THE UNIVERSITY OF CHICAGO

PROBLEMS IN MEAN FIELD GAMES AND MEAN FIELD CONTROL

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

BY
NIKIFOROS MIMIKOS-STAMATOPOULOS

CHICAGO, ILLINOIS
AUGUST 2024

To my advisors, family, and friends

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Professor Panagiotis Souganidis for all his support and guidance throughout my PhD studies. My academic and personal development would not have been possible without his continuous mentorship, support and expertise.

In the same spirit, I have to express my gratitude to Professor Luis Silvestre for all his support and teachings throughout my studies, as well as faculty members Charles Smart and Carlos Kenig.

Moreover, I want to extend my gratitude to all my friends and colleagues Mark Cerenzia, Peter Morfe, Mariya Sardali, Sebastian Munoz, Antonios Zitridis, Henrik Mathiensen, Joe Jackson, Bill Cooperman, Stephen Yearwood, Sam Quin, Andrea Iorga, Oliver Wang, David Bowman, Sehyun Ji, Faidon Andriopoulos, Jason Kountouridis, Kostas Psaromiligos and Santiago Chavez-Aguilar. This list cannot possibly be exhaustive with all the wonderful people I had the luck of meeting, so I apologize in advance.

Outside of the University of Chicago, I have to thank especially Professor Cardaliaguet and Samuel Daudin for all the support and collaborations throughout the past six years.

Finally, I want to thank Ioanna, my parents, and my sister for all the support they have given me.

ABSTRACT

This thesis studies various problems that arise in the study of Mean Field Games and Mean Field Control. First we prove that a hypoelliptic MFG system with local-coupling is well posed, extending results from the the parabolic literature. Next, we show that a first-order local MFG system, as well as the planning problem, admit smooth solutions and characterize their long time behavior in one dimension. Finally, we show that for smooth, but not necessarily convex data, there exists a open and dense region, in which the value function in Mean Field Control, as well as the optimal controls, converge at a rate of $\frac{1}{N}$.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This thesis is concerned with problems arising in Mean Field Games (MFG for short) and Mean Field Control (MFC for short). MFG were introduced by Lasry and Lions [286, 282, 284], and at the same time, in a particular setting, by Caines, Huang, and Malhamé [239]. The theory studies asymptotic equilibria of $N-$player games, for large N, of weakly interacting agents. Such games arise naturally in numerous fields such as economics, networks and traffic control to name a few. In the aforementioned applications, analyzing equilibrium states, such as Nash equilibria, is a notoriously challenging problem. The main idea of MFG is to exploit the fact that under suitable assumptions, as $N-$becomes large, agents become indistinguishable. Consequently, each agent's decision depends on the distribution of the others. In turn, as N tends to infinity, this leads to a more tractable game with a continuum of players. What MFG propose is to study the equilibria in the infinite population game and then infer qualitative and quantitative properties for the finite game. This last step is achieved by establishing asymptotic convergence to the infinite regime.

Broadly speaking, we may separate the study of the aforementioned games in two categories. The first category addresses non-cooperative games and a typical application is the study of Nash equilibria. The second one concerns cooperative games, and a model example is that of the central planner. This thesis will primarily focus on the latter, however in both setups the primary tool for the analysis of the limiting system and the convergence problems, is the MFG system, which we introduce next.

The MFG system is a coupled system of an HJB equation, and a Fokker-Planck equation (FP for short). It typically reads

$$\begin{cases} -\partial_t u - \nu\Delta u + H(x, Du) = F(x, m(t)), \ u(T, x) = G(x, m(T)), x \in \mathbb{R}^d, \\ \partial_t m - \nu\Delta m - \operatorname{div}(mH_p(x, Du)) = 0, \ m(t_0) = m_0. \end{cases} \tag{MFG}$$

Systems like (MFG) capture the equilibrium state for a population, with initial density m_0. The backwards HJB equation describes the evolution in time of a generic agent, while the FP equation provides the evolution of the density of the players. Due to their forward-backward structure, (MFG) falls outside the scope of the classical theory of evolution equations and their study necessitates the development of new approaches.

Although there exist many variants of system (MFG), we mention the following important subcategories:

- When diffusion is present, that is $\nu > 0$, it is called a second order system, for $\nu = 0$ we say its a first order system. The former case corresponds to stochastic games while the latter to deterministic.

- When the coupling terms F, G take inputs in $\mathcal{P}(\mathbb{R}^d)$ it is called a non-local system. In this case, agents interact only through the distribution of the others. Moreover, typically in this case F, G have a regularizing effect on the distribution.

- Lastly, when the couplings F, G depend on the density, that is $F = F(x, m(t, x))$, then the system is called local. We may interpret this coupling as each agent being affected by the surrounding players.

It is important to note that in all the above cases, the MFG system is expected to have uniqueness and stability when the couplings are increasing, in a appropriate sense, in the measure argument. Finally, in all of the above, in order to avoid technical steps related to boundary conditions, when (MFG) is studied in a bounded domain it is common to do so over the torus $\mathbb{T}^d$.

The study of system (MFG) and its variants, whether as a standalone problem or related to regularity in MFC (see Chapter 4), lead to many interesting questions on their own. Here we will focus on:

1. Existence and uniqueness of solutions. In Chapter 2, we study this question for a hypoelliptic version of a local, MFG system.

2. Regularity of solutions and long time behavior. In Chapter 3, we study this question for a one-dimensional, local and first-order MFG system.

3. Stability with respect to m_0. In Chapter 4, we study this question, among others, were we establish convergence rates for Mean Field Control problems.

CHAPTER 2

HYPOELLIPTIC MEAN FIELD GAMES SYSTEM

2.1 Introduction

The goal of this chapter is to establish the well posedness (existence and uniqueness) of solutions of the local, hypoelliptic Mean Field Games system

$$\begin{cases} -\partial_t u - \Delta_v u + v \cdot D_x u + H(D_v u) = F(t,x,v,m(t,x,v)) \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ \partial_t m - \Delta_v m - v \cdot D_x m - \mathrm{div}_v(m H_p(D_v u)) = 0, \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ u(T,x,v) = G(x,v,m(T,x,v)), m(0,x,v) = m_0(x,v). \end{cases} \tag{2.1.1}$$

The Hamiltonian $H : \mathbb{R}^d \to \mathbb{R}$ is convex, the coupling term $F : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ as well as the terminal cost function $G : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ are increasing in m, and $m_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a given probability density.

As described in the introduction, systems like (2.1.1) formally describe the equilibrium of an N-player game, when N tends to infinity, of indistinguishable players, where each player makes decisions based on the distribution of the other co-players. In this setup, it is natural to interpret $x \in \mathbb{R}^d$ as the position and $v \in \mathbb{R}^d$ as the velocity of such players. More precisely, the players control their accelaration in order to minimize the cost introduced by the coupling F and the Hamiltonian H, which leads to the Hamilton-Jacobi-Bellman equation (HJB for short). The optimal feedback is then given by the vector field $-(v, D_p H(D_v u))$, under which, their distribution changes according to the degenerate Fokker-Planck equation (FP for short). As far as applications are concerned, we refer to the flocking model in Carmona and Delarue [97], and for a first order system we refer to Bardi and Cardaliaguet [21], Griffin and Meszaros [225] and Achdou, Mannucci, Marchi and Tchou [2]. Finally, we mention that the general form of (2.1.1) is reminiscent of Boltzmann-type equations, which have been investigated in the MFG context by Burger, Lorz, Wolfram [68] in a setting different to the one used here.

Although there has been extensive study of non-degenerate second-order mean field games, with a local or non-local coupling, less has been done in the degenerate setting, an example of the latter being hypoelliptic MFG. In this setting, when the degeneracy is a sum of squares, Dragoni and Feleqi studied in [157] the ergodic problem; see also Feleqi, Gomes and Tada [175]. When $H(p) = \frac{1}{2}|p|^2$, Camilli in [71], obtained, using the Hopf-Cole transformation, weak solutions to (2.1.1) with uncoupled terminal data. We remark that the assumptions of Camilli appear almost complementary to the ones in this work, as the existence of solutions in [71] is established for terminal data that have to be unbounded since they need to be superquadratic. For results in the case of non-Hörmander degenerate systems, we refer to Cardaliaguet, Graber, Porretta and Tonon in [83], who study, using a variational approach, degenerate MFG systems, for Hamiltonians with super-linear growth and no coupling on the terminal data of the HJB equation.

Our goal is to show existence and uniqueness for quadratic and Lipschitz Hamiltonians, under similar assumptions as that of Porretta in [325], where existence and uniqueness of renormalized solutions was established in the non-degenerate setting. We work with two different types of Hamiltonian H, that is, with linear or quadratic growth. Furthermore, the degeneracy is not a sum of squares, that is, L is not of the form $L := \sum_{i,j}^{k} a_{ij} X_i X_j$, for some vector fields X_i satisfying Hörmanders condition . In the context of hypoelliptic operators, the degenerate operator $L := \partial_t - \Delta_v + v \cdot D_x$ is the simplest and historically the first one to be studied.

The first result addresses the case of a Lipschitz Hamiltonian, whereas the latter the case of quadratic Hamiltonian.

Theorem 1. *Assume that H, F, G, and m_0 satisfy 2.2,2.2,2.2, and 2.2. Then, there exists a unique weak solution (u, m) of* (2.1.1)*, according to Definition 1. Moreover, there exists a constant $C > 0$, such that,*

$$\| - \partial_t u + v \cdot D_x u\|_{L^2([0,t]\times\mathbb{R}^d\times\mathbb{R}^d)} + \|\Delta_v u\|_{L^2([0,t]\times\mathbb{R}^d\times\mathbb{R}^d)}$$

$$+\| - \partial_t m + v \cdot D_x m\|_{L^2([0,t]\times\mathbb{R}^d\times\mathbb{R}^d)} + \|\Delta_v m\|_{L^2([0,t]\times\mathbb{R}^d\times\mathbb{R}^d)} \leq \frac{C}{T-t}.$$

Furthermore, if F also satisfies 2.2, there exists a constant $C = C(F, G, H, T, m_0) > 0$, such that

$$\sup_{t\in[0,T]} \|m(t)\|_2 + \sup_{t\in[0,T]} \|Dm(t)\|_2 + \|D^2_{v,v}m\|_2 + \|D_v D_x m\|_2 \le C,$$

and

$$\sup_{t\in[0,T]} \|u(t)\|_2 + \sup_{t\in[0,T]} \|Du(t)\|_2 + \|D^2_{v,v}u\|_2 + \|D_v D_x u\|_2 \le C.$$

The second result is about renormalized solutions as in Definition 4.

Theorem 2. *Assume that H, F, G, and m_0 satisfy* (2.2),(2.2),(2.2), *and* (2.2). *Then, there exists a unique pair (u, m), of renormalized solutions of the MFG system* (2.1.1). *Furthermore, assume that F, G are only functions of m. Then, there exists a constant $C = C(m_0, F, G, T) > 0$, such that*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} G'(m(T,x,v))|Dm(T,x,v)|^2 dxdv + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} F'(m(t,x,v))|Dm(t,x,v)|^2$$

$$+m \sum_{k=1}^{2d} mD_v u_k H_{pp}(D_v u) D_v u_k dxdv \le C.$$

The existence of a solution, in the case of Lipschitz Hamiltonians, is established using a Schauder fixed point theorem as follows. Fix a probability density m_0. Given $\mu \in X := C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d))$, let $u^\mu \in C([0,T]; L^2(\mathbb{R}^d\times\mathbb{R}^d))$, with $D_v u \in L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, be the unique, distributional solution of

$$\begin{cases} -\partial_t u - \Delta_v u + v\cdot D_x u + H(D_v u) = F(t,x,v,\mu) \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ u(T,x,v) = G(\mu(T,x,v)) \text{ in } \mathbb{R}^d\times\mathbb{R}^d, \end{cases}$$

and m the unique distributional solution of

$$\begin{cases} \partial_t m - \Delta_v m - v\cdot D_x m - \operatorname{div}_v(mD_pH(D_v u^\mu)) = 0 \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ m(0,x,v) = m_0(x,v) \text{ in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases}$$

Set $\Phi(\mu) = m$. We need to show that Φ is $X-$valued, continuous, and compact. The two aforementioned properties follow easily once we show that $\Phi(m) \in L^\infty$ with appropriate bounds. Compactness does not follow immediately, because of the degenerate $x-$direction. To work with that, we localize in time the results in Bouchut [59].

For Theorem 2, we rely on the work in [325] and mostly adapt the arguments in the hypoelliptic setting. In particular, given a Hamiltonian H with quadratic growth (exact assumptions are given later), we consider a sequence of Lipschitz pointwise-approximations and the corresponding solutions provided by Theorem 1 and show compactness in the appropriate spaces. The main technical difficulties and deviations from [325] are the gradient estimates in hypoelliptic equations with L^1-data, which are briefly described next. Let H^ϵ be a suitable pointwise Lipschitz approximation of a quadratic Hamiltonian H and (m^ϵ, u^ϵ) the corresponding weak solutions. In order to show that there exists a limit which is a renormalized solution, we must show the convergence (up to a subsequence) of u^ϵ, m^ϵ in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ and of the gradients $D_v(u^\epsilon \wedge k), D_v(m^\epsilon \wedge k)$ of the truncations in $L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$. The compactness of u^ϵ in L^1 follows by the results of DiPerna and Lions in [151], while the convergence of the gradients is due to an appropriate transformation similar to the one used by Porretta in [323] and the references therein. This important transformation is studied in the Appendix. Finally, for the FP equation, the crucial bound as pointed out in [325] is that, for some independent of ϵ, $C > 0$,

$$\|m^\epsilon |H_p^\epsilon(D_v u^\epsilon)|^2\|_1 \leq C. \tag{2.1.2}$$

This estimate is crucial in the following way. If m^ϵ is a solution to the FP equation (2.1.1), a priori, the best independent of ϵ estimate for $m^\epsilon H_p^\epsilon(D_v u^\epsilon)$ is in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$. However, to obtain fractional gradient estimates we need bounds in L^r for some $r > 1$. The main observation that allows us to obtain this under condition (2.1.2), is the following: Due to hypoellipticity, higher integrability of $m^\epsilon H_p^\epsilon(D_v u^\epsilon)$ should yield higher integrability for m^ϵ, while under condition (2.1.2) higher integrability of m^ϵ should also yield higher integrability for $m^\epsilon H_p^\epsilon(D_v u^\epsilon)$. We show that it is possible to combine the above gains and obtain higher integrability with bounds independent of ϵ and therefore use the results from [59].

2.1.1 Organization of Chapter 2

In section 1, we state all the assumptions and definitions used throughout the chapter. In section 2, we study the backwards HJB and FP equations with L^2-terminal/initial data respectively. The main estimates come from Theorems 17 and 18. We also obtain results regarding the hypoelliptic FP equation and, in particular, we establish fractional gradient bounds. Finally, we establish Theorem 1. Section 3 is devoted to the proof of Theorem 2. Finally, in the appendix (section 4) we show an important technical result for the hypoelliptic HJB equation and we give the statements of the theorems we will use from [59].

2.1.2 Notation and Terminology

Throughout the chapter, $d \in \mathbb{N} := \{1, \cdots, \infty\}$, $T > 0$ is the terminal time, $t \in [0,T]$ is the time variable, $x \in \mathbb{R}^d$ and $v, \nu \in \mathbb{R}^d$, and vectors in $[0,T] \times \mathbb{R}^d \times \mathbb{R}^d$ always appear in the order (t,x,v). For $p \in [1,\infty]$, $L^p([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)_+$ and $L^p(\mathbb{R}^d \times \mathbb{R}^d)_+$, are the non-negative functions of $L^p([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ and $L^p(\mathbb{R}^d \times \mathbb{R}^d)$ respectively. For $s > 0, W^{s,p}(\mathbb{R}^d \times \mathbb{R}^d)$ is the usual fractional Sobolev space and $D^s = (-\Delta)^{s/2}$, we refer for example to [148] for the definition of fractional Sobolev spaces. If $\phi = \phi(t,x,v) : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ or $\phi = \phi(x,v) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, we write $D^2\phi = D^2_{x,v}\phi$, for the hessian in the space variables, $\Delta_v \phi := \sum_{i=1}^{d} \partial_{v_i v_i}\phi$, $D\phi := (D_x\phi, D_v\phi)$ and $\text{div}_v(\phi) := \sum_{i=1}^{d} \partial_{v_i}\phi$. For a function $F(t,x,v,m) : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ or $G(x,v,m) : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, we use the notations $D_{(x,v)}F = (\partial_{x_1}F, \cdots, \partial_{x_d}F, \partial_{v_1}F, \cdots, \partial_{v_d}F)$, $F_m = \partial_m F$, and similarly for G. Throughout the chapter when we reference a standard sequence of mollifiers $\rho_n : \mathbb{R}^d \times \mathbb{R}^d \to [0,\infty)$ we mean that $\rho_n(x,v) := n^{2d}\rho(\frac{x}{n}, \frac{v}{n})$ where $\rho \in C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, such that $\rho \geq 0$ and $\int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(x,v)dxdv = 1$. Moreover in all the proofs constants are subject to change from line to line and they only depend on the quantities/functions stated in the statement of the result.

Finally, we will often use the terminology dimensional constant referring to a constant that only depends on the dimension.

2.2 Assumptions/Definitions

We split this section in two subsections, one for Lipschitz Hamiltonian and one for quadratic.

Lipschitz Hamiltonian and weak solutions

As far as the data are concerned, we assume the following, for the case of Lipschitz Hamiltonian:

[H1] (Lipschitz Hamiltonian) The Hamiltonian $H : \mathbb{R}^d \to \mathbb{R}$, is $C^1(\mathbb{R}^d)$, convex, $H \geq 0$, $H(0) = 0$, and there exists an $L_H > 0$, such that,

$$|H(p_2) - H(p_1)| \leq L_H |p_2 - p_1| \text{ for all } p_1, p_2 \in \mathbb{R}^d. \tag{H1.1}$$

[F1] (Coupling term) The coupling term $F = F(t, x, v, m) : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, is continuous, strictly increasing and locally Lipschitz in m, that is, for all $L > 0$, there exists a constant $c_L > 0$ such that $|F(t, x, v, m_2) - F(t, x, v, m_1)| \leq c_L |m_2 - m_1|$ for all $0 \leq m_1, m_2 \leq L$, and $F(t, x, v, 0) \in L^2([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)$. Finally, we assume that $F \geq 0$.

[G1] (Terminal data for u) The coupling term $G = G(x, v, m) : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, is continuous, strictly increasing and locally Lipschitz in m (in the same sense as F above), and $G(x, v, 0) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$. Finally, we assume that $G \geq 0$.

[M1] (Initial density) The initial density $m_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, satisfies $m_0 \in L^\infty([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)_+$, $\sqrt{m_0} \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$, $(|x|^2 + |v|^4) m_0 \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$, $\log(m_0) \in L^1_{loc}(\mathbb{R}^d \times \mathbb{R}^d)$, $Dm_0 \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$ and $\int_{\mathbb{R}^d \times \mathbb{R}^d} m_0(x, v) dx dv = 1$.

[R1] (Regularity) Assume that F, G satisfy 2.2,2.2 and that for every $L > 0$, there exists a $c_0 = c_0(L) > 0$, such that,

$$c_0 \leq |F_m(t,x,v,m)|, |G_m(x,v,m)|, \text{ for all } (t,x,v,m) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times [0,L].$$

Furthermore, we assume that there exists a constant $C > 0$, such that,

$$|D_{(x,v)}F(t,x,v,m)| + |D_{(x,v)}G(x,v,m)| \leq C|m| \text{ for all } (t,x,v,m) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}.$$

Remark 1. *We note that assumption 2.2 implies in particular that* $m_0 \log(m_0) \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$.

Next we state the definition of a weak solution.

Definition 1. *Assume that* H, G, F, *and* m_0 *satisfy 2.2,2.2,2.2 and 2.2. A pair* $(u,m) \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d) \times L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ *is a weak solution of the system* (2.1.1)*, if*

$$u \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d), D_v u \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)),$$

$$m \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d)), D_v m \in L^2, D_x^{1/3} m \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d), m \in L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d),$$

the system (2.1.1) *holds in a distributional sense.*

Quadratic Hamiltonian and renormalized solutions

For the case of a quadratic Hamiltonian H, we assume the following:

[H2] For the Hamiltonian $H : \mathbb{R}^d \to \mathbb{R}$ we assume that it is convex, continuous and there exist constants $c > 0, C > 0$ such that, for all $p \in \mathbb{R}^d$,

$$0 \leq H(p) \leq C|p|^2, \qquad \text{(H2.1)}$$

$$H_p(p) \cdot p - H(p) \geq cH(p), \tag{H2.2}$$

$$|H_p(p)| \leq C|p|. \tag{H2.3}$$

[F2] For the coupling term $F : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, we assume that it satisfies 2.2 and with bounds that possibly depend on $L > 0$, one of the following hold:

$$f_L(t,x,v) := \sup_{m \in [0,L]} F(t,x,v,m) \in L^1(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d), \tag{F2.1}$$

$$f_L(t,x,v) := \sup_{m \in [0,L]} F(t,x,v,m)/m \in L^\infty(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d). \tag{F2.2}$$

[G2] For the coupling term $G : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, we assume that it satisfies 2.2 and with bounds that possibly depend on $L > 0$, one of the following hold:

$$g_L(x,v) := \sup_{m \in [0,L]} G(x,v,m) \in L^1(\mathbb{R}^d \times \mathbb{R}^d), \tag{G2.1}$$

$$g_L(x,v) := \sup_{m \in [0,L]} G(x,v,m)/m \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d). \tag{G2.2}$$

Remark 2. *The above conditions on F, G yield that if* (F2.1) *and* (G2.1) *hold, then*

$$F(x,v,t,m) \leq f_L(t,x,v) + \frac{m}{L} F(t,x,v,m), \;\; G(x,v,m) \leq g_L(x,v) + \frac{m}{L} G(x,v,m),$$

for every $m \geq 0, L > 0$. While if (F2.2) *and* (G2.2) *hold, then,*

$$F(t,x,v,m) \leq f_L(t,x,v)m + \frac{m}{L} F(x,v,m), G(x,v,m) \leq g_L(x,v)m + \frac{m}{L} G(x,v,m).$$

Conditions (F2.1)*,* (G2.1) *do not allow for F, G to depend only on m due to the unbounded domain, while conditions* (F2.2)*,* (G2.2) *do allow for dependence only on m. Typical examples for the* *coupling are of the form*

$$F(t,x,v,m) = a(t,x,v)h_1(m) + h_2(m)$$

where for assumption (2.2) *we need* $h_2(0) = 0$*,* $h_1 \geq 0$*, strictly increasing and locally Lipschitz continuous and finally* $a \geq 0$*,* $a \in L^2 \cap L^\infty$ *and continuous. For assumption* (2.2) *we need to also assume that*

- *In the case of* (F2.1)*,* $a \in L^1$ *and* $h_2(m) = 0$*.*
- *While for the case of* (F2.2)*, we may also impose* $a \in L^\infty$ *and that* $h_1(m) = m^{q_1}, h_2(m) = m^{q_2}$ *for some* $q_1, q_2 \in [1,\infty)$*.*

Next, we define renormalized solutions for equations of the form

$$\begin{cases} \partial_t m - \Delta_v m - v \cdot D_x m - \mathrm{div}_v(mb) = 0 \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ m(0,x,v) = m_0(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases} \tag{2.2.1}$$

where $b : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $m_0 : \mathbb{R}^d \times \mathbb{R}^d$, and equations of the form

$$\begin{cases} -\partial_t u - \Delta_v u + v \cdot D_x u + H(D_v u) = f \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ u(T,x,v) = g(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.2.2}$$

Remark 3. *Regarding our notation, in the rest of the chapter, we will follow the convention that capital letters* F, G *are be used when referring to the MFG system, while lower case letters* f, g *will be used for general HJB equations.*

Our definitions are in the same spirit as in [325].

Definition 2. *Let* $m \in C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d)_+)$ *and* $b : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$*, such that* $m|b|^2 \in L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*. We say that* m *is a renormalized solution of* (2.2.1)*, if*

$$\lim_{n\to\infty} \frac{1}{n} \int_{n<m<2n} |D_v m|^2 dx dv dt = 0,$$

and for each $S \in W^{2,\infty}(\mathbb{R}), S(0) = 0$, *the function* $S(m)$ *satisfies in the distributional sense,*

$$\partial_t S(m) - \Delta_v S(m) - v \cdot D_x S(m) - div_v(S'(m)mb) + S''(m)|D_v m|^2 + S''(m)mbD_v m = 0,$$

$$S(m)(0) = S(m_0).$$

Definition 3. *Let* $u \in C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d)_+)$, *with* $D_v u \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $f \in L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $g \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$. *We say that u is a renormalized solution of* (2.2.2), *if*

$$\lim_{n\to\infty} \frac{1}{n} \int_{n<m<2n} |D_v u|^2 dxdvdt = 0,$$

and for each $S \in W^{2,\infty}(\mathbb{R}^d), S(0) = 0$, *the function* $S(u)$ *satisfies in the distributional sense,*

$$-\partial_t S(u) - \Delta_v S(u) + v \cdot D_x S(u) + S'(u)H(D_v u) = S'(u)f, \;\; S(u(T)) = S(g).$$

Definition 4. *Assume that* H, G, F, *and* m_0 *satisfy 2.2,2.2,2.2, and 2.2. A pair* $(m,u) \in C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d)_+) \times C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d)_+)$, *is a renormalized solution of the MFG system* (2.1.1), *if* m, u *are renormalized solutions to the corresponding equations according to Definitions* (2.2.1), (2.2.2), *respectively.*

Remark 4. *In general the notions of renormalized and distributional solutions are distinct. However under suitable conditions we may show they are equivalent. We do not explore this direction in the present work, although it should follow with similar methods as in the non degenerate case, see Porretta [325] and for results on the whole space Porretta [326].*

2.3 The well possedness in the case of Lipschitz Hamiltonian

All the equations in the rest of the section should be understood in the distributional sense, unless stated otherwise. We divide this section in four parts. In the first two we study the HJB equation and the FP equation separately, in the third section we use these bounds to obtain weak solutions to the MFG problem, and in the last part we show a regularity result for these weak solutions.

2.3.1 Estimates for the Hamilton-Jacobi-Bellman equation

Theorem 3. *Let* $g \in L^2(\mathbb{R}^d \times \mathbb{R}^d) \cap L^\infty(\mathbb{R}^d \times \mathbb{R}^d)_+$, $f \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d)) \cap L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)_+$, *and a Hamiltonian* $H : \mathbb{R}^d \to \mathbb{R}$, *which satisfies 2.2. Then, there exists a unique solution* $u \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d))$, *with* $D_v u \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ *of* (2.2.2). *Furthermore, there exists a* $C = C(T, Lip_H) > 0$, *such that*

$$\sup_{t \in [0,T]} \|u(t)\|_2 + \|D_v u\|_2 \le C(\|g\|_2 + \|f\|_2)$$

and for each $t \in [0,T]$,

$$\|\partial_t u - v \cdot D_x u\|_{L^2([0,t] \times \mathbb{R}^d \times \mathbb{R}^d)} + \|\Delta_v u\|_{L^2([0,t] \times \mathbb{R}^d \times \mathbb{R}^d)} \le \frac{C}{T-t}\Big(\|f\|_2 + \|g\|_2\Big).$$

Finally, there exists a constant $C = C(T, d, \|f\|_\infty, \|g\|_\infty) > 0$, *such that* $\|u\|_\infty \le C$, *in particular* C *does not depend on the Lipschitz constant of the Hamiltonian* H.

Proof. First we address the issue of existence. Consider the Banach space $X := \{v \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)\} : \|v\|_X < \infty\}$, where

$$\|v\|_X = \sup_{0 \le t \le T} e^{-\lambda t}\|v(t)\|_2 + \Big(\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{-\lambda s}|D_v v|^2 dx dv dt\Big)^{\frac{1}{2}},$$

for some $\lambda > 0$ to be determined later. We define the map $T : X \to X$ by $T(w) = u$, where u is the solution to

$$\begin{cases} \partial_t u - \Delta_v u + v \cdot D_x u = f - H(D_v w) \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ u(0,x,v) = g(x,v), \end{cases} \tag{2.3.1}$$

where in the above we took the equation forward in time only for notational simplicity. The goal now is to show that T is a contraction on X if λ is large enough. But indeed if $C_H > 0$ is the Lipschitz constant of H, then for $T(w^1) = u^1, T(w^2) = u^2$ by testing against $u^2 - u^1$ in the equation of their difference (see the end of this proof on how we justify this), we have

$$\partial_t \int_{\mathbb{R}^d\times\mathbb{R}^d} (u^2 - u^1)^2(t,x,v)dxdv + 2\int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u^2 - u^1)|^2 dxdv$$

$$\leq C_H \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(w^2-w^1)||u^2-u^1|dxdv \leq C_H\epsilon \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(w^2-w^1)|^2 dxdv + \frac{C_H}{4\epsilon}\int_{\mathbb{R}^d\times\mathbb{R}^d} |u^2-u^1|^2 dxdv.$$

The above imply

$$\partial_t\Big(e^{-\frac{C_H}{4\epsilon}t}\int_{\mathbb{R}^d\times\mathbb{R}^d} |u^2-u^1|^2 dxdv\Big) + 2e^{-\frac{C_H}{4\epsilon}t}\int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u^2-u^1)|^2 dxdv \leq C_H\epsilon e^{-\frac{C_H}{4\epsilon}t}\int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(w^2-w^1)|^2 dxdv,$$

and thus by Grönwall, if we let $\lambda = \frac{C_H}{4\epsilon}$ we have

$$\|u^2 - u^1\|_X^2 \leq 4C_H\epsilon\|w^2 - w^1\|_X^2.$$

Therefore for $\epsilon > 0$ small enough the above map is a contraction in X and thus has a unique fixed point. Regarding the estimates, we need to test against u in 2.2.2. First need to establish integrability for u. To this end, note that since $H, f, g \geq 0$, if w is the solution of

$$\begin{cases} -\partial_t w - \Delta_v w + v\cdot D_x w = f \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ w(0,x,v) = g(x,v) \text{ in } \mathbb{R}^d\times\mathbb{R}^d, \end{cases}$$

then by standard comparison we have that

$$0 \leq u(t,x,v) \leq w(t,x,v) \text{ for all } (t,x,v) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^d.$$

Finally, note that from our assumptions on f, g

$$w \in L^p, \text{ for all } p \in [2,\infty] \text{ therefore } u \in L^p \text{ for all } p \in [2,\infty].$$

Now that we may test against u in the equation, the fact that $u \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d))$ is easy to see due to our assumptions on f. The first estimate is obtained by simply testing against u and using the fact that H is Lipschitz with $H(0) = 0$. To justify this however, we need to address the integration by parts that occurs. To this end let $\phi : [0,\infty) \to [0,1]$, be a smooth function such that $\phi(s) = 1$ for $0 \le s \le 1$ and $\phi(s) = 0$ for $s \ge 2$. For $R > 0$ we consider the function $\psi_R(x,v) = \phi(\frac{|x|^2+|v|^2}{R})$. Testing against $u\psi_R^2$ in equation 2.2.2 yields

$$-\partial_t \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|u|^2\psi_R^2 dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v u|^2\psi_R^2 dxdv$$

$$+ \int_{\mathbb{R}^d\times\mathbb{R}^d} 2uD_v uD_v\psi_R\psi_R + 2\psi_R v\cdot D_x\psi_R u^2 + H(D_v u)u\psi_R^2 dxdv = \int_{\mathbb{R}^d\times\mathbb{R}^d} fu\psi_R^2 dxdv.$$

In what follows the constant $C > 0$ may change from line to line, however it is independent of $R > 0$. We note that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \left|2\psi_R v\cdot D_x\psi_R u^2\right| = \int_{\mathbb{R}^d\times\mathbb{R}^d} \left|\frac{2x\cdot v}{R}\psi_R\phi'(\frac{|x|^2+|v|^2}{R})u^2\right| \le \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|x|^2+|v|^2}{R}\psi_R\phi'(\frac{|x|^2+|v|^2}{R})u^2$$

$$\le \int_{\mathbb{R}^d\times\mathbb{R}^d} 2\psi_R\phi'(\frac{|x|^2+|v|^2}{R})u^2 \le C\int_{|x^2|+|v|^2\ge R} u^2.$$

Moreover,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \left|2uD_v uD_v\psi_R\psi_R\right| dxdv \le \frac{1}{4}\int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v u|^2\psi_R^2 dxdv + C\int_{|x|^2+|v|^2\ge R} u^2 dxdv$$

and

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} H(D_v u)u\psi_R^2 dxdv \le \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{4}|D_v u|^2\psi_R^2 + Cu^2\psi_R^2 dxdv,$$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} fu\psi_R^2 dxdv \le \|f\|_2\|u\|_2.$$

Collecting all the above estimates we have that

$$-\partial_t \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|u|^2\psi_R dxdv+\int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v u|^2\psi_R^2 dxdv \le C(\|f\|_2\|u\|_2+\int_{\mathbb{R}^d\times\mathbb{R}^d} u^2\psi_R^2 dxdv+\int_{|x^2|+|v|^2\ge R} u^2),$$

recalling that $0 \le u \le w \in L^2$, the result follows by Grönwall and letting $R \to \infty$.

The second estimates are due to Theorems 17 and 19 in the Appendix. Finally the $L^\infty-$bounds follow by similar arguments as in [136], Proposition A.3.

□

2.3.2 Degenerate Fokker-Planck equation

All the equations should be understood in the distributional sense, unless stated otherwise. In this subsection we study the following equation

$$\begin{cases} \partial_t m - \Delta_v m - v\cdot D_x m - \mathrm{div}_v(mb) = 0 \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ m(0,\cdot,\cdot) = m_0(\cdot,\cdot) \text{ in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases} \tag{2.3.2}$$

The purpose of this subsection is to show the following theorem:

Theorem 4. *Let $b \in L^\infty([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ and m_0 a density which satisfies 2.2. Then, there exists a unique distributional solution $m \in C([0,T];L^2(\mathbb{R}^d\times\mathbb{R}^d))$ of equation* (2.3.2)*. Furthermore, there exists a $C = C(T,\|b\|_\infty) > 0$, such that*

$$\sup_{t\in[0,T]} \|m(t)\|_2+\|D_v m\|_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)}+\|D_x^{1/3} m\|_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)}+\|D_t^{1/3} m\|_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \le C\|(1+|v|^2)m_0\|_2$$

and a $C_0 = C_0(\|b\|_\infty,T,\|m_0\|_2,\|m_0\|_\infty) > 0$, so that

$$\|m\|_\infty \le C_0.$$

Moreover, $m(t)$ is a probability density for all $t \in (0,T]$. Finally, if $(T-t)div_v(b) \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, it follows that

$$[m_t - v \cdot D_x m], (T-t)\Delta_v m \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d).$$

The main two assertions in the theorem above are, firstly, the fractional gradient estimates and, secondly, the $L^\infty-$bounds. The gradient estimates are the result of Theorem 18, in the appendix. The $L^\infty-$bounds can be obtained with a De Giorgi type argument similar to the one found for example in F. Golse, C. Imbert, C. Mouhot and A. Vasseur in [211], thus we only provide the main steps in Proposition 11 at the Appendix. For a survey on the De Giorgi type arguments we refer to Mouhot [315]. First a proposition.

Proposition 1. *Assume that $m \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $b \in L^\infty \cap L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ and m_0, which satisfies 2.2, satisfy equation* (2.3.2) *in the distributional sense. Then, $|v|^2 m, |v|^2 D_v m \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$.*

Proof. We may assume that the data are smooth and bounded and obtain the general case by approximation. We test the equation with $|v|^4 m$ (see Lemma (1), on how we may justify this) to obtain

$$\frac{d}{dt}\int_{\mathbb{R}^d\times\mathbb{R}^d}|v|^4|m|^2dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d}|v|^4|D_v m|^2dxdv$$

$$= -4\int_{\mathbb{R}^d\times\mathbb{R}^d} m|v|^2 v\cdot D_v m dxdv - 4\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2|v|^2 v\cdot b dxdv - \int_{\mathbb{R}^d\times\mathbb{R}^d} m|v|^4 D_v m\cdot b dxdv$$

$$\leq \frac{1}{4}\int_{\mathbb{R}^d\times\mathbb{R}^d}|v|^4|D_v m|^2dxdv + C\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2(1+|v|^4)dxdv$$

$$+4\|b\|_\infty\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2(1+|v|^4)dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2|v|^4dxdv + \frac{1}{4}\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_v m|^2|v|^4dxdv.$$

It is easy to see that $\sup_{t\in[0,T]}\|m(t)\|_2 \leq C\|m_0\|_2$, therefore the result follows by Grönwall since,

$$\frac{d}{dt}\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2|v|^4dxdv \leq C\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^2|v|^4dxdv + C\|m_0\|_2^2.$$

□

Proof. (Theorem 4) Proposition 1, together with Theorem 18, gives us the result. □

2.3.3 Existence of Solutions via the fixed point argument

In this section we show the main theorem.

Theorem 5. *Let G, H, F and m_0 satisfy 2.2,2.2,2.2 and 2.2. Then, there exists a unique solution to system* (2.1.1)*, according to definition* (1).

Proof. As mentioned in the introduction, we apply Schauder in the following setting. Let $C_0 > 0$ be the constant from Theorem 4 and consider the closed convex subset $X := C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d)) \cap \{m : \|m\|_\infty \leq L\}$ of $C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d)_+)$, where $L > 0$, such that $\frac{1}{C_0} \max\{\|m_0\|_\infty, \|m_0\|_2\} \leq L$. For $\mu \in X$, let u_μ be the solution of

$$\begin{cases} -\partial_t u_\mu - \Delta_v u_\mu - v \cdot D_x u_\mu + H(D_v u_\mu) = F(t, x, v, \mu(t, x, v)) \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ u_\mu(T, x, v) = G(\mu(T, x, v)) \text{ in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases}$$

provided by Theorem 3. For this u_μ, we then solve

$$\partial_t m - \Delta_v m - v \cdot D_x m - \operatorname{div}_v(m H_p(D_v u)) = 0, \;\; m(0) = m_0.$$

We set $\Phi(\mu) = m$ which due to the choice of L and the bounds on m implies that $m \in X$. It remains to show that the map is continuous and compact in order to apply Schauder's Fixed Point Theorem. Continuity is straightforward to check with our given assumptions and will be omitted. For compactness, we proceed as follows. Due to the domain being unbounded we first show that $\lim_{N\to\infty} \sup_{\mu \in X} \|\Phi(\mu) \mathbf{1}_{B(0,N)^c}\|_2 = 0$, where $B(0,N) := \{(t,x,v) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d : |(x,v)| \leq N\}$. This follows directly by the same argument as in Lemma 1. Furthermore, from Theorem 4, we have

$$\|m\|_2 + \|D^s_{t,x,v} m\|_2 \leq C\|m_0\|_2 \text{ for some } s > 0.$$

Thus, by Kolmogorov–M. Riesz–Fréchet (see for example Brezis [61], Theorem 4.26 and Corollary 4.27) we have compactness of the map. Uniqueness follows from the by-now classical Lasry-Lions monotonicity argument, which we omit. □

We conclude this section with some crucial estimates, which follow directly from the by-now classical Lasry-Lions argument under assumptions 2.2 and 2.2, so we omit the proof. The computations can be found for example in [325].

Proposition 2. *Assume that $H : \mathbb{R}^d \to \mathbb{R}$, $F : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$, $m_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ satisfy 2.2,2.2,2.2 and 2.2. Let (u,m) be the weak solution of the MFG system provided by Theorem 1. Then, there exists a constant $C = C(\|m_0\|_1, \|m_0\|_\infty, T)$, such that*

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} G(x,v,m(T))dxdv + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} F(x,v,t,m)m dxdvds \\ + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} m\Big[H_p(D_v u)\cdot D_v u - H(D_v u)\Big]dxdv \leq C.\end{aligned} \tag{2.3.3}$$

Furthermore, we have the following L^1 estimates

$$\sup_{t\in[0,T]} \|u(t)\|_1 + \|F(\cdot,m)\|_1 + \|F(\cdot,m)m\|_1 + \|G(\cdot,m(T))\|_1 + \|G(\cdot,m(T))m(T)\|_1$$

$$+\|H(D_v u)\|_1 + \|m|H_p(D_v u)|^2\|_1 \leq C.$$

2.3.4 *Further Regularity of Solutions to the Mean Field Games System, for Lipschitz Hamiltonian*

In this section we study the gain of regularity for solutions to the MFG system (2.1.1). In particular, we derive appropriate energy estimates by taking advantage of the coupling.

Theorem 6. *Let F, G satisfy 2.2,2.2 with constant c_0, H, m_0 satisfy 2.2,2.2 and (u, m) be a weak solution to system* (1)*, according to Definition* (1)*. Then, there exists a constant $C = C(c_0, F, G, H, m_0) > 0$, such that*

$$\sup_{t\in[0,T]} \|m(t)\|_2 + \sup_{t\in[0,T]} \|Dm(t)\|_2 + \|D^2_{v,v}m\|_2 + \|D_vD_xm\|_2 \leq C$$

and

$$\sup_{t\in[0,T]} \|u(t)\|_2 + \sup_{t\in[0,T]} \|Du(t)\|_2 + \|D^2_{v,v}u\|_2 + \|D_vD_xu\|_2 \leq C.$$

Proof. For $i \in \{1, \cdots, d\}$ and $h \in \mathbb{R} \setminus \{0\}$, we denote

$$\delta^h(u)(t,x,v) := \frac{u(t,x+he_i,v) - u(t,x,v)}{h}, \delta^h(m)(t,x,v) := \frac{m(t,x+he_i,v) - m(t,x,v)}{h}$$

$$m^h := m(t,x+he_i,v), m^0 := m(t,x+he_i,v), D_vu^h := D_vu(t,x+he_i,v), D_vu^0 := D_vu(t,x,v)$$

$$H^h := H(D_vu(t,x+he_i,v)), H^0 := H(D_vu(t,x,v)),$$

$$F^h := F(t,x,v,m(t,x+he_i,v)), F^0 := F(t,x,v,m(t,x,v)),$$

$$\delta_{x,h}F := \frac{F(t,x+he_i,v,m(t,x+he_i,v)) - F(t,x,v,m(t,x+he_i,v))}{h},$$

$$\delta_{x,h}G := \frac{G(x+he_i,v,m(T,x+he_i,v)) - G(x,v,m(T,x+he_i,v))}{h}.$$

The equations for δ^hu, δ^hm read as follows,

$$\begin{cases} -\partial_t\delta^hu - \Delta_v\delta^hu + v\cdot D_x\delta^hu + \frac{H^h-H^0}{h} = \frac{F^h-F^0}{h} + \delta_{x,h}F, \\ \delta^hu(T) = \frac{G^h-G^0}{h} + \delta_{x,h}G. \end{cases} \tag{2.3.4}$$

$$\begin{cases} \partial_t \delta^h m - \Delta_v \delta^h m - v \cdot D_x \delta^h m - \mathrm{div}_v\left(\frac{m^h H_p^h - m^0 H_p^0}{h}\right) = 0, \\ \delta^h m(0) = \delta^h m_0 \end{cases} \tag{2.3.5}$$

Testing against $\delta^h u$ in (2.3.5), yields

$$\Big[\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{G^h - G^0}{h}\delta^h m(T)dxdv\Big]_1 + \Big[\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} \delta^h m[\frac{F^h - F^0}{h}]dxdvdt\Big]_2$$

$$+\Big[\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} -\delta^h m\frac{H^h - H^0}{h}dxdvdt + \int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} D_v\delta^h u\frac{m^h H_p^h - m^0 H_p^0}{h}\Big]_3$$

$$= -\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{x,h}F\delta^h mdxdvdt + \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta^h m_0\delta^h u(0)dxdv - \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{x,h}G\delta m^h(T)dxdv$$

In the following, we refer to the terms based on the enumeration of the brackets. For the first bracketed term using the monotonicity of G we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{G^h - G^0}{h}\delta^h m(T)dxdv =$$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\int_0^1 G'(m^0(T) + s(m^h - m^0)(T))ds|\delta^h m|^2(T)dxdv \geq c_0\int_{\mathbb{R}^d\times\mathbb{R}^d} |\delta^h m|^2(T)dxdv,$$

while for the second term again using the monotonicity of F

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} \delta^h m\frac{F^h - F^0}{h}dxdvdt = \int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} |\delta^h m|^2(t)\int_0^1 F'(m^0(t) + s(m^h - m^0)(t))dsdxdvdt$$

$$\geq c_0\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} |\delta^h m|^2(t)dxdvdt.$$

We may rewrite the third term as in the proof of uniqueness to see that it is non-negative by the convexity of H, indeed it can be written as

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{m^h}{h^2}\big[H(D_v u) - H(D_v u^h) - H_p(D_v u^h)D_v(u - u^h)\big]$$

$$+\frac{m}{h^2}\Big[H(D_v u^h) - H(D_v u) - H_p(D_v u)D_v(u^h - u)\Big]dxdvdt \geq 0.$$

Continuing, for the right hand side we estimate as follows

$$\delta_{x,h}F = \int_0^1 \partial_{x_i}F(t, x + she_i, v, m(t, x + he_i, v))hds,$$

hence,

$$\|\delta_{x,h}F\|_2 \leq C\|m\|_2,$$

and similarly for $\delta_{x,h}G$. Thus,

$$-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{x,h}F\delta^h m dxdvdt + \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta^h m_0 \delta^h u(0)dxdv - \int_{\mathbb{R}^d\times\mathbb{R}^d} \delta_{x,h}G\delta m^h(T)dxdv$$

$$\leq \frac{c_0}{2}\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |\delta m^h|^2 dxdvdt + \frac{c_0}{2}\|\delta m^h(T)\|_2 + C\sup_{t\in[0,T]}\|m(t)\|_2^2 + \|\delta^h m_0\|_2\|\delta^h u(0)\|_2.$$

Gathering everything together we obtain

$$\|\delta^h m(T)\|_2^2 + \|\delta^h m\|_2^2 \leq C\|\delta^h m_0\|_2\|\delta^h u(0)\|_2. \tag{2.3.6}$$

We now turn to (2.3.4). Test, against $\delta^h u$ to obtain

$$\sup_{t\in[0,T]}\|\delta^h u(t)\|_2 + \|D_v\delta^h u\|_2 \leq C(\|\delta^h m(T)\|_2 + \|\delta^h m\|_2)$$

and using this estimate in (2.3.6) provides

$$\|\delta^h m(T)\|_2 + \|\delta^h m\|_2 \leq C = C(\inf F', \inf G', T, \mathrm{Lip}_H, \mathrm{Lip}_F, \mathrm{Lip}_G, \|D_x m_0\|_2).$$

Testing against $\delta^h m$ in (2.3.5) yields

$$\sup_{t\in[0,T]} \|\delta^h m(t)\|_2 + \|D_v \delta^h m\|_2 \le C(\|\delta^h m_0\|_2 + \|D_v \delta^h u\|_2) \le C.$$

Since the bounds are independent of h, we have shown that

$$\sup_{t\in[0,T]} \|D_x m(t)\|_2 + \sup_{t\in[0,T]} \|D_x u(t)\|_2 + \|D_v D_x u\|_2 + \|D_v D_x m\|_2 \le C.$$

Now, using these bounds, we repeat the process for the derivatives with respect to v. We use completely symmetric notation as in the above case, for example $\delta_v^h u := \frac{u(t,x,v+he_i)-u(t,x,v)}{h}$. The equations satisfied by $\delta_v^h u, \delta_v^h m$ are similar with the exception of the $v \cdot D_x$ term. They read

$$\begin{cases} -\partial_t \delta_v^h u - \Delta_v \delta_v^h u + e_{v,i} D_x u^h + v \cdot D_x \delta_v^h u + \frac{H^h - H^0}{h} = \frac{F^h - F^0}{h} + \delta_{v,h} F, \\ u^h(T) = \frac{G^h - G^0}{h} + \delta_{v,h} G \end{cases}$$

and

$$\begin{cases} \partial_t \delta_v^h m - \Delta_v \delta_v^h m - e_{v,i} D_x m^h - v \cdot D_x \delta_v^h m - \operatorname{div}_v(m^h \frac{H_p^h - H_p^0}{h} + \delta^h m H_p^0) = 0, \\ \delta_v^h m^0 = \delta_v^h m_0. \end{cases}$$

The argument is completely symmetrical with the only difference being the presence of $D_x u^h, D_x m^h$. However, these terms are bounded from the previous case. We thus obtain bounds of the form

$$\sup_{t\in[0,T]} \|D_v m(t)\|_2 + \sup_{t\in[0,T]} \|D_v u(t)\|_2 + \|D_{v,v}^2 u\|_2 + \|D_{v,v}^2 m\|_2 \le C.$$

□

2.4 Quadratic Hamiltonian

In this section we will show existence and uniqueness for renormalized solutions to the MFG system. All the ideas and proofs in this section are entirely motivated or even parallel to the original work in [325].

To motivate some of the technical steps we outline the strategy. The plan is to approximate a given Hamiltonian H with quadratic growth by a sequence of Lipschitz Hamiltonians H^{ϵ}(see bellow for definition), for which we have shown the existence of solutions $(u^{\epsilon}, m^{\epsilon})$ in the previous section and show that these solutions converge to a renormalized solution. A crucial structural estimate, as pointed out in [325], is that $\sup_{\epsilon} \|m^{\epsilon}|H_p^{\epsilon}(D_v u^{\epsilon})|^2\|_1 < \infty$, which is shown in Proposition 2. This estimate, along with L^2-bounds on $D_v u^{\epsilon}$, allows us to conclude the convergence (up to a subsequence) to a renormalized solution of $\{m^{\epsilon}\}_{\epsilon}$. The bounds for the HJ equation are straightforward and mostly follow the classical techniques of the non-degenerate case, with the exception of the L^1-compactness for the u^{ϵ} which is due to Theorem 10 in [151] and the technical Lemma in the Appendix.

In the rest of the chapter we consider a fixed Hamiltonian H that satisfies 2.2. Furthermore, following [325], we consider the following Lipschitz approximations

$$H^{\epsilon}(p) := \frac{H(p)}{1 + \epsilon H^{\frac{1}{2}}(p)} \text{ for } \epsilon > 0. \tag{2.4.1}$$

The following are shown in [325]

Proposition 3. *The functions H^{ϵ} are Lipschitz in p and satisfy*

$$H_p^{\epsilon} \cdot p - H^{\epsilon}(p) \geq cH^{\epsilon}(p), |H_p^{\epsilon}|^2 \leq CH^{\epsilon},$$

for some constants $c > 0$,$C > 0$ independent of ϵ.

2.4.1 Analysis of Degenerate Fokker-Planck equation

In this subsection, we study the following Fokker-Planck equation

$$\begin{cases} \partial_t m - \Delta_v m - v \cdot D_x m - \operatorname{div}_v(mb) = 0 \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ m(0,x,v) = m_0(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases} \tag{2.4.2}$$

Our approach is a parallel of the techniques from [55] in the Hypoelliptic case.

Definition 5. *We say that m is a weak solution of* (2.4.2)*, if* $m \in L^1 \cap L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*, with* $D_v m \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*,* m_0 *satisfies 2.2,* $m|b|^2 \in L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*, and* (2.4.2) *is satisfied in the distributional sense.*

Lemma 1. *Let* (m, b, m_0) *be a weak solution of* (2.4.2) *according to definition 5. Then, there exists a constant* $C = C(d, T, \|m|b|^2\|_1, \|(1 + |x|^2 + |v|^2)m_0\|_1)$*, such that for all* $t \in [0,T]$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} (|x|^2 + |v|^2 + 1) m(t,x,v) dx dv \leq C.$$

Proof. Formally the result follows immediately by testing against $(|x|^2+|v|^2)$ and applying standard methods. However, this needs to be justified given that $(|x|^2+|v|^2)$ is unbounded. This requires some technical steps which we present in detail, hence the lengthy computations. First assume that b, m_0 are smooth and compactly supported. For $R > 0$ consider a bump function $\psi_R : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$, such that $\psi_R\Big|_{B(0,R)} \equiv 1$ and $\operatorname{spt}(\psi_R) \subset B(0, R+1)$. Fix a $t_0 \in [0,T]$ and let $\phi_R : [0,t_0] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be the smooth solution of the adjoint equation (see for example E. Priola [330], Theorem 5.3)

$$\begin{cases} -\partial_t \phi_R - \Delta_v \phi_R + v \cdot D_x \phi_R + b \cdot D_v \phi_R = 0 \text{ on } [0,t_0) \times \mathbb{R}^d \times \mathbb{R}^d, \\ \phi_R(t_0,x,v) = (|x|^2 + |v|^2)\psi_R(x,v) \text{ on } \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.4.3}$$

A priori, ϕ_R is bounded by a constant depending only on R, b, T. We claim that there exists a constant $C > 0$ independent of $R > 0$, such that

$$\phi_R(t, x, v) \leq C(1 + |x|^2 + |v|^2) \text{ for all } (t, x, v) \in [0, t_0] \times \mathbb{R}^d \times \mathbb{R}^d.$$

Indeed, for $A, B > 0$ large enough to be determined later, let $w(t, x, v) = Ce^{-At}(1 + |x|^2 + |v|^2) - B(t - t_0)$, which satisfies

$$-\partial_t w - \Delta_v w + v \cdot D_x w + b \cdot D_v w = Ce^{-At}(A(1 + |x|^2 + |v|^2) - 2d + 2v \cdot x + b \cdot v) + B$$

$$\geq (B - 2dCe^{-At} - ||b||_\infty^2) + Ce^{-At}(A - \frac{3}{2})(|x|^2 + |v|^2) \geq 1,$$

if $A, B > 0$ are large enough. In particular let $A = 2$ and for any choice of $C > 0$ we set $B = 1 + 2dCe^{-2t} - ||b||_\infty^2$, so that the above inequality is satisfied. Furthermore, at $t = t_0$ we have that

$$w(t_0, x, v) = Ce^{-2t_0}(|x|^2 + |v|^2) \geq (|x|^2 + |v|^2)\psi_R(x, v) = \phi_R(t_0, x, v) \text{ for all } (x, v) \in \mathbb{R}^d \times \mathbb{R}^d,$$

if say $C > e^{2t_0}$, in particular however C can be chosen independent of $R > 0$. Finally, for each $R > 0$ fixed, the function

$$E(t, x, v) = w - \phi_R$$

is coercive in (x, v), that is for each fixed $t \in [0, t_0]$,

$$\lim_{|(x,v)| \to \infty} E(t, x, v) = \infty.$$

Thus by classical arguments we find that the minimum of E is achieved at $t = t_0$, which shows the claim. To conclude the proof of the Lemma, we test against ϕ_R in equation (2.4.2), which yields

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} m(t_0)(|x|^2 + |v|^2)\psi_R(x, v)dxdv = \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi_R(0, x, v)m_0(x, v)dxdv$$

$$\leq C \int_{\mathbb{R}^d\times\mathbb{R}^d} m_0(|x|^2 + |v|^2 + 1)dxdv = C\|m_0(1 + |x|^2 + |v|^2)\|_1.$$

With the above bounds we may now test equation (2.4.2) against $(1 + |x|^2 + |y|^2)$, which yields

$$\partial_t \int_{\mathbb{R}^d\times\mathbb{R}^d} (1 + |x|^2 + |v|^2)m(t)dxdv = \int_{\mathbb{R}^d\times\mathbb{R}^d} 2dm(t) - 2x\cdot vm(t) + 2m(t)v\cdot bdxdv$$

$$\leq 2dm(t) + \int (1+|x|^2+|v|^2)m(t)+m|v|^2+m|b|^2dxdv \leq (2d+2)\int_{\mathbb{R}^d\times\mathbb{R}^d}(1+|x|^2+|v|^2)m(t)dxdv+\|m|b|^2\|_1,$$

and so by Grönwall we obtain that for some constant $C = C(d, T, \|m|b|^2\|_1, \|(1+|x|^2+|v|^2)m_0\|_1) > 0$

$$\|(|x|^2 + |v|^2 + 1)m_0\|_1 \leq C.$$

The general case follows by approximation with smooth data. □

In the following Proposition we will need the following estimate, which may be found in in Folland [195].

Proposition 4. *[[195], Theorem 5.14] Let* Γ *denote the fundamental solution of the operator* $\partial_t - \Delta_v - v\cdot D_x$ *in the space* $\mathbb{R}^d\times\mathbb{R}^d$*. Assume that* $p, q \in (1,\infty)$ *are such that*

$$\frac{1}{p} = \frac{1}{q} - \frac{1}{Q+2},$$

where $Q = d + 2$*. For a function* $f \in L^q$ *we define*

$$g(t, x, v) := \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} D_v\Gamma(t-s, x, v, y, w)f(s, w, y)dydwds.$$

Then, there exists a constant $C = C(p, q, d)$ *such that*

$$\|g\|_p \leq C\|f\|_q.$$

Proposition 5. *Let* (m, b, m_0) *be a weak solution of* (2.4.2)*, according to definition 5. Then, there exists a dimensional constant* $C = C(d) > 0$ *and a constant* $C_0 = C_0(m_0) > 0$*, such that*

$$\|m|b|^2\|_{\frac{d+4}{d+3}} + \|m\|_{\frac{d+4}{d+2}} \leq C\|m|b|^2\|_1 + C_0.$$

Proof. Let Γ denote the fundamental solution of the operator $\partial_t - \Delta_v - v \cdot D_x$. From the equation satisfied by m we obtain

$$m(x, v, t) = -\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} D_v\Gamma(t - s, x, v, y, w)mb(s, w, y)dydwds + C(m_0)(t, x, v)$$

where

$$C(m_0)(t, x, v) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \Gamma(t, x, v, y, w)m_0(y, w)dydw.$$

From Proposition (4) above, we have that

$$\|m\|_p \leq C\|mb\|_q$$

where

$$\frac{1}{p} = \frac{1}{q} - \frac{1}{Q+2},$$

and $Q = d + 2$. Moreover, by Hölder

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |m|^q|b|^q dxdvdt$$

$$\leq \Big(\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |m|^{\frac{q}{2-q}} dxdvdt\Big)^{\frac{2-q}{2}} \Big(\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} m|b|^2 dxdv\Big)^{\frac{q}{2}} = C\|m\|^{\frac{q}{2}}_{\frac{q}{2-q}}.$$

Hence, we can have a gain of integrability if we require that

$$p = \frac{q}{2-q} \iff \frac{2-q}{q} = \frac{1}{q} - \frac{1}{Q+2} \iff \frac{1}{q} - 1 = -\frac{1}{Q+2} \iff \frac{1}{q} = \frac{Q+1}{Q+2},$$

therefore

$$q = \frac{Q+2}{Q+1} \text{ and } p = \frac{Q+2}{2Q+2-Q-2} = \frac{Q+2}{Q}.$$

□

Proposition 6. *Let* (m, b, m_0) *be a weak solution of* (2.4.2) *according to Definition 5, with* $b \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$*. Then, there exists a constant* $C(\|m_0 \log(m_0)\|_1, \|m|b|^2\|_1) > 0$ *such that*

$$\sup_{t \in [0,T]} \|m(t)\log(m(t))\|_1 + \|D_v(\sqrt{m})\|_2 \leq C.$$

Proof. For $\delta > 0$, define $w(x) = \log(x+\delta)$ and $W(x) = (x+\delta)\log(x+\delta) - \delta\log(\delta)$. Test against $w(m)$ in (2.4.2) ($m \in L^\infty \cap L^1$ and so $w(m) \in L^\infty, W(m) \in L^1$) to obtain that for each $t \in [0,T]$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} W(m(t))dxdv + \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|D_v m|^2}{(m+\delta)} dxdvds = -\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{m}{m+\delta} D_v m \cdot b dxdvdt$$

$$+ \int_{\mathbb{R}^d \times \mathbb{R}^d} W(m_0)dxdv$$

$$\leq \frac{1}{2}\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|D_v m|^2}{(m+\delta)} dxdvdt + \frac{1}{2}\|m|b|^2\|_1 + \int_{\mathbb{R}^d \times \mathbb{R}^d} W(m_0)dxdv.$$

Letting $\delta \to 0$ yields

$$\int m(t)\log(m(t))dxdv + \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|D_v m|^2}{m} dxvds \leq C(\|m|b|^2\|_1 + \|m_0 \log(m_0)\|_1)$$

where $C > 0$ is a universal constant. It remains to show that $m(t)\log(m(t) \in L^1$. This is shown for example in [151], under the conditions

1. $\|m(t)(1 + |x|^2 + |v|^2)\|_1 < \infty$

2. $\int_{\mathbb{R}^d \times \mathbb{R}^d} m(t)\log(m(t)) < \infty$.

Condition 1 follows from Lemma 1, while condition 2 is shown above. □

We now proceed with gradient estimates for the measure.

Theorem 7. *Let* (m, b, m_0) *be a weak solution of* (2.4.2) *according to Definition 5. Then, there exist* $s \in (0,1)$ *and* $q \in (1,\infty)$*, such that*

$$\|D^s m\|_q \leq C,$$

where C *depends only on* $m_0, d, T, \|m|b|^2\|_1$ *and in particular not on* $\|D_v m\|_2$.

Proof. The constant $C > 0$ that appears in this proof is subject to change from line to line and depends only on m_0, d, T. The technique that follows is the same as in [55]. In the original equation (2.4.2) we test against $\phi(m)$ for $\phi(s) = s$ for $s \in [0,1]$ and $\phi(s) = 1, s \geq 1, \phi(s) = 0, s \leq 0$. This yields

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi(m(t))dxdv + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(m)|D_v m|^2 =$$

$$-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(m) D_v m H_p m dxdvdt + \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi(m_0)dxdv$$

$$\leq \frac{1}{2}\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(m)|D_v m|^2 dxdv + \int_{|m|\leq 1} |m|^2|b|^2 dxdv + C(m_0).$$

Since $|m|^2 \leq |m|$ on $|m| \leq 1$, we obtain

$$\int_{\{|m|\leq 1\}} |D_v m|^2 dxdv \leq C.$$

For $k \in \mathbb{N}$ we define ϕ_k by

$$\phi_k(s) := \begin{cases} 0, s \leq k-1, \\ s-(k-1), s \leq k, \\ 1, s \geq 1, \end{cases} \tag{2.4.4}$$

and $\Phi_k(t) := \int_0^T \phi_k(s)ds$. Testing against $\phi_k(m)$ in the equation yields

$$\begin{aligned}&\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_k(m(T)) + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_k'(m)|D_v m|^2 dxdvdt \\ &= -\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_k' D_v mbmdxdvdt + \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_k(m_0)dxdv.\end{aligned} \tag{2.4.5}$$

Additionally

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_k(m_0)dxdv \leq \|m_0\|_2 + \|m_0\|_1 \leq C$$

and

$$0 \leq \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_k(m(T)).$$

For $A_k := \{k-1 \leq |m| \leq k\}, k \in \mathbb{N}$, equation (2.4.5) yields

$$\int_{A_k} |D_v m|^2 dxdvdt \leq \frac{1}{2k}\int_{A_k} m|D_v m|^2 dxdv + Ck\int_{A_k} m|b|^2 dxdv + C, \text{ for all } k \in \mathbb{N}.$$

Moreover,

$$\int_{A_k} m|D_v m|^2 dxdvdt \leq k\int_{A_k} |D_v m|^2 dxdvdt$$

hence, by summing for $k = 2, \cdots$, for $\lambda > 1$, we obtain

$$\int_{|m|\geq 1} \frac{|D_v m|^2}{(1+m)^\lambda} dxdvdt \leq \sum_{k=1}^{\infty} \frac{k}{(1+k)^\lambda}\int_{A_k} m|b|^2 dxdvdt + \frac{C}{k^\lambda} < \infty.$$

Thus,

$$\int_{m>1} |D_v m|^q dxdv \leq \Big[\int_{m>1} \frac{|D_v m|^2}{(1+m)^\lambda}\Big]^{q/2}\Big[\int_{m>1} (1+m)^{\frac{\lambda q}{2-q}} dxdv\Big]^{\frac{2-q}{2}}.$$

Next, using that

$$(a+b)^\lambda \leq 2^\lambda \max\{a^\lambda, b^\lambda\} \leq C(a^\lambda + b^\lambda)$$

and

$$|\{|m| > 1\}| \leq \|m\|_1 = 1,$$

we obtain

$$\int_{|m|>1}(1+m)^{\frac{\lambda q}{2-q}}dxdv \leq C\Big(|\{m>1\}|^{\frac{\lambda q}{2-q}}+\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^{\frac{\lambda q}{2-q}}dxdv\Big) \leq C(1+\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^{\frac{\lambda q}{2-q}}dxdv).$$

Hence,

$$\int_{m>1}|D_v m|^q dxdv \leq \Big[\int_{m>1}\frac{|D_v m|^2}{(1+m)^\lambda}\Big]^{q/2}\Big(1+\int_{\mathbb{R}^d\times\mathbb{R}^d}|m|^{\frac{\lambda q}{2-q}}dxdv\Big)^{\frac{2-q}{2}}. \tag{2.4.6}$$

Integrate in time inequality (2.4.6), and apply Hölders inequality for $\frac{2}{q}, \frac{2}{2-q}$, to obtain for some $C = C(T, \lambda, q, \|\frac{D_v m}{(1+m)^{\frac{\lambda}{2}}}\mathbf{1}_{m\leq 1}\|_2) > 0$

$$\int_{m>1}\|D_v m(t)\|_q^q dxdvdt \leq \Big(\int_{m>1}\frac{|D_v m|^2}{(1+m)^\lambda}dxdvdt\Big)^{\frac{q}{2}}\Big(1+\int_0^T\|m(t)\|_{\frac{\lambda q}{2-q}dt}^{\frac{\lambda q}{2-q}}dt\Big)^{\frac{2-q}{q}}$$

$$\leq C(1+\Big(\int_0^T\|m(t)\|_{\frac{\lambda q}{2-q}}^{\frac{\lambda q}{2-q}}dt\Big)^{\frac{2-q}{2}})$$

The Fractional Gagliardo-Niremberg inequality gives us

$$\|m(t)\|_\sigma \leq C\|D^s m\|_q^\theta \|m(t)\|_\rho^{1-\theta},$$

where

$$\frac{1}{\sigma} = \theta(\frac{1}{q}-\frac{s}{n})+\frac{1-\theta}{\rho}, \tag{2.4.7}$$

and $C = C(s, q, n, \theta, \rho) > 0$, we refer for example to [62]. We can easily obtain the following time dependent version,

$$\int_0^T\|m(t)\|_\sigma^\sigma dt \leq C\sup_t\|m(t)\|_1^{\sigma(1-\theta)}\int_0^T\|D^s m\|_q^{\theta\sigma}dt \leq C\int_0^T\|D^s m\|_q^{\theta\sigma}dt.$$

Set

$$\theta = \frac{q}{\sigma}, \rho = 1, \sigma = \frac{\lambda q}{2-q},$$

which implies that

$$\frac{1}{\sigma} = \frac{q}{\sigma}(\frac{1}{q} - \frac{s}{n}) + 1 - \frac{q}{\sigma} = \frac{1}{\sigma} - \frac{qs}{\sigma n} + 1 - \frac{q}{\sigma}$$

thus,

$$\frac{qs}{\sigma n} = 1 - \frac{q}{\sigma} \implies \sigma = \frac{qs}{n} + q \implies \sigma = q(\frac{s}{n} + 1)$$

and so

$$q(\frac{s}{n} + 1) = \frac{\lambda q}{2 - q} \implies \lambda = (2 - q)(1 + \frac{s}{n})$$

which is a valid choice as long as

$$(2 - q)(1 + \frac{s}{n}) > 1 \implies q < 2 - \frac{n}{n + s}$$

thus our restrictions on q is that

$$1 < q < 2 - \frac{n}{n + s}.$$

Continuing with the above analysis for the above choices of parameters we obtain

$$\int_{m>1} \|D_v m(t)\|_q^q dx dv dt \leq C(1 + \Big(\int_0^T \|m(t)\|_\sigma^\sigma dt\Big)^{\frac{2-q}{2}}) \leq C\Big(1 + \int_0^T \|D^s m\|_q^q dt\Big)^{\frac{2-q}{2}}.$$

Therefore for some $\alpha \in (0, 1)$

$$\|D_v m\|_q \leq C(\|D_v m \mathbf{1}_{m\leq 1}\|_q + \|D_v m \mathbf{1}_{m>1}\|_q)$$

$$\leq C(1 + \|D_v m\|_1^\alpha \|D_v m \mathbf{1}_{m\leq 1}\|_2^{1-\alpha} + \|D^s m\|_q^{\frac{2-q}{2}}),$$

and by using the estimate from Proposition 6, we obtain

$$\|D_v m\|_1 = \|\sqrt{m} D_v \sqrt{m}\|_1 \leq \|D_v \sqrt{m}\|_2,$$

therefore

$$\|D_v m\|_q \le C(1 + \|D^s m\|_q^{\frac{2-q}{2}}).$$

By Theorem 19, we have that

$$\|D_x^s m\|_q \le C(1 + \|D_v m\|_q + \|m|b|^2\|_q + \|m\|_q)$$
$$\le C(1 + \|D^s m\|_q^{\frac{2-q}{2}})$$

Thus by choosing q so that $\|m|b|^2\|_q + \|m\|_q \le C$ from Proposition 5, the result follows.

□

Theorem 8. *Let* $\{(m^n, b^n, m_0)\}_{n\in\mathbb{N}}$ *be a sequence of weak solutions to* (2.4.2) *according to definition 5, such that*

$$\sup_{n\in\mathbb{N}} \left(\|m^n|b^n|^2\|_1 + \|b^n\|_2\right) < \infty.$$

Then, the set $\{m^n\}_{n\in\mathbb{N}}$ *is compact in* $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$.

Proof. From Proposition 7, we have that

$$\|m^n\|_r + \|D^s m^n\|_q \le C \text{ for all } n \in \mathbb{N} \text{ and some } r > 1, s \in (0,1).$$

The result about the compactness in $L^1([0,T] \times L^1(\mathbb{R}^d \times \mathbb{R}^d))$ now follows by the results in [338], with a slight modification due to the unbounded domain. We sketch the argument. For $R > 0$, let $\phi_R(x,v) := \psi_R(x)\psi_R(v)$, where ψ_R are standard non-negative cutoff functions with support in $B(0,R) \subset \mathbb{R}^d$. The, equation satisfied by $m^R := m\phi_R$, reads

$$\partial_t m^R - \Delta_v m^R - v \cdot D_x m^R - \mathrm{div}_v(m^R b) = D_v \phi^R m b - m\Delta_v(\phi^R) - 2D_v\phi^R D_v m - mv \cdot D_x \phi^R.$$

Next for $\frac{1}{p}+\frac{1}{q} = 1$, we set $X := W^{s,q}(B_{\mathbb{R}^d\times\mathbb{R}^d}(0,R))$, $B := L^q(B_{\mathbb{R}^d\times\mathbb{R}^d}(0,R))$ and $Y := W^{-1,p}(B_{\mathbb{R}^d\times\mathbb{R}^d}(0,R))$. Space X embeds compactly in B and B embeds continuously in Y. Since m_n^R are bounded in

$L^q(0,T,X)$ and $\partial_t m_n^R$ is bounded in $L^q(0,T,Y) \subset L^1((0,T),Y)$. Therefore from Corollary 4 in [338], for each fixed $R > 0$ the sequence m_n^R is compact in $L^q(0,T,B) = L^q(0,T,B_{\mathbb{R}^d\times\mathbb{R}^d}(0,R)) \subset L^1(0,T,B_{\mathbb{R}^d\times\mathbb{R}^d}(0,R))$. Combining the above with the estimate $\sup_{n,t}\int_{B(0,R)^c} m^n(t,x,v)dxdv \to 0$ as $R \to \infty$, from Lemma 1, yields the strong convergence in $L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$. □

Proposition 7. *Let $\{(m^n, b^n, m_0)\}_{n\in\mathbb{N}}$ be a sequence of weak solutions to* (2.4.2) *according to definition 5, such that*

$$\sup_{n\in\mathbb{N}}\left(\|m^n|b^n|^2\|_1 + \|b^n\|_2\right) < \infty$$

and

$$b^n \to b \text{ almost everywhere, for some } b \in L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d).$$

Then, there exists a $m \in L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, such that up a subsequence $m^n \to m, m^n b^n \to mb$ in $L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$. Furthermore, the set $\{m^n\}_{n\in\mathbb{N}}$ is compact in $C([0,T];\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d))$. Finally, m is a distributional solution of (2.4.2).

Proof. From Theorem 8, there exists an $m \in L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ and a subsequence(still denoted by $\{m_n\}_{n\in\mathbb{N}}$) such that $\|m_n - m\|_1 \to 0$. Furthermore, from Lemma 1 we have that

$$\limsup_{R\to\infty}\sup_{n\in\mathbb{N}}\int_0^T\int_{B_R^c}|m^n||b^n|dxdv$$

$$\leq \limsup_{R\to\infty}\sup_{n\in\mathbb{N}}\Big(\int_0^T\int_{B_R^c}|m^n|dxdvdt\Big)^{\frac{1}{2}}\Big(\int_0^T\int|m^n||b^n|^2dxdvdt\Big)^{\frac{1}{2}} = 0.$$

The above combined with Proposition 5 yields that the sequence $\{m^n b^n\}_{n\in\mathbb{N}}$ is uniformly integrable, which together with the almost everywhere convergence gives us that the limit m is in fact a distributional solution of (2.4.2).

Next, we show the claim about the compactness in $C([0,T];\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d))$. From Lemma 1 the set $\{m^n(t)\}_{n\in\mathbb{N}}$ is compact in $\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d)$ for each $t \in [0,T]$. The result about compactness in $C([0,T];\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d))$, will follow once we obtain Hölder time continuity. However this follows by typical arguments such as the one found in the notes of Cardaliaguet [73]. □

Theorem 9. *Let* $\{(m^n, b^n, m_0)\}_{n\in\mathbb{N}}$ *be a sequence of weak solutions to* (2.4.2) *according to definition 5. Assume furthermore that* $\sup_n \|b^n\|_2 < \infty$, *and that the assumptions of Proposition 7 are satisfied. Then, the limit m provided by Proposition 7 is a renormalized solution according to Definition 2.*

Proof. Let $S : \mathbb{R} \to \mathbb{R}$, such that $S \in W^{1,\infty}(\mathbb{R})$ and that S' has compact support. Then, for each $n \in \mathbb{N}$ we have

$$\partial_t S(m^n) - \Delta_v S(m^n) - v \cdot D_x S(m^n) - \mathrm{div}_v(S'(m^n)m^n b^n) + S''(m^n)D_v m^n m^n b^n + S''(m^n)|D_v m^n|^2 = 0. \tag{2.4.8}$$

Since $\{m^n|b^n|^2\}_{n\in\mathbb{N}}$ is uniformly bounded in $L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, we obtain that

$$\lim_{k\to\infty} \sup_{n\in\mathbb{N}} \frac{1}{k} \int_{k<m^n<2k} |D_v m^n|^2 dx dv ds = 0,$$

just as in Theorem 6.1 of [325]. It remains to show that for a fixed $k \in \mathbb{N}$, we have the following convergence $D_v(m^n \wedge k) \to D_v(m \wedge k)$ strongly in L^2. To show the strong convergence of the truncations, it is enough to show that

$$\|D_v \log(1+m_n) - D_v \log(1+m)\|_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \to 0.$$

The argument that follows is entirely due to DiPerna-Lions in [151]. We only present some of the main estimates since we have a slightly different setup. We look at $g^n = \log(1+m_n)$ and the corresponding equation they satisfy. From Proposition 6 we have that $\sup_{n\in\mathbb{N}} \|D_v g^n\|_2 < \infty$ and so without loss of generality we may assume that $D_v g^n$ converges weakly in L^2 to $D_v g$, where $g = \log(1+m)$. Therefore, there exists a non-negative bounded measure μ (in the sense that $\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} d\mu < \infty$) on $(0,T)\times\mathbb{R}^d\times\mathbb{R}^d$ such that

$$|D_v g^n|^2 \to |D_v g|^2 + \mu$$

in the distributional sense. It remains to show that μ is identically zero. First, for each $n \in \mathbb{N}$ we let $\beta = \log(1+t)$ and $g^n = \beta(m^n)$. The functions g^n satisfy

$$\partial_t g^n - \Delta_v g^n - v \cdot D_x g^n - \mathrm{div}_v(\frac{m^n}{1+m^n} b^n) = |D_v g^n|^2 + \frac{m^n}{1+m^n} b^n D_v g^n$$

$$g^n(0) = \log(1+m_0).$$

Again, just as in [151], we set $\Phi^n_{s,R}(t) = \exp(st \wedge R)$ and $\Psi^n_{s,R}(t) := \int_0^T \Phi^n_{s,R}(\theta) d\theta$, for some $0 < s < 1$. Test the equation against $\Phi^n_{s,R}(g^n)\phi$, where $\phi \in C_c((0,T))$, which yields

$$-\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \Psi^n_{s,R}(g^n)\phi'(t) dxdvdt$$

$$+\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} s\phi |D_v g^n|^2 \mathbf{1}_{g^n \leq R} \Phi^n_{s,R}(g^n) + s\Phi^n_{s,R}(g^n) \mathbf{1}_{g^n \leq R} D_v g^n \frac{m^n}{1+m^n} b^n dxdvdt$$

$$= \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi^n_{s,R}(g^n)\phi |D_v g^n|^2 + \phi \Phi^n_{s,R}(g^n) \frac{m^n}{1+m^n} b^n D_v g^n,$$

or equivalently

$$-\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \Psi^n_{s,R}(g^n)\phi'(t) dxdvdt =$$

$$\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi \Phi^n_{s,R}(g^n) \Big[\Big(|D_v g^n|^2 - s|D_v g^n|^2 \mathbf{1}_{g^n \leq R} \Big) + \frac{m^n \cdot b^n}{1+m^n} \Big(D_v g^n - s D_v g^n \mathbf{1}_{g^n \leq R} \Big) \Big] dxdvdt \tag{2.4.9}$$

$$= (I) + (II).$$

Now we bound each term,

$$|(I)| \leq \|\phi\|_\infty \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} (1-s) \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |D_v g^n|^2 \Phi^n_{s,R}(g^n) dxdvdt$$

$$+ \exp(sR) \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi^n_{s,R}(g^n) |D_v g^n|^2 \mathbf{1}_{g^n > R}.$$

Using the fact that

$$|\Phi^n_{s,R}(g^n)| \leq (1+m_n)^s,$$

we obtain

$$|D_v g^n|^2 \Phi^n_{s,R}(g^n) \leq \frac{|D_v m^n|^2}{(1+m^n)^{2-s}} \leq \frac{|D_v m^n|^2}{m^n}.$$

Furthermore,

$$\Phi^n_{s,R}(g^n)|D_v g^n|^2 \mathbf{1}_{g^n>R} \leq \exp(sR)\exp(-R)\frac{|D_v m^n|^2}{m^n},$$

where in the last inequality we used that

$$\Phi_{s,R}(t) = \exp(sR) \text{ for } t > R, \text{ and } \frac{1}{1+m^n}\mathbf{1}_{g^n>R} \leq \exp(-R).$$

Thus, from Proposition 6, for some $C = C(\|m_0\|_1, \|m_0 \log(m_0)\|_1, \|\log(1+m_0)\|_1, \sup_n(\|b^n\|_2 + \|m^n|b^n|^2\|_1))$ we have the bound

$$|(I)| \leq \Big((1-s)\|\phi\|_\infty$$

$$+\exp(-(1-s)R)\Big)\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{|D_v m^n|^2}{m^n} dxdvdt \leq C\Big((1-s)\|\phi\|_\infty + \exp(-(1-s)R)\Big),$$

where in the last inequality is due to Proposition 6. For the second term we work as follows

$$|(II)| \leq (1-s)\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_{s,R}(g^n)\frac{|m^n||b^n|}{(1+m^n)}|D_v g^n| dxdvdt$$

$$+\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_{s,R}(g^n)\frac{|m^n||b^n|}{(1+m^n)}|D_v g^n|\mathbf{1}_{m^n>R} dxdvdt.$$

For the first term above we use

$$\Phi_{s,R}(g^n)\frac{|m^n||b^n|}{(1+m^n)}|D_v g^n| \leq \frac{|m^n||b^n|}{(1+m^n)^{2-s}}|D_v m^n| \leq m_n|b^n|^2 + \frac{|D_v m^n|^2}{m^n},$$

while for the second integral

$$\Phi_{s,R}(g^n)\frac{|m^n||b^n|}{(1+m^n)}|D_v g^n|\mathbf{1}_{g^n>R} \leq \exp(-(1-s)R)\Big(m^n|b^n|^2 + \frac{|D_v m^n|^2}{m^n}\Big),$$

hence

$$|(II)| \leq C\Big((1-s) + \exp(-(1-s)R)\Big).$$

Thus passing to the limit in (2.4.9), we obtain

$$\Big|\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(t)\Psi_{s,R}(g)dxdvdt\Big| \leq C(1+\|\phi\|_\infty)\Big((1-s)+e^{-(1-s)R}\Big). \tag{2.4.10}$$

Now that we have obtained these bounds we obtain the result just as in [151], section III. The only difference in the proof is the divergence term, which however causes no technical difficulties. We provide the details next. For $\epsilon > 0$ let ρ_ϵ be a standard sequence of mollifiers. The equation satisfied by $g^\epsilon := \rho_\epsilon \star g$ where g solves

$$\partial_t g - \Delta_v g - v\cdot D_x g - \mathrm{div}_v(\frac{m}{1+m}b) = |D_v g|^2 + \mu + \frac{m}{1+m}bD_v g$$

reads

$$\partial_t g^\epsilon - \Delta_v g^\epsilon - v\cdot D_x g^\epsilon - \mathrm{div}_v(\rho_\epsilon\star(\frac{m}{1+m}b)) = \rho_\epsilon\star|D_v g|^2 + \rho_\epsilon\star(\frac{m}{1+m}bD_v g) + \rho_\epsilon\star\mu + r_\epsilon. \tag{2.4.11}$$

Testing against $\phi\Phi_{s,R}(g^\epsilon)$ in (2.4.11) yields,

$$-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(t)\Psi_{s,R}(g^\epsilon)dxdvdt$$

$$\geq \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(t)\Big[-|D_v g^\epsilon|^2\Phi'_{s,R}(g^\epsilon) + \Phi'_{s,R}(g^\epsilon)D_v g^\epsilon\rho_\epsilon\star(\frac{m}{1+m}b) + \rho_\epsilon\star|D_v g|^2\Phi_{s,R}(g^\epsilon) +$$
$$\rho_\epsilon\star(\frac{m}{1+m}bD_v g)\Phi_{s,R}(g^\epsilon)\Big]\phi(t)\Phi_{s,R}(g^\epsilon)\rho_\epsilon\star\mu dxdvdt - \|r_\epsilon\|_1\|\phi\|_\infty\|\Phi_{s,R}(g^\epsilon)\|_\infty.$$

We let $\epsilon \to 0$ and using that $\Phi_{s,R} \geq 1$ obtain

$$-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'(t)\Psi_{s,R}(g)dxdvdt$$

$$\geq \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(t)\Big[|D_v g|^2\Phi_{s,R}(g) - |D_v g|^2\Phi'_{s,R}(g)\Big]$$

$$+\phi(t)\Big[\frac{m}{1+m}bD_v g\Phi_{s,R}(g) - \Phi'_{s,R}(g)D_v g\frac{m}{m+1}b\Big]dxdv + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(t)d\mu$$

$$\geq \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (1-s)\phi(t)|D_v g|^2(g)\mathbf{1}_{g\leq R} + \phi(t)|D_v g|^2\mathbf{1}_{g>R}$$

$$+\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (1-s)\phi(t)\frac{m}{m+1}bD_v g\Phi_{s,R}(g)\mathbf{1}_{g\leq R} + \phi(t)\frac{m}{m+1}bD_v g\Phi_{s,R}(g)\mathbf{1}_{g>R},$$

where in the last equality we used that $\Phi_{s,R} \geq 1$. Next we bound the terms in the RHS

$$\Big|\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (1-s)\phi(t)|D_v g|^2(g)\mathbf{1}_{g\leq R} + \phi(t)|D_v g|^2\mathbf{1}_{g>R}\Big|$$

$$\leq (1-s)C\|\phi\|_\infty\|D_v\sqrt{m}\|_2 + \|\phi\|_\infty e^{-R}\|D_v\sqrt{m}\|_2,$$

while for the rest of the terms

$$\Big|\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (1-s)\phi(t)\frac{m}{m+1}bD_v g\Phi_{s,R}(g)\mathbf{1}_{g\leq R} + \phi(t)\frac{m}{m+1}bD_v g\Phi_{s,R}(g)\mathbf{1}_{g>R}\Big|$$

$$\leq (1-s)\|\phi\|_\infty\Big(\|m|b|^2\|_1 + \|D_v\sqrt{m}\|_2\Big) + \|\phi\|_\infty e^{-R(1-s)}\Big(\|m|b|^2\|_1 + \|D_v\sqrt{m}\|_2\Big).$$

Hence combining the estimates above with estimate (2.4.10), we obtain

$$\int \phi d\mu \leq C((1-s) + e^{-R(1-s)})$$

letting $R \to \infty$ and then $s \uparrow 1$ yields $\int \phi d\mu \leq 0$, for all $\phi \geq 0$ and since $\mu \geq 0$ it follows that $\mu \equiv 0$. Finally, we show that $m \in C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d))$. Let ρ_n be a standard sequence of mollifiers (see section 1 for definition) and $m_n := \rho_n \star m$. The functions m_n satisfy

$$\partial_t m_n - \Delta_v m_n - v \cdot D_x m_n - \mathrm{div}_v(\rho_n \star (mb)) = r_n, \ m_n(0) = \rho_n \star m_0, \tag{2.4.12}$$

where $r_n = K_n \star m$ and K_n is given by

$$K_n := n^{2d} \frac{v}{n} D_x \rho(\frac{x}{n}, \frac{v}{n}),$$

and so $r_n \to 0$ strongly in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$. From Lemma A.1 in [136], we have that $m_n \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d))$. For any $S \in C_c^\infty(\mathbb{R})$, the function $S(m_n)$ satisfies

$$\begin{cases} \partial_t S(m_n) - \Delta_v S(m_n) - v \cdot D_x S(m_n) - \mathrm{div}_v(S'(m_n)\rho_n \star (mb)) = -S''(m_n)|D_v m_n|^2 \\ -S''(m_n) D_v m_n \rho_n \star (mb) + S'(m_n) r_n, \\ S(m_n)(0) = S(\rho_n \star m_0). \end{cases}$$

For $n, k \in \mathbb{N}$, we test against $S(m_n) - S(m_k)$ in the equation satisfied by their difference which yields for all $t \in [0,T]$

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |S(m_n) - S(m_k)|^2(t) dx dv + \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} |D_v(S(m_n) - S(m_k))|^2 dx dv dt$$

$$= -\boxed{\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} D_v(S(m_n) - S(m_k))\Big(S'(m_n)\rho_n \star (mb) - S'(m_k)\rho_k \star (mb)\Big) dx dv dt}_1$$

$$-\boxed{\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(S(m_n) - S(m_k)\Big)\Big(S''(m_n)|D_v m_n|^2 - S''(m_k)|D_v m_k|^2\Big) dx dv dt}_2$$

$$-\boxed{\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(S(m_n) - S(m_k)\Big)\Big(S''(m_n) D_v m_n \rho_n \star (mb) + S'(m_n) r_n - S''(m_k) D_v m_k \rho_k \star (mb)}_3$$

$$\boxed{+S'(m_k) r_k\Big) dx dv dt}_3 + \boxed{\int_{\mathbb{R}^d \times \mathbb{R}^d} |S(m_n) - S(m_k)|^2(0) dx dv}_4 .$$

As noted in [325] (Remark 3.9) we have that

$$|\rho_n \star (mb)|^2 \leq [\rho_n \star (m|b|^2)]m_n. \tag{2.4.13}$$

For the first boxed term the following hold

$$D_v(S(m_n)) \to D_v S(m) \text{ strongly in } L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d) \text{ as } n \to \infty,$$

while from (2.4.13), we obtain

$$|S'(m_n)\rho_n \star (mb)|^2 \leq (S'(m_n))^2 m_n[\rho_n \star (m|b|^2)] \leq C_S[\rho_n \star (m|b|^2)],$$

where $C_S := \|(S'(x))^2 x\|_\infty$. Since $[\rho_n \star (m|b|^2)] \to m|b|^2$ strongly in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ by Dominated Convergence Theorem we obtain

$$S'(m_n)\rho_n \star (mb) \to S'(m)mb \text{ strongly in } L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d),$$

therefore the first term can be bounded by a function $\omega(n,k)$ such that $\lim\limits_{n,k} \omega(n,k) = 0$ independently of t. For the second term we note that

$$S''(m_n)|D_v m_n|^2 \to S''(m)|D_v m|^2 \text{ strongly in } L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d),$$

while $S(m_n) \to S(m)$ strongly in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ with $\sup\limits_n \|S(m_n)\|_\infty < \infty$ therefore it can also be bounded like the first term. For the third term, from (2.4.13) we have

$$|S''(m_n)D_v m_n \rho_n \star (mb)| \leq \frac{1}{2}|S''(m_n)||D_v m_n|^2 + |S''(m_n)m_n|[\rho_n \star (m|b|^2)]$$

and since the right hand side of the above inequality converges strongly in L^1 by Dominated Convergence we obtain that

$$S''(m_n)D_v m_n \rho_n \star (mb) \to S''(m)D_v m \cdot mb \text{ strongly in } L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d),$$

while $S'(m_n)r_n$ converges strongly to 0 in $L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ just as in step 3, section III of [151]. Finally the fourth term clearly vanishes as $n,k\to\infty$. Thus taking the sup over t we obtain

$$\lim_{t,n,k}\int_{\mathbb{R}^d\times\mathbb{R}^d}|S(m_n)-S(m_k)|^2(t)dxdv = 0.$$

The above show that $S(m)\in C([0,T];L^2(\mathbb{R}^d\times\mathbb{R}^d))$ for all $S\in C_c^\infty(\mathbb{R}^d\times\mathbb{R}^d)$ and so $T_k(m)\in C([0,T];L^2(\mathbb{R}^d\times\mathbb{R}^d))$ for all $k\in\mathbb{N}$ where T_k is the truncation at k. To conclude, since for all $R>0$

$$\|m(t)-m(s)\|_{L^1(\mathbb{R}^d\times\mathbb{R}^d)} \le \|m(t)-m(s)\|_{L^1(B_R)} + \|m(t)-m(s)\|_{L^1(B_R^c)}$$

and due to the bounds of Lemma 1, we obtain that for some $C=C(R)>0$ and $C_1=C_1(m_0,b)>0$

$$\|m(t)-m(s)\|_{L^1(\mathbb{R}^d\times\mathbb{R}^d)} \le C(R)\|T_k(m(t))-T_k(m(s))\|_2 + 2\sup_{\theta\in[0,T]}\|m(\theta)-T_k(m(\theta))\|_1 + \frac{C_1}{R^2}.$$

Furthermore by Proposition 6,

$$\|m(\theta)-T_k(m(\theta))\|_1 = \int_{m(\theta)>k}|m|(\theta)dxdv \le \frac{A(\|m_0\log(m_0)\|_1)}{\log(k)},$$

where $A>0$ is the constant provided by Proposition 6. Putting everything together we obtain

$$\|m(t)-m(s)\|_1 \le C_R\|T_k(m(t))-T_k(m(s))\|_2 + \frac{A}{\log(k)}\|m_0\log(m_0)\|_1 + \frac{C_1}{R^2}.$$

Thus given an $\epsilon>0$, first we fix an $R>0$ such that $\frac{C_1}{R^2}\le\frac{\epsilon}{3}$ and a $k\in\mathbb{N}$ such that

$$\frac{A}{\log(k)}\|m_0\log(m_0)\|_1 < \frac{\epsilon}{3},$$

then we find a $\delta > 0$ such that

$$|t-s| < \delta \implies C_R \|T_k(m(t)) - T_k(m(s))\|_2 < \frac{\epsilon}{3}$$

and so $m \in C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d))$. ◻

2.4.2 Analysis of the Hamilton-Jacobi-Bellman equation

In this section we will study the bounds for the HJB equation

$$\begin{cases} -\partial_t u - \Delta_v u + v \cdot D_x u + H(D_v u) = f(t,x,v) \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ u(T,x,v) = g(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.4.14}$$

Definition 6. *Let $H : \mathbb{R}^d \to \mathbb{R}$ be a convex Lipschitz function such that $H \geq 0$, $f \in L^1 \cap L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $f \geq 0$, $(|x|^2 + |v|^2) f \in L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ $g \in L^1 \cap L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, $g \geq 0$, $(|x|^2 + |v|^2) g \in L^1(\times \mathbb{R}^d \times \mathbb{R}^d)$ and $u \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d)) \cap L^1(\mathbb{R}^d \times \mathbb{R}^d)$ with $D_v u \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $u \geq 0$. We say that (u, H, f, g) is a weak solution of* (2.4.14)*, if the equation is satisfied in the distributional sense.*

Our starting point is the following compactness theorem found in the Appendix of [151].

Theorem 10 (Appendix of P.-L. Lions, DiPerna [151])**.** *Assume that $u^n, f^n \in L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$, $g^n \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ satisfy in the distributional sense*

$$\partial_t u_n - \Delta_v u_n + v \cdot D_x u_n = f_n, \;\; u_n(0) = g^n.$$

If g^n, f_n are uniformly bounded in L^1 with

$$\lim_{R \to \infty} \sup_{n \in \mathbb{N}} \int_0^T \int_{|(x,v)| \geq R} |f^n| dx dv dt = 0 \tag{2.4.15}$$

and

$$\lim_{R\to\infty}\sup_{n\in\mathbb{N}}\int_{|(x,v)|\geq R}|g_0^n|dxdv=0, \tag{2.4.16}$$

then the sequence $\{u_n\}_{n\in\mathbb{N}}$ *is compact in* $L^1((0,T)\times\mathbb{R}^d\times\mathbb{R}^d)$.

Theorem 11. *Let* $f^n\in L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, $g^n\in L^1(\mathbb{R}^d\times\mathbb{R}^d)$ *be non-negative, uniformly integrable sequences and* $H^n:\mathbb{R}^d\to\mathbb{R}$ *Lipschitz convex Hamiltonians. Assume that* $\{(u^n,H^n,f^n,g^n)\}_{n\in\mathbb{N}}$ *are weak solution to* (2.4.14) *according to definition 6. Then, the sequence* $\{u^n\}$ *is compact in* $L^1((0,T)\times\mathbb{R}^d\times\mathbb{R}^d)$ *and*

$$\sup_{n\in\mathbb{N}}\Big(\sup_{t\in[0,T]}\|u^n(t)\|_1+\|H^n(D_vu^n)\|_1\Big)<\infty,$$

$$\lim_{R\to\infty}\sup_{n}\Big(\sup_{t\in[0,T]}\int_{B(0,R)^c}|u^n|(t)dxdv+\int_{B(0,R)^c}H^n(D_vu^n)dxdvdt\Big)=0.$$

Proof. By the same arguments as in Lemma 1, we can justify testing against 1 in the HJB equation to obtain the uniform L^1 estimates on $u^n, H^n(D_vu^n)$. To show compactness in L^1 we work as follows. Let $L:=-\partial_t-\Delta_v+v\cdot D_x$ and since $H^n\geq 0, f^n\geq 0, g^n\geq 0$ the functions u^n are non-negative and satisfy

$$Lu^n\leq f^n \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d,\ u^n(T)=g^n \text{ in } \mathbb{R}^d\times\mathbb{R}^d.$$

For each $n\in\mathbb{N}$, let w^n be the solution of

$$Lw^n=f^n \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d,\ w^n(T)=g^n \text{ in } \mathbb{R}^d\times\mathbb{R}^d.$$

Since $L(w^n-u^n)\geq 0$ and $w^n(T)=u^n(T)$ we have that

$$0\leq u^n\leq w^n. \tag{2.4.17}$$

Since f^n, g^n are uniformly integrable, by Theorem 10 the set $\{w^n\}_{n\in\mathbb{N}}$ is compact in L^1 and so in particular uniformly integrable and from (2.4.17) we see that $\{u^n\}_{n\in\mathbb{N}}$ are also uniformly integrable. For $R > 0$, let $\phi_R : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ be cutoff functions defined just as in Lemma 1. Testing against ϕ_R in

$$Lu^n + H(D_v u^n) = f^n, \; u^n(T) = g^n$$

yields for some dimensional constant $C > 0$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} u^n(t)\phi_R dxdv + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} H^n(D_v u^n)\phi_R dxdvdt \leq$$

$$\frac{C}{R}\|u^n\|_1 + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} f^n \phi_R dxdvdt + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} g^n\phi_R + C\int_{R<|(x,v)|<2R} u^n dxdv$$

and since the sequence $\{u^n\}_{n\in\mathbb{N}}$ is uniformly integrable we see that the terms on the right vanish uniformly in n as $R \uparrow \infty$. Finally with the estimate

$$\lim_{R\to\infty} \sup_{n\in\mathbb{N}} \int_{R<|(x,v)|} H^n(D_v u^n) dxdvdt = 0$$

the compactness for u^n in L^1 follows immediately by Theorem 10 with $\widetilde{f^n} = f^n - H^n(D_v u^n)$. □

Theorem 12. *Let* (u, H, f, g) *be a weak solution of* (2.4.14)*, according to Definition 6. Then, there exists a constant* $C = C(d, T) > 0$*, such that*

$$\sup_{t\in[0,T]} \|u(t)\|_2 + \|uH(D_v u)\|_1 + \|D_v u\|_2 \leq C\Big(\|f\|_\infty\|f\|_1 + \|g\|_1\|g\|_\infty\Big). \tag{2.4.18}$$

Proof. The result follows by testing against u in (2.4.14) and applying Grönwall. □

Proposition 8. *Let* $\{(u^n, H^n, f^n, g^n)\}_{n\in\mathbb{N}}$*, be weak solutions of* (2.4.14)*, according to Definition 6, such that*

$$\|f^n\|_1 + \|g^n\|_1 \leq C \textit{ for all } n \in \mathbb{N},$$

and

$$u^n \to u \text{ strongly in } L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d).$$

Then, the limit u belongs to $L^2([0,T] \times \mathbb{R}^d; H^1(\mathbb{R}^d_v))$ *and*

$$D_v u^n \to D_v \text{ in } L^q_{loc}([0,T] \times \mathbb{R}^d \times \mathbb{R}^d),$$

for all $q < 2$*, up to a subsequence almost everywhere.*

Proof. The equation for $u^n - u^m$ is

$$-\partial_t(u^n - u^m) - \Delta_v(u^n - u^m) + v \cdot D_x(u^n - u^m) = f^n - f^m,$$

$$(u^n - u^m)(T) = g^n - g^m.$$

For $\epsilon > 0$, we define

$$\phi(s) := \begin{cases} s, \text{ for } s \in [-\epsilon, \epsilon], \\ -\epsilon, \text{ for } s \leq -\epsilon, \\ \epsilon, \text{ for } s \geq \epsilon, \end{cases}$$

and $\Phi(t) := \int_0^t \phi(s)ds \geq 0$. We test against $\phi(u^n - u^m)$ in the equation for the difference, which yields

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(u^n - u^m)(t) dxdv + \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi'(u^n - u^m) |D_v(u^n - u^m)|^2 dxdv$$

$$\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi(u^n - u^m)(T) dxdv + \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(u^n - u^m)(f^n - f^m) dxdvdt$$

$$\leq C\epsilon \|g^n - g^m\|_1 + \epsilon \|f^n - f^m\|_1 \leq C\epsilon.$$

Therefore,

$$\int_{|u^n - u^m| \leq \epsilon} |D_v(u^n - u^m)|^2 dxdvdt \leq C\epsilon.$$

Thus, fixing a radius $R > 0$ and a $q < 2$ we obtain

$$\int_{B(0,R)} |D_v(u^n - u^m)|^q dxdv \leq \int_{B(0,R)\cap\{|u^n-u^m|\leq\epsilon\}} |D_v(u^n - u^m)|^q dxdvdt$$

$$+ \int_{B(0,R)\cap|u^n-u^m|>\epsilon} |D_v(u^n - u^m)|^q dxdvdt \leq CR^d\epsilon + CR^d|\{|u^n - u^m| > \epsilon\}|^\theta$$

for some $\theta = \theta(q) \in (0,1)$. Since u^n converges in L^1, we have that $\lim_{n,m\to\infty} |\{|u^n - u^m| > \epsilon\}| = 0$ and so $D_v u^n \to D_v u$ in $L^q([0,T] \times B(0,R))$ for all $R > 0$. □

Proposition 9. *Assume that* $\{(u^n, H^n, f^n, g^n)\}_{n\in\mathbb{N}}$ *are weak solutions to* (2.4.14) *according to Definition 6, such that* $\{g^n\}_{n\in\mathbb{N}} \subset L^1(\mathbb{R}^d \times \mathbb{R}^d)$ *is uniformly integrable,* $\{f^n\}_{n\in\mathbb{N}}$ *and* $\{g^n\}_n$ *are bounded subsets of their respective* L^∞ *spaces, and for some* u, f, $u^n \to u, f^n \to f, f^n \to f$, *in* $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ *and almost everywhere. Then, up to a subsequence, for each* $\tau \in [0,T)$*, we have that*

$$H^n(D_v u^n) \to H(D_v u) \text{ in } L^1([0,\tau] \times \mathbb{R}^d \times \mathbb{R}^d)$$

and,

$$D_v u^n \to D_v u \text{ in } L^2([0,\tau] \times \mathbb{R}^d \times \mathbb{R}^d).$$

Proof. From Proposition 8 by choosing a subsequence if necessary we can assume that $H^n(D_v u^n) \to H(D_v u)$ almost everywhere, furthermore since $\sup_n \|f^n\|_\infty + \|g^n\|_\infty < \infty$, for some $C > 0$ we have that $\|u^n\|_\infty \leq C$ for all $n \in \mathbb{N}$. Denote by $L := -\partial_t - \Delta_v + v \cdot D_x$ and test against $(T-t)e^{\lambda(u^n - u^k)}$ in the equation

$$L(u^n - u^k) + [H^n(D_v u^n) - H^k(D_v u^k)] = f^n - f^k.$$

Which yields,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} T\frac{1}{\lambda}(e^{\lambda(u^n-u^k)} - 1)(0)dxdv - \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{\lambda}(e^{\lambda(u^n-u^k)} - 1)(s)dxdvds$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)e^{\lambda(u^n-u^k)}|D_v(u^n-u^k)|^2+(T-s)e^{\lambda(u^n-u^k)}\Big(H^n(D_vu^n)-H^k(D_vu^k)\Big)dxdvds$$

$$=\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}e^{\lambda(u^n-u^k)}\Big(f^n-f^k\Big)dxdvds.$$

Next using the strong convergence of u^n, f^n and that u^n is uniformly bounded in L^∞, we obtain that for some function $\omega(n,k)$ such that $\lim_{n,k\to\infty}\omega(n,k)=0$

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)\lambda e^{\lambda(u^n-u^k)}|D_v(u^n-u^k)|^2dxdvds$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)e^{\lambda(u^n-u^k)}\Big(H^n(D_vu^n)-H^k(D_vu^k)\Big)dxdvds\le\omega(n,k)$$

If $n>k$ we have that $H^k\le H^n$, hence by the convexity of H

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)\lambda e^{\lambda(u^n-u^k)}|D_v(u^n-u^k)|^2dxdvds$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)e^{\lambda(u^n-u^k)}\Big(H^n(D_vu^n)-H^n(D_vu^k)\Big)dxdvds\le\omega(n,k)$$

$$\implies\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)\lambda e^{\lambda(u^n-u^k)}|D_v(u^n-u^k)|^2dxdvds$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)e^{\lambda(u^n-u^k)}H^n_p(D_vu^k)D_v(u^n-u^k)dxdvds\le\omega(n,k).$$

Letting $n\to\infty$ and using that $D_vu^n\to D_vu$ almost everywhere and weakly in L^2, while $u^n\to u$ strongly in L^1 with $\|u^n\|_\infty\le C$ and $|H^n_p(D_vu^k)|\le|H_p|(D_vu^k)$ thus $H^n_p(D_vu^k)\to H_p(D_vu^k)$ strongly in L^2, yields

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)\lambda e^{\lambda(u-u^k)}|D_v(u-u^k)|^2dxdvds$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(T-s)e^{\lambda(u-u^k)}H_p(D_vu^k)D_v(u-u^k)dxdvds\le\omega(k).$$

From 2.2, there exists a constant $C > 0$ such that

$$H_p(D_v u^k) D_v(u - u^k) = -(H_p(D_v u) - H_p(D_v u^k)) \cdot D_v(u - u^k) + H_p(D_v u) D_v(u - u^k)$$

$$\geq -C|D_v(u - u^k)|^2 + H_p(D_v u) D_v(u - u^k)$$

$$\implies \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} (T - s) e^{\lambda(u - u^k)} (\lambda - C) |D_v(u - u^k)|^2 dxdvds$$
$$+ \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} (T - s) e^{\lambda(u - u^k)} H_p(D_v u) D_v(u - u^k) dxdvds \leq \omega(k)$$

and again by the weak convergence of $D_v(u - u^k)$ in L^2 and the strong convergence of u^k to u in L^1 with uniform bounds we obtain

$$\int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} (T - s) e^{\lambda(u - u^k)} (\lambda - C) |D_v(u - u^k)|^2 dxdvds \leq \omega(k).$$

Finally, the result follows since by choosing $\lambda > C$ and using that $\|u - u^k\|_\infty \leq C$ we obtain that for some constant $c_0 > 0$ depending only on H

$$c_0 \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} (T - s) |D_v(u - u^k)|^2 dxdvds \leq \omega(k).$$

□

Theorem 13. *Assume that* $\{(u^n, H^n, f^n, g^n)\}_{n \in \mathbb{N}}$ *are weak solutions to* (2.4.14) *according to Definition 6, such that* $f^n \to f$ *in* L^1, $g^n \to g$, *weakly in* L^1, $u^n \to u$ *in* L^1 *and* $D_v u^n \to D_v u$ *almost everywhere and* $H^n(D_v u^n) \to H(D_v u)$ *in* $L^1_{loc}((0, T]; L^1(\mathbb{R}^d \times \mathbb{R}^d))$, *where* $H(D_v u) \in L^1([0, T] \times \mathbb{R}^d \times \mathbb{R}^d)$. *Then, we have that* $u \in C((0, T]; L^1((\mathbb{R}^d \times \mathbb{R}^d))$.

Proof. The result follows by the fact that $Lu \in L^1$, where $L := -\partial_t - \Delta_v + v \cdot D_x$, see for example [151]. □

2.4.3 Existence and uniqueness for the quadratic case

In this subsection, we will establish the existence and uniqueness of renormalized solutions for the MFG system.

Theorem 14. *Assume that* $H : \mathbb{R}^d \to \mathbb{R},\ F : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R},\ m_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ *and* $G : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ *satisfy 2.2,2.2,2.2 and 2.2. Then, there exists a unique renormalized solution* (m, u) *of system* (2.1.1)*, according to Definition 4.*

Proof. The proof is divided in two steps. First we show the result for F, G bounded in their respective $L^\infty-$spaces and let the Hamiltonians H^ϵ vary. While in the second case we show the result for a fixed quadratic Hamiltonian H while letting F^n, G^n vary. The reason for this approach is so that we can always have bounds on $D_v u^n$ in L^2. Indeed in the first case the bounds follow by Theorem 12 and are due to the Δ_v term while in the second case the bounds are a result of Theorem 11 and are due to $\|H(D_v u^n)\|_1 \le C$.

First Case: For H^ϵ, as defined in (2.4.1), we consider the solutions $(m^\epsilon, u^\epsilon, m_0)$ provided by Theorem 1. From Proposition 2 above, we have that for some $C > 0$ independent of ϵ

$$\|m^\epsilon |H_p^\epsilon(D_v u^\epsilon)|^2\|_1 \le C, \text{ for all } \epsilon > 0, \tag{2.4.19}$$

furthermore, by Theorem 12 and our assumptions on H^ϵ we have that

$$\|H_p^\epsilon(D_v u^\epsilon)\|_2 \le C, \text{ for all } \epsilon > 0.$$

Therefore, from Theorem 8, we may extract a subsequence m^n, which is convergent in $L^1([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ and almost everywhere to some m. From Remark 2, we have that the sequence $\{F(t,x,v,m^n)\}_{n \in \mathbb{N}}$ is uniformly integrable, indeed in the case $f_L := \sup\limits_{m \in [0,L]} F(t,x,v,m) \in L^1$ the claim holds just as

[325], while in the case $f_L := \sup_{m\in[0,L]} F(t,x,v,m)/m \in L^\infty$ since

$$0 \le F(t,x,v,m^n) \le f_L(t,x,v)m^n + \frac{m^n}{L}F(t,x,v,m^n)$$

the result follows due to uniform bound on $\|F(t,x,v,m^n)m^n\|_1$ from Proposition 2 and the convergence of m^n in L^1. Since $m^n \to m$ almost everywhere, we obtain

$$F(\cdot,m^n(\cdot)) \to F(\cdot,m(\cdot)) \text{ strongly in } L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d).$$

By choosing a further subsequence if necessary, Theorem 11, Lemma 1 and Proposition 8, yield a $u \in C([0,T];L^1(\mathbb{R}^d\times\mathbb{R}^d)) \cap L^2([0,T]\times\mathbb{R}^d;H^1(\mathbb{R}^d_v))$, such that

$$u^n \to u \text{ almost everywhere and strongly in } L^1([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$$

$$D_v u^n \to D_v u \text{ almost everywhere and in } L^1_{loc}([0,T]\times\mathbb{R}^d\times\mathbb{R}^d).$$

Furthermore, again by taking subsequences if needed, by Proposition 9 we have that for each $\tau \in [0,T)$,

$$H^{\epsilon_n}(D_v u^n) \to H(D_v u) \text{ in } L^1([0,\tau]\times\mathbb{R}^d\times\mathbb{R}^d)$$

and for each $k \in \mathbb{N}$,

$$D_v(u^n \wedge k) \to D_v(u \wedge k) \text{ in } L^2([0,\tau]\times\mathbb{R}^d\times\mathbb{R}^d).$$

By inequality (2.4.19) and the fact that $H^{\epsilon_n}_p(D_v u^n) \to H^{\epsilon_n}_p(D_v u)$ almost everywhere, Proposition 7 implies that

$$m^n \to m \text{ in } C([0,T];\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)),$$

and by Theorem 9, m is a renormalized solution of

$$\partial_t m - \Delta_v m - v \cdot D_x m - \mathrm{div}_v(mH_p(D_v u)) = 0 \text{ in } (0,T] \times \mathbb{R}^d \times \mathbb{R}^d, m(0) = m_0 \text{ in } \mathbb{R}^d \times \mathbb{R}^d. \quad (2.4.20)$$

It remains to show the convergence of the terminal data in the HJB equation. This follows exactly as in [325]. Thus, we have that $m^n(T) \to m(T)$ in $L^1(\mathbb{R}^d \times \mathbb{R}^d)$ which from Remark 2 implies that $G(\cdot, m^n(T,\cdot)) \to G(\cdot, m(T,\cdot))$ in $L^1(\mathbb{R}^d \times \mathbb{R}^d)$. Thus, the limit u is also a renormalized solution.

Second Case: Next, given F, G that satisfy 2.2 and 2.2 respectively, consider $F^n := F \wedge n, G^n := G \wedge n$ for $n \in \mathbb{N}$. The functions F^n, G^n clearly also satisfy 2.2 and 2.2 respectively. Let (u^n, m^n) be the solutions provided for the data (H, F^n, G^n) by the first case. The rest of the proof follows exactly the first case only now we use Theorem 16 to obtain the convergence of $D_v T_k(u^n)$.

Finally, we address the issue of uniqueness whose proof follows the same exact arguments as [325] once we establish that $m(t,x,v) > 0$ almost everywhere. But this will follow from assumption (2.2) and in particular $\log(m_0) \in L^1_{loc}(\mathbb{R}^d \times \mathbb{R}^d)$. Indeed, let $R > 0$ and define $\phi_R : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ such that

$$\phi_R(x,v) := \begin{cases} 1 \text{ if } |(x,v)| \le R, \\ 0 \text{ if } |(x,v)| \ge R+1. \end{cases}$$

Then since

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \log(m(t))\phi_R^2 dxdv \le \int_{\mathbb{R}^d \times \mathbb{R}^d} m(t)\phi_R^2 dxdv \le 1,$$

it is enough to bound $\int_{\mathbb{R}^d \times \mathbb{R}^d} \log(m(t))\phi_R dxdv$ from bellow, since that would imply $m(t,x,v) > 0$ almost everywhere. To show the lower bound we test the equation satisfied by m with $\phi_R^2 \frac{1}{m}$ (technically we would need to fix a $\delta > 0$ and test against $\phi_R^2 \frac{1}{m+\delta}$ and let $\delta \to 0$ but we skip the approximation for simplicity). This yields

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \log(m(t))\phi_R^2 dxdv + \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} -\frac{|D_v m|^2}{m^2}\phi_R^2 + 2\frac{D_v m}{m}\phi_R D_v \phi_R - \frac{D_v m}{m^2} m H_p \phi_R^2$$

$$+2\phi_R D_v\phi_R H_p dxdvdt = \int_{\mathbb{R}^d\times\mathbb{R}^d} \log(m_0)\phi_R^2 dxdv.$$

Next we use the following inequalities

$$2\Big|\frac{D_v m}{m}\phi_R D_v\phi_R\Big| \leq \frac{1}{4}\frac{|D_v m|^2}{m^2}\phi_R^2 + 4|D_v\phi_R|^2$$

$$\Big|\frac{D_v m}{m^2} m H_p \phi_R^2\Big| \leq \frac{1}{4}|\frac{D_v m}{m}|^2\phi_R^2 + |H_p|^2\phi_R^2$$

$$\Big|2\phi_R D_v\phi_R H_p\Big| \leq |H_p|^2 + 2|\phi_R|^2|D_v\phi_R|^2$$

and thus combining everything we obtain that for some constant $C = C(R, d) > 0$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \log(m(t))\phi_R^2 \geq \int_{\mathbb{R}^d\times\mathbb{R}^d} C(R, d) + \|\log(m_0)\phi_R^2\|_1 - \|H_p(D_v u)\|_2$$

which proves the claim. □

2.4.4 Further regularity for quadratic Hamiltonian

Theorem 15. *Let (H, F, G, m_0) be as in Theorem 14 with $F = F(m)$, $G = G(m)$ and m_0 also satisfying $\|D^2 m_0\|_\infty \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$. Then, there exists a constant $C(F, G, H, T, m_0)$, such that*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} G'(m(T, x, v))|Dm(T, x, v)|^2 dxdv + \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} F'(m(t, x, v))|Dm(t, x, v)|^2$$

$$+m\sum_{k=1}^{2d} mD_v u_k H_{pp}(D_v u)D_v u_k dxdv \leq C.$$

Proof. The proof is almost identical to the one in the case of Lipschitz Hamiltonian. The only difference is now instead of using the HJB equation we estimate

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \delta^h m_0 \delta^h u(0) dxdv = \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{m_0(x+h, v) - 2m_0(x, v) + m_0(x-h, v)}{h^2} u(0, x, v)dxv$$

$$\leq \|D^2 m_0\|_{L^\infty(\mathbb{R}^d\times\mathbb{R}^d)} \|u(0,\cdot,\cdot)\|_1,$$

and conclude due to the estimate in Proposition 2.3.3. □

2.5 Appendix of Chapter 2

2.5.1 Technical results

In this sub-section we show some important properties about the convergence of u^n where u^n solves

$$\begin{cases} Lu^n + H(D_v u^n) = f^n \text{ in } (0,T]\times\mathbb{R}^d\times\mathbb{R}^d, \\ u^n(0) = g^n \text{ in } \mathbb{R}^d\times\mathbb{R}^d, \end{cases} \tag{2.5.1}$$

for $L := \partial_t - \Delta_v + v\cdot D_x$ and f^n, g^n strongly convergent sequences in their respective L^1-spaces. We show an analogue of the convergence results in [323] from which all our techniques are motivated and parallel to. In particular we show that if u^n solves (2.5.1) and are strongly convergent in L^1 to some function u, then $D_v T_k(u^n) \to D_v T_k(u)$ strongly in $L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, where T_k is the truncation at k, namely, for $k\in\mathbb{N}$ $T_k(x) = s$ for $|s|\leq k$ and $T_K(s) = sign(s)k$ otherwise. A crucial technical step in [323] is the following transformation which allows the authors to deal with the degenerate ∂_t direction. Given a function u, for $\nu > 0$ define

$$\partial_t(u)_\nu = \nu(T_k(u) - (u)_\nu).$$

This transformation enjoys many nice properties such as $(u)_\nu \to u$ and $D((u)_\nu) \to Du$ as $\nu\to\infty$ in appropriate spaces. In our setup the above transformation does not seem to work due to the extra degenerate operator $v\cdot D_x$. In order to deal with this, we consider a slightly different transformation. Fix $\alpha > 0$ and consider the solution of

$$L\Phi_\alpha = \alpha(T_k(u) - \Phi_\alpha).$$

We will show that under the condition $u \in L^1$ the transformation Φ_α converges to $T_k(u)$ in L^1, however, we cannot show in general, even if $D_v u \in L^2$, that $D_v\Phi_\alpha \to D_v T_k(u)$ strongly in L^2, with no assumptions on $D_x u$. However the fact that $Lu^n + H(D_v u^n) = f^n$ and $u^n \to u$ strongly in L^1, is enough to show the strong convergence of $D_v\Phi_\alpha$. With this, we can follow the rest of the argument found in [323].

Lemma 2. *Let $u \in L^1 \cap L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d) \cap C([0,T]; L^1(\mathbb{R}^d \times \mathbb{R}^d))$ and $\alpha > 0$. Then, there exists a unique function $\Phi_\alpha \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ with $D_v\Phi_\alpha \in L^2([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ which solves*

$$\begin{cases} \partial_t\Phi_\alpha - \Delta_v\Phi_\alpha + v \cdot D_x\Phi_\alpha = \alpha(u - \Phi_\alpha) \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ \Phi_\alpha(0,x,v) = u(0,x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.5.2}$$

Furthermore, the functions Φ_α have the following properties

1. $u \geq 0 \implies \Phi_\alpha \geq 0$ *almost everywhere,*

2. $\|\Phi_\alpha\|_\infty \leq \|u\|_\infty$,

3. $\lim_{\alpha\to\infty} \|\Phi_\alpha - u\|_2 = 0$

4. $\|\Phi_\alpha\|_1 \leq \|u\|_1 + \frac{1}{\alpha}\|u_0\|_1$

Proof. First we assume that $u \in C^\infty([0,T]; C_c^\infty(\mathbb{R}^d \times \mathbb{R}^d))$. Let Γ denote the fundamental solution of $L = \partial_t - \Delta_v + v \cdot D_x$. Then, it is easy to check that the solution of equation (2.5.2) is given by

$$\Phi_\alpha(t,x,v) = \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} \alpha e^{-\alpha(t-s)}\Gamma(t-s,x,v,y,w)u(s,y,w)dydwds$$

$$+ \int_{\mathbb{R}^d\times\mathbb{R}^d} \alpha e^{-\alpha t}\Gamma(t,x,v,y,w)u(0,y,w)dydw,$$

see for example [302]. Furthermore, the solution Φ_α is also C^∞ since L is hypoelliptic. Let $f := L(u) \in C^\infty([0,T] \times C_c(\mathbb{R}^d \times \mathbb{R}^d))$. In the equation

$$L(u - \Phi_\alpha) = -\alpha(u - \Phi_\alpha) + f, \quad (u - \Phi_\alpha)(0) = 0,$$

we test against $(u - \Phi_\alpha)$, which yields

$$\frac{d}{dt}\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |u - \Phi_\alpha|^2 dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u - \Phi_\alpha)|^2 dxdv$$

$$= -\alpha \int_{\mathbb{R}^d\times\mathbb{R}^d} |u - \Phi_\alpha|^2 dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d} f(u - \Phi_\alpha) dxdv \leq \frac{1}{4\alpha}\int_{\mathbb{R}^d\times\mathbb{R}^d} |f|^2 dxdv.$$

Hence, we obtain that

$$\sup_{t\in[0,T]} \|u(t) - \Phi_\alpha(t)\|_2 + \|D_v(u - \Phi_\alpha)\|_2 \leq \frac{C}{\alpha}$$

where $C = C(T, f) > 0$. Furthermore, by testing against $p|u - \Phi_\alpha|^{p-2}(u - \Phi_\alpha)$ for $p > 1$ yields

$$\frac{d}{dt}\int |u - \Phi_\alpha|^p dxdv + \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u - \Phi_\alpha)|^2 |u - \Phi_\alpha|^{p-2} p(p-1) dxdv$$

$$\leq -\alpha p \int_{\mathbb{R}^d\times\mathbb{R}^d} |u - \Phi_\alpha|^p + p\int_{\mathbb{R}^d\times\mathbb{R}^d} |f||u - \Phi_\alpha|^{p-1} dxdv \leq \frac{p}{4a}\int_{\mathbb{R}^d\times\mathbb{R}^d} |f|^p dxdv,$$

where $1/p + 1/q = 1$. Letting $p \to 1$ yields

$$\sup_{t\in[0,1]} \|u - \Phi_\alpha\|_1 \leq \frac{C}{\alpha}\|f\|_1,$$

where $C = C(T, f) > 0$. The first two claims now follow easily by the Maximum Principle. For the general case we work as follows. Testing against $p|\Phi_\alpha|^{p-2}\Phi_\alpha$ in (2.5.2) for $p > 1$ and letting $p \to 1$ just as above we obtain

$$-\int_{\mathbb{R}^d\times\mathbb{R}^d} |u_0| dxdv \leq \frac{\alpha}{2}\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |u| dxdvdt - \frac{\alpha}{2}\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |\Phi_\alpha| dxdvdt.$$

Hence,

$$\|\Phi_\alpha\|_1 \leq \|u\|_1 + \frac{2}{\alpha}\|u_0\|_1,$$

and so by linearity of the map $(u, u_0) \to \Phi_\alpha$ and the fact that $|u| \leq k \implies |\Phi_\alpha| \leq k$ the result holds in the general case. ◻

Theorem 16. *Let $H : \mathbb{R}^d \to \mathbb{R}$ be a Hamiltonian satisfying 2.2. Assume that $\{f^n\}_{n\in\mathbb{N}} \subset L^1 \cap L^\infty([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, $\{g^n\}_{n\in\mathbb{N}} \subset L^1 \cap L^\infty([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ such that $f^n \to f$ and $g^n \to g$ strongly in the respective L^1 spaces (the limits need not be in L^∞). Let $u^n \in L^1 \cap L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ with $D_v u^n \in L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$ solve*

$$\begin{cases} \partial_t u^n - \Delta_v u^n + v \cdot D_x u^n + H(D_v u^n) = f^n, \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ u^n(0,x,v) = g^n(x,v) \text{ in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases} \tag{2.5.3}$$

Finally, assume that $u^n \to u$ strongly in L^1 and that $D_v u^n \to D_v u$ almost everywhere. Then, the limit u is a renormalized solution of

$$\begin{cases} \partial_t u - \Delta_v u + v \cdot D_x u + H(D_v u) = f(t,x,v) \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d, \\ u(0,x,v) = g(x,v) \text{ in } \mathbb{R}^d\times\mathbb{R}^d, \end{cases}$$

according to Definition 3.

Proof. Following [323], we see that the result will hold once we show that for some increasing sequence $0 \leq m_k \in \mathbb{R}, k \in \mathbb{N}$ with $m_k \uparrow \infty$ as $k \to \infty$, $D_v(T_{m_k}(u^n)) \to D_v(T_{m_k}(u))$ strongly in $L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$, where

$$T_k(s) := \begin{cases} s, \text{ if } |s| \leq k, \\ k, \text{ if } s > k, \\ -k, \text{ if } s < -k. \end{cases} \tag{2.5.4}$$

Note that for almost all $\beta \in \mathbb{R}$, we have that $|\{u = \beta\}| = 0$ ($|A|$ denotes the Lebesgue measure), therefore in order to keep the notation lighter we may assume that $|\{u = k\}| = 0$ and thus choose the sequence $m_k = k$. The reason for this choice will become apparent later; in particular to prove that $\chi_{u^n > m_k} \to \chi_{u > m_k}$ almost everywhere, it is convenient to know that $|\{u = m_k\}| = 0$. In the rest of the proof we will use the notation $\omega(n)$ and $\omega(n, \alpha)$, for quantities that satisfy $\lim_{n\to\infty} \omega(n) = 0$ and $\lim_{\alpha\to\infty} \lim_{n\to\infty} \omega(n, \alpha) = 0$ respectively, furthermore these quantities are subject to change from line to line. Just as in [323] and the references therein, for $\lambda > 0$ we define $\phi_\lambda(s) := s \exp(\lambda s^2)$. For $\alpha > 0$ and $k \in \mathbb{N}$, consider the solution $\Phi_{\alpha,k}$ of

$$\begin{cases} \partial_t \Phi_{\alpha,k} - \Delta_v \Phi_{\alpha,k} + v \cdot D_x \Phi_{\alpha,k} = \alpha(T_k(u) - \Phi_{\alpha,k}), \\ \Phi_{\alpha,k}(0) = T_k(g). \end{cases} \tag{2.5.5}$$

Denote by $L := \partial_t - \Delta_v + v \cdot D_x$ and test equation (2.5.3) against $\phi_\lambda(u^n - \Phi_{\alpha,k})^-$ which yields

$$\boxed{\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n - \Phi_{\alpha,k}), \phi_\lambda(u^n - \Phi_{\alpha,k})^- \rangle dxdvdt}_1$$

$$+ \boxed{\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L\Phi_{\alpha,k}, \phi_\lambda(u^n - \Phi_{\alpha,k})^- \rangle dxdvdt}_2$$

$$+ \boxed{\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} H(D_v u^n)\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt}_3 = \boxed{\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} f^n \phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt}_4 .$$

Let $\Phi_\lambda(s) := \int_0^s \phi_l(\theta)^- d\theta$, then the first boxed term gives us

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \Phi_\lambda(u^n - \Phi_{\alpha,k})(T)\Phi_\lambda(g^n - T_k(g))dxdv - \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'_\lambda(u^n - \Phi_{\alpha,k})^- |D_v(u^n - \Phi_{\alpha,k})|^2 dxdvdt$$

$$\leq \omega(n) - \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi'_\lambda(u^n - \Phi_{\alpha,k})^- |D_v(u^n - \Phi_{\alpha,k})|^2 dxdvdt,$$

where in the last inequality we used that $\Phi_\lambda(s) := \int_0^s \phi_\lambda(u)^- du \leq 0$ and that $g^n \to g$ strongly in

L^1. For the second boxed term we obtain

$$\alpha \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (T_k(u) - \Phi_{\alpha,k})\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt \leq \alpha\omega(n),$$

since $u^n \to u$ strongly in L^1, $\phi_\lambda(u^n - \Phi_{\alpha,k})^- = \phi_\lambda(T_k(u^n) - \Phi_{\alpha,k})^-$ and $s\phi_\lambda(s)^- \leq 0$. For the third boxed term we have that for some constant $C > 0$, depending only on H

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} H(D_v u^n)\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt \leq C \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u^n)|^2\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt,$$

and using that for all $p, q \in \mathbb{R}^d$, we have $|p|^2 \leq 2|p-q|^2 + 2|q|^2$ the third boxed term is bounded by

$$2C \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(u^n - \Phi_{\alpha,k})|^2\phi_\lambda(u^n - \Phi_{\alpha,k})^- + 2C|D_v(\Phi_{\alpha,k})|^2\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt.$$

Finally, the last boxed term vanishes as $n \to \infty$ and then $\alpha \to \infty$ due to Lemma 2. Putting everything together we obtain

$$2C \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \left[\phi'_\lambda(u^n - \Phi_{\alpha,k})^- - \phi_\lambda(u^n - \Phi_{\alpha,k})^-\right]|D_v(u^n - \Phi_{\alpha,k})|^2 dxdvdt$$

$$\leq \omega(n,\alpha) + 2C \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(\Phi_{\alpha,k})|^2\phi_\lambda(u^n - \Phi_{\alpha,k})^- dxdvdt.$$

By choosing λ large enough depending only on $\|H_{pp}\|_\infty$, we have that $\phi'_\lambda(u^n - \Phi_{\alpha,k})^- - \phi_\lambda(u^n - \Phi_{\alpha,k})^- \geq 0$ thus by Fatous Lemma on the LHS and the strong convergence of $u^n \to u$ in L^1, as $n \to \infty$ we obtain

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \left[\phi'_\lambda(u - \Phi_{\alpha,k})^- - 2C\phi_\lambda(u - \Phi_{\alpha,k})^-\right]|D_v(u - \Phi_{\alpha,k})|^2 dxdvdt$$

$$\leq \omega(\alpha) + 2C \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(\Phi_{\alpha,k})|^2\phi_\lambda(u - \Phi_{\alpha,k})^- dxdvdt.$$

Furthermore,

$$2C\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_v(\Phi_{\alpha,k})|^2\phi_\lambda(u-\Phi_{\alpha,k})^-dxdvdt$$

$$\leq 4C\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_v(\Phi_{\alpha,k}-u)|^2\phi_\lambda(u-\Phi_{\alpha,k})^-dxdvdt+4C\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_vu|^2\phi_\lambda(u-\Phi_{\alpha,k})^-dxdvdt.$$

Hence,

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}\left[\phi'_\lambda(u-\Phi_{\alpha,k})^- - 6C\phi_\lambda(u-\Phi_{\alpha,k})^-\right]|D_v(u-\Phi_{\alpha,k})|^2dxdvdt$$

$$\leq \omega(\alpha)+4C\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_vu|^2\phi_\lambda(u-\Phi_{\alpha,k})^-dxdvdt,$$

now we may fix $\lambda>0$ such that $\phi'_\lambda(s)^- - 6C\phi_\lambda(s)^- \geq \frac{1}{2}$ and so letting $\alpha\to\infty$ yields

$$\lim_{\alpha\to\infty}\|D_v(T_k(u)-\Phi_{\alpha,k})^-\|_2=0.$$

We now show the convergence on the set $T_k(u)\geq\Phi_{\alpha,k}$. Since $H\geq 0$ the functions u^n are subsolutions of

$$\begin{cases} Lu^n \leq f^n(t,x,v) \text{ in } (0,T)\times\mathbb{R}^d\times\mathbb{R}^d \\ u^n(0,x,v)=g^n(x,v). \end{cases} \tag{2.5.6}$$

Define $w^n=(T_k(u^n)-\Phi_{\alpha,k})_+$ which may also be written as

$$w_n=(u^n-\Phi_{\alpha,k})_+-(u^n-T_k(u^n)).$$

Indeed if $u^n\leq k$ then

$$(u^n-\Phi_{\alpha,k})_+-(u^n-T_k(u^n))=(u^n-\Phi_{\alpha,k})_+=(T_k(u^n)-\Phi_{\alpha,k})_+,$$

while if $u^n>k$ since $0\leq\Phi_{\alpha,k}\leq k$

$$(u^n-\Phi_{\alpha,k})_+-(u^n-T_k(u^n))=u^n-\Phi_{\alpha,k}-u^n+k=k-\Phi_{\alpha,k}=T_k(u^n)-\Phi_{\alpha,k}=(T_k(u^n)-\Phi_{\alpha,k})_+.$$

Thus testing against w_n in equation (2.5.6) yields

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n), w_n\rangle dxdvdt \leq \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} f^n w_n dxdvdt \implies$$

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \boxed{\langle L(u^n-\Phi_{\alpha,k}), (T_k(u^n)-\Phi_{\alpha,k})_+\rangle}_1 + \boxed{\langle L(\Phi_{\alpha,k}), (T_k(u^n)-\Phi_{\alpha,k})_+\rangle dxdvdt}_2$$

$$\boxed{-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n), u^n - T_k(u^n)\rangle dxdvdt} \leq \boxed{\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} f^n w_n dxdvdt}_4 .$$

The first boxed term equals

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n-\Phi_{\alpha,k}), (u^n-\Phi_{\alpha,k})_+\rangle dxdvdt = \int_{\mathbb{R}^d\times\mathbb{R}^d} (u^n-\Phi_{\alpha,k})_+^2/2(T) - ((g^n - T_k(g))_+^2/2dxdv$$

$$+\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} D_v(u^n-\Phi_{\alpha,k})D_v(u^n-\Phi_{\alpha,k})_+ dxdvdt$$

and since $g^n \in L^1 \cap L^\infty$ the quantities that appear make sense. The second boxed term is bounded by

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(\Phi_{\alpha,k}), (u^n-\Phi_{\alpha,k})_+\rangle dxdvdt = \alpha \int_0^T (T_k(u)-\Phi_{\alpha,k})(u^n-\Phi_{\alpha,k}) \geq \omega(n).$$

The third boxed term equals

$$-\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n), u^n - T_k(u^n)\rangle dxdvdt$$

$$= -\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} (u^n(T) - T_k(u^n)(T))^2/2 - (g^n - T_k(g^n))^2/2 D_v u^n D_v(u^n - T_k(u^n))dxdvdt.$$

Putting everything together yields

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle L(u^n), w_n\rangle dxdvdt \geq \omega(n) + \int_{\mathbb{R}^d\times\mathbb{R}^d} (u^n-\Phi_{\alpha,k})_+^2/2(T) - ((g^n - T_k(g))_+^2/2dxdv$$

$$-\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}(u^n(T)-T_k(u^n)(T))^2/2-(g^n-T_k(g^n))^2/2dxdv$$

$$+\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_v(u^n-\Phi_{\alpha,k})D_v(u^n-\Phi_{\alpha,k})_+dxdvdt-\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_vu^nD_v(u^n-T_k(u^n))dxdvdt.$$

The first line equals

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}(u^n-\Phi_{\alpha,k})_+^2/2(T)-((g^n-T_k(g))_+^2/2-(u^n(T)-T_k(u^n)(T))^2/2-(g^n-T_k(g^n))^2/2dxdv$$

$$=\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\Big((u^n-\Phi_{\alpha,k})_+(T)-(u^n(T)-T_k(u^n)(T))((u^n-\Phi_{\alpha,k})_++(u^n-T_k(u^n))(T))\Big)dxdv$$

$$-\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}((g^n-T_k(g))_+-(g^n-T_k(g^n)))((g^n-T_k(g))_++(g^n-T_k(g^n)))$$

$$\geq -2\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}(T_k(g^n)-T_k(g))_+(g^n-T_k(g))_+dxdv=\omega(n).$$

For the last line

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_v(u^n-\Phi_{\alpha,k})D_v(u^n-\Phi_{\alpha,k})_+dxdvdt-\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_vu^nD_v(u^n-T_k(u^n))dxdvdt=$$

$$\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_v(u^n-\Phi_{\alpha,k})D_v\Big((u^n-\Phi_{\alpha,k})_+-(u^n-T_k(u^n))\Big)+D_v(u^n-\Phi_{\alpha,k})D_v(u^n-T_k(u^n))_+dxdvdt$$

$$-\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_vu^nD_v(u^n-T_k(u^n))dxdvdt$$

$$=\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_v(u^n-\Phi_{\alpha,k})D_v(T_k(u^n)-\Phi_{\alpha,k})_+dxdvdt-\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}D_v\Phi_{\alpha,k}D_v(u^n-T_k(u^n))dxdvdt$$

$$=\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_v(u^n-\Phi_{\alpha,k})_+|^2+D_v(u^n-T_k(u^n))D_v\Big((T_k(u^n)-\Phi_{\alpha,k})_+-\Phi_{\alpha,k}\Big)dxdvdt$$

$$=\int_0^T\int_{\mathbb{R}^d\times\mathbb{R}^d}|D_v(T_k(u^n)-\Phi_{\alpha,k})_+|^2dxdvdt-2\int_0^T\int_{u^n>k}D_v(u^n)D_v\Big(\Phi_{\alpha,k}\Big)dxdvdt,$$

where in the last equality we used that $D_v(u^n-T_k(u^n))=D_vu^n\chi_{u^n>k}$ and $0\leq\Phi_{\alpha,k}\leq k$. Finally, we clearly have that

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} f^n w_n dxdvdt \leq \omega(n,\alpha).$$

Hence, putting everything together

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(T_k(u^n) - \Phi_{\alpha,k})_+|^2 dxdvdt \leq 2\int_0^T \int_{u^n>k} D_v(u^n) D_v\Big(\Phi_{\alpha,k}\Big) dxdvdt + \omega(n,\alpha).$$

Since $D_v u^n \to D_v u$ weakly in L^2 while $\chi_{u^n>k}\Phi_{\alpha,k} \to \chi_{u>k}\Phi_{\alpha,k}$ strongly in L^2 (here is where the discussion in the beginning of the proof is relevant) we may use Fatous Lemma which yields

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(T_k(u) - \Phi_{\alpha,k})_+|^2 dxdvdt \leq 2\int_0^T \int_{u>k} D_v(u) D_v\Big(\Phi_{\alpha,k}\Big) dxdvdt + \omega(\alpha).$$

Furthermore,

$$\|D_v\Phi_{\alpha,k}\|_2 \leq \|D_v(T_k(u) - \Phi_{\alpha,k})_+\|_2 + \|D_v(T_k(u) - \Phi_{\alpha,k})_-\|_2 + \|D_v T_k(u)\|_2 \leq C,$$

for some $C > 0$ independent of α (due to $\omega(\alpha) \to 0$ as $\alpha \to \infty$). Therefore, we may assume WLOG that $D_v\Phi_\alpha \to D_v T_k(u)$ weakly in L^2. Thus, taking the limit as $\alpha \to \infty$ we find that

$$\limsup_{\alpha\to\infty} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(T_k(u) - \Phi_{\alpha,k})_+|^2 dxdvdt \leq$$

$$\lim_{\alpha\to\infty}\Big(2\int_0^T \int_{u>k} D_v(u) D_v\Big(\Phi_{\alpha,k}\Big) dxdvdt + \omega(\alpha)\Big) = 2\int_0^T \int_{u>k} D_v(u) D_v\Big(D_v T_k(u)\Big) dxdvdt = 0.$$

Now that we have $D_v\Phi_{\alpha,k} \to D_v T_k(u)$ strongly in L^2, we may conclude since by the previous estimates

$$\int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |D_v(T_k(u) - \Phi_{\alpha,k})_+|^2 dxdvdt \leq \omega(n,\alpha).$$

□

We conclude this subsection with a sketch of the proof for the upper bound in Theorem 4. We recall the Fractional Gagliardo-Niremberg inequality.

Proposition 10. *(Fractional Gagliardo-Niremberg inequality). Let* $z \in H^s(\mathbb{R}^d \times \mathbb{R}^d)$, *where* $s > 0$. *If* $\theta \in (0,1)$ $p \in (1,\infty)$ *are such that*

$$\theta\Big(\frac{1}{2} - \frac{s}{d}\Big) + \frac{1-\theta}{2} = \frac{1}{p} \iff \frac{1}{p} = \frac{1}{2} - \frac{\theta s}{d},$$

then

$$\|z\|_p \leq C\|D^s z\|_2^{\theta}\|z\|_2^{1-\theta},$$

where $D^s z_a = (D_v^s z_a, D_x^s z_a)$.

Corollary 1. *Let* $z \in L^2((0,T); H^s(\mathbb{R}^d \times \mathbb{R}^d))$. *Then, for* $p = 2(1 + \frac{2s}{d})$ *and* $\theta p = 2$, *we have*

$$\Big(\int_0^T \|z(t)\|_p^p dt\Big)^{1/p} \leq \sup_{t\in[0,T]} \|z(t)\|_2^{1-\theta}\|D^s z\|_2^{2/p} = \sup_{t\in[0,T]} \|z(t)\|_2^{1-\theta}\|D^s z\|^{\theta}_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)}.$$

Proposition 11. *Let* $b \in L^\infty([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$ *and* m_0 *a density which satisfies 2.2. Furthermore, let* $m \in C([0,T]; L^2(\mathbb{R}^d \times \mathbb{R}^d))$ *be the distributional solution to*

$$\begin{cases} \partial_t m - \Delta_v m + v \cdot D_x m - div_v(mb) = 0 \text{ in } (0,T) \times \mathbb{R}^d \times \mathbb{R}^d, \\ m(0,x,v) = m_0(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d. \end{cases} \tag{2.5.7}$$

Then, there exists a $C_0 = C_0(\|b\|_\infty, T, \|m_0\|_2, \|m_0\|_\infty) > 0$, *such that*

$$\|m\|_\infty \leq C_0.$$

Proof. The proof follows the work of F. Golse, C. Imbert, C. Mouhot and A. Vasseur in [211]. To simplify the notation we define the operator

$$\mathcal{L}^* m := \partial_t m - \Delta_v m + v \cdot D_x m - \mathrm{div}_v(mb).$$

Furthermore, to reduce the technical steps we make the following reduction. By linearity it is enough to show the result in the case of $\|m_0\|_\infty \leq 1$. Moreover, we assume that b is smooth with compact support, since the general case may be handled by approximation. We note that once b is smooth and compactly supported, the density m is bounded above, however this bound depends on $\|\mathrm{div}_v(b)\|_\infty$. Nonetheless, at the level of smooth b the functions m, m^2 are integrable. For $\alpha > 1 \geq \|m_0\|_\infty$ we set $m_\alpha := (m-\alpha)_+$. Then, we have that m_α is a subsolution of

$$\mathcal{L}^* m_\alpha - (1+\alpha)\mathbf{1}_{m>\alpha}\mathrm{div}_v(b) \leq 0, m_\alpha(0) = 0. \tag{2.5.8}$$

Moreover, for technical reasons we will also require the function m_α^2, which is a subsolution of

$$\partial_t m_\alpha^2 - \Delta_v m_\alpha^2 - v \cdot D_x m_\alpha^2 - \mathrm{div}_v(m_\alpha^2 b) - m_\alpha^2 \mathrm{div}_v(b) - 2\alpha m_\alpha \mathrm{div}_v(b) \leq 0, \;\; m_\alpha^2(0) = 0, \tag{2.5.9}$$

or equivalently

$$\partial_t m_\alpha^2 - \Delta_v m_\alpha^2 - v \cdot D_x m_\alpha^2 - 2\mathrm{div}_v(m_\alpha^2 b) + D_v(m_\alpha^2)\cdot b - 2\alpha \mathrm{div}_v(m_\alpha b) + 2\alpha Dv(m_\alpha)\cdot b \leq 0, m_\alpha^2(0) = 0. \tag{2.5.10}$$

The typical energy estimates required in the De Giorgi argument for improvement of integrability, are not suitable for this setting. Namely testing against m_a^2 in 2.5.9, only yields bounds on $D_v m_a^2$. To obtain an increase in integrability we first look at the solution w_α of

$$\partial_t w_\alpha - \Delta_v w_\alpha - v \cdot D_x w_\alpha - \mathrm{div}_v(m_\alpha^2 b) - m_\alpha^2 \mathrm{div}_v(b) - 2\alpha m_\alpha \mathrm{div}_v(b) = 0, \;\; w_\alpha(0) = 0, \tag{2.5.11}$$

and we note that $w_a \geq m_a^2 \geq 0$. Testing against w_a in (2.5.11) yields by Grönwall

$$\begin{aligned} &\sup_{t\in[0,T]} \|w_\alpha(t)\|_2^2 + \|D_v w_\alpha\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \\ &\leq C(\|m_\alpha\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} + \|m_\alpha^2\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} + \|D_v m_\alpha\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} + \|D_v(m_\alpha^2)\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)}). \end{aligned} \tag{2.5.12}$$

For the estimates on m_α^2 we test (2.5.8) against m_α^2 and integrate in space to obtain by Grönwall

$$\sup_{t\in[0,T]}\int |m_\alpha(t)|^4 + \int_0^T \int |D_v m_\alpha^2|^2 \le C(\int_0^T \int |m_\alpha|^2 + \int_0^T \int |D_v m_\alpha|^2). \tag{2.5.13}$$

We need an estimate for $\int_0^T \int |D_v m_\alpha|^2$, so we test against m_α in (2.5.8) and integrate in space to obtain by Grönwall

$$\sup_{t\in[0,T]} \|m_\alpha(t)\|_2^2 + \int_0^T \int |D_v m_\alpha|^2 \le C \int_0^T |\{m_\alpha(t) > 0\}|. \tag{2.5.14}$$

Using estimates (2.5.14),(2.5.13) on (2.5.12) yields

$$\sup_{t\in[0,T]} \|w_\alpha(t)\|_2^2 + \int_0^{T+2} \int |D_v w_\alpha|^2 \le C \int_0^T |\{m_\alpha(t) > 0\}|dt. \tag{2.5.15}$$

From the above and Theorem 19, we obtain

$$\|D^s w_\alpha\|^2_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \le C \int_0^T |\{m_\alpha(t) > 0\}|dt. \tag{2.5.16}$$

From (2.5.16) and Corollary 1, we obtain

$$\|w_\alpha\|_{L^p([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \le C\|D^s w_\alpha\|^\theta_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \sup_{t\in[0,T]} \|w_\alpha(t)\|^{1-\theta}_{L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \le C\|D^s w_\alpha\|_2^\theta \sup_{t\in[0,T]} \|w_\alpha(t)\|_2^{1-\theta}$$

from (2.5.15) and (2.5.16) we have

$$\|w_\alpha\|_{L^p([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \le C \int_0^T |\{m_\alpha(t) > 0\}|. \tag{2.5.17}$$

We may now setup the De-Giorgi iteration. For $k \in \mathbb{N}$, let $\alpha_k = (2 + \frac{1}{2^{k-1}})$ and $m_k := m_{\alpha_k}$. Since

$$|\{m_k(t) > 0\}| = |\{m_{k-1}(t) > \frac{1}{2^k}\}| \le 16^k \int |m_{k-1}(t)|^4, \tag{2.5.18}$$

if we define $U_k := \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} |m_k|^4 dxdvdt$, and use estimate (2.5.18) in (2.5.17), we obtain

$$\|w_k\|_{L^p([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)} \leq C16^k U_{k-1}. \tag{2.5.19}$$

Recall that $m_\alpha^2 \leq w_\alpha$, thus from (2.5.19) we have

$$\|m_\alpha^2\|_p \leq \|w_\alpha\|_p \leq C16^k U_{k-1}.$$

Therefore,

$$U_k = \int_0^T \int |m_k|^4 dxdvdt = \|m_k^2\|_2^2 \leq C\|w_k\|_p^2 |\{m_k > 0\}|^\epsilon \leq C16^k U_{k-1}^{1+\epsilon},$$

for some $\epsilon = \epsilon(p) > 0$ and the result follows. □

2.5.2 Prerequisites

We rely on the following minor modifications of three results from [59]. We modify these Theorems slightly, to be used for a finite time interval $[0,T]$.

Theorem 17. *(Theorem 1.5,[59]) Let* $f, g \in L^2([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$*,* $D_v f \in L^2(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d)$ *and* $f_0 \in L^2(\mathbb{R}^d\times\mathbb{R}^d)$*, such that*

$$\begin{cases} \partial_t f - v\cdot D_x f - \Delta_v f = g \text{ in } [0,T]\times\mathbb{R}^d\times\mathbb{R}^d, \\ f(0,x,v) = f_0(x,v) \text{ in } \mathbb{R}^d\times\mathbb{R}^d. \end{cases}$$

Then, there exists a dimensional constant $C > 0$*, such that*

$$\|\partial_t f - v\cdot D_x f\|_2 + \|\Delta_v f\|_2 \leq \frac{C}{t}\Big(\|g\|_2 + \|f_0\|_2\Big).$$

Theorem 18. *(Theorem 1.3, [59]) Assume that* $f, g, g_0 \in L^p([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*, with* $D_v f \in L^p([0,T]\times\mathbb{R}^d\times\mathbb{R}^d)$*,* $(1+|v|^2)g \in L^p(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d)$*,* $(1+|v|)g_0 \in L^p(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d)$ *and* $f_0 \in L^p(\mathbb{R}^d\times\mathbb{R}^d)$ *for some* $p \in (1,\infty)$*, such that they solve*

$$\begin{cases} \partial_t f - v \cdot D_x f = div_v(g) + g_0 \text{ in } (0,T] \times \mathbb{R}^d \times \mathbb{R}^d, \\ f(0,x,v) = f_0(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases}$$

in the distributional sense. Then, there exists a constant $C > 0$*, such that*

$$\|D_x^{1/3} f\|_p + \|D_t^{1/3} f\|_p \le C(\|f\|_p + \|D_v f\|_p + \|(1+|v|^2)g\|_p + \|(1+|v|)g_0\|_p + \|f_0\|_p).$$

Theorem 19. *(Theorem 2.1, [59]) Assume that* $f, g, g_0 \in L^p([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*, with* $D_v f \in L^p([0,T] \times \mathbb{R}^d \times \mathbb{R}^d)$*, and* $f_0 \in L^p(\mathbb{R}^d \times \mathbb{R}^d)$ *for some* $p \in (1,\infty)$*, such that they solve*

$$\begin{cases} \partial_t f - v \cdot D_x f = div_v(g) + g_0 \text{ in } (0,T] \times \mathbb{R}^d \times \mathbb{R}^d, \\ f(0,x,v) = f_0(x,v) \text{ in } \mathbb{R}^d \times \mathbb{R}^d, \end{cases}$$

in the distributional sense . Then, there exists a constant $C > 0$*, such that*

$$\|D_x^{1/3} f\|_p \le C(\|D_v f\|_p + \|f_0\|_p + \|g\|_p + \|g_0\|_p),$$

where $\alpha, \alpha' \in (0,1)$ *and depend only on the dimension d.*

CHAPTER 3

FIRST-ORDER MEAN FIELD GAMES SYSTEM

3.1 Introduction

The work presented in this chapter refers to the joint work with S. Munoz in [312]. The main purpose of this chapter is to establish that, under very general conditions, the solutions to the one-dimensional first-order mean field games system with local coupling are smooth, and to fully characterize their long time behavior. Specifically, we study the regularity of the solutions to standard MFG with a prescribed terminal condition,

$$\begin{cases} -u_t(x,t) + H(u_x(x,t), m(x,t)) = 0 & (x,t) \in Q_T = \mathbb{T} \times (0,T), \\ m_t(x,t) - (m(x,t) H_p(u_x(x,t), m(x,t)))_x = 0 & (x,t) \in Q_T, \\ m(x,0) = m_0(x),\ u(x,T) = g(m(x,T)) & x \in \mathbb{T}, \end{cases} \tag{MFG}$$

as well as to the so-called planning problem with a prescribed terminal density,

$$\begin{cases} -u_t(x,t) + H(u_x(x,t), m(x,t)) = 0 & (x,t) \in Q_T, \\ m_t(x,t) - (m(x,t) H_p(u_x(x,t), m(x,t)))_x = 0 & (x,t) \in Q_T, \\ m(x,0) = m_0(x),\ m(x,T) = m_T(x) & x \in \mathbb{T}, \end{cases} \tag{MFGP}$$

where $\mathbb{T}$ denotes the 1-dimensional torus, $-H(p,m) : \mathbb{R} \times (0,\infty) \to \mathbb{R}$ and $g(m) : (0,\infty) \to \mathbb{R}$ are strictly increasing in m, H has super-linear growth in p, and $m_0, m_T : \mathbb{T} \to [0,+\infty)$ are probability densities. We also show convergence of the solutions to each of these problems, as $T \to \infty$, to the solution of the infinite time horizon MFG system,

$$\begin{cases} -v_t(x,t) + \lambda + H(v_x(x,t),\mu(x,t)) = 0 & (x,t) \in \mathbb{T} \times (0,\infty), \\ \mu_t(x,t) - (\mu(x,t)H_p(v_x(x,t),\mu(x,t)))_x = 0 & (x,t) \in \mathbb{T} \times (0,\infty), \\ \mu(x,0) = m_0(x) & x \in \mathbb{T}, \end{cases} \tag{MFGL}$$

where $\lambda = -H(0,1)$.

Classical solutions to (MFG), in arbitrary dimension, were previously obtained by the second author in [316, 317], when the initial density is bounded away from 0, and under the blow-up assumption

$$\lim_{m \to 0^+} H(p,m) = +\infty, \tag{3.1.1}$$

which, from the optimal control point of view, corresponds to placing a very strong incentive for players to occupy low-density regions and precludes the appearance of empty regions. A similar regularity result was recently obtained in [328] by A. Porretta for the case of (MFGP), when the Hamiltonian has the separated form $H(p,m) \equiv H(p) - f(m)$, and the terminal density m_T is also bounded away from 0.

Our first contribution is the following theorem, which shows that, in the one-dimensional problem, assumption (3.1.1) can be completely removed. We refer to Section 3.2 for assumptions (M), (H) (G), (E), (W), and (L), and to the notation subsection for the definition of the function spaces mentioned below.

Theorem 3.1.1. *Let* $0 < \alpha < 1$*, and assume that* (M), (H), (G)*, and* (E) *hold. Then the following statements hold:*

(i) There exists a classical solution $(u,m) \in C^{3,\alpha}(\overline{Q_T}) \times C^{2,\alpha}(\overline{Q_T})$ *to* (MFGP)*. The function m is unique, and u is unique up to a constant.*

(ii) There exists a unique classical solution $(u,m) \in C^{3,\alpha}(\overline{Q_T}) \times C^{2,\alpha}(\overline{Q_T})$ *to* (MFG).

Our second result establishes interior smoothness of the solutions when, besides removing the assumption (3.1.1), one also weakens the lower bound assumptions for given densities m_0 and m_T, replacing the latter with the integrability conditions

$$\frac{1}{m_0^\kappa} \in L^1(\mathbb{T}), \quad \frac{1}{m_T^\kappa} \in L^1(\mathbb{T}) \text{ for some } \kappa > 0. \tag{3.1.2}$$

We observe that, in particular, (3.1.2) allows the initial density to vanish in a set of measure zero. In spite of this fact, our result also shows that m becomes strictly positive instantly after the initial time. Moreover, in the case of (MFG), the density remains bounded below, and the solution remains smooth up to and including $t = T$. We refer to Section 3.6 for the definition of a weak solution.

Theorem 3.1.2. *Let* $0 < \alpha < 1$*, and assume that* (W), (H) (G)*, and* (E) *hold. Then the following statements hold:*

(i) There exists a weak solution

$$(u, m) \in (BV(Q_T) \cap L^\infty(Q_T)) \times (C([0, T], H^{-1}(\mathbb{T})) \cap L^\infty_+(Q_T))$$

to (MFGP)*. Moreover,* $(u, m) \in C^{3,\alpha}_{\text{loc}}(Q_T) \times C^{2,\alpha}_{\text{loc}}(Q_T)$ *and* $m > 0$ *in* $(0, T)$*. The function* m *is unique, and* u *is unique up to a constant.*

(ii) Assume, further, that the function H *satisfies, for each* $(p, m) \in \mathbb{R} \times (0, \infty)$*,*

$$H_p(p, m)p \geq 0. \tag{3.1.3}$$

Then there exists a unique weak solution

$$(u, m) \in (BV(Q_T) \cap L^\infty(Q_T)) \times (C([0, T], H^{-1}(\mathbb{T})) \cap L^\infty_+(Q_T))$$

to (MFG). *Moreover,* $(u,m) \in C^{3,\alpha}_{\mathrm{loc}}(\mathbb{T}\times(0,T])\times C^{2,\alpha}_{\mathrm{loc}}(\mathbb{T}\times(0,T])$*, and* $m>0$ *in* $(0,T]$.

Concerning the long time behavior of (3.1.1), it was shown by P. Cardaliaguet and P.J. Graber in [81, Thm 5.1] that the rescaled solution $(x,s)\mapsto u(x,sT)/T$ converges, in a certain space $L^p(\mathbb{T}\times(\delta,1))$, to the map $\lambda(1-s)$, while the rescaled density $(x,s)\mapsto m(x,sT)$ converges in $L^p(\mathbb{T}\times(0,1))$ to the invariant measure $\mu\equiv 1$. Our third result shows that, when the marginals are strictly positive, a much stronger statement holds. That is, the solutions satisfy the turnpike property with an exponential rate of convergence, and the limit as $T\to\infty$ of the pair $(u(t)-\lambda(T-t),m(t))$ can be fully characterized as the solution to (MFGL). We emphasize that this is a convergence result *at the original time scale* (cf. [86, Thm 2.6, Thm. 5.1], [122, Thm 4.1, Thm. 5.3]).

Theorem 3.1.3. *Assume that* (M), (H), (G), (E), *and* (L)*, hold, and let* $T>1$. *Assume that* (u^T,m^T) *is either the solution to* (MFG)*, or the solution to* (MFGP) *that satisfies* $\int_{\mathbb{T}} v^T(\cdot,\frac{T}{2})=0$*, where*

$$v^T(x,t):=u^T(x,t)-\lambda(T-t).$$

Then the following holds:

(i) There exist constants $C,\omega>0$*, independent of* T*, such that*

$$\|m^T(t)-1\|_{L^\infty(\mathbb{T})}+\|u^T_x(t)\|_{L^\infty(\mathbb{T})}\le C(e^{-\omega t}+e^{-\omega(T-t)}),\quad t\in[0,T].$$

Moreover, if (u^T,m^T) *solves* (MFG)*, and* (3.1.3) *holds, we have*

$$\|m^T(t)-1\|_{L^\infty(\mathbb{T})}+\|u^T_x(t)\|_{L^\infty(\mathbb{T})}\le Ce^{-\omega t},\quad t\in[0,T].$$

(ii) There exist functions (v,μ) *such that, for each* $T_0>0$*,*

$$v^T\to v \text{ in } C^{3,\alpha}(\mathbb{T}\times[0,T_0]) \text{ as } T\to\infty,$$

and

$$m^T \to \mu \text{ in } C^{2,\alpha}(\mathbb{T}\times[0,T_0]) \text{ as } T\to\infty.$$

Moreover, one has

$$\lim_{t\to\infty} v(\cdot,t) = c, \quad \lim_{t\to\infty}\mu(\cdot,t) = 1 \text{ uniformly in } \mathbb{T}, \tag{3.1.4}$$

where

$$c = \begin{cases} g(1) & \text{if } (u^T, m^T) \text{ solves (MFG)}, \\ 0 & \text{if } (u^T, m^T) \text{ solves (MFGP)}. \end{cases}$$

Finally, (v,μ) *is the unique classical solution to* (MFGL) *satisfying* (3.1.4) *and*

$$v \in W^{1,\infty}(\mathbb{T}\times(0,\infty)), \quad \mu^{-1}\in L^\infty(\mathbb{T}\times(0,\infty)),$$
$$\mu - 1 \in L^1(\mathbb{T}\times(0,\infty)) \cap L^\infty(\mathbb{T}\times(0,\infty)). \tag{3.1.5}$$

In particular, since the Hamiltonian $H(p,m)$ is non-separated, our results yield well-posedness and regularity of MFG systems with congestion, such as

$$\begin{cases} -u_t + \frac{|u_x|^2}{2(m+c_0)^\alpha} = f(m) & \text{in } Q_T, \\ m_t - \left(\frac{m}{(m+c_0)^\alpha}u_x\right)_x = 0 & \text{in } Q_T, \end{cases} \tag{3.1.6}$$

where $0<\alpha<2$, $c_0 \geq 0$, and $f'>0$. Some of the key techniques used in [316, 317, 328], as well as in the present work, were developed by P.-L. Lions in his lectures at Collège de France [297], where he obtained several a priori estimates for the solutions to (MFGP), in the special case of a separated, quadratic Hamiltonian. The most important of these is the observation that the problems (MFG) and (MFGP) can be transformed into a single quasilinear elliptic equation in u

after eliminating the variable m. Indeed, if one defines H^{-1} by

$$m = H^{-1}(p, H(p, m)),$$

then $m = H^{-1}(u_x, u_t)$ and the problem becomes

$$\begin{cases} Qu := -\mathrm{Tr}(A(Du)D^2u) = 0 & \text{in } Q_T, \\ Nu := B(x, t, u, Du) = 0 & \text{on } \partial Q_T, \end{cases} \tag{Q}$$

where $Du = (u_x, u_t)$ and, for $(x, z, p, s) \in \mathbb{T} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$,

$$A(p, s) = \left(H_p + \frac{1}{2}mH_{mp}, -1\right) \otimes \left(H_p + \frac{1}{2}mH_{mp}, -1\right) - \begin{pmatrix} \frac{1}{4}m^2H_{mp}^2 + mH_mH_{pp} & 0 \\ 0 & 0 \end{pmatrix}, \tag{Q1}$$

$$B(x, 0, z, p, s) = -\, s + H(p, m_0(x)), \tag{B1}$$

and

$$B(x, T, z, p, s) = \begin{cases} s - H(p, g^{-1}(z)) & \text{in the case of (MFG)} \\ s - H(p, m_T(x)) & \text{in the case of (MFGP).} \end{cases} \tag{B2}$$

The condition for ellipticity, that is, for the matrix A to be positive, is

$$-4mH_mH_{pp} > m^2H_{mp}^2, \tag{3.1.7}$$

which is also the well-known condition for uniqueness to (MFG) that follows from the Lasry-Lions monotonicity method (see, for instance, Lions, Souganidis [301]). We remark from (3.1.7) that, in particular, the strict positivity of the density is crucial for the regularizing properties of the system. The lower bounds on m obtained in Corollary 3.3.2 and Proposition 3.6.3 both heavily rely on the one-dimensionality assumption, and this is the main obstacle to generalizing our results to higher dimensions. Indeed, in dimensions $d > 1$, it remains an open question whether the existence of smooth solutions to local first order MFG systems can still be established if one removes or significantly weakens (3.1.1), or if m_0 is not assumed to be bounded away from 0. Even for $d = 1$, it is still unknown whether one can allow m_0 or m_T to vanish in a set of positive measure.

Outline of Chapter 3

Section 3.2 contains all the assumptions that will be in place about the Hamiltonian H, as well as the initial and terminal data. In Section 3.3, we establish an integral displacement convexity formula (see Proposition 3.3.1) that will allow us to bound the density m in terms of its initial and terminal values. Section 3.4 contains the necessary a priori estimates that are needed to prove the existence of classical solutions. In particular, we obtain, in Section 3.4.1, estimates for an ϵ–approximation of (MFGP) via standard MFG systems with a terminal condition of the type $u(\cdot, T) = g(\cdot, m(\cdot, T))$, which we require to prove existence for (MFGP). Finally, we provide a counterexample to existence of solutions to (MFG) when the terminal cost function g is also allowed to depend on the space variable (see Proposition 3.4.5). In Sections 3.5, 3.6, and 3.7, we prove our main results, Theorems 3.1.1, 3.1.2, and 3.1.3, respectively.

Previous literature and related questions

It is natural to ask whether our results continue to hold when the Hamiltonian H, or the terminal coupling g, are allowed to depend on the space variable x. Indeed, the previous results [316, 317, 324] on classical solutions under the blow-up assumption (3.1.1) allowed for very general assumptions on the x-dependence. However, when one removes this assumption as we did in this work, classical solutions fail to exist unless one imposes quite severe restrictions on this dependence. Even in the simple case $H(p, m, x) \equiv H(p) - f(m) - V(x)$, it was shown in [297] that one can construct examples with smooth data where even the stationary solution is non-smooth and vanishes in a set of positive measure (see also [213, 81]). However, a regularity result analogous to Theorem 3.1.1 may be established when the potential $V(x)$ is small in a suitable sense, as this represents a perturbation of the present problem (as far as we are aware, the only references exploring this matter are [213, 18]). A similar situation occurs when the terminal coupling g is still strictly monotone, but allowed to depend on x. As mentioned earlier, we construct such a counterexample in Proposition 3.4.5.

In the special case of a separated Hamiltonian, the estimates of Section 3.3 were first obtained by Gomes and Seneci in [214]. Further estimates on the density using displacement convexity were also obtained by Bakaryan, Ferreira, and Gomes in [18], and by Porretta in [328] (see also Lavenant, Santambrogio [289]). Weak solutions, as defined in Section 3.6, have been widely studied for both (MFG) (see [74, 81, 82, 87, 316]) and (MFGP) (see [222, 328, 319]). For classical solutions in the time-independent case we refer to Evans [170] and Gomes, Mitake [213]. Concerning the study of the long time behavior of solutions, specifically the second part of Theorem 3.1.3, we follow the program developed by Cirant and Porretta in [122], where a similar analysis was performed for second-order MFG systems, and, unlike the earlier work [86], does not involve the use of the master equation (see also [85, 327]). Finally, on the matter of regularity of solutions to first order MFG systems, it is also worth comparing with the recent results of [311]. Unlike the present work, that paper deals with the different setting of MFG systems with a non-local coupling (i.e. when $H : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$), under the so-called displacement monotonicity assumption on the data. The authors prove existence of a solution (u, m), with u being a classical (rather than merely viscosity) solution to the HJ equation, $C^{1,1}$ in the space variable, and m being a distributional solution to the continuity equation.

Notation

Let $d, k \in \mathbb{N}$. For $T > 0$, we denote by $Q_T := \mathbb{T} \times (0, T)$, $\overline{Q_T} := \mathbb{T} \times [0, T]$ and $\partial Q_T := \mathbb{T} \times \{0, T\}$. For $\alpha \in (0, 1]$ and $\Omega \subset \mathbb{R}^d$ we denote by $C^{k,a}(\Omega)$, the standard space of k times differentiable scalar functions with α-Hölder continuous k^{th} order derivatives, with the usual norm. Furthermore, we denote by $C^{k,\alpha}_{\text{loc}}(\Omega)$ the functions u that belong to $C^{k,\alpha}(K)$, for all compact sets $K \subset \Omega$. For functions $u : \mathbb{T} \times [0,T] \to \mathbb{R}$, we denote by $\operatorname{osc} u := \max_{(x,t)\in\mathbb{T}\times[0,T]} u(x,t) - \min_{(x,t)\in\mathbb{T}\times[0,T]} u(x,t)$, $Du(x,t) := (u_x(x,t), u_t(x,t))$. We denote by $H^{-1}(\mathbb{T})$ the dual space of the Sobolev space $H^1(\mathbb{T})$, and by $C^{0,\alpha}([0,T]; H^{-1}(\mathbb{T}^d))$ the space of $H^{-1}(\mathbb{T}^d)$–valued functions that are α–Hölder continuous. We write $C = C(K_1, K_2, \ldots, K_M)$ for a positive constant C depending monotonically on the non-negative quantities $K_1, \ldots, K_M$. $\mathrm{BV}(Q_T)$ denotes the space of functions of bounded variation, and $L^\infty_+(Q_T)$ consists of the functions $m \in L^\infty(Q_T)$ such that $m \geq 0$ a.e. in Q_T.

3.2 Assumptions

In what follows, C_0 and γ, α are positive constants, with $\gamma > 1$, and $0 < \alpha < 1$. Moreover, $\overline{C} : (0,\infty) \to [1,\infty)$ is a continuous, strictly positive function. We note that $\overline{C} = \overline{C}(m)$ should be interpreted simply as a positive bound that may blow up both as $m \downarrow 0$ and as $m \uparrow \infty$. Except when explicitly stated, assumptions (M), (H), (G), and (E) will be in place throughout the chapter.

(M) (Assumptions on m_0 and m_T for classical solutions) The given functions m_0 and m_T satisfy

$$m_0, m_T \in C^{2,\alpha}(\mathbb{T}),\ m_0, m_T > 0, \text{ and } \int_{\mathbb{T}} m_0 = \int_{\mathbb{T}} m_T = 1. \tag{M1}$$

(H) (Assumptions on H) The functions H, H_p, and H_{pp} are in $C^4(\mathbb{R} \times (0,\infty))$, and $H_m < 0$. Moreover, for $(p,m) \in \mathbb{R} \times (0,\infty)$,

$$\frac{1}{C_0}(1+|p|)^{\gamma-2} \leq H_{pp} \leq \overline{C}(m)(1+|p|)^{\gamma-2}, \tag{H1}$$

$$pH_p \geq (1 + \frac{1}{C_0})H - \overline{C}(m), \tag{H2}$$

$$|H_{ppp}| \leq \overline{C}(m)(1+|p|)^{\gamma-3}, \tag{H3}$$

$$|H_m| \leq \overline{C}(m)(1+|p|)^{\gamma}, \tag{HM1}$$

$$m|H_{mm}| \leq -\overline{C}(m)H_m, \quad m|p||H_{mmp}| \leq -\overline{C}(m)H_m, \tag{HM2}$$

$$|H_{mpp}| \leq \overline{C}(m)(1+|p|)^{\gamma-2} \tag{HM3}$$

(G) (Assumptions on g) The function $g:(0,\infty)\to\mathbb{R}$ is four times continuously differentiable and satisfies, for all $m>0$,

$$g'(m) > 0. \tag{G1}$$

(E) (Ellipticity of the system) The function H satisfies, for $m>0$, the condition

$$-4mH_mH_{pp} \geq \left(1+\frac{1}{C_0}\right)m^2H_{mp}^2. \tag{E1}$$

(W) (Assumptions on m_0, m_T, H, and g for weak solutions) The functions m_0 and m_T satisfy, for some $\kappa>0$,

$$m_0,\ m_T \in L^{\infty}(\mathbb{T}),\ m_0,\ m_T \geq 0,\ \int_{\mathbb{T}} m_0 = \int_{\mathbb{T}} m_T = 1,\ \text{and}\ \frac{1}{m_0^{\kappa}}, \frac{1}{m_T^{\kappa}} \in L^1(\mathbb{T}), \tag{MW}$$

H satisfies, for some constant $s\in(-\kappa-1,\kappa-1)$, and for $(p,m)\in\mathbb{R}\times(0,\frac{1}{C_0})$,

$$-H_m(0,m) \leq C_0m^s,\ -H_m(p,m) \geq \frac{1}{C_0}m^s, \tag{HW}$$

and g satisfies

$$\lim_{m\to 0^+} g(m) > -\infty. \tag{GW}$$

(L) (Assumption on H for the long time behavior) The function H satisfies, for $m > 0$,

$$-4mH_m H_{pp} \geq \frac{1}{\overline{C}(m)}. \tag{HL}$$

Remark 5. *We will impose assumptions* (W) *and* (L) *only in the sections discussing weak solutions and long time behavior, respectively.*

Assumption (W) *significantly weakens the positivity assumption on m but, in exchange, requires a more precise control on the behavior of H and g near small densities.*

On the other hand, in the context of our result on the long time behavior of strictly positive classical solutions, no such control near small (or large) densities is needed. However, a different issue arises here: the gradient bounds used throughout the rest of the chapter may degenerate as $T \to \infty$. *Indeed, with* (E) *in place, we could rephrase assumption* (L) *as the requirement that the eigenvalues of the elliptic operator* (Q1) *remain bounded below as* $|p| \to \infty$, *locally uniformly in* $m \in (0, \infty)$[1]. *This will allow us to obtain gradient bounds that are uniform in T (see Lemma 3.7.1, where this assumption is used). For example, for the case of a separated Hamiltonian* $H(p,m) \equiv H(p) - f(m)$, (L) *simply reduces to the assumption that H is uniformly convex, which follows automatically from* (H1) *for* $\gamma \geq 2$.

3.3 Displacement convexity and estimates on the density

To obtain estimates for the density at interior times, we will prove an integral formula which, in particular, implies that the quantity

$$\int_{\mathbb{T}} h(m(x,\cdot))dx$$

is a convex function in $[0,T]$ whenever h is convex, provided that (3.1.7) holds.

1. The factor $4m$ in (HL) is, of course, inconsequential, because it is a positive function of m. It is only included to emphasize the comparison with (E).

Proposition 3.3.1. *Let* $(u,m) \in C^2(\overline{Q}_T) \times C^1(\overline{Q}_T)$ *be a classical solution to*

$$\begin{cases} -u_t + H(u_x, m) = 0, & \text{in } Q_T \\ m_t - (mH_p(u_x, m))_x = 0, & \text{in } Q_T \\ m(\cdot, 0) = m_0, & \text{in } \mathbb{T}, \end{cases} \tag{3.3.1}$$

and let $h \in W^{2,\infty}(\mathbb{R})$. *Then*

$$\frac{d^2}{dt^2}\int_{\mathbb{T}} h(m(x,t))dx = \int_{\mathbb{T}} h''(m)\Big(m_t - m_x(H_p + \frac{m}{2}H_{pm})\Big)^2 dx - \int_{\mathbb{T}} h''(m)(m_x)^2\Big(\frac{m^2}{4}H_{pm}^2 + mH_{pp}H_m\Big)dx. \tag{3.3.2}$$

Moreover, there exists $C = C(C_0)$ *such that, if* $h'' > 0$,

$$\frac{d^2}{dt^2}\int_{\mathbb{T}} h(m(x,t))dx \geq \frac{1}{C}\int_{\mathbb{T}} h''(m)(-mH_mH_{pp}m_x^2 + m^2H_{pp}^2u_{xx}^2)dx. \tag{3.3.3}$$

Proof. Let $\widetilde{h} : \mathbb{R} \to \mathbb{R}$, be a smooth function. Since m satisfies the continuity equation, the following holds for each $t \in [0,T]$:

$$\int_{\mathbb{T}}\Big(m_t(x,t) - (m(x,t)H_p(u_x, m(x,t)))_x\Big)\Big(\partial_t\widetilde{h}(m(x,t)) - (\widetilde{h}(m(x,t))H_p(u_x, m(x,t)))_x\Big)dx = 0. \tag{3.3.4}$$

Expanding equation (3.3.4), we obtain

$$
\begin{aligned}
0 = \int_{\mathbb{T}} &(m_t - m_x(H_p + mH_{pm}) - mH_{pp}u_{xx})(\widetilde{h}'(m)m_t - m_x(\widetilde{h}'(m)H_p + \widetilde{h}(m)H_{pm}) - \widetilde{h}(m)H_{pp}u_{xx})dx \\
= \int_{\mathbb{T}} &\widetilde{h}'(m)(m_t)^2 - m_t m_x\Big[2\widetilde{h}'(m)H_p + \Big(\widetilde{h}'(m)m + \widetilde{h}(m)\Big)H_{pm}\Big] \\
&+ m_x H_{pp}u_{xx}\Big[H_p\Big(\widetilde{h}'(m)m + \widetilde{h}(m)\Big) + 2\widetilde{h}(m)mH_{pm}\Big] \\
&+ m_x^2\Big[\Big(H_p + mH_{pm}\Big)\Big(\widetilde{h}'(m)H_p + \widetilde{h}(m)H_{pm}\Big)\Big] \\
&- m_t H_{pp}u_{xx}\Big[\widetilde{h}'(m)m + \widetilde{h}(m)\Big] \\
&+ \widetilde{h}(m)m\Big(H_{pp}u_{xx}\Big)^2 dx = A_1 - A_2 + A_3 + A_4 - A_5 + A_6.
\end{aligned}
$$

We split term A_3 as follows

$$
A_3 = \int_{\mathbb{T}} m_x H_{pp} H_p u_{xx}\Big(\widetilde{h}'(m)m + \widetilde{h}(m)\Big)dx + 2\int_{\mathbb{T}} \widetilde{h}(m) m_x m H_{pm} H_{pp} u_{xx} dx = A_{3.1} + A_{3.2}.
$$

From the continuity equation, we have that

$$
mH_{pp}u_{xx} = m_t - m_x(H_p + mH_{pm}).
$$

Hence, terms $A_{3.2}$ and A_6 can be written as

$$
A_{3.2} = 2\int_{\mathbb{T}} m_t m_x H_{pm}\widetilde{h}(m)dx - 2\int_{\mathbb{T}} (m_x)^2 H_{pm}\Big(\widetilde{h}(m)H_p + m\widetilde{h}(m)H_{pm}\Big)dx = A_{3.2.1} - A_{3.2.2}
$$

$$
A_6 = \int_{\mathbb{T}} \frac{\widetilde{h}(m)}{m}\Big[m_t - m_x\Big(H_p + mH_{pm}\Big)\Big]^2 dx
$$

$$
= \int_{\mathbb{T}} \frac{\widetilde{h}(m)}{m}(m_t)^2 - 2\frac{\widetilde{h}(m)}{m} m_t m_x\Big(H_p + mH_{pm}\Big) + \frac{\widetilde{h}(m)}{m}(m_x)^2\Big(H_p + mH_{pm}\Big)^2 dx = A_{6.1} - A_{6.2} + A_{6.3}.
$$

From the Hamilton-Jacobi (HJ for short) equation, we have that

$$H_p u_{xx} = u_{xt} - H_m m_x.$$

Therefore, $A_{3.1}$ may be written as

$$A_{3.1} = \int_{\mathbb{T}} m_x H_{pp} u_{xt}\Big(\widetilde{h}'(m)m + \widetilde{h}(m)\Big)dx - \int_{\mathbb{T}} (m_x)^2 H_{pp} H_m\Big(\widetilde{h}'(m)m + \widetilde{h}(m)\Big)dx = A_{3.1.1} - A_{3.1.2}$$

We now begin by grouping together terms A_5, and $A_{3.1.1}$, which yields, for $L(m) = \widetilde{h}(m)m,\ L'(m) = \widetilde{h}(m) + m\widetilde{h}'(m)$,

$$\begin{aligned}
A_{3.1.1} - A_5 &= \int_{\mathbb{T}} m_x\Big(\widetilde{h}(m) + m\widetilde{h}'(m)\Big)H_{pp}u_{xt} - \Big(\widetilde{h}(m) + m\widetilde{h}'(m)\Big)m_t H_{pp} u_{xx} dx \\
&= \int_{\mathbb{T}} -\partial_t(L(m))(H_p)_x + L'(m)m_t H_{pm} m_x + (L(m))_x \partial_t(H_p) - L'(m) m_x m_t H_{pm} dx \\
&= \int_{\mathbb{T}} \partial_t((L(m))_x)H_p + (L(m))_x \partial_t(H_p) dx = \frac{d}{dt}\int_{\mathbb{T}} (L(m))_x H_p dx.
\end{aligned}$$

Next, we group together all the terms with $m_t m_x$ factor, namely A_2, $A_{3.2.1}$, and $A_{6.2}$, which yields

$$-A_2 + A_{3.2.1} - A_{6.2} = -\int_{\mathbb{T}} 2m_t m_x \left(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\right)\left(H_p + \frac{m}{2}H_{pm}\right)dx.$$

Collecting the terms involving $(m_t)^2$, namely terms A_1 and $A_{6.1}$, we obtain

$$A_1 + A_{6.1} = \int_{\mathbb{T}} (m_t)^2 \left(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\right)dx.$$

Finally, we group together the terms involving m_x^2, namely A_4, $A_{3.2.2}$, $A_{6.3}$, and $A_{3.1.2}$:

$$
\begin{aligned}
A_4 - A_{3.2.2} + A_{6.3} - A_{3.1.2} = \\
&\int_{\mathbb{T}} (m_x)^2 \Big[\Big(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\Big)\Big(H_p + \frac{m}{2}H_{pm}\Big)^2\Big] dx \\
&\quad - \int_{\mathbb{T}} (m_x)^2 \Big[\Big(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\Big)\Big(\frac{m^2}{4}H_{pm}^2 + mH_{pp}H_m\Big)\Big] dx.
\end{aligned}
$$

Thus, putting everything together, we obtain

$$
\begin{aligned}
-\frac{d}{dt}\int_{\mathbb{T}} (L(m))_x H_p dx = &\int_{\mathbb{T}} \Big(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\Big)\Big(m_t - m_x\Big(H_p + \frac{m}{2}H_{pm}\Big)\Big)^2 dx \\
&- \int_{\mathbb{T}} m_x^2 \left(\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m}\right)\Big(\frac{m^2}{4}H_{pm}^2 + mH_{pp}H_m\Big) dx. \quad (3.3.5)
\end{aligned}
$$

Next, notice that for a smooth function $h : \mathbb{R} \to \mathbb{R}$, we have

$$
\frac{d}{dt}\int_{\mathbb{T}} h(m) dx = \int_{\mathbb{T}} (h(m))_x H_p + mh'(m)(H_p)_x dx = \int_{\mathbb{T}} (h(m) - h'(m)m)_x H_p dx.
$$

Thus, if we require that

$$
-L(m) = h(m) - h'(m)m,
$$

we obtain

$$
-\frac{d}{dt}\int_{\mathbb{T}} (L(m))_x H_p dx = \frac{d^2}{dt^2}\int_{\mathbb{T}} h(m) dx.
$$

The relation between $h, \widetilde{h}$ is

$$
m\widetilde{h}(m) = h'(m)m - h(m),
$$

therefore

$$
\widetilde{h}(m) = -\frac{h(m)}{m} + h'(m),
$$

and, thus,

$$\widetilde{h}'(m) + \frac{\widetilde{h}(m)}{m} = -\frac{h'(m)}{m} + \frac{h(m)}{m^2} + h''(m) - \frac{h(m)}{m^2} + \frac{h'(m)}{m} = h''(m),$$

from which (3.3.2) follows.

Now, setting $r = 1 - \frac{1}{1+C_0^{-1}}$, we have

$$-\frac{m^2}{4}H_{pm}^2 - mH_mH_{pp} = -\frac{m^2}{4}H_{pm}^2 - (1-r)mH_mH_{pp} - rmH_mH_{pp},$$

and so, applying (E), and multiplying by $h''(m)m_x^2$, (3.3.2) yields

$$\frac{d^2}{dt^2}\int_{\mathbb{T}} h(m(x,t))dx \geq \int_{\mathbb{T}} -rh''(m)mH_mH_{pp}m_x^2. \tag{3.3.6}$$

On the other hand, we infer from (E) that

$$\begin{aligned}
&\Big(m_t - m_x(H_p + \frac{m}{2}H_{pm})\Big)^2 - m_x^2\left(\frac{m^2}{4}H_{pm}^2 + mH_mH_{pp}\right) \\
&\geq \Big(m_t - m_xH_p - \frac{m_xm}{2}H_{pm}\Big)^2 + \frac{1}{C_0}\left(\frac{m_xm}{2}H_{pm}\right)^2 = (m_t - m_xH_p)^2 - 2(m_t - m_xH_p)\frac{m_xm}{2}H_{pm} \\
&+ (1-r)^{-1}\left(\frac{m_xm}{2}H_{pm}\right)^2 = r(m_t - m_xH_p)^2 + \left((1-r)^{\frac{1}{2}}(m_t - m_xH_p) - (1-r)^{-\frac{1}{2}}\frac{m_xm}{2}H_{pm}\right)^2 \\
&\geq r(m_t - m_xH_p)^2 = rm^2H_{pp}^2u_{xx}^2.
\end{aligned} \tag{3.3.7}$$

where the last equality follows from the equation of m. As before, multiplying by $h''(m)$ then yields

$$\frac{d^2}{dt^2}\int_{\mathbb{T}} h(m(x,t))dx \geq \int_{\mathbb{T}} rh''(m)m^2H_{pp}^2u_{xx}^2. \tag{3.3.8}$$

Combining (3.3.6) and (3.3.8), we conclude that (3.3.3) holds. □

It now follows readily that the density of the solution is bounded above and below in terms of the initial and terminal densities.

Corollary 3.3.2. *Let* $(u, m) \in C^2(\overline{Q}_T) \times C^1(\overline{Q}_T)$ *be a classical solution to* (MFG) *or* (MFGP). *Then, if* $c_1 := \min(\min m_0, \min m(\cdot, T))$, $C_1 = \max(\max m_0, \max m(\cdot, T))$, *one has*

$$c_1 \leq m(x,t) \leq C_1, \text{ for all } (x,t) \in \overline{Q}_T. \tag{3.3.9}$$

Proof. The proof follows directly from Proposition 3.3.1 above. Indeed, note that, in view of (E), for any convex function h, the map

$$C(t) := \int_{\mathbb{T}} h(m(x,t))dx$$

is convex, and thus

$$C(t) \leq \max(C(0), C(T)), \text{ for all } t \in [0, T].$$

Hence, setting $h_p(m) = m^p$ and letting $p \to -\infty$ yields the result for the lower bound, whereas letting $p \to +\infty$ yields the upper bound. □

Remark 6. *For dimensions* $d > 1$, *formula (3.3.2) is no longer true. If one repeats the same argument, the issue will arise at the term* $A_{6.2}$. *However, in the case of a separated Hamiltonian, i.e.* $H(p, m) \equiv H(p) - f(m)$, *one still obtains the weaker formula*

$$\begin{aligned}\frac{d^2}{dt^2} \int_{\mathbb{T}} h(m(x,t))dx = \int_{\mathbb{T}} ((h''(m)m^2 - h'(m)m + h(m))(\mathrm{Tr}(D^2_{pp}HD^2_{xx}u))^2 \\ + (h'(m)m - h(m))\mathrm{Tr}((D^2_{pp}HD^2_{xx}u)^2) + h''(m)mf'(m)|Dm|^2)dx. \end{aligned} \tag{3.3.10}$$

In this higher-dimensional setting, it is no longer true that the left hand side is convex whenever h *is convex. In particular, the statement is false for negative powers of* m, *but true for positive powers. Thus, from the proof of Corollary 3.3.2 we see that the upper bound on* m *still holds (see [214]).*

3.4 Estimates on the solution and the terminal density

In this section we obtain the necessary a priori $L^\infty-$bounds on u, Du, and $m(\cdot, T)$ for solutions to both (MFG) and (MFGP). Combined with the results of the previous section, this will yield global upper and lower bounds on the density. In order to treat the setting of Theorem 3.1.2, where the density may vanish at $\{0, T\}$, we also obtain L^∞-bounds on u that do not depend on the quantities $(\min m_0)^{-1}$, $(\min m_T)^{-1}$.

Proposition 3.4.1. *Let* $(u, m) \in C^2(\overline{Q}_T) \times C^1(\overline{Q}_T)$ *be a classical solution to* (MFG)*, and let* $c_1 = \min m_0, C_1 = \max m_0$. *Then, for each* $(x, t) \in \overline{Q}_T$,

$$c_1 \leq m(x, T) \leq C_1, \tag{3.4.1}$$

$$H(0, c_1)(t - T) + g(c_1) \leq u(x, t) \leq H(0, C_1)(t - T) + g(C_1), \tag{3.4.2}$$

and

$$-\int_t^T H(0, \min_{\mathbb{T}}(m(\cdot, s))ds + g(c_1) \leq u(x, t) \leq -\int_t^T H(0, \max_{\mathbb{T}}(m(\cdot, s))ds + g(C_1). \tag{3.4.3}$$

Proof. We will only show the lower bounds, since the argument for the upper bounds is completely symmetrical. Since $H_m < 0$, we may fix $\delta > 0$ and $\epsilon > 0$, such that

$$H(0, c_1) - H(0, c_1 - \delta) < -\epsilon T. \tag{3.4.4}$$

We define

$$w^{\epsilon,\delta}(t) := H(0, c_1 - \delta)(t - T) + \frac{\epsilon}{2}(t - T)^2 + g(c_1 - \delta),$$

and note that

$$w_{xx} = 0, w_{x,t} = 0, w_{tt} = \epsilon.$$

The function $v^{\epsilon,\delta}(x,t) := u(x,t) - w^{\epsilon,\delta}(t)$ has a minimum at some $(x_0, t_0) \in \overline{Q}_T$. If we first assume that $t_0 \in (0,T)$, then it follows that

$$D^2 u - D^2 w^{\epsilon,\delta} \geq 0,$$

which, in view of (Q), implies

$$0 = -\mathrm{Tr}(AD^2 u) \leq -\mathrm{Tr}(AD^2 w^{\epsilon,\delta}) = -\epsilon < 0,$$

a contradiction. On the other hand, assume that $t_0 = 0$. Then,

$$u_t(x_0, 0) \geq w_t^{\epsilon,\delta}(x_0, 0), \;\; u_x(x_0, 0) = w_x^{\epsilon,\delta}(0) = 0,$$

and thus, using the monotonicity of H and (3.4.4),

$$0 = -u_t(x_0, t_0) + H(0, m_0(x_0)) \leq -w_t^{\epsilon,\delta}(0) + H(0, m_0(x_0)) = -H(0, c_0 - \delta) + H(0, m_0(x_0)) + \epsilon T$$

$$\leq -H(0, c_1 - \delta) + H(0, c_1) + \epsilon T < 0,$$

which is again a contradiction. Hence, the minimum must be achieved at $t_0 = T$. At that point, we have

$$u_t(x_0, T) \leq w_t^{\epsilon,\delta}(T), \;\; u_x(x_0, T) = w_x^{\epsilon,\delta}(T) = 0.$$

Consequently, from (G1) and the monotonicity of H, we have

$$u(x_0, T) = g(H^{-1}(0, u_t(x_0, T))) \geq g(H^{-1}(0, w_t^{\epsilon,\delta}(T))) = g(H^{-1}(0, H(0, c_1 - \delta)))$$

$$= g(c_1 - \delta) = w^{\epsilon,\delta}(T).$$

We have thus shown that

$$u(x,t) \geq w^{\epsilon,\delta}(t), \text{ for all } (x,t) \in \overline{Q}_T.$$

Letting $\epsilon \to 0$, and then $\delta \to 0$, yields the lower bound in (3.4.2). In particular, for $t = T$, we have

$$g(m(x,T)) \geq g(c_1) \text{ for all } x \text{ in } \mathbb{T},$$

which proves the lower bound in (3.4.1), in view of (G1). Now, we define

$$w(t) = -\int_t^T H(0, c(s))ds + g(c_1),$$

where $c(s) := \min_{\mathbb{T}}\{m(\cdot, s)\}$ is the running minimum of the density. We observe that the function $v(x,t) = u(x,t) - w(t)$ satisfies $v_t = u_t - H(0, c(t))$, $v_x = u_x$. Thus, for any $\epsilon > 0$, at any extremum point of $v - \epsilon t$, the monotonicity of H implies that $v_t = H(0,m) - H(0,c(t)) - \epsilon < 0$. Letting $\epsilon \to 0$ thus implies that v achieves its minimum at $t = T$. Therefore, using (3.4.1), we obtain

$$u(x,t) - w(t) \geq \min_{\mathbb{T}} g(m(\cdot, T)) - g(c_1) \geq 0,$$

and this is precisely the lower bound in (3.4.3). □

Now, for solutions to (MFGP), we do not need to estimate the terminal density, as it is part of the given data. Concerning u, since the solution is only unique up to a constant, we may only bound the oscillation of u, and this is done in the following proposition.

Proposition 3.4.2. *Let $(u,m) \in C^2(\overline{Q}_T) \times C^1(\overline{Q}_T)$ solve (3.3.1). There exists a constant $C > 0$, with*

$$C = C\left(C_0, \int_0^T |H(0, \min_{\mathbb{T}} m(\cdot, s))| ds, \overline{C}(\max_{\overline{Q}_T} m)\right),$$

such that

$$\operatorname{osc}_{\overline{Q}_T} u \leq C\left(T + T^{-\frac{1}{\gamma-1}} + \int_0^T |H(0, \min_{\mathbb{T}} m(\cdot, s)| ds\right).$$

Proof. We define the functions c and w, for $t \in [0, T]$, by

$$c(t) = \min_{\mathbb{T}} m(\cdot, t), \quad w(t) = -\int_t^T H(0, c(s))ds.$$

Arguing as in the proof of (3.4.3), we obtain

$$\max_{\overline{Q}_T}(u - w) = \max_{\mathbb{T}} (u(\cdot, 0) - w(0)), \ \min_{\overline{Q}_T}(u - w) = \min_{\mathbb{T}} (u(\cdot, T) - w(T)). \tag{3.4.5}$$

Now, in view of (H1) and Proposition 3.4.1, $0 = -u_t + H(u_x, m) \geq -u_t + \frac{1}{C}|u_x|^\gamma - C$. Next, we define γ' by $\frac{1}{\gamma} + \frac{1}{\gamma'} = 1$. By the Hopf-Lax formula, the function

$$v(x, t) = \min_{y \in \mathbb{R}} \Big(\left(\frac{C}{\gamma}\right)^{\frac{\gamma'}{\gamma}} (T - t) \frac{|x - y|^{\gamma'}}{\gamma'(T - t)^{\gamma'}} + C(T - t) + u(y, T) \Big)$$

then solves, in $\overline{Q}_T$,

$$-v_t(x, t) + \frac{1}{C}|v_x|^\gamma - C = 0, \ v(\cdot, T) = u(\cdot, T),$$

and, thus, by the comparison principle,

$$u \leq v.$$

On the other hand, up to increasing the constant C,

$$v(x, 0) \leq \frac{C}{T^{\gamma' - 1}} + CT + \min_{\mathbb{T}} u(\cdot, T),$$

and so

$$\max_{\mathbb{T}} u(\cdot, 0) \leq \max_{\mathbb{T}} v(\cdot, 0) \leq \frac{C}{T^{\gamma' - 1}} + CT + \min_{\mathbb{T}} u(\cdot, T).$$

In view of (3.4.5), we obtain

$$\text{osc}_{\overline{Q}_T}(u - w) \leq \frac{C}{T^{\gamma' - 1}} + CT + w(T) - w(0),$$

and, thus,

$$\operatorname{osc}_{\overline{Q}_T} u \leq \frac{C}{T^{\gamma'-1}} + CT + 2 \cdot \operatorname{osc}_{\overline{Q}_T} w \leq \frac{C}{T^{\gamma'-1}} + CT + 2 \int_0^T |H(0, c(s))| ds.$$

□

We finally obtain a priori estimates on the gradient of u, while simultaneously treating the case of (MFG) and (MFGP). The proof closely follows [316, Lem. 3.8] and [317, Lem 3.3], but allows for weaker assumptions due to the $d = 1$ setting (see (3.4.9)). In fact, in the special case of a separated Hamiltonian, this proof can be seen to yield a gradient bound that is independent of $\min(m)$.

Proposition 3.4.3. *Let* $(u, m) \in C^3(\overline{Q}_T) \times C^2(\overline{Q}_T)$ *be a classical solution to* (MFG) *or* (MFGP). *There exists a constant* $C > 0$*, with*

$$C = C\Big(C_0, T, T^{-1}, \operatorname{osc} u, \gamma, \|m\|_{L^\infty(\overline{Q}_T)}, \|m^{-1}\|_{L^\infty(\overline{Q}_T)}, \\ \|(m_0)_x\|_{L^\infty(\mathbb{T})}, \|(m_T)_x\|_{L^\infty(\mathbb{T})}, \|\overline{C}\|_{L^\infty[\min m, \max m]}\Big)$$

such that

$$\|Du\|_{L^\infty(\overline{Q}_T)} \leq C.$$

Proof. Since $u_t = H(u_x, m)$, and m is bounded above and below, we infer from (H1) and (H2) that it is enough to show that

$$\|u_x\|_{L^\infty(\overline{Q}_T)} \leq CT^2.$$

We let

$$\widetilde{u} = u - \min u + 1 - \frac{(\operatorname{osc} u + 2)}{T}(T - t),$$

and note that the function $\widetilde{u}$ has been constructed to satisfy

$$|\widetilde{u}| \leq 1 + \operatorname{osc} u, \quad \widetilde{u}(\cdot, 0) \leq -1, \ \widetilde{u}(\cdot, T) \geq 1.$$

Define

$$v(x,t) = \frac{1}{2}u_x^2 + \frac{k}{2}\widetilde{u}^2,$$

where $k = \|u_x\|_{\overline{Q}_T}^{\frac{3}{2}}$. Let $(x_0, t_0) \in \overline{Q}_T$ be a point where v achieves its maximum value. With no loss of generality, we may assume that $p = u_x(x_0, t_0)$ satisfies

$$|p| \geq 1, \ |p|^2 \geq \frac{1}{2}\|u_x\|^2. \tag{3.4.6}$$

We remark here that throughout the proof, the constant C is subject to increase from line to line.

Case 1: $t_0 = T$. For this case we consider the linearization of the HJ equation,

$$T_u v = -v_t + H_p(u_x, m)v_x.$$

Since $v_x = 0$ and $v_t \geq 0$,

$$\begin{aligned} 0 \geq T_u v = T_u\left(\frac{1}{2}|u_x|^2\right) + k\widetilde{u}(-\widetilde{u}_t + H_p u_x) \\ = -H_m u_x m_x + k\widetilde{u}(-u_t + H_p p - C) \geq -H_m u_x m_x + k\widetilde{u}(\frac{1}{C_0}H) - Ck\widetilde{u} \\ \geq -H_m u_x m_x + k\widetilde{u}\frac{1}{C_0}\left(\frac{1}{\overline{C}(m)}|p|^\gamma - \overline{C}(m)\right) - C|p|^{\frac{3}{2}} \geq -H_m u_x m_x + \frac{1}{C}|p|^{\gamma+\frac{3}{2}} - C|p|^{\frac{3}{2}}. \end{aligned} \tag{3.4.7}$$

If (u, m) solves (MFG), then

$$-H_m u_x m_x = -\frac{H_m}{g'}|p|^2 > 0.$$

On the other hand, if (u, m) solves (MFGP), then

$$|-H_m u_x m_x| \leq C\|(m_T)_x\|_\infty |p|^{\gamma+1}. \tag{3.4.8}$$

In either case, (3.4.7) then implies

$$|p| \leq C.$$

Case 2: $t_0 = 0$. Regardless of whether (u, m) solves (MFG) or (MFGP), this case is dealt with in the same way as was done for $t_0 = T$ when (u, m) solved (MFGP), because, in view of (HM2), we then have the bound

$$|-H_m u_x m_x| \leq C \|(m_0)_x\|_\infty |p|^{\gamma+1}.$$

Case 3: $0 < t_0 < T$. We first observe that, since $v_x = 0$, we have

$$u_x u_{xx} = -k\widetilde{u} u_x,$$

and, thus,

$$|u_{xx}| \leq Ck. \tag{3.4.9}$$

We consider the linearization of (Q), namely

$$L_u(w) = -\mathrm{Tr}(A(Du)D^2 w) - D_q \mathrm{Tr}(A(Du)D^2 u) \cdot Dw.$$

Through direct computation, using (Q1), one obtains

$$\begin{aligned} L_u\left(\frac{1}{2}u_x^2\right) &= -\left|-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right|^2 + \frac{1}{4}m^2 H_{mp}^2 u_{xx}^2 - mH_m H_{pp} u_{xx}^2 \\ &\leq -\left|-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right|^2, \end{aligned} \tag{3.4.10}$$

where (E1) was used in the last inequality. Similarly,

$$L_u\left(k\frac{1}{2}\widetilde{u}^2\right) = -k\left|-\widetilde{u}_t + (H_p + \frac{1}{2}mH_{mp})u_x\right|^2 + k\frac{1}{4}m^2 H_{mp}^2 u_x^2 - kmH_m H_{pp} u_x^2 + E_1 + E_2 + E_3 + E_4, \tag{3.4.11}$$

where

$$E_1 = 2\left(-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right)\left(H_{pp} + \frac{1}{2}mH_{mpp}\right)k\widetilde{u}u_x,$$

$$E_2 = \left(\frac{1}{2}H_{mp}H_{mpp} + mH_{mp}H_{pp} + mH_mH_{ppp}\right)u_{xx}k\widetilde{u}u_x,$$

$$E_3 = \left(-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right)\frac{2}{H_m}\left(H_{pm} + \frac{1}{2}\left(mH_{mmp} + H_{mp}\right)\right)k\widetilde{u}(-\widetilde{u}_t + H_pu_x)$$

$$E_4 = \frac{1}{H_m}\bigg(\frac{1}{2}(mH_{mp}^2 + m^2H_{mp}H_{mmp}) + mH_{mm}H_{pp} + mH_mH_{mpp} + H_mH_{pp}\bigg)u_{xx}k\widetilde{u}(-\widetilde{u}_t + H_pu_x).$$

Now we estimate each of the E_i. By Young's inequality, we obtain

$$|E_1| \leq \frac{1}{4}\left|-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right|^2 + C\left|H_{pp} + \frac{1}{2}mH_{mpp}\right|^2 k^2u_x^2\widetilde{u}^2.$$

As a result of (H1) and (HM3), we thus obtain

$$|E_1| \leq \frac{1}{4}\left|-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)\right|^2 + C|p|^{2\gamma+1}. \tag{3.4.12}$$

Next, to estimate $|E_2|$, we use (3.4.9), (H1) (H3), (HM1), (HM3) and (E1) to obtain

$$|E_2| \leq C|p|^{2\gamma+1}. \tag{3.4.13}$$

For E_3, we have

$$|E_3| \leq \frac{1}{4}\left|-u_{xt} + \left(H_p + \frac{1}{2}mH_{mp}\right)u_{xx}\right|^2 + \frac{Ck^2}{H_m^2}\left(H_{pm}^2 + m^2H_{mmp}^2 + H_{mp}^2\right)|-\widetilde{u}_t + H_pu_x|^2. \tag{3.4.14}$$

Now, recalling that $u_t = H(p,m)$, we infer from (H1), (H2), and (3.4.6) that

$$\frac{1}{C}|p|^\gamma \leq |-\widetilde{u}_t + H_p u_x| \leq C|p|^\gamma. \tag{3.4.15}$$

Therefore, in view of (H1), (HM1), (HM2), and (E1), as well as the HJ equation, we obtain

$$|E_3| \leq \frac{1}{4}\left|-u_{xt} + (H_p + \frac{1}{2}mH_{mp})u_{xx}\right|^2 + C|p|^{2\gamma+1}. \tag{3.4.16}$$

Finally, for E_4, we observe that (3.4.9), (H1), (HM2), (HM3), (E1), and (3.4.15) yield

$$|E_4| \leq C|p|^{2\gamma+1}. \tag{3.4.17}$$

Now, (E) implies that

$$\begin{aligned}
&\left|-\widetilde{u}_t + (H_p + \frac{1}{2}mH_{mp})u_x\right|^2 - \frac{1}{4}m^2H_{mp}^2p^2 + mH_mH_{pp}p^2 \\
&\quad \geq \left|-\widetilde{u}_t + (H_p + \frac{1}{2}mH_{mp})u_x\right|^2 + \frac{1}{4C_0}m^2H_{mp}^2p^2 = \left|-\widetilde{u}_t + \left(H_pu_x + \frac{1}{2}mH_{mp}\right)p\right|^2 \\
&\qquad + \frac{1}{C_0}\left(\frac{1}{2}mH_{mp}p\right)^2 \geq \frac{1}{2}\left|-\widetilde{u}_t + \left(H_p + \frac{1}{2}mH_{mp}\right)u_x\right|^2 + \frac{1}{C}|-\widetilde{u}_t + H_pu_x|^2.
\end{aligned} \tag{3.4.18}$$

So, as a result of (3.4.11) and (3.4.15) we get

$$L_u\left(k\frac{1}{2}\widetilde{u}^2\right) \leq -\frac{1}{2}k\left|-\widetilde{u}_t + \left(H_p + \frac{1}{2}mH_{mp}\right)u_x\right|^2 - \frac{1}{C}|p|^{2\gamma+\frac{3}{2}} + E_1 + E_2 + E_3 + E_4. \tag{3.4.19}$$

Now, since (x_0, t_0) is an interior maximum point of v, we have $L_u(v) \geq 0$. Thus, combining (3.4.12), (3.4.13), (3.4.16), (3.4.17), (3.4.10) and (3.4.19), we conclude

$$0 \leq -\frac{1}{C}|p|^{2\gamma+\frac{3}{2}} + C|p|^{2\gamma+1},$$

which implies

$$|p| \leq C.$$

□

3.4.1 Estimates for MFG with ϵ–penalized terminal condition

In order to obtain classical solutions to (MFGP), it will be necessary to use a natural approximation method, which was previously used in [324] to obtain weak solutions to the second-order planning problem. The solution will be obtained as the limit of solutions to standard MFG systems with a penalized terminal condition. Specifically, we will need to prove estimates for solutions (u^ϵ, m^ϵ) to

$$\begin{cases} -u^\epsilon_t + H(u^\epsilon_x, m^\epsilon) = 0 \text{ in } Q_T, \\ m^\epsilon_t - (m^\epsilon H_p(u^\epsilon_x, m^\epsilon))_x = 0 \text{ in } Q_T, \\ m^\epsilon(x,0) = m_0(x),\ \epsilon u^\epsilon(x,T) = m^\epsilon(x,T) - m_T(x) \text{ on } \partial Q_T. \end{cases} \tag{MFG$_\epsilon$}$$

As long as u^ϵ is bounded in $L^\infty(Q_T)$, the limit is expected to solve (MFGP). This estimate is obtained in the following lemma. While treating this system, we will temporarily assume that $H(0,0)$ is finite. This assumption will be removed in the proof of Theorem 3.1.1.

Lemma 3.4.4. *For $\epsilon > 0$, let $(u^\epsilon, m^\epsilon) \in C^2(\overline{Q_T}) \times C^1(\overline{Q_T})$ be a classical solution to system (MFG$_\epsilon$), and set $c_1 = \min\{\min_{\mathbb{T}} m_0, \min_{\mathbb{T}} m_T\}$, $C_1 = \max\{\max_{\mathbb{T}} m_0, \max_{\mathbb{T}} m_T\}$. Assume that $H(0,0) < \infty$. Then there exists a constant $C > 0$, independent of ϵ, such that*

$$\|u^\epsilon\|_{L^\infty(\overline{Q_T})} \leq C. \tag{3.4.20}$$

Furthermore, for all $\epsilon < \frac{1}{C}$, we have

$$\frac{c_1}{2} \leq m^\epsilon(x,t) \leq 2C_1 \text{ for all } (x,t) \in \overline{Q_T}, \tag{3.4.21}$$

and

$$\|m^{\epsilon}(T,\cdot) - m_T(\cdot)\|_{\infty} \leq \epsilon C. \tag{3.4.22}$$

Proof. As a result of Proposition 3.4.2, since $H(0, \min_{\overline{Q}_T} m^{\epsilon}) \leq H(0,0)$, there exists

$$C = C(C_0, T, |H(0,0)|, |H(0, \max_{\overline{Q}_T} m^{\epsilon})|, \overline{C}(\max_{\overline{Q}_T} m^{\epsilon}))$$

such that

$$\operatorname{osc}_{\overline{Q}_T}(u^{\epsilon}) \leq C.$$

To make this bound on the oscillation independent of ϵ, we must obtain upper bounds on the density m^{ϵ}. Note that, from Corollary 3.3.2, it is enough to bound $m^{\epsilon}(T,\cdot)$ from above. To this end, let $M_0 := \max_{\mathbb{T}} m_0$ and, for $\delta > 0$, define

$$v^{\delta}(x,t) = u^{\epsilon}(x,t) + H(0, M_0 + \delta)(T - t).$$

Since $D^2 v^{\delta} = D^2 u^{\epsilon}$, we have that v^{δ} also solves the elliptic equation (Q) in Q_T. Therefore, the maximum of v^{δ}, must occur at $t = 0$ or $t = T$. If the maximum occurred at $t = 0$, then at that point

$$u_t^{\epsilon} - H(0, M_0 + \delta) = v_t^{\delta} \leq 0, \; v_x^{\delta} = u_x^{\epsilon} = 0,$$

and, hence,

$$0 \geq u_t^{\epsilon} - H(0, M_0 + \delta) = H(0, m_0) - H(0, M_0 + \delta),$$

which is a contradiction because $H_m < 0$. Therefore, for every $\delta > 0$, the maximum occurs at $t = T$, and, letting $\delta \to 0$, we see that the same is true for $\delta = 0$. The maximum value of $v(x,t) := u^{\epsilon}(x,t) + H(0, M_0)(T - t)$ equals the maximum of $u^{\epsilon}(x,T)$, since $v(x,T) = u^{\epsilon}(x,T)$.

Letting $x_0 \in \mathbb{T}$ be a point at which this maximum occurs, it follows that $v_t(x_0, T) \geq 0$, and therefore

$$H(0, m^\epsilon(x_0, T)) \geq H(0, M_0),$$

which implies that

$$m^\epsilon(x_0, T) \leq M_0.$$

But, since

$$\epsilon u^\epsilon(x, T) = m^\epsilon(x, T) - m_T(x),$$

we obtain, for each $x \in \mathbb{T}$,

$$\epsilon u^\epsilon(x, T) \leq \epsilon u^\epsilon(x_0, T) = (m^\epsilon(x_0, T) - m_T(x_0)) \leq (M_0 - m_T(x_0)),$$

and, consequently,

$$m^\epsilon(x, T) = \epsilon u^\epsilon(x, T) + m_T(x) \leq M_0 + m_T(x) - m_T(x_0) \leq M_0 + \operatorname{osc}_{\mathbb{T}}(m_T).$$

We have thus shown that the bound on the oscillation of u^ϵ does not depend on ϵ. Furthermore, since

$$\epsilon u^\epsilon(x, T) = m^\epsilon(x, T) - m_T(x),$$

and $m^\epsilon(T, \cdot), m_T(\cdot)$ are both probability densities, we have $\int_{\mathbb{T}} u^\epsilon(\cdot, T) = 0$, so there must exist some $x^\epsilon \in \mathbb{T}$ such that

$$u^\epsilon(x^\epsilon, T) = 0.$$

This implies that, for any $(x, t) \in \overline{Q}_T$,

$$-\operatorname{osc}_{\overline{Q}_T}(u^\epsilon) \leq u^\epsilon(x, t) - u^\epsilon(x^\epsilon, T) \leq \operatorname{osc}_{\overline{Q}_T}(u^\epsilon),$$

which shows (3.4.20). To prove (3.4.21), we require C to be large enough to satisfy $\frac{1}{C}\|u^\epsilon\|_\infty < \frac{1}{2}c_1$. Then for all $\epsilon < \frac{1}{C}$, we have

$$m^\epsilon(x,T) = m_T(x) + \epsilon u^\epsilon(x,T) \geq m_T(x) - \frac{1}{2}c_1 \geq \frac{1}{2}c_1.$$

The upper bound for $m^\epsilon(x,T)$ is obtained similarly. We now conclude by Corollary 3.3.2, since the maxima and minima of m^ϵ both occur at $t = 0, t = T$. Finally, (3.4.22) follows immediately from the terminal condition in (MFG$_\epsilon$) and (3.4.20). □

While the usefulness of (MFG$_\epsilon$) will mainly be as a tool to obtain existence for (MFGP), it can also be used to provide an interesting counterexample. Indeed, one should note that (MFG$_\epsilon$) is not itself a planning problem, but rather a special case of a standard MFG system, which would fit in the framework of (MFG) if the terminal cost function g were allowed to depend on x. Such terminal conditions are treated in [316, 317] under the blow-up assumption (3.1.1), as well as the requirement that

$$g(x,0) \text{ is constant, or } \lim_{m\to 0^+} g(x,m) = -\infty,$$

which is a slightly weaker version of (3.1.1). The following proposition illustrates the fact that, when such assumptions do not hold, the solution may fail to exist.

Proposition 3.4.5. *Assume that $H(0,0) < \infty$, and that the condition $m_T > 0$ in* (M1) *does not hold, and $m_T(x_0) < 0$ for some $x_0 \in \mathbb{T}$. Then there exists $C > 0$ such that, for all $0 < \epsilon < \frac{1}{C}$, there exists no classical solution to (MFG$_\epsilon$).*

Proof. We assume, by contradiction, that there exists a decreasing sequence $\epsilon_n > 0$, with $\lim_{n\to\infty} \epsilon_n = 0$, such that, for each positive integer n, there exists a solution (u^n, m^n) to system (MFG$_{\epsilon_n}$). Since $H(0,0) < \infty$, the proof of Lemma 3.4.4 shows that, for some constant $C > 0$ independent of $n \in \mathbb{N}$, we have $\|u^n\|_\infty \leq C$. However, this implies that

$$\|m^n(T,\cdot) - m_T(\cdot)\|_\infty \leq C\epsilon_n,$$

while $m^n(x_0, T) \geq 0 > m_T(x_0)$, which is a contradiction. $\square$

We finish our estimates for the ϵ–penalized problem with an analogue of Proposition 3.4.3.

Lemma 3.4.6. *For $\epsilon > 0$, let $(u^\epsilon, m^\epsilon) \in C^{3,\alpha}(\overline{Q_T}) \times C^{2,\alpha}(\overline{Q_T})$ be a classical solution to system (MFG$_\epsilon$), and assume that $H(0,0) < \infty$. Let c_1 and C_1 be as in Corollary 3.3.2. There exists a constant $C > 0$, independent of ϵ, such that, for $\epsilon < \frac{1}{C}$,*

$$\|Du^\epsilon\|_\infty \leq C.$$

Proof. We first observe that, by Corollary 3.3.2 and Lemma 3.4.4, $\|m^\epsilon\|_{\overline{Q}_T}$ and $\|(m^\epsilon)^{-1}\|_{\overline{Q}_T}$ are bounded a priori in terms of C_1 and c_1^{-1}. The proof of Proposition 3.4.3 may thus be repeated here, with Lemma 3.4.4 replacing the use of Proposition 3.4.2, with one exception. Namely, the term $-H_m u_x^\epsilon m_x^\epsilon$ in (3.4.7) should be estimated as

$$-H_m u_x^\epsilon m_x^\epsilon = -\epsilon H_m (u_x^\epsilon)^2 - H_m u^\epsilon (m_T)_x \geq -H_m u^\epsilon (m_T)_x,$$

which, in view of (3.4.8), yields the gradient bound in the case $t_0 = T$. The rest of the argument follows unchanged. $\square$

3.5 Existence of classical solutions

In the previous sections, a priori L^∞–bounds were obtained for u, Du, m, and m^{-1}. This is already sufficient to obtain classical solutions to (MFG), following the arguments of [316, 317]. The existence of solutions to (MFGP), on the other hand, is a more delicate issue, because the Neumann type boundary condition that appears in the linearization makes the latter non–invertible. Namely, the linearization of (Q) is

$$\begin{cases} L_u(w) = f & \text{in } Q_T, \\ (-1, H_p(u_x, m)) \cdot Dw = g_1(x) & \text{at } t = 0, \\ (1, -H_p(u_x, m)) \cdot Dw = g_2(x) & \text{at } t = T, \end{cases}$$

which is an oblique boundary value problem that is only solvable for certain functions f, g_1, g_2 satisfying a compatibility condition that itself depends on u. This failure of invertibility precludes the direct use of the implicit function theorem and thus of the method of continuity, which means a different approach is needed. Indeed, we will obtain the solution as the limit as $\epsilon \to 0$ of the solution to the ϵ–penalized problem (MFG$_\epsilon$). We begin by noting, in the following lemma, that for ϵ small enough, the solutions to (MFG$_\epsilon$) are a priori uniformly bounded in $C^{1,\beta}(\overline{Q}_T)$, for some $0 < \beta < 1$, and that the system thus has a classical solution.

Lemma 3.5.1. *Let C be as in Lemma 3.4.4. For all $0 < \epsilon < \frac{1}{C}$, (MFG$_\epsilon$) has a unique smooth solution $(u^\epsilon, m^\epsilon) \in C^{3,\alpha}(\overline{Q_T}) \times C^{2,\alpha}(\overline{Q_T})$. Moreover, there exist constants $K > 0$, $0 < \beta < 1$, independent of ϵ, such that*

$$\|u^\epsilon\|_{C^{1,\beta}} \leq K. \tag{3.5.1}$$

Proof. The a priori C^1–bounds on u^ϵ, as well as L^∞–bounds on m^ϵ and $(m^\epsilon)^{-1}$ (and thus on the ellipticity constants of the system), were all established in Lemmas 3.4.4 and 3.4.6. The Hölder estimate for the gradient then follows in the same way as in [316, Lem. 4.1], by directly applying the classical $C^{1,\alpha}$–estimates for quasilinear elliptic equations with oblique boundary conditions (see [292, Lem. 2.3]). Indeed, it suffices to verify that, for $(x, t, z, p, s) \in \mathbb{T} \times \{0, T\} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, the boundary condition

$$B^\epsilon(x, 0, z, p, s) = -s + H(p, m_0(x)), \quad B^\epsilon(x, T, z, p, s) = s - H(p, \epsilon z + m_T(x)),$$

is oblique. For this purpose, we let $\nu(x,t)$ denote the outward unit normal vector at $(x,t) \in \partial Q_T$. Then we have

$$D_{(p,s)}B^{\epsilon}(x,0,z,p,s) \cdot \nu(x,0) = -B^{\epsilon}_s(x,0,z,p,s) = 1 > 0,$$

$$D_{(p,s)}B^{\epsilon}(x,T,z,p,s) \cdot \nu(x,T) = -B^{\epsilon}_s(x,T,z,p,s) = 1 > 0$$

and thus the a priori estimate (3.5.1) follows. The proof of existence is then the same as in [316, Thm. 1.1] through the method of continuity. □

We now have enough information on the ϵ–penalized problem to prove our first theorem.

Proof of Theorem 3.1.1. We initially assume that $m_0, m_T \in C^{\infty}(\mathbb{T})$. The proof of part (ii), corresponding to (MFG), is identical to the one carried out in [316, Thm. 1.1]. We simply note that the condition $\lim_{m\to 0^+} H(p,m) = +\infty$ in that proof was only used to guarantee the existence of a positive lower bound for the density, which in turn makes the equation (Q) uniformly elliptic. In our case, the lower bound is a consequence of Corollary 3.3.2 and Proposition 3.4.1.

Now, for the case of (MFGP), we remark first that uniqueness of u, up to a constant, follows by the standard Lasry-Lions monotonicity method. To establish existence, we consider first the approximate system (MFG$_\epsilon$), under the assumption $H(0,0) < \infty$. We assume that $\epsilon > 0$ is small enough for Lemma 3.5.1 to guarantee the existence of solutions $(u^{\epsilon}, m^{\epsilon})$. Letting $0 < \beta < 1$ be as in Lemma 3.5.1, we also have (3.5.1), for some constant $K > 0$ independent of ϵ. We infer that there exist a subsequence $\{u_n\}_n \subset \{u^{\epsilon}\}_{\epsilon}$, and $u \in C^{1,\alpha}(\overline{Q_T})$, such that $u_n \to u$ uniformly. Furthermore, in view of Lemma 3.4.4, there exists $C > 0$, independent of ϵ, such that

$$\frac{1}{C} \le m^{\epsilon}(x,t) \le C \text{ for all } (x,t) \in \overline{Q_T}.$$

We let (A, B) and (A_n, B_n), be the quasilinear operators and boundary conditions corresponding, respectively, to u and u_n. Then one has

$$(A_n, B_n) \to (A, B) \text{ locally uniformly,}$$

$$D_q B_n \cdot \nu = 1.$$

Hence, by Fiorenza's convergence theorem for elliptic equations with oblique boundary conditions (see [316, Thm. 2.5], [210, Chapter 17, Lemma 17.29]), we obtain $u_n \to u$ in $C^{2,\alpha}(\overline{Q}_T)$, and u solves (Q), with the boundary condition corresponding to (MFGP). The $C^{3,\alpha}$ regularity (and, in fact, uniform convergence in $C^{3,\alpha}$) then follows readily from the standard Schauder estimates for linear oblique problems, as in [316, Thm. 1.1].

The last step will be to remove the assumption that $m_0 \in C^\infty(\mathbb{T})$ and, for (MFGP), the assumptions that $m_T \in C^\infty(\mathbb{T})$ and $H(0,0) < \infty$. We will explain the argument for (MFGP), with the treatment of (MFG) being completely analogous. Consider, for $\delta > 0$, the modified Hamiltonians $H^\delta(p,m) := H(p, m+\delta)$, which satisfy (H) and (E), uniformly in δ, as well as $H^\delta(0,0) < \infty$, and a sequence of C^∞ densities (m_0^δ, m_T^δ), uniformly bounded in $C^{2,\alpha}$ and bounded away from 0, converging uniformly to (m_0, m_T). Let (u^δ, m^δ) be the corresponding solutions to

$$\begin{cases} -u_t^\delta + H^\delta(u_x^\delta, m^\delta) = 0 & \text{in } Q_T, \\ \int_0^T \fint_{\mathbb{T}} u^\delta = 0, & \\ m_t^\delta - (m^\delta H_p^\delta(u_x^\delta, m^\delta))_x = 0 & \text{in } Q_T, \\ m^\delta(\cdot, 0) = m_0^\delta, \; m^\delta(\cdot, T) = m_T^\delta & \text{on } \mathbb{T}. \end{cases} \tag{3.5.2}$$

Propositions 3.4.3 and 3.4.2, and Corollary 3.3.2, yield uniform C^1-bounds on u^δ, and thus, as in the proof of Lemma 3.5.1, uniform $C^{1,\beta}$ bounds for some $0 < \beta < 1$. We may thus conclude by letting $\delta \to 0$ and applying Fiorenza's convergence result as above. □

3.6 Regularity of weak solutions

We now study the existence and regularity of solutions to (MFG) and (MFGP) under the weaker assumption that, for some $\kappa > 0$

$$\int_{\mathbb{T}} \frac{1}{m_0^{\kappa}(x)} dx < \infty, \int_{\mathbb{T}} \frac{1}{m_T^{\kappa}(x)} dx < \infty.$$

We note that, in particular, the above conditions allow for the densities to vanish at a set of measure zero. This, in general, creates significant issues, because (Q) is no longer uniformly elliptic. The key estimate that will allow us to prove smoothness in this setting is an interior lower bound on the density which depends only on t^{-1}, $\|m_0^{-\kappa}\|_1$ (and $(T-t)^{-1}, \|m_T^{-\kappa}\|_1$, in the case of (MFGP)). Indeed, this yields uniform ellipticity of (Q) away from $t = 0$ and $t = T$.

We begin by giving the standard definition of a weak solution (see, for instance, [81, 316, 328]).

Definition 3.6.1 (Definition of weak solution)**.** *A pair $(u, m) \in BV(Q_T) \times L^{\infty}_{+}(Q_T)$ is called a weak solution to* (MFG) *(respectively* (MFGP)*) if the following conditions hold:*

(i) $u_x \in L^2(Q_T), u \in L^{\infty}(Q_T), m \in C^0([0,T]; H^{-1}(\mathbb{T}))$.

(ii) u satisfies the HJ inequality

$$-u_t + H(u_x, m) \leq 0 \quad \text{in } Q_T,$$

in the distributional sense.

(iii) m satisfies the continuity equation

$$m_t - (mH_p(u_x, m))_x = 0 \text{ in } Q_T, \tag{3.6.1}$$

in the distributional sense.

(iv) We have $m(\cdot,T) \in L^\infty(\mathbb{T})$. Moreover, $m(\cdot,0) = m_0$ in $H^{-1}(\mathbb{T})$ and $u(\cdot,T) = g(m(\cdot,T))$ in the sense of traces (respectively, $m(\cdot,T) = m_T$ in $H^{-1}(\mathbb{T})$).

(v) The following identity holds:

$$\int\int_{Q_T} m(x,t)(H(u_x,m) - H_p(u_x,m)u_x)dxdt = \int_{\mathbb{T}}(m(x,T)u(x,T) - m_0(x)u(x,0))dx.$$

The following lemma will be needed to show that, for solutions to (MFG), our interior regularity results may be extended up to time $t = T$.

Lemma 3.6.2. *Let (u,m) be a smooth solution to* (MFG) *under the assumptions of Theorem 3.1.1 and assume that* (3.1.3) *holds. Then, for every convex function $h \in C^2(0,\infty)$, the map*

$$t \to \int_{\mathbb{T}} h(m(x,t))dx$$

is decreasing. Moreover, there exists a constant $C = C(C_0, \|g'\|^{-(\gamma-1)}_{L^\infty([\min m_0, \max m_0])})$ such that

$$\frac{d}{dt}\int_{\mathbb{T}} h(m(x,T))dx + \frac{1}{C}\int_{\mathbb{T}} h''(m(x,T))|m_x(x,T)|^\gamma \leq 0.$$

Proof. In view of Proposition 3.3.1, we have that

$$\frac{d^2}{dt^2}\int_{\mathbb{T}} h(m(x,t))dx \geq 0,$$

and, thus, the function

$$d(t) := \frac{d}{dt}\int_{\mathbb{T}} h(m(x,t))dx$$

is increasing. We then infer that the monotonicity will follow if we show that

$$d(T) \leq 0.$$

Since $u(\cdot, T) = g(m(\cdot, T))$, and m satisfies the continuity equation, we have

$$d(T) = \int_{\mathbb{T}} h'(m(x,T))m_t(x,T)dx = \int_{\mathbb{T}} h'(m)(mH_p(u_x, m))_x dx = -\int_{\mathbb{T}} h''(m)m_x H_p(m_x g'(m), m).$$

Now, as a result of (3.1.3) and (H1),

$$H_p(m_x g'(m), m)(m_x g'(m)) \geq \frac{1}{C}|m_x g'(m)|^{\gamma},$$

and, therefore,

$$d(T) \leq -\frac{1}{C}\int_{\mathbb{T}} h''(m)|m_x|^{\gamma}.$$

□

We are now ready to obtain the interior lower bounds on m. Our method of proof relies on the displacement convexity formula (3.3.2), and uses similar techniques to [328, Prop. 5.2].

Proposition 3.6.3. *Let* (u, m) *be a smooth solution to* (MFG) *or* (MFGP)*, under the same assumptions as in Theorem 3.1.1. Assume, furthermore, that* (HW) *holds and, in the case of* (MFG)*, assume that* (3.1.3) *holds. Let*

$$\beta = \frac{2}{\kappa - s - 1},$$

and let $\delta > 0$*. Then, there exist a constant* $C = C(C_0\|m_0^{-\kappa}\|_{L^1}, \|m_T^{-\kappa}\|_{L^1}, \delta^{-1})$ *such that*

$$m(x,t) \geq \frac{1}{C}\left(\frac{1}{t^{\beta+\delta}} + \frac{1}{(T-t)^{\beta+\delta}}\right)^{-1}. \tag{3.6.2}$$

Furthermore, in the case of (MFG)*, one has*

$$m(x,t) \geq \frac{1}{C}t^{\beta+\delta}. \tag{3.6.3}$$

Proof. Using the displacement convexity formula (3.3.2) for $h(m) = \frac{1}{m^{\kappa}}$, we have, for each $t \in$

$[0, T]$,

$$\int_{\mathbb{T}} \frac{1}{m^{\kappa}(x,t)} dx \leq \max\left(\int_{\mathbb{T}} \frac{1}{m_0^{\kappa}(x)} dx, \int_{\mathbb{T}} \frac{1}{m^{\kappa}(x,T)} dx \right). \tag{3.6.4}$$

Combined with Lemma 3.6.2 (for the case of (MFG) where $m(\cdot, T)$ is not prescribed), this yields

$$\sup_{t \in [0,T]} \|m^{-\kappa}(t)\|_1 \leq C. \tag{3.6.5}$$

Next, for any $p > 1$, we define the function

$$\phi(t) := \int_{\mathbb{T}} m^{-p\kappa}(t) dx.$$

Using Proposition 3.3.1 with $h(m) = m^{-p\kappa}$, as a result of (E), we obtain

$$\begin{aligned} \frac{d^2}{dt^2} \int_{\mathbb{T}} \frac{m^{-p\kappa}(t)}{p\kappa(p\kappa+1)} dx &\geq -\frac{1}{C} \int_{\mathbb{T}} m^{-p\kappa-1} H_{pp} H_m (m_x)^2 dx \geq \int_{\mathbb{T}} \frac{1}{C} m^{-p\kappa-1+s} (m_x)^2 dx \\ &\geq \frac{1}{C \frac{-p\kappa+s+1}{2})^2} \int_{\mathbb{T}} \left(\left(m^{\frac{-p\kappa+s+1}{2}} \right)_x \right)^2 dx. \end{aligned}$$

As a result, letting

$$C_p := \frac{C(p\kappa - s - 1)^2}{4p\kappa(p\kappa+1)},$$

$$\lambda := \frac{-p\kappa + s + 1}{2}, \tag{3.6.6}$$

we have shown that

$$C_p \phi''(t) \geq \int_{\mathbb{T}} (m^{\lambda})_x^2 dx. \tag{3.6.7}$$

From (W), and the fact that $p > 1$, we see that $\lambda < 0$. For each $t \in [0, T]$, since $m(\cdot, t)$ is a probability measure, there exists a point x_0^t such that $m(x_0^t, t) = 1$. By the fundamental theorem of calculus,

$$\left\| m^{\lambda}(t) - 1 \right\|_{\infty}^2 = \left\| m^{\lambda}(t) - m(x_0^t, t)^{\lambda} \right\|_{\infty}^2 \leq C \int_{\mathbb{T}} (m^{\lambda})_x^2 dx, \tag{3.6.8}$$

and therefore

$$\left\|\frac{1}{m}\right\|_{\infty}^{2|\lambda|} \leq C\Big(\int_{\mathbb{T}}(m^{\lambda})_x^2 dx + 1\Big). \tag{3.6.9}$$

Now, using (3.6.5), we obtain

$$\phi = \int_{\mathbb{T}} \frac{1}{m^{\kappa p}} \leq \int_{\mathbb{T}} \frac{1}{m^{\kappa}} \left\|\frac{1}{m}\right\|_{\infty}^{\kappa(p-1)} \leq C\left\|\frac{1}{m}\right\|_{\infty}^{\kappa(p-1)},$$

and, consequently,

$$C^{-r}\phi^r \leq \left\|\frac{1}{m}\right\|_{\infty}^{2|\lambda|}, \tag{3.6.10}$$

where $r := \frac{2|\lambda|}{\kappa(p-1)}$. From condition (W), we see that $r > 1$. Combining (3.6.7), (3.6.9), and (3.6.10), we obtain

$$C_p\Big(\phi''(t) + 1\Big) - C^{-r}\phi(t)^r \geq 0,$$

that is, for some constant $C = C(p)$,

$$-\phi''(t) + \frac{1}{C}\phi^r \leq C. \tag{3.6.11}$$

A straightforward computation then shows that the functions

$$\psi_1(t) = A_p t^{-p\kappa\beta} + K_p,$$

$$\psi_2(t) = A_p(T-t)^{-p\kappa\beta} + K_p,$$

$$\psi(t) = \psi_1(t) + \psi_2(t),$$

are supersolutions of (3.6.11) for large enough A_p, K_p. Therefore, we have

$$\int_{\mathbb{T}} m^{-p\kappa}(t) \leq A_p(t^{-p\kappa\beta} + (T-t)^{-p\kappa\beta}) + 2K_p. \tag{3.6.12}$$

Now, going back to (3.6.7) and (3.6.9), we may write

$$\left\|\frac{1}{m}\right\|_{\infty}^{2|\lambda|}(t) \le C\left(\frac{d^2}{dt^2}\int_{\mathbb{T}} m^{-p\kappa} + 1\right). \tag{3.6.13}$$

In view of (3.3.2), for $q > 0$, the map

$$t \mapsto \int_{\mathbb{T}} m^{-q}(t) \tag{3.6.14}$$

is convex in $[0, T]$. Thus, fixing $t_0 \in (0, \frac{T}{2}]$, we infer that, for each $t \in [t_0, T - t_0]$,

$$\begin{aligned}\left(\int_{\mathbb{T}} m^{-2|\lambda|q}(t)\right)^{\frac{1}{q}} &\le \frac{2}{t_0}\max\left(\int_{\frac{t_0}{2}}^{t_0}\left(\int_{\mathbb{T}} m^{-2|\lambda|q}\right)^{\frac{1}{q}}, \int_{T-t_0}^{T-\frac{t_0}{2}}\left(\int_{\mathbb{T}} m^{-2|\lambda|q}\right)^{\frac{1}{q}}\right)\\ &\le \frac{2}{t_0}\int_{\frac{t_0}{2}}^{T-\frac{t_0}{2}}\left(\int_{\mathbb{T}} m^{-2|\lambda|q}\right)^{\frac{1}{q}}.\end{aligned}$$

Letting $q \to \infty$, we obtain

$$\left\|m^{-1}\right\|_{L^{\infty}(\mathbb{T}\times[t_0,T-t_0])}^{2|\lambda|} \le \frac{2}{t_0}\int_{\frac{t_0}{2}}^{T-\frac{t_0}{2}}\left\|m^{-1}(t)\right\|_{\infty}^{2|\lambda|} dt. \tag{3.6.15}$$

Now, letting $\zeta \in C^{\infty}(Q_T)$ be a test function, supported in $[\frac{t_0}{4}, T - \frac{t_0}{4}]$, such that $0 \le \zeta \le 1$, $\zeta \equiv 1$ in $[\frac{t_0}{2}, T - \frac{t_0}{2}]$, and $\int_0^T |\zeta''(t)|dt \le \frac{C}{t_0}$, we see that (3.6.15) implies

$$\left\|m^{-1}\right\|_{L^{\infty}(\mathbb{T}\times[t_0,T-t_0])}^{2|\lambda|} \le \frac{2}{t_0}\int_0^T\left\|m^{-1}\right\|_{\infty}^{2|\lambda|}(t)\zeta(t)dt. \tag{3.6.16}$$

Hence, recalling (3.6.13) and integrating by parts twice, we infer from (3.6.12) that

$$\left\|m^{-1}\right\|_{L^{\infty}(\mathbb{T}\times[t_0,T-t_0])}^{2|\lambda|} \le \frac{C}{t_0}\left(\int_0^T\int_{\mathbb{T}}(m^{-p\kappa}\zeta'') + CT\right) \le C\left(\frac{1}{t_0^{2+p\kappa\beta}} + \frac{1}{t_0}\right),$$

which yields

$$\left\| m^{-1} \right\|_{L^\infty(\mathbb{T}\times[t_0,T-t_0])} \leq C\left(\frac{1}{t_0^{\frac{2+p\kappa\beta}{2|\lambda|}}} + \frac{1}{t_0^{\frac{1}{2|\lambda|}}} \right).$$

Now, recalling (3.6.6), we see that

$$\lim_{p\to\infty} \frac{1}{2|\lambda|} = 0 \text{ and } \lim_{p\to\infty} \frac{2+p\kappa\beta}{2|\lambda|} = \beta. \tag{3.6.17}$$

Thus, we may fix p chosen large enough that $\frac{2+\kappa\beta}{2|\lambda|} < \beta + \delta$, and, as a result of (3.6.17),

$$\left\| m^{-1} \right\|_{L^\infty(\mathbb{T}\times[t_0,T-t_0])} \leq C\frac{1}{t_0^{\beta+\delta}}.$$

This implies (3.6.2). Now, for the case of (MFG), we simply observe that, from Lemma 3.6.2, the map (3.6.14) is non-increasing on $[0,T]$, and, thus, (3.6.15) may be strengthened to

$$\left\| m^{-1} \right\|^{2|\lambda|}_{L^\infty(\mathbb{T}\times[t_0,T])} \leq \frac{2}{t_0}\int_{\frac{t_0}{2}}^{T} \left\| m^{-1} \right\|^{2|\lambda|}_{\infty}(t)dt.$$

□

The following lemma is a basic computation exploiting (E1), and will be used in the proof of Theorem 3.1.2 to estimate the terms arising from the Lasry-Lions monotonicity method.

Lemma 3.6.4. *There exists a constant $C = C(C_0) > 0$ such that, given $-\infty < p_0 < p_1 < \infty$ and $0 < m_0 < m_1 < \infty$, we have*

$$\begin{aligned}\Big(m_1H_p(p_1,m_1) - m_0H_p(p_0,m_0)\Big)(p_1-p_0) - (H(p_1,m_1) - H(p_0,m_0))(m_1-m_0)\\ \geq \frac{m_1+m_0}{C}(p_1-p_0)^2 + \frac{k}{C}(m_1-m_0)^2,\end{aligned} \tag{3.6.18}$$

where $k = \min_{[p_0,p_1]\times[m_0,m_1]}(-H_m(p,m))$. *Moreover, if* H *satisfies* (HW)*, then*

$$\Big(m_1 H_p(p_1,m_1) - m_0 H_p(p_0,m_0)\Big)(p_1-p_0) - (H(p_1,m_1)-H(p_0,m_0))(m_1-m_0) \geq \frac{m_1+m_0}{C}(p_1-p_0)^2 + \frac{1}{C(s+1)}(m_1^{s+1}-m_0^{s+1})(m_1-m_0). \quad (3.6.19)$$

Proof. Following the technique carried out in [301], for $z \in [0,1]$, we define

$$\Delta p = p_1 - p_0 \ \Delta m = m_1 - m_0,\ p_z = p_0 + z\Delta p,\ m_s = m_0 + z\Delta m.$$

We then let

$$\phi(z) = (m_z H_p(p_z,m_z) - m_0 H_p(p_0,m_0))\Delta p - (H(p_z,m_z) - H(p_0,m_0))\Delta m,$$

and differentiation yields

$$\phi'(z) = m_z H_{pp}(\Delta p)^2 + m_z H_{mp}\Delta m \Delta p - H_m(\Delta m)^2.$$

Now, in view of (E1), we have, for some constant $C > 0$,

$$-H_m \geq \frac{1}{4H_{pp}} m_z H_{mp}^2 (1+\frac{1}{C}) - \frac{1}{C}H_m.$$

Therefore,

$$\phi'(z) \geq m_z \left(\frac{1}{\sqrt{1+\frac{1}{C}}}\sqrt{H_{pp}}\Delta p + \frac{\sqrt{1+\frac{1}{C}}}{2\sqrt{H_{pp}}} H_{mp}\Delta m \right)^2 + m_z H_{pp}(\Delta p)^2 (1 - \frac{1}{1+\frac{1}{C}}) - \frac{1}{C}H_m(\Delta m)^2. \quad (3.6.20)$$

If (W) holds, then, up to increasing the constant $C > 0$, as well as using (H1) and (HW), we obtain

$$\phi'(z) \geq \frac{1}{C}(m_z(\Delta p)^2 + m_z^s(\Delta m)^2),$$

and integrating over $[0, 1]$ then yields (3.6.19). The proof of (3.6.18) follows from (3.6.20) in the same way. □

Before proving Theorem 3.1.2, we remind the reader that assumption (M) will not be in place, and will be instead replaced by (W).

Proof of Theorem 3.1.2. For $\epsilon \in (0, 1)$, let m_0^ϵ, m_T^ϵ be smooth, positive densities such that, for $\theta \in \{0, T\}$,

$$m_\theta^\epsilon \to m_\theta \text{ a.e. in } \mathbb{T},\ \|m_\theta^\epsilon\|_\infty \leq C \text{ and } \|(m_\theta^\epsilon)^{-\kappa}\|_1 \leq C,$$

where $C > 0$ is a constant independent of ϵ. Let $(u^{\epsilon,1}, m^{\epsilon,1})$ be a smooth solution to (MFGP) obtained from taking m_0^ϵ and m_T^ϵ, respectively, as the initial and terminal densities. Similarly, let $(u^{\epsilon,2}, m^{\epsilon,2})$ be the smooth solution to (MFG) corresponding to the initial density m_0^ϵ. The existence and regularity of such solutions is guaranteed by Theorem 3.1.1. We may further choose the $u^{\epsilon,1}$ to be normalized so that $\fint_{\mathbb{T}} u^{\epsilon,1}(T) = 0$.

As in the proof of Proposition 3.6.3, we obtain, for some $C > 0$ independent of ϵ and for $i \in \{1, 2\}$,

$$\|(m^{\epsilon,i})^{-\kappa}\|_1 \leq C. \tag{3.6.21}$$

On the other hand, Corollary 3.3.2 and Proposition 3.4.1 yield

$$\|m^{\epsilon,i}\|_\infty \leq C, \tag{3.6.22}$$

and (3.6.22), (HW) and Proposition 3.6.3 imply that

$$\int_0^T |H(0, \min_{\mathbb{T}} m^{\epsilon,i}(s)|ds \leq C. \tag{3.6.23}$$

Thus, as a result of (GW), Proposition 3.4.1, and Proposition 3.4.2,

$$\|u^{\epsilon,i}\|_\infty \leq C. \tag{3.6.24}$$

We will first observe that, up to a subsequence, there is convergence to a weak solution. Indeed, given $0 < \epsilon, \epsilon' < 1$, applying the Lasry-Lions monotonicity method to the corresponding systems yields, for $i \in \{1, 2\}$,

$$\begin{aligned}\int_{\mathbb{T}} (u^{\epsilon,i}(T) - u^{\epsilon',i}(T))(m^{\epsilon,i}(T) - m^{\epsilon',i}(T)) - \int_{\mathbb{T}} (u^{\epsilon,i}(0) - u^{\epsilon',i}(0))(m^{\epsilon,i}(0) - m^{\epsilon',i}(0)) \\ + \int\int_{Q_T} \left(m^{\epsilon,i} H_p(u_x^{\epsilon,i}, m^{\epsilon,i}) - m^{\epsilon',i} H_p(u_x^{\epsilon',i}, m^{\epsilon',i}) \right)(u_x^{\epsilon,i} - u_x^{\epsilon',i}) \\ - \left(H(u_x^{\epsilon,i}, m^{\epsilon,i}) - H(u_x^{\epsilon',i}, m^{\epsilon',i}) \right)(m^{\epsilon,i} - m^{\epsilon',i}) = 0.\end{aligned} \tag{3.6.25}$$

Lemma 3.6.4 therefore yields

$$\begin{aligned}\int_{\mathbb{T}} (u^{\epsilon,i}(T) - u^{\epsilon',i}(T))(m^{\epsilon,i}(T) - m^{\epsilon',i}(T)) - \int_{\mathbb{T}} (u^{\epsilon,i}(0) - u^{\epsilon',i}(0))(m^{\epsilon,i}(0) - m^{\epsilon',i}(0)) \\ + \int\int_{Q_T} \left(\frac{m^{\epsilon,i} + m^{\epsilon',i}}{C} (u_x^{\epsilon,i} - u_x^{\epsilon',i})^2 + \frac{1}{C(s+1)} ((m^{\epsilon,i})^{s+1} - (m^{\epsilon',i})^{s+1})(m^{\epsilon,i} - m^{\epsilon',i}) \right) \leq 0.\end{aligned} \tag{3.6.26}$$

Proceeding as in [316, Thm. 1.2], it readily follows that, for $i \in \{1, 2\}$, as $\epsilon \to 0$, $(u^{\epsilon,i}, m^{\epsilon,i})$ converges to a weak solution (u^i, m^i).

It remains to show the interior regularity. For $\delta > 0$, we define

$$I_{1,\delta} = [\delta, T - \delta], \;\; I_{2,\delta} = [\delta, T].$$

By Proposition 3.6.3, there exists $C = C(\delta^{-1})$ such that, for $t_i \in I_{i,\delta/4}$,

$$m^{\epsilon,i}(\cdot, t_i) \geq \frac{1}{C}. \tag{3.6.27}$$

We must first obtain a priori gradient bounds for $u^{\epsilon,i}$ on $I_{i,\delta/2}$. Setting

$$\phi_1(t) = (t - \delta/4)^{-2/(\gamma-1)} + (T - \delta/4 - t)^{-2/(\gamma-1)} \, \phi_2(t) = (t - \delta/4)^{-2/(\gamma-1)},$$

we go through the steps of Proposition 3.4.3, replacing the function v by

$$v_i(x,t) = \frac{1}{2}(u_x^{\epsilon,i})^2 + \frac{1}{2}(\widetilde{u}^{\epsilon,i})^2 - K\phi_i(t),$$

where $K > 0$, $\widetilde{u}^{\epsilon,i}$ is defined as in Proposition 3.4.3. We consider the maximum point (x_0, t_0) of v_i in $\mathbb{T} \times I_{i,\delta/4}$. In the case of (MFGP), namely $i = 1$, this maximum must be attained in the interior of I_i, since ϕ_i is unbounded near the endpoints. When $i = 2$, the maximum may be attained at $t = T$, and the proof that $|p| \leq C$ in this case follows through unchanged from Case 1 of Proposition 3.4.3. If the maximum is achieved at an interior time, the steps of Proposition 3.4.3 yield that if $v_i(x_0, t_0)$ is large enough, then

$$0 \leq -|p|^{2\gamma} + |p|^{2\gamma-2} - K(-\phi_i'' + \frac{1}{C}K^\gamma \phi_i^\gamma - C\phi_i).$$

Similarly to Proposition 3.6.3, we see that, if K is chosen large enough, ϕ_i must be a supersolution to

$$-\phi_i'' + \frac{1}{C}K^\gamma \phi_i^\gamma - C\phi_i = 0,$$

which then implies $p \leq C$, and thus $|u_x^{\epsilon,i}|$ is bounded on $I_{i,\delta/2}$. In view of (3.6.27) and (3.6.22), $|u_t^{\epsilon,i}| = |H(u_x^{\epsilon,i}, m^{\epsilon,i})|$ is also bounded on $I_{i,\delta/2}$. That is, we have

$$\|u^{\epsilon,i}\|_{C^1(\mathbb{T}\times I_{i,\delta/2})} \leq C. \tag{3.6.28}$$

The interior $C^{1,\alpha}$-estimates for quasilinear elliptic equations (see [210, Chapter 13, Thm. 13.6]), followed by the interior Schauder estimates (see [280, Chapter, 2, (1.12)]) then yield, for some $C = C(\delta^{-1})$, and for $i \in \{1, 2\}$,

$$\|u^{\epsilon,i}\|_{C^{3+\alpha}(\mathbb{T}\times I_{1,\delta})} \leq C. \tag{3.6.29}$$

For $i = 1$, by virtue of the Arzelà–Ascoli theorem, we may finish the proof by simply letting $\epsilon \to 0$. On the other hand, for $i = 2$ (that is, the case of (MFG)), we require estimates up to the terminal time T. We first observe that (3.6.27), (3.6.22), and (3.6.29) imply that $u^{\epsilon,2}$ solves, in $I_{2,\delta} \times \mathbb{T}$, a system of the form (MFG), where the initial density $m^{\epsilon,2}(\cdot, \delta)$ is bounded below by a positive constant, and bounded above in $C^{2,\alpha}(\mathbb{T})$. Moreover, as in Lemma 3.5.1, (3.6.28) implies that $u^{\epsilon,2}$ is bounded in $C^{1,\beta}$ for some $0 < \beta < 1$. We may now conclude through the same convergence argument as in the proof of Theorem 3.1.1. □

Finally, by requiring some further regularity on the marginals, we establish additional Sobolev regularity for the weak solutions.

Proposition 3.6.5. *Let m_0, m_T satisfy $(m_0)_{xx}, (m_T)_{xx} \in L^1(\mathbb{T})$. Let (u, m) be a weak solution to* (MFG) *or* (MFGP) *under the assumptions of Theorem 3.1.2. Then, for some constant $C > 0$ we have:*

- *In the case of* (MFG)*,*

$$\int_{\mathbb{T}} g'(m(x,T))|m_x(x,T)|^2 + \int_0^T \int_{\mathbb{T}} mH_{pp}(u_{xx})^2 + m^s(m_x)^2 dxdt \leq C, \tag{3.6.30}$$

where $C = C(\|u\|_\infty, \|(m_0)_{xx}\|_1, C_0)$.

- *In the case of* (MFGP),

$$\int_0^T \int_{\mathbb{T}} mH_{pp}(u_{xx})^2 + m^s(m_x)^2 dxdt \leq C, \tag{3.6.31}$$

where $C = C(\|u\|_\infty, \|(m_0)_{xx}\|_1, \|(m_T)_{xx}\|_1, C_0)$.

Proof. We will show the result in the case where (u, m) is smooth, since the general case follows by considering the approximations employed in the proof of Theorem 3.1.2. Differentiating with respect to x the (MFG) or (MFGP), we obtain

$$\begin{cases} -u_{xt} + H_p(u_x, m)u_{xx} + H_m(u_x, m)m_x = 0 \text{ in } Q_T, \\ m_{xt} - (m_x H_p(u_x, m) + mH_{pp}(u_x, m)u_{xx} + mH_{pm}(u_x, m)m_x)_x = 0 \text{ in } Q_T. \end{cases} \tag{3.6.32}$$

Testing against u_x in the equation for m_x above we obtain

$$\int_{\mathbb{T}} m_x(T)u_x(T) - \int_{\mathbb{T}} m_x(0)u_x(0) + \int_0^T \int_{\mathbb{T}} (m_x(-u_{xt} + u_{xx}H_p(u_x, m)) + mu_{xx}^2 H_{pp}(u_x, m) + mu_{xx}H_{pm}(u_x, m)m_x) = 0, \tag{3.6.33}$$

and, therefore,

$$\int_{\mathbb{T}} m_x(T)u_x(T) + \int_0^T \int_{\mathbb{T}} mu_{xx}^2 H_{pp} - H_m(m_x)^2 = -\int_{\mathbb{T}} u(0)(m_0)_{xx}dx - \int_0^T \int_{\mathbb{T}} mu_{xx}H_{pm}m_x. \tag{3.6.34}$$

Now, observe that

$$\Big|\int_{\mathbb{T}} u(0)(m_0)_{xx}dx\Big| \leq \|u\|_\infty \|(m_0)_{xx}\|_1, \qquad \Big|\int_{\mathbb{T}} u(T)(m_T)_{xx}dx\Big| \leq \|u\|_\infty \|(m_T)_{xx}\|_1. \tag{3.6.35}$$

Additionally, as a result of (E1), we infer that, for $\delta \in (0,1)$,

$$\begin{aligned}\left|mu_{xx}H_{pm}m_x\right| &\leq (1-\delta)mu_{xx}^2H_{pp} + \frac{1}{4(1-\delta)H_{pp}}m|H_{pm}|^2(m_x)^2 \\ &\leq (1-\delta)mu_{xx}^2H_{pp} - \frac{1}{(1-\delta)(1+\frac{1}{C_0})}H_m(m_x)^2. \quad (3.6.36)\end{aligned}$$

We choose $\delta > 0$ small enough so that

$$\frac{1}{(1-\delta)(1+\frac{1}{C_0})} < 1.$$

Using (3.6.35) and (3.6.36) in (3.6.34), we obtain the following. In the case of (MFG), we have

$$\int_{\mathbb{T}} g'(m(T))(m_x(T))^2dx + \int_0^T \int_{\mathbb{T}} mH_{pp}(u_{xx})^2dx - H_m(m_x)^2dx \leq C$$

while in the case of (MFGP), we have

$$\int_0^T \int_{\mathbb{T}} mH_{pp}(u_{xx})^2dx - H_m(m_x)^2dx \leq C.$$

We conclude by using the fact that H satisfies (HW). □

3.7 Long time behavior and the infinite horizon problem

In this section, we will characterize the behavior, as $T \to \infty$, of solutions to (MFG) and (MFGP). First, we establish the turnpike property with an exponential rate of convergence. This property shows that, for large values of T, the players spend most of their time close to the equilibrium $m \equiv 1$.

Lemma 3.7.1. *Let* (u,m) *be a solution to* (MFG) *or* (MFGP)*, let* $T > 1$*, and set*

$$c_1 = \min(\min m_0, \min m_T), \;\; C_1 = \max(\max m_0, \max(m_T)).$$

Then there exist constants $C, \omega > 0$, with

$$C = C(C_0, C_1, c_1^{-1}, \|\overline{C}\|_{L^\infty([c_1,C_1])}, \|(m_0)_x\|_\infty, \|(m_T)_x\|_\infty, \|(g')^{-(\gamma-1)}\|_{L^\infty([\min m_0, \max m_0])})$$

and

$$\omega^{-1} = \omega^{-1}(C_0, c_1^{-1}, C_1, \|\overline{C}\|_{L^\infty([c_1,C_1])}),$$

such that

$$\|m(t) - 1\|_{L^\infty(\mathbb{T})} + \|u_x(t)\|_{L^\infty(\mathbb{T})} \leq C(e^{-\omega t} + e^{-\omega(T-t)}), \quad t \in [0,T]. \tag{3.7.1}$$

If (u, m) *solves* (MFG)*, and* (3.1.3) *holds, we have*

$$\|m(t) - 1\|_{L^\infty(\mathbb{T})} + \|u_x(t)\|_{L^\infty(\mathbb{T})} \leq Ce^{-\omega t}, \quad t \in [0,T]. \tag{3.7.2}$$

Proof. As in previous arguments, we recall that the constant C may increase at each step. For each $k \in \mathbb{N}$, Proposition 3.3.1 yields

$$\frac{d^2}{dt^2}\int_{\mathbb{T}}(m-1)^{2k}dx \geq 0, \tag{3.7.3}$$

and, as a result of (L) and Corollary 3.3.2,

$$\frac{d^2}{dt^2}\int_{\mathbb{T}}(m-1)^2 dx \geq \int_{\mathbb{T}} -2mH_mH_{pp}m_x^2 dx \geq \frac{1}{C}\int_{\mathbb{T}} |(m-1)_x|^2\, dx.$$

Since $⨍_{\mathbb{T}} m(\cdot, t) \equiv 1$, arguing in the same way as in (3.6.8), we obtain

$$\frac{d^2}{dt^2}\int_{\mathbb{T}}(m-1)^2 dx \geq \frac{1}{C}\left\|m-1\right\|_\infty^2.$$

Therefore, setting

$$\phi(t) := \int_{\mathbb{T}}(m(t)-1)^2 dx,$$

we have

$$-\phi'' + \frac{1}{C}\phi \leq 0. \tag{3.7.4}$$

Moreover, if (u, m) solves (MFG) and (3.1.3) holds, up to increasing the value of C, Lemma 3.6.2 implies that

$$\phi'(T) \leq -\frac{1}{\sqrt{C}}\phi(T). \tag{3.7.5}$$

We now fix the choice $\omega = \frac{1}{2\sqrt{C}}$ (the value of C may still increase in subsequent steps, but the value of ω will not). The comparison principle applied to (3.7.4) then implies that, for each $t \in [0, T]$,

$$\phi(t) \leq \phi(0)e^{-2\omega t} + \phi(T)e^{-2\omega(T-t)} \leq C(e^{-2\omega t} + e^{-2\omega(T-t)}). \tag{3.7.6}$$

Similarly, if (u, m) solves (MFG) and (3.1.3), then (3.7.4), coupled with the Robin boundary condition (3.7.5), readily implies that

$$\phi(t) \leq \phi(0)e^{-2\omega t} \leq Ce^{-2\omega t}. \tag{3.7.7}$$

By using the same convexity arguments as in (3.6.16), in view of (3.7.3), we have

$$\|m(t) - 1\|_\infty^2 \leq C \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} \|m(s) - 1\|_\infty(s)^2 ds \leq C \int_{t-1}^{t+1} \int_{\mathbb{T}} (m-1)^2 = C \int_{t-1}^{t+1} \phi(s)ds. \tag{3.7.8}$$

We now turn our attention to estimating u_x. Fixing $t \in [1, T-1]$, as a result of (H1), Proposition 3.3.1, and Corollary 3.3.2, we obtain, for $s \in [t-1, t+1]$,

$$\frac{1}{C}\int_{\mathbb{T}} u_{xx}^2(s) \leq \frac{d^2}{ds^2}\int_{\mathbb{T}} (m(s) - 1)^2.$$

Thus, testing against a bump function $\zeta \geq 0$, which is supported on $[t-1, t+1]$, and identically equals 1 on $[t-\frac{1}{2}, t+\frac{1}{2}]$, we get

$$\int_{t-\frac{1}{2}}^{t+\frac{1}{2}} \int_{\mathbb{T}} u_{xx}^2 \le C \int_{t-1}^{t+1} \int_{\mathbb{T}} (m-1)^2 \zeta'' \le C \int_{t-1}^{t+1} \phi(s) ds. \tag{3.7.9}$$

Differentiating (Q) with respect to x, one sees that $v = u_x$ solves a linear elliptic equation of the form

$$-\mathrm{Tr}(A(x,t)D^2 v) + b(x,t) \cdot Dv = 0.$$

Thus, v satisfies the maximum and minimum principles on compact subsets of $\overline{Q}_T$. Applying this observation to $\mathbb{T} \times [t-s, t+s]$, for $s \in (0, \frac{1}{2})$, as well as the fact that, for every $t \in [0,T]$, $\{x \in \mathbb{T} : u_x(x,t) = 0\} \neq \emptyset$, we have

$$\mathrm{osc}_{\mathbb{T}} v(t) \le \mathrm{osc}_{\mathbb{T}} v(t+s) + \mathrm{osc}_{\mathbb{T}} v(t-s) \le \int_{\mathbb{T}} |u_{xx}(t+s)| + \int_{\mathbb{T}} |u_{xx}(t-s)|.$$

Integrating in s then yields

$$\mathrm{osc}_{\mathbb{T}} u_x(t) \le \int_{t-\frac{1}{2}}^{t+\frac{1}{2}} \int_{\mathbb{T}} |u_{xx}|,$$

and, thus, as a result of (3.7.9) and the Cauchy-Schwarz inequality,

$$\|u_x(t)\|_\infty^2 \le C \int_{t-1}^{t+1} \phi(s) ds. \tag{3.7.10}$$

Now, adding (3.7.8) and (3.7.10), followed by (3.7.6), we obtain (3.7.1) for $t \in [1, T-1]$. Similarly, when (u,m) solves (MFG) and (3.1.3) holds, (3.7.7) yields (3.7.2) for $t \in [1, T-1]$. We observe that, for $t \in [0,T] \backslash [1, T-1]$, the bounds on $\|m(t)-1\|_\infty$ given by (3.7.1) and (3.7.2) hold trivially, up to increasing the value of C. Let us see that the same is true for the bounds on $\|u_x(t)\|_\infty$ on the interval $[0,1]$. Indeed, we may simply follow the proof of Proposition 3.4.3, applied to the MFG system on the domain $\mathbb{T} \times [0,1]$, with the only change being on Case 1 of that proof, that is, when the maximum value is attained at $t = 1$. For this case, we may simply use the fact that, as a result of (3.7.1) holding for $t = 1$, $|u_x(\cdot, 1)|$ is bounded. Thus, if we take $T = 1$ in Proposition 3.4.2, this yields a bound on $\|u_x\|_{\mathbb{T}\times[0,1]}$ that depends only on $C_0, \|m\|_{L^\infty(\overline{Q}_T)}, \|m^{-1}\|_{L^\infty(\overline{Q}_T)}, \|(m_0)_x\|_\infty$, and $\|\overline{C}\|_{L^\infty([\min m, \max m])}$. A similar argument may be followed on $\mathbb{T}\times[T-1, T]$, which completes the proof. ◻

Having established the turnpike property, we now follow the program developed in [122] to study the long time behavior. In order to characterize the limit, as $T \to \infty$, of the functions $(u(t) - \lambda(T - t), m(t))$, we first show a uniqueness result for (MFGL).

Lemma 3.7.2. *Assume that (L) holds. Then, up to adding a constant to v, there exists at most one classical solution* (v, μ) *to* (MFGL) *satisfying* (3.1.5).

Proof. Assume that $(v^1, \mu^1), (v^2, \mu^2)$ are solutions to (MFGL) satisfying (3.1.5). Since $\mu^1 - 1, \mu^2 - 1 \in L^1(\mathbb{T}\times(0,\infty))$, there exists a sequence $T_k \to \infty$ such that

$$\lim_{k\to\infty}\int_{\mathbb{T}}\left(|\mu^1(\cdot, T_k) - 1| + |\mu^2(\cdot, T_k) - 1|\right) = 0.$$

Performing the standard Lasry-Lions computation for v^1, v^2 on Q_{T_k}, using Lemma 3.6.4, and noting that

$$\mu^i, (\mu^i)^{-1}, v^i_x, \in L^\infty(\mathbb{T}\times(0,\infty)), \quad i \in \{1, 2\},$$

we obtain

$$\begin{aligned}\frac{1}{C}\left(\int_0^{T_k}\int_{\mathbb{T}}|v^1_x - v^2_x|^2 + |\mu^1 - \mu^2|^2\right) &\le \int_{\mathbb{T}} -(v^1(T_k) - v^2(T_k))(\mu^1(T_k) - \mu^2(T_k)) \\ &= \int_{\mathbb{T}} -(v^1(T_k) - v^2(T_k))((\mu^1(T_k) - 1) - (\mu^2(T_k) - 1)). \quad (3.7.11)\end{aligned}$$

Now, since $v^1, v^2 \in L^\infty(\mathbb{T}\times(0,\infty))$, the right hand side converges to 0 as $k \to \infty$. Therefore,

$$\int_0^\infty\int_{\mathbb{T}}|v^1_x - v^2_x|^2 + |\mu^1 - \mu^2|^2 = 0.$$

This implies that $\mu^1 = \mu^2$ and $v_x^1 = v_x^2$. From the HJ equations, $v_t^1 = v_t^2$, which concludes the proof. ◻

In the following lemma, we obtain uniform estimates for the solution that are independent of T.

Lemma 3.7.3. *Let* (u^T, m^T) *be a solution to* (MFG) *or* (MFGP) *for* $T > 0$*, and let* $\omega > 0$ *be the constant from Lemma 3.7.1. Set* $v^T = u^T - \lambda(T-t)$*. Then there exists a constant* $C > 0$*, independent of* T*, such that:*

- *If* (3.1.3) *holds and* (u^T, m^T) *solves* (MFG)*, then*

$$|v^T(t) - g(1)| \leq Ce^{-\omega t} \text{ for all } t \in [0, T]. \tag{3.7.12}$$

- *If* (u^T, m^T) *solves* (MFGP)*, and*

$$\int_{\mathbb{T}} v^T\left(\frac{1}{2}T\right) dx = 0, \tag{3.7.13}$$

  then we have

$$\|v^T\|_{L^\infty(Q_T)} \leq C \tag{3.7.14}$$

  and

$$\|v^T(t)\|_\infty \leq Ce^{-\omega t} \text{ for all } t \in \left[0, \frac{T}{2}\right]. \tag{3.7.15}$$

Proof. First we note that in both (MFG) and (MFGP), as a result of Lemma 3.7.1, the function $v_x^T = u_x^T$ is bounded uniformly, independently of T, and, by Corollary 3.3.2, so are $m^T, (m^T)^{-1}$. Therefore, since H is smooth, and thus locally Lipschitz, we have, for some constant $C > 0$ independent of $T > 0$,

$$|v_t^T| \leq C(|v_x^T| + |m^T - 1|). \tag{3.7.16}$$

Assume first that (u^T, m^T) solves (MFG) and (3.1.3) holds. Integrating the HJ equation in $[t, T]$

and using (3.7.16) along with (3.7.2) in Lemma 3.7.1 we obtain

$$|v^T(t) - v^T(T)| \leq C \int_t^T e^{-\omega s} ds.$$

Furthermore, using the fact that

$$v^T(T) = u^T(T) = g(m^T(T)),$$

and

$$|m^T(T) - 1| \leq Ce^{-\omega T},$$

by increasing the constant C if necessary, we obtain

$$|v^T(t) - g(1)| \leq C(e^{-\omega T} + e^{-\omega t}) \leq 2Ce^{-\omega t},$$

which proves (3.7.12). Next, we assume that (u^T, m^T) solves (MFGP) and (3.7.13) holds. Letting $t < \frac{T}{2}$, and integrating the HJ equation in $[t, \frac{T}{2}]$, we obtain from (3.7.16) and (3.7.1) that

$$\Big| \int_{\mathbb{T}} v^T(\cdot, t) \Big| \leq C \int_t^{\frac{T}{2}} e^{-\omega s} + e^{-\omega(T-s)} ds \leq \frac{2C}{\omega} \Big(e^{-\omega t} + e^{-\omega \frac{T}{2}} \Big) \leq \frac{4C}{\omega} e^{-\omega t}. \tag{3.7.17}$$

Similarly, for $t \geq \frac{T}{2}$ integrating the HJ equation in $[\frac{T}{2}, t]$ yields

$$\Big| \int_{\mathbb{T}} v^T(\cdot, t) dx \Big| \leq C. \tag{3.7.18}$$

Now, for every $t \in [0, T]$, there exists a point $x_t \in \mathbb{T}$ such that $v^T(x_t, t) = \int_{\mathbb{T}} v^T(\cdot, t)$. Therefore,

$$|v^T(x, t)| \leq \operatorname{osc}_{\mathbb{T}} v^T(t) + \Big| \int_{\mathbb{T}} v^T(\cdot, t) \Big|.$$

As a result, in view of (3.7.1), the estimates (3.7.18) and (3.7.17) yield, respectively, (3.7.14) and

(3.7.15). ◻

We are now ready to prove our last result.

Proof of Theorem 3.1.3. We set

$$v^T = u^T - \lambda(T - t),$$

and show that v^T is convergent as $T \to \infty$.

In view of Lemmas 3.7.1 and 3.7.3, as well as (3.7.16), we see that $\|v^T\|_{W^{1,\infty}(Q_T)}$ and $\|m^T\|_\infty$ are bounded, independently of T. We may therefore apply the Arzelà–Ascoli theorem to conclude that, up to extracting a subsequence, there exist $v \in W^{1,\infty}(\mathbb{T} \times [0,\infty))$ and $\mu \in L^\infty(\mathbb{T} \times [0,\infty))$ such that

$$v^T \to v \text{ locally uniformly in } \mathbb{T} \times [0,\infty),$$

and

$$m^T \rightharpoonup \mu \text{ weakly–* in } L^\infty(\mathbb{T} \times (0,\infty)).$$

We now fix $T_0 \in (1,\infty)$, and assume that $T > T_0 + 1$. Then (v^T, m^T) solves the system

$$\begin{cases} -v_t^T + \lambda + H(v_x^T, m^T) = 0 & \text{in } Q_{T_0}, \\ m_t^T - (m^T H_p(v_x^T, m^T))_x = 0 & \text{in } Q_{T_0}, \\ m^T(\cdot, 0) = m_0. \end{cases} \tag{3.7.19}$$

Moreover, as a result of the interior $C^{1,\alpha}$ estimates for quasilinear elliptic equations, and the interior Schauder estimates for linear equations, $m^T(\cdot, T_0)$ is uniformly bounded in $C^{2,\alpha+\epsilon}$, where $\epsilon > 0$ is chosen such that $\alpha + \epsilon < 1$. Therefore, as in the proof of Theorem 3.1.1, we conclude that, as $T \to \infty$,

$$(v^T, m^T) \to (v, \mu) \text{ in } C^{3,\alpha}(\mathbb{T} \times [0, T_0]) \times C^{2,\alpha}(\mathbb{T} \times [0, T_0]). \tag{3.7.20}$$

In particular, this implies that $(v,\mu) \in C^{3,\alpha}_{\text{loc}}(\mathbb{T} \times [0,\infty)) \times C^{2,\alpha}_{\text{loc}}(\mathbb{T} \times [0,\infty))$, and that (v,μ) solves

(MFGL). Letting $T \to \infty$ in (3.7.1) yields

$$\|\mu(t) - 1\|_\infty + \|v_x(t)\|_\infty \le Ce^{-\omega t}, \tag{3.7.21}$$

which shows that $\mu - 1 \in L^1(\mathbb{T} \times (0, \infty))$. Moreover, since $\|(m^T)^{-1}\|_\infty$ is bounded, we conclude that (3.1.5) holds.

Now, since a subsequence was extracted, we must verify that the limit is uniquely determined. In view of Lemma 3.7.2, μ is uniquely determined, and v is uniquely determined up to a constant. In the case of (MFG) we see from (3.7.12) that

$$\lim_{t\to\infty} \|v(t) - g(1)\|_\infty = 0.$$

On the other hand, in the case of (MFGP), letting $T \to \infty$ followed by $t \to \infty$ in (3.7.15), we obtain

$$\lim_{t\to\infty} \|v(t)\|_\infty = 0.$$

□

CHAPTER 4

SHARP RATES OF CONVERGENCE IN MEAN FIELD CONTROL

4.1 Introduction

The work presented in this chapter is the collaboration with P. Cardaliaguet, J. Jackson and P. Souganidis in [84].

This chapter is concerned with the convergence of certain high-dimensional stochastic control problems towards their mean field limits. To define these control problems, we fix throughout the chapter a dimension $d \in \mathbb{N}$, a time horizon $T > 0$, and a filtered probability space $(\Omega, \mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$ satisfying the usual conditions and hosting independent d-dimensional Brownian motions W and $(W^i)_{i \in \mathbb{N}}$.

The data consists of nice functions

$$L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad \mathcal{F}, \mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R},$$

where $\mathcal{P}_2(\mathbb{R}^d)$ is the Wasserstein space of Borel probability measures on $\mathbb{R}^d$ with finite second moment. Precise assumptions on L, $\mathcal{F}$, and $\mathcal{G}$ will be introduced in Subsection 4.2.2 below.

The N-particle value function $\mathcal{V}^N : [0, T] \times (\mathbb{R}^d)^N \to \mathbb{R}$ is defined by the formula

$$\mathcal{V}^N(t_0, \boldsymbol{x}_0) = \inf_{\alpha \in \mathcal{A}^N} \mathbb{E}\Big[\int_{t_0}^T \Big(\frac{1}{N} \sum_{i=1}^N L(X_t^i, \alpha_t^i) + \mathcal{F}(m^N_{\boldsymbol{X}_t}) \Big) dt + \mathcal{G}(m^N_{\boldsymbol{X}_T}) \Big], \tag{4.1.1}$$

where $\mathcal{A}^N$ is the set of square-integrable, $\mathbb{F}$-adapted, $(\mathbb{R}^d)^N$-valued processes $\boldsymbol{\alpha} = (\alpha^1, ..., \alpha^N)$ defined on $[t_0, T]$, and $\boldsymbol{X} = (X^1, ..., X^N)$ is the $(\mathbb{R}^d)^N$-valued state process which is determined from the control $\boldsymbol{\alpha}$ by the dynamics

$$dX_t^i = \alpha_t^i dt + \sqrt{2} dW_t^i, \quad t_0 \leq t \leq T, \quad X_{t_0}^i = x_0^i.$$

We recall that under mild conditions on the data (in particular Assumption 1 below), $\mathcal{V}^N$ is the unique classical solution of the Hamilton-Jacobi-Bellman equation

$$\begin{cases} -\partial_t \mathcal{V}^N - \sum_{i=1}^N \Delta_{x^i} \mathcal{V}^N + \frac{1}{N} \sum_{i=1}^N H(x^i, N D_{x^i} \mathcal{V}^N) = \mathcal{F}(m_{\boldsymbol{x}}^N) \text{ in } [0,T] \times (\mathbb{R}^d)^N, \\ \mathcal{V}^N(T, \boldsymbol{x}) = \mathcal{G}(m_{\boldsymbol{x}}^N) \text{ for } \boldsymbol{x} \in (\mathbb{R}^d)^N, \end{cases} \tag{HJB$_N$}$$

with the Hamiltonian $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ given by $H(x,p) = \sup_{a \in \mathbb{R}^d} \{ -a \cdot p - L(x,a) \}$.

Next, we define the value function $\mathcal{U} : [0,T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ for the corresponding mean field problem by

$$\mathcal{U}(t_0, m_0) = \inf_{(m,\alpha)} \Big\{ \int_{t_0}^T \Big(\int_{\mathbb{R}^d} L(x, \alpha(t,x)) m_t(dx) + \mathcal{F}(m_t) \Big) dt + \mathcal{G}(m_T) \Big\}, \tag{4.1.2}$$

where the infimum is taken over all pairs (m, α) consisting of a curve $[t_0, T] \ni t \mapsto m_t \in \mathcal{P}_2(\mathbb{R}^d)$ and a measurable map $\alpha : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\begin{cases} \int_{t_0}^T \int_{\mathbb{R}^d} |\alpha(t,x)|^2 m_t(dx) dt < \infty, \text{ and the Fokker-Planck equation} \\ \partial_t m = \Delta m - \operatorname{div}(m\alpha) \text{ in } [t_0, T] \times \mathbb{R}^d, \quad m_{t_0} = m_0 \\ \text{is satisfied in the sense of distributions.} \end{cases} \tag{4.1.3}$$

We recall that $\mathcal{U}$ is expected to be the unique solution, in an appropriate viscosity sense, to the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t \mathcal{U} - \int_{\mathbb{R}^d} \operatorname{tr}(D_x D_m \mathcal{U}) dm + \int_{\mathbb{R}^d} H(x, D_m \mathcal{U}) dm = \mathcal{F}(m) \text{ in } [0,T] \times \mathcal{P}_2(\mathbb{R}^d), \\ \mathcal{U}(T, m) = \mathcal{G}(m) \text{ in } \mathcal{P}_2(\mathbb{R}^d); \end{cases} \tag{HJB$_\infty$}$$

see e.g. [124, 340, 134, 32, 133] and the references therein for various approaches to the comparison principle for viscosity solutions of (HJB_∞).

4.1.1 Previous convergence results

It is by now well understood that $\mathcal{V}^N$ converges to $\mathcal{U}$ in the sense that

$$\mathcal{V}^N(t, \boldsymbol{x}) \approx \mathcal{U}(t, m_{\boldsymbol{x}}^N), \quad \text{for } N \text{ large.} \tag{4.1.4}$$

This convergence was first established in [277], and later extended in [153] to allow the presence of a common noise. In another direction, [201] and [308] used PDE techniques to obtain similar results in a setting with purely common noise. We refer also to the works [197] and [106] for a study of the deterministic case via Γ-convergence techniques, to [129] for an extension of the methods in [277] to problems with state constraints, and to [348] for a similar convergence result in the setting of mean field optimal stopping. All the works mentioned in the preceding paragraph use techniques based on compactness, and so obtain only qualitative versions of the statement (4.1.4).

More recently, there have been a number of attempts to quantify the convergence of $\mathcal{V}^N$ to $\mathcal{U}$. On the one hand, when $\mathcal{F}$ and $\mathcal{G}$ are convex and sufficiently smooth, the value function $\mathcal{U}$ is smooth, and a standard argument (see the introduction of [78] for a more detailed explanation) shows that $|\mathcal{V}^N(t, \boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq C/N$. On the other hand, when $\mathcal{F}$ and $\mathcal{G}$ are not convex, then the value function $\mathcal{U}$ may fail to be C^1 even if all the data is smooth (see [63] for an example). In this setting the optimizers for the mean field control problem may not be unique, and obtaining quantitative convergence results is much more subtle.

The first general, that is, not requiring convexity or other special structure on $\mathcal{F}$ and $\mathcal{G}$, quantitative version of (4.1.4) was obtained in [78], where the authors proved the estimate

$$|\mathcal{V}^N(t,\boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq C(1 + \frac{1}{N}\sum_{i=1}^{N}|x^i|^2)N^{-\beta_d}, \tag{4.1.5}$$

for C depending on all of the data and the exponent β_d depending only on d. We refer also to [107] for a thorough treatment of the finite state space setting and to [25], which obtains a sharp rate but under a special structural condition on the data. The more recent work [131] attempts to identify the optimal rate of convergence, and shows in particular that the optimal rate depends on the smoothness of the data or, more precisely, the metric with respect to which the data is regular. Theorem 2.7 of [131] shows that, if the data is periodic, that is, the state space $\mathbb{R}^d$ is replaced by the d-dimensional flat-torus $\mathbb{T}^d$, and sufficiently smooth (with the amount of smoothness required depending on the dimension d), (4.1.5) can be improved to

$$|\mathcal{V}^N(t,\boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq CN^{-1/2}, \tag{4.1.6}$$

Example 2 in [131], meanwhile, shows that this rate cannot be improved even if all of the data is C^∞. In summary, we now know that when the data is smooth and convex $\mathcal{U}$ is smooth and the rate is $1/N$, but when the data is not convex, $\mathcal{U}$ may fail to be smooth, and in this case the global rate is at best $1/\sqrt{N}$ even if all the data is very regular.

There have also been some efforts to understand the convergence of the optimal trajectories and the optimal controls. For example, when $\mathcal{U}$ is smooth one can follow the strategy initiated in [79] to show that optimal trajectories of the N-particle control problem converge (with a rate) to optimal trajectories of the mean field problem (see [208] for details on this approach). In the non-convex regime, such questions are much more subtle since, as mentioned already, there may not be a unique optimal trajectory for the limiting problem. The recent work [92] overcomes this issue by identifying an open and dense subset $\mathcal{O}$ of $[0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ where the value function $\mathcal{U}$ is C^1 and such that optimal trajectories started from initial conditions in $\mathcal{O}$ are unique. In particular, it is shown in [92] that a quantitative propagation of chaos can be established when starting from initial conditions in $\mathcal{O}$.

4.1.2 Our results

The open and dense set $\mathcal{O}$ identified in [92] plays a central role in our results. In what follows, we will call the set $\mathcal{O}$ the region of strong regularity by analogy with the terminology in [193], where the same language is used to describe a region where a certain first-order Hamilton-Jacobi equation has a classical solution which can be computed via the method of characteristics; more on this analogy shortly.

Our first result states that, locally inside $\mathcal{O}$, the convergence rate $1/N$ can be achieved even if the data is not convex. More precisely, we show in Theorem 4.2.1 that, for each set $K \subset \mathcal{O}$ which is compact in $\mathcal{P}_p(\mathbb{R}^d)$ for some $p > 2$, with $\mathcal{P}_p(\mathbb{R}^d)$ denoting the p-Wasserstein space, there is a constant $C = C(K)$ such that, for each $N \in \mathbb{N}$ and each $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$ such that $(t, m_{\boldsymbol{x}}^N) \in K$,

$$|\mathcal{V}^N(t, \boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq C/N. \tag{4.1.7}$$

We refer to Remark 8 for a discussion of the role of compactness in $\mathcal{P}_p$ with $p > 2$. Combined with Example 2 in [131], this shows that the optimal global convergence rate is different than the optimal rate of convergence within $\mathcal{O}$.

Example 2 in [131] also explains why we claim that the set $\mathcal{O}$ plays a similar role as the regions of strong regularity in [193]. Indeed, it is explained there that when $\mathcal{F}$ and $\mathcal{G}$ depend on m only through its mean $\overline{m}$, that is, $\mathcal{F}(m) = f(\overline{m})$ and $\mathcal{G}(m) = g(\overline{m})$, and $L = \frac{1}{2}|a|^2$ for simplicity, we have

$$\mathcal{U}(t, m) = u(t, \overline{m}) \text{ and } \mathcal{V}^N(t, \boldsymbol{x}) = v^N(t, \frac{1}{N}\sum_{i=1}^N x^i),$$

where u and v^N are the solutions of the finite-dimensional PDEs

$$-\partial_t u + \frac{1}{2}|Du|^2 = f \text{ in } [0,T) \times \mathbb{R}^d \text{ and } u(T,x) = g,$$

and

$$-\partial v^N - \frac{1}{N}\Delta v^N + \frac{1}{2}|Dv^N|^2 = f \text{ in } [0,T) \times \mathbb{R}^d \text{ and } v^N(T,x) = g.$$

Thus, the convergence problem reduces to vanishing viscosity. Moreover, while the best global estimate for $|v^N - u|$ is $O(1/\sqrt{N})$, the expansion achieved in [193] clearly shows that, locally uniformly on "regions of strong regularity", $|v^N - u| = O(1/N)$. Thus our Theorem 4.2.1 can be viewed as an infinite-dimensional (partial) analogue of the results in [193], with $\mathcal{O}$ playing the role of the regions of strong regularity in [193].

Our second result shows that, when the data is smooth enough, a similar sharp rate of convergence can be obtained for the gradients. More precisely, Theorem 4.2.2 shows that, for each set $K \subset \mathcal{O}$ which is compact in $\mathcal{P}_p(\mathbb{R}^d)$ for some $p > 2$, there is a constant $C = C(K)$ such that, for each $N \in \mathbb{N}$ and each $(t, \boldsymbol{x}) \in [0,T] \times (\mathbb{R}^d)^N$ such that $(t, m_{\boldsymbol{x}}^N) \in K$,

$$|ND_{x^i}\mathcal{V}^N(t,\boldsymbol{x}) - D_m\mathcal{U}(t, m_{\boldsymbol{x}}^N, x^i)| \leq C/N. \tag{4.1.8}$$

The main interest of (4.1.8) is that it demonstrates that optimal feedbacks for the N-particle problem converge toward the optimal feedback for the mean field problem; see Remark 10 for more details.

Using the strong convergence of optimal feedbacks in (4.1.8), we obtain in Proposition 4.2.3 a concentration inequality for the optimal trajectories of the N-particle problems when started from appropriate i.i.d. initial conditions. This result complements the quantitative propagation of chaos results in [92], and can also be compared to similar concentration results for mean field games obtained in [143] under the assumption that the master equation has a smooth solution.

Finally, we mention that in order to obtain our main convergence results, we have to sharpen in various ways the regularity results in [92]. In particular we show in Theorem 4.2.4 that, under appropriate regularity conditions, the second Wasserstein derivative $D_{mm}\mathcal{U}$ exists and is continuous in the region of strong regularity.

4.1.3 Strategy of the proof

We explain here the strategy of proof for the estimate (4.1.4). To avoid unnecessary technicalities related to higher moments of the relevant probability measures, we only discuss here the periodic case, that is, $\mathbb{R}^d$ is replaced by the d-dimensional flat torus $\mathbb{T}^d$.

First, we note that, in view of the semi-concavity but not semi-convexity estimates for $\mathcal{U}$, the estimate

$$\mathcal{V}^N(t, \boldsymbol{x}) \leq \mathcal{U}(t, m_{\boldsymbol{x}}^N) + C/N$$

in fact holds globally.

To complete the proof, we need to show that the symmetric inequality holds locally uniformly in $\mathcal{O}$. For each fixed $(t_0, m_0) \in \mathcal{O}$ we work with small tubes $\mathcal{T}_r(t_0, m_0)$ of radius r around the optimal trajectory for the mean field control problem started from (t_0, m_0); see Subsection 4.3.1 for the precise definition of $\mathcal{T}_r(t_0, m_0)$. The key result is proved in Lemma 4.3.6. It says that, when $0 < r_1 \ll r_2 \ll 1$, the probability that the empirical measure associated to the optimally controlled state process started from $(t, m_{\boldsymbol{x}}^N) \in \mathcal{T}_{r_1}(t_0, m_0)$ exits the larger tube $\mathcal{T}_{r_2}(t_0, m_0)$ decays algebraically in N. In the non-compact setting treated below, this algebraic decay is uniform only over the intersection of $\mathcal{T}_{r_1}(t_0, m_0)$ with a large ball in $\mathcal{P}_p$, but we ignore this subtlety in the introduction.

More precisely, we show that, for each $(t, \boldsymbol{x})$ such that $(t, m_{\boldsymbol{x}}^N) \in \mathcal{T}_{r_1}(t_0, m_0)$,

$$\mathbb{P}\Big[s \mapsto (s, m^N_{\boldsymbol{X}_s^{(t,\boldsymbol{x})}}) \text{ leaves } \mathcal{T}_{r_2}(t_0, m_0)\Big] \leq CN^{-\gamma}, \tag{4.1.9}$$

where $\boldsymbol{X}^{(t,\boldsymbol{x})}$ denotes the optimal trajectory for the N-particle problem started from $(t_0, \boldsymbol{x}_0)$. Lemma 4.3.6 relies crucially on the global convergence rate of $\mathcal{V}^N$ to $\mathcal{U}$ already established in [78] and on an "asymmetric" version of the propagation of chaos arguments in [92].

The next step of the argument is to use the fact that, in view of the regularity of $\mathcal{U}$ in $\mathcal{O}$, $\mathcal{U}^N(t, \boldsymbol{x}) = \mathcal{U}(t, m_{\boldsymbol{x}}^N)$ nearly solves (HJB$_N$) on $\mathcal{T}_r(t_0, m_0)$ for r sufficiently small. In Lemma 4.3.8, we use this fact together with a verification argument to show that, for r small and $(t, \boldsymbol{x})$ such that $(t, m_{\boldsymbol{x}}^N) \in \mathcal{T}_r(t_0, m_0)$,

$$\begin{aligned}\mathcal{U}(t, m_{\boldsymbol{x}}^N) - \mathcal{V}^N(t, \boldsymbol{x}) &\leq C/N \\ &+ \mathbb{P}\Big[s \mapsto (s, m_{\boldsymbol{X}_s^{(t,\boldsymbol{x})}}) \text{ leaves } \mathcal{T}_r(t_0, m_0)\Big] \times \sup_{(s, m_{\boldsymbol{y}}^N) \in \mathcal{T}_r(t_0, m_0)} \Big(\mathcal{U}(s, \boldsymbol{y}^N) - \mathcal{V}^N(s, m_{\boldsymbol{y}}^N)\Big).\end{aligned} \tag{4.1.10}$$

Combining (4.1.10) with (4.1.9), we show that the rate of convergence improves when the radius of the tube shrinks. More precisely, we establish, for $0 < r_1 \ll r_2 \ll 1$, an estimate of the form

$$\begin{aligned}\sup_{(s, m_{\boldsymbol{y}}^N) \in \mathcal{T}_{r_1}(t_0, m_0)} \Big(\mathcal{U}(s, m_{\boldsymbol{y}}^N) - \mathcal{V}^N(s, \boldsymbol{y})\Big) &\leq C/N \\ &+ CN^{-\gamma} \times \sup_{(s, m_{\boldsymbol{y}}^N) \in \mathcal{T}_{r_2}(t_0, m_0)} \Big(\mathcal{U}(s, m_{\boldsymbol{y}}^N) - \mathcal{V}^N(s, \boldsymbol{y})\Big),\end{aligned} \tag{4.1.11}$$

where, crucially, γ is independent of r_1 and r_2. In particular, because γ is uniform we can apply (4.1.11) to a finite sequence of radii $r_2^{(1)} \gg r_1^{(1)} = r_2^{(2)} \gg r_1^{(2)} = r_2^{(3)} \gg ... \gg r_1^{(k)} = r_2^{(k)}$ to get that

$$\sup_{(s, m_{\boldsymbol{y}}^N) \in \mathcal{T}_{r_1^{(k)}}(t_0, m_0)} \Big(\mathcal{U}(s, m_{\boldsymbol{y}}^N) - \mathcal{V}^N(s, \boldsymbol{y})\Big) \leq CN^{-(1 \wedge k\gamma)},$$

and so choosing k large enough we conclude the existence of a small radius $r > 0$ such that the desired rate of convergence holds on the small tube $\mathcal{T}_r(t_0, m_0)$. This is enough to establish the estimate uniformly over compact subsets of $\mathcal{O}$ as desired.

The strategy of proof for the convergence of the gradients in (4.1.8) is similar, but more complicated because without the comparison principle it is harder to conclude an estimate analogous to (4.1.10). In addition, while the argument outlined above requires only minor refinements of the regularity results in [92] to execute, the convergence of the gradients requires a new Lipschitz bound on $D_{mm}\mathcal{U}$ (locally within $\mathcal{O}$), which is obtained in Theorem 4.2.4.

4.1.4 Organization of Chapter 4

In Section 4.2, we discuss notations and some preliminaries, and then state precisely our main results. Section 4.3 contains the proof of our first main convergence result, Theorem 4.2.1. Section 4.4 contains the proof of the convergence of the gradients (Theorem 4.2.2), and Section 4.5 contains the proof of the concentration inequality (Proposition 4.2.3). Finally, in Section 4.6 we state and prove a number of regularity results which are used in the earlier sections.

4.2 Preliminaries and main results

4.2.1 Basic notation

We fix throughout the chapter numbers $d \in \mathbb{N}$, $T > 0$. We work on a fixed filtered probability space $(\Omega, \mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$, which hosts independent d-dimensional Brownian motions $(W^i)_{i \in \mathbb{N}}$. We use bold to write elements of $(\mathbb{R}^d)^N$ or processes taking values in $(\mathbb{R}^d)^N$, that is, we write $\boldsymbol{x} = (x^1, ..., x^N) \in (\mathbb{R}^d)^N$ for a general element of $(\mathbb{R}^d)^N$. We denote by $\mathcal{P} = \mathcal{P}(\mathbb{R}^d)$ the space of probability measures on $\mathbb{R}^d$, and, for $q \in (1, \infty)$, we denote by $\mathcal{P}_q = \mathcal{P}_q(\mathbb{R}^d)$ the q-Wasserstein space, that is, the set of $m \in \mathcal{P}(\mathbb{R}^d)$ such that $M_q(m) < \infty$, where

$$M_q(m) := \int_{\mathbb{R}^d} |x|^q m(dx)$$

is the $q^{\text{th}}-$moment of the measure m. We endow $\mathcal{P}_q(\mathbb{R}^d)$ with the usual q-Wasserstein distance, which we denote by $\mathbf{d}_q$.

We will make use of the calculus on the Wasserstein space as explained in [79] and [96]. In particular, for a sufficiently smooth $\phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we write $\frac{\delta\phi}{\delta m}(m, x) : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ for the linear derivative of ϕ, which is defined by the formula

$$\phi(m) - \phi(\overline{m}) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta\phi}{\delta m}(sm + (1-s)\overline{m}, x)(m - \overline{m})(dx)ds,$$

together with the normalization convention

$$\int_{\mathbb{R}^d} \frac{\delta\phi}{\delta m}(m, x)m(dx) = 0.$$

When $\frac{\delta\phi}{\delta m}$ exists and is differentiable in its second argument, we denote by $D_m\phi = D_x\frac{\delta\phi}{\delta m}$ the Wasserstein or so called Lions derivative

$$D_m\phi(m, x) = D_x\frac{\delta\phi}{\delta m}(m, x) : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d.$$

Higher derivatives are denoted similarly.

For $k \in \mathbb{N}$, we denote by $C^k(\mathcal{P}_2(\mathbb{R}^d))$, the space of functions $\phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, such that for all $i \in \{1, \cdots, k\}$ and multi-index $l \in \{0, 1, ..., i\}^d$ with $|l| + i \leq k$, the derivative $D^{(l)}D_m^i\phi$ exists, and is continuous and uniformly bounded. Finally, for $k, n \in \mathbb{N}$, we denote by $C^k(\mathbb{R}^d; \mathbb{R}^n)$, the functions $\psi : \mathbb{R}^d \to \mathbb{R}^n$ that are $k-$times continuously differentiable. When $n = 1$ we will write C^k. Similar notation will be used for standard Hölder spaces, that is, for $\alpha \in (0, 1)$, $C^{k+\alpha}$ will be the space of functions in C^k with bounded and α-Hölder continuous derivatives up to order k.

4.2.2 Assumptions

The data for our problem consists of the three functions

$$L = L(x, a) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad \mathcal{F} = \mathcal{F}(m) : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \text{ and } \mathcal{G} = \mathcal{G}(m) : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}.$$

The Lagrangian L determines the Hamiltonian $H = H(x, p) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by the formula

$$H(x, p) = \sup_{a \in \mathbb{R}^d} \Big\{ -a \cdot p - L(x, a) \Big\}.$$

For part of the chapter, we will work with essentially the same assumptions as in [92], which we record here.

Assumption 1. *The Hamiltonian H is in $C^2(\mathbb{R}^d \times \mathbb{R}^d)$, and, for some $c, C > 0$ and all $(x, p) \in \mathbb{R}^d \times \mathbb{R}^d$,*

$$-C + c|p|^2 \leq H(x, p) \leq C + \frac{1}{c}|p|^2, \tag{4.2.1}$$

and

$$|D_x H(x, p)| \leq C(1 + |p|) \tag{4.2.2}$$

Moreover, H is locally strictly convex with respect to the last variable, that is, for each $R > 0$, there exists $c_R > 0$ such that, for all $(x, p) \in \mathbb{R}^d \times B_R$,

$$D^2_{pp} H(x, p) \geq c_R I_d. \tag{4.2.3}$$

Meanwhile, $\mathcal{F} \in C^2(\mathcal{P}_2(\mathbb{R}^d))$, and $\mathcal{F}$, $D_m\mathcal{F}$, $D^2_{ym}\mathcal{F}$ and $D^2_{mm}\mathcal{F}$ are uniformly bounded. Finally,

$\mathcal{G} \in C^4(\mathcal{P}_2(\mathbb{R}^d))$ with all derivatives up to order 4 uniformly bounded.

In order to obtain more regularity and to study the convergence of the gradients of $\mathcal{V}^N$, we require some additional smoothness, recorded here.

Assumption 2. *The data $\mathcal{F}, \mathcal{G}$ and H satisfy Assumption 1. In addition $\mathcal{F} \in C^3(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ and, for $i = 1, 2, 3$ and some $\delta \in (0, 1)$,*

$$\sup_{m \in \mathcal{P}_2(\mathbb{R}^d)} \|\frac{\delta^{(i)}\mathcal{F}}{\delta m^{(i)}}(m, \cdot)\|_{C^{2+\delta}((\mathbb{R}^d)^i;\mathbb{R})} + \sup_{m \in \mathcal{P}_2(\mathbb{R}^d)} \|\frac{\delta^{(i)}\mathcal{G}}{\delta m^{(i)}}(m, \cdot)\|_{C^{2+\delta}((\mathbb{R}^d)^i;\mathbb{R})} < \infty.$$

4.2.3 Preliminaries

In this section we recall some of the main results from the recent papers [78] and [92].

First, the main result of [78] (Theorem 2.5 therein) shows that, under Assumption 1, there exist constants $C > 0$ depending on the data, $\beta_d \in (0, 1)$ depending only on d, such that, for all $N \in \mathbb{N}$ and $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$,

$$|\mathcal{V}^N(t, \boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq CN^{-\beta_d}\Big(1 + \frac{1}{N}\sum_{i=1}^{N} |x^i|^2\Big). \tag{4.2.4}$$

In [92], meanwhile, the authors show that under Assumption 1, there exists an open and dense set $\mathcal{O} \subset [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ such that $\mathcal{U}$ is C^1 and satisfies (HJB$_\infty$) in a classical sense on $\mathcal{O}$.

To define $\mathcal{O}$ precisely, we first need some additional notation and terminology. If (m, α) is an optimizer for the problem defining $\mathcal{U}(t_0, m_0)$, then we call the curve $[t_0, T] \ni t \mapsto m_t \in \mathcal{P}_2(\mathbb{R}^d)$ an optimal trajectory starting from (t_0, m_0). Assumption 1 is enough to guarantee the validity of a standard result from the theory of mean field control, namely, that, if m is an optimal trajectory started from (t_0, m_0), then there exists a unique function $u : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}$, which is called the multiplier associated to the optimal trajectory m, such that the pair (u,m) solves the MFG system

$$\begin{cases} -\partial_t u - \Delta u + H(x,Du) = \dfrac{\delta \mathcal{F}}{\delta m}(m_t,x) \ \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(mD_pH(x,Du)) = 0 \ \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ m_{t_0} = m_0 \text{ and } u(T,x) = \dfrac{\delta \mathcal{G}}{\delta m}(m_T,x) \text{ in } \mathbb{R}^d. \end{cases}$$

The set $\mathcal{O}$ is defined as the set of $(t_0,m_0) \in [0,T)\times\mathcal{P}_2(\mathbb{R}^d)$ such that there is a unique optimal trajectory m started from (t_0,m_0), which is stable (see [92, Definition 2.5]) in the sense that, if u is the corresponding multiplier, then $(z,w)=(0,0)$ is the only solution in the space

$$C^{(1+\delta)/2,1+\delta}\times C([t_0,T];(C^{2+\delta})')$$

to the linear system

$$\begin{cases} -\partial_t z - \Delta z + D_pH(x,Du)\cdot Dz = \dfrac{\delta \mathcal{F}}{\delta m}(x,m(t))(\mu(t)) \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ \partial_t \mu - \Delta\mu - \operatorname{div}(D_pH(x,Du)\mu) - \operatorname{div}(D_{pp}H(x,Du)Dzm) = 0 \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ \mu(t_0) = 0 \text{ and } z(T,x) = \dfrac{\delta \mathcal{G}}{\delta m}(x,m(T))(\mu(T)) \text{ in } \ \mathbb{R}^d. \end{cases}$$

The set $\mathcal{O}$ will play a crucial role in the results of the present chapter, as well.

4.2.4 Main results

Our first main result shows that on compact with respect to $\mathcal{P}_p$, for some $p>2$, subsets of $\mathcal{O}$, the rate (4.2.4) can be substantially sharpened.

Theorem 4.2.1. *Let Assumption 1 hold, and assume that $p>2$. Then, for each subset K of $\mathcal{O}$ which is compact in $\mathcal{P}_p(\mathbb{R}^d)$, there is a constant $C=C(K)$ such that, for each $(t,\boldsymbol{x})\in[0,T]\times(\mathbb{R}^d)^N$ such* *that* $(t, m_{\boldsymbol{x}}^N) \in K$*,*

$$|\mathcal{U}(t, m_{\boldsymbol{x}}^N) - \mathcal{V}^N(t, \boldsymbol{x})| \leq C/N. \tag{4.2.5}$$

Remark 7. *As explained above, Example 2 in [131] clearly shows that, even if all the data is* C^∞*, we cannot expect a global convergence rate (of* V^N *to* U*) better than* $1/\sqrt{N}$*. Thus Theorem 4.2.1 shows that the convergence rate is generally different inside* $\mathcal{O}$ *than it is outside of* $\mathcal{O}$*.*

Remark 8. *Theorem 4.2.1 is not a consequence only of the regularity of* $\mathcal{U}$ *inside* $\mathcal{O}$*, that is, if we know only that* $\mathcal{U}$ *is smooth on some arbitrary open set* $\mathcal{O}'$*, it does not follow that the rate is* $1/N$ *inside* $\mathcal{O}'$ *in the sense of* (4.2.5)*. Indeed, the invariance of* $\mathcal{O}$ *under optimal trajectories, that is the fact that optimal trajectories for the MFC problem which start in* $\mathcal{O}$ *remain there, plays a crucial role throughout the proof of Theorem 4.2.1.*

Remark 9. *The fact that the rate is uniform only over subsets of* $\mathcal{O}$ *which are compact in* $\mathcal{P}_p$ *for some* $p > 2$ *is related to the fact that extra integrability is needed in order to obtain the convergence of empirical measures in the expected Wasserstein distance. More precisely, in order to obtain a key lemma (Lemma 4.3.6 below), we rely on the results of [199] to bound the probability that an auxiliary particle system exits a small "tube" around an optimal trajectory of the limiting problem, with the radius of the tube measured with respect to* $\mathbf{d}_2$*. For this it is necessary to control the* p^{th}*−moment for the initial condition of the particle system.*

Our next result shows that a sharp rate of convergence can also be obtained for the gradients.

Theorem 4.2.2. *Let Assumption 1 hold, and assume that* $p > 2$*. Then, for each subset* K *of* $\mathcal{O}$ *which is compact in* $\mathcal{P}_p$*, there is a constant* $C = C(K)$ *such that, for each* $i = 1, ..., N$ *and* $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$ *such that* $(t, m_{\boldsymbol{x}}^N) \in K$*,*

$$|D_m \mathcal{U}(t, m_{\boldsymbol{x}}^N, x^i) - N D_{x^i} \mathcal{V}^N(t, \boldsymbol{x})| \leq C/N. \tag{4.2.6}$$

Remark 10. *The main interest of Theorem 4.2.2 is that it implies a convergence rate for the optimal feedbacks for the N-particle problem which are given by*

$$\alpha^{N,i}(t,\boldsymbol{x}) = -D_p H(x^i, N D_{x^i} \mathcal{V}^N(t,\boldsymbol{x})).$$

That is, for any initial condition $(t_0, \boldsymbol{x}_0) \in [0,T] \times (\mathbb{R}^d)^N$*, the optimizer* $\boldsymbol{\alpha}$ *for the minimization problem in* (4.1.1) *satisfies* $\alpha_t^i = \alpha^{N,i}(t, \boldsymbol{X}_t)$*,* $\boldsymbol{X}$ *denoting the optimal trajectory starting from* $(t_0, \boldsymbol{x}_0)$*. Meanwhile, the optimal feedback for the mean field problem, at least for initial conditions* $(t_0, m_0) \in \mathcal{O}$*, takes the form*

$$\alpha^{MF}(t,m,x) = -D_p H(x^i, D_m \mathcal{U}(t,m,x)).$$

That is, for any $(t_0, m_0) \in \mathcal{O}$*, the unique optimizer* α *for the minimization problem in* (4.1.2) *satisfies* $\alpha(t,x) = \alpha^{MF}(t, m_t, x)$*,* m_t *denoting the unique optimal trajectory started from* (t_0, m_0)*. Thus we can clearly infer from Theorem 4.2.2 that, for each subset* K *of* $\mathcal{O}$ *which is compact in* $\mathcal{P}_p$*, there exists* $C = C(K)$ *such that*

$$\sup_{\{(t,\boldsymbol{x}):(t,m_{\boldsymbol{x}}^N)\in K\}} |\alpha^{N,i}(t,\boldsymbol{x}) - \alpha^{MF}(t, m_{\boldsymbol{x}}^N, x^i)| \le C/N.$$

In [92], it is shown in Theorem 1.2 that the regularity of $\mathcal{U}$ in $\mathcal{O}$ implies a convergence for the optimal trajectories of $\mathcal{V}^N$, in the spirit of propagation of chaos, at least for appropriate initial conditions. Using the convergence of the gradients obtained in Theorem 4.2.2, we are able to supplement this result with the following concentration inequality.

Proposition 4.2.3. *Let Assumption 2 hold, and fix* $(t_0, m_0) \in \mathcal{O}$ *such that* $m_0 \in \mathcal{P}_p$ *for some* $p > 2$ *and, in addition, satisfies a quadratic transport-entropy inequality, that is, there exists* $\kappa > 0$ *such* *that*

$$\mathbf{d}_2(m_0,\nu) \le \kappa \mathcal{R}(\nu|m_0) \text{ if } \nu \ll m_0, \tag{4.2.7}$$

where $\mathcal{R}$ denotes the relative entropy,

$$\mathcal{R}(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} log \frac{d\nu}{d\mu} d\mu & \text{if } \nu \ll \mu, \\ \\ \infty & \text{otherwise}. \end{cases}$$

For each N, denote by $\boldsymbol{X}^N = (X^{N,1}, ..., X^{N,N})$ the solution of

$$dX_t^{N,i} = -D_pH(X_t^{N,i}, ND_{x^i}V^N(t,\boldsymbol{X}_t^N))dt + \sqrt{2}dW_t^i, \quad X_{t_0}^i = \xi^i, \tag{4.2.8}$$

where $(\xi^i)_{i\in\mathbb{N}}$ are i.i.d. with common law m_0. Then, there exists a constant $r_0 > 0$ (which can depend on (t_0, m_0)) such that the following holds: for each $\eta > 0$, $K > 0$, we can find a constant $C > 0$ (which may depend on (t_0, m_0), in addition to η and K) such that for any $0 < r < r_0$ and $N > C/\min(\epsilon, \epsilon^{d+8})$, we have

$$\mathbb{P}\Big[\sup_{t_0\le t\le T} \mathbf{d}_2(m^N_{\boldsymbol{X}_t^N}, m_t) > r\Big] \le C\exp\big(-\frac{r^2}{C}N\big) + \exp(-KN^{1-\eta}), \tag{4.2.9}$$

where (m,α) denotes the unique optimizer for the problem defining $\mathcal{U}(t_0, m_0)$.

Remark 11. *Concentration results similar to Proposition 4.2.3 were obtained for mean field games in [143] under the assumption that the corresponding master equation has a smooth solution. The analogous condition in our setting would be that the value function $\mathcal{U}$ is globally smooth, which we do not have because we do not assume any convexity on $\mathcal{F}$ and $\mathcal{G}$.*

Remark 12. *The presence of the η in* (4.2.9) *comes from the fact that in Theorem 4.2.2 we have convergence of the optimal feedback strategies only on compact subsets of $\mathcal{P}_p$, with $p > 2$. In* *particular, in the proof of Proposition 4.2.3 we must begin the argument by bounding from above the probability that the empirical measure of the optimal trajectory for the N-particle problem exits a large ball in $\mathcal{P}_p$. This can be estimated from above by a multiple of a probability of the form*

$$\mathbb{P}[\frac{1}{N}\sum_{i=1}^{N}|\xi^i|^p > R] \tag{4.2.10}$$

for some large R, where $(\xi^i)_{i\in\mathbb{N}}$ are i.i.d. sub-Gaussian random variables. When $p > 2$, the random variables $|\xi^i|^p$ are only "sub-Weybull" rather than sub-exponential, that is, they have tails like $\mathbb{P}[|\xi^i|^p > x] \approx \exp(-cx^{2/p})$. Applying a concentration inequality to bound this tail probability leads to the second term in (4.2.9).

A key step in proving Theorem 4.2.2 is showing that, under Assumption 2, the C^1 regularity obtained in [92] can be improved to C^2 regularity. This result is interesting in its own right, and so we state it here, alongside our main convergence results.

Theorem 4.2.4. *Under Assumption 2, the derivative $D_{mm}\mathcal{U}$ exists and is continuous in $\mathcal{O}$. Moreover, for each $(t_0, m_0) \in \mathcal{O}$, there exist constants $\delta, C > 0$ such that, for each t, m_1, m_2 with $|t - t_0| < \delta$, $\mathbf{d}_2(m_0, m_i) < \delta$ and $i = 1, 2$, we have*

$$\sup_{x,y\in\mathbb{R}^d} |D_{mm}\mathcal{U}(t, m_1, x, y) - D_{mm}\mathcal{U}(t, m_2, x, y)| \leq C\mathbf{d}_1(m_1, m_2).$$

4.3 The proof of Theorem 4.2.1

In this section we present the proof of Theorem 4.2.1. For simplicity, we fix throughout the section a $p > 2$.

It will be useful to note that one of the inequalities in Theorem 4.2.1 is relatively easy. Indeed, under the smoothness assumptions on $\mathcal{F}$, $\mathcal{G}$, it is not difficult to show the following.

Proposition 4.3.1. *There is a constant C such that, for all $N \in \mathbb{N}$ and for each $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$,*

$$\mathcal{V}^N(t, \boldsymbol{x}) \leq \mathcal{U}(t, m^N_{\boldsymbol{x}}) + C/N \ .$$

Proof. We omit the proof, since it is almost identical the one of the first inequality in Theorem 2.7 of [131]. □

4.3.1 Tubes around optimal trajectories

We now introduce some notation which will be useful in the proof of Theorem 4.2.1.

Given $(t_0, m_0) \in \mathcal{O}$, we denote by $t \mapsto m_t^{(t_0,m_0)}$ the unique optimal trajectory for the limiting McKean-Vlasov control problem started from (t_0, m_0). For simplicity, we extend $m^{(t_0,m_0)}$ by a constant to $[0, t_0]$, that is, we define $m_t^{(t_0,m_0)} = m_0$ for $0 \leq t < t_0$.

For $r > 0$, $\mathcal{T}_r(t_0, m_0)$ is an open tube around the optimal trajectory started from (t_0, m_0), that is,

$$\mathcal{T}_r(t_0, m_0) = \Big\{(t, m) : t \in (t_0 - r, T] \cap [0, T],\ \mathbf{d}_2(m, m_t^{(t_0,m_0)}) < r\Big\}.$$

When (t_0, m_0) is understood from context, we write simply $\mathcal{T}_r$.

Because we are working in the whole space, we will often have to intersect the "tubes" $\mathcal{T}_r$ with bounded subsets of the Wasserstein space $\mathcal{P}_p$. To facilitate this, we set, for $R > 0$,

$$\mathcal{Q}_R = [0, T] \times B^p_R,$$

where B^p_R is the ball of radius R in $\mathcal{P}_p$, centered at δ_0, and use the notation

$$\mathcal{T}_{r,R}(t_0, m_0) = \mathcal{T}_r(t_0, m_0) \cap \mathcal{Q}_R.$$

We also need to project these sets down to finite-dimensional spaces. In particular, we will write $\mathcal{O}^N = \{(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N : (t, m_{\boldsymbol{x}}^N) \in \mathcal{O}\}$, and likewise we set

$$\mathcal{T}_{r,R}^N(t_0, m_0) = \Big\{(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N : (t, m_{\boldsymbol{x}}^N) \in \mathcal{T}_{r,R}(t_0, m_0)\Big\}.$$

Finally, $\boldsymbol{X}^{N,t,\boldsymbol{x}} = (X^{N,t,\boldsymbol{x},i})_{i=1,\dots,N}$ is the optimal trajectory for the N-particle control problem started from $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$, that is, the solution of

$$\begin{cases} dX_s^{N,t,\boldsymbol{x},i} = -D_p H(X_s^{N,t,\boldsymbol{x},i}, N D_{x^i} \mathcal{V}^N(s, \boldsymbol{X}_s^{N,t,\boldsymbol{x}}))ds + \sqrt{2} dW_s^i \text{ in } [s, T], \\ X_t^{N,t,\boldsymbol{x},i} = x^i. \end{cases} \tag{4.3.1}$$

To prove Theorem 4.2.1, it will suffice to establish the following Proposition.

Proposition 4.3.2. *Let $(t_0, m_0) \in \mathcal{O}$ and $R > 0$. Then there exist $r, C > 0$ such that, for each $N \in \mathbb{N}$ and each $(t, \boldsymbol{x}) \in \mathcal{T}_{r,R}^N(t_0, m_0)$,*

$$|\mathcal{V}^N(t, \boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq C/N.$$

Indeed, once Proposition 4.3.2 is proved we can complete the proof of Theorem 4.2.1 as follows:

Proof of Theorem 4.2.1. Fix $(t_0, m_0) \in \mathcal{O} \cap ([0, T] \times \mathcal{P}_p)$, and choose R large enough that $(t_0, m_0) \in \mathcal{Q}_R$. Thanks to Proposition 4.3.2, there exist $r, C > 0$ such that, for all $(t, \boldsymbol{x})$ such that $(t, m_{\boldsymbol{x}}^N) \in \mathcal{T}_{r,R}(t_0, m_0)$,

$$|\mathcal{V}(t, \boldsymbol{x}) - \mathcal{U}(t, m_{\boldsymbol{x}}^N)| \leq C/N. \tag{4.3.2}$$

In particular, for each $(t_0, m_0) \in \mathcal{O} \cap ([0, T] \times \mathcal{P}_p)$, there exists a subset of $\mathcal{O}$, which is open in $\mathcal{P}_p$ and contains (t_0, m_0), on which the convergence rate is $1/N$, in the sense of (4.3.2). This completes the proof. $\square$

4.3.2 The proof of Proposition 4.3.2

We give here the proof of Proposition 4.3.2. We start by recording a few preliminary facts about the tubes $\mathcal{T}_r(t_0, m_0)$ defined above.

Lemma 4.3.3. *Given* $(t_0, m_0) \in \mathcal{O}$*, there exists* $r_0 > 0$ *such that* $\mathcal{T}_{r_0}(t_0, m_0) \subset \mathcal{O}$*. Moreover, for any* $r_2 > 0$*, there exists* $0 < r_1 < r_2$ *such that optimal optimal trajectories from within* $\mathcal{T}_{r_1}(t_0, m_0)$ *remain in* $\mathcal{T}_{r_2}(t_0, m_0)$*, that is, for each* $(t, m) \in \mathcal{T}_{r_1}(t_0, m_0)$ *and each* $t \le s \le T$*,*

$$\mathbf{d}_2(m_s^{(t,m)}, m_s^{(t_0,m_0)}) < r_2.$$

Proof. The first claim is a consequence of the fact that $\mathcal{O}$ is open, as proved in [92, Theorem 2.8]. The second one can be inferred from [92, Lemma 2.9] together with the uniform 1/2-Hölder continuity of the optimal trajectories $s \mapsto m_s^{(t,m)} \in \mathcal{P}_2$, which is in turn a consequence of the uniform boundedness of optimal controls (see Lemma 3.3 of [78]). □

The next task is to use the regularity of $\mathcal{U}$ in $\mathcal{O}$ to argue that the function

$$\mathcal{U}^N(t, \boldsymbol{x}) = \mathcal{U}(t, m_{\boldsymbol{x}}^N) : [0, T] \times (\mathbb{R}^d)^N \to \mathbb{R} \tag{4.3.3}$$

satisfies a PDE similar to (HJB$_N$). First, notice that, if we assume that $\mathcal{U}$ is $C^{1,2}$, then $\mathcal{U}^N$ is a classical solution in $\mathcal{O}^N$ to

$$-\partial_t \mathcal{U}^N - \sum_{j=1}^N \Delta_{x^j} \mathcal{U}^N + \frac{1}{N} \sum_{j=1}^N H(x^j, N D_{x^j} \mathcal{U}^N) = \mathcal{F}(m_{\boldsymbol{x}}^N) + E^N(t, \boldsymbol{x}) \tag{4.3.4}$$

where

$$E^N(t, \boldsymbol{x}) = -\frac{1}{N^2} \sum_{j=1}^N \mathrm{tr}\,(D_{mm} \mathcal{U}(t, m_{\boldsymbol{x}}^N, x^j, x^j)). \tag{4.3.5}$$

Under Assumption 1, the main results of [92] show that $\mathcal{U}$ is C^1, but not necessarily C^2, so we cannot immediately conclude that $\mathcal{U}^N$ satisfies (4.3.4). Nevertheless, Proposition 4.6.4 shows that, uniformly in x and locally uniformly in (t, m) with respect to $\mathbf{d}_2$, $m \mapsto D_m U(t, m, x)$ is $\mathbf{d}_1$-Lipschitz. In particular, for each $(t_0, m_0) \in \mathcal{O}$, it is clear that we can choose r small enough so that $m \mapsto D_m \mathcal{U}(t, m, x)$ is Lipschitz with respect to $\mathbf{d}_1$ uniformly over $\mathcal{T}_r(t_0, m_0)$. It follows that $\mathcal{U}^N$ is C^1 on $\mathcal{T}_r(t_0, m_0)$ with each partial derivative

$$D_{x^i} \mathcal{U}^N(t, \boldsymbol{x}) = \frac{1}{N} D_m \mathcal{U}(t, m_{\boldsymbol{x}}^N, x^i)$$

being uniformly Lipschitz in $\boldsymbol{x}$. Moreover, arguing as in the proof of [131, Proposition 5.1], one can show that the equation (4.3.4) holds almost everywhere on $\mathcal{T}_r^N(t_0, m_0)$, with an error term E^N satisfying

$$\|E^N\|_{L^\infty(\mathcal{T}_r)} \le C/N. \tag{4.3.6}$$

We record this sequence of observations in the following lemma.

Lemma 4.3.4. *For any* $(t_0, m_0) \in \mathcal{O}$, *there exists* $r > 0$ *such that, for each* $N \in \mathbb{N}$, *the projection* $\mathcal{U}^N$ *defined by* (4.3.3) *lies in* $C^1(\mathcal{T}_r^N(t_0, m_0))$, *with spatial derivatives* $D_{x^i} \mathcal{U}^N$ *being Lipschitz continuous in* $\boldsymbol{x}$, *and such that* (4.3.4) *holds almost everywhere, with the error function*

$$E^N := -\partial_t \mathcal{U}^N - \sum_{j=1}^N \Delta_{x^j} \mathcal{U}^N + \frac{1}{N} \sum_{j=1}^N H(x^j, N D_{x^j} \mathcal{U}^N) - \mathcal{F}(m_{\boldsymbol{x}}^N) \tag{4.3.7}$$

satisfying the estimate (4.3.6).

We now establish a sequence of technical lemmas. The first one explains how to estimate the probability that the empirical measure associated with an optimal trajectory for $\mathcal{V}^N$ grows quickly in $\mathcal{P}_p$.

Lemma 4.3.5. *There is a constant* C_p *depending on* p *and the data with the following property:*

for each $N \in \mathbb{N}$ *and* $(t_0, \boldsymbol{x}_0) \in [0,T] \times (\mathbb{R}^d)^N$,

$$\mathbb{P}\Big[\sup_{t_0 \le t \le T} \mathbf{d}_p(m_{\boldsymbol{X}_t^{N,t_0,\boldsymbol{x}_0}}, m^N_{\boldsymbol{x}_0}) \ge C_p\Big] \le C_p/N. \tag{4.3.8}$$

Proof. For simplicity, we set $\boldsymbol{X} = (X^1, ..., X^N) = \boldsymbol{X}^{N,t_0,x_0}$, and let $\boldsymbol{\alpha} = (\alpha^1, ..., \alpha^N)$ denote the optimal control started from (t_0, x_0). Then, for each fixed $t \in [t_0, T]$, we have

$$\mathbf{d}_p^p(m^N_{\boldsymbol{X}_t}, m^N_{\boldsymbol{x}_0}) \le \frac{1}{N}\sum_{i=1}^N |X^i_t - x^i_0|^p \le \frac{1}{N}\sum_{i=1}^N \Big|\int_{t_0}^t \alpha^i_s ds + W^i_t - W^i_{t_0}\Big|^p.$$

Using the fact that α^i is bounded independently of N (see [92, Lemma 1.7]), we find easily that

$$\sup_{t_0 \le t \le T} \mathbf{d}_p^p(m^N_{\boldsymbol{X}_t}, m^N_{\boldsymbol{x}_0}) \le C\Big(1 + \frac{1}{N}\sum_{i=1}^N \sup_{t_0 \le t \le T} |W^i_t - W^i_{t_0}|^p\Big).$$

The result now follows easily from Chebyshev's inequality.

□

The next lemma shows that the probability of the empirical measure exiting a tube decays algebraically, with a constant that is uniform over a smaller tube or, more precisely, uniform over the intersection of the smaller tube with a ball in $\mathcal{P}_p$.

Lemma 4.3.6. *There exists a constant* $\gamma_{p,d} \in (0,1)$*, which depends only on* p *and* d*, with the following property. Suppose that* $(t_0, m_0) \in \mathcal{O} \cap ([0,T] \times \mathcal{P}_p)$*,* $0 < r_1 < r < r_2$ *and* $R_1, R_2 > 0$ *are such that*

1. $R_2 - R_1 \ge C_p$*,* C_p *being the constant appearing in* (4.3.8).
2. r_2 *is small enough so that* $\overline{\mathcal{T}_{r_2}(t_0, m_0)} \subset \mathcal{O}$*, and the conclusion of Lemma 4.3.4 applies on* $\mathcal{T}_{r_2}(t_0, m_0)$.

3. optimal trajectories started from inside $\mathcal{T}_{r_1}(t_0, m_0)$ remain in $\mathcal{T}_r(t_0, m_0)$, that is,

$$\sup_{t \le s \le T} \mathbf{d}_2\Big(m_s^{(t,m)}, m_s^{(t_0,m_0)}\Big) < r \ \textit{for all} \quad (t,m) \in \mathcal{T}_{r_1}(t_0, m_0).$$

Then there exists a constant C, which is independent of N, such that, for all $N \in \mathbb{N}$ and $(t, \boldsymbol{x}) \in \mathcal{T}^N_{r_1,R_1}$,

$$\mathbb{P}[\tau^{N,t,\boldsymbol{x}} < T] \le C N^{-\gamma_{p,d}},$$

where $\tau^{N,t,\boldsymbol{x}} = \inf\Big\{s > t : (s, \boldsymbol{X}_s^{N,t,\boldsymbol{x}}) \in (\mathcal{T}^N_{r_2,R_2})^c\Big\} \wedge T$.

Proof. We fix (t_0, m_0), $r_1 < r < r_2$ and $R_1 < R_2$ as in the statement and $N \in \mathbb{N}$ and $(t, \boldsymbol{x}) \in \mathcal{T}^N_{r_1,R_1}$. For notational simplicity, we write $\boldsymbol{X} = X^{N,t,\boldsymbol{x}}$ and $\tau = \tau^{N,t,\boldsymbol{x}}$. At this point we have dropped the superscript for simplicity, and we note that in the rest of the argument generic constants C may change from line to line, but they must not depend on $(t, \boldsymbol{x}, N)$.

We are going to break the problem up by writing τ as

$$\tau = \tau_1 \wedge \tau_2,$$

where

$$\tau_1 = \inf\Big\{s > t : (s, \boldsymbol{X}_s^{N,t,\boldsymbol{x}}) \in (\mathcal{T}^N_{r_2})^c\Big\} \wedge T \ \text{ and } \ \tau_2 = \inf\Big\{s > t : d_p(m^N_{\boldsymbol{X}_s}, \delta_0) > R_2\Big\} \wedge T. \quad (4.3.9)$$

In particular, we have

$$\mathbb{P}[\tau < T] \le \mathbb{P}[\tau_1 < T] + \mathbb{P}[\tau_2 < T]. \qquad (4.3.10)$$

Moreover, since Lemma 4.3.5 shows that

$$\mathbb{P}[\tau_2 < T] = \mathbb{P}[\sup_{t\le s\le T} \mathbf{d}_p(m^N_{\boldsymbol{X}_s}, \delta_0) > R_2] \le \mathbb{P}[\sup_{t\le s\le T} \mathbf{d}_p(m^N_{\boldsymbol{X}_s}, m^N_{\boldsymbol{x}}) > R_2 - R_1] \le C_p/N,$$

it suffices to prove a corresponding estimate on $\mathbb{P}[\tau_1 < T]$.

We are now going to employ an argument similar to the proof of Lemma 3.2 in [92], but with non-symmetric initial conditions. In particular, combining Lemma 4.3.4 and the Itô-Krylov formula (see [263] Section 2.10), on $[t, \tau_1]$ we have the dynamics

$$\begin{aligned}
d\mathcal{U}^N(s, \boldsymbol{X}_s) &= \Big(\partial_t \mathcal{U}^N(s, \boldsymbol{X}_s) + \sum_{j=1}^N \Delta_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s) \\
&\quad - \sum_{j=1}^N D_p H(X^i_s, N D_{x^j} \mathcal{V}^N) \cdot D_{x^j} \mathcal{U}^N(x, \boldsymbol{X}_s)\Big) ds + \sqrt{2} \sum_{j=1}^N D_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s) \cdot dW^j_s \\
&= \Big(\frac{1}{N} \sum_{j=1}^N H(X^i_s, N D_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s)) - \sum_{j=1}^N D_p H(X^i_s, N D_{x^j} \mathcal{V}^N) \cdot D_{x^j} \mathcal{U}^N(x, \boldsymbol{X}_s) \\
&\qquad - E_s - \mathcal{F}(m^N_{\boldsymbol{X}_s})\Big) ds + \sqrt{2} \sum_{j=1}^N D_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s) dW^j_s \\
&\ge \Big(\frac{1}{N} \sum_{j=1}^N \Big(- L(X^j_s, -D_p H(X^j_s, N D_{x^j} \mathcal{V}^N(s, \boldsymbol{X}_s))) \\
&\qquad + C^{-1} \Big| D_p H(X^j_s, N D_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s)) - D_p H(X^j_s, N D_{x^j} \mathcal{V}^N(s, \boldsymbol{X}_s)) \Big|^2 \\
&\qquad - C N^{-1} - \mathcal{F}(m^N_{\boldsymbol{X}_s})\Big) ds + \sqrt{2} \sum_{j=1}^N D_{x^j} \mathcal{U}^N(s, \boldsymbol{X}_s) dW^j_s.
\end{aligned}$$

In the last bound, we used the fact that the strict convexity of L in a yields some $C > 0$ such that, for any $x \in \mathbb{R}^d$ and $p, q \in \mathbb{R}^d$,

$$H(x, p) - D_p H(x, q) \cdot p \ge -L(x, -D_p H(x, q)) + \frac{1}{C} |D_p H(x, p) - D_p H(x, q)|^2. \tag{4.3.11}$$

Taking expectations in the inequality above for the dynamics of $\mathcal{U}^N(s, \boldsymbol{X}_s)$ and integrating from t

to τ, we get

$$
\begin{aligned}
&\mathbb{E}\Big[\int_t^\tau \frac{1}{CN}\sum_{j=1}^N \Big|D_pH(X_s^j, ND_{x^j}\mathcal{U}^N(s,\boldsymbol{X}_s)) - D_pH(X_s^j, ND_{x^j}\mathcal{V}^N(s,\boldsymbol{X}_s))\Big|^2\Big] \\
&\quad \le \mathbb{E}[\mathcal{U}^N(\tau,\boldsymbol{X}_\tau)] - \mathcal{U}^N(t,\boldsymbol{x}) + \mathbb{E}\Big[\int_t^\tau \frac{1}{N}\sum_{j=1}^N L(X_s^j, -D_pH(X_s^j, ND_{x^j}\mathcal{V}^N(s,\boldsymbol{X}_s))) \\
&\qquad\qquad + \mathcal{F}(m^N_{\boldsymbol{X}_s})\Big] + C/N \\
&\quad \le C(1+M_2(m^N_{\boldsymbol{x}}))N^{-\beta_d} + \mathbb{E}[\mathcal{V}^N(\tau,\boldsymbol{X}_\tau)] - \mathcal{V}^N(t,\boldsymbol{x}) \\
&\quad + \mathbb{E}\Big[\int_t^\tau \frac{1}{N}\sum_{j=1}^N L(X_s^j, -D_pH(X_s^j, ND_{x^j}\mathcal{V}^N(s,\boldsymbol{X}_s))) + \mathcal{F}(m^N_{\boldsymbol{X}_s})\Big] = CN^{-\beta_d},
\end{aligned}
\tag{4.3.12}
$$

with the last equality following from the fact that $\boldsymbol{X}$ is the optimal trajectory for the N-particle problem, and where β_d is the exponent appearing in (4.2.4). We note that the dependence factor $(1 + M_2(m^N_{\boldsymbol{x}}))$ was absorbed into the constant, keeping in mind that the constant is allowed to depend on the radius r_2 of the largest tube.

Next, we are going to use (4.3.12) to compare the process $\boldsymbol{X}$ to the process $\boldsymbol{Y} = (Y^i)_{i=1,\dots,N}$ defined on a stochastic interval $[t,\sigma]$ by

$$
\begin{cases}
dY_s^i = -D_pH(Y_s^i, D_m\mathcal{U}(s, m^N_{\boldsymbol{Y}_s}, Y_s^i))ds + \sqrt{2}dW_s^i \quad t \le s \le \sigma : \inf\{s > t : (s,\boldsymbol{Y}) \in \big(\mathcal{T}^N_{r_3}\big)^c\} \wedge T, \\
Y_t^i = x^i.
\end{cases}
$$

In particular, noting that we can rewrite the dynamics of $\boldsymbol{X}$ as

$$
dX_s^i = \Big(-D_pH(X_s^i, D_m\mathcal{U}(s, m^N_{\boldsymbol{X}_s}, X_s^i)) + E_s^i\Big)ds + dW_s^i,
$$

with

$$
E_s^i = D_pH(X_s^i, ND_{x^i}\mathcal{U}^N(s,\boldsymbol{X}_s)) - D_pH(X_s^i, ND_{x^i}\mathcal{V}^N(s,\boldsymbol{X}_s)),
$$

we easily get by computing the dynamics of $\frac{1}{N}\sum_{j=1}^N |X_s^i - Y_s^i|^2$ and applying Gronwall's inequality that

$$\mathbb{E}\Big[\sup_{t\le s\le \sigma\wedge\tau}\frac{1}{N}\sum_{i=1}^N |X_s^i - Y_s^i|^2 ds\Big] \le C\mathbb{E}\Big[\int_t^{\sigma\wedge\tau}\frac{1}{N}\sum_{j=1}^N |E_s^j|^2 ds\Big] \le CN^{-\beta_d},$$

the last inequality coming from (4.3.12). It follows that

$$\mathbb{E}\Big[\sup_{t\le s\le \sigma\wedge\tau}\mathbf{d}_2^2(m_{\boldsymbol{X}_s}^N, m_{\boldsymbol{Y}_s}^N)\Big] \le CN^{-\beta_d}. \tag{4.3.13}$$

In particular, if we pick $r' \in (r, r_2)$, and set

$$\sigma' = \inf\{s > t; (s, \boldsymbol{Y}_s) \in (\mathcal{T}_{r'}^N)^c\} \wedge T,$$

then, using Markov's inequality and (4.3.13) , we find

$$\begin{aligned}
\mathbb{P}[\tau_1 < T] &\le \mathbb{P}[\sigma' < T] + \mathbb{P}[\sigma' = T, \text{ and } \tau_1 < T] \\
&\le \mathbb{P}[\sigma' < T] + \mathbb{P}\Big[\sup_{t\le s\le \sigma\wedge\tau}\mathbf{d}_2(m_{\boldsymbol{X}_s}^N, m_{\boldsymbol{Y}_s}^N) > r_2 - r'\Big] \\
&\le \mathbb{P}[\sigma' < T] + CN^{-\beta_d/2}.
\end{aligned}$$

To complete the proof, we need only show that, for some $\gamma'_{d,p} > 0$ depending only on d and p,

$$\mathbb{P}[\sigma' < T] \le CN^{-\gamma'_{d,p}}.$$

We would like to infer this from Lemma 4.3.7 below, which is a non-symmetric version of a standard "propagation of chaos" result for interacting particle systems. We cannot, however, apply

Lemma 4.3.7 directly, since the map

$$(t, x, m) \mapsto -D_p H(x, D_m \mathcal{U}(t, m, x))$$

is not globally defined, let alone globally Lipschitz. To overcome this, we simply extend it, choosing a map $b = b(t, x, m) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ such that

$$\begin{cases} b(t, x, m) = -D_p H(x, D_m \mathcal{U}(t, m, x)) \quad \text{for } (t, x, m) \in \mathcal{T}_{r_2}(t_0, m_0), \\ b \text{ is globally bounded and is Lipschitz in } (x, m) \text{ (with respect to the } \mathbf{d}_2\text{-distance) uniformly in } t. \end{cases}$$

This is possible thanks to the regularity of $\mathcal{U}$ in $\mathcal{T}_{r_2}$. Then, we define the process $\mathbf{Z} = (Z^i)_{i=1,\dots,N}$ on the whole interval $[t, T]$ via

$$\begin{cases} dZ_s^i = b(s, Z_s^i, m_{\mathbf{Z}_s}^N) ds + \sqrt{2} dW_s^i \quad t \le s \le T, \\ Z_t^i = x^i. \end{cases}$$

It is easy to see that $\mathbf{Z} = \boldsymbol{Y}$ on $[t, \sigma)$, and that the solution Z of the corresponding McKean-Vlasov SDE

$$\begin{cases} dZ_s = b(s, Z_s, \mathcal{L}(Z_s)) ds + \sqrt{2} dW_s \quad t \le s \le T, \\ Z_t \sim m_{\boldsymbol{x}}^N \end{cases}$$

is exactly the optimal trajectory for the mean field control problem, that is, $\mathcal{L}(Z_s) = m_s^{(t, m_{\boldsymbol{x}}^N)}$. In particular, we can use Lemma 4.3.7 and Markov's inequality to conclude

$$\mathbb{P}[\sigma' < T] = \mathbb{P}\Big[\sup_{t \le s \le T} \mathbf{d}_2(m_{\mathbf{Z}_s}^N, m_s^{(t_0, m_0)}) > r'\Big] \le \mathbb{P}\Big[\sup_{t \le s \le T} \mathbf{d}_2(m_{\mathbf{Z}_s}^N, \mathcal{L}(Z_s)) > r' - r\Big] \le C N^{-\gamma'_{d,p}},$$

where in the first inequality we used the fact that $\mathbf{d}_2(\mathcal{L}(Z_s), m_s^{(t_0, m_0)}) = \mathbf{d}_2(m_s^{(t, m_{\boldsymbol{x}}^N)}, m_s^{(t_0, m_0)}) < r$, by hypothesis. This completes the proof.

□

The following lemma is a sort of "non-symmetric" version of a classical propagation of chaos estimate for interacting particle systems, which was used in the proof of Lemma 4.3.6.

Lemma 4.3.7. *Fix $b = b(t, x, m) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, and suppose that b is measurable, bounded, and Lipschitz in (x, m) with respect to the $\mathbf{d}_2$-distance, uniformly in t. Then there is a constant $C > 0$ depending only on the L^∞ and Lipschitz bounds on b, and a constant $\gamma'_{d,p} > 0$ depending only on d and p, such that, for any $N \in \mathbb{N}$ and $(t_0, \boldsymbol{x}_0) \in [0, T] \times (\mathbb{R}^d)^N$,*

$$\mathbb{E}\Big[\sup_{t_0 \le t \le T} \mathbf{d}_2(m_t^{(t_0, m^N_{\boldsymbol{x}_0})}, m^N_{\boldsymbol{X}_t^{(t_0,\boldsymbol{x}_0)}})\Big] \le C M_p^{1/p}(m^N_{\boldsymbol{x}_0}) N^{-\gamma'_{d,p}},$$

where $\boldsymbol{X}_t^{(t_0,\boldsymbol{x}_0)} = (X_t^{(t_0,\boldsymbol{x}_0),i})_{i=1,\dots,N}$ denotes the solution to the SDE

$$dX_t^i = b(t, X_t^i, m_{\boldsymbol{X}_t})dt + \sqrt{2}dW_t^i \;\; \text{and} \;\; X_{t_0}^i = x_0^i, \tag{4.3.14}$$

and $m_t^{(t_0, m^N_{\boldsymbol{x}_0})} = \mathcal{L}(Y_t)$, where Y solves

$$dY_t = b(t, Y_t, \mathcal{L}(Y_t))dt + \sqrt{2}dW_t \;\; \text{and} \;\; Y_{t_0} \sim m^N_{\boldsymbol{x}}.$$

Proof. We fix $(t_0, \boldsymbol{x}_0) \in [0, T] \times (\mathbb{R}^d)^N$. Thanks to the Lipschitz continuity of b, it is straightforward to check that the map $\Phi = \Phi(\boldsymbol{x}) : (\mathbb{R}^d)^N \to \mathbb{R}$ given by

$$\Phi(\boldsymbol{x}) = \mathbb{E}\Big[\sup_{t_0 \le t \le T} \mathbf{d}_2(m_t^{(t_0, m^N_{\boldsymbol{x}_0})}, m^N_{\boldsymbol{X}_t^{(t_0,\boldsymbol{x})}})\Big]$$

is Lipschitz continuous with respect to $\mathbf{d}_2$ with a constant C depending only on L, that is,

$$|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y})| \le C \mathbf{d}_2(m^N_{\boldsymbol{x}}, m^N_{\boldsymbol{y}}).$$

As a consequence, we can find a function $\widetilde{\Phi} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ with the same $\mathbf{d}_2$-Lipschitz constant and such that $\widetilde{\Phi}(m^N_{\boldsymbol{x}}) = \Phi(\boldsymbol{x})$ for $\boldsymbol{x} \in (\mathbb{R}^d)^N$.

Now define the "lift" $\widehat{\Phi} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ by

$$\widehat{\Phi}(m) = \int_{(\mathbb{R}^d)^N} \Phi(\boldsymbol{y}) dm^{\otimes N}(d\boldsymbol{y}).$$

It follows from Theorem 1 of [199] that, for some $\gamma'_{p,d} > 0$ depending explicitly on p and d.

$$\begin{aligned} |\Phi(\boldsymbol{x}^N) - \widehat{\Phi}(m^N_{\boldsymbol{x}})| &= |\widetilde{\Phi}(m^N_{\boldsymbol{x}}) - \int_{(\mathbb{R}^d)^N} \widetilde{\Phi}(m^N_{\boldsymbol{y}}) d(m^N_{\boldsymbol{x}})^{\otimes N}(d\boldsymbol{y})| \\ &\leq C \int_{(\mathbb{R}^d)^N} \mathbf{d}_2(m^N_{\boldsymbol{y}}, m^N_{\boldsymbol{x}}) m^N_{\boldsymbol{x}}(d\boldsymbol{y}) \leq C M_p^{1/p}(m^N_{\boldsymbol{x}}) N^{-\gamma'_{p,d}}. \end{aligned} \tag{4.3.15}$$

We can also write

$$\widehat{\Phi}(m_{\boldsymbol{x}_0}) = \mathbb{E}\Big[\sup_{t_0 \leq t \leq T} \mathbf{d}_2(m_t^{(t_0, m^N_{\boldsymbol{x}_0})}, m^N_{\boldsymbol{Y}_t})\Big],$$

where $\boldsymbol{Y} = (Y^1, .., Y^N)$ is defined using the same dynamics as in (4.3.14) but with initial conditions $Y^i_{t_0} = \xi^i$, the $(\xi^i)_{i=1,...,N}$ being i.i.d. with common law $m^N_{\boldsymbol{x}_0}$.

It follows, from a standard "asynchronous coupling argument" (see, for example the Proof of Theorem 1.10 in [93]), that

$$\widehat{\Phi}(m_{\boldsymbol{x}_0}) \leq C(1 + M_q(m^N_{\boldsymbol{x}_0})) N^{-\gamma'_{p,d}},$$

which completes the proof. □

The next lemma shows how the estimate from Lemma 4.3.6 can be used to improve the rate of convergence on small tubes.

Lemma 4.3.8. *Fix* $(t_0, m_0) \in \mathcal{O}$ *and assume that* $r > 0$ *be small enough so that the conclusion of*

Lemma 4.3.4 holds. Then, for $R > 0$, there exists a constant $C > 0$ independent of N such that, for all $(t, \boldsymbol{x}) \in \mathcal{T}^N_{r,R}$,

$$\mathcal{U}^N(t, \boldsymbol{x}) - \mathcal{V}^N(t, \boldsymbol{x}) \leq \frac{C}{N} + \sup_{(s,\boldsymbol{y}) \in \mathcal{T}^N_{r,R}} \Big(\mathcal{U}^N(s, \boldsymbol{y}) - \mathcal{V}^N(s, \boldsymbol{y}) \Big) \mathbb{P}[\tau^{N,t,\boldsymbol{x}} < T],$$

where

$$\tau^{N,t,\boldsymbol{x}} := \inf\{s > t : \boldsymbol{X}^{N,t,\boldsymbol{x}}_s \in (\mathcal{T}^N_{r,R})^c\} \wedge T.$$

Proof. Recall that U^N satisfies (4.3.4) on $\mathcal{T}^N_r = \mathcal{T}^N_r(t_0, m_0)$. Therefore, using a standard verification argument, we find, for any $(t, \boldsymbol{x}) \in \mathcal{T}^N_r$, the formula

$$\begin{aligned} \mathcal{U}^N(t, \boldsymbol{x}) = \inf_{\boldsymbol{\alpha}=(\alpha^1,\ldots,\alpha^N)} \mathbb{E}\Big[\int_t^\tau \Big(\frac{1}{N} \sum_{i=1}^N L(X^i_s, \alpha^i_s) + \mathcal{F}(m^N_{\boldsymbol{X}_s}) \\ + E^N(s, \boldsymbol{X}_s) \Big) ds + \mathcal{U}^N(\tau, \boldsymbol{X}_\tau) \Big], \end{aligned} \tag{4.3.16}$$

subject to the dynamics

$$dX^i_s = \alpha^i_s ds + \sqrt{2} dW^i_s \quad t \leq s \leq \tau = \inf\{s > t : X^i_s \in (\mathcal{T}^N_r)^c\} \text{ and } X^i_t = x^i.$$

We note that both X and τ are impacted by the choice of $\boldsymbol{\alpha}$.

Similarly, we have

$$\mathcal{V}^N(t, \boldsymbol{x}) = \inf_{\boldsymbol{\alpha}} \mathbb{E}\Big[\int_t^\tau \frac{1}{N} \sum_{i=1}^N L(X^i_s, \alpha^i_s) + \mathcal{F}(m^N_{\boldsymbol{X}_s}) ds + \mathcal{V}^N(\tau, \boldsymbol{X}_\tau) \Big], \tag{4.3.17}$$

subject to the same dynamics.

By dynamic programming, the optimal control for the optimization problem in (4.3.17) is the feedback $\alpha^i_s = -D_p H(X^i_s, N D_{x^i} \mathcal{V}^N(s, \boldsymbol{X}_s))$, and when this control is played $\boldsymbol{X} = \boldsymbol{X}^{N,t,\boldsymbol{x}}$ and

$\tau = \tau^{N,t,\boldsymbol{x}}$.

Thus testing this control in the optimization problem (4.3.16), we find

$$\begin{aligned}
\mathcal{U}^N(t,\boldsymbol{x}) - \mathcal{V}^N(t,\boldsymbol{x}) &\leq \mathbb{E}\Big[\int_t^{\tau^{N,t,\boldsymbol{x}}} E_N(s,\boldsymbol{X}_s)ds\Big] + \mathbb{E}\Big[\mathcal{U}^N(\tau^{N,t,\boldsymbol{x}},\boldsymbol{X}_{\tau^{N,t,x}}) - \mathcal{V}^N(\tau^{N,t,\boldsymbol{x}},\boldsymbol{X}_{\tau^{N,t,x}})\Big] \\
&\leq \frac{C}{N} + \mathbb{E}\Big[\Big(\mathcal{U}^N(\tau^{N,t,\boldsymbol{x}},X_\tau) - \mathcal{V}^N(\tau^{N,t,\boldsymbol{x}},X_{\tau^{N,t,x}})\Big)\mathbf{1}_{\tau^{N,t,x}<T} \\
&\qquad + \Big(\mathcal{U}^N(\tau^{N,t,\boldsymbol{x}},X_{\tau^{N,t,x}}) - \mathcal{V}^N(\tau^{N,t,\boldsymbol{x}},X_{\tau^{N,t,x}})\Big)\mathbf{1}_{\tau^{N,t,x}=T}\Big] \\
&\leq \frac{C}{N} + \sup_{(t,\boldsymbol{x})\in\mathcal{T}_r^N}\Big(\mathcal{U}^N - \mathcal{V}^N\Big)\mathbb{P}\Big[\tau^{N,t,\boldsymbol{x}} < T\Big],
\end{aligned}$$

where we used in the first inequality the fact that $|E^N| \leq C/N$ and in the last inequality the fact that $U^N(T,\boldsymbol{x}) = \mathcal{G}(m^N_{\boldsymbol{x}}) = V^N(T,\boldsymbol{x})$.

□

We now proceed with the proof of Proposition 4.3.2.

Proof of Proposition 4.3.2. Let $\gamma_{d,p}$ be the exponent appearing in Lemma 4.3.6, and choose $M \in \mathbb{N}$ with $M-1 \geq 1/\gamma_{d,p}$, $R_M > 0$ large enough so that $m_0 \in B^p_{R_M}$, and, for $m = 2,...,M$, define R_{m-1} inductively by $R_{m-1} = R_m + C_p$, C_p being the constant appearing in (4.3.8) or, in other words, $C_m = C_M + (M-m)C_p$ for $m = 1,...,M$. The important point is that we have

$$m_0 \in B^p_{R_M} \subset B^p_{R_{M-1}} \subset ... \subset B^p_{R_1} \text{ and } R_m - R_{m-1} = C_P.$$

Next, choose $r_2^{(1)} > 0$ small enough so that the conclusion of Lemma 4.3.4 holds, and then construct, for $i = 1,2$ and $m = 1,...,M,$, $r_i^{(m)}$, in such a way that

$$\begin{cases} r_2^{(m+1)} = r_1^{(m)} < r_2^{(m)} \text{ for } m = 1,...,M-1, \text{and} \\ \text{there exists } r^{(m)} \text{ such that } r_1^{(m)} < r^{(m)} < r_2^{(m)} \text{ satisfy the hypotheses of Lemma 4.3.6.} \end{cases}$$

Lemma 4.3.3 makes it clear that we can choose $r_i^{(m)}$ satisfying these conditions. To be clear, the important point is that we have

$$0 < r_1^{(M)} < r_2^{(M)} = r_1^{(M-1)} < \dots < r_2^{(2)} = r_1^{(1)} < r_2^{(1)}, \tag{4.3.18}$$

and that, for $m = 1, \dots, M$, there is a $r^{(m)}$ such that the triple $r_1^{(m)} < r^{(m)} < r_2^{(m)}$ satisfies the hypotheses of Lemma 4.3.6.

Combining Lemma 4.3.8 and Lemma 4.3.6, we find that, for each $m = 2, \dots, M$,

$$\begin{aligned} \sup_{(t,\boldsymbol{x})\in\mathcal{T}^N_{r_1^{(m)},R_m}} \Big(\mathcal{U}^N(t,\boldsymbol{x}) - \mathcal{V}^N(t,\boldsymbol{x})\Big) &\le \frac{C}{N} + \sup_{(t,\boldsymbol{x})\in\mathcal{T}^N_{r_2^{(m)},R_{m-1}}} \Big(\mathcal{U}^N - \mathcal{V}^N\Big) N^{-\gamma_{d,p}} \\ &= \frac{C}{N} + \sup_{(t,\boldsymbol{x})\in\mathcal{T}^N_{r_1^{(m-1)},R_{m-1}}} \Big(\mathcal{U}^N - \mathcal{V}^N\Big) N^{-\gamma_{d,p}} \end{aligned} \tag{4.3.19}$$

Then (4.3.19) and the global estimate (4.2.4) easily yield, through an inductive argument starting with $m = 1$, that, for each $m = 1, \dots, M$,

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}^N_{r_1^{(m)},R_m}} \Big(\mathcal{U}^N(t,\boldsymbol{x}) - \mathcal{V}^N(t,\boldsymbol{x})\Big) \le C/N^{1\wedge\left(\beta_d+(m-1)\gamma_{d,p}\right)}$$

Since $M - 1 > 1/\gamma_{d,p} \ge 1$ by design, we see in particular that

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}^N_{r_1^{(M)},R_M}} \Big(\mathcal{U}^N(t,\boldsymbol{x}) - \mathcal{V}^N(t,\boldsymbol{x})\Big) \le C/N,$$

which together with Proposition 4.3.1 completes the proof.

□

4.4 The proof of Theorem 4.2.2

In this section we prove the convergence result for the gradients. Once again, we fix a $p > 2$ throughout the section. As in the proof of Theorem 4.2.1, a useful first step is to observe that it suffices to prove the following proposition.

Proposition 4.4.1. *Let* $(t_0, m_0) \in \mathcal{O}$ *and* $R > 0$. *Then there exists* $r, C > 0$ *such that, for each* $N \in \mathbb{N}$, $i = 1, ..., N$ *and* $(t, \boldsymbol{x}) \in \mathcal{T}^N_{r,R}(t_0, m_0)$,

$$|ND_{x^i}\mathcal{V}^N(t, \boldsymbol{x}) - D_m\mathcal{U}(t, m^N_{\boldsymbol{x}})| \leq \frac{C}{N}.$$

It is easy to see that Proposition 4.4.1 implies Theorem 4.2.2.

Proof of Theorem 4.2.2. Theorem 4.2.2 follows from Proposition 4.4.1 exactly as Theorem 4.2.1 followed from Proposition 4.3.2.

□

The rest of this section is devoted to proving Proposition 4.4.1. Before we proceed, we fix some of the notations that will be used throughout the section.

First we define

$$\mathcal{V}^{N,i} = D_{x^i}\mathcal{V}^N \text{ and } \mathcal{U}^{N,i} = D_{x^i}\mathcal{U}^N \text{ for } i \in \{1, \cdots, N\}.$$

Using the the PDEs for $\mathcal{V}^N$ and $\mathcal{U}^N$, it is easy to check that $\mathcal{V}^{N,i}$ satisfies, in $[0, T) \times (\mathbb{R}^d)^N$,

$$\begin{aligned} &-\partial_t\mathcal{V}^{N,i} - \sum_j \Delta_{x^j}\mathcal{V}^{N,i} + \sum_j D_pH(x^j, NV^{N,j})D_{x^j}\mathcal{V}^{N,j} \\ &\qquad + \frac{1}{N}D_{x^i}H(x^i, N\mathcal{V}^{N,i}) = \mathcal{F}^{N,i} := \frac{1}{N}D_m\mathcal{F}(t, m^N_{\boldsymbol{x}}, x^i). \end{aligned} \tag{4.4.1}$$

Meanwhile, if $\mathcal{U}$ were smooth enough on some tube $\mathcal{T}_r = \mathcal{T}_r(t_0, m_0)$, then we would have

$$\begin{aligned} -\partial_t \mathcal{U}^{N,i} - \sum_j \Delta_{x^j} \mathcal{U}^{N,i} + \sum_j D_p H(x^j, N\mathcal{U}^{N,j}) D_{x^j} \mathcal{U}^{N,i} \\ + \frac{1}{N} D_{x^i} H(x^i, N\mathcal{U}^{N,i}) = \mathcal{F}^{N,i} + E^{N,i} \text{ in } \mathcal{T}^N, \end{aligned} \tag{4.4.2}$$

with

$$E^{N,i}(t, \boldsymbol{x}) = -\frac{2}{N^2} D_x \operatorname{tr}[D_{mm}\mathcal{U}(t, m^N_{\boldsymbol{x}}, x^i, x^i)] - \frac{1}{N^3} \sum_{j=1}^N D_m[\operatorname{tr}(D_{mm}\mathcal{U}(t, m^N_{\boldsymbol{x}}, x^j, x^j)](x^j).$$

Of course, Theorem 4.2.4 does not give enough regularity to immediately justify the computation above, since it only shows that $D_{mm}\mathcal{U}$ Lipschitz in an appropriate local sense. Nevertheless, when combining the regularity results Theorem 4.2.4 and Proposition 4.6.8 with the reasoning in the proof of [131, Proposition 5.1], one easily obtains the following analogue of Lemma 4.3.4.

Lemma 4.4.2. *For any $(t_0, m_0) \in \mathcal{O}$, there is $r > 0$ such that $\overline{\mathcal{T}_r(t_0, m_0)} \subset \mathcal{O}$, and, for each $N \in \mathbb{N}$, the projection $\mathcal{U}^{N,i}$ defined by (4.3.3) is Lipschitz in time and C^1 in space on the set $\mathcal{T}^N_r(t_0, m_0)$, with spatial derivatives $D_{x^j}\mathcal{U}^{N,i}$ being Lipschitz continuous in $\boldsymbol{x}$ and such that the L^∞−function*

$$E^{N,i} := -\partial_t \mathcal{U}^{N,i} - \sum_j \Delta_{x^j} \mathcal{U}^{N,i} + \sum_j D_p H(x^j, N\mathcal{U}^{N,j}) D_{x^j} \mathcal{U}^{N,i} + \frac{1}{N} D_{x^i} H(x^i, N\mathcal{U}^{N,i}) - \mathcal{F}^{N,i}$$

satisfies, for each $N \in \mathbb{N}$, the bound

$$\|E^{N,i}(t, \boldsymbol{x})\|_{L^\infty(\mathcal{T}^N_r(t_0,m_0))} \le C/N.$$

The next Lemma gives an analogue of Lemma 4.3.8, but at the level of the gradients of $\mathcal{U}^N$ and $\mathcal{V}^N$.

Lemma 4.4.3. *Let $(t_0, m_0) \in \mathcal{O}$ and $r > 0$ be small enough so that the conclusion of Lemma 4.4.2* *holds. Moreover, let*

$$\tau_r^{N,t,\boldsymbol{x}} = \inf\{s > t : (s, \boldsymbol{X}_x^{N,t,\boldsymbol{x}}) \in (\mathcal{T}_r^N(t_0, m_0))^c\} \wedge T.$$

Then, there exists an independent of N constant $C > 0$ *such that, for all* $(t, \boldsymbol{x}) \in \mathcal{T}_r^N$,

$$\sum_{i=1}^N |\mathcal{U}^{N,i}(t, \boldsymbol{x}) - \mathcal{V}^{N,i}(t, \boldsymbol{x})|^2 \leq \frac{C}{N^3} + \sup_{(s,\boldsymbol{y}) \in \mathcal{T}_r^N} \Big(\sum_{i=1}^N |\mathcal{U}^{N,i}(s, \boldsymbol{y}) - \mathcal{V}^{N,i}(s, \boldsymbol{y})|^2 \Big) \mathbb{P}[\tau_r^{N,t,\boldsymbol{x}} < T].$$

Proof. For the sake of notational simplicity, we give the proof for $d = 1$. The general case follows the same steps.

Fix $(t, \boldsymbol{x}) \in \mathcal{T}_r^N$ and consider the optimal trajectory $\boldsymbol{X} = \boldsymbol{X}^{(t,\boldsymbol{x})}$ for the N-particle control problem started from $(t, \boldsymbol{x})$. Let $\tau = \tau_r^{N,t,\boldsymbol{x}}$, drop the superscript N and write E_s^i for $E_s^{N,i}$ and set

$$\begin{aligned} Y_s^i &= \mathcal{V}^{N,i}(s, \boldsymbol{X}_s), \quad Z_s^i = \sqrt{2} D\mathcal{V}^{N,i}(s, \boldsymbol{X}_s), \\ \overline{Y}_s^i &= \mathcal{U}^{N,i}(s, \boldsymbol{X}_s), \quad \overline{Z}_s^i = \sqrt{2} D\mathcal{U}^{N,i}(s, \boldsymbol{X}_s). \end{aligned}$$

On the stochastic interval $[t, \tau)$, we have the dynamics

$$dY_s^i = \Big(\frac{1}{N} D_{x^i} H(X_s^i, N Y_s^i) - \mathcal{F}^{N,i}(s, \boldsymbol{X}_s) \Big) ds + Z_s^i d\boldsymbol{W}_s,$$

and

$$\begin{aligned} d\overline{Y}_s^i = \Big(& \frac{1}{N} D_{x^i} H(X_s^i, N \overline{Y}_s^i) - \mathcal{F}^{N,i}(s, \boldsymbol{X}_s) - E^{N,i}(s, \boldsymbol{X}_s) \\ & + \sum_j \Big(D_p H(X_s^j, N \overline{Y}^j) - D_p H(X_s^j, N Y^j) \Big) D_{x^j} \mathcal{U}^{N,i}(s, \boldsymbol{X}_s) \Big) ds + \overline{Z}_s^i d\boldsymbol{W}_s, \end{aligned}$$

where we have written $\boldsymbol{W} = (W^1, ..., W^N)$ for simplicity.

We set

$$\Delta Y_s^i = Y_s^i - \overline{Y}_s^i \text{ and } \Delta Z_s^i = Z_s^i - \overline{Z}_s^i,$$

and, for each $i, j \in \{1, \cdots, N\}$, define the process $A^{i,j}$ by

$$A_s^{i,j} = \frac{1}{|\Delta Y^j|^2} D_{x^j} \mathcal{U}^{N,i}(s, \boldsymbol{X}_s)\Big(D_p H(X_s^j, NY^j) - D_p H(X_s^j, N\overline{Y}^j)\Big)(\Delta Y^j)^T 1_{\Delta Y^j \neq 0}$$
$$+ \frac{1}{N|\Delta Y^i|^2}\Big(D_{x^i} H(X_s^i, NY_s^i) - D_{x^i} H(X_s^i, N\overline{Y}_s^i)\Big)(\Delta Y_s^i)^T 1_{\Delta Y^i \neq 0} 1_{i=j}.$$

We may now write

$$d\Delta Y_s^i = \Big(\sum_{j=1}^N A_s^{i,j} \Delta Y_s^j + E_s^i\Big) ds + \Delta Z_s^i d\boldsymbol{W}_s.$$

The key point about the coefficients $A^{i,j}$ is that, for some C independent of N,

$$|A^{i,j}| \leq C/N + C 1_{i=j}.$$

Next, we compute

$$d(\sum_{i=1}^N |\Delta Y_s^i|^2) = 2(\sum_{i,j=1}^N A_s^{i,j} \Delta Y_s^i \Delta Y_s^j + \sum_{i=1}^N \Delta Y_s^i E_s^i + \frac{1}{2}\sum_{i=1}^N |\Delta Z_s^i|^2) ds + 2\sum_{i=1}^N \Delta Y_s^i \Delta Z_s^i d\boldsymbol{W}_s.$$

Integrating and taking expectations, we find, for some, independent of N, C,

$$\begin{aligned}
\mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{s\wedge\tau}|^2\Big] &= \mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{\tau}|^2\Big] - 2\mathbb{E}\Big[\int_{s\wedge\tau}^{\tau}\Big(\sum_{i,j=1}^{N}A^{i,j}_r\Delta Y^i_r\Delta Y^j_r + \sum_{i=1}^{N}\Delta Y^i_r E^i_r + \frac{1}{2}\sum_{i=1}^{N}|\Delta Z^i_r|^2\Big)dr\Big] \\
&\leq \mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{\tau}|^2\Big] + C\mathbb{E}\Big[\int_{s\wedge\tau}^{\tau}\Big(\sum_{i=1}^{N}|\Delta Y^i_r|^2 + \sum_{i=1}^{N}|E^i_r|^2\Big)dr\Big] \\
&\leq \mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{\tau}|^2\Big] + C\mathbb{E}\Big[\int_{s}^{T}\Big(\sum_{i=1}^{N}|\Delta Y^i_{r\wedge\tau}|^2 + \sum_{i=1}^{N}|E^i_{r\wedge\tau}|^2\Big)dr\Big] \\
&= \mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{\tau}|^2\Big] + C\int_{s}^{T}\mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{r\wedge\tau}|^2 + \sum_{i=1}^{N}|E^i_{r\wedge\tau}|^2\Big]dr,
\end{aligned}$$

with the first bound coming from Young's inequality and the bounds on $A^{i,j}$.

Applying Grönwall's inequality to the function $\psi(s) = \mathbb{E}[\sum_i|\Delta Y^i_s|^2]$ and using the bounds on E^i, we get

$$\begin{aligned}
\sum_{i=1}^{N}|\mathcal{U}^{N,i}(t,\boldsymbol{x}) - \mathcal{V}^{N,i}(t,\boldsymbol{x})|^2 &= \mathbb{E}[\sum_{i=1}^{N}|\Delta Y^i_t|^2] \leq \frac{C}{N^3} + C\mathbb{E}\Big[\sum_{i=1}^{N}|\Delta Y^i_{\tau}|^2\Big] \\
&= \frac{C}{N^3} + C\mathbb{E}\Big[\sum_{i=1}^{N}|\mathcal{V}^{N,i}(\tau,\boldsymbol{X}_\tau) - \mathcal{U}^{N,i}(\tau,\boldsymbol{X}_\tau)|^2\Big] \\
&\leq \frac{C}{N^3} + \sup_{(s,\boldsymbol{y})\in\mathcal{T}^N_r}\Big(\sum_{i=1}^{N}|\mathcal{U}^{N,i}(s,\boldsymbol{y}) - \mathcal{V}^{N,i}(s,\boldsymbol{y})|^2\Big)\mathbb{P}[\tau < T],
\end{aligned}$$

which completes the proof.

□

The following Lemma can be proved using Lemma 4.4.3, exactly as Proposition 4.3.2 is proved using Lemma 4.3.8.

Lemma 4.4.4. *Let $(t_0, m_0) \in \mathcal{O}$ and $R > 0$. Then there exist $r, C > 0$ such that, for each $N \in \mathbb{N}$ and* *each* $(t,\boldsymbol{x}) \in \mathcal{T}^N_{r,R}(t_0,m_0)$,

$$\sum_{i=1}^N |ND_{x^i}V^N(t,\boldsymbol{x}) - D_mU(t,m^N_{\boldsymbol{x}})|^2 \leq \frac{C}{N}.$$

We may now proceed with the proof of Proposition 4.4.1.

Proof of Proposition 4.4.1. Fix $(t_0,m_0) \in \mathcal{O}$ and $R > 0$ and let $r > 0$ be given by Lemma 4.4.4. With the notation of Lemma 4.4.3, we now have, for each $i \in \{1,\cdots,N\}$, the dynamics

$$d\Delta Y^i_s = \Big(\sum_j^N A^{i,j}_s \Delta Y^j_s + E^i_s\Big)ds + \Delta Z^i_s d\boldsymbol{W}_s, \tag{4.4.3}$$

Therefore,

$$d(|\Delta Y^i_s|^2) = (\sum_{j=1}^N A^{i,j}_s \Delta Y^i_s \Delta Y^j_s + \Delta Y^i_s E^i_s + \frac{1}{2}|\Delta Z^i_s|^2)ds + \Delta Y^i_s \Delta Z^i_s d\boldsymbol{W}_s,$$

which yields

$$\begin{aligned}
\mathbb{E}\Big[|\Delta Y^i_{s\wedge\tau}|^2\Big] &= \mathbb{E}\Big[|\Delta Y^i_\tau|^2\Big] - \mathbb{E}\Big[\int_{s\wedge\tau}^\tau (\sum_{j=1}^N A^{i,j}_r \Delta Y^i_r \Delta Y^j_r + \Delta Y^i_r E^i_r + \frac{1}{2}|\Delta Z^i_r|^2)dr\Big] \\
&\leq \mathbb{E}\Big[|\Delta Y^i_\tau|^2\Big] - \mathbb{E}\Big[\int_{s\wedge\tau}^\tau (\sum_{j\neq i}^N A^{i,j}_r \Delta Y^i_r \Delta Y^j_r + A^{i,i}\Delta Y^i_r \Delta Y^i_r + \Delta Y^i_r E^i_r)dr\Big] \\
&\leq \mathbb{E}\Big[|\Delta Y^i_\tau|^2\Big] + C\mathbb{E}\Big[\int_{s\wedge\tau}^\tau \Big(|\Delta Y^i_r|^2 + N\sum_{j\neq i}|A^{i,j}|^2|\Delta Y^j_r|^2 + |E^i_r|^2\Big)dr\Big] \\
&\leq \mathbb{E}\Big[|\Delta Y^i_\tau|^2\Big] + C\mathbb{E}\Big[\int_s^T \Big(|\Delta Y^i_{r\wedge\tau}|^2 + \frac{1}{N}\sum_{j\neq i}|\Delta Y^j_{r\wedge\tau}|^2 + \frac{1}{N^4}\Big)dr\Big] \\
&\leq \mathbb{E}\Big[|\Delta Y^i_\tau|^2\Big] + C\int_s^T \mathbb{E}\Big[|\Delta Y^i_{r\wedge\tau}|^2 + \frac{1}{N^4}\Big]dr,
\end{aligned}$$

where in the second inequality we used Young's inequality in the form

$$A_r^{i,j}\Delta Y_r^i \Delta Y_r^j \leq \frac{1}{N}|\Delta Y_r^i|^2 + CN|A_r^{i,j}|^2|\Delta Y_r^j|^2,$$

in the third inequality we used the bounds on $|E_r^i|$ as well as the bounds on $A^{i,j}$ for $i \neq j$, and, finally, in the last inequality we used Lemma 4.4.4.

Thus by Gronwall and the fact that

$$|\Delta Y_\tau^i|^2 \leq \begin{cases} 0 & \text{if } \tau = T, \\ \sup_{(s,\boldsymbol{y})\in\mathcal{T}_r^N}\left(|\mathcal{U}^{N,i}(s,\boldsymbol{y}) - \mathcal{V}^{N,i}(s,\boldsymbol{y})|^2\right) & \text{if } \tau < T, \end{cases}$$

we have that

$$|\mathcal{U}^{N,i}(t,\boldsymbol{x}) - \mathcal{V}^{N,i}(t,\boldsymbol{x})|^2 \leq \frac{C}{N^4} + \sup_{(s,\boldsymbol{y})\in\mathcal{T}_r^N}\left(|\mathcal{U}^{N,i}(s,\boldsymbol{y}) - \mathcal{V}^{N,i}(s,\boldsymbol{y})|^2\right)\mathbb{P}[\tau_r^{N,t,\boldsymbol{x}} < T].$$

To conclude, we follow the same procedure as in the proof of Proposition 4.3.2 to show that there exists a $r_0 < r$ such that

$$\sup_{(s,\boldsymbol{y})\in\mathcal{T}_{r_0}^N}\left(|\mathcal{U}^{N,i}(s,\boldsymbol{y}) - \mathcal{V}^{N,i}(s,\boldsymbol{y})|^2\right) \leq \frac{C}{N^4},$$

which proves the result.

□

4.5 The proof of Proposition 4.2.3

The proof of Proposition 4.2.3 follows from the following two Lemmas.

Lemma 4.5.1. *Let Assumption 2 hold. Fix $(t_0, m_0) \in \mathcal{O}$, assume that m_0 satisfies the condition* (4.2.7) *from the statement of Proposition 4.2.3, and, for each $N \in \mathbb{N}$, denote by $\widetilde{\mathbf{X}}^N =$*

$(\widetilde{X}^{N,1}, ..., \widetilde{X}^{N,N})$ *the solution to*

$$d\widetilde{X}_t^{N,i} = -D_pH(\widetilde{X}^{N,i}, D_mU(t, m^N_{\widetilde{\mathbf{X}}_t^N}, \widetilde{X}_t^{N,i}))dt$$
$$+ \sqrt{2}dW_t^i \quad for \quad t_0 \le t \le \tau = \inf\left\{t > t_0 : (t, m^N_{\widetilde{\boldsymbol{X}}_t}) \in O^c\right\} \wedge T, \tag{4.5.1}$$
$$\widetilde{X}_{t_0}^{N,i} = \xi^i,$$

where $(\xi^i)_{i\in\mathbb{N}}$ *are i.i.d. with common law* m_0*. Then, there exists an* $r_0 > 0$ *and a constant* $C > 0$ *such that for each* $0 < r < r_0$ *and* $N \ge C/\min(r, r^{d+8})$*, we have*

$$\mathbb{P}\Big[\sup_{t_0\le t\le\tau} \mathbf{d}_2(m^N_{\widetilde{\boldsymbol{X}}_t^N}, m_t^{(t_0,m_0)}) > r\Big] \le Ce^{-\frac{r^2}{C}N}.$$

Proof. By Proposition 4.6.4 below, we can choose r_0 small enough so that $\overline{\mathcal{T}_{r_0}(t_0, m_0)} \subset O$ and $(x, m) \mapsto -D_pH(x, D_m\mathcal{U}(t, m, x))$ is uniformly Lipschitz on $\mathcal{T}_r(t_0, m_0)$ with m endowed with the $\mathbf{d}_2-$metric. We then extend, as in the proof of Lemma 4.3.6, to find a measurable map

$$b(t, x, m) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

which is globally Lipschitz in (x, m) and such that

$$b(t, x, m) = -D_pH(x, D_mU(t, m, x)) \text{ for } (t, m) \in \mathcal{T}_{r_0}.$$

Let $\boldsymbol{Y}^N = (\boldsymbol{Y}^{N,1}, ..., \boldsymbol{Y}^{N,N})$ be the unique solution on $[t_0, T]$ of the SDE

$$dY_t^{N,i} = b(t, Y_t^{N,i}, m^N_{\boldsymbol{Y}_t^N})dt + \sqrt{2}dW_t^i \text{ in } (t_0, T] \text{ and } Y_{t_0} = \xi^i, \tag{4.5.2}$$

and notice that $\boldsymbol{Y} = \widetilde{\boldsymbol{X}}$ on $[0, \sigma)$, where $\sigma = \inf\{t \ge t_0 : (t, m^N_{\boldsymbol{Y}_t}) \notin \mathcal{T}_r\}$.

Now for $r < r_0$, we can use [143, Corollary 3.5] to conclude

$$\mathbb{P}\Big[\sup_{t_0 \le t \le \tau} \mathbf{d}_2(m^N_{\widetilde{\boldsymbol{X}}^N_t}, m_t^{(t_0,m_0)}) > r\Big] = \mathbb{P}\Big[\sup_{t_0 \le t \le \tau} \mathbf{d}_2(m^N_{\boldsymbol{Y}^N_t}, m_t^{(t_0,m_0)}) > r\Big] \le Ce^{-\frac{r^2}{C}N}.$$

□

Lemma 4.5.2. *Let $\widetilde{X}^{N,i}$ and $X^{N,i}$ be defined by* (4.2.8) *and* (4.5.1) *respectively, fix $R > 0$, choose r small enough so that the conclusion of Proposition 4.4.1 holds on $\mathcal{T}_r(t_0, m_0)$, and let*

$$\sigma = \inf\{t > t_0 : (t, m^N_{\boldsymbol{Y}^N_t}) \notin \mathcal{T}^N_{r,R}(t_0, m_0))\ or\ (t, m^N_{\boldsymbol{X}^N_t}) \notin \mathcal{T}^N_{r,R}(t_0, m_0))\} \wedge T. \tag{4.5.3}$$

Then there exists a constant $C > 0$ such that, a.s.,

$$\sup_{t_0 \le t \le \sigma} \Big\{\frac{1}{N}\sum_{i=1}^N |X_t^{N,i} - Y_t^{N,i}|\Big\} \le C/N.$$

Proof. For simplicity of notation, we write

$$b^{N,i}(t, \boldsymbol{x}) = -D_pH(x^i, ND_{x^i}\mathcal{V}^N(t, \boldsymbol{x})) \text{ and } \widetilde{b}^{N,i}(t, \boldsymbol{x}) = -D_pH(x^i, D_m\mathcal{U}(t, m^N_{\boldsymbol{x}}, x^i)).$$

On $[t_0, \sigma)$ we can rewrite the dynamics of $\boldsymbol{X}^N$ as

$$dX_t^{N,i} = b^{N,i}(t, \boldsymbol{X}_t^{N,i})dt + \sqrt{2}dW_t^i = \Big(\widetilde{b}^{N,i}(t, \boldsymbol{X}_t^{N,i}) + E_t^i\Big)dt + \sqrt{2}dW_t^i,$$

where, in view or Theorem 4.2.2,

$$|E_t^i| = |b^{N,i}(t, \boldsymbol{X}_t^{N,i}) - \widetilde{b}^{N,i}(t, \boldsymbol{X}_t^{N,i})| \le C/N.$$

Thus setting $\Delta X_t^{N,i} = X_t^{N,i} - \widetilde{X}_t^{N,i}$, we have, for $t_0 \le t \le \sigma$,

$$\Delta X_t^{N,i} = \int_{t_0}^t \Big(\widetilde{b}^{N,i}(t, X_t) - \widetilde{b}^{N,i}(t, \widetilde{X}_t) + E_t^{N,i}\Big).$$

Using the bounds on $E^{N,i}$ and the fact that the regularity of U implies $D_{x^j}\widetilde{b}^{N,i} \le C/N + C1_{i=j}$, we easily get

$$\frac{1}{N}\sum_{i=1}^N |\Delta X_t^{N,i}|^2 \le C/N^2 + C\int_{t_0}^t \frac{1}{N}\sum_{i=1}^N |\Delta X_s^{N,i}|^2 ds,$$

and we conclude with Gronwall's inequality.

□

The proof of Proposition 4.2.3. We fix $R > 0$ to be chosen later, choose r_0 small enough that the conclusions of the preceding two Lemmas are valid, and let $\widetilde{\boldsymbol{X}}^N$ and σ be defined as (4.5.1) and (4.5.3) respectively.

We note that to prove Proposition 4.2.3, it suffices to show an estimate of the form

$$\mathbb{P}[\sigma < T] \le C\exp(-\frac{r^2}{C}N^{1-\eta}).$$

To this end, we remark that

$$\sigma \ge \sigma_R \wedge \sigma_T \wedge \widetilde{\sigma}_R \wedge \widetilde{\sigma}_T \wedge T,$$

where

$$\sigma_R = \inf\{t \ge t_0 : (t_0, m^N_{\boldsymbol{X}_t^N}) \notin B_R^p\} \text{ and } \widetilde{\sigma}_R = \inf\{t \ge t_0 : (t_0, m^N_{\widetilde{\boldsymbol{X}^N}_t}) \notin B_R^p\},$$

and

$$\sigma_{\mathcal{T}} = \inf\{t \geq t_0 : (t_0, m^N_{\boldsymbol{X}^N_t}) \notin \mathcal{T}_r\} \text{ and } \widetilde{\sigma}_{\mathcal{T}} = \inf\{t \geq t_0 : (t_0, m^N_{\widetilde{\boldsymbol{X}}^N_t}) \notin \mathcal{T}_{r/2}\}.$$

Then we have

$$\begin{aligned}
\mathbb{P}\Big[\sup_{t_0 \leq t \leq T} \mathbf{d}_2(m^N_{\boldsymbol{X}^N_t}, m^{(t_0,m_0)}_t) > r\Big] &\leq \mathbb{P}[\sigma_R < T] + \mathbb{P}[\widetilde{\sigma}_R < T] + \mathbb{P}[\widetilde{\sigma}_{\mathcal{T}} < T] \\
&\quad + \mathbb{P}\Big[\sup_{t_0 \leq t \leq T} \mathbf{d}_2(m^N_{\boldsymbol{X}^N_t}, m^{(t_0,m_0)}_t) > r \text{ and } \sigma_R = \widetilde{\sigma}_R = \widetilde{\sigma}_{\mathcal{T}} = T\Big] \\
&\leq \mathbb{P}[\sigma_R < T] + \mathbb{P}[\widetilde{\sigma}_R < T] + \mathbb{P}[\widetilde{\sigma}_{\mathcal{T}} < T] + \mathbb{P}\Big[\sup_{t_0 \leq t \leq \sigma} \mathbf{d}_2(m^N_{\boldsymbol{X}^N_t}, m^N_{\widetilde{\boldsymbol{X}}^N_t}) > \frac{r}{2}\Big].
\end{aligned}$$

By Lemma 4.5.2, the last term in the final line above vanishes when $N \geq \frac{C}{r}$.

To bound $\mathbb{P}[\sigma_R < T]$, we argue as in the proof of Lemma 4.5.1 to conclude that

$$\sup_{t_0 \leq t \leq T} \mathbf{d}_p(m^N_{\boldsymbol{X}_t}, m^N_{\boldsymbol{x}_0}) \leq C\Big(1 + \frac{1}{N}\sum_{i=1}^N \sup_{t_0 \leq t \leq T} |W^i_t - W^i_{t_0}|^p\Big).$$

We conclude that

$$\mathbb{P}[\sigma_R < T] \leq \mathbb{P}\Big[C\Big(1 + \frac{1}{N}\sum_{i=1}^N \sup_{t_0 \leq t \leq T} |W^i_t - W^i_{t_0}|^p\Big) > R^p - C\Big],$$

and so, applying Corollary 3.1 of [356], we see that for any K we can choose R large enough so that

$$\mathbb{P}[\sigma_R < T] \leq \exp(-KN^{2/p})$$

An identical argument shows the same estimate for $\mathbb{P}[\widetilde{\sigma}_R < T]$.

Finally, by Lemma 4.5.1, we have

$$\mathbb{P}[\widetilde{\sigma}_{\mathcal{T}} < T] \leq C \exp(-\frac{r^2}{C} N).$$

Since $p > 2$ is arbitrary, this completes the proof. □

4.6 Regularity

In this section we show the necessary regularity results for the value function U.

4.6.1 Terminology and notation

Throughout this part, we set

$$F(x, m) = \frac{\delta \mathcal{F}}{\delta m}(m, x) \text{ and } G(x, m) = \frac{\delta \mathcal{G}}{\delta m}(m, x).$$

We also make use of the notation

$$\frac{\delta F}{\delta m}(x, m)(\rho) = \langle \frac{\delta F}{\delta m}(x, m, y), \rho(dy) \rangle, \quad \frac{\delta^2 F}{\delta m^2}(x, m)(\rho_1)(\rho_2) = \langle \langle \frac{\delta^2 F}{\delta m^2}(x, m, y, z), \rho_1(dy) \rangle, \rho_2(dz) \rangle,$$

whenever ρ, ρ_1, ρ_2 are distributions such that the above pairings make sense, and similar notations with G replacing F.

As in [92], we need to study a number of linear equations, which are obtained by linearizing the MFG system describing the optimal trajectories. For the reader's convenience, we list here all of the relevant equations.

We say that (u, m) is a solution to (MFG) with initial condition $m(t_0) = m_0$, if

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t)) \ \text{ in } \ (t_0, T) \times \mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(m D_p H(x, Du)) = 0 \ \text{ in } \ \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0, \cdot) = m_0 \text{ and } u(T, \cdot) = G(\cdot, m(T)) \ \text{ in } \ \mathbb{R}^d. \end{cases} \tag{MFG}$$

We say that (z, ρ) is a solution to (MFGL) driven by (u, m) and with initial condition $\rho(t_0) = \xi$, if

$$\begin{cases} -\partial_t z - \Delta z + D_p H(x, Du) \cdot Dz = \dfrac{\delta F}{\delta m}(x, m(t))(\rho(t)) \ \text{ in } \ (t_0, T) \times \mathbb{R}^d, \\ \partial_t \rho - \Delta \rho - \operatorname{div}(\rho D_p H(x, Du)) = \operatorname{div}(m D_{pp} H(x, Du) Dz) \ \text{ in } \ (t_0, T) \times \mathbb{R}^d, \\ \rho(t_0, \cdot) = \xi \text{ and } z(T, \cdot) = \dfrac{\delta G}{\delta m}(\cdot, m(T))(\rho(T)) \ \text{ in } \ \mathbb{R}^d. \end{cases} \tag{MFGL}$$

We say that (z, ρ) is a solution to (MFGLE) driven by (u, m) with initial condition $\rho(t_0) = \xi$ and forcing terms R^1, R^2, R^3 if

$$\begin{cases} -\partial_t z - \Delta z + D_p H(x, Du) \cdot Dz = \dfrac{\delta F}{\delta m}(x, m(t))(\rho) + R^1 \ \text{ in } \ (t_0, T) \times \mathbb{R}^d, \\ \partial_t \rho - \Delta \rho - \operatorname{div}(\rho D_p H(x, Du)) = \operatorname{div}(m D_{pp} H(x, Du) Dz) \\ \qquad\qquad\qquad\qquad\qquad + \operatorname{div}(R^2) \ \text{ in } \ (t_0, T) \times \mathbb{R}^d, \\ \rho(t_0, \cdot) = \xi \text{ and } z(T, \cdot) = \dfrac{\delta G}{\delta m}(\cdot, m(T))(\rho(T)) + R^3 \ \text{ in } \ \mathbb{R}^d. \end{cases} \tag{MFGLE}$$

Given a sufficiently smooth vector field $V : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and a bounded map $\Gamma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$, we say that that (z, ρ) is a solution to (MFGLG) driven by (u, m) with initial condition

$\rho(t_0) = \xi$ and forcing terms R^1, R^2, R^3 if

$$\begin{cases} -\partial_t z - \Delta z + V(t,x)\cdot Dz = \dfrac{\delta F}{\delta m}(x, m(t))(\rho) + R^1 \ \text{ in } \ (t_0, T)\times\mathbb{R}^d, \\ \partial_t \rho - \Delta\rho - \operatorname{div}(\rho V) = \sigma \operatorname{div}(m\Gamma Dz) + \operatorname{div}(R^2) \ \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ \rho(t_0,\cdot) = \xi \text{ and } z(T,\cdot) = \dfrac{\delta G}{\delta m}(\cdot, m(T))(\rho(T)) + R^3 \ \text{ in } \ \mathbb{R}^d, \end{cases} \tag{MFGLG}$$

where m solves

$$\begin{cases} \partial_t m - \Delta m - \operatorname{div}(mV) = 0 \ \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ m(t_0,\cdot) = m_0 \ \text{ in } \ \mathbb{R}^d. \end{cases} \tag{4.6.1}$$

We provided separate definitions for the above systems due to their frequent use. We note, however, that $MFGL$ is a special case of $MFGLE$, which in turn is a special case of MFGLG.

Finally, we recall the notion of strong stability used in [92] for the system

$$\begin{cases} -\partial_t z - \Delta z + V(t,x)\cdot Dz = \dfrac{\delta F}{\delta m}(x, m(t))(\rho) \ \text{ in } \ (t_0, T)\times\mathbb{R}^d, \\ \partial_t \rho - \Delta\rho - \operatorname{div}(\rho V) = \sigma \operatorname{div}(m\Gamma Dz) \ \text{ in } \ (t_0,T)\times\mathbb{R}^d, \\ \rho(t_0,\cdot) = \xi \ \text{ and } \ z(T,\cdot) = \dfrac{\delta G}{\delta m}(\cdot, m(T))(\rho(T)) \ \text{ in } \ \mathbb{R}^d. \end{cases} \tag{4.6.2}$$

We say that

$$\begin{aligned} &\text{the system (4.6.2) is strongly stable if, for any } \sigma \in [0,1], \\ &\text{its unique solution is } (z,\rho) = (0,0). \end{aligned} \tag{4.6.3}$$

4.6.2 Refinement of the results in [92]

The purpose of this subsection is to present sharpened versions of the results in [92] under the increased regularity of the data. The majority of them require only small adjustments. The only critical extensions are Lemma 4.6.1 for estimates on the linearized system MFGLE and Lemma

4.6.3 for the stability of controls. For this reason we include detailed proofs of these two results.

The following is a generalization of [92, Lemma 2.1]. The only improvement is in the norm of dependence on ξ.

Lemma 4.6.1. *Assume* (2) *and* (4.6.3). *There exists a neighborhood* $\mathcal{V}$ *of* (V, Γ) *in the topology of locally uniform convergence, and* $\eta, C > 0$ *such that, for any* $(V', t_0', \Gamma', R^{1,'}, R^{2,'}, R^{3,'}, \xi', \sigma')$ *with*

$$\begin{cases} (V', \Gamma') \in \mathcal{V},\ |t_0' - t_0| + d_2(m_0', m_0) \leq \eta,\ \|V'\|_{C^{1,3}} + \|\Gamma'\|_\infty \leq 2C_0,\ \sigma' \in [0,1], \\ R^{1,'} \in C^{\delta/2,\delta},\ R^{2,'} \in L^\infty([t_0, T], (W^{1,\infty})'(\mathbb{R}^d, \mathbb{R}^d)), R^{3,'} \in C^{2+\delta},\ \xi \in (C^{1+\delta})', \end{cases} \tag{4.6.4}$$

any solution (z', ρ') *to* (MFGLG) *associated with these data on* $[t_0', T]$ *and* m' *the solution to* (4.6.1) *with drift* V' *and initial condition* m_0' *at time* t_0' *satisfies*

$$\|z'\|_{C^{(2+\delta)/2,2+\delta}} + \sup_{t \in [t_0', T]} \|\rho'(t, \cdot)\|_{(C^{2+\delta})'} + \sup_{t' \neq t} \frac{\|\rho'(t', \cdot) - \rho'(t, \cdot)\|_{(C^{2+\delta})'}}{|t' - t|^{\delta/2}} \leq CM', \tag{4.6.5}$$

where

$$M' = \|\xi'\|_{(C^{1+\delta})'} + \|R^{1,'}\|_{C^{\delta/2,\delta}} + \sup_{t \in [t_0', T]} \|R^{2,'}(t)\|_{(W^{1,\infty})'} + \|R^{3,'}\|_{C^{2+\delta}}. \tag{4.6.6}$$

The proof is identical to the one in [92], where a careful inspection of the proofs of [92, Lemma 2.3] and [92, Lemma 2.1] shows that we may in fact use $\|\xi\|_{(C^{1+\delta})'}$ instead of $\|\xi\|_{(W^{1,\infty})'}$.

The following is a restatement of [92, Lemma 1.3] for measures in $\mathcal{P}_1(\mathbb{R}^d)$.

Lemma 4.6.2. *Assume* (2) *and let* (u, m) *be a solution of MFG. Then there exists* $C > 0$*, which is independent of* (t_0, m_0)*, such that*

$$\|u\|_{C^{(3+\delta)/2,3+\delta}} + \sup_{t \neq t'} \frac{d_1(m(t), m(t'))}{|t' - t|^{\frac{1}{2}}} \leq C \tag{4.6.7}$$

and

$$\sup_{t\in[t_0,T]}\int_{\mathbb{T}^d}|x|m(t,dx)\leq C\int_{\mathbb{T}^d}|x|m_0(dx). \tag{4.6.8}$$

We now present the critical improvement of [92, Lemma 2.9].

Lemma 4.6.3. *Assume* (2) *and fix* $(t_0, m_0) \in \mathcal{O}$*. There exist* $\theta, C > 0$ *such that, for any* t_0', m_0^1, m_0^2 *satisfying* $|t_0' - t_0| < \theta$ *and* $\mathbf{d}_2(m_0, m_0^i) < \theta$*, if* (m^i, α^i) *is the unique minimizer starting from* (t_0', m_0^i) *with associated multiplier* u^i *for* $i = 1$ *and* $i = 2$*, then*

$$\begin{aligned}&\|u^2-u^1\|_{C^{(2+\delta)/2,2+\delta}}+\sup_{t\in[t_0',T]}\mathbf{d}_1(m^2(t),m^1(t))\\&+\sup_{t'\neq t}\frac{\|(m^2-m^1)(t')-(m^2-m^1)(t)\|_{(C^{2+\delta})'}}{|t'-t|^{\delta/2}}\leq C\mathbf{d}_1(m_0^2,m_0^1).\end{aligned}$$

Remark 13. *We note that in the statement of Lemma 4.6.3* m_0, m_0^i *are close with respect to* $\mathbf{d}_2$ *while the result uses* $\mathbf{d}_1(m_0^2, m_0^1)$*. This is done to be consistent with the results already proven in [92], where the set* $\mathcal{O}$ *was shown to be open in the* $\mathbf{d}_2$*-topology. However, as we show below the function* $\mathcal{U}$ *is in fact regular in* $\mathbf{d}_1$ *inside the set* $\mathcal{O}$*. The use of the metrics* $\mathbf{d}_1$ *and* $\mathbf{d}_2$ *will appear throughout this section.*

Proof. Let (m, α) be the unique stable minimizer starting from (t_0, m_0) with multiplier u. Then [92, Lemma 2.6], yields that system MFGL is strongly stable.

Let $V = -D_pH(x, Du), \Gamma = -D_{pp}H(x, Du)$ and $\mathcal{V}$ be the corresponding neighborhood as described in Lemma 4.6.1. With the same argument as in [92, Lemma 2.9], by choosing $\theta > 0$ small enough, we have that

$$(V^i, \Gamma^i) \in \mathcal{V}$$

where $V^i = -D_pH(x, Du^i), \Gamma^i = -D_{pp}H(x, Du^i)$.

Given $\eta > 0$, by choosing $\theta > 0$ even smaller if necessary, we have that under our assumptions

$$\|u^2 - u^1\|_{C^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0',T]} \mathbf{d}_2(m^2(t), m^1(t)) < \eta. \tag{4.6.9}$$

Moreover, it is easy to check that, for a constant $C = C(T, H, \|D^2u^1\|_\infty, \|D^2u^2\|_\infty) > 0$ which is bounded by Lemma 4.6.2, we have

$$\sup_{t_0'\le t\le T} \mathbf{d}_1(m^2(t), m^1(t)) \le C(\mathbf{d}_1(m_0^2, m_0^1) + \|Du^2 - Du^1)\|_\infty). \tag{4.6.10}$$

Consider the pair

$$(v, \rho) = (u^2 - u^1, m^2 - m^1),$$

which solves (MFGLE) with

$$\begin{aligned}
R^1(t,x) &= H(x, Du^2) - H(x, Du^1) - D_pH(x, Du^1)\cdot(Du^2 - Du^1) \\
&\qquad + F(x, m^2) - F(x, m^1) - \frac{\delta F}{\delta m}(x, m^1(t))(m^2(t) - m^1(t)), \\
R^2(t,x) &= D_pH(x, Du^2)m^2 - D_pH(x, Du^1)m^1 - D_pH(x, Du^1)(m^2 - m^1) \\
&\qquad - D_{pp}H(x, Du^1)(Du^2 - Du^1)m^1 \\
&= \Big(D_pH(x, Du^2) - D_pH(x, Du^1)\Big)(m^2 - m^1) \\
&\qquad + \Big(D_pH(x, Du^2) - D_pH(x, Du^1) - D_{pp}H(x, Du^1)\cdot(Du^2 - Du^1)\Big)m^1 \\
&= \Big(D_pH(x, Du^2) - D_pH(x, Du^1)\Big)(m^2 - m^1) \\
&\qquad + m^1 \int_0^1 D(u^2 - u^1)\cdot\Big(D_{pp}H(\lambda Du^2 + (1-\lambda)Du^1) - D_{pp}H(x, Du^2)\Big)d\lambda, \\
R^3(x,T) &= G(x, m^2(T)) - G(x, m^1(T)) - \frac{\delta G}{\delta m}(x, m^1(T))(m^2(T) - m^1(T)), \\
\xi &= m_0^2 - m_0^1.
\end{aligned}$$

Note that we have the estimates

$$
\begin{aligned}
\|\xi\|_{(C^{1+\delta})'} &\leq \sup_{\|Df\|_\infty \leq 1} \int_{\mathbb{R}^d} f(x)(m_0^2 - m_0^1) = \mathbf{d}_1(m_0^2, m_0^1), \\
\sup_{t\in[t_0,T]} \|R^2(t)\|_{(W^{1,\infty})'} &\leq C\Big(\|Du^2 - Du^1\|_{C^{0,2}} \sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) + \|Du^2 - Du^1\|^2_{C^{0,2}}\Big) \\
&\leq C\Big(\|u^2 - u^1\|^2_{C^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t))\Big). \\
\|R^3\|_{C^{2+\delta}} &\leq C \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t)).
\end{aligned}
$$

It remains to estimate the quantity $\|R^1\|_{C^{\delta/2,+\delta}}$. For this, we rewrite R^1 as

$$
R^1(t,x) = A(t,x) + B(t,x),
$$

where

$$
A(t,x) = -\int_0^1 \Big(D_pH(x, \lambda Du^2 + (1-\lambda)Du^1) - D_pH(Du^1)\Big) \cdot (Du^2 - Du^1) d\lambda,
$$

and

$$
B(t,x) = \int_0^1 \Big(\frac{\delta F}{\delta m}(x, \lambda m^2(t) + (1-\lambda)m^1(t)) - \frac{\delta F}{\delta m}(x, m^1(t))\Big)(m^2(t) - m^1(t)) d\lambda\Big).
$$

Bounding A is relatively straightforward, since

$$
\|A\|_{C^{\delta/2,\delta}} \leq C\|D(u^2 - u^1)\|^2_{C^{\delta/2,\delta}} \leq C\|u^2 - u^1\|^2_{C^{\delta/2,2+\delta}}.
$$

The bound B is a bit more involved. Given $\lambda, \theta \in [0,1]$, let

$$
[m]_\lambda(t) = \lambda m^2(t) + (1-\lambda)m^1(t)), \quad [m]_{\lambda,\theta}(t) = \theta[m]_\lambda(t) + (1-\theta)m^2(t), \quad \rho(t) = m^2(t) - m^1(t).
$$

We rewrite B as

$$B(t,x) = \int_0^1 \int_0^1 \lambda \frac{\delta^2 F}{\delta m^2}\Big(x, [m]_{\lambda,\theta}(t)\Big)(\rho(t))(\rho(t))d\lambda d\theta,$$

and look at the functions

$$C_{\lambda,\theta}(t,x) = \frac{\delta^2 F}{\delta m^2}\Big(x, [m]_{\lambda,\theta}(t)\Big)(\rho(t))(\rho(t)).$$

For every $\lambda, \theta \in [0,1]$, we have the estimates

$$\begin{aligned}
\|C_{\lambda,\theta}(\cdot,\cdot)\|_{\delta/2,\delta} \leq C\Big(&\sup_{t\neq s} \frac{\|[m]_{\lambda,\theta}(t) - [m]_{\lambda,\theta}(s)\|_{(C^1)'}}{|t-s|^{\delta/2}} \sup_{t\in[t_0,T]} \|\rho(t)\|^2_{(C^{2+\delta})'} \\
&+ \sup_{t\neq s} \frac{\|\rho(t)-\rho(s)\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} \sup_{t\in[t_0,T]} \|\rho(t)\|_{(C^1)'}\Big) \\
\leq C\Big(&\Big(\sup_{t\neq s} \frac{\mathbf{d}_1(m^2(t), m^2(s))}{|t-s|^{\delta/2}} + \sup_{t\neq s} \frac{\mathbf{d}_1(m^1(t), m^1(s))}{|t-s|^{\delta/2}}\Big) \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t)) \\
&+ \sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \frac{\|(m^2(t)-m^1(t)) - (m^2(s)-m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}}\Big)
\end{aligned}$$

and thus

$$\begin{aligned}
\|B(\cdot,\cdot)\|_{\delta/2,\delta} \leq &C\Big(\sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \frac{\|(m^2(t)-m^1(t)) - (m^2(s)-m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} \\
&+ \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t))\Big).
\end{aligned}$$

Combining the upper bounds on A and B, we finally conclude that

$$\begin{aligned}
\|R^1\|_{C^{\delta/2,\delta}} \leq C\Big(&\sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \frac{\|(m^2(t)-m^1(t)) - (m^2(s)-m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} \\
&+ \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t)) + \|u^2 - u^1\|^2_{C^{\delta/2, 2+\delta}}\Big),
\end{aligned}$$

and, hence, using Lemma 4.6.1 we get

$$\|u^2 - u^1\|_{C^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0,T]} \|m^2(t) - m^1(t)\|_{(C^{2+\delta})'} + \sup_{t\neq s} \frac{\|(m^2(t) - m^1(t)) - (m^2(s) - m^1(s)\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}}$$
$$\leq C\Big(\|u^2 - u^1\|^2_{C^{\delta/2,2+\delta}} + \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t))$$
$$+ \sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \frac{\|(m^2(t) - m^1(t)) - (m^2(s) - m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}}\Big).$$

Thus choosing $\eta > 0$ small enough in (4.6.9), we find

$$\|u^2 - u^1\|_{C^{(2+\delta)/2,2+\delta}} + \frac{\|(m^2(t) - m^1(t)) - (m^2(s) - m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}}$$
$$\leq C\Big(\mathbf{d}_1(m_0^2, m_0^1) + \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t))\Big)$$
$$\leq C\Big(\mathbf{d}_1(m_0^2, m_0^1) + \|u^2 - u^1\|^2_{C^{(2+\delta)/2,2+\delta}} + \mathbf{d}_1^2(m_0^2, m_0^1)\Big),$$

where in the last inequality we used (4.6.10).

Therefore, choosing $\eta > 0$ even smaller if necessary, we obtain

$$\|u^2 - u^1\|_{C^{(2+\delta)/2,k+\delta}} + \frac{\|(m^2(t) - m^1(t)) - (m^2(s) - m^1(s))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} \leq C\mathbf{d}_1(m_0^2, m_0^1).$$

Using the last inequality in (4.6.10) yields

$$\sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \leq C\mathbf{d}_1(m_0^2, m_0^1).$$

□

Next, we give a sharpened version of the main regularity result in [92].

Proposition 4.6.4. *Let Assumption 1 hold. Then, for each $(t_0, m_0) \in \mathcal{O}$, there exist constants $\delta, C > 0$ such that, for t, m_1, m_2 with $|t - t_0| < \delta$, $\mathbf{d}_2(m_0, m_i) < \delta$ and $i = 1, 2$, we have*

$$\sup_{x \in \mathbb{R}^d} |D_m\mathcal{U}(t, m_1, x) - D_m\mathcal{U}(t, m_2, x)| \leq C\mathbf{d}_1(m_1, m_2).$$

Proof. For $i = 1, 2$, let (u^i, m^i) be solutions to MFG with initial conditions $m^i(t) = m_i$. We have that

$$D_m\mathcal{U}(t, m^i, x) = Du^i(t, x)$$

and the result follows from Lemma 4.6.3.

□

4.6.3 The C^2−regularity of $\mathcal{U}$

In this subsection we show that the function $\mathcal{U}$ is twice differentiable in m and, moreover, the map $m \to D_{mm}\mathcal{U}(t, m, x, y)$ is Lipschitz in d_1 locally within $\mathcal{O}$. This is the assertion of Theorem 4.2.4. The proof follows closely the one developed [79].

We introduce next some notation, and also give a roadmap showing how the arguments of [79] will be adapted to the present setting.

For $(t_0, m_0) \in \mathcal{O}$, let $m^{(t_0,m_0)}$ denote the unique optimal trajectory started from (t_0, m_0) and by $u^{(t_0,m_0)}$ its corresponding multiplier, and consider the map $\Phi : \mathcal{O} \times \mathbb{R}^d \to \mathbb{R}$ given by

$$\Phi(t_0, m_0, x) = u^{(t_0,m_0)}(t_0, x).$$

It was shown in the proof of Lemma 2.9 in [92] that, for $(t, m) \in \mathcal{O}$, we have

$$\frac{\delta\mathcal{U}}{\delta m}(t, m, x) = \Phi(t, m, x) - \int_{\mathbb{R}^d} \Phi(t, m, z)m(dz). \tag{4.6.11}$$

Given a multi-index $l \in \{0,1\}^d$ with $|l| = \sum_{i=1}^{d} l_i \leq 1$, that is, either $l = (0, \cdots, 0)$ or $l = e_i$ for some $i \in \{1, \cdots, d\}$, where e_i is the standard basis in $\mathbb{R}^d$, and $y \in \mathbb{R}^d$, let $(w^{(l),y}, \rho^{(l),y})$ be the solution to (MFGL) driven by (u, m), with initial condition $\rho^{(l),y}(t_0, \cdot) = (-1)^{|l|} D^{(l)} \delta_y$.

We also define the function $K^{(l)} : \mathcal{O} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ given by

$$K^{(l)}(t_0, m_0, x, y) = w^{(l),y}(t_0, x),$$

and, for simplicity of notation, we write $K = K^{(0)}$.

Following the arguments in [79], we show below that

$$\frac{\delta \Phi}{\delta m}(t, m, x, y) = \frac{\delta}{\delta m}[\Phi(t, \cdot, x)](y) = K(t_0, m_0, x, y). \tag{4.6.12}$$

We note that the normalization convention

$$\int_{\mathbb{T}^d} K(t_0, m_0, x, y) m_0(dy) = 0$$

is satisfied, since $(z, \rho) = (0, m^{(t_0, m_0)})$ is the unique solution to (MFGL).

Combining (4.6.12) and (4.6.11) and keeping in mind the normalization convention for linear derivatives, it follows that $\frac{\delta^2 \mathcal{U}}{\delta m^2} = \mathcal{K}$, where $\mathcal{K}$ is the “normalized" version of K given by

$$\begin{aligned}\mathcal{K}(t_0, m_0, x, y) = K(t_0, m_0, x, y) - \int_{\mathbb{T}^d} K(t_0, m_0, z, y) m_0(dz) - u^{(t_0, m_0)}(t_0, y) \\ + \int_{\mathbb{T}^d} u^{(t_0, m_0)}(t_0, z) m_0(dz).\end{aligned} \tag{4.6.13}$$

The existence and regularity of $D_{mm}\mathcal{U}$ thus follows from the regularity properties of the map K, which we investigate next.

Proposition 4.6.5. *Let Assumption 2 hold and fix $(t_0, m_0) \in \mathcal{O}$. Then, the function $K^{(0)}(t_0, m_0, x, y)$*

is differentiable in y. Moreover, for any $l \in \{0,1\}^d$ *with* $|l| \le 1$*, the derivative* $x \to D_y^{(l)} K^{(0)}(t_0, m_0, x, y)$ *belongs to* $C^{2+\delta}(\mathbb{R}^d)$ *and is given by*

$$D_y^{(l)} K^{(0)}(t_0, m_0, x, y) = K^{(l)}(t_0, m_0, x, y). \tag{4.6.14}$$

Furthermore, there exists a constant $C > 0$*, which depends on the data, such that*

$$\sup_{y \in \mathbb{R}^d} \|D_y^{(l)} K^{(0)}(t_0, m_0, \cdot, y)\|_{C^{2+\delta}} \le C, \tag{4.6.15}$$

and

$$\|D_y^{(l)} K^{(0)}(t_0, m_0, \cdot, y') - D_y^{(l)} K^{(0)}(t_0, m_0, \cdot, y)\|_{2+\delta} \le C|y' - y|^{\delta}. \tag{4.6.16}$$

Finally, given a finite signed measure ξ *on* $\mathbb{R}^d$*, the unique solution* (z, ρ) *to* (MFGL) *driven by* (u, m) *and with initial condition* $\rho(t_0) = \xi$ *satisfies*

$$z(t_0, x) = \langle K^{(0)}(t_0, m_0, x, \cdot), \xi \rangle. \tag{4.6.17}$$

Proof. Fix $y \in \mathbb{R}^d$. Since $|l| \le 1$, Lemma 4.6.1 yields

$$\begin{aligned} &\|w^{(l),y}\|_{C^{(2+\delta)/2, 2+\delta}} + \sup_{t \in [t_0, T]} \|\rho^{(l),y}(t)\|_{(C^{2+\delta})'} + \sup_{t' \neq t} \frac{\|\rho^{(l),y}(t) - \rho^{(l),y}(s)\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} \\ &\quad \le C\|D^{(l)}\delta_y\|_{(C^{1+\delta})'} \le C. \end{aligned}$$

Note that in the estimate above we used that

$$\|D^{(l)}\delta_y\|_{(C^{1+\delta})'} = \sup_{\|f\|_{C^{1+\delta}} \le 1} D_y^{(l)} f(y) \le \sup_{\|f\|_{C^{1+\delta}} \le 1} \|f\|_{C^{1+\delta}} \le 1.$$

Let $e_1, \ldots, e_d$ be the standard basis vectors in $\mathbb{R}^d$. Then, for $\epsilon > 0$ and $i \in \{1, \cdots, d\}$, the pair

$$(z_i^\epsilon, \lambda_i^\epsilon) = \left(\frac{1}{\epsilon}\Big(w^{(0),y+\epsilon e_i} - w^{(0),y}\Big) - w^{(e_i),y}, \frac{1}{\epsilon}\Big(\rho^{(0),y+\epsilon e_i} - \rho^{(0),y}\Big) - \rho^{(e_i),y}\right)$$

is the unique solution to (MFGL) driven by (u, m) with initial condition

$$\lambda_i^\epsilon = \frac{1}{\epsilon}\Big(\delta_{y+\epsilon e_i} - \delta_y\Big) - (-1)D^{(e_i)}\delta_y.$$

Since

$$\begin{aligned}
&\|\frac{1}{\epsilon}\Big(\delta_{y+\epsilon e_i} - \delta_y\Big) - (-1)D^{(e_i)}\delta_y\|_{(C^{1+\delta})'} \\
&\quad = \sup_{\|f\|_{C^{1+\delta}}\le 1} \frac{f(y+\epsilon e_i) - f(y) - \epsilon D^{(e_i)} f(y)}{\epsilon} \\
&\quad = \sup_{\|f\|_{C^{1+\delta}}\le 1} \int_0^1 D^{(e_i)}\Big(f(s(y+\epsilon e_i) + (1-s)y) - f(y)\Big) ds \\
&\quad \le C \sup_{\|f\|_{C^{1+\delta}}\le 1} \|D^{e_i} f\|_{C^\delta} \int_0^1 s^\delta \epsilon^\delta ds \le C \sup_{\|f\|_{C^{1+\delta}}\le 1} \|f\|_{C^{1+\delta}} \epsilon^\delta \le C\epsilon^\delta,
\end{aligned}$$

Lemma 4.6.1 yields

$$\begin{aligned}
&\|\frac{1}{\epsilon}\Big(w^{(0),y+\epsilon e_i} - w^{(0),y}\Big) - w^{(e_i),y}\|_{C^{(2+\delta)/2,2+\delta}} \\
&\qquad \le C\|\frac{1}{\epsilon}\Big(\delta_{y+\epsilon e_i} - \delta_y\Big) - (-1)D^{(e_i)}\delta_y\|_{(C^{1+\delta})'} \le C\epsilon^\delta,
\end{aligned}$$

and, hence, (4.6.15) and (4.6.14) follow.

Furthermore, given y', y, (4.6.16) is an application of Lemma 4.6.1 to the pair $(w^{(l),y'} - w^{(l),y}, \rho^{(l),y'} - \rho^{(l),y})$.

Finally, (4.6.17) follows from the fact that the pair

$$(z^{\xi}, \mu^{\xi}) = (\int_{\mathbb{T}^d} w^{(0),y}(t,x)d\xi(y), \int_{\mathbb{T}^d} \rho^{(0),y} d\xi(y))$$

is the unique solution to (MFGL) driven by (u, m) with initial condition $\rho^{\xi}(t_0) = \xi$.

□

We now show the Lipschitz continuity of $K^{(l)}$ with respect to m.

Proposition 4.6.6. *Let Assumption 2 hold. Given* $(t_0, m_0^1) \in \mathcal{O}$*, there exist* $\eta > 0$ *and* $C > 0$ *such that, if* $m_0^2 \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $\mathbf{d}_2(m_0^2, m_0^1) \le \eta$*, then*

$$\|K^{(l)}(t_0, m_0^2, \cdot, y) - K^{(l)}(t_0, m_0^1, \cdot, y)\|_{2+\delta} \le C\mathbf{d}_1(m_0^2, m_0^1).$$

Proof. Let (u^1, m^1) and (u^2, m^2) be the unique stable solutions to (MFG) with initial conditions $m^1(t_0) = m_0^1$ and $m^2(t_0) = m_0^2$ respectively. In addition, let $(z^1, \rho^1), (z^2, \rho^2)$ be solutions to (MFGL) driven by $(u^1, m^1), (u^2, m^2)$ respectively and with initial conditions $\rho^1(t_0) = \rho^2(t_0) = (-1)^{|l|} D^{(l)} \delta_y$. The pair

$$(w, \lambda) = (z^2 - z^1, \rho^2 - \rho^1)$$

solves (MFGLE) driven by (u^1, m^1) with

$$\begin{aligned}
R^1 &= \Big(D_p H(x, Du^1) - D_p H(x, Du^2)\Big) \cdot Du^2 + \Big(\frac{\delta F}{\delta m}(x, m^2(t))(\rho^2(t)) - \frac{\delta F}{\delta m}(x, m^1(t))(\rho^2(t))\Big),\\
R^2 &= \rho^2(t)\Big(D_p H(x, Du^2) - D_p H(x, Du^1)\Big) + \Big(m^2 D_{pp} H(x, Du^2) - m^1 D_{pp} H(x, Du^1)\Big) \cdot Du^2,\\
R^3 &= \frac{\delta G}{\delta m}(x, m^2(T))(m^2(T)) - \frac{\delta G}{\delta m}(x, m^1(T))(m^2(T)),\\
\xi &= 0.
\end{aligned}$$

From Lemma 4.6.3 we have

$$\|R^3\|_{C^{2+\delta}} \le C \sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t)) \le C\mathbf{d}_1(m_0^2, m_0^1),$$

and

$$\sup_{t\in[t_0,T]} \|R^2(t)\|_{(W^{1,\infty})'} \le C\Big(\sup_{t\in[t_0,T]} \|Du^2(t) - Du^1(t)\|_{C^1} + \sup_{t\in[t_0,T]} \mathbf{d}_1(m^2(t), m^1(t))\Big)$$
$$\le C\mathbf{d}_1(m_0^2, m_0^1).$$

It remains to estimate $\|R^1\|_{C^{\delta/2,\delta}}$. To this end, we rewrite it as

$$R^1 = A + B,$$

with

$$A = \Big(D_pH(x, Du^1) - D_pH(x, Du^2)\Big) \cdot Du^2,$$

and

$$B = \frac{\delta F}{\delta m}(x, m^2(t))(\rho^2(t)) - \frac{\delta F}{\delta m}(x, m^1(t))(\rho^2(t)).$$

It follows from Lemma 4.6.3 that

$$\|A\|_{C^{\delta/2,\delta}} \le C\|D(u^2 - u^1)\|_{C^{\delta/2,\delta}} \le C\|u^2 - u^1\|_{C^{(2+\delta),2+\delta}} \le C\mathbf{d}_1(m_0^2, m_0^1).$$

Finally, we write

$$B(t,x) = \int_0^1 \frac{\delta^2 F}{\delta m^2}(x, \lambda m^2(t) + (1-\lambda)m^1(t))(\rho^2(t))(m^2(t) - m^1(t))d\lambda.$$

An argument similar to the one in the proof of Lemma 4.6.3 yields

$$\|B(\cdot,\cdot)\|_{C^{\delta/2,\delta}} \leq C\Big(\sup_{t\neq s}\frac{\|(m^2(t)-m^1(t)-(m^2(s)-m^1(s)))\|_{(C^{2+\delta})'}}{|t-s|^{\delta/2}} + \sup_{t\in[t_0,T]}\mathbf{d}_1(m^2(t),m^1(t))\Big),$$

and, hence, by Lemma 4.6.3, we have

$$\|R^1\|_{C^{\delta/2,\delta}} \leq C\mathbf{d}_1(m_0^2,m_0^1).$$

Finally, Lemma 4.6.1 and the definition of $K^{(l)}$ imply that

$$\|K^{(l)}(t_0,m_0^2,\cdot,y)-K^{(l)}(t_0,m_0,\cdot,y)\|_{2+\delta} \leq \|z^2-z\|_{C^{(2+\delta)/2,2+\delta}} \leq C\mathbf{d}_1(m_0^2,m_0^1).$$

□

Theorem 4.6.7. *Let Assumption 2 hold. The map $\mathcal{U}$ is C^2 in the set $\mathcal{O}$ and satisfies*

$$\frac{\delta^2\mathcal{U}}{\delta^2 m}(t_0,m_0,x,y) = \mathcal{K}(t_0,m_0,x,y).$$

Moreover, given $(t_0,m_0^1)\in\mathcal{O}$, there exist an $\eta>0$ and $C>0$ such that, if $m_0^2\in\mathcal{P}_2(\mathbb{R}^d)$ with $d_2(m_0^2,m_0^1)\leq\eta$, then

$$\Big\|\frac{\delta\mathcal{U}}{\delta m}(t_0,m_0^2,\cdot)-\frac{\delta\mathcal{U}}{\delta m}(t_0,m_0^1,\cdot)-\int_{\mathbb{T}^d}\mathcal{K}(t_0,m_0,\cdot,y)d(m_0^2-m_0^1)(y)\Big\|_{2+\delta} \leq C\mathbf{d}_1^2(m_0^2,m_0^1).$$

Proof. Let (u^1,m^1) and (u^2,m^2) be the unique stable solutions to (MFG) with initial conditions $m^1(t_0)=m_0^1, m^2(t_0)=m_0^2$ respectively, and (z,μ) be the solution to (MFGL) driven by (u,m) with initial condition $\rho(t_0)=m_0^2-m_0^1$.

The pair

$$(z,\rho) = (u^2-u^1-z, m^2-m^1-\mu)$$

solves the linearized system (MFGLE) driven by (u^1, m^1), with $\xi = 0$ and

$$\begin{aligned}
R^1 &= -(H(x, Du^2) - H(x, Du^1) - D_pH(x, Du^1) \cdot D(u^2 - u^1)) \\
&\quad + F(x, m^2) - F(x, m^1) - \frac{\delta F}{\delta m}(x, m^1(t))(m^2(t) - m^1(t)), \\
R^2 &= (D_pH(x, Du^2) - D_pH(x, Du^1) - D_{pp}H(x, Du^1) \cdot (Du^2 - Du^1))m^1 \\
&\quad + (D_pH(x, Du^2) - D_pH(x, Du^1))(m^2 - m^1), \\
R^3 &= G(x, m^2(T)) - G(x, m^1(T)) - \frac{\delta G}{\delta m}(x, m^1(T))(m^2(T) - m^1(T)).
\end{aligned}$$

Arguments similar to those used to bound R^1 in the proof of 4.6.3 yield

$$\begin{aligned}
&\|R^1\|_{C^{\delta/2,\delta}} + \sup_{t\in[t_0,T]} \|R^2\|_{(W^{1,\infty})'} + \|R^3\|_{(C^{2+\delta})'} \\
&\quad \le C\Big(\|Du^2 - Du^1\|^2_{C^{\delta/2,\delta}} + \sup_{t\in[t_0,T]} \mathbf{d}_1^2(m^2(t), m^1(t)) + \Big(\sup_{t\neq s} \frac{\|(m^2(t) - m^1(t)) - (m^2(s) - m^1(s))\|}{|t-s|^{\delta/2}}\Big)^2\Big),
\end{aligned}$$

and thus Lemma 4.6.1 and Lemma 4.6.3 imply that

$$\|u^2 - u^1 - z\|_{C^{(2+\delta)/2,2+\delta}} \le C\mathbf{d}_1^2(m_0^2, m_0^1).$$

The above shows that

$$\frac{\delta \Phi}{\delta m}(t_0, m_0^1, x, y) = K(t_0, m_0, x, y).$$

The result now follows from the representations (4.6.13), (4.6.11) and (4.6.17).

□

We may now show Theorem 4.2.4.

The proof of Theorem 4.2.4. The claim follows from Proposition 4.6.6 and Theorem 4.6.7.

□

We conclude this section by showing that the function $D_m\mathcal{U}$ is Lipschitz continuous with respect to time.

Proposition 4.6.8. *Let Assumption 2 hold. Given $(t_0, m_0) \in \mathcal{O}$, there exist $C > 0$ and $\delta > 0$ depending on the data such that, for all $|h| < \delta$,*

$$\left\| D_m\mathcal{U}((t_0+h)\wedge T, m_0, \cdot) - D_m\mathcal{U}(t_0, m_0, \cdot) \right\|_{L^\infty} \leq C|h|.$$

Proof. Fix $(t_0, m_0) \in \mathcal{O}$ and $x \in \mathbb{R}^d$.

We write

$$|D_m\mathcal{U}(t_0+h, m_0 x) - D_m\mathcal{U}(t_0, m_0, x)| \leq I + II,$$

with

$$I = |D_m\mathcal{U}(t_0+h, m_{t_0}, x) - D_m\mathcal{U}(t_0+h, m_{t_0+h}, x)|,$$

and

$$II = |D_m\mathcal{U}(t_0+h, m_{t_0+h}, x) - D_m\mathcal{U}(t_0+h, m_{t_0}, x)|,$$

where $t \mapsto m_t$ is the unique optimal trajectory started from (t_0, m_0), and proceed obtaining bounds for I and II.

To estimate II, we note that the regularity of $u^{(t_0,m_0)}$ yields

$$II = |D_x u^{(t_0,m_0)}(t_0+h, x) - D_x u^{(t_0,m_0)}(t_0, x)| \leq Ch.$$

For the term I, we first note that, in view of Theorem 4.6.7 and the regularity of K proved in

Proposition 4.6.5, we can find δ small enough that, if (t, m) is such that $|t - t_0| < \delta$ and $\mathbf{d}_2(m, m_0) < \delta$, then, for some $C > 0$,

$$\|D_{mm}\mathcal{U}(t, m, \cdot, y)\|_{C^{1+\delta}} \leq C.$$

Since $D_{mm}\mathcal{U}(t, m, x, y) = D_{mm}\mathcal{U}(t, m, y, x)^T$ by Corollary 5.89 in [96], it follows that

$$\|D_{mm}\mathcal{U}(t, m, x, \cdot)\|_{C^{1+\delta}} \leq C.$$

From here it is straightforward to check that there is a constant C such that, if $|t - t_0| < \delta$ and $\mathbf{d}_2(m, m_0) < \delta$, we have

$$\|\frac{\delta}{\delta m}[D_m\mathcal{U}(t, m, x)](\cdot)\|_{C^2} \leq C.$$

It follows that $D_m\mathcal{U}$ is locally Lipschitz in $(C^2)'$, and, in particular, for $|t - t_0| < \delta$, $\mathbf{d}_2(m, m_0) < \delta$, we find

$$|D_m\mathcal{U}(t, m, x) - D_m\mathcal{U}(t, m_0, x)| \leq C\|m - m_0\|_{(C^2)'}.$$

Note also that standard estimates yield that $t \mapsto m_t$ is Lipschitz with respect to the $(C^2)'$-metric. Thus, choosing, if necessary, δ even smaller so that $\mathbf{d}_2(m_{t_0+h}, m_{t_0})$ is small enough, we find that

$$I \leq C\|m_{t_0+h} - m_{t_0}\|_{(C^2)'} \leq Ch,$$

which completes the proof. □

REFERENCES

[1] Y. Achdou and I. Capuzzo-Dolcetta. Mean field games: Numerical methods. *SIAM J. Numer. Anal.*, 48(3):1136–1162, 2010.

[2] Y. Achdou, P. Mannucci, C. Marchi, and N. Tchou. Deterministic mean field games with control on the acceleration. *Nonlinear Differential Equations and Applications NoDEA*, 27(3):1–32, 2020.

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

[4] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.

[5] A. A. Albanese and E. M. Mangino. Analyticity of a class of degenerate evolution equations on the canonical simplex of $\mathbb{R}^d$ arising from Fleming-Viot processes. *J. Math. Anal. Appl.*, 379(1):401–424, 2011.

[6] C. Alekos, D. P. Paolo, F. Markus, and P. Guglielmo. On the convergence problem in mean field games: a two state model without uniqueness. *SIAM J. Control Optim.*, 57(4):2443–2466, 2019.

[7] S. A. Alimov, R. R. Ashurov, and A. K. Pulatov. Multiple Fourier series and Fourier integrals [MR1027847 (91b:42022)]. In *Commutative harmonic analysis, IV*, volume 42 of *Encyclopaedia Math. Sci.*, pages 1–95. Springer, Berlin, 1992.

[8] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis: A hitchhiker's guide*. Springer, Berlin, third edition, 2006.

[9] D. M. Ambrose and A. R. Mészáros. Well-posedness of mean field games master equations involving non-separable local hamiltonians. *arXiv*, https://arxiv.org/abs/2105.03926, 2021.

[10] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

[11] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[12] N. Antunes, C. Fricker, P. Robert, and D. Tibi. Stochastic networks with multiple stable points. *The Annals of Probability*, 36(1):255–278, 2008.

[13] B. Ata, J. M. Harrison, and L. A. Shepp. Drift rate control of a Brownian processing system. *Ann. Appl. Probab.*, 15(2):1145–1160, 2005.

[14] R. Atar. A diffusion regime with nondegenerate slowdown. *Oper. Res.*, 60(2):490–500, 2012.

[15] R. J. Aumann. Markets with a continuum of traders. *Econometrica*, 32:39–50, 1964.

[16] F. Baccelli, F. Karpelevich, M. Y. Kelbert, A. Puhalskii, A. Rybko, and Y. M. Suhov. A mean-field limit for a class of queueing networks. *Journal of Statistical Physics*, 66(3-4):803–825, 1992.

[17] K. Bahlali, M. Mezerdi, and B. Mezerdi. Existence and optimality necessary conditions for general stochastic mean# field control problems! 2014.

[18] T. Bakaryan, R. Ferreira, and D. Gomes. Some estimates for the planning problem with potential. *NoDEA Nonlinear Differential Equations Appl.*, 28(2):1–23, 2021.

[19] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham. Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem. *Stochastic Process. Appl.*, 129(2):674–711, 2019.

[20] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1997. With appendices by Maurizio Falcone and Pierpaolo Soravia.

[21] M. Bardi and P. Cardaliaguet. Convergence of some mean field games systems to aggregation and flocking models. *Nonlinear Analysis*, 204:112199, 2021.

[22] G. Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*. Springer Berlin, Heidelberg, 1994.

[23] E. Bayraktar, A. Budhiraja, and A. Cohen. A numerical scheme for a mean field game in some queueing systems based on Markov chain approximation method. *ArXiv e-prints*, 2017.

[24] E. Bayraktar, A. Budhiraja, and A. Cohen. Rate control under heavy traffic with strategic servers. *Ann. Appl. Probab.*, 2017. to appear.

[25] E. Bayraktar, A. Cecchin, and P. Chakraborty. Mean field control and finite agent approximation for regime-switching jump diffusions. *Appl. Math. Optim.*, 88(2), May 2023.

[26] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue. Finite state MFGs with Wright–Fisher common noise. *arXiv e-prints*, 2019.

[27] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue. Finite state mean field games with Wright Fisher common noise as limits of *N*-player weighted games. *arXiv e-prints*, page arXiv:2012.04845, Dec. 2020.

[28] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue. Finite state mean field games with Wright-Fisher common noise as limits of *N*-player weighted games. *arXiv e-prints*, arXiv:2012.04845, Dec 2020.

[29] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue. Finite state mean field games with Wright-Fisher common noise. *J. Math. Pures Appl.*, 147:98–162, 2021.

[30] E. Bayraktar and A. Cohen. Analysis of a finite state many player game using its master equation. *SIAM Journal on Control and Optimization*, 56(5):3538–3568, 2018.

[31] E. Bayraktar, A. Cosso, and H. Pham. Randomized dynamic programming principle and feynman-kac representation for optimal control of mckean-vlasov dynamics. *Transactions of the American Mathematical Society*, 370(3):2115–2160, 2018.

[32] E. Bayraktar, I. Ekren, and X. Zhang. Comparison of viscosity solutions for a class of second order PDEs on the wasserstein space. Preprint,arXiv:2309.05040 [math.AP].

[33] E. Bayraktar, I. Ekren, and X. Zhang. A smooth variational principle on Wasserstein space. *Proceedings of the American Mathematical Society*, 151:4089–4098, 2023.

[34] E. Bayraktar, U. Horst, and R. Sircar. A limit theorem for financial markets with inert investors. *Math. Oper. Res.*, 31(4):789–810, 2006.

[35] E. Bayraktar, U. Horst, and R. Sircar. Queuing theoretic approaches to financial price fluctuations. *Handbooks in Operations Research and Management Science*, 15:637–677, 2007.

[36] E. Bayraktar and M. Ludkovski. Optimal trade execution in illiquid markets. *Math. Finance*, 21(4):681–701, 2011.

[37] E. Bayraktar and M. Ludkovski. Liquidation in limit order books with controlled intensity. *Math. Finance*, 24(4):627–650, 2014.

[38] E. Bayraktar and X. Zhang. On non-uniqueness in mean field games. *arXiv e-prints*, arXiv:1908.06207, Aug 2019.

[39] E. Bayraktar and X. Zhang. On non-uniqueness in mean field games. *Proc. Amer. Math. Soc.*, 148(9):4091–4106, 2020.

[40] E. Bayraktar and X. Zhang. Corrigendum to "On non-uniqueness in mean field games". *Proc. Amer. Math. Soc.*, 149(3):1359–1360, 2021.

[41] E. Bayraktar and Y. Zhang. A rank-based mean field game in the strong formulation. *Electron. Commun. Probab.*, 21:Paper No. 72, 12, 2016.

[42] S. L. Bell and R. J. Williams. Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: asymptotic optimality of a threshold policy. *Ann. Appl. Probab.*, 11(3):608–649, 2001.

[43] M. Benaim and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11):823–838, 2008.

[44] A. Bensoussan, J. Frehse, and P. Yam. *Mean field games and mean field type control theory*. SpringerBriefs in Mathematics. Springer, New York, 2013.

[45] A. Bensoussan, J. Frehse, and S. C. P. Yam. The master equation in mean field theory. *J. Math. Pures Appl.*, 103(6):1441–1474, 2015.

[46] A. Bensoussan, J. Frehse, and S. C. P. Yam. The master equation in mean field theory. *J. Math. Pures Appl. (9)*, 103(6):1441–1474, 2015.

[47] J. Bergin and D. Bernhardt. Anonymous sequential games with aggregate uncertainty. *J. Math. Econom.*, 21(6):543–562, 1992.

[48] J. Bergin and D. Bernhardt. Anonymous sequential games: existence and characterization of equilibria. *Econom. Theory*, 5(3):461–489, 1995.

[49] C. Bertucci. Monotone solutions for mean field games master equations : continuous state space and common noise. *arXiv*, https://arxiv.org/abs/2107.09531, 2021.

[50] C. Bertucci. Monotone solutions for mean field games master equations: finite state space and optimal stopping. *J. Éc. polytech. Math.*, 8:1099–1132, 2021.

[51] C. Bertucci, J.-M. Lasry, and P.-L. Lions. Some remarks on mean field games. *Communications in Partial Differential Equations*, 44(3):205–227, 2019.

[52] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[53] J. Blair, P. Johnson, and P. Duck. Analysis of optimal liquidation in limit order books. http://eprints.ma.man.ac.uk/2299/01/covered/MIMS_ep2014_23.pdf, 2015. Preprint.

[54] A. Bobbio, M. Gribaudo, and M. Telek. Analysis of large scale interacting systems by mean field method. In *Quantitative Evaluation of Systems, 2008. QEST'08. Fifth International Conference on*, pages 215–224. IEEE, 2008.

[55] L. Boccardo, A. Dall'Aglio, T. Gallouët, and L. Orsina. Nonlinear parabolic equations with measure data. *journal of functional analysis*, 147(1):237–258, 1997.

[56] V. I. Bogachev. *Differentiable measures and the Malliavin calculus*, volume 164 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2010.

[57] V. S. Borkar. Optimal control of diffusion processes. volume 203 of *Pitman Research Notes in Mathematics Series*, pages vi+196. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1989.

[58] K. A. Borovkov. Propagation of chaos for queueing networks. *Theory Probab. Appl.*, 42(3):385–394, 1998.

[59] F. Bouchut. Hypoelliptic regularity in kinetic equations. *Journal de mathématiques pures et appliquées*, 81(11):1135–1159, 2002.

[60] M. Bowe, S. Hyde, and I. Johnson. Determining the intensity of buy and sell limit order submissions: A look at the market pre-opening period. Preprint.

[61] H. Brezis and H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.

[62] H. Brezis and P. Mironescu. Gagliardo–nirenberg inequalities and non-inequalities: the full story. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 35, pages 1355–1376. Elsevier, 2018.

[63] A. Briani and P. Cardaliaguet. Stable solutions in potential mean field game systems. *NoDEA Nonlinear Differential Equations Appl.*, 25(1):Paper No. 1, 26, 2018.

[64] G. Brunick and S. Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *Annals of Applied Probability*, 23(4):1584–1628, 2013.

[65] R. Buckdahn, J. Li, S. Peng, and C. Rainer. Mean-field stochastic differential equations and associated PDEs. *The Annals of Probability*, 45(2):824 – 878, 2017.

[66] A. Budhiraja and E. Friedlander. Diffusion approximations for controlled weakly interacting large finite state systems with simultaneous jumps. *preprint, arXiv:1603.09001*, 2016.

[67] A. Budhiraja, A. P. Ghosh, and C. Lee. Ergodic rate control problem for single class queueing networks. *SIAM J. Control Optim.*, 49(4):1570–1606, 2011.

[68] M. Burger, A. Lorz, and M.-T. Wolfram. Balanced growth path solutions of a boltzmann mean field game model for knowledge growth. *arXiv preprint arXiv:1602.01423*, 2016.

[69] M. Burzoni, V. Ignazio, A. M. Reppen, and H. M. Soner. Viscosity solutions for controlled McKean-Vlasov jump-diffusions. *SIAM J. Control Optim.*, 58(3):1676–1699, 2020.

[70] J. Calder. Lecture notes on viscosity solutions. University of Minnesota., 2018.

[71] F. Camilli. A quadratic mean field games model for the langevin equation. *Axioms*, 10(2):68, 2021.

[72] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, volume 58 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 2004.

[73] P. Cardaliaguet. Notes on mean field games. Technical report, Technical report, 2010.

[74] P. Cardaliaguet. Weak solutions for first order mean field games with local coupling. In *Analysis and geometry in control theory and its applications*, pages 111–158. Springer, 2015.

[75] P. Cardaliaguet. Weak solutions for first order mean field games with local coupling. *Analysis and Geometry in Control Theory and its Applications. Springer INdAM Series*, 11:111–158, 2015.

[76] P. Cardaliaguet, M. Cirant, and A. Porretta. Splitting methods and short time existence for the master equations in mean field games. *arXiv*, https://arxiv.org/abs/2001.10406, 2020.

[77] P. Cardaliaguet, S. Daudin, J. Jackson, and P. Souganidis. An algebraic convergence rate for the optimal control of mckean-vlasov dynamics. *arXiv preprint arXiv:2203.14554*, 2022.

[78] P. Cardaliaguet, S. Daudin, J. Jackson, and P. E. Souganidis. An algebraic convergence rate for the optimal control of mckean–vlasov dynamics. *SIAM Journal on Control and Optimization*, 61(6):3341–3369, 2023.

[79] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The master equation and the convergence problem in mean field games*, volume 201 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2019.

[80] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The master equation and the convergence problem in mean field games:(ams-201)*. Princeton University Press, 2019.

[81] P. Cardaliaguet and J. Graber. Mean field games systems of first order. *ESAIM: Control, Optimisation and Calculus of Variations*, 21:690–722, 2015.

[82] P. Cardaliaguet, J. Graber, A. Porretta, and D. Tonon. Second order mean field games with degenerate diffusion and local coupling. *NoDEA*, 22:1287–1317, 2015.

[83] P. Cardaliaguet, P. J. Graber, A. Porretta, and D. Tonon. Second order mean field games with degenerate diffusion and local coupling. *Nonlinear Differential Equations and Applications NoDEA*, 22(5):1287–1317, 2015.

[84] P. Cardaliaguet, J. Jackson, N. Mimikos-Stamatopoulos, and P. E. Souganidis. Sharp convergence rates for mean field control in the region of strong regularity. *arXiv preprint arXiv:2312.11373*, 2023.

[85] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, and A. Porretta. Long time average of mean field games. *Networks & Heterogeneous Media*, 7(2):279, 2012.

[86] P. Cardaliaguet and A. Porretta. Long time behavior of the master equation in mean field game theory. *Analysis & PDE*, 12(6):1397–1453, 2019.

[87] P. Cardaliaguet and A. Porretta. An introduction to mean field game theory. In *Mean Field Games*, pages 1–158. Springer, 2020.

[88] P. Cardaliaguet and A. Porretta. An introduction to mean field game theory. In *Mean Field Games, chapter 1, Cetraro, Italy 2019, Cardaliaguet, Pierre, Porretta, Alessio (Eds.)*, LNM 2281, pages 203–248. Springer, 2021.

[89] P. Cardaliaguet and M. Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(1):1–16, 2008.

[90] P. Cardaliaguet and P. Souganidis. On first order mean field game systems with a common noise. *To appear on Annals of applied probability*, 2020.

[91] P. Cardaliaguet and P. Souganidis. Weak solutions of the master equation for mean field games with no idiosyncratic noise. *arXiv*, https://arxiv.org/abs/2109.14911, 2021.

[92] P. Cardaliaguet and P. Souganidis. Regularity of the value function and quantitative propagation of chaos for mean field control problems. *arXiv*, 2204.01314, 2022.

[93] R. Carmona. *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2016.

[94] R. Carmona and F. Delarue. Probabilistic analysis of mean-field games. *SIAM J. Control Optim.*, 51(4):2705–2734, 2013.

[95] R. Carmona and F. Delarue. The master equation for large population equilibriums. In *Stochastic analysis and applications 2014*, volume 100 of *Springer Proc. Math. Stat.*, pages 77–128. Springer, Cham, 2014.

[96] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I : Mean Field FBSDEs, Control, and Games*. Springer, 2018.

[97] R. Carmona and F. Delarue. Probabilistic theory of mean field games with applications. i, volume 83 of probability theory and stochastic modelling, 2018.

[98] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications. II*, volume 84 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field games with common noise and master equations.

[99] R. Carmona, F. Delarue, and A. Lachapelle. Control of McKean-Vlasov dynamics versus mean field games. *Math. Financ. Econ.*, 7(2):131–166, 2013.

[100] R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. *Annals of Probability*, 44:3740–3803, 2016.

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

[102] R. Carmona and P. Wang. Finite state mean field games with major and minor players. *ArXiv e-prints*, Oct. 2016.

[103] R. Carmona and P. Wang. A Probabilistic Approach to Extended Finite State Mean Field Games. *arXiv e-prints*, arXiv:1808.07635, Aug 2018.

[104] R. Carmona and P. Wang. Finite-State Contract Theory with a Principal and a Field of Agents. *arXiv e-prints*, page arXiv:1808.07942, Aug 2018.

[105] G. Cavagnari, S. Lisini, C. Orrieri, and G. Savaré. Lagrangian, Eulerian and Kantorovich formulations of multi-agent optimal control problems: equivalence and gamma-convergence. *J. Differential Equations*, 322:268–364, 2022.

[106] G. Cavagnari, S. Lisini, C. Orrieri, and G. Savaré. Lagrangian, Eulerian and Kantorovich formulations of multi-agent optimal control problems: Equivalence and gamma-convergence. *Journal of Differential Equations*, 322:268–364, 2022.

[107] A. Cecchin. Finite state N-agent and mean field control problems. *ESAIM Control Optim. Calc. Var.*, 27:Paper No. 31, 33, 2021.

[108] A. Cecchin and F. Delarue. Selection by vanishing common noise for potential finite state mean field games. *Comm. Partial Differential Equations*, 47(1):89–168, 2022.

[109] A. Cecchin and F. Delarue. Weak solutions to the master equation of potential mean field games. *arXiv*, 2204.04315, 2022.

[110] A. Cecchin and F. Delarue. Weak solutions to the master equation of potential mean field games, 2022.

[111] A. Cecchin and F. Delarue. Selection by vanishing common noise for potential finite state mean field games. In preparation.

[112] A. Cecchin and M. Fischer. Probabilistic approach to finite state mean field games. *Appl. Math. Optim.*, 81(2):253–300, 2020.

[113] A. Cecchin and G. Pelino. Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stochastic Processes and their Applications*, 129(11):4510 – 4555, 2019.

[114] A. Cecchin and G. Pelino. Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stochastic Processes and their Applications*, 129(11):4510–4555, 2019.

[115] J.-F. Chassagneux, D. Crisan, and F. Delarue. Numerical method for FBSDEs of McKean-Vlasov type. *Ann. Appl. Probab.*, 29(3):1640–1684, 2019.

[116] J.-F. Chassagneux, D. Crisan, and F. Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria. *Memoirs of the AMS*, To appear.

[117] P.-E. Chaudru de Raynal and N. Frikha. From the backward kolmogorov PDE on the wasserstein space to propagation of chaos for mckean-vlasov sdes. *Journal de Mathématiques Pures et Appliquées*, 156:1–124, 2021.

[118] P.-E. Chaudru de Raynal and N. Frikha. Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space. *Journal de Mathématiques Pures et Appliquées*, 159:1–167, 2022.

[119] H. Chen and D. D. Yao. *Fundamentals of queueing networks*, volume 46 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001. Performance, asymptotics, and optimization, Stochastic Modelling and Applied Probability.

[120] L. Chen and D. W. Stroock. The fundamental solution to the Wright-Fisher equation. *SIAM J. Math. Anal.*, 42(2):539–567, 2010.

[121] M. F. Chen and S. F. Li. Coupling methods for multidimensional diffusion processes. *Ann. Probab.*, 17(1):151–177, 1989.

[122] M. Cirant and A. Porretta. Long time behaviour and turnpike solutions in mildly non-monotone mean field games. *ESAIM: Control Optim. Calc. Var.*, 27, 2021.

[123] G. Conforti, R. Kraaij, and D. Tonon. Hamilton–jacobi equations for controlled gradient flows: the comparison principle. *arXiv*, https://arxiv.org/abs/2111.13258, 2021.

[124] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato. Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions. *arXiv*, https://arxiv.org/abs/2107.10535, 2021.

[125] A. Cosso and H. Pham. Zero-sum stochastic differential games of generalized McKean-Vlasov type. *J. Math. Pures Appl. (9)*, 129:180–212, 2019.

[126] M. G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comp.*, 43(167):1–19, 1984.

[127] M. Cranston. A probabilistic approach to gradient estimates. *Canad. Math. Bull.*, 35(1):46–55, 1992.

[128] J. G. Dai and R. J. Williams. Existence and uniqueness of semimartingale reflecting brownian motions in convex polyhedrons. *Theory Probab. Appl.*, 40(1):1–40, 1996.

[129] S. Daudin. Mean-field limit for stochastic control problems under state constraint. *arXiv preprint arXiv:2306.00949*, 2023.

[130] S. Daudin. Mean-field limit for stochastic control problems under state constraint. *arXiv preprint arXiv:2306.00949*, 2023.

[131] S. Daudin, F. Delarue, and J. Jackson. On the optimal rate for the convergence problem in mean field control. *arXiv preprint arXiv:2305.08423*, 2023.

[132] S. Daudin, F. Delarue, and J. Jackson. On the optimal rate for the convergence problem in mean field control. *arXiv preprint arXiv:2305.08423*, 2023.

[133] S. Daudin, J. Jackson, and B. Seeger. Well-posedness of Hamilton-Jacobi equations in the Wasserstein space: non-convex Hamiltonians and common noise. *arXiv preprint arXiv:2312.02324*, 2023.

[134] S. Daudin and B. Seeger. A comparison principle for semilinear Hamilton-Jacobi-Bellman equations in the Wasserstein space. Preprint,arXiv:2308.15174 [math.AP], 2023.

[135] D. A. Dawsont and J. Gärtner. Large deviations from the mckean-vlasov limit for weakly interacting diffusions. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(4):247–308, 1987.

[136] P. Degond. Global existence of smooth solutions for the vlasov-fokker-planck equation in 1 and 2 space dimensions. In *Annales scientifiques de l'École Normale Supérieure*, volume 19, pages 519–542, 1986.

[137] F. Delarue. On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. *Stochastic Processes and their Applications*, 99(2):209 – 286, 2002.

[138] F. Delarue. Restoring uniqueness to mean-field games by randomizing the equilibria. *Stochastics and Partial Differential Equations: Analysis and Computations*, 7:598–678, 2019.

[139] F. Delarue. Master equation for finite state mean field games with additive common noise. In *Mean Field Games, Cetraro, Italy 2019, Cardaliaguet, Pierre, Porretta, Alessio (Eds.)*, LNM 2281, pages 203–248. Springer, 2021.

[140] F. Delarue and F. Flandoli. The transition point in the zero noise limit for a 1D Peano example. *Discrete Contin. Dyn. Syst.*, 34(10):4071–4083, 2014.

[141] F. Delarue and R. Foguen Tchuendom. Selection of equilibria in a linear quadratic mean field game. *Stochastic Processes and their Applications*, 130(2):1000–1040, 2020.

[142] F. Delarue, D. Lacker, and K. Ramanan. The master equation and asymptotics for mean field games.
http://www.math.lsa.umich.edu/seminars_events/fileupload/4354_Lacker.pdf/, 2017.

[143] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: Large deviations and concentration of measure. *The Annals of Probability*, 2018.

[144] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem. *Electron. J. Probab.*, 24:Paper No. 51, 54, 2019.

[145] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: Large deviations and concentration of measure. 2020.

[146] F. Delarue and A. Tse. Uniform in time weak propagation of chaos on the torus. *arXiv*, https://arxiv.org/abs/2104.14973, 2021.

[147] L. Dello Schiavo. A Rademacher-type theorem on L^2-Wasserstein spaces over closed Riemannian manifolds. *J. Funct. Anal.*, 278(6):108397, 57, 2020.

[148] F. Demengel, G. Demengel, F. Demengel, and G. Demengel. Fractional sobolev spaces. *Functional spaces for the theory of elliptic partial differential equations*, pages 179–228, 2012.

[149] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(4):1183–1203, 2013.

[150] E. Di Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional sobolev spaces. *Bulletin des Sciences Mathématiques*, 136(5):521–573, 2012.

[151] R. J. DiPerna and P.-L. Lions. On the fokker-planck-boltzmann equation. *Communications in mathematical physics*, 120(1):1–23, 1988.

[152] M. F. Djete. Extended mean field control problem: a propagation of chaos result, 2022.

[153] M. F. Djete, D. Possamai, and X. Tan. Mckean-Vlasov optimal control: Limit theory and equivalence between different formulations. *Mathematics of Operations Research*, 2022.

[154] M. F. Djete, D. Possamaï, and X. Tan. Mckean-vlasov optimal control: the dynamic programming principle. *arXiv preprint arXiv:1907.08860*, 2019.

[155] J. Doncel, N. Gast, and B. Gaujal. Discrete Mean Field Games: Existence of Equilibria and Convergence. *Journal of Dynamics and Games*, 6(3):1–19, 2019.

[156] A. Douglis. The continuous dependence of generalized solutions of non-linear partial differential equations upon initial data. *Comm. Pure Appl. Math.*, 14:267–284, 1961.

[157] F. Dragoni and E. Feleqi. Ergodic mean field games with hörmander diffusions. *Calculus of Variations and Partial Differential Equations*, 57(5):1–22, 2018.

[158] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[159] P. Dupuis, K. Ramanan, and W. Wu. Large deviation principle for finite-state mean field interacting particle systems. *ArXiv e-prints*, Jan. 2016.

[160] R. Durrett. *Stochastic calculus*. Probability and Stochastics Series. CRC Press, Boca Raton, FL, 1996. A practical introduction.

[161] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.

[162] N. El Karoui, D. Hůů Nguyen, and M. Jeanblanc-Picqué. Compactification methods in the control of degenerate diffusions: Existence of an optimal control. *Stochastics*, 20(3):169–219, 1987.

[163] C. L. Epstein and R. Mazzeo. Wright-Fisher diffusion in one dimension. *SIAM J. Math. Anal.*, 42(2):568–608, 2010.

[164] C. L. Epstein and R. Mazzeo. Analysis of degenerate diffusion operators arising in population biology. In *From Fourier analysis and number theory to Radon transforms and geometry*, volume 28 of *Dev. Math.*, pages 203–216. Springer, New York, 2013.

[165] C. L. Epstein and R. Mazzeo. *Degenerate diffusion operators arising in population biology*, volume 185 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2013.

[166] C. L. Epstein and R. Mazzeo. Harnack inequalities and heat kernel estimates for degenerate diffusion operators arising in population biology. *Appl. Math. Res. Express. AMRX*, (2):217–280, 2016.

[167] C. L. Epstein and C. A. Pop. The Feynman-Kac formula and Harnack inequality for degenerate diffusions. *Ann. Probab.*, 45(5):3336–3384, 2017.

[168] S. N. Ethier. A class of degenerate diffusion processes occurring in population genetics. *Comm. Pure Appl. Math.*, 29(5):483–493, 1976.

[169] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[170] L. C. Evans. Some new PDE methods for weak KAM theory. *Calc. Var. Partial Differential Equations*, 17(2):159–177, 2003.

[171] L. C. Evans. Adjoint and compensated compactness methods for Hamilton-Jacobi PDE. *Arch. Ration. Mech. Anal.*, 197(3):1053–1088, 2010.

[172] L. C. Evans and H. Ishii. A pde approach to some asymptotic problems concerning random differential equations with small noise intensities. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 2, pages 1–20. Elsevier, 1985.

[173] L. C. Evans and P. E. Souganidis. A pde approach to certain large deviation problems for systems of parabolic equations. In *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, volume 6, pages 229–258. Elsevier, 1989.

[174] P. M. N. Feehan and C. A. Pop. On the martingale problem for degenerate-parabolic partial differential operators with unbounded coefficients and a mimicking theorem for Itô processes. *Trans. Amer. Math. Soc.*, 367(11):7565–7593, 2015.

[175] E. Feleqi, D. A. Gomes, and T. Tada. Hypoelliptic mean-field games—a case study. 2020.

[176] W. Feller. Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 227–246. University of California Press, Berkeley and Los Angeles, 1951.

[177] W. Feller. *An introduction to probability theory and its applications. Vol. I.* John Wiley & Sons, Inc., New York-London-Sydney, third edition, 1968.

[178] K. W. Fendick and M. A. Rodrigues. Asymptotic analysis of adaptive rate control for diverse sources with delayed feedback. *Information Theory, IEEE Transactions on*, 40(6):2008–2025, 1994.

[179] J. Feng. Large deviation for diffusions and hamilton–jacobi equation in hilbert spaces. *The Annals of Probability*, 34(1):321–385, 2006.

[180] J. Feng and M. Katsoulakis. A comparison principle for hamilton–jacobi equations related to controlled gradient flows in infinite dimensions. *Archive for rational mechanics and analysis*, 192(2):275–310, 2009.

[181] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*. Number 131. American Mathematical Soc., 2006.

[182] B. n. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Process. Appl.*, 71(1):33–53, 1997.

[183] A. F. Filippov. *Differential equations with discontinuous righthand sides*, volume 18 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1988. Translated from the Russian.

[184] R. Fiorenza. Sui problemi di derivata obliqua per le equazioni ellittiche. *Ricerche Mat*, 8:83–110, 1959.

[185] R. Fiorenza. Sui problemi di derivata obliqua per le equazioni ellittiche quasi lineari. *Ricerche Mat*, 15:74–108, 1966.

[186] M. Fischer. On the connection between symmetric N-player games and mean field games. *Ann. Appl. Probab.*, 127(2):757–810, 2017.

[187] F. Flandoli, F. Russo, and J. Wolf. Some SDEs with distributional drift. I. General calculus. *Osaka J. Math.*, 40(2):493–542, 2003.

[188] F. Flandoli, F. Russo, and J. Wolf. Some SDEs with distributional drift. II: Lyons-Zheng structure, Itô's formula and semimartingale characterization. *Random Oper. Stoch. Equ.*, 12(2):145–184, 2004.

[189] W. H. Fleming. Stochastic control for small noise intensities. *SIAM J. Control*, 9:473–517, 1971.

[190] W. H. Fleming. Generalized solutions in optimal stochastic control. Technical report, DTIC Document, 1976.

[191] W. H. Fleming. A stochastic control approach to some large deviations problems. In *Recent Mathematical Methods in Dynamic Programming*, pages 52–66. Springer, 1985.

[192] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition, 2006.

[193] W. H. Fleming and P. E. Souganidis. Asymptotic series and the method of vanishing viscosity. *Indiana Univ. Math. J.*, 35(2):425–447, 1986.

[194] W. H. Fleming and P. E. Souganidis. Pde-viscosity solution approach to some problems of large deviations. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 13(2):171–192, 1986.

[195] G. B. Folland. Subelliptic estimates and function spaces on nilpotent lie groups. *Arkiv för matematik*, 13(1):161–207, 1975.

[196] H. Föllmer and U. Horst. Convergence of locally and globally interacting Markov chains. *Stochastic Process. Appl.*, 96(1):99–121, 2001.

[197] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-field optimal control as gamma-limit of finite agent controls. *European Journal of Applied Mathematics*, 30(6):1153–1186, 2019.

[198] M. Fornasier, S. Lisini, C. Orrieri, and G. Savaré. Mean-field optimal control as gamma-limit of finite agent controls. *European J. Appl. Math.*, 30(6):1153–1186, 2019.

[199] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.

[200] C. Fricker and N. Gast. Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *EURO Journal on Transportation and Logistics*, pages 1–31, 2014.

[201] W. Gangbo, S. Mayorga, and A. Swiech. Finite dimensional approximations of Hamilton-Jacobi-Bellman equations in spaces of probability measures. *SIAM J. Math. Anal.*, 53(2):1320–1356, 2021.

[202] W. Gangbo and A. R. Mészáros. Global well-posedness of Master equations for deterministic displacement convex potential mean field games. *arXiv e-prints*, 2020.

[203] W. Gangbo and A. R. Mészáros. Global well-posedness of master equations for deterministic displacement convex potential mean field games. *arXiv*, https://arxiv.org/abs/2004.01660, 2020.

[204] W. Gangbo, A. R. Mészáros, C. Mou, and J. Zhang. Mean field games master equations with non-separable hamiltonians and displacement monotonicity. *arXiv*, https://arxiv.org/abs/2101.12362, 2021.

[205] W. Gangbo and A. Świech. Existence of a solution to an equation arising from the theory of mean field games. *J. Differential Equations*, 259(11):6573–6643, 2015.

[206] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 125:119–174, 2019.

[207] B. Gashi and J. Li. Backward stochastic differential equations with unbounded generators. *Stoch. Dyn.*, 19(1):1950008, 30, 2019.

[208] M. Germain, H. Pham, and X. Warin. Rate of convergence for particle approximation of PDEs in Wasserstein space. *Journal of Applied Probability*, 59(4):992–1008, 2022.

[209] R. Gibbens, P. Hunt, and F. Kelly. Bistability in communication networks. *Disorder in physical systems*, pages 113–128, 1990.

[210] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*, volume 224. springer, 2015.

[211] F. Golse, C. Imbert, C. Mouhot, and A. Vasseur. Harnack inequality for kinetic fokker-planck equations with rough coefficients and application to the landau equation. *arXiv preprint arXiv:1607.08068*, 2016.

[212] B. I. Golubov. Multiple series and Fourier integrals. In *Mathematical analysis, Vol. 19*, pages 3–54, 232. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Informatsii, Moscow, 1982.

[213] D. Gomes, H. Mitake, and K. Terai. The selection problem for some first–order stationary mean-field games. *Netw. Heterog. Media*, 15(4):681–710, 2020.

[214] D. Gomes and T. Seneci. Displacement convexity for first–order mean-field games. *Minimax Theory Appl.*, 3(2):261–284, 2018.

[215] D. A. Gomes, J. Mohr, and R. R. Souza. Discrete time, finite state space mean field games. *J. Math. Pures Appl.*, 93(3):308–328, 2010.

[216] D. A. Gomes, J. Mohr, and R. R. Souza. Continuous time finite state mean field games. *Appl. Math. Optim.*, 68(1):99–143, 2013.

[217] D. A. Gomes and J. a. Saúde. Mean field games models—a brief survey. *Dyn. Games Appl.*, 4(2):110–154, 2014.

[218] D. A. Gomes, R. M. Velho, and M.-T. Wolfram. Dual two-state mean-field games. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 2703–2708. IEEE, 2014.

[219] D. A. Gomes, R. M. Velho, and M.-T. Wolfram. Socio-economic applications of finite state mean field games. *Phil. Trans. R. Soc. A*, 372(2028):20130405, 2014.

[220] R. Gopalakrishnan, S. Doroudi, A. R. Ward, and A. Wierman. Routing and staffing when servers are strategic. *Oper. Res.*, 2016. to appear.

[221] C. Goulaouic and N. Shimakura. Regularité hölderienne de certains problèmes aux limites elliptiques dégénérés. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 10(1):79–108, 1983.

[222] P. J. Graber, A. R. Mészáros, F. J. Silva, and D. Tonon. The planning problem in mean field games as regularized mass transport. *Calc. Var. Partial Differential Equations*, 58(3):1–28, 2019.

[223] C. Graham. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability*, 37(1):198–211, 2000.

[224] C. Graham and P. Robert. Interacting multi-class transmissions in large stochastic networks. *The Annals of Applied Probability*, 19(6):2334–2361, 2009.

[225] M. Griffin-Pickering and A. R. Mészáros. A variational approach to first order kinetic mean field games with local couplings. *Communications in Partial Differential Equations*, 47(10):1945–2022, 2022.

[226] M. Gubinelli. Infinite dimensional rough dynamics. In *The Abel Symposium*, pages 401–413. Springer, 2016.

[227] O. Guéant. From infinity to one: The reduction of some mean field games to a global control problem. *Cahier de la Chaire Finance et Développement Durable*, 42, 2011.

[228] O. Guéant. Existence and uniqueness result for mean field games with congestion effect on graphs. *Appl. Math. Optim.*, 72(2):291–303, 2015.

[229] O. Guéant, J.-M. Lasry, and P.-L. Lions. Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 205–266. Springer, Berlin, 2011.

[230] O. Guéant and C.-A. Lehalle. General intensity shapes in optimal liquidation. *Math. Finance*, 25(3):457–495, 2015.

[231] S. Halfin and W. Whitt. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.*, 29(3):567–588, 1981.

[232] J. M. Harrison. *Brownian motion and stochastic flow systems*. Robert E. Krieger Publishing Co., Inc., Malabar, FL, 1990. Reprint of the 1985 original.

[233] J. Horowitz and R. L. Karandikar. Mean rates of convergence of empirical measures in the Wasserstein metric. *J. Comput. Appl. Math.*, 55(3):261–273, 1994.

[234] U. Horst. Stationary equilibria in discounted stochastic games with weakly interacting players. *Games Econom. Behav.*, 51(1):83–108, 2005.

[235] Y. Hu and S. Peng. A stability theorem of backward stochastic differential equations and its application. *C. R. Acad. Sci. Paris Sér. I Math.*, 324(9):1059–1064, 1997.

[236] M. Huang, P. E. Caines, and R. P. Malhamé. The Nash certainty equivalence principle and McKean-Vlasov systems: An invariance principle and entry adaptation. In *Decision and Control, 2007 46th IEEE Conference on*, pages 121–126. IEEE, 2007.

[237] M. Huang, R. P. Malhamé, and P. E. Caines. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. *Commun. Inf. Syst.*, 6(3):221–252, 2006.

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

[239] M. Huang, R. P. Malhamé, P. E. Caines, et al. Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[240] M. Huang, C. PE, and R. Malhame. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and nash equilibrium solutions, 2003.

[241] P. J. Hunt and T. G. Kurtz. Large loss networks. *Stochastic Process. Appl.*, 53(2):363–378, 1994.

[242] M. Iseri and J. Zhang. Set values for mean field games. *arXiv*, https://arxiv.org/abs/2107.01661, 2021.

[243] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987.

[244] F. Jean, O. Jerhaoui, and H. Zidani. Deterministic optimal control on Riemannian manifolds under probability knowledge of the initial condition. *HAL*, https://hal-ensta-paris.archives-ouvertes.fr//hal-03564787, Feb. 2022. working paper or preprint.

[245] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab*, pages 20–65, 1986.

[246] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[247] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 34:727–766, 12 1998.

[248] B. Jourdain, S. Méléard, and W. A. Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:1–29, 2008.

[249] B. Jovanovic and R. W. Rosenthal. Anonymous sequential games. *J. Math. Econom.*, 17(1):77–87, 1988.

[250] E. Kalai. Large robust games. *Econometrica*, 72(6):1631–1665, 2004.

[251] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[252] K. H. Karlsen and N. H. Risebro. A note on front tracking and equivalence between viscosity solutions of Hamilton-Jacobi equations and entropy solutions of scalar conservation laws. *Nonlinear Anal.*, 50(4, Ser. A: Theory Methods):455–469, 2002.

[253] F. P. Kelly. Loss networks. *Ann. App. Probab.*, pages 319–378, 1991.

[254] D. G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Ann. Math. Statistics*, 24:338–354, 1953.

[255] M. A. Khan and Y. Sun. Non-cooperative games with many players. *Handbook of game theory with economic applications*, 3:1761–1808, 2002.

[256] M. Kimura. Diffusion models in population genetics. *J. Appl. Probability*, 1:177–232, 1964.

[257] V. Kolokoltsov. Nonlinear markov games on a finite state space (mean-field and binary interactions). *Int. J. Stat. Prob.*, 1:77–91, 2012.

[258] V. Kolokoltsov and W. Yang. Sensitivity analysis for HJB equations with an application to coupled backward-forward systems. *ArXiv e-prints*, Mar. 2013.

[259] P. M. Kotelenez and T. G. Kurtz. Macroscopic limits for stochastic partial differential equations of McKean-Vlasov type. *Probab. Theory Related Fields*, 146(1-2):189–222, 2010.

[260] L. Kruk, J. Lehoczky, K. Ramanan, and S. Shreve. An explicit formula for the Skorokhod map on $[0,a]$. *Ann. Probab.*, 35(5):1740–1768, 2007.

[261] S. N. Kružkov. The Cauchy problem in the large for certain non-linear first order differential equations. *Soviet Math. Dokl.*, 1:474–477, 1960.

[262] S. N. Kružkov. Generalized solutions of nonlinear equations of the first order with several independent variables. II. *Mat. Sb. (N.S.)*, 72 (114):108–134, 1967.

[263] N. V. Krylov. *Controlled Diffusion Processes*. Springer-Verlag Berlin Heidelberg, 1980.

[264] N. V. Krylov. *Lectures on elliptic and parabolic equations in Hölder spaces*, volume 12 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[265] N. V. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, 131(2):154–196, 2005.

[266] T. G. Kurtz. *Approximation of population processes*, volume 36 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1981.

[267] T. G. Kurtz and J. Xiong. Numerical solutions for a class of SPDEs with application to filtering. In *Stochastics in finite and infinite dimensions*, Trends Math., pages 233–258. Birkhäuser Boston, Boston, MA, 2001.

[268] T. G. Kurtz and J. Xiong. A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems. *Commun. Math. Sci.*, 2(3):325–358, 2004.

[269] T. G. Kurtz and J. Xiong. A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems. *Commun. Math. Sci.*, 2(3):325–358, 2004.

[270] H. J. Kushner. *Heavy traffic analysis of controlled queueing and communication networks*, volume 47 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2001. Stochastic Modelling and Applied Probability.

[271] H. J. Kushner and P. G. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1992.

[272] A. Lachapelle, J. Lasry, C. Lehalle, and P. Lions. Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis. http://arxiv.org/abs/1305.6323, 2015.

[273] A. Lachapelle, J.-M. Lasry, C.-A. Lehalle, and P.-L. Lions. Efficiency of the price formation process in presence of high frequency participants: A mean field game analysis, May 2013.

[274] D. Lacker. A general characterization of the mean field limit for stochastic differential games. *Probab. Theory Related Fields*, pages 1–68, 2015.

[275] D. Lacker. Mean field games via controlled martingale problems: existence of Markovian equilibria. *Stochastic Process. Appl.*, 125(7):2856–2894, 2015.

[276] D. Lacker. A general characterization of the mean field limit for stochastic differential games. *Probab. Theory Related Fields*, 165(3-4):581–648, 2016.

[277] D. Lacker. Limit theory for controlled mckean-vlasov dynamics. *SIAM J. Control Optim.*, 55:1641–1672, 2017.

[278] D. Lacker. On the convergence of closed-loop Nash equilibria to the mean field game limit. *Ann. Appl. Probab.*, 30(4):1693–1761, 2020.

[279] D. Lacker, M. Shkolnikov, and J. Zhang. Superposition and mimicking theorems for conditional McKean-Vlasov equations. *arXiv*, https://arxiv.org/abs/2004.00099, 2020.

[280] O. Ladyzhenskaya and N. Ural'tseva. Linear and quasilinear elliptic equations 1st ed, 1968.

[281] J.-M. Lasry and P.-L. Lions. A remark on regularization in Hilbert spaces. *Israel Journal of Mathematics*, 55(3):257–266, 1986.

[282] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. i–le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.

[283] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris*, 343(9):619–625, 2006.

[284] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. ii–horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.

[285] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris*, 343(10):679–684, 2006.

[286] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

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

[288] M. Laurière and O. Pironneau. Dynamic programming for mean-field type control. *J. Optim. Theory Appl.*, 169(3):902–924, 2016.

[289] H. Lavenant and F. Santambrogio. Optimal density evolution with congestion: L^∞ bounds via flow interchange techniques and applications to variational mean field games. *Comm. Partial Differential Equations*, 43(12):1761–1802, 2018.

[290] J. Li, R. Bhattacharyya, S. Paul, S. Shakkottai, and V. Subramanian. Incentivizing sharing in realtime d2d streaming networks: A mean field game perspective. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 2119–2127. IEEE, 2015.

[291] G. M. Lieberman. Solvability of quasilinear elliptic equations with nonlinear boundary conditions. *Trans. Amer. Math. Soc.*, 273(2):753–765, 1982.

[292] G. M. Lieberman. The nonlinear oblique derivative problem for quasilinear elliptic equations. *Nonlinear Anal.*, 8(1):49–65, 1984.

[293] G. M. Lieberman. *Second order parabolic differential equations*. World Scientific Publishing Co., Inc., River Edge, NJ, 1996.

[294] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

[295] P.-L. Lions. Cours au collège de france, equations aux dérivées partielles et applications. https://www.college-de-france.fr/site/pierre-louis-lions/course-2010-2011.htm, 2010-11.

[296] P.-L. Lions. Cours au Collège de France, Equations aux dérivées partielles et applications. https://www.college-de-france.fr/site/pierre-louis-lions/course-2008-2009.htm, 2008-09.

[297] P.-L. Lions. Courses at the collège de france. www.college-de-france.fr.

[298] P.-L. Lions. Estimées nouvelles pour les équations quasilinéaires. Seminar in Applied Mathematics at the Collège de France. http://www.college-de-france.fr/site/pierre-louis-lions/seminar-2014-11-14-1

[299] P.-L. Lions. *Generalized solutions of Hamilton-Jacobi equations*, volume 69 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.

[300] P.-L. Lions. *Generalized solutions of Hamilton-Jacobi equations*, volume 69 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.

[301] P.-L. Lions and P. E. Souganidis. Extended mean-field games. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur.*, 31(3):611–625, 2020.

[302] A. Lunardi. Schauder estimates for a class of degenerate elliptic and parabolic operators with unbounded coefficients in $\mathbb{R}^n$. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 24(1):133–164, 1997.

[303] D. Luo and J. Wang. Coupling by reflection and hölder regularity for non-local operators of variable order. *Trans. Amer. Math. Soc.*, 371(1):431–459, 2019.

[304] J. Ma, P. Protter, and J. Yong. Solving forward-backward stochastic differential equations explicitly – a four step scheme. *Probability Theory and Related Fields*, 98(3):339–359, 1994.

[305] M. Manjrekar, V. Ramaswamy, and S. Shakkottai. A mean field game approach to scheduling in cellular systems. In *INFOCOM, 2014 Proceedings IEEE*, pages 1554–1562. IEEE, 2014.

[306] J. C. M. A. Q. Marc. Optimal control of multiagent systems in the Wasserstein space. *Calc. Var. Partial Differential Equations*, 59(2):no. 58, 2020.

[307] A. Mas-Colell. On a theorem of Schmeidler. *J. Math. Econom.*, 13(3):201–206, 1984.

[308] S. Mayorga and A. Święch. Finite dimensional approximations of Hamilton–Jacobi–Bellman equations for stochastic particle systems with common noise. *SIAM Journal on Control and Optimization*, 61(2):820–851, 2023.

[309] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, 1996.

[310] R. Merlose. Differential analysis on manifolds with corners. Unfinished book. http://www-math.mit.edu/ rbm/book.html.

[311] A. R. Mészáros and C. Mou. Mean field games systems under displacement monotonicity. *arXiv:2109.06687*, 2021.

[312] N. Mimikos-Stamatopoulos and S. Munoz. Regularity and long time behavior of one-dimensional first-order mean field games and the planning problem. *SIAM Journal on Mathematical Analysis*, 56(1):43–78, 2024.

[313] C. Mou and J. Zhang. Wellposedness of second order master equations for mean field games with nonsmooth data. *arXiv*, https://arxiv.org/abs/1903.09907, 2019.

[314] C. Mou and J. Zhang. Mean field game master equations with anti-monotonicity conditions. *arXiv*, https://arxiv.org/abs/2201.10762, 2022.

[315] C. Mouhot. De giorgi–nash–moser and hörmander theories: new interplays. In *Proceedings of the International Congress of Mathematicians (ICM 2018) (In 4 Volumes) Proceedings of the International Congress of Mathematicians 2018*, pages 2467–2493. World Scientific, 2018.

[316] S. Munoz. Classical and weak solutions to local first-order mean field games through elliptic regularity. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 39(1):1–39, 2022.

[317] S. Munoz. Classical solutions to local first-order extended mean field games. *ESAIM Control Optim. Calc. Var.*, 29, 2023.

[318] A. V. Nagaev. Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I. *Theory of Probability & Its Applications*, 14(1):51–64, 1969.

[319] C. Orrieri, A. Porretta, and G. Savaré. A variational approach to the mean field planning problem. *J. Funct. Anal.*, 277(6):1868–1957, 2019.

[320] H. Pham and X. Wei. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control Optim. Calc. Var.*, 24(1):437–461, 2018.

[321] C. A. Pop. C^0-estimates and smoothness of solutions to the parabolic equation defined by Kimura operators. *J. Funct. Anal.*, 272(1):47–82, 2017.

[322] C. A. Pop. Existence, uniqueness and the strong Markov property of solutions to Kimura diffusions with singular drift. *Trans. Amer. Math. Soc.*, 369(8):5543–5579, 2017.

[323] A. Porretta. Existence results for nonlinear parabolic equations via strong convergence of truncations. *Annali di Matematica Pura ed Applicata*, 177(1):143–172, 1999.

[324] A. Porretta. On the planning problem for the mean field games system. *Dyn Games Appl*, 4(2):231–256, 2014.

[325] A. Porretta. Weak solutions to fokker–planck equations and mean field games. *Archive for Rational Mechanics and Analysis*, 216(1):1–62, 2015.

[326] A. Porretta. On the weak theory for mean field games systems. *Bollettino dell'Unione Matematica Italiana*, 10(3):411–439, 2017.

[327] A. Porretta. On the turnpike property for mean field games. *Minimax Theory and Appl.*, 3(2):285–312, 2018.

[328] A. Porretta. Regularizing effects of the entropy functional in optimal transport and planning problems. *J. Funct. Anal.*, 284(3), 2023.

[329] A. Porretta and M. Ricciardi. Mean field games under invariance conditions for the state space. *Comm. Partial Differential Equations*, 45(2):146–190, 2020.

[330] E. Priola. Global schauder estimates for a class of degenerate kolmogorov equations. *arXiv preprint arXiv:0705.2810*, 2007.

[331] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.

[332] J.-F. raynalgneux, L. Szpruch, and A. Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. *The Annals of Applied Probability*, 32(3):1929 – 1969, 2022.

[333] N. Saldi. Large deviations principle for discrete-time mean-field games. *Systems & Control Letters*, 157:105042, 2021.

[334] K.-i. Sato. Diffusion processes and a class of Markov chains related to population genetics. *Osaka Math. J.*, 13(3):631–659, 1976.

[335] J. P. Schauder. Der fixpunktsatz in funktionalr´aumen. *Studia Math*, 2:171–180, 1930.

[336] D. Schmeidler. Equilibrium points of nonatomic games. *J. Statist. Phys.*, 7:295–300, 1973.

[337] N. Shimakura. Équations différentielles provenant de la génétique des populations. *Tôhoku Math. J.*, 29(2):287–318, 1977.

[338] J. Simon. Compact sets in the space $l^p((0,t);b)$. *Annali di Matematica pura ed applicata*, 146(1):65–96, 1986.

[339] D. R. Smart. *Fixed point theorems*. Cambridge University Press, London-New York, 1974. Cambridge Tracts in Mathematics, No. 66.

[340] H. M. Soner and Q. Yan. Viscosity Solutions for McKean-Vlasov Control on a torus. *arxiv:2212.11053*, pages 1–21, 2022.

[341] D. W. Stroock. *Lectures on stochastic analysis: diffusion theory*, volume 6 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1987.

[342] D. W. Stroock and S. R. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, volume 3, pages 333–359, 1972.

[343] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[344] A. Święch. A pde approach to large deviations in hilbert spaces. *Stochastic processes and their applications*, 119(4):1081–1123, 2009.

[345] A. Sznitman. A fluctuation result for nonlinear diffusions. *Infinite Dimensional Analysis and Stochastic Processes*, pages 145–160, 1985.

[346] A.-S. Sznitman. Topics in propagation of chaos. *Ecole d'Eté de Probabilités de Saint-Flour XIX—1989*, pages 165–251, 1991.

[347] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[348] M. Talbi, N. Touzi, and J. Zhang. From finite population optimal stopping to mean field optimal stopping. *arXiv preprint arXiv:2210.16004*, 2022.

[349] H. Tanaka. Limit theorems for certain diffusion processes with interaction. *North-Holland Mathematical Library*, 32:469–488, 1984.

[350] H. Tanaka and M. Hitsuda. Central limit theorem for a simple diffusion model of interacting particles. *Hiroshima Mathematical Journal*, 11(2):415–423, 1981.

[351] R. F. Tchuendom. Uniqueness for linear-quadratic mean field games with common noise. *Dyn. Games Appl.*, 8(1):199–210, 2018.

[352] A. Tse. Higher order regularity of nonlinear Fokker-Planck PDEs with respect to the measure component. *Journal de Mathématiques Pures et Appliquées*, 150:134–180, 2021.

[353] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[354] A. J. Veretennikov. Strong solutions and explicit formulas for solutions of stochastic integral equations. *Mat. Sb. (N.S.)*, 111(153)(3):434–452, 480, 1980.

[355] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[356] M. Vladimirova, S. Girard, H. Nguyen, and J. Arbel. Sub-weibull distributions: Generalizing sub-gaussian and sub-exponential properties to heavier tailed distributions. *Stat*, 9(1):e318, 2020. e318 sta4.318.

[357] N. D. Vvedenskaya and Y. M. Suhov. Dobrushin's Mean-Field Approximation for a Queue With Dynamic Routing. Research Report RR-3328, INRIA, 1997. Projet MEVAL.

[358] P. Wiecek, E. Altman, and A. Ghosh. Mean-field game approach to admission control of an M/M/∞ queue with shared service cost. *Dyn. Games Appl.*, pages 1–29, 2015.

[359] A. Wiśniewski. The structure of measurable mappings on metric spaces. *Proc. Amer. Math. Soc.*, 122(1):147–150, 1994.

[360] C. Wu and J. Zhang. Viscosity solutions to parabolic master equations and McKean-Vlasov SDEs with closed-loop controls. *Ann. Appl. Probab.*, 30(2):936–986, 2020.

[361] H. Ye, J. Gao, and Y. Ding. A generalized Gronwall inequality and its application to a fractional differential equation. *J. Math. Anal. Appl.*, 328(2):1075–1081, 2007.

[362] E. Zeidler. *Applied functional analysis*, volume 108 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1995. Applications to mathematical physics.

[363] F. Zhang and K. Du. Krylov-Safonov estimates for a degenerate diffusion process. *Stochastic Process. Appl.*, 130(8):5100–5123, 2020.

[364] L. Zhizhiashvili. *Trigonometric Fourier series and their conjugates*, volume 372 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996. Revised and updated translation of Some problems of the theory of trigonometric Fourier series and their conjugate series (Russian) [Tbilis. Gos. Univ., Tbilisi, 1993], Translated from the Russian by George Kvinikadze.

[365] C. Zălinescu. On uniformly convex functions. *J. Math. Anal. Appl.*, 95(2):344–374, 1983.

[366] A. K. Zvonkin. A transformation of the phase space of a diffusion process that will remove the drift. *Mat. Sb. (N.S.)*, 93(135):129–149, 152, 1974.