_p(x_0,x,Du,m(t)))-\mathrm{div}(mH_{pp}Dv)\\
\qquad\qquad\qquad\qquad\qquad\qquad -\mathrm{div}(m\dfrac{\delta H_p}{\delta m}(\rho))=\mathrm{div}(R_2(t,x))\qquad\text{in } (t_0,T)\times\mathbb{R}^d\\
(iii)\quad \rho(t_0)=\rho_0,\qquad v(T,x)=\dfrac{\delta G}{\delta m}(x_0,x,m(T))(\rho(T))+R_3(x)\qquad\text{in }\mathbb{R}^d
\end{cases}\tag{45}$$

where (u,m) solves (40) and H and its derivatives are evaluated at $(x_0,x,Du(t,x),m(t))$. In this section, we work under the conditions given in Proposition 5.4 so that (40) admits a unique solution, in particular we always assume that $T\leqslant T_M$, where T_M is given by Proposition 5.4. Our goal now is to establish estimates for (v,ρ) in dependence of G and u; we implicitly assume that $G(x_0,\cdot,m)$ is sufficiently regular (say, C_b^n) so that u inherits the same order of regularity (from (42)).

The data of equation (45) are $x_0\in\mathbb{R}^{d_0}$, $\rho_0\in C^{-k}$, $R_1\in C^0([0,T],C_b^{n-1})$, $R_2\in C^0([0,T],C^{-(k-1)})$ and $R_3\in C_b^{n-1}$. Here $n\geqslant 2$ and $k\geqslant 1$. By a solution to (45), we mean a pair (v,ρ) such that $v\in C^0([0,T],C_b^{n-1})$ satisfies (45)-(i) (integrated in time) with terminal condition $v(T,\cdot)=\frac{\delta G}{\delta m}(x_0,\cdot,m(T))(\rho(T))+R_3(\cdot)$ and $\rho\in C^0([0,T],C_b^{-(k-1)})$ is a solution in the sense of distributions to (45)-(ii) with initial condition $\rho(t_0)=\rho_0$.

Proposition 5.5. *Let us fix $M>0$, $n\geqslant 2$ and $k\geqslant 1$. Under the assumptions of Proposition 5.4, and if*

$$\|\frac{\delta G}{\delta m}\|_{1;k}\leqslant M,\tag{46}$$

then there exist constants $T_M,C_M>0$, depending on M, n, k, $\sup_{t\in[0,T]}\|u(t)\|_n$, $\sup_{t\in[0,T]}\|u(t)\|_{k+1}$, such that for $T\leqslant T_M$ there exists a unique solution (v,ρ) to (45), and this solution satisfies

$$\begin{aligned}
&\sup_{t\in[t_0,T]}\|v(t)\|_{n-1}\ \leqslant\\
&(1+C_MT)\|\frac{\delta G}{\delta m}(x_0,\cdot_x,m(T),\cdot_y)\|_{n-1;k}\left(\|\rho_0\|_{-k}+T\sup_t\|R_2(t)\|_{-(k-1)}+T\sup_t\|R_1(t)\|_1\right)\\
&\quad+(1+C_MT)\|R_3\|_{n-1}+C_MT\left(1+\sup_t\|R_1(t)\|_{n-1}+\|R_2\|_{-(k-1)}\right),
\end{aligned}\tag{47}$$

as well as

$$\begin{aligned}\sup_{t\in[t_0,T]}\|\rho(t)\|_{-k} \;\leqslant\;& (1+C_MT)\,\|\rho_0\|_{-k}\\ &+C_MT(\sup_t\|R_1(t)\|_1+\sup_t\|R_2(t)\|_{-(k-1)}+\|R_3\|_{n-1}).\end{aligned}\tag{48}$$

Moreover, we have, for any $r\leqslant n-1$,

$$\begin{aligned}\sup_{t\in[0,T]}\|D_x^rv(t)\|_\infty\leqslant(1+C_MT)&\left(\Big\|D_x^r\frac{\delta G}{\delta m}(x_0,\cdot,m(T))(\rho(T))\Big\|_\infty+\|D_x^rR_3\|_\infty\right)\\ &+C_MT\big(\|\rho_0\|_{-k}+\sup_t\|R(t)\|_{n-1}+\sup_t\|R_2(t)\|_{-(k-1)}+\|R_3\|_{n-1}\big).\end{aligned}\tag{49}$$

Proof. After proving the a priori estimates, the existence of a solution can be obtained using a continuation argument (see [9] for details). The uniqueness is an obvious consequence of the estimates. So it remains to prove the estimates. To simplify the expression, we omit the dependence of the constant C with respect to M. Fix $t_1\in[t_0,T]$, $z_1\in C_b^k$ for $k\in\{1,\dots,n-1\}$. Let z be the solution to

$$\begin{cases}-\partial_tz-\mathrm{Tr}(a(t,x)D^2z)+H_p(x_0,x,Du,m(t))\cdot Dz=0 & \text{in } (t_0,t_1)\times\mathbb{R}^d,\\ z(t_1,\cdot)=z_1(x) & \text{in } \mathbb{R}^d.\end{cases}\tag{50}$$

According to Proposition A.7 (with $k=1$), we have

$$\sup_{t\in[t_0,t_1]}\|z(t)\|_k\leqslant(1+CT)\|z_1\|_k,$$

where C depends on the regularity of a and H and on $\sup_t\|u(t)\|_{k+1}$. Then, by duality,

$$\begin{aligned}\int_{\mathbb{R}^d}z_1\rho(t_1) \;=\;& \int_{\mathbb{R}^d}z(t_0)\rho_0-\int_{t_0}^{t_1}\int_{\mathbb{R}^d}(H_{pp}Dv\cdot Dz+\frac{\delta H_p}{\delta m}(\rho)\cdot Dz)m-\int_{t_0}^{t_1}\int_{\mathbb{R}^d}Dz\cdot R_2\\ \leqslant\;& \|z(t_0)\|_k\|\rho_0\|_{-k}+C\|Dz\|_\infty\left(T\|Dv\|_\infty+\int_{t_0}^{t_1}\|\rho(t)\|_{-k}\right)+T\sup_t\|z(t)\|_k\|R_2\|_{-(k-1)}\\ \leqslant\;& (1+CT)\|z_1\|_k\left(\|\rho_0\|_{-k}+C\left(T\|Dv\|_\infty+\int_{t_0}^{t_1}\|\rho(t)\|_{-k}\right)+T\|R_2\|_{-(k-1)}\right),\end{aligned}$$

where $\|R_2\|_{-(k-1)}:=\sup_t\|R_2(t)\|_{-(k-1)}$. Thus, taking the supremum over $\|z_1\|_k\leqslant1$, we obtain:

$$\|\rho(t_1)\|_{-k}\leqslant(1+CT)\|\rho_0\|_{-k}+CT\big(\|Dv\|_\infty+\|R_2\|_{-(k-1)}\big)+C\int_{t_0}^{t_1}\|\rho(t)\|_{-k}.$$

Since this holds for all $t_1\in(t_0,T]$, by Gronwall's inequality we obtain

$$\sup_{t\in[t_0,T]}\|\rho(t)\|_{-k}\leqslant(1+CT)\|\rho_0\|_{-k}+CT\big(\|Dv\|_\infty+\|R_2\|_{-(k-1)}\big).\tag{51}$$

Next we apply Proposition A.7 (with $k=1$) to the HJ equation satisfied by v: we have, for any $r\leqslant n-1$,

$$\sup_t\|v(t)\|_r\;\leqslant\;(1+CT)\|v(T)\|_r+CTC_1,\tag{52}$$

where C depends on $\sup_t\|a(t)\|_{n-1}$, on the regularity of H, on $\sup_t\|u(t)\|_n$, and where C_1 is estimated by

$$C_1=\sup_t\|\frac{\delta H}{\delta m}(x_0,\cdot,Du(t,\cdot),m(t))(\rho(t))\|_{n-1}+\|R_1\|_{n-1}\leqslant C\sup_t\|\rho(t)\|_{-k}+\|R_1\|_{n-1}\,,\tag{53}$$

where we used the inequality

$$\|\frac{\delta H}{\delta m}(x_0,\cdot,Du(t,\cdot),m(t))(\rho(t))\|_{n-1}\leqslant\|\frac{\delta H}{\delta m}(x_0,\cdot,Du(t,\cdot_x),m(t),\cdot)\|_{n-1,k}\|\rho(t)\|_{-k}\,.$$

Again we notice here that the right-hand side is estimated through the regularity of H and $\sup_t \|u(t)\|_n$. Similarly we estimate, for $r \leqslant n-1$,

$$\|v(T)\|_r \leqslant \|\frac{\delta G}{\delta m}(x_0, \cdot, m(T))\|_{r;k} \sup_t \|\rho(t)\|_{-k} + \|R_3\|_r. \tag{54}$$

Collecting the estimates in (51), (52), (53), (54), we find, for $r \leqslant n-1$:

$$\begin{aligned}\sup_t \|v(t)\|_r \leqslant & (1+CT)\|\frac{\delta G}{\delta m}(x_0, \cdot, m(T), \cdot)\|_{r;k}\Big\{(1+CT)\|\rho_0\|_{-k} + CT(\|Dv\|_\infty + \|R_2\|_{-(k-1)})\Big\} \\ & + \|R_3\|_r(1+CT) + CT\big(\|\rho_0\|_{-k} + T(\|Dv\|_\infty + \|R_2\|_{-(k-1)}) + \|R_1\|_{n-1}\big).\end{aligned} \tag{55}$$

We first consider this inequality for $r=1$. Recall that $\|\frac{\delta G}{\delta m}\|_{1;k} \leqslant M$. So, if we choose $T_M > 0$ such that

$$(1+CT_M)MCT_M + CT_M^2 < 1,$$

we obtain (47) for $T \leqslant T_M$ and $n=2$. Then from (51) we infer (48) (with a constant only depending on $\sup_t \|u(t)\|_{k+1}$). Having now estimated $\|Dv\|_\infty$, we deduce from (55) that (47) holds.

To obtain (49), we apply again Proposition A.7 to the HJ equation satisfied by v, together with estimates (53) and (48). □

5.3 The second order linearized system

Next we study the second order linearization of the MFG system. Given (u,m) a solution to (40) and (v,ρ) and (v',ρ') two solutions to (45) with arbitrary R_1, R_2, R_3 and R_1', R_2', R_3', we consider a solution (w,μ) to

$$\begin{cases}
(i) & -\partial_t w - \mathrm{Tr}(aD^2 w) + H_p \cdot Dw + \dfrac{\delta H}{\delta m}(\mu(t)) \\
& \quad + \dfrac{\delta^2 H}{\delta m^2}(\rho(t), \rho'(t)) + H_{pp} Dv \cdot Dv' + \dfrac{\delta H_p}{\delta m}(\rho) \cdot Dv' + \dfrac{\delta H_p}{\delta m}(\rho') \cdot Dv = \tilde R_1(t,x) \\
& \qquad\qquad \text{in } (t_0,T)\times\mathbb{R}^d \\
(ii) & \partial_t \mu - \displaystyle\sum_{i,j} D_{ij}(a_{i,j}\mu) - \mathrm{div}(\mu H_p) - \mathrm{div}(m H_{pp} Dw) - \mathrm{div}(m \dfrac{\delta H_p}{\delta m}(\mu)) - \mathrm{div}(\rho H_{pp} Dv') \\
& \quad - \mathrm{div}(\rho \dfrac{\delta H_p}{\delta m}(\rho')) - \mathrm{div}(\rho' H_{pp} Dv) - \mathrm{div}(m H_{ppp} Dv Dv') - \mathrm{div}(m \dfrac{\delta H_{pp}}{\delta m}(\rho') Dv) \\
& \quad - \mathrm{div}(\rho' \dfrac{\delta H_p}{\delta m}(\rho)) - \mathrm{div}(m \dfrac{\delta H_{pp}}{\delta m}(\rho) Dv') - \mathrm{div}(m \dfrac{\delta^2 H_p}{\delta m^2}(\rho,\rho')) = \mathrm{div}(\tilde R_2(t,x)) \\
& \qquad\qquad \text{in } (t_0,T)\times\mathbb{R}^d \\
(iii) & \mu(t_0) = 0,\ w(T,x) = \dfrac{\delta^2 G}{\delta m^2}(x_0, x, m(T))(\rho(T), \rho'(T)) \\
& \qquad + \dfrac{\delta G}{\delta m}(x_0, x, m(T))(\mu(T)) + \tilde R_3(x) \qquad \text{in } \mathbb{R}^d
\end{cases} \tag{56}$$

where H and its derivatives are evaluated at $(x_0, x, Du(t,x), m(t))$. Here again we work under the conditions assumed in the previous Sections which guarantee the existence, uniqueness and enough regularity for (u,m) as well as for the solutions of the linearized system. In particular, we always assume that $T \leqslant T_M$, where T_M is now given by Proposition 5.5. The goal now is to establish estimates for (w,μ) in terms of G as well as of (u,m) and (v,ρ), (v',ρ').

The data of the problem are $\tilde R_1 \in C^0([0,T], C_b^{n-2})$, $\tilde R_2 \in C^0([0,T], C^{-(k-1)})$ and $\tilde R_3 \in C_b^{n-2}$. By a solution to (56), we mean a pair (w,μ) such that $w \in C^0([0,T], C_b^{n-2})$ satisfies (56)-(i) (integrated in time) with the terminal condition in (56)-(iii) and $\mu \in C^0([0,T], C^{-k})$ solves (56)-(ii) in the sense of distributions with vanishing initial condition. Here we assume $n \geqslant 3$ and $k \geqslant 2$; the reason for this condition is just because we wish to keep the regularity threshold of (w,μ) consistent with what stated previously for (u,m) and for (v,ρ). In general, the estimates below apply to any degree of k,n but this is obviously a cascade regularity: an estimate of w in C_b^{n-2} requires an estimate of v in C_b^{n-1} and of u in C_b^n, while an estimate of μ in C^{-k} requires an estimate of ρ in $C^{-(k-1)}$.

Proposition 5.6. *Let us fix $M > 0$, $n \geqslant 3$ and $k \geqslant 1$. Under the assumptions of Proposition 5.4, and if* (46) *holds, there exist $T_M > 0$, depending on M and the regularity of H, such that for any $T \in (0, T_M]$, system* (56) *has a unique solution which satisfies*

$$\begin{aligned}
&\sup_t \|w(t)\|_{n-2} \\
&\quad \leqslant (1 + C_M T)\Big(\|\frac{\delta^2 G}{\delta m^2}(x_0, \cdot, m(T), \cdot, \cdot)\|_{n-2;k-1,k-1} \|\rho(T)\|_{-(k-1)} \|\rho'(T)\|_{-(k-1)} + \|\tilde R_3\|_{n-2} \Big) \\
&\quad + C_M T (1 + \|\frac{\delta G}{\delta m}\|_{n-2,k}) \Big(\sup_t \|\tilde R_1(t)\|_{n-2} + \sup_t \|\tilde R_2(t)\|_{-(k-1)} \\
&\qquad\qquad + \mathcal{R}_{k-1,k} \mathcal{R}'_{k-1,k} + \mathcal{R}_{k-1,n-1} \mathcal{R}'_{k-1,n-1} \Big)
\end{aligned} \tag{57}$$

for some C_M depending on M, on the regularity of H as well as on $n, k, \sup_{t\in[0,T]} \|u\|_{n-1}, \sup_{t\in[0,T]} \|u\|_{k+1}$, and

$$\begin{aligned}
\sup_t \|\mu(t)\|_{-k} \leqslant \tilde C_M T \Big(&\Big(1 + \|\frac{\delta^2 G}{\delta m^2}(x_0, \cdot, m(T), \cdot, \cdot)\|_{1;k-1,k-1} \Big) \|\rho(T)\|_{-(k-1)} \|\rho'(T)\|_{-(k-1)} \\
&+ \sup_{t\in[0,T]} \|\tilde R_1(t)\|_1 + \sup_{t\in[0,T]} \|\tilde R_2(t)\|_{-(k-1)} + \|\tilde R_3\|_1 + \mathcal{R}_{k-1,k} \mathcal{R}'_{k-1,k} + \mathcal{R}_{k-1,2} \mathcal{R}'_{k-1,2} \Big) ,
\end{aligned} \tag{58}$$

where $\tilde C_M$ depends on M, the regularity of H, $n, k, \sup_{t\in[0,T]} \|u\|_{k+1}$, and where we have set, for $k, j \geqslant 1$:

$$\mathcal{R}_{k-1,j} := \sup_t (\|\rho(t)\|_{-(k-1)} + \|v(t)\|_j) , \qquad \mathcal{R}'_{k-1,j} := \sup_t (\|\rho'(t)\|_{-(k-1)} + \|v'(t)\|_j) .$$

In addition, if

$$\|\frac{\delta G}{\delta m}\|_{n-2;k} \leqslant M,$$

then we have, for any $r \leqslant n-2$, $(t, x_0) \in [0, T] \times \mathbb{R}^{d_0}$,

$$\begin{aligned}
\|D^r w(t, \cdot)\|_\infty \leqslant &\Big(\Big\| D_x^r \frac{\delta^2 G}{\delta m^2}(x_0, \cdot, m(T))(\rho(T), \rho'(T)) \Big\|_\infty + \|D_x^r \tilde R_3(\cdot)\|_\infty \Big) \\
&+ C_M T \Big(\|\frac{\delta^2 G}{\delta m^2}(x_0, \cdot, m(T), \cdot, \cdot)\|_{n-2;k-1,k-1} \|\rho(T)\|_{-(k-1)} \|\rho'(T)\|_{-(k-1)} \\
&\quad + \sup_t \|\tilde R_1(t)\|_{n-2} + \sup_t \|\tilde R_2(t)\|_{-(k-1)} + \|\tilde R_3\|_{n-2} + \mathcal{R}_{k-1,k} \mathcal{R}'_{k-1,k} + \mathcal{R}_{k-1,n-1} \mathcal{R}'_{k-1,n-1} \Big).
\end{aligned} \tag{59}$$

Remark 5.7. *We recall that the quantities $\|\rho(T)\|_{-(k-1)}$ and $\mathcal{R}_{k-1,j}$ are estimated from* (47) *and* (48)*. In particular, we have*

$$\mathcal{R}_{k-1,k} \leqslant (1 + C_M T) C \left(\|\rho_0\|_{-(k-1)} + \sup_t \|R_2(t)\|_{-(k-2)} + \sup_t \|R_1(t)\|_k + \|R_3\|_k \right) ,$$

for some constant C depending on $\|\frac{\delta G}{\delta m}\|_{k;k-1}$ and $\sup_t \|u(t)\|_{k+1}$, and similarly

$$\mathcal{R}_{k-1,n-1} \leqslant (1 + C_M T) C \left(\|\rho_0\|_{-(k-1)} + \sup_t \|R_2(t)\|_{-(k-2)} + \|R_3\|_{n-1} + \sup_t \|R_1(t)\|_{n-1} \right)$$

for a constant C depending on $\|\frac{\delta G}{\delta m}\|_{n;k-1}$ and $\sup_t \|u(t)\|_n$. Of course the same holds for ρ', v' accordingly.

Proof. We omit the proof of the well-posedness of the system, which is a consequence of the estimates (as for Proposition 5.5). To simplify the expression, we also omit the dependence of the constant C with respect to M. We first estimate μ by duality. Fix $t_1 \in [t_0, T]$, $z_1 \in C_b^k$ for $k \in \{1, \dots, n-1\}$. Let z be the solution to (50). Recall that Proposition A.7 (with $k = 1$) implies that there is a constant $C > 0$, depending on $\sup_t \|u(t)\|_{k+1}$, such that

$$\sup_{t\in[t_0,t_1]} \|z(t)\|_k \leqslant (1 + CT) \|z_1\|_k.$$

Then

$$\int_{\mathbb{R}^d}\mu(t_1)z_1 = -\Big\{\int_{t_0}^{t_1}\int_{\mathbb{R}^d} Dz\cdot\Big(mH_{pp}Dw + m\frac{\delta H_p}{\delta m}(\mu) + \rho H_{pp}Dv' + \rho' H_{pp}Dv \\ +\rho\frac{\delta H_p}{\delta m}(\rho') + \rho'\frac{\delta H_p}{\delta m}(\rho) + mH_{ppp}DvDv' \\ +m\frac{\delta H_{pp}}{\delta m}(\rho')Dv + m\frac{\delta H_{pp}}{\delta m}(\rho)Dv' + m\frac{\delta^2 H_p}{\delta m^2}(\rho,\rho') + \tilde R_2(t,x)\Big)\Big\}.$$

Hence

$$\int_{\mathbb{R}^d}\mu(t_1)z_1 \leqslant CT\|Dw\|_\infty\|Dz\|_\infty + C\|Dz\|_\infty\int_{t_0}^{t_1}\|\mu(s)\|_{-k}ds \\ +CT\Big(\sup_t\|\rho(t)\|_{-(k-1)}\sup_t\|v'(t)\|_k + \sup_t\|\rho'(t)\|_{-(k-1)}\sup_t\|v(t)\|_k\Big)\sup_t\|z(t)\|_k \\ +CT\Big(\sup_t\|\rho(t)\|_{-(k-1)}\sup_t\|\rho'(t)\|_{-(k-1)}\Big)\sup_t\|z(t)\|_k + CT\,\|Dv\|_\infty\|Dv'\|_\infty\|Dz\|_\infty \\ +CT\Big(\sup_t\|\rho(t)\|_{-(k-1)}\|Dv'\|_\infty + \sup_t\|\rho'(t)\|_{-(k-1)}\|Dv\|_\infty + \sup_t\|\rho(t)\|_{-(k-1)}\sup_t\|\rho'(t)\|_{-(k-1)}\Big)\|Dz\|_\infty \\ +CT\,\|\tilde R_2\|_{-(k-1)}\sup_t\|z(t)\|_k\,,$$

where the constant C depends on the regularity of the function H and on $\sup_t\|u(t)\|_k$. Taking the supremum over $\|z_1\|_k\leqslant 1$, we infer that:

$$\|\mu(t_1)\|_{-k} \leqslant C\int_{t_0}^{t_1}\|\mu(s)\|_{-k}ds + CT\Big\{\|Dw\|_\infty + \|\tilde R_2\|_{-(k-1)} \\ + \Big(\sup_t\|\rho(t)\|_{-(k-1)} + \sup_t\|v(t)\|_k\Big)\Big(\sup_t\|\rho'(t)\|_{-(k-1)} + \sup_t\|v'(t)\|_k\Big)\Big\}.$$

By Gronwall's inequality, we obtain

$$\sup_t\|\mu(t)\|_{-k} \leqslant CT\Big\{\|Dw\|_\infty + \sup_t\|\tilde R_2(t)\|_{-(k-1)} + \mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k}\Big\}, \tag{60}$$

where C depends on the regularity of the function H and on $\sup_t\|u(t)\|_{k+1}$. From Proposition A.7 (with $k=1$), we have

$$\sup_t\|w(t)\|_{n-2} \leqslant (1+CT)\Big(\|\frac{\delta^2 G}{\delta m}(\rho(T),\rho'(T))\|_{n-2} + \|\frac{\delta G}{\delta m}(\mu(T))\|_{n-2} + \|\tilde R_3\|_{n-2}\Big) + CT\sup_t\|f(t)\|_{n-2}, \tag{61}$$

where

$$f(t,x) \;=\; \frac{\delta H}{\delta m}(\mu(t)) + \frac{\delta^2 H}{\delta m^2}(\rho(t),\rho'(t)) + H_{pp}Dv\cdot Dv' + \frac{\delta H_p}{\delta m}(\rho)\cdot Dv' + \frac{\delta H_p}{\delta m}(\rho')\cdot Dv - \tilde R_1(t,x).$$

We estimate

$$\sup_t\|f(t)\|_{n-2} \;\leqslant\; \Big(\|\frac{\delta H}{\delta m}(x_0,\cdot_x,Du(t,\cdot_x),m(t),\cdot_y)\|_{n-2;k}\sup_t\|\mu(t)\|_{-k} + \|\tilde R_1\|_{n-2} \\ +C\sup_t(\|\rho(t)\|_{-(k-1)} + \|v(t)\|_{n-1})(\|\rho'(t)\|_{-(k-1)} + \|v'(t)\|_{n-1})\Big)$$

for a constant C depending on the regularity of H and on $\sup_t\|u(t)\|_{n-1}$. So we conclude, using also (60),

$$\sup_t\|f(t)\|_{n-2} \;\leqslant\; C\,T\Big(\|Dw\|_\infty + \mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k} + \|\tilde R_2\|_{-(k-1)}\Big) \\ +\sup_t\|\tilde R_1\|_{n-2} + C\,\mathcal{R}_{k-1,n-1}\mathcal{R}'_{k-1,n-1}\,.$$

Similarly, again from (60) we get

$$\|\frac{\delta G}{\delta m}(\mu(T))\|_{n-2} \leqslant CT\,\|\frac{\delta G}{\delta m}(x_0,\cdot,m(T),\cdot)\|_{n-2;k}\Big(\|Dw\|_\infty + \mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k} + \|\tilde R_2\|_{-(k-1)}\Big)$$

and

$$\|\frac{\delta^2 G}{\delta m^2}(\rho(T),\rho'(T))\|_{n-2} \leqslant \|\frac{\delta^2 G}{\delta m^2}(x_0,\cdot,m(T),\cdot,\cdot)\|_{n-2;k-1,k-1}\|\rho(T)\|_{-(k-1)}\|\rho'(T)\|_{-(k-1)}\,.$$

Then, we find

$$\begin{aligned}\sup_t \|w(t)\|_{n-2} \leqslant\ & (1+CT)\left(\|\frac{\delta^2 G}{\delta m^2}(x_0,\cdot,m(T),\cdot,\cdot)\|_{n-2;k-1,k-1}\|\rho(T)\|_{-(k-1)}\|\rho'(T)\|_{-(k-1)}+\|\tilde R_3\|_{n-2}\right)\\ &+CT\,(1+\|\frac{\delta G}{\delta m}\|_{n-2;k})\left(\|Dw\|_\infty+\|\tilde R_2\|_{-(k-1)}+\mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k}\right)\\ &+CT\left(\|\tilde R_1\|_{n-2}+\mathcal{R}_{k-1,n-1}\mathcal{R}'_{k-1,n-1}\right)\end{aligned}$$

where now the constant C depends on both $\sup_t \|u(t)\|_{k+1}$ and $\sup_t \|u(t)\|_{n-1}$.

For $n=3$, if we choose T small enough (depending on $\|\frac{\delta G}{\delta m}\|_{1,k}$ and $\sup_t \|u(t)\|_2$) we estimate $\|Dw\|_\infty$. Then, plugging this estimate into (60) gives (58) (with a constant only depending on $\sup_t \|u(t)\|_{k+1}$). Finally, we deduce (57) for $n>3$.

For any $r \leqslant n-2$, $x_0 \in \mathbb{R}^{d_0}$ and $t\in[0,T]$, the estimate (59) on $D_x^r w$ follows again from Proposition A.7 (with $k=1$), that gives, arguing as before,

$$\begin{aligned}\|D_x^r w(t,\cdot)\|_\infty &\leqslant (1+CT)\Big(\|D_x^r\frac{\delta^2 G}{\delta m^2}(\rho(T),\rho'(T))\|_\infty+\|D_x^r\frac{\delta G}{\delta m}(\mu(T))\|_\infty+\|D_x^r\tilde R_3\|_\infty\Big)\\ &\quad+CT\sup_t\|f(t)\|_{n-2}\\ &\leqslant \Big(\|D_x^r\frac{\delta^2 G}{\delta m^2}(\rho(T),\rho'(T))\|_\infty+\|D_x^r\tilde R_3\|_\infty\Big)+(1+CT)\|\frac{\delta G}{\delta m}(x_0,\cdot,m(t),\cdot)\|_{n-2;k}\sup_t\|\mu(t)\|_{-k}\\ &\quad+CT\Big(\|\tilde R_1\|_{n-2}+\|\tilde R_2\|_{-(k-1)}+\|\tilde R_3\|_{n-2}+\|\frac{\delta^2 G}{\delta m^2}(x_0,\cdot,m(T),\cdot,\cdot)\|_{n-2;k-1,k-1}\|\rho(T)\|_{-(k-1)}\|\rho'(T)\|_{-(k-1)}\\ &\qquad+\mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k}+\mathcal{R}_{k-1,n-1}\mathcal{R}'_{k-1,n-1}\Big),\end{aligned}$$

that yields the desired claim using (58). □

By gathering together Proposition 5.5 and Proposition 5.6, we deduce the following three corollaries, which will be useful in the derivation of second order estimates for the solution of the master equation.

Corollary 5.8. *Let $M>0$, $n\geqslant 3$ and $k\in\{2,\dots,n-1\}$, and assume that*

$$\|G\|_n+\|\frac{\delta G}{\delta m}\|_{n-1;k}+\|\frac{\delta^2 G}{\delta m^2}\|_{n-2;k-1,k-1}\leqslant M\,.$$

Let (u,m) be the unique solution to (40) in some interval $[0,T_M]$ given by Proposition 5.4, and let (v,ρ) and (v',ρ') be two solutions to (45) with $R_1=R_2=R_3=0$ and initial conditions ρ_0, ρ'_0 respectively.

Then there exists a constant C_M such that the solution (w,μ) to (56) corresponding to (u,m), (v,ρ) and (v',ρ') and with $\tilde R_1=\tilde R_2=\tilde R_3=0$ satisfies, for any $T\in(0,T_M)$, $r\leqslant n-2$:

$$\sup_{t,x}|D_x^r w(t,x)|\leqslant \sup_x\Big|D_x^r\frac{\delta^2 G}{\delta m^2}(x_0,x,m(T))(\rho(T),\rho'(T))\Big|+C_M T\|\rho_0\|_{-(k-1)}\|\rho'_0\|_{-(k-1)}$$

where C_M depends on M, as well as on $\|a\|_n$ and the regularity of H.

Proof. We first notice that

$$\mathrm{Lip}_{0,1}(G)\leqslant\sup_{x_0,m}\|\frac{\delta G}{\delta m}(x_0,\cdot,m,\cdot)\|_{1,1}\leqslant M$$

hence we are in the position to apply Proposition 5.4, and there exists a time $T_M>0$ such that the unique solution (u,m) to (40) satisfies $u\in C_b^n$ with an estimate depending on M and $\sup_{x_0,m}\|G(x_0,\cdot,m)\|_n$.

From Proposition 5.6, we have

$$\begin{aligned}\|D_x^r w(t,\cdot)\|_\infty &\leqslant \Big\|D_x^r \tfrac{\delta^2 G}{\delta m^2}(x_0,\cdot,m(T))(\rho(T),\rho'(T))\Big\|_\infty \\ &\quad + C_M T\Big(\|\rho(T)\|_{-(k-1)}\|\rho'(T)\|_{-(k-1)} + \mathcal{R}_{k-1,k}\mathcal{R}'_{k-1,k} + \mathcal{R}_{k-1,n-1}\mathcal{R}'_{k-1,n-1}\Big).\end{aligned}$$

On the other hand, we know from Proposition 5.5 that

$$\begin{aligned}\sup_t \|v(t)\|_{n-1} &\leqslant (1+C_M T)\|\tfrac{\delta G}{\delta m}\|_{n-1;k-1}\,\|\rho_0\|_{-(k-1)} \\ \sup_t \|\rho(t)\|_{-(k-1)} &\leqslant (1+C_M T)\,\|\rho_0\|_{-(k-1)}\,,\end{aligned}$$

which allows us to estimate $\mathcal{R}_{k-1,k}$ and $\mathcal{R}_{k-1,n-1}$. Here the constant depends on $\sup_t \|u(t)\|_n$. A similar estimate holds for (v',ρ'). Therefore, we conclude the desired estimate. □

Corollary 5.9. *Under the assumptions of Corollary 5.8, suppose in addition that*

$$\|D_{x_0}\frac{\delta G}{\delta m}\|_{n-2;k-1} \leqslant M.$$

Let (u,m) be the unique solution to (40) in $[0,T_M]$, let (v,ρ) be a solution to (45) with $R_1=R_2=R_3=0$ and initial condition ρ_0, and, for any $|l|=1$, $l\in\mathbb{R}^{d_0}$, (v^l,ρ^l) be a solution to (45) with zero initial condition and with

$$\begin{aligned}R_1(t,x) &= -\partial^l_{x_0}H(y_0,x,Du(t,x),m(t)) \\ R_2(t,x) &= m(t,x)\partial^l_{x_0}H_p(y_0,x,Du(t,x),m(t)) \\ R_3(t,x) &= \partial^l_{x_0}G(y_0,x,m(T)).\end{aligned} \tag{62}$$

Then there exists a constant C_M such that the solution (w^l,μ^l) to (56) corresponding to (u,m), (v,ρ) and (v^l,ρ^l) and with

$$\begin{aligned}\tilde R_1(t,x) &= -\partial^l_{x_0}H_p(x_0,x,Du,m(t))Dv - \partial^l_{x_0}\frac{\delta H}{\delta m}(x_0,x,Du,m(t))(\rho(t)), \\ \tilde R_2(t,x) &= \rho\partial^l_{x_0}H_p(x_0,x,Du,m(t)) + m\partial^l_{x_0}H_{pp}(x_0,x,Du,m(t))Dv + m\partial^l_{x_0}\frac{\delta H_p}{\delta m}(\rho), \\ \tilde R_3(x) &= \partial^l_{x_0}\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho(T)),\end{aligned} \tag{63}$$

satisfies, for any $T\in(0,T_M)$, $r\leqslant n-2$,

$$\sup_{t,x}\Big(\sum_{|l|=1}|D_x^r w^l(t,x)|^2\Big)^{1/2} \leqslant \sup_x\Big|D_x^r D_{x_0}\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho(T))\Big| + C_M T\|\rho_0\|_{-(k-1)},$$

where C_M depends on M, as well as on $\|a\|_n$ and the regularity of H.

Proof. We first notice that

$$\sup_t\|\tilde R_1(t)\|_{n-2} + \sup_t\|\tilde R_2(t)\|_{-(k-1)} \leqslant C\sup_t\big(\|v(t)\|_{n-1} + \|\rho(t)\|_{-(k-1)}\big)$$

for a constant depending on the regularity of H, on $\sup_t\|u(t)\|_{n-1}$ and on $\sup_t\|u(t)\|_k$. However, the latter term is bounded by $\sup_t\|u(t)\|_{n-1}$ since $k\leqslant n-1$. Next we estimate the terms (v,ρ), (v^l,ρ^l) and μ^l: we have, from Proposition 5.5 and Proposition 5.6,

$$\begin{aligned}\sup_t\|v(t)\|_{n-1} &\leqslant (1+C_M T)\|\tfrac{\delta G}{\delta m}\|_{n-1;k-1}\,\|\rho_0\|_{-(k-1)} \leqslant C_M\|\rho_0\|_{-(k-1)} \\ \sup_t\|\rho(t)\|_{-(k-1)} &\leqslant (1+C_M T)\,\|\rho_0\|_{-(k-1)}\,,\end{aligned}$$

and

$$\sup_t \|v^l(t)\|_{n-1} \leqslant \Big\{(1+C_MT)\|\frac{\delta G}{\delta m}\|_{n-1;k-1} + C_MT\Big\} \leqslant C_M,$$
$$\sup_t \|\rho^l(t)\|_{-(k-1)} \leqslant C_MT, \qquad \|\mu^l\|_{-k} \leqslant C_MT\|\rho_0\|_{-(k-1)}.$$

We note that the w^l solve linear equations with the same diffusion and the same drift. So, combining Proposition A.7 with the inequalities above and arguing as in the proof of Proposition 5.6 gives, for any $r \leqslant n-2$,

$$\sup_x\Big(\sum_{|l|=1}|D_x^r w^l(t,x)|^2\Big)^{1/2} \leqslant$$
$$(1+CT)\sup_x\Big(\sum_{|l|=1}\Big(|D_x^r\frac{\delta^2 G}{\delta m^2}(\rho(T),\rho^l(T))| + |D_x^r\frac{\delta G}{\delta m}(\mu^l(T))| + |D_x^r\partial_{x_0}^l\frac{\delta G}{\delta m}(\rho^l(T))|\Big)^2\Big)^{1/2} + C_MT\|\rho_0\|_{-(k-1)}$$
$$\leqslant \sup_x\Big(\sum_{|l|=1}\Big(|D_x^r\partial_{x_0}^l\frac{\delta G}{\delta m}(\rho^l(T))| + C_MT\|\rho_0\|_{-(k-1)}\Big)^2\Big)^{1/2} + C_MT\|\rho_0\|_{-(k-1)},$$

where we have omitted the dependence of G with respect to $(x_0, x, m(T))$. This gives the result. □

Corollary 5.10. *Under the assumptions of Corollary 5.9, suppose in addition that $\|D_{x_0}^2 G(x_0,\cdot,m)\|_{n-2} \leqslant M$. Fix $l, l' \in \mathbb{N}^{d_0}$ with $|l| = |l'| = 1$. Let (u,m) be the unique solution to (40) in $[0,T_M]$ and let (v^l,ρ^l), $(v^{l'},\rho^{l'})$ be the solution to (45) with zero initial condition and with R_1, R_2, R_3 and R_1', R_2', R_3' given by (62) for l and l' respectively.*

Let $(w^{l,l'},\mu^{l,l'})$ be the solution to (56) corresponding to (u,m), (v^l,ρ^l) and $(v^{l'},\rho^{l'})$ and with

$$\begin{aligned}
\tilde R_1^{l,l'}(t,x) &= -\Big(\partial_{x_0}^{l+l'}H + \partial_{x_0}^l H_p Dv^{l'} + \partial_{x_0}^{l'}H_p Dv^l + \partial_{x_0}^l\frac{\delta H}{\delta m}(\rho^{l'}(t)) + \partial_{x_0}^{l'}\frac{\delta H}{\delta m}(\rho^l(t))\Big)\\
\tilde R_2^{l,l'}(t,x) &= \rho^{l'}\partial_{x_0}^l H_p + \rho^l\partial_{x_0}^{l'}H_p + m(\partial_{x_0}^l H_{pp}Dv^{l'} + \partial_{x_0}^{l'}H_{pp}Dv^l)\\
&\quad + m(\partial_{x_0}^l\frac{\delta H_p}{\delta m}(\rho^{l'}) + \partial_{x_0}^{l'}\frac{\delta H_p}{\delta m}(\rho^l)) + m\partial_{x_0}^{l+l'}H_p\\
\tilde R_3^{l,l'}(t,x) &= \partial_{x_0}^{l+l'}G(x_0,x,m(T)) + D_{x_0}^l\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho^{l'}(T)) + \partial_{x_0}^{l'}\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho^l(T))\ ,
\end{aligned} \tag{64}$$

where H and its derivatives are computed at $(x_0,x,Du(t,x),m(t))$. Then there exists a constant C_M such that, for any $T \in (0,T_M)$, $r \leqslant n-2$:

$$\sup_{t,x}\Big(\sum_{l,l'}|D_x^r w^{l,l'}(t,x)|^2\Big)^{1/2} \leqslant \sup_x|D_x^r D_{x_0}^2 G(x_0,\cdot,m(T))| + C_MT,$$

where C_M depends on M, as well as on $\|a\|_n$ and the regularity of H.

Proof. We can estimate (v^l,ρ^l) and $(v^{l'},\rho^{l'})$ and $\mu^{l,l'}$—and therefore $\tilde R_1^{l,l'}$ and $\tilde R_2^{l,l'}$—exactly as in the previous Corollary. Moreover, as the $w^{l,l'}$ solve a HJ with the same diffusion and the same drift term, we can use Proposition A.7 to bound the sum $(\sum_{l,l'}|D_x^r w^{l,l'}(t,x)|^2)^{1/2}$:

$$\sup_{t,x}\Big(\sum_{l,l'}|D_x^r w^{l,l'}(t,x)|^2\Big)^{1/2} \leqslant \sup_x\Big(\sum_{l,l'}(|D_x^r\partial_{x_0}^{l+l'}G(x_0,\cdot,m(T))| + C_MT)^2\Big)^{1/2} + C_MT,$$

which gives the required estimate after rearranging. □

6 Estimates on the first order master equation

In this section, we complete our program by proving regularity results for the solutions of the various first order master equations encountered in the previous sections. We mainly consider the first order master

equation:

$$\begin{cases} -\partial_t U(t,x_0,x,m) - \mathrm{Tr}(a(t,x)D^2_{xx}U(t,x_0,x,m)) + H(x_0,x,D_xU(t,x_0,x,m),m) \\ \qquad - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}(a(t,y)D^2_{ym}U(t,x_0,x,m,y))\ m(dy) \\ \qquad + \displaystyle\int_{\mathbb{R}^d} D_mU(t,x_0,x,m,y)\cdot H_p(x_0,y,D_xU(t,x_0,y,m),m)\ m(dy) = 0 \text{ in } (0,T)\times\mathbb{R}^d\times\mathcal{P}_2 \\ U(T,x_0,x,m) = G(x_0,x,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2\,. \end{cases} \tag{65}$$

In the above equation, $x_0 \in \mathbb{R}^{d_0}$ is considered as a parameter. Our aim is to build a solution to this equation and study its regularity. The method for finding a solution to (65) is well-known: if we set

$$U(t_0,x_0,x,m_0) := u(t_0,x) \tag{66}$$

where (u,m) is the solution to (40), then U is a solution to (65).

In order to study the Major-Minor agents' problem, we also have to consider a linear master equation

$$\begin{cases} -\partial_t U^0 - \displaystyle\int_{\mathbb{R}^d} \mathrm{Tr}(a(t,y)D^2_{ym}U^0(t,x_0,m,y))\ m(dy) \\ \qquad + \displaystyle\int_{\mathbb{R}^d} D_mU^0(t,x_0,m,y)\cdot H_p(x_0,y,D_xU(t,x_0,y,m),m)dm(y) = 0 \\ U^0(T,x_0,m) = G^0(x_0,m) \qquad \text{in } \mathbb{R}^d\times\mathcal{P}_2, \end{cases} \tag{67}$$

where U is the solution to (65). In this case, we build the solution U^0 by the simple formula:

$$U^0(t_0,x_0,m_0) = G^0(x_0,m(T)), \tag{68}$$

where (u,m) is also the solution to (40).

Our aim is to show that, if G and G^0 are regular enough, then (65) and (67) have classical solutions, given by the above representation formulas. Moreover, we show that the regularity of these solutions only deteriorate linearly in time. This last point is the key result in order to build later solutions to the second order master equation and to the master equation for the Major-Minor agents' problem.

Throughout the section, the assumptions of Subsection 2.3 on a, H, G and G^0 are in force.

6.1 First order differentiability of U and U^0

Proposition 6.1. *For any $M > 0$, there exists $T_M > 0$ and $K_M > 0$, depending on C_0 and γ and $\|Da\|_\infty$, and there exists $C_M > 0$, depending also on n, $k \in \{2,\dots,n-1\}$, $\sup_t \|a(t)\|_n$ and the regularity of H such that, if*

$$\|G\|_n + \left\|\frac{\delta G}{\delta m}\right\|_{n-1;k} \leqslant M, \tag{69}$$

and if $T \in (0,T_M]$, then the map U defined by (66) is a classical solution to (65), and satisfies

$$\sup_{t\in[0,T]} \|U(t)\|_n \leqslant \|G\|_n + C_M T$$

Moreover, for any $|\alpha| \leqslant n-1$, $\partial_x^\alpha \frac{\delta U}{\delta m}$ is of class C^1 in m, and for $k \in \{2,\dots,n-1\}$,

$$\sup_{t\in[0,T]} \left\|\frac{\delta U}{\delta m}(t)\right\|_{n-1;k} \leqslant \left\|\frac{\delta G}{\delta m}\right\|_{n-1;k} + C_M T.$$

Remark 6.2. *We show in the proof the following representation:*

$$\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(t_0,x_0,x,m_0,y)\rho_0(dy) = v(t_0,x) \tag{70}$$

where (u,m) is the solution of the MFG system (40) and (v,ρ) is the solution of the linearized system (45) with right-hand side $R_1 = R_2 = R_3 = 0$ and with initial condition (t_0,ρ_0). Note that the normalization condition (9) is satisfied because, if one chooses $\rho_0 = m_0$, then $(v,\rho) = (0,m)$.

The proof relies on the following lemma, in which we also provide estimates to obtain later one the differentiability of U with respect to x_0.

Lemma 6.3. *Under the assumptions of Proposition 6.1, we fix $(t_0, m_0), (t_0, m_1) \in [0,T) \times \mathcal{P}_2$, $y_0, \xi \in \mathbb{R}^d$ with $|\xi| \leqslant 1$. Let (u,m) be the solution to (40) with $x_0 = y_0$ and with initial condition (t_0, m_0), and, for $h \in (0,1)$, let (u_h, m_h) be the solution to (40) with $x_0 = y_0 + \xi h$ and with initial condition $(t_0, (1-h)m_0 + hm_1)$. Let also (v,ρ) be the solution to (45) associated with (u,m), $x_0 = y_0$ and with*

$$\begin{aligned} R_1(t,x) &= -H_{x_0}(y_0, x, Du(t,x), m(t)) \cdot \xi \\ R_2(t,x) &= m(t,x) H_{x_0 p}(y_0, x, Du(t,x), m(t)) \cdot \xi \\ R_3(t,x) &= G_{x_0}(y_0, x, m(T)) \cdot \xi, \end{aligned} \tag{71}$$

and initial condition $(t_0, m_1 - m_0)$. Then there exists a constant C (independent of h) such that

$$\sup_{t \in [t_0,T]} \|u_h(t) - u(t) - hv(t)\|_{n-1} \leqslant Ch^2 \tag{72}$$

and

$$\sup_{t \in [t_0,T]} \|m_h(t) - m(t) - h\rho(t)\|_{-k} \leqslant Ch^2 \,. \tag{73}$$

Remark 6.4. *The goal of this Lemma is to identify the first order derivatives $\frac{\delta U}{\delta m}$ and $D_{x_0} U$. The constant C above will depend on the regularity of H and G, as well as on $\sup_{t \in [t_0,T]} \|u(t)\|_n$; however this is not detailed later since it will not be relevant; indeed, (72) and (73) are only used for letting $h \to 0$.*

Proof. We set

$$v_h(t,x) = u_h(t,x) - u(t,x) - hv(t,x), \quad \rho_h(t,x) = m_h(t,x) - m(t,x) - h\rho(t,x).$$

Then the pair (v_h, ρ_h) solves

$$\begin{cases} -\partial_t v_h - \mathrm{Tr}(a(t,x) D^2 v_h) + H_p(y_0, x, Du, m(t)) \cdot Dv_h + \dfrac{\delta H}{\delta m}(y_0, x, Du, m(t))(\rho_h(t)) = R_{h,1}(t,x) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{in } (t_0,T) \times \mathbb{R}^d \\ \partial_t \rho_h - \displaystyle\sum_{i,j} D_{ij}(a_{i,j} \rho_h) - \mathrm{div}(\rho_h H_p(x, Du, m(t)) - \mathrm{div}(m H_{pp}(x, Du, m(t)) Dv_h) \\ \qquad\qquad\qquad\qquad -\mathrm{div}(m \dfrac{\delta H_p}{\delta m}(x, Du, m(t))(\rho_h)) = \mathrm{div}(R_{h,2}(t,x)) \qquad \text{in } (t_0,T) \times \mathbb{R}^d \\ \rho_h(t_0) = 0, \qquad v_h(T,x) = \dfrac{\delta G}{\delta m}(x, m(T))(\rho_h(T)) + R_{h,3}(x) \qquad \text{in } \mathbb{R}^d \end{cases}$$

where

$$\begin{aligned} R_{h,1}(t,x) = -\Big(& H(y_0 + \xi h, x, Du_h(t,x), m_h(t)) - H(y_0, x, Du(t,x), m(t)) \\ & - H_p(y_0, x, Du(t,x), m(t)) \cdot D(u_h(t,x) - u(t,x)) - \frac{\delta H}{\delta m}(y_0, x, Du(t,x), m(t))(m_h(t) - m(t)) \\ & - hH_{x_0}(y_0, x, Du(t,x), m(t)) \cdot \xi \Big), \\ R_{h,2}(t,x) = {} & m_h(t,x) H_p(y_0 + \xi h, x, Du_h(t,x), m_h(t)) - m(t,x) H_p(y_0, x, Du(t,x), m(t)) \\ & - (m_h(t,x) - m(t,x)) H_p(y_0, x, Du(t,x), m(t)) \\ & - m(t,x) H_{pp}(y_0, x, Du(t,x), m(t)) D(u_h - u)(t,x) \\ & - hm(t,x) H_{x_0 p}(y_0, x, Du(t,x), m(t)) \cdot \xi \\ & - m(t,x) \frac{\delta H_p}{\delta m}(y_0, x, Du(t,x), m(t))(m_h(t) - m(t)), \\ R_{h,3}(x) = {} & G(y_0 + \xi h, x, m_h(T)) - G(y_0, x, m(T)) - \frac{\delta G}{\delta m}(y_0, x, m(T))(m_h(T) - m(T)) \\ & - hG_{x_0}(y_0, x, m(T)) \cdot \xi. \end{aligned}$$

Next we estimate $R_{h,1}$, $R_{h,2}$ and $R_{h,3}$. As

$$\begin{aligned}R_{h,1} = -\int_0^1 \Big\{&(H_p(x_\tau, x, p_\tau(t,x), m_\tau(t)) - H_p(y_0, x, Du(t,x), m(t))) \cdot D(u_h(t,x) - u(t,x)) \\ &+ (H_{x_0}(x_\tau, x, p_\tau(t,x), m_\tau(t)) - H_{x_0}(y_0, x, Du(t,x), m(t))) \cdot h\xi \\ &+ \int_{\mathbb{R}^d} (\frac{\delta H}{\delta m}(x_\tau, x, p_\tau(t,x), m_\tau(t), y) - \frac{\delta H}{\delta m}(y_0, x, Du(t,x), m(t), y))(m_h(t) - m(t))(dy)\Big\}\, d\tau\end{aligned}$$

where $x_\tau := (1-\tau)y_0 + \tau(y_0 + \xi h)$, $p_\tau := (1-\tau)Du(t,x) + \tau Du_h(t,x)$ and $m_\tau(t,x) := (1-\tau)m(t,x) + \tau m_h(t,x)$, we have

$$\|R_{h,1}(t)\|_{n-1} \leqslant C\big(\|u_h(t) - u(t)\|_n^2 + h^2 + \mathbf{d}_2^2(m_h(t), m(t))\big),$$

In the same way,

$$\begin{aligned}\|R_{h,3}\|_{n-1} &\leqslant C(\mathbf{d}_2^2(m_h(T), m(T)) + h^2) \\ &\leqslant C(\|u_h(T) - u(T)\|_n^2 + \mathbf{d}_2^2(m_h(T), m(T)) + h^2).\end{aligned}$$

Finally, for $k \geqslant 2$, we have

$$\begin{aligned}&\|R_{h,2}(t)\|_{-(k-1)} \\ &\quad= \sup_{\|\phi\|_{k-1}\leqslant 1} \int_{\mathbb{R}^d} \phi(t,x)\Big(H_p(x_0, x, Du_h(t,x), m_h(t)) - H_p(y_0, x, Du(t,x), m(t))\Big)(m_h(t,dx) - m(t,dx)) \\ &\quad+ \int_{\mathbb{R}^d} \phi(t,x)\Big(H_p(x_0, x, Du_h(t,x), m_h(t)) - H_p(y_0, x, Du(t,x), m(t)) \\ &\qquad- H_{x_0 p}(y_0, x, Du(t,x), m(t)) \cdot h\xi \\ &\qquad- H_{pp}(y_0, x, Du(t,x), m(t))D(u_h - u)(t,x) - \frac{\delta H_p}{\delta m}(y_0, x, Du(t,x), m(t))(m_h(t) - m(t))\Big) m(t,dx) \\ &\quad\leqslant C(\|u_h - u\|_2^2 + \mathbf{d}_2^2(m_h(t), m(t)) + h^2).\end{aligned}$$

By Proposition 5.5, there exist constants $T_M, C_M > 0$, depending on M, n, k, $\sup_{t\in[0,T]} \|u\|_n$, such that, if $T \leqslant T_M$ and if (69) holds, then

$$\begin{aligned}\sup_{t\in[0,T]} \|v_h(t)\|_{n-1} &\leqslant (1 + C_M T)\|R_{h,3}\|_{n-1} + C_M T \left(\sup_t \|R_{h,1}(t)\|_{n-1} + \sup_t \|R_{h,2}(t)\|_{-(k-1)}\right) \\ &\leqslant C\left(\sup_t \|u_h(t) - u(t)\|_n^2 + \sup_t \mathbf{d}_2^2(m_h(t), m(t)) + h^2\right).\end{aligned}$$

We then infer by Proposition 5.4 and the definition of v_h that

$$\sup_{t\in[t_0,T]} \|u_h(t) - u(t) - hv(t)\|_{n-1} \leqslant C(\mathbf{d}_2^2((1-h)m_0 + hm_1, m_0) + h^2) \leqslant Ch^2.$$

The estimate of ρ_h comes from Proposition 5.5 in the same way. □

Proof of Proposition 6.1. Proposition 5.4 and the representation formula (66) imply the estimate on $\|U(t,\cdot,m)\|_n$. Let us now show that the map U given by (66) is differentiable with respect to m. Fix $x_0 \in \mathbb{R}^{d_0}$, $(t_0, m_0), (t_0, m_1) \in [0,T) \times \mathcal{P}_2$, let (u, m), (u_h, m_h) and (v, ρ) be as in Lemma 6.3 with $\xi = 0$, so $R_1 = R_2 = R_3 = 0$. Then

$$\sup_{t\in[t_0,T]} \|u_h(t) - u(t) - hv(t)\|_{n-1} \leqslant o(h).$$

Taking $t = t_0$, this implies that

$$\|U(t_0, x_0, \cdot, (1-h)m_0 + hm_1) - U(t_0, x_0, \cdot, m_0) - hv(t_0, \cdot)\|_{n-1} \leqslant o(h).$$

So, if we choose $m_1 = \delta_y$ for a fixed $y \in \mathbb{R}^d$, we have just proved that the map $\hat{U}(h; m_0, y) = U(t_0, x_0, x, (1-h)m_0 + h\delta_y)$ has a derivative at $h = 0$ and that this derivative is given by $v(t_0, x)$. Note that the map $(m_0, y) \mapsto v(t_0, x; m_0, y)$ is continuous and bounded thanks to the estimates in Proposition 5.5 and the uniqueness of the solution. So we can apply Lemma B.1 which says that U is C^1 in m with

$$v(t_0, x) = \frac{\delta U}{\delta m}(t_0, x_0, x, m_0, y).$$

Then by linearity and continuity one easily checks that (70) and the normalization condition (9) hold. A similar argument applies to derivatives of $\frac{\delta U}{\delta m}$ with respect to x.

Next we check that U solves (65). Let us start with $m(t_0) = m_0$ with a smooth density. Then (u, m) is a classical solution and, as

$$U(t, x_0, x, m(t)) = u(t, x) \qquad \forall (t, x) \in [t_0, T] \times \mathbb{R}^d,$$

we have, for any $h > 0$ and in view of the equation for m:

$$\begin{aligned}
&u(t_0 + h, x) - u(t_0, x) = U(t_0 + h, x_0, x, m(t_0 + h)) - U(t_0, x_0, x, m(t_0)) \\
&= \int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(t_0 + h, x_0, x, m(t), y) \partial_t m(t, y) dy dt + U(t_0 + h, x_0, x, m(t_0)) - U(t_0, x_0, x, m(t_0)) \\
&= -\int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} D_m U(t_0 + h, x_0, x, m(t), y) \cdot H_p(x_0, y, D_x u(t, y), m(t)) m(t, y) dy dt \\
&\quad + \int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} \mathrm{Tr}(a(t, y) D^2_{ym} U(t, x_0, x, m, y))\ m(dy) dt \\
&\quad + U(t_0 + h, x_0, x, m(t_0)) - U(t_0, x_0, x, m(t_0)).
\end{aligned}$$

On the other hand, by the equation for u,

$$\begin{aligned}
u(t_0 + h, x) - u(t_0, x) &= \int_{t_0}^{t_0+h} \Big(-\mathrm{Tr}(a(t, x) D^2 u(t, x)) + H(x_0, x, Du(t, x), m(t)) \Big) dt \\
&= \int_{t_0}^{t_0+h} \Big(-\mathrm{Tr}(a(t, x) D^2_{xx} U(t, x_0, x, m(t))) + H(x_0, x, D_x U(t, x_0, x, m(t)), m(t)) \Big) dt.
\end{aligned}$$

So

$$\begin{aligned}
&U(t_0 + h, x_0, x, m_0) - U(t_0, x_0, x, m_0) \\
&= \int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} D_m U(t_0 + h, x_0, x, m(t), y) \cdot H_p(x_0, y, D_x U(t, x_0, y, m(t)), m(t)) m(t, y) dy dt \\
&\quad - \int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} \mathrm{Tr}(a(t, y) D^2_{ym} U(t, x_0, x, m, y))\ m(dy) dt \\
&\quad + \int_{t_0}^{t_0+h} \Big(-\mathrm{Tr}(a(t, x) D^2_{xx} U(t, x_0, x, m(t))) + H(x_0, x, D_x U(t, x_0, x, m(t)), m(t)) \Big) dt.
\end{aligned}$$

Therefore U has a time-derivative at (t_0, x_0, x, m_0) and

$$\begin{aligned}
\partial_t U(t_0, x_0, x, m_0) = &\int_{\mathbb{R}^d} D_m U(t_0, x_0, x, m_0, y) \cdot H_p(x_0, y, D_x U(t_0, x_0, y, m_0), m_0)) m_0(y) dy \\
&- \int_{\mathbb{R}^d} \mathrm{Tr}(a(t_0, y) D^2_{ym} U(t_0, x_0, x, m, y))\ m(dy) \\
&- \mathrm{Tr}(a(t_0, x) D^2_{xx} U(t_0, x_0, x, m_0)) + H(x_0, x, D_x U(t, x_0, x, m_0), m_0).
\end{aligned}$$

This shows that U satisfies (65) at any point (t_0, x_0, x, m_0) where m_0 has a smooth density. The general case can be treated by a density argument, since the right-hand side of the above equation is continuous in (t_0, x_0, x, m_0).

Let us now explain the estimates on $\frac{\delta U}{\delta m}$. In view of (69), (70) and Proposition 5.5, we have, for any $r \leqslant n-1$,

$$\begin{aligned}\Big\|D_x^r \frac{\delta U}{\delta m}(t_0, x_0, \cdot, m_0)(\rho_0)\Big\|_\infty &= \|D_x^r v(t_0, x_0, \cdot)\|_\infty \\ &\leqslant (1 + C_M T)\|D_x^r \frac{\delta G}{\delta m}(x_0, \cdot, m(T), \cdot)\|_{0;k}\|\rho_0\|_{-k} + C_M T\|\rho_0\|_{-k}.\end{aligned}$$

Taking the sup over ρ_0 with $\|\rho_0\|_{-k} \leqslant 1$, $x_0 \in \mathbb{R}^{d_0}$, summing over $r \leqslant n-1$ and then taking the sup over t, m gives the estimate on $\frac{\delta U}{\delta m}$. Notice that the estimate given by Proposition 5.5 depends on $\sup_t \|u(t)\|_n$ (we use here that $k \leqslant n-1$); but this latter term is estimated in terms of M only, because of Proposition 5.4 and since $\|G(x_0, \cdot, m)\|_n \leqslant M$. □

Proposition 6.5. *Under the assumptions of Proposition 6.1, let $M, T_M, C_M > 0$ be given accordingly. Assume, in addition, that $T \in (0, T_M]$ and*

$$\sup_{x_0,m} \big|G^0(x_0, m)\big| + \big|D_{x_0} G^0(x_0, m)\big| + \Big\|\frac{\delta G^0}{\delta m}(x_0, m, \cdot)\Big\|_{n-1;k} + \|D_{x_0} G(x_0, \cdot, m)\|_{n-1} \leqslant M. \tag{74}$$

Then, the map U^0 defined by (68) is a classical solution to (67). In addition, U^0 and U are differentiable with respect to x_0 and satisfy

$$\sup_t \Big\|(U^0, U)(t)\Big\|_n \leqslant \Big\|(G^0, G)\Big\|_n + C_M T. \tag{75}$$

$$\sup_t \Big\|D_{x_0}(U^0, U)(t)\Big\|_{n-1} \leqslant \Big\|(D_{x_0} G^0, D_{x_0} G)\Big\|_{n-1} + C_M T. \tag{76}$$

and

$$\sup_t \Big\|\frac{\delta(U^0, U)}{\delta m}(t)\Big\|_{n-1;k} \leqslant \Big\|\frac{\delta(G^0, G)}{\delta m}\Big\|_{n-1;k} + C_M T. \tag{77}$$

As we will see in the proof, it is possible to estimate U^0 and U separately. However we will need the specific form of the estimate in the analysis of the MFG problem with a major player.

Proof. Differentiability of U with respect to x_0 can be checked as for its differentiability with respect to m: let ξ be any unit vector of $\mathbb{R}^{d_0}$, (u, m), (u_h, m_h) and (v, ρ) be as in Lemma 6.3 with $m_1 = m_0$. Then, by Proposition 5.5 and the fact that

$$\sup_t \|R_1(t)\|_{n-1} + \sup_t \|R_2(t)\|_{-(k-1)} \leqslant C, \qquad \|R_3\|_{n-1} \leqslant \sup_{x_0,m} \|G_{x_0}(x_0, \cdot, m(T))\|_{n-1} \tag{78}$$

one has

$$\|U(t_0, x_0 + h\xi, \cdot, m_0) - U(t_0, x_0, \cdot, m_0) - hv(t_0, \cdot)\|_{n-1} \leqslant o(h),$$

and so

$$U_{x_0}(t_0, x_0, x, m_0) \cdot \xi = v(t_0, x). \tag{79}$$

To show the differentiability of U^0 with respect to m we proceed as in the proof of Proposition 6.1. Fix $x_0 \in \mathbb{R}^d$, $(t_0, m_0), (t_0, m_1) \in [0, T) \times \mathcal{P}_2$, let (u, m), (u_h, m_h) and (v, ρ) be as in Lemma 6.3 with $\xi = 0$, so $R_1 = R_2 = R_3 = 0$. Then

$$\sup_{t \in [t_0, T]} \|\rho_h(t)\|_{-k} \leqslant o(h),$$

where $\rho_h(t, x) = m_h(t, x) - m(t, x) - h\rho(t, x)$. This inequality and Proposition 5.4 imply

$$\begin{aligned}\Big|G^0(x_0, m_h(T)) - G^0(x_0, m(T)) - h\frac{\delta G^0}{\delta m}(x_0, m(T))(\rho(T))\Big| &\leqslant \Big|\frac{\delta G^0}{\delta m}(x_0, m(T))(\rho_h(T))\Big| \\ + \Big|\int_0^1 \int_{\mathbb{R}^d} \Big(\frac{\delta G^0}{\delta m}(x_0, (1-\tau)m(T) + \tau m_h(T), y) - \frac{\delta G^0}{\delta m}(x_0, m(T), y)\Big)(m_h(t) - m(t))(dy)\, d\tau\Big| \\ \leqslant o(\mathbf{d}_2(m_h(T), m(T)) + h) \leqslant o(h).\end{aligned} \tag{80}$$

For $y \in \mathbb{R}^d$ choose now $m_1 = \delta_y$, then

$$\left|U^0(t_0, x_0, (1-h)m_0 + h\delta_y) - U^0(t_0, x_0, m_0) - h\frac{\delta G^0}{\delta m}(x_0, m(T))(\rho(T))\right| \leqslant o(h).$$

Note that $\rho_0 \mapsto \rho(T)$ is linear and continuous as a map from C^{-k} onto itself. Apply then Lemma B.1 to get that U^0 is C^1 in m with

$$\frac{\delta U^0}{\delta m}(t_0, x_0, m_0, y) = \frac{\delta G^0}{\delta m}(x_0, m(T))(\rho(T)). \tag{81}$$

Moreover, one can check as in the proof of Proposition 6.1 that U^0 solves (67) (here it is even simpler, and based on the fact that by definition of U^0, $U^0(t_0 + h, x_0, m(t_0 + h)) - U^0(t_0, x_0, m(t_0)) = 0$).

Concerning the differentiability of U^0 with respect to x_0, let ξ be any unit vector of $\mathbb{R}^d$, (u, m), (u_h, m_h) and (v, ρ) be as in Lemma 6.3 with $m_1 = m_0$. Then,

$$\begin{aligned}\Big|G^0(y_0 + \xi h, m_h(T)) - G^0(y_0, m(T)) - hG^0_{x_0}(y_0, m(T)) \cdot \xi - h\frac{\delta G^0}{\delta m}(y_0, m(T))(\rho(T))\Big| \leqslant \\ \left|G^0(y_0 + \xi h, m_h(T)) - G^0(y_0, m_h(T)) - hG^0_{x_0}(y_0, m_h(T)) \cdot \xi\right| + \\ h\left|G^0_{x_0}(y_0, m_h(T)) - G^0_{x_0}(y_0, m(T))\right| + \\ \left|G^0(y_0, m_h(T)) - G^0(y_0, m(T)) - h\frac{\delta G^0}{\delta m}(y_0, m(T))(\rho(T))\right|.\end{aligned}$$

The third term of this inequality can be treated as in (80). Therefore,

$$\left|U^0(t_0, y_0 + \xi h, m_0) - U^0(t_0, y_0, m_0) - hG^0_{x_0}(y_0, m(T)) \cdot \xi - h\frac{\delta G^0}{\delta m}(y_0, m(T))(\rho(T))\right| \leqslant o(h),$$

hence it follows that

$$D_{x_0}U^0(t_0, x_0, m_0) \cdot \xi = G^0_{x_0}(x_0, m(T)) \cdot \xi + \frac{\delta G^0}{\delta m}(x_0, m(T))(\rho(T)). \tag{82}$$

We now prove the estimates. By Proposition 5.1 and the representation formulas (66) and (68), we have, for any $x_0 \in \mathbb{R}^{d_0}$, $m \in \mathcal{P}_2$ and $r \leqslant n$,

$$\begin{aligned}|U^0(t, x_0, m)|^2 + |D^r_x U(t, x_0, x, m)|^2 &= |G^0(x_0, m(T))|^2 + |D^r_x u(t, x)|^2 \\ &\leqslant |G^0(x_0, m(T))|^2 + (\sup_x |D^r_x G(x_0, x, m(T))| + C_M T)^2 \\ &\leqslant \left(\left(|G^0(x_0, m(T))|^2 + \sup_x |D^r_x G(x_0, x, m(T)|^2\right)^{1/2} + C_M T\right)^2,\end{aligned}$$

(where we used that $x^2 + (y + z)^2 \leqslant ((x^2 + y^2)^{1/2} + z)^2$ for nonnegative reals x, y, z) which gives (75). Next we prove (76). For $|l| = 1$, $l \in \mathbb{N}^{d_0}$, we represent $\partial^l_{x_0}U^0$ and $\partial^l_{x_0}U$ by (82) and (79) respectively, where (v^l, ρ^l) is as in Lemma 6.3 with $\xi = e_l$, $m_1 = m_0$ (so that $\rho^l_0 = 0$). Then we have, for $r \leqslant n - 1$,

$$\begin{aligned}\sum_{|l|=1} |\partial^l_{x_0}U^0(t, x_0, m)|^2 + |D^r_x \partial^l_{x_0}U(t, x_0, x, m)|^2 \\ = \sum_{|l|=1} \left|\partial^l_{x_0}G^0(x_0, m(T)) + \frac{\delta G^0}{\delta m}(x_0, m(T))(\rho^l(T))\right|^2 + |D^r_x v^l(t, x)|^2\end{aligned}$$

Note that $\sup_t \|\rho(t)\|_{-k} \leqslant C_M T$ by Proposition 5.6. As the v^l solve HJ equations with the same diffusion

and the same drift, Proposition A.7, (74) and (78) imply that

$$\sup_x\Big(\sum_{|l|=1}|D^r v^l|^2\Big)^{1/2}\leqslant(1+CT)\sup_x\Big(\sum_{|l|=1}|D^r v^l(T)|^2\Big)^{1/2}+C_MT$$
$$\leqslant(1+CT)\sup_x\Big(\sum_{|l|=1}(\|D^r_x\frac{\delta G}{\delta m}(x_0,\cdot,m(T),\cdot)\|_{0;k}\|\rho(T)\|_{-k}+|D^r_x\partial^l_{x_0}G(x_0,x,m(T))|)^2\Big)^{1/2}+C_MT$$
$$\leqslant\sup_x\Big(\sum_{|l|=1}(|D^r_x\partial^l_{x_0}G(x_0,x,m(T))|+C_MT)^2\Big)^{1/2}+C_MT\leqslant\sup_x(\sum_{|l|=1}|D^r_x\partial^l_{x_0}G(x_0,x,m(T))|^2)^{1/2}+C_MT,$$

while

$$\sum_{|l|=1}|\partial^l_{x_0}U^0(t,x_0,m)|^2\leqslant\sum_{|l|=1}\Big(|\partial^l_{x_0}G^0(x_0,m(T))|+|\frac{\delta G^0}{\delta m}(x_0,m(T))(\rho^l(T))|\Big)^2$$
$$\leqslant\sum_{|l|=1}(|\partial^l_{x_0}G^0(x_0,m(T))+C_MT)^2\leqslant\Big(\big(\sum_{|l|=1}|\partial^l_{x_0}G^0(x_0,m(T))|^2\big)^{1/2}+C_MT\Big)^2.$$

Using that $((x+z)^2+(y+z)^2)^{1/2}\leqslant(x^2+y^2)^{1/2}+\sqrt{2}z$, we obtain

$$\sup_x\left(\sum_{|l|=1}|\partial^l_{x_0}U^0(t,x_0,m)|^2+|D^r_x\partial^l_{x_0}U(t,x_0,x,m)|^2\right)^{1/2}$$
$$\leqslant\sup_x\Big(\sum_{|l|=1}\big|\partial^l_{x_0}G^0(x_0,m(T))\big|^2+|D^r_x\partial^l_{x_0}G(x_0,x,m(T))|^2\Big)^{1/2}+C_MT,$$

from which we derive (76), by taking the sup over x_0, summing over r and finally taking the sup over m.

For (77), let (v,ρ) be as in Lemma 6.3 with $m_1-m_0=\rho_0\in C^{-k}$ and $\xi=0$, as in (81) and (70). We have, for any $r\leqslant n-1$,

$$\Big|\frac{\delta U^0}{\delta m}(t,x_0,m)(\rho_0)\Big|^2+\Big|D^r_x\frac{\delta U}{\delta m}(t,x_0,x,m)(\rho_0)\Big|^2=\Big|\frac{\delta G^0}{\delta m}(x_0,m(T))(\rho(T))\Big|^2+|D^r_xv(t,x)|^2.$$

So again by Proposition 5.5,

$$\Big|\frac{\delta U^0}{\delta m}(t,x_0,m)(\rho_0)\Big|^2+\Big|D^r_x\frac{\delta U}{\delta m}(t,x_0,x,m)(\rho_0)\Big|^2$$
$$\leqslant\Big|\frac{\delta G^0}{\delta m}(x_0,m(T))(\rho(T))\Big|^2+\Big(\sup_x\Big|D^r_x\frac{\delta G}{\delta m}(x_0,\cdot,m(T))(\rho(T))\Big|+C_MT\|\rho(T)\|_{-k}\Big)^2$$
$$\leqslant\left[\sup_x\frac{1}{\|\rho(T)\|_{-k}}\left(\Big|\frac{\delta G^0}{\delta m}(x_0,m(T))(\rho(T))\Big|^2+\Big|D^r_x\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho(T))\Big|^2\right)^{1/2}+C_MT\right]^2\|\rho(T)\|^2_{-k}$$
$$\leqslant(1+C_MT)^2\left[\sup_{x,\|\rho\|_{-k}=1}\left(\Big|\frac{\delta G^0}{\delta m}(x_0,m(T))(\rho)\Big|^2+\Big|D^r_x\frac{\delta G}{\delta m}(x_0,x,m(T))(\rho)\Big|^2\right)^{1/2}+C_MT\right]^2\|\rho_0\|^2_{-k},$$

This gives (77). □

6.2 Second order differentiability of U and U^0

Proposition 6.6. *Let U be the solution of (65) given by (66). Let $n\geqslant 3$ and $k\in\{2,\dots,n-1\}$. Suppose, in addition to the assumptions of Proposition 6.1, that G is of class C^2 and that $\|\frac{\delta^2G}{\delta m^2}(x_0,\cdot,m,\cdot,\cdot)\|_{n-2;k-1,k-1}\leqslant M$. Then there exists $T_M>0$ (depending on M and on the data but not on G) such that, if $T\in(0,T_M]$, the map U is C^2 with respect to the measure variable and the parameter x_0, and satisfies*

$$\sup_{t\in[0,T]}\|\frac{\delta^2U}{\delta m^2}(t)\|_{n-2;k-1,k-1}\leqslant\|\frac{\delta^2G}{\delta m^2}\|_{n-2;k-1,k-1}+C_MT.$$

Proof. Our first goal is to show that $\delta U/\delta m$ is differentiable with respect to m. Let $(t_0, m_0) \in [0,T) \times \mathcal{P}_2$, $y, y' \in \mathbb{R}^d$ and

- (u, m) (respectively (u_h, m_h)) be the solution of the MFG system (40) with initial condition (t_0, m_0) (respectively $(t_0, (1-h)m_0 + h\delta_{y'})$),
- (v, ρ) (respectively (v', ρ')) be the solution of the first order linearized system (45) with zero right-hand side, initial condition (t_0, δ_y) (respectively $(t_0, \delta_{y'})$) and where the Hamiltonian and its derivatives are evaluated at $(x_0, x, Du(t,x), m(t))$,
- $(\tilde{v}_h, \tilde{\rho}_h)$ be the solution to the first order linearized system (45) with zero right-hand side, with initial condition (t_0, δ_y) and where the Hamiltonian and its derivatives are evaluated at $(x_0, x, Du_h(t,x), m_h(t))$,
- (w, μ) be the solution to the second order linearized system (56) associated with (u, m), (v, ρ), (v', ρ') and with right-hand side 0.

Recall (see (70)) that

$$\begin{aligned} \tilde{v}_h(t_0, x) &= \frac{\delta U}{\delta m}(t_0, x_0, x, (1-h)m_0 + h\delta_{y'}, y), \\ v(t_0, x) &= \frac{\delta U}{\delta m}(t_0, x_0, x, m_0, y), \qquad \text{and} \quad v'(t_0, x) = \frac{\delta U}{\delta m}(t_0, x_0, x, m_0, y') \end{aligned} \tag{83}$$

so we expect $w(t_0, \cdot)$ to represent the derivative in m of $\delta U/\delta m$, namely $\frac{\delta^2 U}{\delta m^2}(t_0, x_0, x, m_0, y, y')$.

We consider

$$(\hat{v}_h, \hat{\rho}_h) := (\tilde{v}_h, \tilde{\rho}_h) - (v, \rho) - h(w, \mu).$$

Let us first note that, by Proposition 5.4, we have

$$\sup_{t \in [t_0, T]} \Big(\|\tilde{u}_h(t,x) - u(t,x)\|_{n-1} + \mathbf{d}_2(m_h(t), m(t)) \Big) \leqslant C \mathbf{d}_2((1-h)m_0 + h\delta_{y'}, m_0) \leqslant Ch. \tag{84}$$

Next we claim that

$$\sup_{t \in [t_0, T]} \|\tilde{v}_h(t,x) - v(t,x)\|_{n-2} + \|\tilde{\rho}_h(t) - \rho(t)\|_{-(k-1)} \leqslant Ch. \tag{85}$$

Indeed, the pair $(\tilde{v}_h, \tilde{\rho}_h) - (v, \rho)$ solves the first order linearized system (45), associated with (u, m), initial condition $(t_0, 0)$ and with a right-hand side given by

$$\begin{aligned} R_{h,1}(t,x) =& - \Big(\big(H_p(x_0, x, Du_h, m_h(t)) - H_p(x_0, x, Du, m(t)) \big) \cdot D\tilde{v}_h \\ &+ \big(\frac{\delta H}{\delta m}(x_0, x, Du_h, m_h(t)) - \frac{\delta H}{\delta m}(x_0, x, Du, m(t)) \big)(\tilde{\rho}_h(t)) \Big) \\ R_{h,2}(t,x) =& \tilde{\rho}_h (H_p(x_0, x, Du_h, m_h(t)) - H_p(x_0, x, Du, m(t))) \\ &+ (m_h H_{pp}(x_0, x, Du_h, m_h) - m H_{pp}(x_0, x, Du, m)) \cdot D\tilde{v}_h \\ &+ \Big(m_h \frac{\delta H_p}{\delta m}(x_0, x, Du_h, m_h) - m \frac{\delta H_p}{\delta m}(x_0, x, Du, m) \Big)(\tilde{\rho}_h) \\ R_{h,3}(t,x) =& \left(\frac{\delta G}{\delta m}(x_0, x, m_h(T)) - \frac{\delta G}{\delta m}(x_0, x, m(T)) \right)(\tilde{\rho}_h(T)). \end{aligned}$$

Applying Proposition 5.5 and using (84) we infer that (85) holds.

In view of the equations satisfied by $(\tilde{v}_h, \tilde{\rho}_h)$, (v, ρ) and (w, μ), the pair $(\hat{v}_h, \hat{\rho}_h)$ solves the first order linearized system (45), associated with (u, m), initial condition $(t_0, 0)$ and with

$$\begin{aligned} R_{h,1}(t,x) = - \Big[& \big(H_p(x_0, x, Du_h, m_h(t)) - H_p(x_0, x, Du, m(t)) \big) \cdot D\tilde{v}_h \\ & - h H_{pp}(x_0, x, Du, m(t)) Dv \cdot Dv' - h \frac{\delta H_p}{\delta m}(x_0, x, Du, m(t))(\rho'(t)) \cdot Dv \\ & + \Big(\frac{\delta H}{\delta m}(x_0, x, Du_h, m_h(t)) - \frac{\delta H}{\delta m}(x_0, x, Du, m(t)) \Big)(\tilde{\rho}_h(t)) \\ & - h \frac{\delta^2 H}{\delta m^2}(x_0, x, Du, m(t))(\rho(t), \rho'(t)) - h \frac{\delta H_p}{\delta m}(x_0, x, Du, m(t))(\rho(t)) \cdot Dv' \Big], \end{aligned}$$

$$R_{h,2}(t,x) = \tilde\rho_h\Big(H_p(x_0,x,Du_h,m_h(t)) - H_p\Big) - h\rho\Big(H_{pp}Dv' + \frac{\delta H_p}{\delta m}(\rho')\Big) +$$
$$D\tilde v_h\cdot\Big(m_h H_{pp}(x_0,x,Du_h,m_h(t)) - mH_{pp}\Big) - hDv\cdot\Big(\rho' H_{pp} + m\frac{\delta H_{pp}}{\delta m}(\rho') + mH_{ppp}Dv'\Big) +$$
$$\Big(m_h\frac{\delta H_p}{\delta m}(x_0,x,Du_h,m_h(t)) - m\frac{\delta H_p}{\delta m}\Big)(\tilde\rho_h) - h\Big(\rho'\frac{\delta H_p}{\delta m} + mDv'\cdot\frac{\delta H_{pp}}{\delta m}\Big)(\rho) - hm\frac{\delta^2 H_p}{\delta m^2}(\rho,\rho'),$$

$$R_{h,3}(x) = \frac{\delta G}{\delta m}(x_0,x,m_h(T))(\tilde\rho_h(T)) - \frac{\delta G}{\delta m}(x_0,x,m(T))(\tilde\rho_h(T)) - h\frac{\delta^2 G}{\delta m^2}(x_0,x,m(T))(\rho(T),\rho'(T))$$

(for brevity, H_p and its derivatives in $R_{h,2}$ are evaluated at $(x_0,x,Du,m(t))$, unless otherwise specified). Using

$$\sup_{t\in[t_0,T]}\|u_h(t)-u(t)-hv'(t)\|_{n-2}\leqslant Ch^2,\qquad \sup_{t\in[t_0,T]}\|m_h(t)-m(t)-h\rho'(t)\|_{-(k-1)}\ \leqslant Ch^2 \tag{86}$$

(see (72) and (73) in Lemma 6.3) as well as the above estimate (84), we have

$$\sup_t\big(\|R_{h,1}(t,\cdot)\|_{n-2}+\|R_{h,2}(t,\cdot)\|_{-(k-1)}+\|R_{h,3}(t,\cdot)\|_{n-2}\big)\leqslant Ch^2.$$

Then Proposition 5.5 and the representation formula (83) implies that

$$\begin{aligned}&\|\frac{\delta U}{\delta m}(t_0,x_0,\cdot,(1-h)m_0+h\delta_{y'},y) - \frac{\delta U}{\delta m}(t_0,x_0,\cdot,m_0,y) - hw(t_0,\cdot)\|_{n-2}\\ &\quad=\|\tilde v_h(t_0,\cdot)-v(t_0,\cdot)-hw(t_0,\cdot)\|_{n-2}\leqslant\sup_t\|\hat v_h(t)\|_{n-2}\leqslant Ch^2.\end{aligned}$$

Note that we also have

$$\sup_{t\in[t_0,T]}\|\tilde\rho_h(t)-\rho(t)-h\mu(t)\|_{-k}\ \leqslant Ch^2. \tag{87}$$

Hence, we can apply Lemma B.1 as in the proof of Proposition 6.1 and infer that $\delta U/\delta m$ has a derivative in m given by w:

$$\frac{\delta^2 U}{\delta m^2}(t_0,x_0,x,m_0,y,y') = w(t_0,x).$$

If, in general, w is the solution to the second order linearized system (56) associated with $(v,\rho),(v',\rho')$ (having initial data (t_0,ρ_0) and (t_0,ρ_0') respectively) and with $R_i=0$, $\tilde R_i=0$, $i=1,\dots 3$, then by a linearity argument one may also conclude that

$$\int_{\mathbb{R}^d}\frac{\delta^2 U}{\delta m^2}(t_0,x_0,x,m_0,y,y')\rho_0(dy)\rho_0'(dy') = w(t_0,x). \tag{88}$$

Thus, the estimate on $\frac{\delta^2 U}{\delta m^2}$ follows from Corollary 5.8, which gives

$$\|\frac{\delta^2 U}{\delta m^2}(t_0,x_0,\cdot,m_0,\cdot,\cdot)\|_{n-2;k-1,k-1}\leqslant\|\frac{\delta^2 G}{\delta m^2}(x_0,\cdot,m(T),\cdot,\cdot)\|_{n-2;k-1,k-1}+C_MT\,, \tag{89}$$

using the fact that $\sup_t\|\rho(t)\|_{-(k-1)}\leqslant(1+C_MT)\,\|\rho_0\|_{-(k-1)}$ and that the same holds for ρ'. □

Next we discuss the second order regularity of U and U^0 with respect to m and x_0.

Proposition 6.7. *Let U^0 and U be the solutions of* (67) *and* (65) *respectively. Suppose, in addition to the assumptions of Propositions 6.5 and 6.6, that we have*

$$\Big\|D^2_{x_0}(G^0,G)\Big\|_{n-2}+\Big\|D_{x_0}\frac{\delta(G^0,G)}{\delta m}\Big\|_{n-2;k-1}+\Big\|\frac{\delta^2(G^0,G)}{\delta m^2}\Big\|_{n-2;k-1,k-1}\leqslant M.$$

Then there exists $T_M > 0$ (depending on M and on the data but not on G) such that, if $T \in (0, T_M]$, the maps U^0 and U are C^2 with respect to the measure variable and x_0, and

$$\sup_t \left\| D^2_{x_0}(U^0, U)(t) \right\|_{n-2} \leqslant \left\| (D^2_{x_0} G^0, D^2_{x_0} G) \right\|_{n-2} + C_M T,$$
$$\sup_t \left\| D_{x_0} \frac{\delta(U^0, U)}{\delta m}(t) \right\|_{n-2;k-1} \leqslant \left\| D_{x_0} \frac{\delta(G^0, G)}{\delta m} \right\|_{n-2;k-1} + C_M T.$$

Moreover,

$$\sup_t \left\| \frac{\delta^2(U^0, U)}{\delta m^2}(t) \right\|_{n-2;k-1,k-1} \leqslant \left\| \frac{\delta^2(G^0, G)}{\delta m^2} \right\|_{n-2;k-1,k-1} + C_M T.$$

Proof. Step 1. The differentiability of $\delta U/\delta m$ with respect to x_0 can be achieved exactly as for its differentiability with respect to m in Proposition 6.6. For any direction $\xi \in \mathbb{R}^{d_0}$, let

- (u, m) (respectively (u_h, m_h)) be the solution of the MFG system (40) with initial condition (t_0, m_0) and parameters x_0 and $x_0 + h\xi$ respectively,
- (v, ρ) (respectively (v', ρ')) be the solution of the first order linearized system (45) with zero right-hand side (respectively right-hand side as in (71)), initial condition (t_0, δ_y) (respectively $(t_0, 0)$) and where the Hamiltonian and its derivatives are evaluated at $(x_0, x, Du(t, x), m(t))$,
- $(\tilde{v}_h, \tilde{\rho}_h)$ be the solution to the first order linearized system (45) with zero right-hand side, with initial condition (t_0, δ_y) and where the Hamiltonian and its derivatives are evaluated at $(x_0 + h\xi, x, Du_h(t, x), m_h(t))$,
- (w, μ) be the solution to the second order linearized system (56) associated with $(v, \rho), (v', \rho')$ (and (u, m)), and with right-hand side

$$\tilde{R}_1(t, x) = -H_{x_0 p}(x_0, x, Du, m(t))\xi \cdot Dv - \frac{\delta H_{x_0}}{\delta m}(x_0, x, Du, m(t))(\rho(t)) \cdot \xi,$$
$$\tilde{R}_2(t, x) = \rho H_{x_0 p}(x_0, x, Du, m(t))\xi + m H_{x_0 pp}(x_0, x, Du, m(t))\xi \; Dv + m \frac{\delta H_{x_0 p}}{\delta m}(\rho)\xi,$$
$$\tilde{R}_3(x) = \frac{\delta G_{x_0}}{\delta m}(x_0, x, m(T))(\rho(T)) \cdot \xi,$$

so that

$$\tilde{v}_h(t_0, x) = \frac{\delta U}{\delta m}(t_0, x_0 + h\xi, x, m_0, y),$$
$$v(t_0, x) = \frac{\delta U}{\delta m}(t_0, x_0, x, m_0, y), \qquad \text{and} \quad v'(t_0, x) = U_{x_0}(t_0, x_0, x, m_0) \cdot \xi.$$

Then we find $\frac{\delta U_{x_0}}{\delta m}(t_0, x_0, x, m_0, y) \cdot \xi = w(t_0, x)$, and if one replaces δ_y by an arbitrary $\rho_0 \in C^{-(k-1)}$ as the initial datum for ρ, the following representation holds:

$$\frac{\delta U_{x_0}}{\delta m}(t_0, x_0, x, m_0)(\rho_0) \cdot \xi = w(t_0, x). \tag{90}$$

Step 2. The second order differentiability of U with respect to x_0 can be checked in a similar way: let (u, m) and (u_h, m_h) be as before,

- (v, ρ), $(\tilde{v}_h, \tilde{\rho}_h)$ be the solutions of the first order linearized system (45) with right-hand side as in (71), initial condition $(t_0, 0)$, and Hamiltonian and its derivatives evaluated at $(x_0, x, Du(t, x), m(t))$ and $(x_0 + h\xi, x, Du_h(t, x), m_h(t))$ respectively,
- (w, μ) be the solution to the second order linearized system (56) associated with $(v, \rho), (v', \rho') = (v, \rho)$ (and (u, m)), and with right-hand side $\tilde{R}_1, \tilde{R}_2, \tilde{R}_3$ given by (64).

Then we find

$$D^2_{x_0}U(t_0,x_0,x,m_0)\xi\cdot\xi = w(t_0,x). \tag{91}$$

Step 3. We now prove the regularity of U^0. To show that $\delta U^0/\delta m$ is differentiable with respect to m, let $(t_0,m_0)\in[0,T)\times\mathcal{P}_2$, $y,y'\in\mathbb{R}^d$ and

- (u,m) (respectively (u_h,m_h)) be the solution of the MFG system (40) with initial condition (t_0,m_0) (respectively $(t_0,(1-h)m_0+h\delta_{y'})$),
- (v,ρ) (respectively (v',ρ')) be the solution of the first order linearized system (45) with zero right-hand side, initial condition (t_0,δ_y) (respectively $(t_0,\delta_{y'})$) and where the Hamiltonian and its derivatives are evaluated at $(x_0,x,Du(t,x),m(t))$,
- $(\tilde v_h,\tilde\rho_h)$ be the solution to the first order linearized system (45) with zero right-hand side, with initial condition (t_0,δ_y) and where the Hamiltonian and its derivatives are evaluated at $(x_0,x,Du_h(t,x),m_h(t))$,
- (w,μ) be the solution to the second order linearized system (56) associated with $(v,\rho),(v',\rho')$ (and (u,m)), and with right-hand side 0,

as in the proof of differentiability of $\delta U/\delta m$ with respect to m in Proposition 6.6. Note that

$$\begin{aligned}
&\frac{\delta U^0}{\delta m}(t_0,x_0,(1-h)m_0+h\delta_{y'},y) = \frac{\delta G^0}{\delta m}(x_0,m_h(T))(\tilde\rho_h(T)),\\
&\frac{\delta U^0}{\delta m}(t_0,x_0,m_0,y) = \frac{\delta G^0}{\delta m}(x_0,m(T))(\rho(T)).
\end{aligned}$$

Therefore, using (86) and (87)

$$\begin{aligned}
\Big|&\frac{\delta G^0}{\delta m}(x_0,m_h(T))(\tilde\rho_h(T)) - \frac{\delta G^0}{\delta m}(x_0,m(T))(\rho(T))\\
&\quad - h\Big(\frac{\delta^2 G^0}{\delta m^2}(x_0,m(T))(\rho(T),\rho'(T)) + \frac{\delta G^0}{\delta m}(x_0,m(T))(\mu(T))\Big)\Big| \leqslant Ch^2\,.
\end{aligned}$$

Lemma B.1 then implies that $\frac{\delta U^0}{\delta m}(t_0,x_0,\cdot,y)$ has a derivative, and by linearity, if μ is the solution to the second order linearized system (56) associated with $(v,\rho),(v',\rho')$ (that in turn have initial data (t_0,ρ_0) and (t_0,ρ_0') respectively and with zero right-hand side), then

$$\int_{\mathbb{R}^d}\frac{\delta^2U^0}{\delta m^2}(t_0,x_0,m_0,y,y')\rho_0(dy)\rho_0'(dy') = \frac{\delta^2G^0}{\delta m^2}(x_0,m(T))(\rho(T),\rho'(T)) + \frac{\delta G^0}{\delta m}(x_0,m(T))(\mu(T)). \tag{92}$$

Hence, by the representation formula (88) for δ^2U/δ^2m, Proposition 5.5, 5.6 and Corollary 5.8, we have, for $r\leqslant n-2$,

$$\begin{aligned}
&\Big|\frac{\delta^2U^0}{\delta m^2}(t,x_0,m_0)(\rho_0,\rho_0')\Big|^2 + \Big|D_x^r\frac{\delta^2U}{\delta m^2}(t,x_0,x,m_0)(\rho_0,\rho_0')\Big|^2\\
&= \Big(\Big|\frac{\delta^2G^0}{\delta m^2}(x_0,m(T))(\rho(T),\rho'(T))\Big| + \Big|\frac{\delta G^0}{\delta m}(x_0,m(T))(\mu(T))\Big|\Big)^2 + |D_x^rw(t,x)|^2\\
&\leqslant \Big(\Big|\frac{\delta^2G^0}{\delta m^2}(x_0,m(T))(\rho(T),\rho'(T))\Big| + C_MT\|\rho_0\|_{-(k-1)}\|\rho_0'\|_{-(k-1)}\Big)^2\\
&\qquad\qquad + \Big(\sup_x\Big|D_x\frac{\delta^2G}{\delta m^2}(x_0,x,m(T))(\rho(T),\rho'(T))\Big| + C_MT\|\rho_0\|_{-(k-1)}\|\rho_0'\|_{-(k-1)}\Big)^2\\
&\leqslant \Big\{\sup_x\frac{1}{\|\rho(T)\|_{-(k-1)}\|\rho'(T)\|_{-(k-1)}}\Big(\Big|\frac{\delta^2G^0}{\delta m^2}(x_0,m(T))(\rho(T),\rho'(T))\Big|^2 + \Big|D_x^r\frac{\delta^2G}{\delta m^2}(x_0,x,m(T))(\rho(T),\rho'(T))\Big|^2\Big)^{1/2}\\
&\qquad\qquad \times(1+C_MT) + C_MT\Big\}^2\|\rho_0\|^2_{-(k-1)}\|\rho_0'\|^2_{-(k-1)},
\end{aligned}$$

(where we use that $(x+z)^2+(y+z)^2\leqslant((x^2+y^2)^{1/2}+2z)^2$, for $x,y,z\geqslant0$). Taking the square root, then sup over x_0, ρ_0 and ρ_0' and summing over $r\leqslant n-2$ gives the estimate on $\Big\|\frac{\delta^2(U^0,U)}{\delta m^2}\Big\|_{n-2;k-1,k-1}$.

Differentiability of $\delta U^0/\delta m$ with respect to x_0 follows analogous lines: (v,ρ), (v',ρ'), $(\tilde v_h,\tilde\rho_h)$ and (w,μ) have to be changed according to Step 1. By (90), we have, using the notations of Corollary 5.9 and for any $r \leqslant n-2$:

$$\sum_{|l|=1} \left|\partial^l_{x_0}\frac{\delta U^0}{\delta m}(t_0,x_0,m_0,y)(\rho_0)\right|^2 + \left|D^r_x\partial^l_{x_0}\frac{\delta U}{\delta m}(t,x_0,x,m_0)(\rho_0)\right|^2 = \sum_{|l|=1}\left|\partial^l_{x_0}\frac{\delta G^0}{\delta m}(\rho(T)) + \frac{\delta^2 G^0}{\delta m^2}(\rho(T),\rho^l(T)) + \frac{\delta G^0}{\delta m}(\mu^l(T))\right|^2 + |D^r_x w^l(t,x)|^2,$$

where G^0 and its derivatives are all evaluated at $(x_0,m(T))$. We obtain the bounds on $(\frac{\delta U^0_{x_0}}{\delta m},\frac{\delta U_{x_0}}{\delta m})$ by using Propositions 5.5, 5.6 and Corollary 5.9.

Finally, second order differentiability of U^0 with respect to x_0, and the corresponding bound, can be obtained similarly: let $l,l'\in\mathbb{R}^{d_0}$ with $|l|=|l'|=1$, (v^l,ρ^l), $(v^{l'},\rho^{l'})$ and $(w^{l,l'},\mu^{l,l'})$ be as in Corollary 5.10. Note that

$$\partial^{l+l'}_{x_0}U^0(t_0,x_0,m_0) = \partial^{l+l'}_{x_0}G^0 + \partial^l_{x_0}\frac{\delta G^0_{x_0}}{\delta m}(\rho^{l'}(T)) + \partial^{l'}_{x_0}\frac{\delta G^0_{x_0}}{\delta m}(\rho^l(T)) + \frac{\delta^2 G^0}{\delta m^2}(\rho^l(T),\rho^{l'}(T)) + \frac{\delta G^0}{\delta m}(\mu^{l,l'}(T)),$$

while $\partial^{l+l'}_{x_0}U^0(t_0,x_0,m_0)$ is given by polarizing the representation formula (91). We can then conclude by Propositions 5.5, 5.6 and Corollary 5.10. □

6.3 Uniform continuity estimates on second order derivatives

Proposition 6.8. *Let U be the solution of (65) given by (66) and $n\geqslant 4$, $k\in\{3,\dots,n-1\}$. Suppose, in addition to the assumptions of Proposition 6.6, that*

$$\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2 G}{\delta m^2}\right)\leqslant M. \tag{93}$$

Then there exists $T_M>0$ (depending on M and on the data but not on G), such that

$$\sup_t \mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2 U}{\delta m^2}(t)\right)\leqslant \sup_{x_0}\mathrm{Lip}_{n-3;k-2,k-2}\left(\frac{\delta^2 G}{\delta m^2}\right) + C_M T.$$

Proof. We establish for later use a slightly stronger estimate involving the dependence with respect to x_0. This is used in Proposition 6.9 below. Let $(t_0,m_1,m_2)\in[0,T]\times\mathcal{P}_2^2$ and $x_0^1,x_0^2\in\mathbb{R}^{d_0}$ be fixed. We use the representation formula (88) for $\delta^2U/\delta m^2(t_0,x_0^1,m_1)$ and $\delta^2U/\delta m^2(t_0,x_0^2,m_2)$. In particular we let, for $i=1,2$,

- (u^i,m^i) be the solution to the MFG system (40) starting from m_i at time t_0 with H (and G) evaluated at $(x_0^i,x,Du^i(t,x),m^i(t))$ (and $(x_0^i,x,m^i(T))$),
- (v_i,ρ_i) (respectively (v_i',ρ_i')) be the solution of the first order linearized system (45) with zero right-hand side, initial condition (t_0,ρ_0) (respectively (t_0,ρ_0')) and where the Hamiltonian and its derivatives are evaluated at $(x_0^i,x,Du^i(t,x),m^i(t))$,
- (w^i,μ^i) be the solution to the second order linearized system (56) associated with $(v_i,\rho_i),(v_i',\rho_i')$ (and x_0^i,u^i,m^i), and with zero right-hand side.

We aim at estimating $(\bar w,\bar\mu):=(w^1-w^2,\mu^1-\mu^2)$, since

$$\bar w(t_0,x) = \frac{\delta^2U}{\delta m^2}(t_0,x_0^1,x,m_1)(\rho_0,\rho_0') - \frac{\delta^2U}{\delta m^2}(t_0,x_0^2,x,m_2)(\rho_0,\rho_0'). \tag{94}$$

We first set $(\bar v,\bar\rho):=(v_1-v_2,\rho_1-\rho_2)$ and $(\bar v',\bar\rho'):=(v_1'-v_2',\rho_1'-\rho_2')$. The pair $(\bar v,\bar\rho)$ solves the first order linearized system (45) with zero initial datum, H and its derivatives evaluated at $(x_0^1,x,Du^1(t,x),m^1(t))$,

and right-hand side

$$
\begin{aligned}
R_1(t,x) &= -(H_p^1 - H_p^2)\cdot Dv_2 - (\frac{\delta H^1}{\delta m} - \frac{\delta H^2}{\delta m})(\rho_2(t)),\\
R_2(t,x) &= \rho_2(H_p^1 - H_p^2) + (m^1 H_{pp}^1 - m^2 H_{pp}^2)Dv_2 + (m^1\frac{\delta H_p^1}{\delta m} - m^2\frac{\delta H_p^2}{\delta m})(\rho_2),\\
R_3(x) &= (\frac{\delta G^1}{\delta m} - \frac{\delta G^2}{\delta m})(\rho_2(T)),
\end{aligned}
$$

where H^i and its derivatives correspond to H and its derivatives evaluated at $(x_0^i, x, Du^i(t,x), m^i(t))$.

By Proposition 5.5 we have

$$
\sup_{t\in[t_0,T]} \|v_i(t)\|_{n-1} \leqslant C\|\rho_0\|_{-(k-2)}, \qquad \sup_{t\in[t_0,T]} \|\rho_i(t)\|_{-(k-2)} \leqslant (1+CT)\,\|\rho_0\|_{-(k-2)}, \tag{95}
$$

where C depends on the regularity of $\delta G/\delta m$, H_{x_0}, H_{x_0p}, m^i and $\sup_t \|u^i\|_n$. Note that, by the above estimates and Proposition 5.4,

$$
\sup_t \|R_1(t)\|_{n-2} + \sup_t \|R_2(t)\|_{-(k-2)} + \|R_3\|_{n-2} \leqslant C\big(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|\big)\|\rho_0\|_{-(k-2)},
$$

and therefore by Proposition 5.5 (applied to $n-1 \geqslant 2$ and $k-2 \geqslant 1$) we obtain

$$
\sup_t \|\bar v(t)\|_{n-2} \leqslant C\,T\big(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|\big)\|\rho_0\|_{-(k-2)}, \tag{96}
$$

$$
\sup_t \|\bar\rho(t)\|_{-(k-1)} \leqslant CT\big(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|\big)\|\rho_0\|_{-(k-2)}. \tag{97}
$$

Completely analogous estimates hold for v_i', ρ_i' and their differences $\bar v', \bar\rho'$.

We now proceed by estimating $(\bar w, \bar\mu)$, which solves the first order linearized system with zero initial datum, H and its derivatives evaluated at $(x_0^1, x, Du^1(t,x), m^1(t))$, and right-hand side

$$
\begin{aligned}
\overline{R}_1(t,x) := -\Big((H_p^1 - H_p^2)\cdot Dw^2 &+ (\frac{\delta H^1}{\delta m} - \frac{\delta H^2}{\delta m})(\mu^2(t))\\
&+ \frac{\delta^2 H^1}{\delta m^2}(\rho_1(t),\rho_1'(t)) - \frac{\delta^2 H^2}{\delta m^2}(\rho_2(t),\rho_2'(t)) + H_{pp}^1 Dv_1\cdot Dv_1' - H_{pp}^2 Dv_2\cdot Dv_2'\\
&+ \frac{\delta H_p^1}{\delta m}(\rho_1)\cdot Dv_1' - \frac{\delta H_p^2}{\delta m}(\rho_2)\cdot Dv_2' + \frac{\delta H_p^1}{\delta m}(\rho_1')\cdot Dv_1 - \frac{\delta H_p^2}{\delta m}(\rho_2')\cdot Dv_2\Big),
\end{aligned}
$$

$$
\begin{aligned}
\overline{R}_2(t,x) := \mu^2(H_p^1 - H_p^2) &+ (m^1 H_{pp}^1 - m^2 H_{pp}^2)Dw^2 + (m^1\frac{\delta H_p^1}{\delta m} - m^2\frac{\delta H_p^2}{\delta m})(\mu^2) + \rho_1 H_{pp}^1 Dv_1'\\
&- \rho_2 H_{pp}^2 Dv_2' + \rho_1' H_{pp}^1 Dv_1 - \rho_2' H_{pp}^2 Dv_2 + m^1 H_{ppp}^1 Dv_1 Dv_1' - m^2 H_{ppp}^2 Dv_2 Dv_2'\\
&+ m^1\frac{\delta^2 H_p^1}{\delta m^2}(\rho_1,\rho_1') - m^2\frac{\delta^2 H_p^2}{\delta m^2}(\rho_2,\rho_2') + \rho_1\frac{\delta H_p^1}{\delta m}(\rho_1') - \rho_2\frac{\delta H_p^2}{\delta m}(\rho_2')\\
&+ \rho_1'\frac{\delta H_p^1}{\delta m}(\rho_1) - \rho_2'\frac{\delta H_p^2}{\delta m}(\rho_2) + m^1\frac{\delta H_{pp}^1}{\delta m}(\rho_1')Dv_1\\
&- m^2\frac{\delta H_{pp}^2}{\delta m}(\rho_2')Dv_2 + m^1\frac{\delta H_{pp}^1}{\delta m}(\rho_1)Dv_1' - m^2\frac{\delta H_{pp}^2}{\delta m}(\rho_2)Dv_2'
\end{aligned}
$$

and

$$
\overline{R}_3(x) := \frac{\delta^2 G^1}{\delta m^2}(\rho_1(T),\rho_1'(T)) - \frac{\delta^2 G^2}{\delta m^2}(\rho_2(T),\rho_2'(T)) + (\frac{\delta G^1}{\delta m} - \frac{\delta G^2}{\delta m})(\mu^2(T)).
$$

Recall also that Proposition 5.6 and Remark 5.7 (applied to $n-1$ and $k-1$) yield

$$
\sup_t \|w^i(t)\|_{n-3} \leqslant (1+CT)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)}, \quad \sup_t \|\mu^i(t)\|_{-(k-1)} \leqslant CT\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)}. \tag{98}
$$

By the previous inequalities, (95), (96) and (97) we get

$$\sup_t \|\overline{R}_1(t)\|_{n-3} + \sup_t \|\overline{R}_2(t)\|_{-(k-1)} \leqslant CT(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)}.$$

Similarly, using also the Lipschitz regularity of $\delta G/\delta m$,

$$\begin{aligned}\|\overline{R}_3\|_{n-3} \leqslant &(1+CT)\Big\|\frac{\delta^2 G}{\delta m^2}(x_0^2, m^2(T)) - \frac{\delta^2 G}{\delta m^2}(x_0^1, m^1(T))\Big\|_{n-3;k-2,k-2}\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)} \\ &+ CT(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)},\end{aligned}$$

Then, recalling that $\bar{w} = w^1 - w^2$ satisfies (94), we obtain in view of (49) in Proposition 5.5 and for any $r \leqslant n-3$,

$$\begin{aligned}&\Big\|D_x^r \frac{\delta^2 U}{\delta m^2}(t_0, x_0^2, m_2) - D_x^r \frac{\delta^2 U}{\delta m^2}(t_0, x_0^1, m_1)\Big\|_{0;k-2,k-2} \\ &\leqslant (1+C_M T)\Big\|D_x^r \frac{\delta^2 G}{\delta m^2}(x_0^2, m_2(T)) - D_x^r \frac{\delta^2 G}{\delta m^2}(x_0^1, m_1(T))\Big)\Big\|_{0;k-2,k-2} + C_M T(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|).\end{aligned} \tag{99}$$

Choosing $x_0^1 = x_0^2$, summing over $r \leqslant n-3$ and recalling Proposition 5.4 and (93) then gives the claim.

Note that we have also the following inequality for $\bar{\mu} = \mu^1 - \mu^2$, that will be useful in the next proposition:

$$\sup_{t\in[t_0,T]} \|\mu^1(t) - \mu^2(t)\|_{-k} \leqslant CT(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)}. \tag{100}$$

□

Finally we establish the Lipschitz regularity of the second order derivatives of G^0 and G with respect to x_0 and m.

Proposition 6.9. *Let U be the solution of* (65) *given by* (66) *and U^0 be the solution to* (67) *given by* (68). *Suppose that the assumptions of Proposition 6.8 hold and that in addition:*

$$\mathrm{Lip}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) + \mathrm{Lip}_{n-3;k-2}\Big(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}\Big) + \mathrm{Lip}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) \leqslant M$$

and

$$\mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) + \mathrm{Lip}^{x_0}_{n-3;k-2}\Big(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}\Big) + \mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) \leqslant M\,,$$

for some $n \geqslant 4$ and $k \in \{3,\dots,n-1\}$. Then

$$\sup_t \mathrm{Lip}_{n-3;k-2,k-2}\Big(\frac{\delta^2 U^0(t)}{\delta m^2}, \frac{\delta^2 U(t)}{\delta m^2}\Big) \leqslant \mathrm{Lip}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) + C_M T,$$

$$\sup_t \mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\Big(\frac{\delta^2 U^0(t)}{\delta m^2}, \frac{\delta^2 U(t)}{\delta m^2}\Big) \leqslant \mathrm{Lip}^{x_0}_{n-3;k-2,k-2}\Big(\frac{\delta^2 G^0}{\delta m^2}, \frac{\delta^2 G}{\delta m^2}\Big) + C_M T,$$

$$\sup_t \mathrm{Lip}_{n-3;k-2}\Big(\frac{\delta U^0_{x_0}(t)}{\delta m}, \frac{\delta U_{x_0}(t)}{\delta m}\Big) \leqslant \mathrm{Lip}_{n-3;k-2}\Big(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}\Big) + C_M T,$$

$$\sup_t \mathrm{Lip}^{x_0}_{n-3;k-2}\Big(\frac{\delta U^0_{x_0}(t)}{\delta m}, \frac{\delta U_{x_0}(t)}{\delta m}\Big) \leqslant \mathrm{Lip}^{x_0}_{n-3;k-2}\Big(\frac{\delta G^0_{x_0}}{\delta m}, \frac{\delta G_{x_0}}{\delta m}\Big) + C_M T,$$

and

$$\mathrm{Lip}_{n-3}(D^2_{x_0}U^0(t), D^2_{x_0}U(t)) \leqslant \mathrm{Lip}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) + C_M T,$$

$$\mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}U^0(t), D^2_{x_0}U(t)) \leqslant \mathrm{Lip}^{x_0}_{n-3}(D^2_{x_0}G^0, D^2_{x_0}G) + C_M T.$$

Proof. We will detail only the proof of Lipschitz estimates of $(\delta^2U^0/\delta m^2, \delta^2U/\delta m^2)$. Lipschitz regularity of $\delta U^0_{x_0}/\delta m$ and $D^2_{x_0}U^0$, $\delta U_{x_0}/\delta m$ and $D^2_{x_0}U$ can be proven by following identical lines using the representation formulas that appear in the proof of Proposition 6.7.

Let us start with $\delta^2U^0/\delta m^2$. Let $(t_0, m_1, m_2) \in [0,T] \times \mathcal{P}_2^2$ and $x_0^1, x_0^2 \in \mathbb{R}^d$ be fixed. Let also, as in the proof of the previous Proposition 6.7, for $i = 1, 2$

- (u^i, m^i) be the solution to the MFG system (40) starting from m_i at time t_0 with H (and G) evaluated at $(x_0^i, x, Du^i(t,x), m^i(t))$ (and $(x_0^i, x, m^i(T))$),
- (v_i, ρ_i) (respectively (v_i', ρ_i')) be the solution of the first order linearized system (45) with zero right-hand side, initial condition (t_0, ρ_0) (respectively (t_0, ρ_0')) and where the Hamiltonian and its derivatives are evaluated at $(x_0^i, x, Du^i(t,x), m^i(t))$,
- (w^i, μ^i) be the solution to the second order linearized system (56) associated with $(v_i, \rho_i), (v_i', \rho_i')$ (and (u^i, m^i)), and with zero right-hand side.

Recall that (92) provides a representation formula for $\delta^2U^0/\delta m^2$, that is

$$\frac{\delta^2U^0}{\delta m^2}(t_0, x_0^i, m_i)(\rho_0, \rho_0') = \frac{\delta^2G^0}{\delta m^2}(x_0^i, m^i(T))(\rho_i(T), \rho_i'(T)) + \frac{\delta G^0}{\delta m}(x_0^i, m^i(T))(\mu^i(T)),$$

and $\frac{\delta^2U}{\delta m^2}(t_0, x_0^i, x, m_i)(\rho_0, \rho_0') = w^i(t_0, x)$. Let us recall the following inequalities

$$\begin{aligned}
&\sup_{t\in[t_0,T]} \mathbf{d}_2(m^1(t), m^2(t)) \leqslant (1+CT)\mathbf{d}_2(m_0^1, m_0^2) + CT|x_0^1 - x_0^2|,\\
&\sup_{t\in[t_0,T]} \|\rho_i(t)\|_{-(k-2)} \leqslant (1+CT)\,\|\rho_0\|_{-(k-2)},\\
&\sup_{t\in[t_0,T]} \|\rho_i'(t)\|_{-(k-2)} \leqslant (1+CT)\,\|\rho_0'\|_{-(k-2)},\\
&\sup_{t\in[t_0,T]} \|\rho_1(t) - \rho_2(t)\|_{-(k-1)} \leqslant CT\big(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|\big)\|\rho_0\|_{-(k-2)},\\
&\sup_{t\in[t_0,T]} \|\rho_1'(t) - \rho_2'(t)\|_{-(k-1)} \leqslant CT\big(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|\big)\|\rho_0'\|_{-(k-2)},\\
&\sup_{t\in[t_0,T]} \|\mu^i(t)\|_{-(k-1)} \leqslant CT\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)},\\
&\sup_{t\in[t_0,T]} \|\mu^1(t) - \mu^2(t)\|_{-k} \leqslant CT(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)},
\end{aligned}$$

that are consequences of Proposition 5.4, (95), (97), (98) and (100) respectively. Setting

$$\theta_T := CT(\mathbf{d}_2(m_1,m_2) + |x_0^1 - x_0^2|)\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)},$$

we obtain, using (99) also, for any $r \leqslant n-3$,

$$\begin{aligned}
&\Big|\Big(\frac{\delta^2U^0}{\delta m^2}(t, x_0^1, m_1) - \frac{\delta^2U^0}{\delta m^2}(t, x_0^2, m_2)\Big)(\rho_0, \rho_0')\Big|^2 + \sup_x \Big|D_x^r\Big(\frac{\delta^2U}{\delta m^2}(t, x_0^1, x, m_1) - \frac{\delta^2U}{\delta m^2}(t, x_0^2, x, m_2)\Big)(\rho_0, \rho_0')\Big|^2\\
&\leqslant (1+CT)\Big\{\Big|\frac{\delta^2G^0}{\delta m^2}(x_0^1, m^1(T))(\rho_1(T), \rho_1'(T)) - \frac{\delta^2G^0}{\delta m^2}(x_0^2, m^2(T))(\rho_1(T), \rho_1'(T))\Big| + \theta_T\Big\}^2\\
&+ (1+CT)\Big\{\sup_x \Big|D_x^r\frac{\delta^2G}{\delta m^2}(x_0^1, x, m^1(T))(\rho_1(T), \rho_1'(T)) - D_x^r\frac{\delta^2G}{\delta m^2}(x_0^2, x, m^2(T))(\rho_1(T), \rho_1'(T))\Big| + \theta_T\Big\}^2.
\end{aligned}$$

Choosing $m_1 = m_2 = m$ and rearranging gives the Lipschitz estimates in x_0:

$$\begin{aligned}
&\Big|\Big(\frac{\delta^2U^0}{\delta m^2}(t, x_0^1, m) - \frac{\delta^2U^0}{\delta m^2}(t, x_0^2, m)\Big)(\rho_0, \rho_0')\Big|^2 + \sup_x \Big|D_x^r\Big(\frac{\delta^2U}{\delta m^2}(t, x_0^1, x, m) - \frac{\delta^2U}{\delta m^2}(t, x_0^2, x, m)\Big)(\rho_0, \rho_0')\Big|^2\\
&\leqslant (1+CT)\Big\{\Big(\Big|\frac{\delta^2G^0}{\delta m^2}(x_0^1, m^1(T))(\rho_1(T), \rho_1'(T)) - \frac{\delta^2G^0}{\delta m^2}(x_0^2, m^1(T))(\rho_1(T), \rho_1'(T))\Big|^2 +\\
&\qquad \sup_x \Big|D_x^r\frac{\delta^2G}{\delta m^2}(x_0^1, x, m^1(T))(\rho_1(T), \rho_1'(T)) - D_x^r\frac{\delta^2G}{\delta m^2}(x_0^2, x, m^1(T))(\rho_1(T), \rho_1'(T))\Big|^2\Big)^{1/2} +\\
&\quad + CT|x_0^1 - x_0|\|\rho_0\|_{-(k-2)}\|\rho_0'\|_{-(k-2)}\Big\}^2,
\end{aligned}$$

while the choice $x_0^1 = x_0^2$ gives similarly the Lipschitz estimates in m. □

A Estimates for solutions to HJ equations

A.1 Main estimates

In this section, we assume that the data a, h and g are smooth and we are looking for a priori estimates on the smooth and globally bounded solution u to the HJ equation

$$\begin{cases} -\partial_t u(t,x) - \mathrm{Tr}(a(t,x)D^2u(t,x)) + h(t,x,Du(t,x)) = 0 & \text{in } (0,T)\times\mathbb{R}^d \\ u(T,x) = g(x) & \text{in } \mathbb{R}^d \end{cases} \tag{101}$$

We always assume below that there exists $C_0 > 0$ and $\gamma \geqslant 1$ such that

$$a(t,x) \geqslant C_0^{-1} I_d, \qquad \|Da\|_\infty \leqslant C_0$$

and

$$|D_x h(t,x,p)| \leqslant C_0(1+|p|^\gamma)$$

for every $(t,x,p) \in (0,T)\times\mathbb{R}^d\times\mathbb{R}^d$.

Proposition A.1. *(Lipschitz estimates.) For any $M > 0$ there exists $T_M, C_M > 0$, depending on M, C_0 and γ, such that, if $T \in (0,T_M)$ and $\|Dg\|_\infty \leqslant M$, then*

$$\sup_{t\in[0,T]} \|Du(t)\|_\infty \leqslant \|Dg\|_\infty + C_M T$$

Proof. We use a standard Bernstein method. Let $v(t,x) = \sum_i u_i^2(t,x)$. Then

$$\partial_t v(t,x) = 2\sum_i u_i(t,x)u_{i,t}(t,x), \qquad v_j(t,x) = 2\sum_i u_i(t,x)u_{ij}(t,x),$$

$$v_{jk}(t,x) = 2\sum_i (u_{ik}(t,x)u_{ij}(t,x) + u_i(t,x)u_{ijk}(t,x)).$$

Thus

$$\begin{aligned} &-\partial_t v - \mathrm{Tr}(a(t,x)D^2v(t,x)) \\ &= -2\sum_i u_i(t,x)u_{i,t}(t,x) - 2\sum_{i,j,k} a_{jk}(t,x)(u_{ik}(t,x)u_{ij}(t,x) + u_i(t,x)u_{ijk}(t,x) \\ &= -2\sum_{i,j,k} a_{jk}(t,x)u_{ik}(t,x)u_{ij}(t,x) - 2\sum_i u_i(t,x)D_i\left(\partial_t u + \mathrm{Tr}(a(t,x)D^2u(t,x))\right) \\ &\qquad + \sum_{i,j,k} u_i(t,x)(a_{jk})_i(t,x)u_{jk}(t,x) \end{aligned}$$

where $(a_{jk})_i$ denotes the x_i-derivative of the element a_{jk} of the matrix $a(t,x)$.

Using the equation for u we find

$$\begin{aligned} &-\partial_t v - \mathrm{Tr}(a(t,x)D^2v(t,x)) \\ &= -2\sum_{i,j,k} a_{jk}(t,x)u_{ik}(t,x)u_{ij}(t,x) - 2\sum_i u_i(t,x)\left(h_i(t,x,Du(t,x)) + h_p(t,x,Du(t,x))\cdot Du_i(t,x)\right) \\ &\qquad + \sum_{i,j,k} u_i(t,x)(a_{jk})_i(t,x)u_{jk}(t,x). \end{aligned} \tag{102}$$

Using our assumptions on a and h, we infer that

$$\begin{aligned} &-\partial_t v - \mathrm{Tr}(a(t,x)D^2v(t,x)) + h_p(t,x,Du(t,x))\cdot Dv(t,x) \\ &\quad \leqslant -2C_0^{-1}|D^2u|^2 + 2C_0|Du|(1+|Du|^\gamma) + \|Da\|_\infty|Du|\,|D^2u| \\ &\quad \leqslant 2C_0|Du|(1+|Du|^\gamma) + c_d\|Da\|_\infty^2 C_0|Du|^2 \end{aligned}$$

for some constant c_d only depending on the dimension d. In particular, by maximum principle we estimate

$$\|v\|_{L^\infty(Q_T)} \leqslant \|Dg\|^2_{L^\infty(Q_T)} + T\left[2C_0\|Du\|_{L^\infty(Q_T)}(1+\|Du\|^\gamma_{L^\infty(Q_T)}) + c_d\|Da\|^2_\infty C_0\|Du\|^2_{L^\infty(Q_T)}\right] \tag{103}$$

which implies

$$\begin{aligned}\|v\|_{L^\infty(Q_T)} \leqslant \|Dg\|^2_{L^\infty(Q_T)} + 4T^2C_0^2 + \frac{1}{4}\|Du\|^2_{L^\infty(Q_T)} \\ +\hat{C}\,T\,\|Du\|^2_{L^\infty(Q_T)}\left[\|Du\|^{\gamma-1}_{L^\infty(Q_T)}+1\right]\end{aligned} \tag{104}$$

for some $\hat{C}$ only depending on d and C_0. Recall that $\|v\|_{L^\infty(Q_T)} = \|Du\|^2_{L^\infty(Q_T)}$, and define T_M as

$$T_M = \min\left\{\frac{1}{2C_0}M, \frac{1}{4\hat{C}(1+(2M)^{\gamma-1})}\right\}.$$

Then it is easy to see that

$$\|Du\|_{L^\infty(Q_T)} \leqslant 2M \qquad \forall T \leqslant T_M\,. \tag{105}$$

Indeed, for $T < T_M$ and $\|Du\|_{L^\infty(Q_T)} \leqslant 2M$, (104) implies

$$\begin{aligned}\|Du\|^2_{L^\infty(Q_T)} &\leqslant \|Dg\|^2_{L^\infty(Q_T)} + 4T_M^2C_0^2 + \frac{1}{4}\|Du\|^2_{L^\infty(Q_T)} \\ &\quad +\hat{C}\,T_M\,\|Du\|^2_{L^\infty(Q_T)}\left[(2M)^{\gamma-1}+1\right] \\ &< \|Dg\|^2_{L^\infty(Q_T)} + M^2 + \frac{1}{2}\|Du\|^2_{L^\infty(Q_T)}\end{aligned}$$

hence

$$\|Du\|_{L^\infty(Q_T)} < 2M$$

whenever $T < T_M$ and $\|Du\|_{L^\infty(Q_T)} \leqslant 2M$. A continuity argument implies that

$$\sup\{T\,:\,\|Du\|_{L^\infty(Q_T)} \leqslant 2M\} = T_M$$

so (105) holds true. Using this information, we deduce from (103) that

$$\|Du\|^2_{L^\infty(Q_T)} \leqslant \|Dg\|^2_{L^\infty(Q_T)} + C_M\,T\,\|Du\|_{L^\infty(Q_T)}$$

where $C_M = 2C_0(1+(2M)^\gamma) + c_d\|Da\|^2_\infty C_0\,2M$. Hence

$$\left(\|Du\|_{L^\infty(Q_T)} - \frac{1}{2}C_M\,T\right)^2 \leqslant \|Dg\|^2_{L^\infty(Q_T)} + \frac{1}{4}C_M^2\,T^2$$

which implies

$$\|Du\|_{L^\infty(Q_T)} \leqslant C_M\,T + \|Dg\|_{L^\infty(Q_T)}\,.$$

□

Proposition A.2. *(Lipschitz estimates, linear case.) We now assume that $T \leqslant 1$ and that*

$$|D_x h(t,x,p)| \leqslant C_1 + C_2|p| \qquad \forall (t,x,p) \in (0,T)\times\mathbb{R}^d\times\mathbb{R}^d\,,$$

for some constants $C_1, C_2 > 0$. Then there exists a constant C, depending on C_0, C_2 and $\|Da\|_\infty$ only, such that

$$\sup_{t\in[0,T]}\|Du(t)\|_\infty \;\leqslant\; \|Dg\|_\infty(1+CT) + CC_1T.$$

Proof. Our starting point is inequality (102) in the previous proof. Using our assumptions on a and h we get:

$$\begin{aligned}&-\partial_t v - \mathrm{Tr}(a(t,x)D^2v(t,x)) + h_p(t,x,Du(t,x))\cdot Dv(t,x) \\ &\quad\leqslant -2C_0^{-1}|D^2u|^2 + 2|Du|(C_1 + C_2\,|Du|) + \|Da\|_\infty|Du|\,|D^2u| \\ &\quad\leqslant 2|Du|(C_1 + C_2\,|Du|) + c_d\|Da\|^2_\infty C_0|Du|^2\end{aligned}$$

which implies

$$-\partial_t v - \mathrm{Tr}(a(t,x)D^2v(t,x)) + h_p(t,x,Du(t,x))\cdot Dv(t,x) \leqslant \lambda\, v + 2\,C_1\, v^{1/2}$$

where $\lambda = 2C_2 + c_d\|Da\|_\infty^2 C_0$. By the maximum principle we get

$$\|v\|_{L^\infty(Q_T)} \leqslant e^{\lambda T}\left(2C_1\,T\|v\|_{L^\infty(Q_T)}^{1/2} + \|Dg\|_\infty^2\right),$$

from which we derive that

$$\|v\|_{L^\infty(Q_T)}^{1/2} \leqslant 2\,C_1\,T\,e^{\lambda T} + e^{\lambda T/2}\,\|Dg\|_\infty.$$

Since $T \leqslant 1$ (and so $e^{\lambda T/2} \leqslant 1 + c_\lambda T$), the conclusion follows. □

Proposition A.3. *(Second order estimate.) Assume that h and a are of class C_b^2. Then, for any $M > 0$, there are constants $T_M, C_M > 0$, depending on M and on*

$$\sup_{t\in[0,T]}\|a(t)\|_2 + \sup_{|p|\leqslant\|Du\|_\infty}\|D_{xp}^2 h(\cdot,\cdot,p)\|_\infty \tag{106}$$

such that, if $\|D^2 g\|_\infty \leqslant M$ and $T \in (0, T_M)$, then

$$\sup_{t\in[0,T]}\|D^2u(t)\|_\infty \leqslant \|D^2 g\|_\infty + C_M T.$$

If, in addition, h is affine in p, then there is a constant C, depending only on C_0, $\sup_{t\in[0,T]}\|a(t)\|_2$ and on $\|D_{xp}^2 h\|_\infty$, such that, for any $T \in (0,1]$,

$$\sup_{t\in[0,T]}\|D^2u(t)\|_\infty \leqslant (1+CT)\|D^2 g\|_\infty + CT\sup_{|p|\leqslant\|Du\|_\infty}\|D_{xx}^2 h(\cdot,\cdot,p)\|_\infty.$$

Proof. We use the Bernstein method again. Let $w(t,x) = \sum_{i,j} u_{ij}^2$. Then

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x))\\
&= -2\sum_{i,j,k,l} a_{kl}(t,x)u_{ijk}(t,x)u_{ijl}(t,x) - 2\sum_{i,j} u_{ij}(t,x)D_{i,j}\left(\partial_t u + \sum_{k,l} a_{kl}u_{kl}\right)\\
&\qquad +2\sum_{i,j,k,l} u_{ij}(t,x)\left((a_{kl})_i(t,x)u_{jkl}(t,x) + (a_{kl})_j(t,x)u_{ikl}(t,x) + (a_{kl})_{ij}u_{kl}\right).
\end{aligned}$$

So

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x))\\
&= -2\sum_{i,j,k,l} a_{kl}u_{ijk}u_{ijl} - 2\sum_{i,j} u_{ij}\left(h_{ij} + h_{i,p}\cdot Du_j + h_{j,p}\cdot Du_i + h_{pp}Du_i\cdot Du_j + h_p Du_{ij}\right)\\
&\qquad +2\sum_{i,j,k,l} u_{ij}(t,x)\left((a_{kl})_i(t,x)u_{jkl}(t,x) + (a_{kl})_j(t,x)u_{ikl}(t,x) + (a_{kl})_{ij}u_{kl}\right)
\end{aligned} \tag{107}$$

which yields, using the ellipticity of $a(t,x)$,

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x)) + h_p(t,x,Du(t,x))\cdot Dw(t,x)\\
&\leqslant -2C_0^{-1}|D^3u|^2 + C_h|D^2u|\left(1 + |D^2u| + |D^2u|^2\right) + C|D^2u|\left(\|a\|_1\,|D^3u| + \|a\|_2\,|D^2u|\right)
\end{aligned}$$

for some constant C_h depending on $\sup_{|p|\leqslant\|Du\|_\infty}\|D_{x,p}^2 h(\cdot,\cdot,p)\|_\infty$. Young's inequality leads to

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x)) + h_p(t,x,Du(t,x))\cdot Dw(t,x)\\
&\leqslant C|D^2u|\left(1 + |D^2u| + |D^2u|^2\right)
\end{aligned}$$

where now C depends on $\|a\|_2$ as well. We conclude using maximum principle as in the proof of Proposition A.1.

If h is affine in p, then with the same estimates we deduce from (107):

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x)) + h_p(t,x,Du(t,x))\cdot Dw(t,x)\\
&\leqslant |D^2u|\left(2\|D_{xx}h\|_\infty + C|D^2u|\right)\\
&\leqslant Cw + 2\,\|D_{xx}h\|_\infty\,|D^2u|\,,
\end{aligned}$$

where C depends on $\|a\|_2$, C_0 and $\sup_{|p|\leqslant\|Du\|_\infty}\|D^2_{x,p}h(\cdot,\cdot,p)\|_\infty$. The conclusion follows as in Lemma A.2. □

Proposition A.4. *(Third order estimate) Assume that h and a (and the solution u) are of class C^3_b. Then there is a constant C, depending on $\|D^2u\|_\infty$, on $\|Da\|_\infty + \|D^2a\|_\infty + \|D^3a\|_\infty$ and on*

$$\sup_{|p|\leqslant\|Du\|_\infty}\left\{\|D^3_{(x,p)}h(\cdot,\cdot,p)\|_\infty + \|h_{pp}(\cdot,\cdot,p)\|_\infty\right\},$$

such that, for any $T\in(0,1]$,

$$\sup_{t\in[0,T]}\|D^3u(t)\|_\infty \leqslant (1+CT)\|D^3g\|_\infty + CT,$$

Proof. Let $w=\sum_{ijk}u^2_{ijk}$. Then

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x))\\
&= -2\sum_{i,j,k,l,m} a_{lm}(t,x)u_{ijkl}(t,x)u_{ijkm}(t,x) - 2\sum_{i,j}u_{ijk}(t,x)D_{i,j,k}\left(\partial_t u + \sum_{l,m}a_{lm}u_{lm}\right)\\
&\quad +2\sum_{i,j,k,l,m}u_{ijk}(t,x)\Big((a_{lm})_{ijk}u_{lm} + (a_{lm})_{ij}u_{klm} + (a_{lm})_{ik}u_{jlm} + (a_{lm})_{jk}u_{ilm}\\
&\qquad\qquad +(a_{lm})_i u_{jklm} + (a_{lm})_j u_{iklm} + (a_{lm})_k u_{ijlm}\Big)
\end{aligned}$$

So

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x))\\
&= -2\sum_{i,j,k,l,m} a_{lm}(t,x)u_{ijkl}(t,x)u_{ijkm}(t,x) - 2\sum_{i,j}u_{ijk}(t,x)D_{i,j,k}\{h\}\\
&\quad +2\sum_{i,j,k,l,m}u_{ijk}(t,x)\Big((a_{lm})_{ijk}u_{lm} + (a_{lm})_{ij}u_{klm} + (a_{lm})_{ik}u_{jlm} + (a_{lm})_{jk}u_{ilm}\\
&\qquad\qquad +(a_{lm})_i u_{jklm} + (a_{lm})_j u_{iklm} + (a_{lm})_k u_{ijlm}\Big)\,.
\end{aligned}\tag{108}$$

As before, the coercivity of a implies

$$-2\sum_{i,j,k,l,m} a_{lm}(t,x)u_{ijkl}(t,x)u_{ijkm}(t,x) \leqslant -2C_0^{-1}|D^4u|^2\,,$$

whereas last term in (108) is estimated as

$$\begin{aligned}
&2\textstyle\sum_{i,j,k,l,m}u_{ijk}(t,x)\Big((a_{lm})_{ijk}u_{lm} + (a_{lm})_{ij}u_{klm} + (a_{lm})_{ik}u_{jlm} + (a_{lm})_{jk}u_{ilm}\\
&\qquad\qquad +(a_{lm})_i u_{jklm} + (a_{lm})_j u_{iklm} + (a_{lm})_k u_{ijlm}\Big)\\
&\leqslant C_0^{-1}|D^4u|^2 + |D^3u|\left(2\|D^3a\|_\infty|D^2u| + C\,|D^3u|\right),
\end{aligned}$$

for some C depending on C_0 and $\|D^2a\|_\infty$. Finally, a direct computation of $D_{i,j,k}\{h\}$ and a straightforward estimate of all terms involved imply

$$\begin{aligned}
&-2\textstyle\sum_{i,j}u_{ijk}(t,x)D_{i,j,k}\{h\} \leqslant -h_p(t,x,Du(t,x))\cdot Dw(t,x)\\
&\quad +C\,|D^3u|\left[\|D^2h\|_\infty|D^3u|(1+|D^2u|) + \|D^3h\|_\infty(1+|D^2u|^3)\right].
\end{aligned}$$

Hence, putting all together we deduce from (108):

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x)) + h_p(t,x,Du(t,x))\cdot Dw(t,x)\\
&\leqslant C\,|D^3u|\left[\|D^2h\|_\infty\,|D^3u|(1+|D^2u|) + \|D^3h\|_\infty(1+|D^2u|^3)\right]\\
&\quad +|D^3u|\left(2\|D^3a\|_\infty\,|D^2u| + C\,|D^3u|\right)\\
&\quad \leqslant C|D^3u|^2 + C|D^3u|\,,
\end{aligned}$$

where C depends also on $\|D^k u\|_\infty$ for $k\leqslant 2$. We conclude as in Proposition A.2. □

Lemma A.5. *(Higher order estimate) Let $n\in\mathbb{N}$ with $n\geqslant 3$ and assume that h and a (and the solution u) are of class C_b^n. There is a constant C, depending on n, d, $\sup_t\|u(t)\|_{n-1}$, $\sup_t\|a(t)\|_n$ and on*

$$\sup_{|p|\leqslant\|Du\|_\infty}\sum_{k=0}^{n}\|D^k_{(x,p)}h(\cdot,\cdot,p)\|_\infty, \tag{109}$$

such that, for any $T\in(0,1]$,

$$\sup_t\|D^nu(t)\|_\infty\leqslant(1+CT)\|D^ng\|_\infty + CT.$$

Proof. Let $w:=\sum_{|k|=n}u_k^2$ where the multi-index $k=(k_1,\dots,k_d)$ belongs to $\mathbb{N}^d$ and $|k|=\sum_i k_i$. Then

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x))\\
&= -2\sum_{|k|=n}\sum_{i,j}a_{ij}(t,x)u_{k,i}(t,x)u_{k,j}(t,x) - 2\sum_{|k|=n}u_k(t,x)D_k\left\{\partial_t u + \mathrm{Tr}(aD^2u)\right\}\\
&\quad +2\sum_{|k|=n}u_k\left(D_k(\mathrm{Tr}(aD^2u)) - \mathrm{Tr}(aD^2u_k)\right).
\end{aligned}$$

As $n\geqslant 3$, a simple induction argument shows that $D_k\{h\}$ is of the form

$$D_k\{h\} = f_k + g_k\cdot D^nu + h_p\cdot Du_k$$

where the map

$$f_k = f_k(t,x,Du(t,x),\dots,D^{n-1}u(t,x))$$

is a polynomial function of the derivatives of u up to order $n-1$ with coefficients involving derivatives of h with respect to (x,p) up to order n computed at $(t,x,Du(t,x))$, while

$$g_k\cdot D^nu = \sum_{|\xi|=n-1}\sum_{z+\xi=k}D_{z,p}h(t,x,Du(t,x))Du_\xi + h_{pp}(t,x,Du(t,x))Du_zDu_\xi\,,$$

where ξ is any multi-index of length $n-1$, z is a multi-index of length 1 ($z=e_j$ for some $j\in\{1,\dots,d\}$) and $\xi+z=k$.

Therefore

$$\begin{aligned}
&-\partial_t w - \mathrm{Tr}(a(t,x)D^2w(t,x)) + h_p\cdot Dw\\
&= -2\sum_{i,j}\sum_{|k|=n}a_{ij}(t,x)u_{k,i}(t,x)u_{k,j}(t,x) - 2\sum_{|k|=n}u_k(t,x)\left(f_k + g_k\cdot D^nu\right)\\
&\quad +2\sum_{|k|=n}u_k\left(D_k(\mathrm{Tr}(aD^2u)) - \mathrm{Tr}(aD^2u_k)\right)\\
&\leqslant -2C_0^{-1}\sum_{|k|=n}|Du_k|^2 + C|u_k|(1+|u_k|)\\
&\quad +2\sum_{|k|=n}u_k\left(D_k(\mathrm{Tr}(aD^2u)) - \mathrm{Tr}(aD^2u_k)\right)
\end{aligned}$$

where C depends on $\sup_t \|u(t)\|_{n-1}$ and the quantity in (109). Last term can be estimated as before: the higher order quantity involves Du_k, so we have through Young's inequality

$$2\sum_{|k|=n} u_k \left(D_k(\mathrm{Tr}(aD^2u)) - \mathrm{Tr}(aD^2u_k)\right) \leqslant 2C_0^{-1}\sum_{|k|=n}|Du_k|^2 + C\,|u_k|(1+|u_k|),$$

for some C depending on $\sup_t \|a(t)\|_n$ and $\sup_t \|u(t)\|_{n-1}$. Finally, we conclude with maximum principle, as in Lemma A.2. □

Proposition A.6. *(Higher order estimate, further informations) Let $n \in \mathbb{N}$ with $n \geqslant 3$ and assume that h and a (and the solution u) are of class C_b^n. For any $M > 0$, there are constants $K_M, T_M > 0$, depending on M, C_0 and γ, and a constant $C_M > 0$ depending on*

$$\sup_{t\in[0,T_M]} \|a(t)\|_n + \sup_{|p|\leqslant K_M}\sum_{k=0}^{n} \|D^k_{(x,p)}h(\cdot,\cdot,p)\|_\infty,$$

such that, if $\|g\|_n \leqslant M$, then, for any $T \in (0, T_M)$ and any $r \leqslant n$, we have

$$\sup_{t\in[0,T]} \|D_x^r u(t)\|_\infty \leqslant \|D_x^r g\|_\infty + C_M T$$

and therefore

$$\sup_{t\in[0,T]} \|u(t)\|_n \leqslant \|g\|_n + C_M T. \tag{110}$$

Proof. The proof is a straightforward combination of Propositions A.1, A.3 and Lemma A.5. □

We finally address the same issue for (uncoupled) systems of linear parabolic equations: let $(u^l)_{l=1,\dots,k}$ solve the system

$$\begin{cases} -\partial_t u^l - \mathrm{Tr}(a(t,x)D^2u^l) + V(t,x)\cdot Du^l + f^l(t,x) = 0 & \text{in } (0,T)\times\mathbb{R}^d \\ u^l(T,x) = g^l(x) & \text{in } \mathbb{R}^d \end{cases}$$

where a, V and the f^l are bounded in C_b^n independently of $t \in [0,1]$, for some $n \in \mathbb{N}^*$. Note that the diffusion and the drift terms are independent of l.

Proposition A.7 (Higher order estimate, systems of affine equations)**.** *There is a constant C, depending on k, d, $\sup_t \|a(t)\|_n$ and on $\sup_t \|V(t)\|_n$, such that, for any $T \in (0,1]$ and for any $r \leqslant n$,*

$$\sup_{t,x}\left(\sum_{l=1}^k |D_x^r u^l(t,x)|^2\right)^{1/2} \leqslant (1+CT)\sup_x\left(\sum_{l=1}^k |D_x^r g^l(x)|^2\right)^{1/2} + CT\sup_l(\|g^l\|_r + \sup_t \|f^l(t)\|_r).$$

In particular, if $k = 1$, for any $r \leqslant n$

$$\sup_{t\in[0,T]} \|D_x^r u(t)\|_\infty \leqslant (1+CT)\|D_x^r g\|_\infty + CT\sup_t \|D_x^r f(t)\|_\infty.$$

The only small point here is that the supremum over x is outside the sum (and not inside as it would be given by simply applying to each u^l the previous Propositions).

Proof. The proof runs exactly along the same lines as before and so we just explain briefly the idea for $r = 0$. Let us consider $v(t,x) = \sum_{l=1}^k (u^l(t,x))^2$. Then v solves

$$-\partial_t v - \mathrm{Tr}(aD^2v) + V\cdot Dv = -2\sum_{l=1}^k u^l f^l - \sum_{i,j,l} a_{ij}u_i^l u_j^l$$

We infer the result by using the positivity of a and the maximum principle.

A.2 Systems with parameters

In this section we revisit the above estimates for specific systems of Hamilton-Jacobi equations involving a parameter y. The motivation for the specific form of the system is the analysis of the MFG problems with a major player. Note that here the variables-parameter couple $(x;y)$ plays the role of $(x_0;x)$ in the HJ system (32) analyzed throughout Section 4.1. As usual, we discuss linear and nonlinear systems separately.

A.2.1 Nonlinear systems

Here we consider the system consisting in a coupling of a non-linear HJ equation with a linear one:

$$\begin{cases} -\partial_t u^0(t,x) - \Delta u^0(t,x) + h^0(t,x,Du^0(t,x)) = 0 \qquad \text{in } (0,T)\times\mathbb{R}^d, \\ -\partial_t u(t,x;y) - \Delta u(t,x;y) + h^0_p(t,x,Du^0(t,x))\cdot Du(t,x;y) + f(t,x;y) = 0 \text{ in } (0,T)\times\mathbb{R}^d, \\ u^0(T,x) = g^0(x),\ u(T,x) = g(x;y) \qquad \text{in } \mathbb{R}^d, \end{cases} \tag{111}$$

where $h^0 : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ and $f : [0,T]\times\mathbb{R}^d\times\mathbb{R}^{d_1} \to \mathbb{R}$ (d_1 being the space parameter of the variable y) are smooth maps satisfying in addition the bounds:

$$|D_{x,p}h^0(t,x,p)| + |D^2_{x,p}h^0(t,x,p)| \leqslant C_0(|p|^\gamma + 1), \tag{112}$$

for some $\gamma > 0$ and $C_0 > 0$.

Proposition A.8. *Let $r,n \in \mathbb{N}$ and assume (in addition to (112)) that h^0, h^0_p are of class C^r_b and that f is bounded in $C^{r,n}_b$ independently of $t \in [0,1]$ for some $n \in \mathbb{N}$. For any $M > 0$, there are constants $K_M, T_M > 0$, depending on M, C_0 and γ in (112), and a constant $C_M > 0$ depending on*

$$\sup_{|p|\leqslant K_M} \sum_{k=0}^r \|D^k_{(x,p)}h^0(\cdot,\cdot,p)\|_\infty + \sum_{k=0}^r \|D^k_{(x,p)}h^0_p(\cdot,\cdot,p)\|_\infty, \quad \sup_t \|f(t)\|_{r,n}$$

such that, if $\|g^0\|_r + \|g\|_{r,n} \leqslant M$ and $T \in (0,T_M)$, and if (u^0,u) is the solution to (111), then we have, for $l \leqslant n$,

$$\sup_{t,x,y}\Big(|D^r u^0(t,x)|^2 + |D^r_x D^l_y u(t,x;y)|^2\Big)^{1/2} \leqslant \sup_{x,y}\Big(|D^r g^0(x)|^2 + |D^r_x D^l_y g(x;y)|^2\Big)^{1/2} + C_M T.$$

Let us recall that $D^r_x D^l_y u = (\partial^\beta_x \partial^\alpha_y u)_{|\beta|=r,|\alpha|=l}$, hence $|D^r_x D^l_y u|^2 = \sum_{|\beta|=r,|\alpha|=l}(\partial^\beta_x\partial^\alpha_y u)^2$. Let us also point out that the main difference compared to Proposition A.6 is that we need to estimate u^0 and u at the same time.

Proof. The proof uses the same technique as for a single Hamilton-Jacobi equation without parameter. We explain only the main changes. We first prove the result for $l = 0$.

By the maximum principle we can first bound $|u^0|^2 + |u|^2$ by $\|(g^0)^2 + g^2\|_\infty + CT$. Next we address the Lipschitz estimate. We claim that, for any $M > 0$ and any $n \in \mathbb{N}$, if $\|Dg^0\|_\infty + \|D_x g\|_\infty \leqslant M$, then there exists T_M and C_M (depending on M, C_0, n and γ in (112) only) such that

$$\sup_{t,x,y}\Big(|Du^0(t,x)|^2 + |D_x u(t,x;y)|^2\Big)^{1/2} \leqslant \sup_{x,y}\Big(|Dg^0(x)|^2 + |D_x g(x;y)|^2\Big)^{1/2} + C_M T(1 + \sup_t \|D_x f(t)\|_\infty).$$

To this aim, let us set: $v(t,x) = \sum_{i=1}^d ((u^0_i)^2 + (u_i)^2)$. Then following the computation in the proof of Proposition A.1, we find:

$$\begin{aligned} &-\partial_t v - \Delta v(t,x) \\ &= -2\sum_i \big(u^0_i D_{x_i}(\partial_t u^0 + \Delta u^0) + u_i D_{x_i}(\partial_t u + \Delta u)\big) - 2(|D^2u^0|^2 + |D^2u|^2) \\ &= -2\sum_i \big(u^0_i(h^0_{x_i} + h^0_p\cdot Du^0_i) + u_i(h^0_{p,x_i}\cdot Du + h^0_{pp}Du^0_i\cdot Du + h^0_p\cdot Du_i + f_i)\big) - 2(|D^2u^0|^2 + |D^2u|^2), \end{aligned}$$

so that

$$\begin{aligned}&-\partial_t v-\Delta v(t,x)+h_p^0\cdot Dv\\&=-2\sum_i\big(u_i^0h_{x_i}^0+u_i(h_{p,x_i}^0\cdot Du+h_{pp}^0Du_i^0\cdot Du+f_i)\big)-2(|D^2u^0|^2+|D^2u|^2).\end{aligned}$$

Using our assumption on h^0 we get

$$-\partial_t v-\Delta v(t,x)+h_p^0\cdot Dv\leqslant Cv^{1/2}(|v|^\theta+1+\|D_xf\|_\infty),$$

for some $C>0$ and $\theta>0$ which depend on C_0 and γ only. We derive from this the Lipschitz estimate thanks to the maximum principle exactly as in the proof of Proposition A.1.

The higher order estimates can be checked exactly as in Propositions A.3 and Lemma A.5, so we omit the proof. Note that higher order estimates on D^ru^0 and D^ru depend on $D^{r-1}u^0$ and $D^{r-1}u$, but this dependance affects the constant C_M only.

Let us finally explain how to handle the derivative with respect to y: we note that $\partial_y^\alpha u$ satisfies the same linear equation as u with f replaced by $\partial_y^\alpha f$, and the final datum g is replaced by $\partial_y^\alpha g$. So, in order to estimate $D_xD_y^l(u^0,u)$ for instance, we just set $c_l=(\sum_{|\alpha|=l}1)^{-1}$, $v^\alpha=\sum_{i=1}^d(c_l(u_i^0)^2+(\partial_y^\alpha u_i)^2)$ and $w=\sum_{|\alpha|=l}v^\alpha$. As above,

$$-\partial_t v^\alpha-\Delta v^\alpha(t,x)+h_p^0\cdot Dv^\alpha\leqslant C(v^\alpha)^{1/2}(|v^\alpha|^\theta+1+\|D_x\partial_y^\alpha f\|_\infty)\leqslant Cw^{1/2}(|w|^\theta+1+\|D_xD_y^lf\|_\infty),$$

and summing up one concludes the desired inequality, noting that $w=\sum_{|\beta|=1}(\partial_x^\beta u^0)^2+\sum_{|\beta|=1,|\alpha|=l}(\partial_x^\beta\partial_y^\alpha u)^2$. □

A.2.2 Linear systems

We also need to quantify the regularity of linear systems of the form

$$\left\{\begin{array}{ll}-\partial_t u^0(t,x)-\Delta u^0(t,x)+V^0(t,x)\cdot Du^0(t,x)+f^0(t,x)=0 & \text{in } (0,T)\times\mathbb{R}^d,\\-\partial_t u(t,x;y)-\Delta u(t,x;y)+V^0(t,x)\cdot Du(t,x;y)+V(t,x;y)\cdot Du^0(t,x)&\\\qquad+f(t,x;y)=0 & \text{in } (0,T)\times\mathbb{R}^d,\\u^0(T,x)=g^0(x),\ u(T,x)=g(x;y) & \text{in } \mathbb{R}^d\end{array}\right.\tag{113}$$

Proposition A.9. *Assume that, independently on $t\in(0,1]$, V^0,f^0 are bounded in C^r, and V,f are bounded $C_b^{r,n}$ for some $r,n\geqslant 0$. Then, if (u^0,u) is a solution of (113) which is bounded in $C_b^r\times C_b^{r,n}$ and if $\|g^0\|_r+\|g\|_{r,n}\leqslant M$, we have, for any $T\in(0,1]$, $l\leqslant n$,*

$$\begin{aligned}&\sup_{t,x,y}\Big(|D_x^ru^0(t,x)|^2+|D_x^rD_y^lu(t,x;y)|^2\Big)^{1/2}\\&\qquad\leqslant\sup_{x,y}\Big(|D_x^rg^0(x)|^2+|D_x^rD_y^lg(x;y)|^2\Big)^{1/2}+C_MT,\end{aligned}$$

where C_M depends on M, the bounds on V^0,f^0 and V,f in C^r and $C_b^{r,n}$ respectively.

In addition, for $r=0$ and $l\leqslant n$, we have

$$\sup_{t,x,y}\Big(|u^0(t,x)|^2+|D_y^lu(t,x;y)|^2\Big)^{1/2}\leqslant(1+CT)\sup_{x,y}\Big(|g^0(x)|^2+|D_y^lg(x;y)|^2\Big)^{1/2}+CT(\|f^0\|_\infty+\|D_y^lf\|_\infty),$$

where C depends just on the bound of V^0 and V.

Proof. We first note that the derivatives of u with respect to the parameter y solve a system which has the same structure as the one for u: so we just need to check the result for $n=0$, and proceed as in the proof of Proposition A.8 for $n>0$.

Let us start with the L^∞ bounds: We consider $\tilde v := (u^0)^2 + u^2$. Then v satisfies

$$\begin{aligned}-\partial_t\tilde v - \Delta\tilde v &= -2u^0\left(\partial_t u^0 + \Delta u^0\right) - 2u\left(\partial_t u + \Delta u\right) - 2(|Du^0|^2 + |Du|^2)\\ &= -2u^0\Big(V^0(t,x)\cdot Du^0(t,x) + f^0(t,x)\Big) - 2u\Big(V^0(t,x)\cdot Du(t,x)\\ &\qquad + V(t,x;y)\cdot Du^0(t,x) + f(t,x;y)\Big) - 2(|Du^0|^2 + |Du|^2)\\ &\leqslant C\tilde v + \tilde v^{1/2}(\|f^0\|_\infty + \|f\|_\infty),\end{aligned}$$

where C depends on $\|V^0\|_\infty$ and $\|V\|_\infty$ only. This implies the result for $r = n = 0$.

We now check the C^1 estimate. Let us set as usual $v(t,x) = \sum_{i=1}^d((u_i^0)^2 + (u_i)^2)$. Then

$$\begin{aligned}&-\partial_t v - \Delta v(t,x) = -2\sum_i\left(u_i^0 D_{x_i}(\partial_t u^0 + \Delta u^0) + u_i D_{x_i}(\partial_t u + \Delta u)\right) - 2(|D^2u^0|^2 + |D^2u|^2)\\ &= -2\sum_i\Big(u_i^0(V^0_{x_i}\cdot Du^0 + V^0\cdot Du_i^0 + f_i^0) + u_i(V^0_{x_i}\cdot Du + V^0\cdot Du_i + V_{x_i}\cdot Du^0 + V\cdot Du_i^0 + f_i)\Big)\\ &\qquad - 2(|D^2u^0|^2 + |D^2u|^2)\\ &\leqslant Cv + v^{1/2}(\|Df^0\|_\infty + \|D_x f\|_\infty),\end{aligned}$$

where C depends on the C^1 bound on V^0 and on V and on d only. This implies the estimate for $r = 1$ and $n = 0$.

As for the C^2 estimate, let us set as usual $w(t,x) = \sum_{i,j=1}^d((u_{ij}^0)^2 + (u_{ij})^2)$. Then

$$-\partial_t w - \Delta w(t,x) \leqslant Cw + Cw^{1/2}(1 + \|D^2 f^0\|_\infty + \|D_x^2 f\|_\infty + \|Du^0\|_\infty + \|D_x u\|_\infty),$$

where C depends on the C^1 bound on V^0 and on V and on d only. We then get the estimate for $r = 2$ and $n = 0$ by the maximum principle and using the previous bounds for Du^0, Du.

The estimate on higher order derivatives can be checked in a way similar and we omit the proof. □

B Functions on $\mathcal{P}_2$

B.1 A criterium of differentiability

Here we introduce a simple criterium for a map U, depending of the measure, to be of class C^1.

Lemma B.1. *Let $U : \mathcal{P}_2 \to \mathbb{R}$ be continuous. For $(s,m,y) \in [0,1]\times\mathcal{P}_2\times\mathbb{R}^d$ we set*

$$\hat U(s;m,y) := U((1-s)m + s\delta_y).$$

If the map $s \to \hat U(s;m,y)$ has a derivative at $s = 0$ and if its derivative at 0, $\dfrac{d}{ds}_{|s=0}\hat U : \mathcal{P}_2\times\mathbb{R}^d \to \mathbb{R}$ is continuous and bounded, then U is of class C^1 with

$$\frac{\delta U}{\delta m}(m,y) = \frac{d}{ds}\hat U(0;m,y).$$

Proof. We have to show that, for any $m_0, m_1 \in \mathcal{P}_2$, we have

$$U(m_1) - U(m_0) = \int_0^1\int_{\mathbb{R}^d}\frac{d}{ds}\hat U(0;(1-s)m_0 + sm_1, y)(m_1 - m_0)(dy).$$

Before starting the proof, let us note that the continuity assumption of $\frac{d}{ds}\hat U$ at $s = 0$ implies its continuity at any $s \in [0,1]$, replacing m by $(1-s)m + s\delta_y$.

Let us start by considering the case where m_0 is fixed and m_1 is an empirical measure: $m_1 = m_{\mathbf{y}}^N := \frac{1}{N}\sum_{k=1}^N\delta_{y_k}$ for some $N \in \mathbb{N}$, $N \geqslant 1$, $y_k \in \mathbb{R}^d$. The general case will be treated next by approximation.

All the measures we are going to manipulate in the next lines belong to the set

$$K := \{\alpha_0 m_0 + \sum_{k=1}^{N} \alpha_k \delta_{y_k},\ \alpha_k \geqslant 0, \sum_{k=0}^{N} \alpha_k = 1\}$$

which is compact in $\mathcal{P}_2$. So, by continuity of $\frac{d}{ds}\hat{U}$, if we fix $\epsilon > 0$, there exists $\delta \in (0, 1/2)$ such that, if $m', m'' \in K$ with $\mathbf{d}_2(m, m') < \delta$ and $s \in [0, \delta]$, then

$$\sup_k \left| \frac{d}{ds}\hat{U}(s; m, y_k) - \frac{d}{ds}\hat{U}(0; m', y_k) \right| \leqslant \epsilon. \tag{114}$$

Our first step consists in showing that, for $s > 0$ small enough (to be defined below) and for any $m \in K$, we have

$$\left| U((1-s)m + sm_{\mathbf{y}}^N) - U(m) - s \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0; m, y) m_{\mathbf{y}}^N(dy) \right| \leqslant C(\epsilon s + s^2), \tag{115}$$

where C depends on the sup norm of $\frac{d}{ds}\hat{U}$ on $[0,1] \times K \times \{y_k,\ k = 1, \dots, N\}$. In order to prove (115), we define $\alpha_k = \frac{s}{N-(N-k)s}$ for $k = 0, \dots, N$ and note that

$$\prod_{l=k}^{N} (1 - \alpha_l) = 1 - \frac{(N+1-k)s}{N}\,. \tag{116}$$

We now define by induction

$$m_0 = m, \qquad m_k = (1 - \alpha_k) m_{k-1} + \alpha_k \delta_{y_k} \tag{117}$$

and using (116) we get

$$\begin{aligned} m_N &= \prod_{k=1}^{N} (1-\alpha_k) m + \alpha_n \delta_{y_N} + \sum_{k=1}^{N-1} \alpha_k \delta_{y_k} \prod_{l=k+1}^{N} (1 - \alpha_l) \\ &= (1-s)m + \sum_{k=1}^{N} \delta_{y_k} \frac{s}{N-(N-k)s} (1 - \frac{(N-k)s}{N}) = (1-s)m + sm_{\mathbf{y}}^N. \end{aligned}$$

So, by the definition of m_{k+1} in function of m_k in (117),

$$\begin{aligned} U((1-s)m + sm_{\mathbf{y}}^N) - U(m) &= \sum_{k=0}^{N-1} U(m_{k+1}) - U(m_k) \\ = \sum_{k=0}^{N-1} \hat{U}(\alpha_{k+1}; m_k, y_{k+1}) - \hat{U}(0; m_k, y_{k+1}) &= \sum_{k=0}^{N-1} \int_0^{\alpha_{k+1}} \frac{d}{ds}\hat{U}(\tau; m_k, y_{k+1}) d\tau\,. \end{aligned}$$

Let us assume that $s \in (0, \delta)$. As $s < 1/2$, we have $\alpha_k \leqslant 2s/N$ for any k, and thus

$$\mathbf{d}_2(m_k, m) \ \leqslant Cs$$

for a constant C which depends on m_0 and on the y_k (but not on $m \in K$ nor on $s \in (0, \delta)$). We now require that s is so small that $Cs < \delta$. Then, for any k and any $\tau \in (0, \alpha_k)$, we have by (114):

$$\left| \frac{d}{ds}\hat{U}(\tau; m_k, y_{k+1}) - \frac{d}{ds}\hat{U}(0; m, y_{k+1}) \right| \leqslant \epsilon.$$

We infer from this that

$$\left| U((1-s)m + sm_{\mathbf{y}}^N) - U(m) - \sum_{k=0}^{N-1} \alpha_{k+1} \frac{d}{ds}\hat{U}(0; m, y_{k+1}) \right| \leqslant C\epsilon \sum_{k=0}^{N-1} \alpha_{k+1}.$$

As $|\alpha_k - s/N| \leqslant Cs^2/N$, we conclude that (115) holds.

The next step in the proof consists in showing that

$$U(e^{-1}m_0 + (1-e^{-1})m_{\mathbf{y}}^N) - U(m_0) = \int_0^{e^{-1}} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_{\mathbf{y}}^N, y) m_{\mathbf{y}}^N(dy) \frac{d\tau}{1-\tau}. \tag{118}$$

For this, let us now choose $T \in \mathbb{N}$ large and let

$$m_n = \left(1 - \frac{1}{T}\right)^n m_0 + \left(1 - \left(1 - \frac{1}{T}\right)^n\right) m_{\mathbf{y}}^N \qquad n \in \{0, \dots, T\}.$$

We have

$$m_{n+1} = \left(1 - \frac{1}{T}\right) m_n + \frac{1}{T} m_{\mathbf{y}}^N \qquad n \in \{0, \dots, T\}.$$

So, by (115),

$$\begin{aligned}
&\left| U(m_T) - U(m_0) - T^{-1} \sum_{n=0}^{T-1} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0; m_n, y) m_{\mathbf{y}}^N(dy) \right| \\
&\leqslant \sum_{n=0}^{T-1} \left| U((1-1/T)m_n + (1/T)m_{\mathbf{y}}^N) - U(m_n) - T^{-1} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0; m_n, y) m_{\mathbf{y}}^N(dy) \right| \\
&\leqslant C \sum_{n=0}^{T-1} (\epsilon/T + (1/T)^2) \leqslant C(\epsilon + T^{-1}).
\end{aligned}$$

We let $T \to +\infty$ and then $\epsilon \to 0$ to conclude by continuity of U and of $\frac{d}{ds}\hat{U}$ that

$$\begin{aligned}
U(e^{-1}m_0 + (1-e^{-1})m_{\mathbf{y}}^N) - U(m_0) &= \int_0^1 \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0; e^{-s}m_0 + (1-e^{-s})m_{\mathbf{y}}^N, y) m_{\mathbf{y}}^N(dy) ds \\
&= \int_0^{e^{-1}} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_{\mathbf{y}}^N, y) m_{\mathbf{y}}^N(dy) \frac{d\tau}{1-\tau}.
\end{aligned}$$

This is (118).

By continuity of U and of $\frac{d}{ds}\hat{U}$ and by density of the empirical measures, one obtains from (118) that, for any measure $m_0, m_1 \in \mathcal{P}_2$:

$$U(e^{-1}m_0 + (1-e^{-1})m_1) - U(m_0) = \int_0^{e^{-1}} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_1, y) m_1(dy) \frac{d\tau}{1-\tau}. \tag{119}$$

Choosing $m_1 = m_0$ then implies the normalization convention

$$\int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0; m_0, y) m_0(dy) = 0$$

for any $m_0 \in \mathcal{P}_2$. In particular, this yields

$$\int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_1, y) m_1(dy) = (1-\tau) \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_1, y)(m_1 - m_0)(dy)\,.$$

Inserting this relation in (119) gives the more standard form:

$$U(e^{-1}m_0 + (1-e^{-1})m_1) - U(m_0) = \int_0^{e^{-1}} \int_{\mathbb{R}^d} \frac{d}{ds}\hat{U}(0;(1-\tau)m_0 + \tau m_1, y)(m_1 - m_0)(dy) d\tau.$$

Using again the continuity of U and of $\frac{d}{ds}\hat{U}$, one easily deduce from this the desired equality. □

B.2 Interpolation and Ascoli Theorem in $\mathcal{P}_2$

In the proof of Lemma 3.7, we have used two interpolation Lemmas. The first one is standard (see, for instance, [22, Lemma II.3.1]): we recall it because we need a specific setting. The second one is an adaptation to $\mathcal{P}_2$ of the same techniques.

Lemma B.2. *Let $W : [0,1] \times \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ be Holder continuous in time locally uniformly in space: for any $R > 0$, there exists $C_{0,R} > 0$ and $\alpha_R > 0$ such that*

$$|W(t,y) - W(s,y)| \leqslant C_{0,R}|t-s|^{\alpha} \qquad \forall (s,t,y) \in [0,1] \times [0,1] \times \mathbb{R}^{d_1} \text{ with } |y| \leqslant R \text{ and } |t-s| \leqslant \alpha_R,$$

and such that $D_y W$ is Holder continuous in space uniformly in time: there exists $C_1 > 0$ such that

$$|D_y W(t,y_0) - D_y W(t,y_1)| \leqslant C_1 |y_0 - y_1|^{\delta} \qquad \forall (t,y_1,y_2) \in [0,1] \times \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}.$$

Then $D_y W$ is Holder continuous in time locally uniformly in space:

$$|D_y W(t,y) - D_y W(s,y)| \leqslant C_R \, |t-s|^{\frac{\alpha\delta}{(1+\delta)}}$$
$$\forall (s,t,y) \in [0,1] \times [0,1] \times \mathbb{R}^{d_1} \text{ with } |y| \leqslant R \text{ and } |t-s| \leqslant \alpha'_R,$$

for some constants $C_R > 0$ and α'_R only depending on $C_{0,R+1}$, α_{R+1}, C_1, α and δ.

Remark B.3. The proof below also shows that, if in addition W is Holder continuous in time uniformly in space (i.e., $C_{0,R}$ and α_R do not depend on R) and if $D_y W$ is bounded, then $D_y W$ is also Holder continuous in time uniformly in space.

Proof. Fix $y_0, y_1 \in \mathbb{R}^d$ with $|y_0| \leqslant R$ and $|y_1| \leqslant R+1$. Let $y_\tau = (1-\tau)y_0 + \tau y_1$ for $\tau \in [0,1]$. We have

$$\left| \int_0^1 (D_y W(t,y_\tau) - D_y W(s,y_\tau)) \cdot (y_1 - y_0) d\tau \right|$$
$$= |W(t,y_1) - W(t,y_0) - W(s,y_1) + W(s,y_0)| \leqslant 2C_{0,R+1} \, |t-s|^{\alpha} .$$

So

$$|(D_y W(t,y_0) - D_y W(s,y_0)) \cdot (y_1 - y_0)| \leqslant \left| \int_0^1 (D_y W(t,y_0) - D_y W(t,y_\tau)) \cdot (y_1 - y_0) d\tau \right|$$
$$+ \left| \int_0^1 (D_y W(t,y_\tau) - D_y W(s,y_\tau)) \cdot (y_1 - y_0) d\tau \right| + \left| \int_0^1 (D_y W(s,y_\tau) - D_y W(s,y_0)) \cdot (y_1 - y_0) d\tau \right|$$
$$\leqslant 2C_{0,R+1}|t-s|^{\alpha} + 2C_1 |y_1 - y_0|^{1+\delta},$$

using also the Hölder continuity of $D_y W$. Choosing $y_1 = y_0 + hv$, with $|v| = 1$, we get

$$|[D_y W(t,y) - D_y W(s,y)] \cdot v| \leqslant \frac{2C_{0,R+1}}{|h|} |t-s|^{\alpha} + 2C_1 |h|^{\delta} .$$

Optimizing with respect to $h \in (0, \alpha_{R+1}]$ and $|v| = 1$, we find the result for $|t-s| \leqslant \alpha'_R$ for a suitable constant α'_R depending on $C_{0,R+1}$, α, C_1 and δ. □

Lemma B.4. *Let $W : [0,1] \times \mathcal{P}_2 \to \mathbb{R}^{d_2}$ be Holder continuous, locally in time and uniformly in measure: there exists $\alpha \in (0,1]$ and, for any $R > 0$ there exists $C_{0,R} > 0$ such that*

$$|W(t,m) - W(s,m)| \leqslant C_{0,R}|t-s|^{\alpha} \qquad \forall m \in \mathcal{P}_2 \text{ with } M_2(m) \leqslant R, \ \forall s,t \in [0,1],$$

(where $M_2(m) = (\int_{\mathbb{R}^d} |y|^2 m(dy))^{1/2}$) and such that $\frac{\delta W}{\delta m}$ and $D_m W$ are bounded and $D_m W$ is Holder continuous with respect to the measure uniformly in time: there exists $\gamma, \delta \in (0,1]$ and $C_1 > 0$ such that

$$|D_m W(t,m_0,y_0) - D_m W(t,m_1,y_1)| \leqslant C_1 \left(\mathbf{d}_2^{\gamma}(m_0,m_1) + |y_0 - y_1|^{\delta} \right)$$

for any $t \in [0,1]$ and any $(m_i, y_i) \in \mathcal{P}_2 \times \mathbb{R}^d$. Then $D_m W$ is Holder continuous in time locally uniformly in $(m,y) \in \mathcal{P}_2 \times \mathbb{R}^d$: for any $R > 0$, there exists a constant $C_R > 0$, depending on R, $\|D_m W\|_\infty$, $C_{0,R+1}$, C_1, α, γ and δ, such that

$$|D_m W(t,m,y) - D_m W(s,m,y)| \leqslant C_R |t-s|^{\alpha\gamma/((2+\gamma)(1+\delta))},$$

for any $s,t \in [0,1]$ and any $(m,y) \in \mathcal{P}_2 \times \mathbb{R}^d$ with $|y| \leqslant R$ and $M_2(m) \leqslant R$.

Proof. Let $R \geqslant 1$. Fix $m_0, m_1 \in \mathcal{P}_2$ with $M_2(m_i) \leqslant R$ and set $m_\tau = (1-\tau)m_0 + \tau m_1$. Then

$$\begin{aligned}
&\left| \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_\tau, y) - \frac{\delta W}{\delta m}(s, m_\tau, y) \right) (m_1 - m_0)(dy) d\tau \right| \\
&\quad = |W(t,m_1) - W(t,m_0) - W(s,m_1) + W(s,m_0)| \\
&\quad \leqslant 2C_{0,R} |t-s|^\alpha .
\end{aligned}$$

As

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_0, y) - \frac{\delta W}{\delta m}(s, m_0, y) \right) (m_1 - m_0)(dy) \right| \\
&\quad \leqslant \left| \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_\tau, y) - \frac{\delta W}{\delta m}(s, m_\tau, y) \right) (m_1 - m_0)(dy) d\tau \right| \\
&\qquad + \left| \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_\tau, y) - \frac{\delta W}{\delta m}(t, m_0, y) \right) (m_1 - m_0)(dy) d\tau \right| \\
&\qquad + \left| \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(s, m_\tau, y) - \frac{\delta W}{\delta m}(s, m_0, y) \right) (m_1 - m_0)(dy) d\tau \right| ,
\end{aligned}$$

we obtain, by our Holder continuity assumption on $D_m W$:

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_0, y) - \frac{\delta W}{\delta m}(s, m_0, y) \right) (m_1 - m_0)(dy) \right| \\
&\quad \leqslant 2C_0 |t-s|^\alpha + \sup_{\tau \in [0,1]} \|D_m W(t, m_\tau, \cdot) - D_m W(t, m_0, \cdot)\|_\infty \, \mathbf{d}_1(m_0, m_1) \\
&\qquad + \sup_{\tau \in [0,1]} \|D_m W(s, m_\tau, \cdot) - D_m W(s, m_0, \cdot)\|_\infty \, \mathbf{d}_1(m_0, m_1) \\
&\quad \leqslant 2C_0 |t-s|^\alpha + 2C_1 \mathbf{d}_2^\gamma(m_0, m_1) \mathbf{d}_1(m_0, m_1).
\end{aligned}$$

For any $y_0 \in \mathbb{R}^d$ with $|y_0| \leqslant R$, let $m_1 = (1-\theta)m_0 + \theta \delta_{y_0}$ for some $\theta \in (0,1]$ to be chosen below. Note that

$$\mathbf{d}_1(m_1, m_0) \leqslant \theta \int_{\mathbb{R}^d} |y_0 - x| m_0(dx) \leqslant \theta (|y_0| + (M_2(m_0))^{1/2}) \leqslant 2\theta R,$$

(since $R \geqslant 1$) while

$$\mathbf{d}_2(m_1, m_0) \leqslant (\theta \int_{\mathbb{R}^d} |y_0 - x|^2 m_0(dx))^{1/2} \leqslant (2\theta)^{1/2} (|y_0|^2 + M_2^2(m_0))^{1/2} \leqslant 2\theta^{1/2} R.$$

We get, by the convention on the derivative and our previous estimates:

$$\begin{aligned}
&\left| \frac{\delta W}{\delta m}(t, m_0, y_0) - \frac{\delta W}{\delta m}(s, m_0, y_0) \right| = \frac{1}{\theta} \left| \int_{\mathbb{R}^d} \left(\frac{\delta W}{\delta m}(t, m_0, y) - \frac{\delta W}{\delta m}(s, m_0, y) \right) (m_1 - m_0)(dy) \right| \\
&\quad \leqslant \frac{1}{\theta} \left[2C_{0,R} |t-s|^\alpha + c C_1 R^{1+\gamma} \theta^{1+\gamma/2} \right] ,
\end{aligned}$$

where c is universal. If $|t-s|$ is small enough such that $C_{0,R}|t-s|^\alpha / (cC_1 R^{1+\gamma}) \leqslant 1$, then we choose $\theta^{1+\gamma/2} := C_{0,R}|t-s|^\alpha / (cC_1 R^{1+\gamma})$ and obtain

$$\left| \frac{\delta W}{\delta m}(t, m_0, y_0) - \frac{\delta W}{\delta m}(s, m_0, y_0) \right| \leqslant c C_{0,R}^{\gamma/(2+\gamma)} C_1^{1/(1+\gamma/2)} R^{2(1+\gamma)/(2+\gamma)} |t-s|^{\alpha\gamma/(2+\gamma)} , \tag{120}$$

where c is another universal constant.

To show the regularity in time of $D_m W$, we just need to apply Lemma B.2 to $\delta W/\delta m$ since, by (120), $\delta W/\delta m$ is locally Holder in time locally uniformly in space (the constant depending also on the measure) and $D_y \delta W/\delta m = D_m W$ is globally bounded and Holder in y uniformly in time by assumption. We can remove the smallness restriction on $|t-s|$ by using the fact that $D_m W$ is globally bounded. □

In the proof of Theorem 3.2 we also used the following version of Arzela-Ascoli Theorem.

Lemma B.5. *Let (X,d) be a locally compact space and $W^N : X \times \mathcal{P}_2 \to \mathbb{R}$ be a family of uniformly bounded and locally uniformly continuous maps: there exists $x_0 \in X$ such that, for any $R>0$, there exists a continuous nondecreasing modulus $\omega_R : [0,+\infty) \to [0,+\infty)$ with $\omega_R(0)=0$ such that*

$$|W^N(x,m) - W^N(x',m')| \leqslant \omega_R(d(x,x') + \mathbf{d}_2(m,m')), \tag{121}$$

for any $x, x' \in X$ and $m, m' \in \mathcal{P}_2$ with $d(x,x_0) \leqslant R$, $d(x',x_0) \leqslant R$, $M_2(m) \leqslant R$, $M_2(m') \leqslant R$.

Then there exists a continuous map $W : X \times \mathcal{P}_2 \to \mathbb{R}$ and a subsequence (denoted in the same way) such that (W^N) converges to W pointwisely in m and locally uniformly in x: for any $R>0$ and any $m \in \mathcal{P}_2$,

$$\lim_{N \to +\infty} \sup_{d(x,x_0) \leqslant R} |W^N(x,m) - W(x,m)| = 0\,. \tag{122}$$

The only (very small) issue in the result is that $\mathcal{P}_2$ is not locally compact, so that the standard Arzela-Ascoli Theorem cannot be applied.

Proof. Let D be an enumerable dense family of $X \times \mathcal{P}_2$. By a diagonal argument we can find a subsequence (denoted in the same way) such that, for any $(x,m) \in D$, $(W^N(x,m))$ converges to some $W(x,m)$. Let us note that, by our regularity assumption (121) and using the fact that $X \times \mathcal{P}_2$ is complete, W can be extended to the whole space $X \times \mathcal{P}_2$ into a continuous map which satisfies

$$|W(x,m) - W(x',m')| \leqslant \omega_R(d(x,x') + \mathbf{d}_2(m,m')), \tag{123}$$

for any $x, x' \in X$ and $m, m' \in \mathcal{P}_2$ with $d(x,x_0) \leqslant R$, $d(x',x_0) \leqslant R$, $M_2(m) \leqslant R$, $M_2(m') \leqslant R$.

We claim that, for any $(x,m) \in X \times \mathcal{P}_2$, $(W^N(x,m))$ converges to $W(x,m)$. Indeed, fix $\epsilon > 0$, $R = 2(1 + d(x,x_0) + M_2(m))$. Then there is $(x',m') \in D$ such that $d(x',x_0) \leqslant R$, $M_2(m') \leqslant R$ and $\omega_R((d(x,x') + \mathbf{d}_2(m,m')) \leqslant \epsilon/3$. Let also N_0 be so large that $|W^N(x',m') - W(x',m')| \leqslant \epsilon/3$ for $N \geqslant N_0$. Then, for $N \geqslant N_0$, we have

$$\begin{aligned} &|W^N(x,m) - W(x,m)| \\ &\quad \leqslant |W^N(x,m) - W^N(x',m')| + |W^N(x',m') - W(x',m')| + |W(x',m') - W(x,m)| \leqslant \epsilon\,, \end{aligned}$$

where we used (121) and (123) in the last inequality.

It remains to show that (122) holds. Fix $\epsilon > 0$ and let $\eta > 0$ be such that $\omega(\eta) \leqslant \epsilon/3$. As X is locally compact, we can find $x_1, \dots x_n$ such that any point $x \in B_X(x_0,R)$ is at a distance at most η from one of the $(x_i)_{i=1,\dots,n}$. Let N_0 be so large that $|W^N(x_i,m) - W(x_i,m)| \leqslant \epsilon/3$ for any $i = 1, \dots, n$. Then, for any $x \in B_X(x_0,R)$ and any $N \geqslant N_0$, we have (for i such that $d(x,x_i) \leqslant \eta$, so that $\omega_R(d(x,x_i)) \leqslant \epsilon/3$):

$$\begin{aligned} &|W^N(x,m) - W(x,m)| \\ &\quad \leqslant |W^N(x,m) - W^N(x_i,m)| + |W^N(x_i,m) - W(x_i,m)| + |W(x_i,m) - W(x,m)| \leqslant \epsilon\,, \end{aligned}$$

where we used again (121) and (123) in the last inequality. This shows (122). □

References

[1] Achdou, Y., Lasry, J.-M., Lions, P.-L., Rostand, A., Scheinkman, J., *Models for the Economy of Oil*, in preparation.

[2] Ahuja, S. (2016). *Wellposedness of mean field games with common noise under a weak monotonicity condition.* SIAM Journal on Control and Optimization, 54(1), 30-48.

[3] Albeverio, S., Kondratiev, Y. G., and Röckner, M. (1998). *Analysis and geometry on configuration spaces.* Journal of functional analysis, 154(2), 444-500.

[4] Ambrosio, L., Gigli, N., Savaré, G. Gradient flows in metric spaces and in the space of probability measures. Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008.

[5] Bensoussan, A., Chau, M. H. M., Yam, S. C. P. (2015). *Mean field Stackelberg games: Aggregation of delayed instructions.* SIAM Journal on Control and Optimization, 53(4), 2237-2266.

[6] Bensoussan, A., Chau, M. H. M., Yam, S. C. P. (2016). *Mean field games with a dominating player.* Applied Mathematics & Optimization, 74(1), 91-128.

[7] Buckdahn, R., Li, J., Peng, S. (2014). *Nonlinear stochastic differential games involving a major player and a large number of collectively acting minor agents.* SIAM Journal on Control and Optimization, 52(1), 451-492.

[8] Caines, P. E., Kizilkale, A. C. (2014). *Mean field estimation for partially observed LQG systems with major and minor agents.* IFAC Proceedings Volumes, 47(3), 8705-8709.

[9] Cardaliaguet P., Delarue F., Lasry J.-M., Lions P.-L. (2019). The Master Equation and the Convergence Problem in Mean Field Games. (AMS-201) (Vol. 381). Princeton University Press.

[10] Cardaliaguet P., Cirant M., Porretta A. *Remarks on Nash equilibria in mean field game models with a major player.* Preprint.

[11] Carmona R., Delarue F. Probabilistic Theory of Mean Field Games with Applications I & II. 2017 - Springer Verlag. DOI 10.1007/978-3-319-58920-6

[12] Carmona, R., Delarue, F., and Lacker, D. (2016). *Mean field games with common noise.* The Annals of Probability, 44(6), 3740-3803.

[13] Carmona, R., Wang, P. (2016). Finite state mean field games with major and minor players. arXiv preprint arXiv:1610.05408.

[14] Carmona, R., Wang, P. (2017). An alternative approach to mean field game with major and minor players, and applications to herders impacts. Applied Mathematics & Optimization, 76(1), 5-27.

[15] Carmona, R., Zhu, X. (2016). *A probabilistic approach to mean field games with major and minor players.* The Annals of Applied Probability, 26(3), 1535-1580.

[16] Chassagneux, J. F., Crisan, D., and Delarue, F. (2014). *Classical solutions to the master equation for large population equilibria.* arXiv preprint arXiv:1411.3009.

[17] Gangbo, W., Swiech A. (2015) *Existence of a solution to an equation arising from the theory of mean field games.* Journal of Differential Equations, 259(11), 6573-6643.

[18] Gilbarg, D. and Trudinger, N.S. Elliptic partial differential equations of second order. Springer, 2015.

[19] Huang, M., Malhamé, R.P., Caines, P.E. (2006). *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle.* Communication in information and systems. Vol. 6, No. 3, pp. 221-252.

[20] Huang, M. (2010). *Large-population LQG games involving a major player: the Nash certainty equivalence principle.* SIAM Journal on Control and Optimization, 48(5), 3318-3353.

[21] Lacker, D. and Webster, K. (2015). *Translation invariant mean field games with common noise.* Electronic Communications in Probability, 20.

[22] Ladyženskaja O.A., Solonnikov V.A. and Ural'ceva N.N. Linear and quasilinear equations of parabolic type. Translations of Mathematical Monographs, Vol. 23 American Mathematical Society, Providence, R.I. 1967.

[23] Lasry, J.-M., Lions, P.-L. *Jeux à champ moyen. I. Le cas stationnaire.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 9, 619-625.

[24] Lasry, J.-M., Lions, P.-L. *Jeux à champ moyen. II. Horizon fini et contrôle optimal.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 10, 679-684.

[25] Lasry, J.-M., Lions, P.-L. *Mean field games.* Jpn. J. Math. 2 (2007), no. 1, 229–260.

[26] Lasry, J. M., Lions, P. L. (2018). *Mean-field games with a major player.* C. R. Math. Acad. Sci. Paris.

[27] Lions, P.L. Cours au Collège de France. www.college-de-france.fr.

[28] Mayorga S. *Short time solution to the master equation of a first order mean field game.* arXiv:1811.08964, 2018

[29] Lions P.-L. *Estimées nouvelles pour les équations quasilinéaires.* Seminar in Applied Mathematics at the Collège de France. http://www.college-de-france.fr/site/pierrelouis-lions/seminar-2014-11-14-11h15.htm, 2014.

[30] Rachev, S.T., and Schendorf, L. R. Mass Transportation problems. Vol. I: Theory; Vol. II : Applications. Springer-Verlag, 1998.

[31] Villani, C. Optimal transport, old and new. Springer-Verlag, 2008.