Università degli studi di Padova

DIPARTIMENTO DI MATEMATICA "Tullio Levi-Civita"

Corso di Laurea Magistrale in Matematica

Wasserstein regularity in mean field control problems

Relatore:

Prof. Marco Cirant

Laureanda:
Camilla Bonesini
Matricola:
2055481

21 luglio 2023

Anno Accademico 2022-2023

Abstract

In this thesis we deal with a class of mean field control problems that are obtained as limits of optimal control problems for large particle systems. Developing on [Cardaliaguet, P., Souganidis, P. E.; *Regularity of the value function and quantitative propagation of chaos for mean field control problem*, Nonlinear Differ. Equ. Appl., 2023], we analyze the value function $\mathcal{U}$ in Wasserstein metric and we prove its smoothness in an open and dense set of the space, time and probability measures using the strategy of the *linearized system.* The definition of this set exploits the concept of *strong stability.* Finally, we focus on chaos propagation: we study the properties of the optimal solutions of the interacting particle system starting from the aforementioned open and dense set. We also show some classical results on flows of probability measures via simple analytical tools.

Contents

Introduction

Optimal control is a field of study that deals with finding the best control law for a dynamic system in order to achieve a specific optimality condition.
In deterministic optimal control, by dynamical system we mean a set of differential equations for the evolution over time of the variable $y(t)$ that represents the state of the system at time t:

$$\begin{cases} \dot{y}(t) = f(y(t), \alpha(t)) & t > t_0 \geq 0, \\ y(t_0) = x_0 \in \mathbb{R}^N, \end{cases} \tag{1}$$

where $\alpha \in \mathcal{A}$ is the control taking values in the set of admissible controls

$$\mathcal{A} = \{\alpha : \mathbb{R} \to A \mid \alpha \text{ is measurable and the solution of (1) exists unique in } [t_0, T]\},$$

A is a topological space and $T > 0$ a finite time horizon.
The optimality condition in optimal control is defined via an objective cost function, e.g.

$$J(x_0, t_0, \alpha) = \int_{t_0}^{T} l(y(t), \alpha(t))dt + g(y(T))$$

with $l : \mathbb{R}^N \times A \to \mathbb{R}$ and $g : \mathbb{R}^N \to \mathbb{R}$ two given functions. This function quantifies the quality of the control inputs and the resulting system behavior. The objective is to find the control law that minimizes (or maximize) this cost function over the set of admissible controls. When the control is governed by a stochastic equation we call this problem stochastic optimal control.

Mean field game (MFG for short) theory is closely related to optimal control, and extends it to a setting where identical players simultaneously try to minimize some cost function.
Its birth dates back to 2006 when paper [18] by Lasry and Lions was published.
The main idea behind MFGs is the following: when the number of players N in a game is so big that it can be approximated by one with infinitely many indistinguishable agents, the analysis can be reduced to the study of a control problem with a single player representing the whole system.

In particular, in our case of interest the formalization of such a problem is the following. The dynamics of the state of the population is affected by the movement of an average control α via

$$dX_t = \alpha_t dt + \sqrt{2}B_t,$$

where B_t is a standard Brownian motion.
The representative agent aims to minimize the cost functional

$$\mathbb{E}\left[\int_{t_0}^{T} L(X_s, \alpha_s, m(s))ds + G(X_T, m(T))\right].$$

The following system of equations can be used to model a typical mean field game:

$$\begin{cases} \partial_t u - \Delta u + H(x, m, Du) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(H_p(x, m, Du)m) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0) = m_0, \;\; u(T, x) = G(x, m(T)) \text{ in } \mathbb{R}^d, \end{cases} \tag{2}$$

where $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the Lagrangian and its convex conjugate, $H(p) = \max_{a \in \mathbb{R}^d}\{a \cdot p - L(a)\}$, is the Hamiltonian, and $G : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ is the terminal cost.
In the above system (2), the map u can be interpreted as the value function of a player, while $m(t)$ is understood as the evolving probability density of the player at time t.

Mean field control (MFC) is a concept closely related to mean field games. In this case, a social planner controls the distribution of states and chooses a control strategy. MFC problems are control problems where the dynamic of the state X_t satisfies an equation depending on the law of the state $\mathcal{L}(X_t)$ itself. This is known as a Mckean-Vlasov equation:

$$dX_t = b(t, X_t, \mathcal{L}(X_t))dt + a(t, X_t, \mathcal{L}(X_t))dB_t,$$

where $\sigma(t, X_t, \mathcal{L}(X_t))$ and $b(t, X_t, \mathcal{L}(X_t))$ are measurable functions, $a(t, X_t, \mathcal{L}(X_t)) = \sigma(t, X_t, \mathcal{L}(X_t))\sigma(t, X_t, \mathcal{L}(X_t))^T$ and B_t denotes a Brownian motion.
Furthermore the optimality conditions for MFC and MFG coincide in the case of potential MFG, i.e. when the costs come from a derivative.

In this thesis we investigate and discuss the regularity of the value function for a class of MFC problems naturally arising as limits of some large particle systems. Then, using the regularity of this function, we obtain a propagation of chaos property that connects the behaviour of the optimal trajectories for the N-particle system and the one of the limit problem. We also show a convergence rate

for the value functions of the two problems.
In particular, we focus on the following setting. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and let $T > 0$ be a fixed time horizon. Set $t_0 \in [0, T]$ be the initial time and $x_0 = (x_0^1, ..., x_0^N) \in (\mathbb{R}^d)^N$ be the initial position of the system at time t_0. In order to control a N-particle system, we first focus on the minimization, over the set $\mathcal{A}^N$ of admissible controls $\alpha = (\alpha^k)_{k=1}^N \in L^2([0, T] \times \Omega; (\mathbb{R}^d)^N)$, of the following functional:

$$J^N(t_0, x_0, \alpha) = \mathbb{E}\left[\frac{1}{N}\int_{t_0}^T \sum_{i=1}^N L(X_t^i, \alpha_t^i)dt + \int_{t_0}^T \mathcal{F}(m_{X_t}^N)dt + \mathcal{G}(m_{X_T}^N)\right], \tag{3}$$

where the process X_t satisfies

$$X_t^k = x_0^k + \int_{t_0}^t \alpha_s^k ds + \sqrt{2}(B_t^k - B_{t_0}^k), \qquad \text{for } t \in [t_0, T],$$

with $(B^i)_{i=1,...,N}$ independent d-dimensional Brownian motions. In Equation (3), $\mathcal{F}$ and $\mathcal{G}$ are given data, $m_{X_t}^N$ corresponds to the empirical measure of the process X_t, and $L = L(x, \alpha) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the Legendre transform of the Hamiltonian H and it is assumed to be convex with quadratical growth in α.
Then, the value function for the problem is defined as

$$\mathcal{V}^N(t_0, x_0) = \inf_{\alpha \in \mathcal{A}^N} J^N(t_0, x_0, \alpha). \tag{4}$$

We show the convergence of the empirical measure of the optimal trajectories of (4), $m_{X_t^N}^N$, to the optimal solutions of the MFC problem. This last corresponds to minimizing the following functional

$$J^\infty(t_0, m_0, \alpha) = \mathbb{E}\left[\int_{t_0}^T L(X_t, \alpha_t) + \mathcal{F}(\mathcal{L}(X_t))dt + \mathcal{G}(\mathcal{L}(X_T))\right],$$

where m_0 is the initial distribution of the particles at time t_0, $\mathcal{L}(X_t)$ is the law of X_t, α is an admissible control square integrable $\mathbb{R}^d$-valued processes adapted to a Brownian motion B and to an initial condition $\overline{X_0}$, which is independent of B and of the law m_0 and the process $(X_t)_{t\in[t_0,T]}$ satisfies the following equation

$$X_t = \overline{X_0} + \int_{t_0}^t \alpha_s ds + \sqrt{2}(B_t - B_{t_0}) \qquad \text{for } t \in [t_0, T].$$

Finally, the value function of this last problem is

$$\mathcal{U}(t_0, x_0) = \inf_{\alpha \in \mathcal{A}} J^\infty(t_0, m_0, \alpha).$$

The rest of the thesis is structured as follows.
Chapter 1 is devoted to a recap of some basic definitions on Wasserstein distances, derivatives in the space of measures and basic results on MFC problems. In particular, we prove that any minimizer (m, α) of J^∞ corresponds to a solution to the following potential MFG system (i.e. the coupling functions F and G derive from potentials, that is $F = \frac{\delta \mathcal{F}}{\delta m}(m, x)$ and $G = \frac{\delta \mathcal{G}}{\delta m}(m, x)$):

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t)) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(H_p(x, Du)m) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0) = m_0, \;\; u(T, x) = G(x, m(T)) \text{ in } \mathbb{R}^d. \end{cases} \tag{5}$$

In chapter 2 we show a result on the convergence rate of $\mathcal{V}^N$ to $\mathcal{U}$, as N tends to infinity. The rate is obtained in a setting where the value function $\mathcal{U}$ does not need to be smooth.
First of all, we present some regularity estimates for $\mathcal{V}^N$ and $\mathcal{U}$. The second step consists in bounding from above $\mathcal{V}^N$ by $\mathcal{U}$: we define the function

$$\widehat{\mathcal{V}}^N(t, m) = \int_{(\mathbb{R}^d)^N} \mathcal{V}^N(t, x) \prod_{j=1}^N m(dx_j),$$

and, comparing $\mathcal{U}$ with $\widehat{\mathcal{V}}^N$ and $\widehat{\mathcal{V}}^N$ with $\mathcal{V}^N$, we obtain the thesis. The core of this chapter is the proof of the opposite inequality: we divide, via an appropriate partition, the players into subgroups in order to get that the optimal controls for the agents in each of them are close. We show the inequality holds for each subgroup.
So, we obtain that there exist $\beta \in (0, 1]$ and a constant $C > 0$ such that, for any $(t, x) \in [0, T] \times (\mathbb{R}^d)^N$,

$$|\mathcal{V}^N(t, x) - \mathcal{U}(t, m_x^N)| \leq C N^{-\beta}(1 + M_2^{1/2}(m_x^N)),$$

where $M_2^{1/2}(m_x^N) = \frac{1}{N}\sum_{i=1}^N |x_i|^2$.

Chapter 3 represents the core of the work. We introduce the linearized version of the system (5)

$$\begin{cases} -\partial_t z - \Delta z + H_p(x, Du) \cdot Dz = \frac{\delta F}{\delta m}(x, m(t))(\mu(t)) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t \mu - \Delta \mu - div(H_p(x, Du)\mu) - \sigma div(H_{pp}(x, Du)Dzm) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \mu(t_0) = 0 \text{ and } z(T, x) = \frac{\delta G}{\delta m}(x, m(T))(\mu(T)) \text{ in } \mathbb{R}^d. \end{cases} \tag{6}$$

We exploit the notion of stability, that is when the only solution to (6) is the trivial one, to show that, in an open and dense set $\mathcal{O}$ of initial times and measures, the value function $\mathcal{U}$ is smooth and it is a classical solution in $\mathcal{O}$ of the master Hamilton-Jacobi equation

$$-\partial_t \mathcal{U}(t,m) - \int_{\mathbb{R}^d} \operatorname{div}(D_m \mathcal{U}(t,m,y))m(dy) + \int_{\mathbb{R}^d} H(y, D_m \mathcal{U}(t,m,y))m(dy) = \mathcal{F}(m).$$

Chapter 4 focuses on the proof of the propagation of chaos property, that describes the limit behavior of the particle system when the number of particles grows to infinity.
We study the properties of the optimal trajectories of the interacting N-particle system exploiting the results obtained in the previous chapter. So, for every $(t_0, m_0) \in \mathcal{O}$, taking a sequence of independent random variables with law m_0, $Z = (Z^k)_{k=1,\dots,N}$, a sequence of independent Brownian motions independent of Z, $B = (B^k)_{k=1,\dots,N}$, and a sequence of optimal trajectories for $\mathcal{V}^N$, $Y^N = (Y^{N,k})_{k=1,\dots,N}$, such that

$$Y_t^{N,k} = Z^k - \int_{t_0}^{t} H_p(Y_s^k, D\mathcal{V}^N(s, Y_s^N))ds + \sqrt{2}(B_t^k - B_{t_0}^k),$$

we show that there exist $\gamma \in (0,1)$ and $C > 0$ such that

$$\mathbb{E}[\sup_{t\in[t_0,T]} d_1(m^N_{Y_t^N}, m(t))] \le CN^{-\gamma}.$$

Some additional material is collected in the Appendix. In particular, in Appendix A we show a result on the Wasserstein distance using only analytical tools. In Appendix B we present the Kantorovich duality theorem that we use to prove the Kantorovich-Rubistein theorem. Finally, in Appendix C we recall the Lions-Malgrange-type argument: a result to prove uniqueness of solution for general linear forward-backward systems given the initial data.

Chapter 1

Definitions and preliminary facts

In this chapter we provide the necessary notations and the consequent definitions that will then be used in the rest of the work.

1.1 General notation

Let $d \in \mathbb{N}$. We work on $\mathbb{R}^d$. Let B_R be the ball of radius R centred at the origin. We use the notation $\mathcal{P}(\mathbb{R}^d)$ to denote the *set of Borel probability measures* on $\mathbb{R}^d$. Let m in $\mathcal{P}(\mathbb{R}^d)$ and $p \geq 1$ and we call $M_p(m) = \int_{\mathbb{R}^d} |x|^p dm$ the p^{th}-moment of m. We denote by $\mathcal{P}_p(\mathbb{R}^d)$ the set of m in $\mathcal{P}(\mathbb{R}^d)$ such that $M_p(m) < \infty$.
In what follow, we use positive constants C that may change from line to line.

1.1.1 The Wasserstein distance

Let $p \geq 1$, we endow $\mathcal{P}_p(\mathbb{R}^d)$ with the Wasserstein metric d_p defined in the following way:

$$d_p^p(m, m') := \inf_{\pi \in \Pi(m,m')} \int_{\mathbb{R}^d} |x - y|^p d\pi(x, y),$$

where $\Pi(m, m')$ is the set of all $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals m and m'.
In particular, for $p = 1$, we have

Theorem 1.1 (Kantorovich-Rubinstein Theorem)**.** *For any $m, m' \in \mathcal{P}_1(\mathbb{R}^d)$, then*

$$d_1(m, m') = \sup_{\phi:\ 1-Lip} \int_{\mathbb{R}^d} \phi \, d(m - m').$$

For the proof, we refer to the last part of Appendix B.

Proposition 1.2. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and X_μ, X_ν be a stochastic variable whose law is μ and ν respectively. It holds that*

$$d_1(\mu,\nu) \le \mathbb{E}[|X_\mu - X_\nu|].$$

Proof. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function with its Lipschitz constant $C \le 1$. We have

$$|f(x) - f(y)| \le C|x-y| \le |x-y|.$$

Hence,

$$\begin{aligned}\int_{\mathbb{R}^d} f(x)\mu(dx) - \int_{\mathbb{R}^d} f(y)\nu(dy) &= \mathbb{E}[f(X_\mu)] - \mathbb{E}[f(X_\nu)] \le |\mathbb{E}[f(X_\mu) - f(X_\nu)]| \\ &\le \mathbb{E}[|f(X_\mu) - f(X_\nu)|] \le \mathbb{E}[|X_\mu - X_\nu|].\end{aligned}$$

This means that

$$\begin{aligned}d_1(\mu,\nu) &= \sup_{f:\,1-Lip} \int_{\mathbb{R}^d} f\, d(\mu-\nu) = \sup_{f:\,1-Lip} \left\{ \int_{\mathbb{R}^d} f(x)\mu(dx) - \int_{\mathbb{R}^d} f(y)\nu(dy) \right\} \\ &\le \mathbb{E}[|X_\mu - X_\nu|].\end{aligned}$$

□

Remark 1.3. *One can show that*

$$d_1(\mu,\nu) = \inf_{X_\mu, X_\nu} \mathbb{E}[|X_\mu - X_\nu|],$$

where the infimum is taken over random variables X_μ and X_ν such that the law of X_μ (resp. X_ν) is μ (resp. ν).

Remark 1.4. *Another useful distance on the space of measures is the Wasserstein distance d_2. It is defined on the space $\mathcal{P}_2(\mathbb{R}^d)$ of Borel probability measures m with a finite second order moment (i.e., $\int_{\mathbb{R}^d} |x|^2 m(dx) < +\infty$) by*

$$d_2(m_1, m_2) = \inf_\pi \left(\int_{\mathbb{R}^{2d}} |x-y|^2 d\pi(x,y) \right)^{1/2},$$

where the infimum is taken over the Borel probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal given by m_1 and second marginal by m_2:

$$\int_{\mathbb{R}^{2d}} \phi(x) d\pi(x,y) = \int_{\mathbb{R}^d} \phi(x) dm_1(x) \quad \text{and} \quad \int_{\mathbb{R}^{2d}} \phi(y) d\pi(x,y) = \int_{\mathbb{R}^d} \phi(y) dm_2(y).$$

The distance can be defined equivalently by

$$d_2(m_1, m_2) = \inf_{X,Y} \left(\mathbb{E}[|X-Y|^2] \right)^{1/2},$$

where the infimum is taken over random variables X and Y with law m_1 and m_2 respectively.

Remark 1.5. *By Hölder inequality* $\mathcal{P}_r(\mathbb{R}^d) \subset \mathcal{P}_s(\mathbb{R}^d)$*, for any* $1 \le s \le r$*, and*

$$d_s(m, m') \le d_r(m, m') \qquad \forall m, m' \in \mathcal{P}_r(\mathbb{R}^d).$$

1.1.2 Derivatives in the space of measures

Let $U : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$. We, first, restrict the function U to the elements $m \in \mathcal{P}(\mathbb{R}^d)$ which have a density in $L^2(\mathbb{R}^d)$ and assume that the function is defined in a neighborhood $\mathcal{O} \subset L^2(\mathbb{R}^d)$ of $L^2(\mathbb{R}^d) \cap \mathcal{P}(\mathbb{R}^d)$. We write

$$\frac{\delta U}{\delta m}(p)(q) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(U(p + \epsilon q) - U(p)) \qquad p \in \mathcal{O};\ q \in L^2(\mathbb{R}^d).$$

We can identify $\frac{\delta U}{\delta m}(p)$ with an element of $L^2(\mathbb{R}^d)$. We set, when possible

$$D_m U(m, y) = D_y \frac{\delta U}{\delta m}(m, y) \quad \text{and} \quad D^2_{mm} U(m, \cdot, y, y') = D^2_{y,y'} \frac{\delta U}{\delta m}(m, y, y').$$

In the same way, we denote by $\frac{\delta^2 U}{\delta^2 m}$ the second derivative of U

$$\frac{\delta^2 U}{\delta m^2}(p)(q, q') = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(\frac{\delta U}{\delta m}(p + \epsilon q)(q') - \frac{\delta U}{\delta m}(p)(q') \right) \qquad p \in \mathcal{O};\ q, q' \in L^2(\mathbb{R}^d).$$

We can also consider, in a more general way, the derivatives out of L^2 in $\mathcal{P}(\mathbb{R}^d)$.

Definition 1.6. *We say that* $\mathcal{U} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ *is* $\mathcal{C}^1$ *if there exists a continuous map* $\frac{\delta \mathcal{U}}{\delta m} : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ *such that, for any* $m, m_1 \in \mathcal{P}(\mathbb{R}^d)$*,*

$$\lim_{s \to 0^+} \frac{\mathcal{U}((1-s)m + s m_1) - \mathcal{U}(m)}{s} = \int_{\mathbb{R}^d} \frac{\delta \mathcal{U}}{\delta m}(m, y) d(m_1 - m)(y).$$

Since $\frac{\delta \mathcal{U}}{\delta m}$ *is defined up to an additive constant, we adopt the* ***normalization convention***

$$\int_{\mathbb{R}^d} \frac{\delta \mathcal{U}}{\delta m}(m, y) dm(y) = 0.$$

1.2 Assumptions

We state now some assumptions on the data that will be involved in the work: some are more standard than others but will be important to prove the regularity of the value function $\mathcal{U}$.

Let $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $\mathcal{F} : \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ and $\mathcal{G} : \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$.
We define $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ as the *Legendre transform* of H with respect to the second variable as

$$L(x, \alpha) = \sup_{p \in \mathbb{R}^d} [-a \cdot p - H(x, p)].$$

We assume that:

- $H \in \mathcal{C}^4_{loc}$ and strictly convex with respect to the second variable,
- $\exists C > 0$ constant such that, $\forall (x, p) \in \mathbb{R}^d \times \mathbb{R}^d$ $C^{-1}\, \mathbb{I}_d \leq H_{pp}(x, p) \leq C\, \mathbb{I}_d$, $-C + C^{-1}|p|^2 \leq H(x, p) \leq C(1 + |p|^2)$ and $|D_x H(x, p)| \leq C(|p| + 1)$,
- $\mathcal{F} \in \mathcal{C}^2$ with $\mathcal{F}$, $D_m\mathcal{F}$, $D^2_{ym}\mathcal{F}$ and $D^2_{mm}\mathcal{F}$ uniformly bounded and, moreover, $x \to \frac{\delta \mathcal{F}}{\delta m}(m, x)$ is bounded in $\mathcal{C}^2$ uniformly in m, while $y \to \frac{\delta^2 \mathcal{F}}{\delta m^2}(m, x, y)$ is bounded in $\mathcal{C}^2$ uniformly in (m, x),
- $\mathcal{G} \in \mathcal{C}^4$ with all derivatives up to order 4 uniformly bounded.

For convenience, in what follow, we will call all the assumptions on H, $\mathcal{F}$ and $\mathcal{G}$ (1.1)

An example of a Hamiltonian satisfying (1.1) is

$$H(x, p) = |p|^2$$

and a typical $\mathcal{F}$ is the class of cylindrical functions of the form

$$\mathcal{F}(m) = \mathsf{F}\left(\int_{\mathbb{R}^d} \mathcal{F}_1(x) dm(x), \ldots, \int_{\mathbb{R}^d} \mathcal{F}_k(x) dm(x) \right),$$

where F and the $\mathcal{F}_i$, for $1 \leq i \leq k$, are smooth with bounded derivatives.

Remark 1.7. *We note that the strict convexity of H with respect to the second variable implies that L has the same regularity as H.*
The uniform bounds on $D_m\mathcal{F}$ and $D_m\mathcal{G}$ imply that both maps are Lipschitz continuous in $\mathcal{P}_1(\mathbb{R}^d)$.

Remark 1.8. *Since L is the Legendre transform of H and for $|a| \leq R$, $x \in \mathbb{R}^d$ and $p = D_a L(x, a)$, in view of the hypothesis (1.1), we have $L(x, a) = -a \cdot p - H(x, p)$ and that*

$$-R|p| - C + \frac{1}{c}|p|^2 \leq L(x, a) \leq \sup_{p'} \{-a \cdot p' + C - c|p'|^2\} \leq C + \frac{R^2}{4c}.$$

So, we have that for every $R > 0$, there exists $C_R > 0$ such that

$$|D_a L(x, a)| \leq C_R \ \textit{for all} \ (x, a) \in \mathbb{R}^d \times B_R. \tag{1.2}$$

1.3 Background

In this part we present the mean field control (MFC for short) problems obtained as the limit of optimal control problem for large particle systems.

Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probabilistic space, $X = (X^1, ..., X^N)$ the trajectories that satisfy, for each $k \in 1, ..., N$,

$$X_t^k = x_0^k + \int_{t_0}^t \alpha_s^k ds + \sqrt{2}(B_t^k - B_{t_0}^k) \qquad \text{for } t \in [t_0, T], \tag{1.3}$$

where the $(B^i)_{i=1,...,N}$ are independent d-dimensional Brownian motions and $\alpha = (\alpha^k)_{k=1}^N \in L^2([0,T] \times \Omega; (\mathbb{R}^d)^N)$ are admissible controls adapted to the filtration generated by them.
We define the **empirical measure** of the process X_t, where δ_x is the Dirac mass center in x :

$$m_{X_t}^N := \frac{1}{N} \sum_{k=1}^N \delta_{X_t^k}.$$

Let $T > 0$ be a fixed time horizon, $t_0 \in [0,T]$ the initial time and $x_0 = (x_0^1, ..., x_0^N) \in (\mathbb{R}^d)^N$ the initial position of the system at time t_0.
We want to minimize the following problem over the set $\mathcal{A}^N$ of admissible controls $\alpha = (\alpha^k)_{k=1}^N \in L^2([0,T] \times \Omega; (\mathbb{R}^d)^N)$ in order to control a system of N particles

$$J^N(t_0, x_0, \alpha) := \mathbb{E}\left[\frac{1}{N}\int_{t_0}^T \sum_{i=1}^N L(X_t^i, \alpha_t^i) dt + \int_{t_0}^T \mathcal{F}(m_{X_t}^N) dt + \mathcal{G}(m_{X_T}^N)\right], \tag{1.4}$$

where $L = L(x, \alpha) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a convex function with quadratically growth in α and $(B^i)_{i=1,...,N}$ are independent d-dimensional Brownian motions.

For this problem, the value function is

$$\begin{aligned}\mathcal{V}^N(t_0, x_0) &:= \inf_{\alpha \in \mathcal{A}^N} J^N(t_0, x_0, \alpha) \\ &= \inf_{\alpha \in \mathcal{A}^N} \mathbb{E}\left[\frac{1}{N}\int_{t_0}^T \sum_{i=1}^N L(X_t^i, \alpha_t^i) dt + \int_{t_0}^T \mathcal{F}(m_{X_t}^N) dt + \mathcal{G}(m_{X_T}^N)\right].\end{aligned} \tag{1.5}$$

It can be proved that (see [17]), in a more general framework and under slightly different hypothesis on the data, the empirical measure $m_{X_t^N}^N$ in the optimal trajectories of (1.5) converges to the weak optimal solutions of the mean field control problem which consists of minimizing the following functional

$$J^{\infty}(t_0, m_0, \alpha) = \mathbb{E}\left[\int_{t_0}^{T} L(X_t, \alpha_t)dt + \int_{t_0}^{T} \mathcal{F}(\mathcal{L}(X_t))dt + \mathcal{G}(\mathcal{L}(X_T))\right], \tag{1.6}$$

where m_0 is the initial distribution of the particles at time t_0, $\mathcal{L}(X_t)$ is the law of X_t, $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is the set of admissible controls that are square integrable $\mathbb{R}^d$-valued processes adapted to a Brownian motion B and to an initial condition $\overline{X_0}$, which is independent of B and of the law m_0.

The process $(X_t)_{t\in[t_0,T]}$ satisfies the following equation:

$$X_t = \overline{X_0} + \int_{t_0}^{t} \alpha_s ds + \sqrt{2}(B_t - B_{t_0}) \qquad \text{for } t \in [t_0, T]. \tag{1.7}$$

The value function of this last problem is

$$\begin{aligned} \mathcal{U}(t_0, x_0) &:= \inf_{\alpha\in\mathcal{A}} J^{\infty}(t_0, m_0, \alpha) \\ &= \inf_{\alpha\in\mathcal{A}} \mathbb{E}\left[\int_{t_0}^{T} L(X_t, \alpha_t)dt + \int_{t_0}^{T} \mathcal{F}(\mathcal{L}(X_t))dt + \mathcal{G}(\mathcal{L}(X_T))\right]. \end{aligned} \tag{1.8}$$

1.4 Preliminary facts

We call $\mathcal{M}(t_0, m_0)$ the set of controls defined by

$$\mathcal{M}(t_0, m_0) := \left\{ \begin{array}{l} (m, \alpha) \in \mathcal{C}^0([t_0, T], \mathcal{P}_1(\mathbb{R}^d)) \times L^0([t_0, T] \times \mathbb{R}^d; \mathbb{R}^d) \\ \int_0^T \int_{\mathbb{R}^d} |\alpha|^2 m < \infty \\ \partial_t m - \Delta m + \text{div}(m\alpha) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d \text{ and } m(t_0) = m_0 \text{ in } \mathbb{R}^d \end{array} \right\},$$

for each initial point $(t_0, m_0) \in [0, T) \times \mathcal{P}_1(\mathbb{R}^d)$ with the equation understood in the sense of distributions.

We can rewrite the value function in (1.8) as

$$\mathcal{U}(t_0, m_0) = \inf_{(m,\alpha)\in\mathcal{M}(t_0,m_0)} \left\{\int_{t_0}^{T} \int_{\mathbb{R}^d} L(x, \alpha(t, x))m(t, dx) + \mathcal{F}(m(t))dt + \mathcal{G}(m(T))\right\}. \tag{1.9}$$

We also define the set $\mathcal{O}$

$$\mathcal{O} := \{(t_0, m_0) \in [t_0, T]\times\mathcal{P}_2(\mathbb{R}^d) : \exists! \text{ stable minimizer in the definition of } \mathcal{U}(t_0, m_0)\}. \tag{1.10}$$

In what follows, we will prove that $\mathcal{U}$ is smooth in $\mathcal{O}$.

Let $(t_0, m_0) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$. We have [see section 1.5] that for $\mathcal{U}(t_0, m_0)$ there exists at least one minimizer and, if $(m, \alpha) \in \mathcal{M}(t_0, m_0)$ is a minimizer, there exists a multiplier $u : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}$ such that $\alpha = -H_p(x, Du)$ and the pair (u, m) solves the mean field game system

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t)) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(H_p(x, Du)m) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0) = m_0, \;\; u(T, x) = G(x, m(T)) \text{ in } \mathbb{R}^d, \end{cases} \tag{1.11}$$

where

$$F(x,m) = \frac{\delta \mathcal{F}}{\delta m}(m,x) \quad \text{and} \quad G(x,m) = \frac{\delta \mathcal{G}}{\delta m}(m,x)$$

The system is made by a pair formed by a *backward Hamilton-Jacobi equation*, describing the dynamics of the value function of any of the players, and a *forward Kolmogorov equation*, describing the dynamics of the distribution of the population. We can think of F as a running cost and G as a terminal cost.
Since the hypothesis of strict convexity on H, given (m, α), Du is defined uniquely by the relation $\alpha = -H_p(x, Du)$.

Lemma 1.9. *Assume* (1.1) *and let* (u, m) *be a solution of* (1.11)*.Then,*

i) for any $\delta \in (0,1)$ *and* $t, t' \in [0,T]$ *there is a constant* $C > 0$ *independent of* (t_0, m_0) *such that*

$$||u||_{\mathcal{C}^{(\delta+1)/2, \delta+1}} + \sup_{t \neq t'} \frac{d_2(m(t), m(t'))}{|t-t'|^{1/2}} \leq C, \tag{1.12}$$

ii) there exists a constant C *such that*

$$\sup_{t \in [t_0, T]} \int_{\mathbb{R}^d} |x|^2 m(t, dx) \leq (1 + CT) \int_{\mathbb{R}^d} |x|^2 m_0(t, dx) + CT. \tag{1.13}$$

Proof. i)The $\mathcal{C}^{(1+\delta)/2, 1+\delta}$ local regularity of u is a consequence of the classical parabolic regularity theory, since

- the map $x \to G(x, m)$ is bounded $\mathcal{C}^4$
- the map $x \to \frac{\delta F}{\delta m}(x, m(t))$ is of class $\mathcal{C}^2$
- using the fact that m is a solution (see Proposition 1.13) to

$$\partial_t m - \Delta m - \operatorname{div}(H_p m) = 0$$

and the divergence theorem we have

$$\begin{aligned}\frac{d}{dt}F(x,m(t)) &= \int_{\mathbb{R}^d}\frac{\delta F}{\delta m}(m(t),x,y)\partial_t m(t,y)dy\\ &= \int_{\mathbb{R}^d}\frac{\delta F}{\delta m}(m(t),x,y)\Delta m + \frac{\delta F}{\delta m}(m(t),x,y)\mathrm{div}(H_p(y,Du(t,y))m(t,y))dy\\ &= \int_{\mathbb{R}^d}(\Delta_y\frac{\delta F}{\delta m}(m(t),x,y) - H_p(y,Du(t,y))\cdot D_y\frac{\delta F}{\delta m}(m(t),x,y)m(t,y)dy\end{aligned}$$

that implies the map $t \to F(x,m(t))$ is $\mathcal{C}^{(1+\delta)/2}$ for any $\delta \in (0,1)$ for the fact that $y \to \frac{\delta^2 F}{\delta m^2}(m,x,y)$ is bounded in $\mathcal{C}^2$ uniformly in (m,x).

This regularity holds globally in space since there exists a constant $C > 0$ such that for every $x, y \in \mathbb{R}^d$

$$\sup_{t\in[t_0,T]}|F(x,m(t)) - F(y,m(t))| + |G(x,m(t)) - G(y,m(t))| \le C|x-y|$$

for the uniform boundedness of $D^2_{mm}\mathcal{F}$ and $D^2_{mm}\mathcal{G}$.
Applying the maximum principle, recalling that for T terminal time we have $u(T,x) = G(x,m(T))$, we get that u is Lipschitz continuous in the space variable. The same argument applied to the equation satisfied by u_{x_i} for each $i = 1,\dots,d$, implies that Du is Lipschitz in the space variable.

To prove the bound on d_2, let us assume $t' = 0$. Consider

$$\begin{cases}\partial_t m - \Delta m - \mathrm{div}(H_p(x,Du(t,x)m) = 0 \quad \text{in} \quad \mathbb{R}^d\times(0,T),\\ m(0) = m_0 \quad \text{in} \quad \mathbb{R}^d,\end{cases} \tag{1.14}$$

with

$$m(t,x) = \int_{\mathbb{R}^d}\mu(t,x,y)dy$$

and

$$\partial_t\mu - \Delta_x\mu - \mathrm{div}_x(H_p(x,Du(t,x))\mu(t,x,y)) = 0 \quad \text{in} \quad \mathbb{R}^{2d}\times(0,T). \tag{1.15}$$

Multiplying (1.15) by $|x-y|^2$ and integrating over $\mathbb{R}^{2d}$ we obtain

$$\begin{aligned}&\int_{\mathbb{R}^{2d}}\partial_t\mu\cdot|x-y|^2dxdy - \int_{\mathbb{R}^{2d}}\Delta_x\mu\cdot|x-y|^2\\ &\qquad - \int_{\mathbb{R}^{2d}}\mathrm{div}_x(H_p(x,Du(t,x))\mu(t,x,y))\cdot|x-y|^2dxdy = 0.\end{aligned}$$

Notice that, applying twice the divergence theorem since $\Delta_x \mu = \mathrm{div}_x(\nabla_x \mu)$,

$$-\int_{\mathbb{R}^{2d}} \Delta_x \mu \cdot |x-y|^2 dxdy = 2\int_{\mathbb{R}^{2d}} \nabla_x \mu \cdot (x-y) dxdy = -2d,$$

where we used that $\int_{\mathbb{R}^{2d}} \mu(t,x,y)dxdy = 1$. Using the divergence theorem, we have that

$$-\int_{\mathbb{R}^{2d}} \mathrm{div}_x(H_p(x,Du)\mu(t,x,y)) \cdot |x-y|^2 dxdy = 2\int_{\mathbb{R}^{2d}} H_p(x,Du)\mu(t,x,y)(x-y)dxdy,$$

and so

$$\int_{\mathbb{R}^{2d}} \partial_t \mu \cdot |x-y|^2 dxdy - 2d + 2\int_{\mathbb{R}^{2d}} H_p(x,Du)\mu(t,x,y)(x-y)dxdy = 0.$$

Using the Young's inequality

$$\begin{aligned}\int_{\mathbb{R}^{2d}} \mu(t,x,y)|x-y|dxdy &= \int_{\mathbb{R}^{2d}} (\sqrt{\mu(t,x,y)}|x-y|)\sqrt{\mu(t,x,y)}dxdy \\ &\leq \frac{1}{2}\left(\int_{\mathbb{R}^{2d}} |x-y|^2\mu(t,x,y)dxdy + \int_{\mathbb{R}^{2d}} \mu(t,x,y)dxdy\right),\end{aligned}$$

and that $H_p(Du)$ is bounded, we have

$$\begin{aligned}&2\int_{\mathbb{R}^{2d}} H_p(x,Du)\mu(t,x,y)|x-y|dxdy \leq 2||H_p(x,Du)||_\infty \int_{\mathbb{R}^{2d}} \mu(t,x,y)|x-y|dxdy \\ &\leq ||H_p(x,Du)||_\infty \left(\int_{\mathbb{R}^{2d}} |x-y|^2\mu(t,x,y)dxdy + \int_{\mathbb{R}^{2d}} \mu(t,x,y)dxdy\right) \\ &\leq C\left(\int_{\mathbb{R}^{2d}} |x-y|^2\mu(t,x,y)dxdy + 1\right),\end{aligned}$$

and so, putting everything together, we have

$$\int_{\mathbb{R}^{2d}} \partial_t \mu \cdot |x-y|^2 dxdy - 2d \leq C\left(\int_{\mathbb{R}^{2d}} |x-y|^2\mu(t,x,y)dxdy + 1\right),$$

and applying the Gronwall's lemma and knowing that d_2 is a distance and so there exist μ such that $d_2(m,m) = \int_{\mathbb{R}^{2d}} |x-y|^2\mu(0,x,y)dxdy = 0$ we get

$$\int_{\mathbb{R}^{2d}} \mu(t,x,y) \cdot |x-y|^2 dxdy \leq C(-t-1+e^t) + Ct$$

and, there exists a $K > 0$ such that

$$\int_{\mathbb{R}^{2d}} \mu(t,x,y) \cdot |x-y|^2 dxdy \leq Kt$$

The last inequality follow since we assumed $t \in (0,T)$.
Using that

$$d_2^2(m(t), m(0)) = \inf_{\pi \in \Pi(m(t),m(0))} \int_{\mathbb{R}^{2d}} |x-y|^2 d\pi(x,y) \leq \int_{\mathbb{R}^{2d}} |x-y|^2 \mu(t,x,y) dxdy,$$

and so we have that

$$d_2(m(t), m(0)) \leq C\sqrt{t},$$

and we get the thesis.

ii)Using Lemma 2.2 in [1] with $\psi \in \mathcal{C}^\infty(\mathbb{R}^d)$, $\psi(x) = |x|^2$ outside $B_1(0)$, we get the thesis.

□

In what follow, we state a stability estimate for minimizers of (1.9) when they are unique.

Lemma 1.10. *Assume* (1.1) *and fix* $(t_0, m_0) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$. *Let* u *be the associated multiplier to the unique minimizer* (m, α) *of* $\mathcal{U}(t_0, m_0)$. *If* (t_0^n, m_0^n) *converges to* (t_0, m_0) *and if the associated multiplier to a minimizer* (m^n, α^n) *for* $\mathcal{U}(t_0^n, m_0^n)$ *is* u^n, *then* u^n *converges to* u, Du^n *to* Du *and* D^2u^n *to* D^2u *in* $\mathcal{C}^{\delta/2,\delta}$. *In addition, if,* $\forall n$, $t_0^n = t_0$, *the convergence of* (u^n) *holds in* $\mathcal{C}^{(2+\delta)/2, 2+\delta}$.

The following result will be useful in the proofs of the main results.

Lemma 1.11. *Assume* (1.1) *and let* $t_0 \in [0,T)$ *and* $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$. *Call* (m, α) *a minimizer for* $\mathcal{U}(t_0, m_0)$. *Fix* $\beta \in \mathcal{C}^0([t_0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ *or* $\beta \in L^\infty([t_0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ *with* β *vanishing in a neighborhood of* t_0 *and let* $\rho \in \mathcal{C}^0([t_0,T], (\mathcal{C}^{2+\delta}(\mathbb{R}^d))')$ *be the solution to*

$$\begin{cases} \partial_t \rho - \Delta \rho + \operatorname{div}(m\beta) = 0 \ \textit{in} \ (t_0,T) \times \mathbb{R}^d, \\ \rho(t_0) = 0 \ \textit{in} \ \mathbb{R}^d, \end{cases} \tag{1.16}$$

with the equation understood in the sense of the distributions.
Then

$$\int_{t_0}^T \Big(\int_{\mathbb{R}^d} L_{\alpha,\alpha}(x, \alpha(t,x)) \beta(t,x) \cdot \beta(t,x) m(t,dx) + \langle \frac{\delta F}{\delta m}(\cdot, m(t), \cdot)(\rho(t)), \rho(t) \rangle \Big) dt + \\ + \langle \frac{\delta G}{\delta m}(\cdot, m(T), \cdot)(\rho(T,\cdot)), \rho(T,\cdot) \rangle \geq 0. \tag{1.17}$$

For the proof of the previous two lemmas, see Lemmas 1.5 and 1.6 of [10].

We know that $\mathcal{V}^N$, defined in (1.5), solves the Hamilton-Jacobi-Bellman (HJB) equation [see [5]]

$$\begin{cases} -\partial_t \mathcal{V}^N(t,x) - \sum_{j=1}^N \Delta_{x_j} \mathcal{V}^N(t,x) + \frac{1}{N} H(x_j, N D_{x_j} \mathcal{V}^N(t,x)) = \mathcal{F}(m_x^N) \text{ in } (0,T) \times (\mathbb{R}^d)^N, \\ \mathcal{V}^N(T,x) = \mathcal{G}(m_x^N) \text{ in } (\mathbb{R}^d)^N. \end{cases}$$

1.5 Existence of $\mathcal{U}$

Let us begin with some definitions.

Definition 1.12. *We define*

$$\mathcal{M}(\mathbb{R}^d, \mathbb{R}^d) = \{w : w \text{ is a Borel vector measure with finite mass } |w|\}.$$

Let us call $\mathcal{E}(t_0)$ the set of pairs $(m(t), w(t)) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{M}(\mathbb{R}^d, \mathbb{R}^d)$ such that $t \to m(t)$ is continuous and

$$\int_0^T |w(t)| dt < \infty,$$

and the following equation holds in the sense of distributions

$$\partial_t m - \Delta m + \operatorname{div}(w) = 0 \text{ in } \mathbb{R}^d \times [t_0, T] \text{ and } m(t_0) = m_0.$$

We denote by $\mathcal{E}_2(t_0)$ the subset of $(m(t), w(t)) \in \mathcal{E}(t_0)$ such that $w(t)$ is absolutely continuous with respect to $m(t)$ with a density $\frac{dw(t)}{dm(t)}$ satisfying

$$\int_0^T \int_{\mathbb{R}^d} \left| \frac{dw(t)}{dm(t)}(x) \right|^2 m(dx,t) dt < \infty$$

Then, we define J on $\mathcal{E}(t_0)$ by

$$J(t_0, m_0, \cdot) := \begin{cases} \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \frac{dw(t)}{dm(t)}(x)) m(dx) + \int_{t_0}^T \mathcal{F}(m(t)) dt + \mathcal{G}(m(T)) \text{ if } (m,w) \in \mathcal{E}_2(t_0), \\ +\infty \text{ otherwise.} \end{cases}$$

The next proposition shows that minimizers of the functional J correspond to solution of the MFG system.

Proposition 1.13. *Under our standing assumptions:*

i) For any $t_0 \in [0,T]$ and $m_0 \in \mathcal{P}(\mathbb{R}^d)$ there exists a minimum $(m, w) \in \mathcal{E}_2(t_0)$ of $J(t_0, m_0, \cdot)$, *ii) Let (m, w) be minimum of $J(t_0, m_0, \cdot)$. Then there exists u such that (u, m) is a classical solution to the MFG system*

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = F(x, m(t)) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t m - \Delta m - \operatorname{div}(H_p(x, Du)m) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0) = m_0, \;\; u(T, x) = G(x, m(T)) \text{ in } \mathbb{R}^d, \end{cases}$$

where $F(x, m) = \frac{\delta \mathcal{F}}{\delta m}$ and $G(x, m) = \frac{\delta \mathcal{G}}{\delta m}$, and $w(x, t) = -m(x, t)H_p(x, Du)$. In particular, any minimizer is a classical solution of the above system.

Proof. Let us start proving i).
Let $(m_n, w_n) \in \mathcal{E}_2(t_0)$ be a minimizing sequence. By the definition of J, we have $J(m_n, w_n) \leq C$ and, by the assumption on H in (1.1), there exists $C \geq 0$ such that

$$C^{-1} \, \mathbb{I}_d \leq H_{pp}(x, p) \leq C \, \mathbb{I}_d.$$

This assumption on H implies the following uniform bound

$$\int_{t_0}^T \int_{\mathbb{R}^d} \left| \frac{dw_n(t)}{dm_n(t)} \right|^2 m_n(dx, t)dt \leq C.$$

We can then argue as in [7, Lemma 3.1] to conclude that the sequence (m_n) is uniformly bounded in $\mathcal{C}^{1/2}([0, T], \mathcal{P}(\mathbb{R}^d))$. In particular,

$$\int_{t_0}^T |w_n(t)|dt \leq \left(\int_{t_0}^T \int_{\mathbb{R}^d} \left| \frac{dw_n(t)}{dm_n(t)} \right|^2 m_n(dx, t)dt \right)^{1/2} \left(\int_{t_0}^T \int_{\mathbb{R}^d} m_n(dx, t)dt \right)^{1/2} \leq C.$$

So, up to a subsequence,$w_n \to w$ in $\mathcal{M}((0, T) \times \mathbb{R}^d, \mathbb{R}^d)$ and (m_n) converges in $\mathcal{C}^0([0, T], \mathcal{P}(\mathbb{R}^d))$ to m. Then, we can conclude that the pair (m, w) belongs to $\mathcal{E}_2(t_0)$.
Since J is lower semicontinuous

$$\liminf_{n\to\infty} J(m_n, w_n) = \lim_{n\to\infty} J(m_n, w_n) \leq J(m, w),$$

and we have that the couple (m, w) is a minimum of the functional J.

Let us pass to ii). We define on $\mathcal{E}(t_0)$ the functionals $\Phi(m, w)$ and $\Psi(m)$ such that $J(m, w) = \Phi(m, w) + \Psi(m)$, so

$$\Phi(m, w) := \begin{cases} \int_{t_0}^T \int_{\mathbb{R}^d} L(x, \frac{w(t)}{m(t)})m(dx, t)dt \text{ if } (m, w) \in \mathcal{E}_2(t_0), \\ +\infty \text{ otherwise}, \end{cases}$$

and

$$\Psi(m) := \int_{t_0}^{T} \mathcal{F}(m(t))dt + \mathcal{G}(m(T)).$$

Let $(\overline{m}, \overline{w})$ be a minimum of J. Recall from i) that, since it is a minimum, $\overline{m} \in \mathcal{C}^{1/2}([0,T], \mathcal{P}(\mathbb{R}^d))$.
Let m_λ and w_λ be the convex combination of m and $\overline{m}$ and w and $\overline{w}$ respectively. So we have $m_\lambda := (1-\lambda)\overline{m} + \lambda m$ and $w_\lambda := (1-\lambda)\overline{w} + \lambda w$ with $\lambda \in (0,1)$. By minimality of $(\overline{m}, \overline{w})$

$$\Phi(m_\lambda, w_\lambda) + \Psi(m_\lambda) \geq \Phi(\overline{m}, \overline{w}) + \Psi(\overline{m}),$$

and so

$$\Phi(m_\lambda, w_\lambda) - \Phi(\overline{m}, \overline{w}) \geq \Psi(\overline{m}) - \Psi(m_\lambda).$$

Thus, by the convexity of Φ

$$\Phi(m_\lambda, w_\lambda) \leq \lambda\Phi(m, w) + (1-\lambda)\Phi(\overline{m}, \overline{w}),$$

and so

$$\lambda(\Phi(m, w) - \Phi(\overline{m}, \overline{w})) \geq \Phi(m_\lambda, w_\lambda) - \Phi(\overline{m}, \overline{w}) \geq \Psi(\overline{m}) - \Psi(m_\lambda).$$

Thus, by the regularity assumptions on $\mathcal{F}$ and $\mathcal{G}$, we get

$$\begin{aligned}&\lambda(\Phi(m, w) - \Phi(\overline{m}, \overline{w}))\\ &\geq \lambda\left(-\int_{t_0}^{T}\int_{\mathbb{R}^d} \mathcal{F}(x, \overline{m}(t))(m-\overline{m})(dx,t) - \int_{\mathbb{R}^d} \mathcal{G}(x, \overline{m}(T))(m-\overline{m})(dx,T)\right) + o(\lambda).\end{aligned} \tag{1.18}$$

Then dividing (1.18) by λ and letting λ tends to 0, we obtain, for any $(m, w) \in \mathcal{E}(t_0)$,

$$\begin{aligned}&-\int_{t_0}^{T}\int_{\mathbb{R}^d} \mathcal{F}(x, \overline{m}(t))(m(dx,t) - \overline{m}(dx,t)) - \int_{\mathbb{R}^d} \mathcal{G}(x, \overline{m}(T))(m(dx,T) - \overline{m}(dx,T))\\ &\leq \Phi(m, w) - \Phi(\overline{m}, \overline{w}).\end{aligned} \tag{1.19}$$

Let us define

$$\widetilde{J}(m, w) = \Phi(m, w) + \int_0^T \int_{\mathbb{R}^d} \mathcal{F}(x, \overline{m}(t)m(dx,t)dt + \int_{\mathbb{R}^d} \mathcal{G}(x, \overline{m}(T))m(dx,T).$$

Note that this functional is convex.
By (1.19), the pair $(\overline{m}, \overline{w})$ is a minimizer of $\widetilde{J}$ on $\mathcal{E}(t_0)$.

Note also that the problem of minimizing $\widetilde{J}$ on $\mathcal{E}(t_0)$ is the dual problem (in the sense of the Fenchel-Rockafellar duality theorem B.3 [see [7] for the details]) of the problem

$$\inf_{u\in\mathcal{C}^2}\{-\int_{\mathbb{R}^d} m_0(x)u(x,0)dx : -\partial_t u - \Delta u + H(x,Du) \leq \mathcal{F}(x,\overline{m}(t)) \text{ and } u(x,T) \leq \mathcal{G}(x,\overline{m}(x,T))\}. \tag{1.20}$$

By comparison, there is an obvious minimum to this problem which is the solution $\overline{u}$ to

$$\begin{cases} -\partial_t\overline{u} - \Delta\overline{u} + H(x,D\overline{u}) = \mathcal{F}(x,\overline{m}(t)) \text{ in } \mathbb{R}^d\times(0,T), \\ u(x,T) = \mathcal{G}(x,\overline{m}(x,T)) \text{ in } \mathbb{R}^d. \end{cases}$$

This solution is $\mathcal{C}^{2,\alpha}$ because $\overline{m}\in\mathcal{C}^{1/2}([0,T),\mathcal{P}(\mathbb{R}^d))$. By the Fenchel-Rockafellar duality theorem, we have that

$$0 = \widetilde{J}(\overline{m},\overline{w}) - \int_{\mathbb{R}^d} m_0(x)\overline{u}(x,0)dx.$$

Using the definition of $\widetilde{J}$, the equation for $\overline{u}$ and the equation for $(\overline{m},\overline{w})$, we have that

$$\begin{aligned} 0 &= \int_{t_0}^T\int_{\mathbb{R}^d}\left(L\left(x,\frac{dw(t)}{dm(t)}\right)+\mathcal{F}(x,\overline{m}(t))\right)\overline{m}(dx,t)dt + \int_{\mathbb{R}^d}\mathcal{G}(x,\overline{m}(T))\overline{m}(dx,T) \\ &- \int_{\mathbb{R}^d} m_0(x)\overline{u}(x,0)dx = \int_{t_0}^T\int_{\mathbb{R}^d}\left(L\left(x,\frac{dw(t)}{dm(t)}\right)+(-\partial_t\overline{u}-\Delta\overline{u}+H(x,D\overline{u}))\right)\overline{m}(dx,t)dt \\ &+ \int_{\mathbb{R}^d}\mathcal{G}(x,\overline{m}(T))\overline{m}(dx,T) - \int_{\mathbb{R}^d} m_0(x)\overline{u}(x,0)dx \\ &= \int_{t_0}^T\int_{\mathbb{R}^d}\left(L\left(x,\frac{dw(t)}{dm(t)}\right)+H(x,D\overline{u})+Du\cdot\left(\frac{dw(t)}{dm(t)}\right)\right)\overline{m}(dx,t)dt. \end{aligned}$$

Recalling that L is the convex conjugate of H which is uniformly convex, we find

$$\frac{d\overline{w}(t)}{d\overline{m}(t)} = -H_p(x,D\overline{u}) \quad \overline{m}-\text{a.e.}.$$

This means that $\overline{m}$ solves the Kolmogorov equation

$$\partial_t\overline{m} - \Delta\overline{m} - \operatorname{div}(\overline{m}H_p(x,D\overline{u})) = 0 \quad \text{and} \quad \overline{m}(\cdot,t_0) = m_0$$

which has a regular drift: thus $\overline{m}$ is of class $\mathcal{C}^{2,\alpha}$ by Schauder theory. Therefore $\overline{w}$ is also smooth and the proof of ii) is complete.

□

Chapter 2

The convergence of $\mathcal{V}^N$ to $\mathcal{U}$

This chapter is devoted to an algebraic rate of convergence of the value function $\mathcal{V}^N$ of N-particle control problems to the value function $\mathcal{U}$ of the corresponding MFC problem.
With $\mathcal{V}^N$ defined by (1.5) and $\mathcal{U}$ by (1.8), we have the following result.

Theorem 2.1. *Assume* (1.1)*. There exists* $\beta = \beta(d) \in (0,1)$ *and a constant* $C > 0$ *depending on the smoothness of the data such that,* $\forall (t,x) \in [0,T] \times (\mathbb{R}^d)^N$,

$$|\mathcal{V}^N(t,x) - \mathcal{U}(t, m_x^N)| \leq C \frac{1}{N^\beta}(1 + M_2(m_x^N)). \tag{2.1}$$

The proof of Theorem 2.1 requires several steps: we first obtain uniform in N regularity estimates on $\mathcal{V}^N$, then we show how to bound from above $\mathcal{V}^N$ by $\mathcal{U}$ plus an error term and, finally, we prove the converse estimate.

Here we present a result on the regularity of $\mathcal{V}^N$ and of $\mathcal{U}$.

Lemma 2.2. *Assume* (1.1)*. There exists a constant* $C > 0$*, which depends on the data, such that*

$$||\mathcal{V}^N||_\infty + N \sup_{j=1,\dots,N} ||D_{x_j} \mathcal{V}^N||_\infty + ||\partial_t \mathcal{V}^N||_\infty \leq C, \tag{2.2}$$

and, $\forall (t', m'), \ (t'', m'') \in [0,T] \times \mathcal{P}_1(\mathbb{R}^d)$,

$$|\mathcal{U}(t', m') - \mathcal{U}(t'', m'')| \leq C(|t' - t''| + d_1(m', m'')). \tag{2.3}$$

Proof. The bound on $\mathcal{V}^N$ follows by the assumptions on the data.

Call $w^j = D_{x_j}\mathcal{V}^N$. Then w^j satisfies

$$\begin{cases} -\partial_t w^j(t,x) - \sum_{k=1}^N \Delta_{x_k} w^j(t,x) + \frac{1}{N} D_x H(x^j, NDw^j(t,x)) \\ + \sum_{k=1}^N H_p(x_k, ND_{x_k}\mathcal{V}^N(t,x)) \cdot D_{x_k} w^j(t,x) = \frac{1}{N} D_m \mathcal{F}(m_X^N, x_j) \text{ in } (0,T)\times(\mathbb{R}^d)^N, \\ w^j(T,x) = \frac{1}{N} D_m \mathcal{G}(m_X^N, x_j) \text{ in } (\mathbb{R}^d)^N. \end{cases} \tag{2.4}$$

Using the maximum principle we can conclude that $N||D_{x_j}\mathcal{V}^N||_\infty$ are uniformly bounded in N and j.
In the same way, $w^t = \partial_t \mathcal{V}^N$ satisfies

$$\begin{cases} -\partial_t w^t(t,x) - \sum_{k=1}^N \Delta_{x_k} w^t(t,x) + \\ + \sum_{k=1}^N H_p(x_k, ND_{x_k}\mathcal{V}^N(t,x)) \cdot D_{x_k} w^t(t,x) = 0 \text{ in } (0,T)\times(\mathbb{R}^d)^N, \\ w^t(T,x) = -\frac{1}{N}\sum_{k=1}^N tr\left[D^2_{ym}\mathcal{G}(m_X^N, x_k) + \frac{1}{N} D^2_{mm}\mathcal{G}(m_X^N, x_k, x_k)\right] \\ \qquad + \frac{1}{N}\sum_{k=1}^N H(x_k, D_m\mathcal{G}(m_X^N, x_k) - \mathcal{F}(m_X^N) \text{ in } (\mathbb{R}^d)^N, \end{cases}$$

and the uniform bound on $||\partial_t \mathcal{V}^N||_\infty$ follows again from the maximum principle.

For the second part, fix $(t', m') \in [0,T]\times\mathcal{P}_1(\mathbb{R}^d)$. So there exist at least (see Proposition 1.13) a pair (m,α) optimal in the definition of $\mathcal{U}(t',m')$ and a multiplier $u : [t_0,T]\times\mathbb{R}^d \to \mathbb{R}$ and such that (u,m) solves the system

$$\begin{cases} -\partial_t u - \Delta u + H(x, Du) = \frac{\delta\mathcal{F}}{\delta m}(m(t),x) \text{ in } (t_0,T)\times\mathbb{R}^d, \\ \partial_t m - \Delta m - \mathrm{div}(H_p(x,Du)m) = 0 \text{ in } (t_0,T)\times\mathbb{R}^d, \\ m(t') = m' \text{ and } u(T,x) = \frac{\delta\mathcal{G}}{\delta m}(m(T),x) \text{ in } \mathbb{R}^d. \end{cases} \tag{2.5}$$

Arguing as before for $\mathcal{V}^N$, there exists a constant $C>0$ such that

$$||Du||_\infty \le C$$

and, since $\alpha = -H_p(x,Du)$, $||\alpha||_\infty \le C$ and so we have

$$||D\alpha||_\infty = ||D[H_p(\cdot, Du(\cdot,\cdot))]||_\infty \le C.$$

Let $m'' \in \mathcal{P}_1(\mathbb{R}^d)$ and μ be the solution to

$$\partial_t \mu - \Delta \mu + \operatorname{div}(\mu\alpha) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d \text{ with } \mu(t') = m''.$$

Arguing as in Theorem A.1, one can prove that

$$\sup_{t \in [t_0,T]} d_1(\mu(t), m(t)) \leq C d_1(m', m'').$$

Thus, for some constant C depending on the data

$$\begin{aligned}
\mathcal{U}(t', m'') &\leq \int_{t'}^{T} \left(\int_{\mathbb{R}^d} L(x, \alpha(t,x)) \mu(t, dx) + \mathcal{F}(\mu(t)) \right) dt + \mathcal{G}(\mu(T)) \\
&\leq \int_{t'}^{T} \left(\int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t, dx) + \mathcal{F}(m(t)) \right) dt + \\
&+ \mathcal{G}(m(T)) + C \sup_{t \in [t_0,T]} d_1(\mu(t), m(t)) \\
&\leq \mathcal{U}(t', m') + C d_1(m', m'').
\end{aligned}$$

We can conclude that

$$|\mathcal{U}(t', m') - \mathcal{U}(t', m'')| \leq C d_1(m', m'').$$

Finally, fix $t'' > t'$, and we choose (m, α) optimal in the definition of $\mathcal{U}(t'', m'')$. By dynamic programming principle, we have

$$\mathcal{U}(t'', m'') = \mathcal{U}(t', m') + \int_{t''}^{t'} (\int_{\mathbb{R}^d} L(x, \alpha(t,x) m(t, dx) + \mathcal{F}(m(t))) dt),$$

and so

$$\begin{aligned}
|\mathcal{U}(t'', m'') - \mathcal{U}(t', m'')| &\leq | \int_{t''}^{t'} (\int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t, dx) + \mathcal{F}(m(t)) dt)| \\
&+ |\mathcal{U}(t'', m') - \mathcal{U}(t'', m'')| \leq C(t'' - t') + C d_1(m', m'').
\end{aligned} \tag{2.6}$$

This completes the proof.

□

The next lemma is about a key estimate on $\mathcal{V}^N$.

Lemma 2.3. *Assume (1.1). There exists a constant C independent of N, such that, for any $N \geq 1$ $\xi = (\xi^i) \in (\mathbb{R}^d)^N$ and $\xi^0 \in \mathbb{R}$,*

$$\begin{aligned}
\sum_{i,j=1}^{N} D^2_{x_j x_j} \mathcal{V}^N(t,x) \xi^i \cdot \xi^j + 2 \sum_{i=1}^{N} D^2_{x_j t} \mathcal{V}^N(t,x) \cdot \xi^i \xi^0 + D^2_{tt} \mathcal{V}^N(t,x) (\xi^0)^2 \\
\leq \frac{C}{N} \sum_{i=1}^{N} |\xi^i|^2 + C(\xi^0)^2.
\end{aligned} \tag{2.7}$$

Proof. For $1 \le i, j, k \le N$, let

$$\begin{cases} w^i = D_{x_i}\mathcal{V}^N \cdot \xi^i, \quad w^{i,j} = D^2_{x_i x_j}\mathcal{V}^N \xi \cdot \xi^j, \quad w^{0,i} = w^{i,0} = \partial_t D_{x_i}\mathcal{V}^N \cdot \xi^0 \xi^i, \\ w^{0,0} = \partial_{tt}\mathcal{V}^N (\xi^o)^2, \quad w^0 = \partial_t \mathcal{V}^N \xi^0, \quad \widetilde{w} = \sum_{i,j=1}^N w^{i,j} \quad \text{and} \quad \sigma_k = \sum_{i=0}^N D_{x_k} w^i. \end{cases} \tag{2.8}$$

Since $\mathcal{V}^N(t,x)$ solves

$$\begin{cases} -\partial_t \mathcal{V}^N(t,x) - \sum_{j=1}^N \Delta_{x_j}\mathcal{V}^N(t,x) + \frac{1}{N} H(x_j, N D_{x_j}\mathcal{V}^N(t,x)) = \mathcal{F}(m_x^N) \text{ in } (0,T)\times(\mathbb{R}^d)^N, \\ \mathcal{V}^N(T,x) = \mathcal{G}(m_x^N) \;\text{ in }\; (\mathbb{R}^d)^N, \end{cases}$$

a simple computation gives

$$\begin{aligned} &-\partial_t \widetilde{w} - \sum_{k=1}^N \Delta_{x_k}\widetilde{w} + \sum_{k=1}^n D_{x_k}\widetilde{w}\cdot H_p(x_k, N D_{x_k}\mathcal{V}^N(t,x)) \\ &= -N\sum_{k=1}^N H_{pp}(x_k, N D_{x_k}\mathcal{V}^N(t,x))\sigma_k\cdot\sigma_k \\ &- 2\sum_{k=1}^N H_{xp}(x_k, N D_{x_k}\mathcal{V}^N(t,x))\xi^k\cdot\sigma_k - \frac{1}{N}\sum_{i=1}^N H_{xx}(x_i, N D_{x_i}\mathcal{V}^N(t,x)\xi^i\cdot\xi^i \\ &+ \frac{1}{N^2}\sum_{i,j=1}^N D^2_{mm}\mathcal{F}(m_X^N, x_i, x_j)\xi^i\cdot\xi^j + \frac{1}{N}\sum_{i=1}^N D^2_{ym}\mathcal{F}(m_X^N, x_i)\xi^i\cdot\xi^i. \end{aligned} \tag{2.9}$$

Denote by $\star$ the right-hand-side of the equality above.
Recalling that H is strictly convex in the second argument and that, by Lemma 2.2, $N\partial_{x_k}\mathcal{V}^N$ is bounded, we have, for all $1 \le k \le N$,

$$-N H_{pp}\sigma_k\cdot\sigma_k - 2H_{xp}\xi^k\cdot\sigma_k \le \frac{C}{N}|\xi^k|^2.$$

We can use again the Lipschitz bounds on $\mathcal{V}^N$ and the hypothesis on H:
for any $R > 0$, there exists $C_R > 0$ such that

$$|H_{xx}(x,p)| + |H_{xp}(x,p)| \le C_R \quad \forall (x,p)\in\mathbb{R}^d\times B_R,$$

to deduce that

$$\star \le \frac{C}{N}\sum_k |\xi^k|^2. \tag{2.10}$$

Next, fix (t_0, x_0) and consider the weak solution m^N to

$$\begin{cases} \partial_t m^N(t,x) - \sum_{k=1}^N \Delta_{x_k} m^N(t,x) \\ -\sum_{k=1}^N \operatorname{div}(H_p(x_k, N D_{x_k}\mathcal{V}^N(t,x))m^N) = 0 \text{ in } (t_0,T)\times(\mathbb{R}^d)^N, \\ m^N(t_0,\cdot) = \delta_{x_0} \text{ in } (\mathbb{R}^d)^N. \end{cases} \tag{2.11}$$

After summing equation (2.9) multiplied by m^N and equation (2.11) multiplied by $\widetilde{w}$ and integrating in space and time, using (2.10) and that $\int_{\mathbb{R}} m^N(t,x)dx = 1$, we get

$$\widetilde{w}(t_0, x_0) \leq \sup_x ||\widetilde{w}(T,x)||_\infty + \frac{C}{N}\sum_{k=1}^N |\xi^k|^2.$$

In order to bound the right-hand side of the inequality above, we first note that, by the equation satisfied by $\mathcal{V}^N$, we have

$$\partial_t \mathcal{V}^N(T,x) = -\sum_{k=1}^N \Delta_{x_k}\mathcal{G}^N(x) + \frac{1}{N}\sum_{k=1}^N H(x_k, ND_{x_k}\mathcal{G}^N(x)) - \mathcal{F}(x),$$

where $\mathcal{F}^N(x) := \mathcal{F}(m_X^N)$ and $\mathcal{G}^N(x) := \mathcal{G}(m_X^N)$, and, similarly,

$$\partial_{tt}^2 \mathcal{V}^N(T,x) = -\sum_{k=1}^N \Delta_{x_k}\partial_t\mathcal{V}^N(T,x) + \sum_{k=1}^N H_p(x_k, ND_{x_k}\mathcal{G}^N(x)) \cdot D_{x_k}\partial_t\mathcal{V}^N(T,x).$$

Using the following proposition to express the derivatives of $\mathcal{F}^N$ and $\mathcal{G}^N$ in function of the derivatives of $\mathcal{F}$ and $\mathcal{G}$,

Proposition 2.4. *If $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is continuously diffentiable, then its empirical projection u^N, that is the function $u^N = u\left(\frac{1}{N}\sum_{i=1}^N \delta_{X_i}\right)$, is differentiable on $(\mathbb{R}^d)^N$ and, for all $i \in \{1,\dots,N\}$,*

$$\partial_{x_i} u^N(x_1,\dots,x_N) = \frac{1}{N} D_m u\left(\frac{1}{N}\sum_{j=1}^N \delta_{x_j}\right)(x_i),$$

it can be show that, under our standing assumptions on $\mathcal{F}$ and $\mathcal{G}$, for some $C \geq 0$,

$$\sup_x ||\widetilde{w}(T,x)||_\infty \leq \frac{C}{N}\sum_{i=1}^N |\xi^i|^2 + C(\xi^0)^2$$

□

The second step in the proof of Theorem 2.1 is an upper bound of $\mathcal{V}^N$ in terms of $\mathcal{U}$. Our strategy will be to first compare $\mathcal{U}$ to $\widehat{\mathcal{V}}^N$, where

$$\widehat{\mathcal{V}}^N(t,m) := \int_{(\mathbb{R}^d)^N} \mathcal{V}^N(t,x)\prod_{j=1}^N m(dx_j). \tag{2.12}$$

Lemma 2.5. *Let $\widehat{\mathcal{V}}^N$ be given by (2.12). Then $\widehat{\mathcal{V}}^N$ is smooth and satisfies the inequality*

$$\begin{cases} -\partial_t \widehat{\mathcal{V}}^N(t,m) - \int_{\mathbb{R}^d} \mathrm{div}(D_m \widehat{\mathcal{V}}^N(t,m,y))m(dy), \\ + \int_{\mathbb{R}^d} H(y, D_m \widehat{\mathcal{V}}^N(t,m,y))m(dy) \le \widehat{\mathcal{F}}^N(m) \text{ in } (0,T)\times \mathcal{P}_1(\mathbb{R}^d), \\ \widehat{\mathcal{V}}^N(T,m) = \widehat{G}^N(m) \text{ in } \mathcal{P}_1(\mathbb{R}^d), \end{cases} \tag{2.13}$$

where

$$\widehat{\mathcal{F}}^N(m) := \int_{(\mathbb{R}^d)^N} \mathcal{F}(m_X^N) \prod_{j=1}^N m(dx_j) \quad \text{and} \quad \widehat{\mathcal{G}}^N(m) := \int_{(\mathbb{R}^d)^N} \mathcal{G}(m_X^N) \prod_{j=1}^N m(dx_j)$$

For the proof see [8, Proposition 3.1].

Next we prove the easier inequality in Theorem 2.1.

Proposition 2.6. *There exist constants C depending on the data and β depending only on d such that, for all $(t,x_0) \in [0,T]\times(\mathbb{R}^d)^N$*

$$\mathcal{V}^N(t_0, m_{x_0}^N) \le \mathcal{U}(t, m_{x_0}^N) + \frac{C}{N^\beta}(1 + M_2^{1/2}(m_{x_0}^N)). \tag{2.14}$$

Proof. Using the following theorem (for the proof see [15, Theorem 1])

Theorem 2.7. *Let $m \in \mathcal{P}(\mathbb{R}^d$ and let $p > 0$. Assume that for some $q > p$, $M_q(m) \le \infty$. There exists a constant $C = C(p,q,d)$ such that, for all $N \ge 1$,*

$$\int_{\mathbb{R}^d} d_p(m, m_X^N) \le C M_q^{p/q}(m) \begin{cases} N^{-1/2} + N^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \neq 2p \\ N^{-1/2} log(1+N) + N^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \neq 2p \\ N^{-p/d} + N^{-(q-p)/q} & \text{if } p \in (0, d/2) \text{ and } q \neq d/(d-p), \end{cases} \tag{2.15}$$

we have that, there exist β and C constants such that

$$\int_{(\mathbb{R}^d)^N} d_1(m_X^N, m) \prod_{i=1}^N m(dx_i) \le \frac{C}{N^\beta} M_2^{1/2}(m).$$

Let $(t_0, m_0) \in [0,T]\times\mathcal{P}_2(\mathbb{R}^d)$ and the infimum in $\mathcal{U}(t_0,m_0)$ is achieved by α^*. Applying Lemma 2.5 and Itô's formula, we have that

$$\widehat{\mathcal{V}}^N(t_0, m_0) \le \inf_{\alpha\in\mathcal{A}} \left\{ \int_{t_0}^T \left(\int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t,dx) + \widehat{\mathcal{F}}^N(m(t)) \right) dt + \widehat{\mathcal{G}}(m(T), x) \right\},$$

and, using the optimality of α^*,

$$\widehat{\mathcal{V}}^N(t_0, m_0) \le \int_{t_0}^{T} \left(\int_{\mathbb{R}^d} L(x, \alpha^*(t,x)) + \widehat{\mathcal{F}}^N(m(t)) \right) dt + \widehat{\mathcal{G}}(m(T), x). \tag{2.16}$$

Using the Lipschitz continuity of $\mathcal{F}$ with respect to d_1 we get

$$\begin{aligned}\widehat{\mathcal{F}}^N(m(t)) &\le \mathcal{F}(m(t)) + C \int_{(\mathbb{R}^d)^N} d_1(m_X^N, m(t)) \prod_{j=1}^{N} m(t, dx_j) \\ &\le \mathcal{F}(m(t)) + \frac{C}{N^\beta} M_2^{1/2}(m(t)) \le \mathcal{F}(m(t)) + \frac{C}{N^\beta}(1 + M_2^{1/2}(m_0)),\end{aligned}$$

where, in the last inequality, we used the following formula in Lemma 1.9

$$\sup_{t \in [t_0, T]} \int_{\mathbb{R}^d} |x|^2 m(t, dx) \le (1 + CT) \int_{\mathbb{R}^d} |x|^2 m_0(t, dx) + CT. \tag{2.17}$$

Similarly,

$$\widehat{\mathcal{G}}^N(m(T)) \le \mathcal{G}(m(t)) + \frac{C}{N^\beta}(1 + M_2^{1/2}(m_0)).$$

Since α^* is optimal for $\mathcal{U}(t_0, m_0)$, putting together the estimates (2.16) and the ones of $\widehat{\mathcal{F}}^N$ and $\widehat{\mathcal{G}}^N$ we obtain

$$\begin{aligned}\widehat{\mathcal{V}}^N(t_0, m_0) &\le \mathbb{E}\left[\int_{t_0}^{T} (L(X_t, \alpha^*(t)) + \mathcal{F}(\mathcal{L}(X_t)))dt + \mathcal{G}(\mathcal{L}(X_T)) \right] + \frac{C}{N^\beta}(1 + M_2^{1/2}(m_0)) \\ &\le \mathcal{U}(t_0, m_0) + \frac{C}{N^\beta}(1 + M_2^{1/2}(m_0)).\end{aligned} \tag{2.18}$$

Fix now $x_0 \in (\mathbb{R}^d)^N$. Then the Lipschitz estimate on $\mathcal{V}^N$ and the same argument as above yield

$$\left| \mathcal{V}^N(t_0, x_0) - \widehat{\mathcal{V}}^N(t_0, m_{x_0}^N) \right| \le \frac{C}{N^\beta}(1 + M_2^{1/2}(m_{x_0}^N)). \tag{2.19}$$

Putting the equations (2.18) and (2.19) together

$$\begin{aligned}\left| \mathcal{U}(t_0, m_0) - \mathcal{V}^N(t_0, m_0) \right| &\le \left| \mathcal{U}(t_0, m_0) - \widehat{\mathcal{V}}^N(t_0, x_0) \right| + \left| \mathcal{V}^N(t_0, m_0) - \widehat{\mathcal{V}}^N(t_0, m_{x_0}^N) \right| \\ &\le \frac{C}{N^\beta}(1 + M_2^{1/2}(m_{x_0}^N)).\end{aligned}$$

□

The difficult step of the proof is to show the opposite inequality.

We continue estimating in the next lemma the error related to the penalization. For the proof see Lemma 3.9 of [5].

Lemma 2.8. *Assume* (1.1). *There exists $C \geq 0$ such that, for any $i \in \{1, \dots, N\}$ and for $\theta, \lambda \in (0,1)$*

$$\frac{1}{N}\sum_{i=1}^{N} |x_0^i - y_0^i|^2 + |s_0 - t_0|^2 \leq C\theta^2 \quad \textit{and} \quad \frac{1}{N}\sum_{i=1}^{N} |y_0^i|^2 \leq \frac{C}{\lambda}.$$

Lemma 2.9. *For each $\delta > 0$ there exist a constant C depending only on the data, and a partition $(C_j)_{j\in\{1,\dots,J\}}$ of $\{1, \dots, N\}$ and, for $j = 1, \dots, J$, $\overline{\alpha}_j \in \mathbb{R}^d$ such that $J \leq C\delta^{-d}$ and, for all $k \in C_j$,*

$$|H(x_0^k, ND_{x_k}\mathcal{V}^N(t_0, x_0)) + \overline{\alpha}_j \cdot (ND_{x_k}\mathcal{V}^N(t_0, x_0)) + L(x_0^k, \overline{\alpha}_j)| \leq C\delta. \tag{2.20}$$

Proof. Let $\widehat{\alpha}_k(t, x) = -H_p(x_k, ND_{x_k}\mathcal{V}^N(t, x))$ be the optimal feedback for particle k. It follows from the estimate on $D_{x_j}\mathcal{V}^N$ that the optimal feedback of the problem remains uniformly bounded, and so there exists R depending only on the data such that $|\widehat{\alpha}_k(t, x)| \leq R$.

Given $\delta > 0$, we can find a δ-covering of $B_R \subset \mathbb{R}^d$ consisting of $J \leq C\delta^{-d}$ balls of radius δ centered at $(\overline{\alpha}_j)_{j\in\{1,\dots,J\}} \in \mathbb{R}^d$.

We select the partition $(C_j)_{j\in\{1,\dots,J\}}$ so that, for each $k \in C_j$, $|\widehat{\alpha}_k(t, x) - \overline{\alpha}_j| \leq \delta$. Using that L is the Legendre transform of H, we get

$$\begin{aligned}
&|H(x_0^k, ND_{x_k}\mathcal{V}^N(t_0, x_0)) + \overline{\alpha}_j \cdot (ND_{x_k}\mathcal{V}^N(t_0, x_0)) + L(x_0^k, \overline{\alpha}_j)| \\
&= |(\alpha_j - \widehat{\alpha}_k(t, x)) \cdot (ND_{x_k}\mathcal{V}^N(t_0, x_0)) + L(x_k^0, \overline{\alpha}_j) - L(x_0^k, \widehat{\alpha}_k(t, x))| \\
&\leq (ND_{x_k}\mathcal{V}^N(t_0, x_0)) + ||D_\alpha L||_{L^\infty(\mathbb{R}^d\times B_R)})|\widehat{\alpha}_k(t, x) - \overline{\alpha}_j| \leq C\delta.
\end{aligned}$$

where in the last inequality we used the estimate in (1.2).

□

Fix $j \in \{1, \dots, J\}$, set $\alpha^k = \overline{\alpha}^j$ if $k \in C^j$, let

$$X_{t_0+\tau}^k = x_0^k + \tau\alpha^k + \sqrt{2}B_\tau^k \quad \text{and} \quad Y_{s_0+\tau}^k = y_0^k + \tau\alpha^k + \sqrt{2}B_\tau^k,$$

$$m^j_{X_{t_0+\tau}} = \frac{1}{n^j}\sum_{k\in C^j} \delta_{Y^k_{s_0+\tau}} \quad \text{and} \quad m^j_{Y_{s_0+\tau}} = \frac{1}{n^j}\sum_{k\in C^j} \delta_{X^k_{t_0+\tau}},$$

consider the solution m^j to

$$\partial_t m^j - \Delta m^j + \overline{\alpha}_j \cdot Dm^j = 0 \quad \text{in} \quad (s_0, T) \times \mathbb{R}^d \quad \text{and} \quad m^j(s_0, \cdot) = m^j_{y_0} \quad \text{in} \quad \mathbb{R}^d$$

and, finally, set $m(s) = \frac{1}{N}\sum_{j=1}^N n^j m^j(s)$.

We state next the concentration inequality we need for the proof.

Lemma 2.10. *There exist positive constants $\beta \in (0, 1/2)$ depending on d and C, which depends only on $\sup_j |\overline{\alpha}_j|$, d and T , such that, for all $h \geq 0$,*

$$\mathbb{E}\left[d_1(m^j(s_0+h), m^j_{Y_{s_0+h}})\right] \leq C(1 + M_2^{1/2}(m^j(s_0)))\frac{h^\beta}{(n^j)^\beta}, \tag{2.21}$$

and

$$\mathbb{E}\left[d_1(m^j(s_0+h), m^j_{X_{t_0+h}})\right] \leq \frac{1}{n^j}\sum_{k \in C_j} |x_0^k - y_0^k| + C(1 + M_2^{1/2}(m^j(s_0)))\frac{h^\beta}{(n^j)^\beta}, \tag{2.22}$$

and, as a consequence,

$$\mathbb{E}\left[d_1(m(s_0+h), m^N_{Y_{s_0+h}})\right] \leq C\delta^{-d\beta}(1 + \lambda^{-1/2})\frac{h^\beta}{(n^j)^\beta}, \tag{2.23}$$

and

$$\mathbb{E}\left[d_1(m(s_0+h), m^N_{X_{t_0+h}})\right] \leq C\theta + C\delta^{-d\beta}(1 + \lambda^{-1/2})\frac{h^\beta}{(n^j)^\beta}. \tag{2.24}$$

Proof. We define

$$L = \{\phi : \mathbb{R}^d \to \mathbb{R} \mid \phi \text{ is } 1 - \text{Lipschitz}\},$$

$$L_R = \{\phi : B_R \subset \mathbb{R}^d \to [-R, R] \mid \phi \text{ is } 1 - \text{Lipschitz}\},$$

and, for any $\phi \in L_R$, the extension $\widetilde{\phi} : \mathbb{R}^d \to [-R, R]$ given by

$$\widetilde{\phi}(x) = \begin{cases} \phi(x) \text{ if } |x| \leq R \\ \frac{2R - |x|}{R}\phi(\frac{R}{|x|}x) \text{ if } R < |x| < 2R \\ 0 \text{ if } |x| \geq 2R. \end{cases}$$

Without loss of generality, we can assume $s_0 = 0$.
Fix $R > 0$. Note that any $\psi \in L$ with $\psi(0) = 0$ can be rewritten as $\psi = \widetilde{\phi} + \varphi$

with $\phi \in L_R$ and $|\varphi| \le |x|\mathbb{I}_{B_R^C}$.
For any $h \in (0,1]$, we have

$$\begin{aligned}
\mathbb{E}[d_1(m^j(h), m^j_{Y_h})] &= \mathbb{E}[\sup_{\phi \in L} \int_{\mathbb{R}^d} \phi(m^j(h) - m^j_{Y_h})] \\
&\le \mathbb{E}[\sup_{\phi \in L_R} \int_{\mathbb{R}^d} \widetilde{\phi}(m^j(h) - m^j_{Y_h})] + \int_{\mathbb{R}^d} |x|\mathbb{I}_{B_R^C} m^j(h) + \mathbb{E}[\int_{\mathbb{R}^d} |x|\mathbb{I}_{B_R^C} m^j_{Y_h}] \\
&\le \mathbb{E}[\sup_{\phi \in L_R} \int_{\mathbb{R}^d} \widetilde{\phi}(m^j(h) - m^j_{Y_h})] + \frac{M_2(m^j(h))}{R} + \frac{M_2(m^j_{Y_h})}{R} \\
&\le \mathbb{E}[\sup_{\phi \in L_R} \int_{\mathbb{R}^d} \widetilde{\phi}(m^j(h) - m^j_{Y_h})] + C\frac{(1 + M_2(m^j(0))}{R},
\end{aligned}$$

where we used theorem 1.1 in the first equality, the characterization of $\phi \in L$ written above, the definition of $M_2(m)$ and the inequality in (1.13).
Finally, using the following lemma

Lemma 2.11. *There exists a constant C such that, for any $j \in \{1, \dots, J\}$ and $R > 0$,*

$$\mathbb{E}[\sup_{\phi \in L_R} \int_{\mathbb{R}^d} \widetilde{\phi}(m^j(h) - m^j_{Y_h})] \le C(1 + R^{\frac{d}{d+2}})(n^j)^{\frac{-1}{d+2}} h^{\frac{1}{d+2}},$$

we get

$$\begin{aligned}
\mathbb{E}[d_1(m^j(h), m^j_{Y_h})] &\le C(1 + R^{\frac{d}{d+2}})(n^j)^{\frac{-1}{d+2}} h^{\frac{1}{d+2}} + C\frac{(1 + M_2(m^j(0))}{R} \\
&\le C(1 + R)(n^j)^{\frac{-1}{d+2}} h^{\frac{1}{d+2}}.
\end{aligned}$$

Taking $R = (n^j)^{\frac{1}{2d+4}} h^{-\frac{1}{2d+4}} \sqrt{1 + M_2(m^j(0))}$, we get the first thesis with $\beta = \frac{1}{2d+4}$.
We, now, show the proof of the third inequality.

$$\begin{aligned}
&\mathbb{E}\left[d_1(m(s_0 + h), m^N_{X_{t_0+h}})\right] \le \sum_{j=1}^J \frac{n^j}{N} \mathbb{E}\left[d_1(m^j(s_0 + h), m^j_{Y_{s_0+h}})\right] \\
&\le C \sum_{j=1}^J \frac{n^j}{N}(1 + M_2^{1/2}(m^j(s_0)))\frac{h^\beta}{(n^j)^\beta} \\
&\le Ch^\beta \sum_{j=1}^J \frac{(n^j)^{1-\beta}}{N} + Ch^\beta(\sum_{j=1}^J \frac{n^j}{N} M_2(m^j(s_0)))^{1/2}(\sum_{j=1}^J \frac{n^j}{N(n^j)^{2\beta}})^{1/2} \\
&\le Ch^\beta \frac{|J|}{N}\left(\sum_{j=1}^J \frac{n^j}{|J|}\right)^{1-\beta} + CM_2^{1/2}(m(s_0))h^\beta \sqrt{\frac{|J|}{N}}(\sum_{j=1}^J \frac{1}{|J|}(n^j)^{1-2\beta})^{1/2} \\
&\le C\left(\frac{Jh}{N}\right)^\beta (1 + M_2^{1/2}(m(s_0))),
\end{aligned}$$

where we used the definition of $m(s)$, the inequality (2.21), the Cauchy-Schwarz inequality, the concavity of the maps $n \to n^{1-\beta}$ and $n \to n^{1-2\beta}$, the fact that $\sum_j n^j = N$, and the assumption that $\beta \in (0, 1/2)$ and $J \leq C\delta^{-d}$ and the estimate of $M_2(m(s_0))$ in Lemma 2.8.

□

To continue, we need a dynamic programming-type argument, which is stated next.

Lemma 2.12. *With the notation above, we have*

$$\mathcal{U}(s_0 + h, m(s_0 + h)) \geq \mathcal{U}(s_0, m_{y_0^N}) - \int_{s_0}^{s_0+h} (\sum_{j=1}^{J} \int_{\mathbb{R}^d} \frac{1}{N} n^j L(x, \overline{\alpha}_j) m^j(s, x) dx + \mathcal{F}(m(s))) ds. \tag{2.25}$$

Proof. Let $K \in N$ and be $m_0^1, \ldots, m_0^K$ non-negative integrable functions on $\mathbb{R}^d$ such that $\sum_{k=1}^K m_0^k \in \mathcal{P}(\mathbb{R}^d)$.
Define

$$\mathcal{U}^K(t_0, m_0^1, \ldots, m_0^K) := \inf_{(m^1,\beta^1),\ldots,(m^K,\beta^K)} \int_{t_0}^{T} (\int_{\mathbb{R}^d} \sum_{k=1}^{K} L(x, \frac{\beta^k(t,x)}{m^k(t,dx)}) m^k(t,x) dx + \mathcal{F}(\sum_{k=1}^{K} m^k(t))) dt + \mathcal{G}(\sum_{k=1}^{K} m^k(T)),$$

where the infimum is taken over the tuple of measures (m^k, β^k) (the β^k being a vector measure) with β^k absolutely continuous with respect to m^k and such that (m^k, β^k) solve in the sense of distributions,

$$\partial_t m^k - \Delta m^k + \mathrm{div}(\beta^k) = 0 \quad \text{in} \quad (t_0, T] \times \mathbb{R}^d \quad \text{and} \quad m^k(t_0) = m_0^k \quad \text{in} \quad \mathbb{R}^d.$$

We note that

$$\mathcal{U}^K(t_0, m_0^1, \ldots, m_0^K) \leq \mathcal{U}(t_0, m_0^1 + \cdots + m_0^K)$$

since $\mathcal{U}$ is an infimum.
Set $\beta = \sum_{k=1}^K \beta^k$ and $m(t) = \sum_{k=1}^K m^k(t)$, and fix $\epsilon > 0$. Let $(m^1, \beta^1, \ldots, m^K, \beta^K)$ be ϵ-optimal for $\mathcal{U}^K(t_0, m_0^1, \ldots, m_0^K)$.
Then (m, β) solves

$$\partial_t m - \Delta m + \mathrm{div}(\beta) = 0 \quad \text{in} \quad (t_0, T] \times \mathbb{R}^d \quad \text{and} \quad m(t_0) = m_0 \quad \text{in} \quad \mathbb{R}^d,$$

and using the convexity of the map $(\beta, m) \to mL(x, \frac{\beta}{m})$ and the definition of $\mathcal{U}$, we have

$$\begin{aligned}
&\epsilon + \mathcal{U}^K(t_0, m_0^1, \dots, m_0^K) \\
&\geq \int_{t_0}^T (\int_{\mathbb{R}^d} \sum_{k=1}^K L(x, \frac{\beta^k(t,x)}{m^k(t,x)}) \frac{m^k(t,x)}{m(t,x)} m(t,x) dx + \mathcal{F}(\sum_{k=1}^K m^k(t))) dt + \mathcal{G}(\sum_{k=1}^K m^k(T)) \\
&\geq \int_{t_0}^T (\int_{\mathbb{R}^d} L(x, \frac{\sum_{k=1}^K \beta^k(t,x)}{m^k(t,x)}) m(t,x) dx + \mathcal{F}(\sum_{k=1}^K m^k(t))) dt + \mathcal{G}(\sum_{k=1}^K m^k(T)) \\
&\geq \mathcal{U}(t_0, m_0).
\end{aligned} \tag{2.26}$$

We proved that

$$\mathcal{U}^K(t_0, m_0^1, \dots, m_0^K) = \mathcal{U}(t_0, m_0^1 + \cdots + m_0^K).$$

Using the following dynamic programming principle.

Proposition 2.13. *Assume* (1.1). *Then, for any* $0 \leq t_0 \leq t_1 \leq T$,

$$\mathcal{U}(t_0, m_0) = \inf_{(m,\alpha) \in \mathcal{A}} \left\{ \int_{t_0}^{t_1} \int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t) dx + \mathcal{F}(m(t)) dt + \mathcal{U}(t_1, m(t_1)) \right\}.$$

we have,

$$\begin{aligned}
&\mathcal{U}^K(s_0, m_0^1, \dots, m_0^K) = \mathcal{U}(s_0, m_0^1 + \cdots + m_0^K) = \\
&\inf_{(m,\alpha) \in \mathcal{A}} \left\{ \int_{s_0}^{s_0+h} \int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t) dx + \mathcal{F}(m(t)) dt + \mathcal{U}(s_0 + h, m(s_0 + h)) \right\} \\
&\leq \int_{s_0}^{s_0+h} \int_{\mathbb{R}^d} L(x, \alpha^j) m(t) dx + \mathcal{F}(m(t)) dt + \mathcal{U}(s_0 + h, m(s_0 + h)),
\end{aligned}$$

and recalling that $m(t) = \frac{1}{N} \sum_j n^j m^j(t)$ and $m^j(s_0, \cdot) = m^j_{y_0}$, we get the thesis. □

Lemma 2.14. *For any* $(t, x) \in [0, T] \times (\mathbb{R}^d)^N$

$$\begin{aligned}
&\sum_{k=1}^N |D_{x_k} \mathcal{V}^N(t,x) - D_{x_k} \mathcal{V}^N(t_0, x_0)| \\
&\leq \frac{C}{N} \sum_{k=1}^N |x_k - x_0^k| + \left(\frac{C}{N\theta} \sum_{k=1}^N (|x_k - x_0^k| + |x_k - x_0^k|^2) \right)^{1/2} + \frac{C}{\theta^{1/2}} |t - t_0|^{1/2}.
\end{aligned} \tag{2.27}$$

Proof. For $\theta, \lambda \in (0,1)$, we set

$$M := \max_{(t,x),(s,y)\in[0,T]\times(\mathbb{R}^d)^N} e^s(\mathcal{U}(s,m_y^N) - \mathcal{V}^N(t,x)) - \frac{1}{2\theta N}\sum_{i=1}^N |x_i - y_i|^2 - \frac{1}{2\theta}|s-t|^2 - \frac{\lambda}{2N}\sum_{i=1}^N |y_i|^2.$$

We denote by $((t_0,x_0),(s_0,y_0))$ a maximum point in the expression above.

Set $p^k = D_{x_k}\mathcal{V}^N(t_0,x_0)$ and $p^t = \partial_t \mathcal{V}^N(t_0,x_0)$. Using Lemma 2.3, for any $(t_0,m_0),(t,x) \in [0,T]\times(\mathbb{R}^d)^N$, we have,

$$\mathcal{V}^N(t,x) - \mathcal{V}^N(t_0,m_0) - \sum_{k=1}^N p^k\cdot(x_k - x_0^k) - p^t(t-t_0) \le \frac{C}{N}\sum_{k=1}^N |x_k - x_0^k|^2 + C(t-t_0)^2.$$

Since $((t_0,x_0),(s_0,y_0))$ is optimal, for any (t,x),

$$\frac{1}{2N\theta}\sum_{i=1}^N |x_i - y_0^i|^2 + \frac{1}{2\theta}(t-s_0)^2 + \mathcal{V}^N(t,x) \ge \frac{1}{2N\theta}\sum_{i=1}^N |x_0^i - y_0^i|^2 + \frac{1}{2\theta}(t_0-s_0)^2 + \mathcal{V}^N(t_0,x_0). \tag{2.28}$$

In (2.28) putting $x = x_0$ and $t = t_0 + h$, dividing by h and letting h tends to zero, we get that

$$\lim_{h\to 0}\frac{\mathcal{V}^N(t_0+h,x_0) - \mathcal{V}^N(t_0,x_0)}{h} = \partial_t\mathcal{V}^N(t_0,x_0) = p^t \ge \lim_{h\to 0}\left(\frac{|t_0-s_0|^2}{2\theta h} - \frac{|t_0+h-s_0|^2}{2\theta h}\right)$$
$$= \lim_{h\to 0}\frac{2s_0 - 2t_0 - h}{2\theta} = \frac{s_0 - t_0}{\theta},$$

and putting $x = x_0$ and $t = t_0 - h$, dividing by h and letting h tends to zero, we get that

$$p^t \le \frac{s_0 - t_0}{\theta},$$

and so

$$p^t = \frac{s_0 - t_0}{\theta}.$$

Similarly, letting $t = t_0$ and first $x = x_0 + he_k$, and then $x = x_0 - he_k$, letting h going to zero, we get

$$p^k = \frac{y_0^k - x_0^k}{\theta N}.$$

Furthermore, rearranging (2.28), using the triangular inequality and the definition of p^k and p^t, we have

$$\mathcal{V}^N(t,x) - \mathcal{V}^N(t_0,x_0) \geq \frac{1}{2\theta N}\sum_{k=1}^{N}|x_0^k - y_0^k|^2 - \frac{1}{2\theta}|t_0 - s_0|^2 - \frac{1}{2\theta}|t - s_0|^2$$

$$= \frac{1}{2\theta N}\sum_{k=1}^{N}|x_0^k - y_0^k|^2 - \frac{1}{2\theta N}\sum_{k=1}^{N}|(x_k - x_0^k) + (x_0^k - y_0^k)|^2 + \frac{1}{2\theta}|t_0 - s_0|^2$$

$$- \frac{1}{2\theta}|(t - t_0) + (t_0 - s_0)|^2$$

$$\leq \frac{1}{2\theta N}\sum_{k=1}^{N}|x_0^k - y_0^k|^2 - \frac{1}{2\theta N}\sum_{k=1}^{N}|x_k - x_0^k|^2 - \frac{1}{2\theta N}\sum_{k=1}^{N}|x_0^k - y_0^k|^2$$

$$- \frac{1}{2\theta N}\sum_{k=1}^{N}|x_0^k - y_0^k||x_k - x_0^k| + \frac{1}{2\theta}(t_0 - s_0)^2$$

$$- \frac{1}{2\theta}|t_0 - s_0|^2 - \frac{1}{2\theta}(t - t_0)^2 - \frac{1}{\theta}(t_0 - s_0)(t - t_0)$$

$$= \sum_{k=1}^{N} p^k \cdot (x_k - x_0^k) + p^t(t - t_0) - \sum_{k=1}^{N}\frac{1}{2\theta N}|x_k - x_0^k|^2 - \frac{1}{2\theta}(t - t_0)^2,$$

and, putting some addends on the left hand side,

$$\mathcal{V}^N(t,x) - \mathcal{V}^N(t_0,x_0) - \sum_{k=1}^{N} p^k \cdot (x_k - x_0^k) - p^t(t - t_0) \geq -\frac{1}{2N\theta}\sum_{k=1}^{N}|x_k - x_0^k|^2 - \frac{1}{2\theta}(t - t_0)^2.$$

Assuming that $\theta \leq (2C)^{-1}$,we define

$$w(t,x) := \mathcal{V}^N(t_0,x_0) - \mathcal{V}^N(t,x) + \sum_{k=1}^{N} p^k \cdot (x_k - x_0^k) + p^t(t - t_0) + \frac{C}{N}\sum_{k=1}^{N}|x_k - x_0^k|^2 + C(t - t_0)^2.$$

Note that $w(t,x)$, since it is sum of convex and linear functions, is convex and satisfies

$$0 \leq w(t,x) \leq \frac{1}{\theta N}\sum_{k=1}^{N}|x_k - x_0^k|^2 + \frac{1}{\theta}(t - t_0)^2.$$

Thus, using that $w(t,x) \geq 0$ for any (t,x) and any (s,y), we have

$$\begin{aligned}
&\sum_{k=1}^{N} D_{x_k} w(t,x) \cdot (y_k - x_k) + \partial_t w(t,x)(s-t) \\
&\leq w(t,x) + \sum_{k=1}^{N} D_{x_k} w(t,x) \cdot (y_k - x_k) + \partial_t w(t,x)(s-t) \\
&\leq w(s,y) \leq \frac{1}{N\theta} \sum_{k=1}^{N} |y_k - x_0^k|^2 + \frac{1}{\theta}(s-t_0)^2,
\end{aligned}$$

where we used the inequality satisfied by $w(t,x)$.
Let

$$y_k = x_0^k + \frac{1}{2}\theta N D_{x_k} w(t,x) \quad \text{and} \quad s = t_0 + \theta \partial_t w(t,x)$$

in the inequality above, we obtain

$$\frac{\theta N}{4} \sum_{k=1}^{N} |D_{x_k} w(t,x)|^2 \leq \sum_{k=1}^{N} D_{x_k} w(t,x) \cdot (x_k - x_0^k) + \partial_t w(t,x)(t-t_0),$$

and, using the Cauchy-Schwarz inequality and the inequality found above,

$$\begin{aligned}
\sum_{k=1}^{N} |D_{x_k} w(t,x)| &\leq N^{1/2} \left(\sum_{k=1}^{N} |D_{x_k} w(t,x)|^2 \right)^{1/2} \\
&\leq N^{1/2} \left(\frac{4}{N\theta} \sum_{k=1}^{N} |x_0^k - x_k||D_{x_k} w(t,x)| + \frac{4}{N\theta} |\partial_t w(t,x)||t-t_0| \right)^{1/2}.
\end{aligned} \tag{2.29}$$

By the definition of w and since we proved that that $|D_{xk}\mathcal{V}^N| \leq \frac{C}{N}$ and $|\partial_t \mathcal{V}^N| \leq C$, we find

$$|D_{x_k} w(t,x)| = |-D_{x_k}\mathcal{V}^N(t,x) + p^k + \frac{2C}{N}(x_k - x_0^k)| \leq \frac{C}{N} + \frac{2C}{N}|x_k - x_0^k|,$$

and

$$|\partial_t w(t,x)| = |-\partial_t \mathcal{V}^N(t_0,x_0) + p^t + 2C(t-t_0)| \leq C,$$

returning to (2.29), we have

$$\sum_{k=1}^{N} |-D_{x_k}\mathcal{V}(t,x) + p^k + \frac{2C}{N}(x_k - x_0^k)| \leq \frac{C}{N\theta} \left(\sum_{k=1}^{N} (|x_0^k - x_k| + |x_0^k - x_k|^2) + N|t-t_0| \right)^{1/2},$$

and then, by the definition of p^k, we get the thesis.

□

Proposition 2.15. *Assume (1.1). There exists $\beta \in (0,1]$ depending only on the dimension and $C > 0$ depending on the data, such that, for any $N \geq 1$ and any $(t,x) \in [0,T] \times (\mathbb{R}^d)^N$*

$$\mathcal{U}(t, m_{x_0}^N) - \mathcal{V}^N(t,x) \leq \frac{C}{N^\beta}(1 + \frac{1}{N}\sum_{i=1}^N |x^i|^2). \tag{2.30}$$

Remark 2.16. *Dividing the players into subgroups in such a way that the optimal controls for the agents in each subgroup are close and showing a propagation of chaos-type result for each subgroup using a concentration inequality, it can be overcome the main difficulty: transform an optimal control for the $\mathcal{V}^N$ that depends on each particle into a feedback for $\mathcal{U}$.*

Proof of Theorem 2.15. We employ the technique, quite standard for viscosity solutions, to double the variables and we define, for $\theta, \lambda \in (0,1)$,

$$M := \max_{(t,x),(s,y)\in[0,T]\times(\mathbb{R}^d)^N} e^s(\mathcal{U}(s, m_y^N) - \mathcal{V}^N(t,x)) - \frac{1}{2\theta N}\sum_{i=1}^N |x_i - y_i|^2 - \frac{1}{2\theta}|s-t|^2 - \frac{\lambda}{2N}\sum_{i=1}^N |y_i|^2.$$

We denote by $((t_0,x_0),(s_0,y_0))$ a maximum point in the expression above and we continue estimating the error related to the penalization.

We begin creating the subgroups based on an appropriate partition of $\{1,\dots,N\}$ as stated in Lemma 2.9.

Using that M is a max, thanks to the Lipschitz regularity of $\mathcal{U}$ we have

$$\begin{aligned}
M &\geq \mathbb{E}[e^{s_0+h}(\mathcal{U}(s_0+h, m^N_{Y_{s_0+h}}) - \mathcal{V}^N(t_0+h, X_{t_0+h})) \\
&- \frac{1}{2\theta}\left(\frac{1}{N}\sum_{k=1}^N |Y^k_{s_0+h} - X^k_{t_0+h}|^2 + (t_0 - s_0)^2\right) - \frac{\lambda}{2N}\sum_{i=1}^N |Y^i_{s_0+h}|^2] \\
&\geq \mathbb{E}[e^{s_0+h}(\mathcal{U}(s_0+h, m(s_0+h)) - \mathcal{V}^N(t_0+h, X_{t_0+h}))] - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} \\
&- \frac{1}{2\theta}\left(\frac{1}{N}\sum_{k=1}^N |y_0^k - x_0^k|^2 + (s_0 - t_0)^2\right) - \frac{\lambda}{2N}\sum_{i=1}^N (|y_0^i| + Ch^{1/2})^2,
\end{aligned}$$

where we used definition of X_t and Y_t so that $|X^k_{t_0+h} - Y^k_{s_0+h}| = |x_0^k + \tau\alpha^k + \sqrt{2}B^k_\tau - (y_0^k + \tau\alpha^k + \sqrt{2}B^k_\tau)| = |x_0^k - y_0^k|$ and $|t_0 + h - (s_0+h)| = |t_0 - s_0|$ and the equation (2.23).

At this point, we using the dynamic programming-type argument in Lemma 2.12 and Itô's formula for $\mathcal{V}^N$, we find

$$M \geq e^{s_0+h}\mathcal{U}(s_0, m^N_{y_0}) - e^{s_0+h}\int_{s_0}^{s_0+h}(\int_{\mathbb{R}^d}\sum_{j=1}^{J}\frac{1}{N}n^j L(x,\overline{\alpha}^j)m^j(s,x)dx + \mathcal{F}(m(s)))ds$$

$$-e^{s_0+h}\mathbb{E}\left[\mathcal{V}^N(t_0,x_0) + \int_{t_0}^{t_0+h}(\partial_t\mathcal{V}^N(t,X_t) + \sum_{k=1}^{N}(\Delta_{x_k}\mathcal{V}(t,X-t) + \alpha^k\cdot D_{x_k}\mathcal{V}(t,X_t))dt\right]$$

$$-C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} - \frac{1}{2\theta}\left(\frac{1}{N}\sum_{k=1}^{N}|y_0^k - x_0^k|^2 + (s_0-t_0)^2\right) - \frac{\lambda}{2N}\sum_{i=1}^{N}(|y_0^i| + Ch^{1/2})^2.$$

Since the $\overline{\alpha}^j$ are uniformly bounded, the map $L(\cdot,\overline{\alpha}^j)$ is uniformly Lipschitz independently of j. Hence, using lemmas 2.10 and 2.8, a change of variables in the integral and the fact that $\overline{\alpha}^j = \alpha^k$ if $k \in C^j$, we find

$$\int_{s_0}^{s_0+h}\int_{\mathbb{R}^d}\sum_{j=1}^{J}\frac{1}{N}n^j L(x,\overline{\alpha}^j)m^j(s,x)dxds$$

$$\leq \mathbb{E}\left[\int_{s_0}^{s_0+h}\sum_{j=1}^{J}(\sum_{k\in C^j}\frac{1}{N}L(X^k_{t_0-s_0+s},\overline{\alpha}^j) + C\frac{1}{N}n^j d_1(m^j(s), m_{X^j_{t_0-s_0+s}}))ds\right]$$

$$\leq \mathbb{E}\left[\int_{t_0}^{t_0+h}\sum_{k=1}^{N}\frac{1}{N}L(X^k_s,\alpha^k)ds\right] + C\theta h + C\sum_{j=1}^{J}\frac{1}{N}n^j(1+M_2^{1/2}(m^j_{s_0}))\frac{h^\beta}{(n^j)^\beta}$$

$$\leq \mathbb{E}\left[\int_{t_0}^{t_0+h}\sum_{k=1}^{N}\frac{1}{N}L(X^k_s,\alpha^k)ds\right] + C\theta h + C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta}.$$

Note that in the last inequality we used exactly the same argument as for the proof given above for the third inequality of Lemma 2.10.

Hence, recalling the optimality for M of $((t_0,x_0),(s_0,y_0))$ and employing the equation for $\mathcal{V}^N$, we get

$$0 \geq (e^{s_0+h} - e^{s_0})(\mathcal{U}(s_0,m^N_{y_0}) - \mathcal{V}^N(t_0,x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta}$$

$$- C\lambda h^{1/2}N^{-1}\sum_{i=1}^{N}|y_0^i| - C\theta h - e^{s_0+h}\mathbb{E}\left[\int_{s_0}^{s_0+h}(\mathcal{F}(m(s)) - \mathcal{F}(m^N_{X_{s_0-t_0+s}}))ds\right]$$

$$- e^{s_0+h}\mathbb{E}\left[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N}(L(X^k_s,\alpha^k) + \alpha^k\cdot(ND_{x_k}\mathcal{V}^N(s,X_s)) + H(X^k_s, ND_{x_k}\mathcal{V}^N(s,X_s)))ds\right].$$

Using the Lipschitz regularity of $\mathcal{F}$ and Lemma 2.10 to deal with the difference of the $\mathcal{F}$ and Lemma 2.8 to deal with the term in $\sum_i |y_0^i|$, we find

$$0 \geq e^{s_0}h(\mathcal{U}(s_0, m_{y_0}^N) - \mathcal{V}^N(t_0, x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} - C\lambda^{1/2}h^{1/2} - C\theta h - Ch^2$$
$$- e^{s_0+h}\mathbb{E}\left[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N (L(X_s^k, \alpha^k) + \alpha^k \cdot (ND_{x_k}\mathcal{V}^N(s, X_s)) + H(X_s^k, ND_{x_k}\mathcal{V}^N(s, X_s)))ds\right]. \tag{2.31}$$

Noting that, for $s \geq t_0$

$$|X_s^k - x_0^k| = |x_0^k + (s-t_0)\alpha_k + \sqrt{2}B_{s-t_0}^k - x_0^k| = |(s-t_0)\alpha_k + \sqrt{2}B_{s-t_0}^k|,$$

and, thanks to the regularity of L and H, the uniform boundedness of the α^k, that

$$\begin{aligned}
&- e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N (L(X_s^k, \alpha^k) \pm L(x_0^k, \alpha^k)\\
&+ H(X_s^k, ND_{x_k}\mathcal{V}^N(s, X_s)) \pm H(x_0^k, ND_{x_k}\mathcal{V}^N(s, X_s)))ds]\\
&\geq -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N L(x_0^k, \alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s, X_s)))ds]\\
&- e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N (|L(X_s^k, \alpha^k) - L(x_0^k, \alpha^k)|\\
&+ |H(X_s^k, ND_{x_k}\mathcal{V}^N(s, X_s)) - H(x_0^k, ND_{x_k}\mathcal{V}^N(s, X_s))|)ds]\\
&\geq -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N L(x_0^k, \alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s, X_s)))ds]\\
&- e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N (C|X_s^k - x_0^k| + C|X_s^k - x_0^k|)ds]\\
&= -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N L(x_0^k, \alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s, X_s)))ds]\\
&- e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^N (C|(s-t_0)\alpha_k + \sqrt{2}B_{s-t_0}^k| + C|(s-t_0)\alpha_k + \sqrt{2}B_{s-t_0}^k|)ds]
\end{aligned}$$

$$= -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N} L(x_0^k,\alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s,X_s)))ds]$$

$$- e^{s_0+h}\mathbb{E}[|\alpha_k|\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N}(C(s-t_0))ds]$$

$$= -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N} L(x_0^k,\alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s,X_s)))ds]$$

$$- e^{s_0+h}\mathbb{E}[|\alpha_k|\frac{1}{N}NC(\frac{t_0^2}{2}+\frac{h^2}{2}+t_0h-\frac{t_0^2}{2}-t_0^2-t_0h+t_0^2)]$$

$$\geq -e^{s_0+h}\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N} L(x_0^k,\alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s,X_s)))ds]$$

$$- Ce^{s_0+h}\frac{h^2}{2}.$$

Adding and subtracting the term $\sum_k \left(L(x_0^k,\alpha^k) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s,X_s))\right)$ in the inequality (2.31), using the last inequality and that, for $h \leq 1$, $-Ce^h\frac{h^2}{2} \geq -Ch^{3/2}$, we get

$$0 \geq e^{s_0}h(\mathcal{U}(s_0, m_{y_0}^N) - \mathcal{V}^N(t_0,x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta}$$
$$- C\lambda^{1/2}h^{1/2} - C\theta h - Ch^2 - Ch^{3/2}$$
$$- e^{s_0+h}\mathbb{E}\left[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N}(L(x_0^k,\alpha^k) + \alpha^k\cdot(ND_{x_k}\mathcal{V}^N(s,X_s)) + H(x_0^k, ND_{x_k}\mathcal{V}^N(s,X_s)))ds\right],$$

and, in view of (2.20),

$$0 \geq e^{s_0}h(\mathcal{U}(s_0, m_{y_0}^N) - \mathcal{V}^N(t_0,x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} - C\lambda^{1/2}h^{1/2} - C\theta h - Ch^{3/2}$$
$$- C\mathbb{E}[\frac{1}{N}\int_{t_0}^{t_0+h}\sum_{k=1}^{N}|ND_{x_k}\mathcal{V}^N(s,X_s)) - ND_{x_k}\mathcal{V}^N(s,x_0))|ds] - \frac{Ch\delta}{N}. \tag{2.32}$$

The semiconcavity of $\mathcal{V}^N$ and the penalization by the term in θ give the Lemma 2.14.

Inserting the estimate of Lemma 2.14 in (2.32) and using that $\frac{Ch\delta}{N} \leq C\theta h$, we obtain

$$
\begin{aligned}
&0 \geq e^{s_0} h(\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{V}^N(t_0, x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} \\
&- C\lambda^{1/2}h^{1/2} - C\theta h - \frac{Ch\delta}{N} - Ch^{3/2} \\
&- C\mathbb{E}\left[\int_{t_0}^{t_0+h} \left(\frac{C}{N}\sum_{k=1}^N |X^k_s - x^k_0| + \left(\frac{C}{N\theta}\sum_{k=1}^N (|X^k_s - x^k_0| + |X^k_s - x^k_0|^2)\right)^{1/2} + \frac{1}{\theta^{1/2}}|s-t_0|^{1/2}\right) ds\right] \\
&\geq e^{s_0} h(\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{V}^N(t_0, x_0)) - C\delta^{-d\beta}(1+\lambda^{-1/2})\frac{h^\beta}{N^\beta} \\
&C(\theta+\delta)h - C\lambda^{1/2}h^{1/2} - C\theta^{-1/2}h(h^{1/2}+h)^{1/2}.
\end{aligned}
$$

Dividing by h we find, for each choice of $\theta, \lambda, \delta > 0$ and $0 < h \leq (T-s_0)\wedge(T-t_0)$, that

$$
e^{s_0} h(\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{V}^N(t_0, x_0)) \leq C\frac{h^{\beta-1}}{N^\beta \delta^{d\beta}}(1+\lambda^{-1/2}) + C(\theta+\delta) + C\lambda^{1/2}h^{-1/2} + Ch^{1/4}\theta^{-1/2},
$$

we take

$$
\theta = h^{\alpha_1}, \quad \delta = \left(\frac{\lambda^{-1/2}h^{\beta-1}}{N^\beta}\right)^{\alpha_2}, \quad \lambda = N^{-\alpha_3} \text{ and } h = N^{\alpha_4}.
$$

Making appropriate choices of α_1, α_2, α_3 and α_4 we deduce that

$$
e^{s_0} h(\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{V}^N(t_0, x_0)) \leq CN^{-\widetilde{\beta}} \tag{2.33}
$$

holds for some $\widetilde{\beta} = \widetilde{\beta}(\beta) \in (0, 1/2)$ and for all values of N such that $h = N^{-\alpha_4} \leq (T-s_0)\wedge(T-t_0)$.
For those values of N such that $h = N^{-\alpha_4} \geq (T-s_0)\wedge(T-t_0)$, we have by Lemma 2.8 that $(T-s_0)\vee(T-t_0) \leq h + C\theta$, so, noting that $\mathcal{U}(T,m) = \mathcal{G}(m)$ and $\mathcal{V}^N(T,x) = \mathcal{G}(m^N_x)$ and using Lemma 2.2, we find

$$
\begin{aligned}
&|\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{V}^N(t_0, x_0)| \\
&\leq |\mathcal{U}(s_0, m^N_{y_0}) - \mathcal{G}(m^N_{y_0})| + |\mathcal{G}(m^N_{y_0}) - \mathcal{G}(m^N_{x_0})| + |\mathcal{G}(m^N_{x_0}) - \mathcal{V}^N(t_0, x_0)| \\
&\leq C(h+\theta) + C\theta + C(h+\theta) \leq CN^{-\widetilde{\beta}},
\end{aligned}
$$

where in the last line we choose $\widetilde{\beta}$ even smaller if necessary. With this choice of $\widetilde{\beta}$, we have now established that (2.33) holds for all values of N.

Finally, using the optimality of $((t_0, x_0), (s_0, y_0))$ in M, we conclude that, for all $(t, x) \in [0, T] \times (\mathbb{R}^d)^N$,

$$
\begin{aligned}
e^t |\mathcal{U}(t, m_x^N) - \mathcal{V}^N(t, x)| &\leq e^{s_0}(\mathcal{U}(s_0, m_{y_0}^N) - \mathcal{V}^N(t_0, x_0)) + \frac{\lambda}{2N} \sum_{i=1}^N |x_i|^2 \\
&\leq C N^{-\min(\widetilde{\beta}, \alpha_3)} (1 + \frac{1}{N} \sum_{i=1}^N |x_i|^2).
\end{aligned}
$$

□

Proof of Theorem 2.1. Combining Proposition 2.6 and Proposition 2.15 we know that there exist $\beta \in (0, 1]$ depending on dimension and $C > 0$ depending on the data such that, for any $(t, x) \in [0, T] \times (\mathbb{R}^d)^N$,

$$
|\mathcal{U}(t, m_x^N) - \mathcal{V}^N(t, x))| \leq C N^{-\beta}(1 + M_2^{1/2}(m_x^N) + M_2(m_x^N)) \leq C N^{-\beta}(1 + M_2(m_x^N)).
$$

□

Chapter 3

The regularity of $\mathcal{U}$

In this chapter, to show the regularity of $\mathcal{U}$, we analyze a linearized version of the system (1.11). We also introduce the concept of stability and strong stability which are used to define and analyze the open and dense set $\mathcal{O}$ in which the map $\mathcal{U}$ should be smooth.
We use this result to obtain the propagation of chaos property analyzed in Chapter 4.

3.1 The linearized system

Let m be the solution to

$$\begin{cases} \partial_t m - \Delta m - \mathrm{div}(Vm) = 0 \text{ in } (t_0, T) \times \mathbb{R}^d, \\ m(t_0) = m_0 \text{ in } \mathbb{R}^d, \end{cases} \tag{3.1}$$

with $t_0 \in [0, T)$, $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $V : [t_0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $||V||_{\mathcal{C}^{1,3}} \le C_0$, $C_0 > 0$ constant.

In order to prove the regularity of the value function $\mathcal{U}$, we want to show that $\mathcal{U}$ has a derivative with respect to m. To do so we have to differentiate the system (1.11) and we study the new system that we have obtained.
We call it the **inhomogeneous linearized system** of (1.11)

$$\begin{cases} \partial_t z - \Delta z + V(t,x) \cdot Dz = \frac{\delta F}{\delta m}(x, m(t))(\rho(t)) + R^1(t,x) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \partial_t \rho - \Delta \rho - \mathrm{div}(V\rho) - \sigma \mathrm{div}(m \Gamma Dz) = \mathrm{div}(R^2) \text{ in } (t_0, T) \times \mathbb{R}^d, \\ \rho(t_0) = \xi \text{ and } z(T, x) = R^3 + \frac{\delta G}{\delta m}(x, m(T))(\rho(T)) \text{ in } \mathbb{R}^d, \end{cases} \tag{3.2}$$

where

$$\begin{aligned}&\sigma \in [0,1] \text{ and } \delta \in (0,1),\\&\Gamma \in \mathcal{C}^0([0,T]\times\mathbb{R}^d;\mathbb{R}^{d\times d}) \text{ with } ||\Gamma||_\infty \leq C_0,\\&R^1 \in \mathcal{C}^{\delta/2,\delta},\ \ R^2 \in L^\infty([t_0,T],(W^{1,\infty})'(\mathbb{R}^d,\mathbb{R}^d)),\ \ R^3 \in \mathcal{C}^{2+\delta} \text{ and } \xi \in (W^{1,\infty})'.\end{aligned} \tag{3.3}$$

If $z \in \mathcal{C}^{0,1}([t_0,T]\times\mathbb{R}^d)$ and $\rho \in \mathcal{C}^0([0,T],(\mathcal{C}^{2+\delta})')$ satisfy respectively the first and the second equation in the sense of distributions, then the pair (z,ρ) is a solution to (3.2).
Note that, the maps $(t,x) \to \frac{\delta F}{\delta m}(x,m(t))(\rho(t))$ and $x \to \frac{\delta G}{\delta m}(x,m(T))(\rho(T))$ are continuous and bounded because of the regularity of ρ and the assumptions on $\mathcal{F}$ and $\mathcal{G}$.
We will use the system in (3.2) with $V(t,x) = H_p(x,Du(t,x))$ and $\Gamma(t,x) = H_{pp}(x,Du(t,x))$, where (u,m) is a classical solution to (1.11).

The system

$$\begin{cases}\partial_t z - \Delta z + V(t,x)\cdot Dz = \frac{\delta F}{\delta m}(x,m(t))(\rho(t)) \text{ in } (t_0,T)\times\mathbb{R}^d,\\ \partial_t\rho - \Delta\rho - \operatorname{div}(V\rho) - \sigma\operatorname{div}(m\Gamma Dz) = 0 \text{ in } (t_0,T)\times\mathbb{R}^d,\\ \rho(t_0) = 0 \text{ and } z(T,x) = \frac{\delta G}{\delta m}(x,m(T))(\rho(T)) \text{ in } \mathbb{R}^d\end{cases} \tag{3.4}$$

is the homogeneous version of (3.2).

Definition 3.1. *We say that the system* (3.4) *is* strongly stable *if, for any* $\sigma \in [0,1]$*, the unique solution to the system is* $(z,\rho) = (0,0)$*.*

We will use a weaker notion of stability for (3.4) with $\sigma = 1$ when dealing with the optimal control system. We need, however, the notion of strong stability with $\sigma \in [0,1]$ for Proposition 3.5 below in order to prove the existence of a solution to (3.2).

Lemma 3.2. *Assume* (1.1) *and the system* (3.4) *strongly stable.*
There exist a neighborhood $\mathcal{V}$ *of* (V,Γ) *in the topology of locally uniform convergence, and* η*,* $C > 0$ *such that, for any* $(V',t_0',m_0',\Gamma',R^{1,'},R^{2,'},R^{3,'},\xi',\sigma')$ *with*

$$\begin{aligned}&(V',\Gamma')\in\mathcal{V},\ \ |t_0'-t_0| + d_2(m_0',m_0) \leq \eta,\ \ ||V'||_{\mathcal{C}^{1,3}} + ||\Gamma'||_\infty \leq 2C_0,\ \ \sigma' \in [0,1],\\&R^{1,'} \in \mathcal{C}^{\delta/2,\delta},\ \ R^{2,'} \in L^\infty([t_0,T],(W^{1,\infty})'(\mathbb{R}^d,\mathbb{R}^d)),\ \ R^{3,'} \in \mathcal{C}^{2+\delta},\ \ \xi' \in (W^{1,\infty})',\end{aligned} \tag{3.5}$$

any solution (z', ρ') *to* (3.2) *associated with these data on* $[t_0', T]$ *and* m' *the solution to* (3.1) *with drift* V' *and initial condition* m_0' *at time* t_0' *satisfies*

$$||z'||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0',T]} ||\rho'(t,\cdot)||_{(\mathcal{C}^{2+\delta})'} + \sup_{t'\neq t} \frac{||\rho'(t',\cdot) - \rho'(t,\cdot)||_{(\mathcal{C}^{2+\delta})'}}{|t'-t|^{1/2}} \le CM', \quad (3.6)$$

where

$$M' := ||\xi'||_{(W^{1,\infty})'} + ||R^{1,'}||_{\mathcal{C}^{\delta/2,\delta}} + ||R^{3,'}||_{\mathcal{C}^{2+\delta}} + \sup_{t\in t_0',T]} ||R^{2,'}(t)||_{(W^{1,\infty})'}. \quad (3.7)$$

The idea of the proof of Lemma 3.2 follows some of the ideas of [6], where a similar system is studied. For the rigorous proof, see Lemma 2.1 of [10]

An immediate consequence is the following corollary.

Corollary 3.3. *Assume* (1.1) *and that the system* (3.4) *strongly stable.*
Then, for any (V', m_0', Γ') *satisfying* (3.5), *the corresponding homogeneous linearized system is strongly stable.*

Now we state a new lemma very similar to Lemma 3.2: the difference between the estimate below and the one of previous lemma is the right hand side of the former which depends on the solution itself.

Lemma 3.4. *Assume* (1.1) *and let* (z, ρ) *be a solution to* (3.4).
There is a constant $C > 0$, *depending only on the regularity of* $\mathcal{F}$, $\mathcal{G}$ *and on* $||V||_{\mathcal{C}^{1,3}} + ||\Gamma||_\infty$, *such that*

$$||z||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0,T]} \sup_{t'\neq t} \frac{||\rho(t',\cdot) - \rho(t,\cdot)||_{(\mathcal{C}^{2+\delta})'}}{|t'-t|^{\delta/2}} \le C(M + \sup_{t\in[t_0,T]} ||\rho(t)||_{(\mathcal{C}^{2+\delta})'}), \quad (3.8)$$

where

$$M := ||\xi||_{(W^{1,\infty})'} + ||R^1||_{\mathcal{C}^{\delta/2,\delta}} + ||R^3||_{\mathcal{C}^{2+\delta}} + \sup_{t\in t_0,T]} ||R^2(t)||_{(W^{1,\infty})'}.$$

For the proof see [10, Lemma 2.3].

We complete the section with an existence result for system (3.2) given a solution m to (3.1).

Proposition 3.5. *Assume* (1.1) *and the system* (3.4) *strongly stable.*
Then, for any ξ, δ, Γ, R^1, R^2, R^3 *as in* (3.3), *there exists a unique solution to the linearized system* (3.2) *with* $\sigma = 1$.

Proof. let us use a *continuation method*.
Let us call

$$\Sigma = \{\sigma \in [0,1] : \text{the system (3.2) has a solution for any data } \xi, R^1, R^2, R^3 \text{ satisfying (3.3)}\}$$

We want to prove that Σ is *non-empty, open* and *closed* in $[0,1]$, to conclude that $\Sigma = [0,1]$.

Σ is *non-empty* since, letting $\sigma = 0$ the equation for ρ has a unique solution $\rho = 0$ and we get also the solution $z = 0$. Hence, $0 \in \Sigma$.

We now check that Σ is *closed*. Let $\sigma^n \to \sigma \in [0,1]$ and (z^n, ρ^n) be the associated solution to (3.2) given some ξ, R^1, R^2 and R^3. In view of Lemma 3.2, we have

$$||z^n||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0',T]} ||\rho^n(t,\cdot)||_{(\mathcal{C}^{2+\delta})'} + \sup_{t'\neq t} \frac{||\rho^n(t',\cdot) - \rho^n(t,\cdot)||_{(\mathcal{C}^{2+\delta})'}}{|t'-t|^{1/2}} \le C.$$

Then, passing to the limit in system (3.2), we find a solution for the system for σ as a limit (up to subsequence) of the (z^n, ρ^n)'s. By Lemma 3.2, we get that this solution is also unique, and so Σ is closed.

Finally, we have to prove that Σ *is open*. We fix ξ, R^1, R^2 and R^3 and $\sigma \in \Sigma$. Let $\sigma' \in [0,1]$ be close to σ and $(z', \rho') \in \mathcal{C}^{(2+\delta)/2,2+\delta} \times \mathcal{C}^0([t_0,T], (\mathcal{C}^{2+\delta})')$. Let (z'', ρ'') be a solution to

$$\begin{cases} -\partial_t z'' - \Delta z'' + V(t,x)\cdot Dz'' = \frac{\delta F}{\delta m}(x, m(t))(\rho''(t)) + R^1(t,x) \text{ in } (t_0,T)\times\mathbb{R}^d, \\ \partial_t \rho'' - \Delta\rho'' - \mathrm{div}(V\rho'') - \sigma \mathrm{div}(m\Gamma Dz'') = \mathrm{div}(R^2 + (\sigma'-\sigma)m\Gamma Dz') \text{ in } (t_0,T)\times\mathbb{R}^d, \\ \rho''(t_0) = \xi \text{ and } z''(T,x) = R^3 + \frac{\delta G}{\delta m}(x, m(T))(\rho''(T)) \text{ in } \mathbb{R}^d, \end{cases}$$

which is uniquely solvable since $\sigma \in \Sigma$. Call ϕ the map such that $\phi((z', \rho')) = (z'', \rho'')$. ϕ is a contraction, indeed, let (z_1'', ρ_1'') and (z_2'', ρ_2'') be such that $\phi((z_1', \rho_1')) = (z_1'', \rho_1'')$ and $\phi((z_2', \rho_2')) = (z_2'', \rho_2'')$. The difference $(z_2'' - z_1'', \rho_1'' - \rho_2'')$ satisfies an equation of the form (3.2) with σ and $R^1 = R^3 = \xi = 0$ and $R^2 = (\sigma' - \sigma)m\Gamma(z_2' - z_1')$. By Lemma 3.2, we have

$$\begin{aligned} &||z_2'' - z_1''||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0',T]} ||(\rho_1'' - \rho_2'')(t,\cdot)||_{(\mathcal{C}^{2+\delta})'} \\ &+ \sup_{t'\neq t} \frac{||(\rho_1''-\rho_2'')(t',\cdot) - (\rho_1''-\rho_2'')(t,\cdot)||_{(\mathcal{C}^{2+\delta})'}}{|t'-t|^{1/2}} \\ &\le C|\sigma'-\sigma| \sup_t ||(m\Gamma D(z_1'-z_2')(t,\cdot)||_{(W^{1,\infty})'} \le C|\sigma'-\sigma| ||z_2'-z_1'||_{\mathcal{C}^{(2+\delta)/2,2+\delta}}. \end{aligned}$$

If $|\sigma' - \sigma|$ is small enough, ϕ is a contraction and, for Banach-Caccioppoli's Theorem, there exists a unique fixed point for ϕ. Thus, (z', ρ') fixed point for ϕ is a solution to (3.2) with σ' and so $\sigma' \in \Sigma$. Therefore, Σ is open. By the definition of Σ and strongly stable solution, we get the thesis.

□

3.2 The stability property

Let $(t_0, m_0) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^d)$ and the associated multiplier u to (m, α) minimizer for $\mathcal{U}(t_0, m_0)$, that is, the pair (u, m) solves (1.11) and $\alpha(t,x) = -H_p(x, Du(t,x))$.

Definition 3.6. *The solution (u,m) is* strongly stable *(resp. stable), if $\forall \sigma \in [0,1]$ (resp. $\sigma = 1$) the only solution $(z, \mu) \in \mathcal{C}^{(1+\delta)/2, 1+\delta} \times \mathcal{C}^0([t_0,T]; (\mathcal{C}^{2+\delta}(\mathbb{R}^d))')$ to the linearized system*

$$\begin{cases} -\partial_t z - \Delta z + H_p(x, Du) \cdot Dz = \frac{\delta F}{\delta m}(x, m(t))(\mu(t)) \ \textit{in}\ (t_0, T) \times \mathbb{R}^d, \\ \partial_t \mu - \Delta \mu - \mathrm{div}(H_p(x, Du)\mu) - \sigma \mathrm{div}(H_{pp}(x, Du) Dz m) = 0 \ \textit{in}\ (t_0, T) \times \mathbb{R}^d, \\ \mu(t_0) = 0 \ \textit{and}\ z(T,x) = \frac{\delta G}{\delta m}(x, m(T))(\mu(T)) \ \textit{in}\ \mathbb{R}^d \end{cases} \tag{3.9}$$

is $(z, \mu) = (0,0)$.
Moreover, we say that the minimizer (m, α) is strongly stable *(resp. stable) if (u,m) is strongly stable (resp. stable).*

We point out that, for the choice of $V(t,x) = H_p(x, Du(t,x))$ and $\Gamma(t,x) = H_{pp}(x, Du(t,x))$, the system (3.9) is the linearized version of the one studied in the previous subsection. To emphasize that we are working with this particular system and also be consistent with other references, for the solutions, we use the notation (z, μ) instead of (z, ρ).

The following lemma asserts that the minimizers from an initial point in $\mathcal{O}$ are strongly stable.

Lemma 3.7. *Assume* (1.1).
Fix $(t_0, m_0) \in \mathcal{O}$ and let (m, α) be the unique stable minimizer associated to $\mathcal{U}(t_0, m_0)$. Then (m, α) is strongly stable.

Proof. Let (z, ρ) be a solution to (3.4). For $\sigma = 1$, (m, α) by hypothesis is the unique stable minimizer and so $(z, \rho) = (0,0)$ is the unique solution of the system. For $\sigma = 0$, the first equation in (3.4) does not depend on z and $\rho(t_0) = 0$, therefore $\rho = 0$ and thanks to this $z(T,x) = 0$ and $z = 0$.

Let assume $\sigma \in (0,1)$. Thanks to Lemma 3.2 we have that $z \in \mathcal{C}^{(2+\delta)/2, 2+\delta}$.
By duality

$$\langle z(t,\cdot), \mu(t)\rangle = -\int_{t_0}^{t} (\int_{\mathbb{R}^d} (\sigma H_{pp}(x,Du)Dz \cdot Dzmdx) + \langle \frac{\delta F}{\delta m}(\cdot, m(t))(\mu(t)), \mu(t)\rangle)dt,$$

and, for $t = T$,

$$\begin{aligned}&\int_{t_0}^{T} (\int_{\mathbb{R}^d} (\sigma H_{pp}(x,Du)Dz \cdot Dzmdx) + \langle \frac{\delta F}{\delta m}(\cdot, m(t))(\mu(t)), \mu(t)\rangle)dt + \\ &+ \langle \frac{\delta G}{\delta m}(\cdot, m(T))(\mu(T)), \mu(T)\rangle = 0.\end{aligned} \tag{3.10}$$

From (1.17), with $\rho = \mu$ and $\beta = \sigma H_{pp}(x,Du)Dz$ we have

$$\begin{aligned}&\int_{t_0}^{T}\int_{\mathbb{R}^d} (L_{\alpha,\alpha}(x,\alpha(t,x))\sigma H_{pp}(x,Du)Dz \cdot \sigma H_{pp}(x,Du)Dzmdx) + \\ &+ \int_{\mathbb{R}^d} \frac{\delta F}{\delta m}(x, m(t), y)(\mu(t,x)), \mu(t,y))dy)dxdt + \\ &+ \int_{\mathbb{R}^{2d}} \frac{\delta G}{\delta m}(x, m(T), y)(\mu(T,x)), \mu(T,y)dydx = \\ &\int_{t_0}^{T}\int_{\mathbb{R}^d} (\sigma^2 H_{pp}(x,Du)Dz \cdot Dzmdx) + \\ &+ \int_{\mathbb{R}^d} \frac{\delta F}{\delta m}(x, m(t), y)(\mu(t,x)), \mu(t,y))dy)dxdt + \\ &+ \int_{\mathbb{R}^{2d}} \frac{\delta G}{\delta m}(x, m(T), y)(\mu(T,x)), \mu(T,y)dydx \geq 0,\end{aligned} \tag{3.11}$$

where we used that, since $\alpha = -H_p(x,Du)$, $L_{\alpha,\alpha}(x,\alpha(t,x))H_{pp}(x,Du(t,x)) = \mathbb{I}_d$.
Combining the last two results, we obtain

$$(\sigma - \sigma^2)\int_{t_0}^{T}\int_{\mathbb{R}^d} H_{pp}(x,Du)Dz \cdot Dzmdx \leq 0.$$

Since $\sigma < 1$ then $\sigma - \sigma^2 > 0$ and by assumption $H_{pp} > 0$, we have $Dzm = 0$ and we can conclude that $(z,\mu) = (0,0)$. □

The next lemma establishes that $\mathcal{O}$ is not empty.

Lemma 3.8. *Assume* (1.1).
Fix $(t_0, m_0) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$ *and let* (m,α) *be a minimizer for* $\mathcal{U}(t_0,m_0)$. *Then,* $\forall t \in (t_0, m_0)$, *we have* $(t, m(t)) \in \mathcal{O}$.

Proof. Fix $(t_0, m_0) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d)$ and let (m, α) be a minimizer for $\mathcal{U}(t_0, m_0)$ and u its associated multiplier.
For $t_1 \in (t_0, T)$, set $m(t_1) = m_1$ and let $(\widetilde{m}, \widetilde{\alpha})$ be an optimal solution for $\mathcal{U}(t_1, m_1)$ with associated multiplier $\widetilde{u}$. Thanks to the dynamic principle

$$(\hat{m}, \hat{\alpha}) = \begin{cases} (m, \alpha) \text{ on } [t_0, t_1) \times \mathbb{R}^d, \\ (\widetilde{m}, \widetilde{\alpha}) \text{ on } [t_1, T] \times \mathbb{R}^d \end{cases}$$

is optimal for $\mathcal{U}(t_0, m_0)$. By Lemma 1.9, $\hat{\alpha} \in \mathcal{C}^{(1+\delta)/2, 1+\delta}$ and so $\alpha(t_1, \cdot) = \widetilde{\alpha}(t_1, \cdot)$ and $Du(t_1, \cdot) = D\widetilde{u}(t_1, \cdot)$.

Let us call

$$\begin{aligned} g^k(t,x) =& H_{x_k}(x, Du) - H_{x_k}(x, D\widetilde{u}) + H_p(x, Du) \cdot D(\partial_{x_k} u) - H_p(x, D\widetilde{u}) \cdot D(\partial_{x_k} \widetilde{u}) \\ &- F_{x_k}(x, m(t)) + F_{x_k}(x, m_1(t)), \end{aligned}$$

and

$$h = H_p(Du)m - H_p(D\widetilde{u})\widetilde{m}.$$

Then, the pair $((z^k)_{k=1,\dots,d}, \mu = ((\partial_{x_k}(u - \widetilde{u}))_{k=1,\dots,d}, m - \widetilde{m})$ solves the system

$$\begin{cases} -\partial z^k - \Delta z^k + g^k(t,x) = 0 \text{ in } (t_1, T) \times \mathbb{R}^d, \\ \partial\mu - \Delta\mu + \operatorname{div}(h) = 0 \text{ in } (t_1, T) \times \mathbb{R}^d, \\ \mu(t_1) = 0 \textit{ and } z^k(t_1, \cdot) = 0 \text{ on } \mathbb{R}^d. \end{cases} \tag{3.12}$$

Since $t_1 > t_0$ and $m, \widetilde{m} \in \mathcal{C}^{1,2}([t_1, T] \times \mathbb{R}^d)$ and $m(t, \cdot), \widetilde{m}(t, \cdot)$ are bounded in L^2 we have the following estimates :

$$\begin{aligned} \sum_{k=1}^{d} |g^k(t,x)|^2 \leq C(|z(t,x)|^2 + |Dz(t,x)|^2 + ||\mu(t)||_{L^2}^2), \\ |h(t,x)|^2 \leq C(|z(t,x)|^2 + |\mu(t,x)|^2), \\ |\operatorname{div}(h(t,x))|^2 \leq C(|z(t,x)|^2 + |Dz(t,x)|^2 + |\mu(t,x)|^2 + |D\mu(t,x)|^2). \end{aligned} \tag{3.13}$$

Then a *Lions-Malgrange-type argument* [see Appendix C], we can conclude that the solution to (3.12) is $(z_k, \mu) = (0, 0)$ and so the solution starting from (t_1, m_1) is unique.

Let $\sigma = 1$ and (z, μ) be a solution to (3.9) in $[t_1, T] \times \mathbb{R}^d$
As in the previous Lemma in (3.10)

$$\begin{aligned} &\int_{t_1}^{T} (\int_{\mathbb{R}^d} (H_{pp}(x, Du(t,x)) Dz \cdot Dzm dx) + \left\langle \frac{\delta F}{\delta m}(\cdot, m(t))(\mu(t)), \mu(t) \right\rangle) dt + \\ &+ \left\langle \frac{\delta G}{\delta m}(\cdot, m(T))(\mu(T)), \mu(T) \right\rangle = 0. \end{aligned} \tag{3.14}$$

By Lemma 1.11, for (1.17) we have $\forall \beta \in L^\infty([t_0,T]\times\mathbb{R}^d;\mathbb{R}^d)$ such that $\beta = 0$ in a neighborhood of t_0 and ρ solution in the sense of distributions to (1.16) in $[t_0,T]\times\mathbb{R}^d$

$$\begin{aligned}\widetilde{J}(\beta) = \int_{t_0}^{T} (\int_{\mathbb{R}^d} & L_{\alpha,\alpha}(x,\alpha(t,x))\beta(t,x)\cdot\beta(t,x)m(t,dx)+ \\ & + \Big\langle \frac{\delta F}{\delta m}(\cdot,m(t),\cdot)(\rho(t)),\rho(t)\Big\rangle)dt+ \\ & + \Big\langle \frac{\delta G}{\delta m}(\cdot,m(T),\cdot)(\rho(T,\cdot)),\rho(T,\cdot)\Big\rangle \geq 0.\end{aligned} \tag{3.15}$$

Let $\overline{\beta}$ be a map such that

$$\overline{\beta} = \begin{cases} 0 & \text{in } [t_0,t_1), \\ -H_{pp}(x,Du)Dz & \text{on } [t_1,T], \end{cases}$$

and $\overline{\rho}$ be the solution to (1.16) associated to $\overline{\beta}$.

Thus,

$$\overline{\rho}(t) = \begin{cases} 0 & \text{in } [t_0,t_1)\times\mathbb{R}^d, \\ \mu(t) & \text{on } [t_1,T]\times\mathbb{R}^d. \end{cases}$$

It follows from (3.14) that $\widetilde{J}(\overline{\beta}) = 0$, and, $\overline{\beta}$ is a minimizer for $\widetilde{J}$. It can be proved that [see [2]] that this implies $\overline{\beta}$ is a continuous function and so $Dz(t_1,\cdot) = 0$.

Differentiating with respect to the space variable the first equation in (3.9) we have that $((\partial_{x_k} z)_{k=1,\dots,d},\mu)$ solves a system of the form (3.12) with zero initial condition and data g and h satisfying (3.13). Then a *Lions-Malgrange-type argument* [see Appendix C] implies that $((\partial_{x_k} z)_{k=1,\dots,d},\mu) = (0,0)$. Now we obtain that (z,μ) solution to (3.9) is equal to $(0,0)$ and thus the solution is stable.

□

The next theorem establishes the key property of the set $\mathcal{O}$.

Theorem 3.9. *Assume* (1.1).
The set $\mathcal{O}$ *is open and dense in* $[0,T)\times\mathcal{P}_2(\mathbb{R}^d)$.

Proof. $\mathcal{O}$ is a non-empty dense set in $[0,T)\times\mathcal{P}_2(\mathbb{R}^d)$ for Lemma 3.8.
We want to show that $\mathcal{O}$ is open. We argue by contradiction: fix $(t_0,m_0)\in\mathcal{O}$ and assume that $(t^n,m_0^n)\notin\mathcal{O}$ which converge to (t_0,m_0) in $[0,T]\times\mathcal{P}_2(\mathcal{R}^d)$. Let (m,α) be the unique and stable minimizer for $\mathcal{U}(t_0,m_0)$ and u be the associated multiplier.
Since $(t^n,m_0^n)\notin\mathcal{O}$, there are two cases:

1) $\forall n$, there exist several minimizer for $\mathcal{U}(t^n, m_0^n)$,

2) $\forall n$, there exists a unique minimizer not stable.

2) is ruled out by Lemma 3.2 and the strong stability of (m,α).
It remains to consider 1). Let $(m^{n,1}, \alpha^{n,1})$ and $(m^{n,2}, \alpha^{n,2})$ be two distinct minimizer for $\mathcal{U}(t^n, m_0^n)$ with associated multipliers $u^{n,1}$ and $u^{n,2}$ respectively.
Lemma 1.10 and the fact that the problem with initial condition (t_0, m_0) has a unique minimizer imply that, for $i = 1, 2$,

$$(m^{n,i}, \alpha^{n,i}) \to (m, \alpha) \quad \text{in} \quad \mathcal{C}^0([0,T]; \mathcal{P}_1(\mathbb{R}^d)) \times \mathcal{C}^{\delta/2,\delta},$$

and

$$u^{n,i}, Du^{n,i}, D^2u^{n,i} \to u, Du, D^2u \quad \text{in} \quad \mathcal{C}^{\delta/2,\delta} \quad \text{respectively.}$$

Since $m^{n,1}$ and $m^{n,2}$ are distinct, we have $\sup_t d_1(m^{n,1}(t), m^{n,2}(t)) > 0$ and setting

$$\theta^n = ||Du^{n,1} - Du^{n,2}||_{\mathcal{C}^{\delta/2,\delta}} + \sup_{t \in [t_0,T]} d_1(m^{n,1}(t), m^{n,2}(t)),$$

we have $\theta^n > 0$ and $\theta^n \xrightarrow{n\to\infty} 0$.
Since $\sup_{t\in[t_n,T]} d_1(m^{n,1}(t), m^{n,2}(t)) \leq ||Du^1 - Du^2||_\infty$ [the proof is deducible from Appendix A], $\sup_{t\in[t_n,T]} d_1(m^{n,1}(t), m^{n,2}(t))$ is controlled by $C||Du^{n,1} - Du^{n,2}||_{\mathcal{C}^{\delta/2,\delta}}$ and we have

$$\theta^n \leq C||Du^{n,1} - Du^{n,2}||_{\mathcal{C}^{\delta/2,\delta}}. \tag{3.16}$$

Let us call

$$z^n = \frac{u^{n,1} - u^{n,2}}{\theta^n} \quad \text{and} \quad \mu^n = \frac{m^{n,1} - m^{n,2}}{\theta^n}.$$

Observe that (z^n, μ^n) are solution to

$$\begin{cases} -\partial_t z^n - \Delta z^n + H_p(x, Du^{n,1}) \cdot Dz^n = \frac{\delta F}{\delta m}(x, m^{n,1}(t))(\mu^n(t)) + R^{n,1}, \\ \partial_t \mu^n - \Delta \mu^n - \mathrm{div}(H_p(x, Du^{n,1})\mu^n) - \mathrm{div}(H_{pp}(x, Du^{n,1}) Dz^n m^{n,1}) = \mathrm{div}(R^{2,n}), \\ \mu^n(t_0) = 0 \text{ and } z(T,x) = R^{3,n} + \frac{\delta G}{\delta m}(x, m^{n,1}(T))(\mu^n(T)) \end{cases} \tag{3.17}$$

with

$$\begin{aligned} R^{n,1} =& (\theta^n)^{-1}[(H(x, Du^{n,2}) - H(x, Du^{n,1}) - H_p(x, Du^{n,1}) \cdot (Du^{n,2} - Du^{n,1})) + \\ & - \left(F(x, m^{n,2}(t)) - F(x, m^{n,1}(t)) - \frac{\delta F}{\delta m}(x, m^{n,1}(t))(m^{n,2}(t) - m^{n,1}(t)) \right)], \end{aligned}$$

$$\begin{aligned} R^{n,2} =& -(\theta^n)^{-1}[H_p(x, Du^{n,2})m^{n,2} - H_p(x, Du^{n,1})m^{n,1} - H_p(x, Du^{n,1})(m^{n,2} - m^{n,1}) \\ & - H_{pp}(x, Du^{n,1}) \cdot (Du^{n,2} - Du^{n,1})m^{n,1})], \end{aligned}$$

$$R^{n,3} = -\frac{[G(x, m^{n,2}(T)) - G(x, m^{n,1}(T)) - \frac{\delta G}{\delta m}(x, m^{n,1}(T))(m^{n,2}(T) - m^{n,1}(T))]}{\theta^n}.$$

It follows from the regularity of F, G and H and the definition of θ^n that

$$||R^{n,1}||_{\mathcal{C}^{\delta/2,\delta}} + \sup_{t\in[t_0,T]} ||R^{n,2}||_{(W^{1,\infty})'} + ||R^{n,3}||_{\mathcal{C}^{2+\delta}} \leq C\theta^n. \tag{3.18}$$

For Lemma 3.2, $z^n \to 0$ in $\mathcal{C}^{(1+\delta)/2,1+\delta}$ which is a contradiction with (3.16), so $\mathcal{O}$ is open.

□

3.3 The smoothness of $\mathcal{U}$ in $\mathcal{O}$

We prove now a preliminary lemma that is needed to establish the regularity of $\mathcal{U}$, which will allow us to compute its derivative with respect to m.

Lemma 3.10. *Assume* (1.1).
Fix $(t_0, m_0) \in \mathcal{O}$. *There exist* $\delta, C > 0$ *such that* $\forall t_0', m_0^1, m_0^2$ *satisfying* $(t_0', m_0^i) \in \mathcal{O}$, $|t_0' - t_0| < \delta$, $d_2(m_0, m_0^i) < \delta$, *if* (m^i, α^i) *is the unique minimizer starting from* (t_0', m_0^i) *with associated multiplier* u^i *for* $i = 1; 2$, *then*

$$||u^1 - u^2||_{\mathcal{C}^{(\delta+2)/2,\delta+2}} + \sup_{t\in[t_0',T]} d_2(m^1(t), m^2(t)) \leq C d_2(m_0^1, m_0^2). \tag{3.19}$$

Proof. Let u be the multiplier associated to the unique minimizer (m, α) starting from (t_0, m_0).
We call $V = -H_p(x, Du)$ and $\Gamma = -H_{pp}(x, Du)$ and consider the neighborhood $\mathcal{V}$ of (V, Γ) given in Lemma 3.2. Choosing $\delta > 0$, for any m_0^1 such that $d_2(m_0, m_0^1) < \delta$ letting (m^1, α^1) be the unique stable minimizer starting from (t_0, m_0^1) with multiplier u^1 and calling $V^1 = -H_p(x, Du^1)$ and $\Gamma^1 = -H_{pp}(x, Du^1)$, we have that $(V^1, \Gamma^1) \in \mathcal{V}$.
If $\delta > 0$ is small, the above is possible thanks to Lemma 1.10, since u^1 is close to u in $\mathcal{C}^{\delta/2,\delta}$.
At this point, if it is necessary, choosing δ even smaller, for any $t_0', m_0^i, (m^i, \alpha^i)$ and u^i as above such that $|t_0^i - t_0| < \delta$ and $d_2(m_0, m_0^i) < \delta$ for $i = 1, 2$, for some $\eta > 0$ that will be chosen later

$$||u^1 - u^2||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0',T]} d_2(m^1(t), m^2(t)) \leq \eta. \tag{3.20}$$

For $t_0', m_0^i, (m^i, \alpha^i)$ and u^i as above, we have [see Appendix A] for a constant C depending on T, H and $||D^2u^i||_\infty$ which is uniformly bounded by Lemma 3.4

$$\sup_{t\in[t_0',T]} d_2(m^1(t), m^2(t)) \leq C(d_2(m_0^1, m_0^2) + ||Du^1 - Du^2||_\infty). \tag{3.21}$$

Then, the pair

$$(z,\mu) = (u^1 - u^2, m^1 - m^2)$$

satisfies the system (3.2) with

$$V(t,x) = -H_p(x, Du^2),$$

$$\Gamma = -H_{pp}(x, Du^2),$$

$$\xi = m_0^1 - m_0^2,$$

$$\begin{aligned} R^1(t,x) =& -\left(H(x,Du^1) - H(x,Du^2) - H_p(x,Du^2)\cdot(Du^1 - Du^2)\right) \\ &+ F(x,m^1) - F(x,m^2) - \frac{\delta F}{\delta m}(x, m^2(t))(m^1(t) - m^2(t)), \\ R^2(t,x) =& H_p(x,Du^1)m^1 - H_p(x,Du^2)m^2 - H_p(x,Du^2)(m^1 - m^2) \\ &- H_{pp}(x,Du^2)\cdot(Du^1 - Du^2)m^2, \end{aligned}$$

$$R^3(x) = G(x, m^1(T)) - G(x, m^2(T)) - \frac{\delta G}{\delta m}(x, m^2(T))(m^1(T) - m^2(T)).$$

Note that we can rewrite R^2 as

$$\begin{aligned} R^2(t,x) =& (H_p(x,Du^1) - H_p(x,Du^2))(m^1 - m^2) \\ &+ (H_p(x,Du^1) - H_p(x,Du^2) - H_{pp}(x,Du^2)\cdot(Du^1 - Du^2))m^2. \end{aligned}$$

Thus,

$$\begin{aligned} M =& ||\xi||_{(W^{1,\infty})'} + ||R^1||_{\mathcal{C}^{\delta/2,\delta}} + ||R^3||_{\mathcal{C}^{2+\delta}} + \sup_{t\in t_0,T]} ||R^2(t)||_{(W^{1,\infty})'} \\ &\leq d_1(m_0^1, m_0^2) + C\{||Du^1 - Du^2||^2_{\mathcal{C}^{\delta/2,\delta}} + \sup_t d_2^2(m^1(t), m^2(t))\}. \end{aligned}$$

It follows from Lemma 3.2 that

$$||u^1 - u^2||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} \leq C\{d_1(m_0^1, m_0^2) + ||Du^1 - Du^2||^2_{\mathcal{C}^{\delta/2,\delta}} + \sup_t d_2^2(m^1(t), m^2(t))\}.$$

Hence, choosing $\eta > 0$ small enough we find

$$||u^1 - u^2||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} \leq C\{d_1(m_0^1, m_0^2) + \sup_t d_2^2(m^1(t), m^2(t))\},$$

and inserting the last inequality in (3.21) we obtain

$$\sup_t d_2(m^1(t), m^2(t)) \leq C\{d_2(m_0^1, m_0^2) + \sup_t d_2^2(m^1(t), m^2(t))\}.$$

Noting that $\sup_t d_2(m^1(t), m^2(t)) \leq \eta \sup_t d_2(m^1(t), m^2(t))$ then, for $\eta > 0$ such that $1 - C\eta > 0$, we have

$$\sup_t d_2(m^1(t), m^2(t)) \leq C[d_2(m_0^1, m_0^2) + \eta \sup_t d_2(m^1(t), m^2(t))],$$

and so

$$\sup_t d_2(m^1(t), m^2(t)) \leq C d_2(m_0^1, m_0^2).$$

Going back to the previous inequality on $||u^1 - u^2||_{\mathcal{C}^{(2+\delta)/2, 2+\delta}}$, since $d_1 < d_2$ we can conclude the proof.

□

Here we state a simple criterion for the differentiability of the function $\mathcal{U}$.

Lemma 3.11. *Let $\mathcal{U} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be continuous. For $(s, m, y) \in [0,1] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$, we define*

$$\hat{\mathcal{U}}(s, m, y) := \mathcal{U}((1-s)m + s\delta_y).$$

If the map $s \to \hat{\mathcal{U}}(s, m, y)$ has a derivative at $s = 0$ and if its derivative at 0 $\frac{d}{ds}|_{s=0}\hat{\mathcal{U}} : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is continuous and bounded, then $\mathcal{U}$ is of class $\mathcal{C}^1$ with

$$\frac{\delta \mathcal{U}}{\delta m}(m, y) = \frac{d}{ds}\hat{\mathcal{U}}(0, m, y).$$

Remark 3.12. *Let us note that, replacing m by $(1-s)m + s\delta_y$, the continuity assumption of $\frac{d}{ds}|_{s=0}\hat{\mathcal{U}}$ implies its continuity at any $s \in [0,1]$.*

Proof. Let us start by considering the case where m_0 is fixed and, for some $N \in \mathbb{N}$, $N \geq 1$ and $y_k \in \mathbb{R}^d$, $m_1 = \frac{1}{N}\sum_{k=1}^N \delta_{y_k}$ is an empirical measure.
Let us define the set

$$K := \{\alpha_0 m_0 + \sum_{k=1}^N \alpha_k \delta_{y_k} : \alpha_k \geq 0, \sum_{k=0}^N \alpha_k = 1\}.$$

All the measures we will use belong to K, which is compact in $\mathcal{P}_2(\mathbb{R}^d)$.
Since the map $\frac{d}{ds}\hat{\mathcal{U}}$ is continuous, for $\epsilon > 0$, there exist $\delta \in (0, 1/2)$ such that, for $m, m' \in K$ so that $d_2(m, m') < \delta$ and $s \in [0, \delta]$, we have

$$\sup_k |\frac{d}{ds}\hat{\mathcal{U}}(s, m, y_k) - \frac{d}{ds}\hat{\mathcal{U}}(0, m', y_k)| \leq \epsilon. \tag{3.22}$$

Let us define

$$\alpha_k = \frac{s}{N - (N-k)s} \quad \text{for} \quad k = 0, \dots, N,$$

and noting that $1-\alpha_k = \frac{N-(N-k+1)s}{N-(N-k)s}$, we get

$$\begin{aligned}\prod_{l=k}^{N}(1-\alpha_l) &= \frac{N-(N-k+1)s}{N-(N-k)s}\cdot\frac{N-(N-k)s}{N-(N-k-1)s}\cdots\frac{N-s}{N} \\ &= \frac{N-(N-k+1)s}{N} = 1-\frac{(N-k+1)s}{N}.\end{aligned} \tag{3.23}$$

We now define by induction

$$m_0 := m \quad \text{and} \quad m_k := (1-\alpha_k)m_{k-1} + \alpha_k\delta_{y_k}, \tag{3.24}$$

and using (3.23) we get

$$\begin{aligned} m_N &= \prod_{k=1}^{N}(1-\alpha_k)m + \alpha_n\delta_{y_N} + \sum_{k=1}^{N-1}\alpha_k\delta_{y_k}\prod_{l=k+1}^{N}(1-\alpha_l) \\ &(1-s)m + \sum_{k=1}^{N}\delta_{y_k}\frac{s}{N-(N-k)s}\left(1-\frac{(N-k)s}{N}\right) = (1-s)m + sm_y^N. \end{aligned} \tag{3.25}$$

So, by definition of m_{k+1} in function of m_k in (3.24),

$$\begin{aligned} &\mathcal{U}((1-s)m + sm_y^N) - \mathcal{U}(m) = \sum_{k=0}^{N-1}(\mathcal{U}(m_{k+1}) - \mathcal{U}(m_k)) \\ &= \sum_{k=0}^{N-1}\hat{\mathcal{U}}(\alpha_{k+1}, m_k, y_{k+1}) - \hat{\mathcal{U}}(0, m_k, y_{k+1}) = \sum_{k=0}^{N-1}\int_0^{\alpha_{k+1}}\frac{d}{ds}\hat{\mathcal{U}}(\tau, m_k, y_{k+1})d\tau. \end{aligned}$$

Let us assume that $s \in (0,\delta)$. Since $s < 1/2$, we get for any k $\alpha_k \le 2s/N$, and thus there exists a constant $C = C(m_0, y_k) > 0$ such that

$$d_2(m_k, m) \le Cs$$

we now require that s is so small that $Cs < \delta$. Using (3.22), $\forall k$ and $\forall \tau \in (0,\alpha_k)$, we get

$$\left|\frac{d}{ds}\hat{\mathcal{U}}(\tau, m_k, y_{k+1}) - \frac{d}{ds}\hat{\mathcal{U}}(0, m, y_{k+1})\right| \le \epsilon.$$

We deduce from this

$$\left|\mathcal{U}((1-s)m + sm_y^N) - \mathcal{U}(m) - \sum_{k=0}^{N-1}\alpha_{k+1}\frac{d}{ds}\hat{\mathcal{U}}(0, m, y_{k+1})\right| \le C\epsilon\sum_{k=0}^{N-1}\alpha_{k+1}.$$

Since $|\alpha_k - \frac{s}{N}| \leq \frac{Cs^2}{N}$, we conclude that

$$|\mathcal{U}((1-s)m + sm_y^N) - \mathcal{U}(m) - s\int_{\mathbb{R}^d} \frac{d}{ds}\hat{\mathcal{U}}(0,m,y)m_y^N(dy)| \leq C(\epsilon s + s^2). \quad (3.26)$$

At this point, let $T \in \mathbb{N}$ be large and we define, for $n \in \{0, \dots, T\}$

$$m_n = \left(1 - \frac{1}{N}\right)^n m_0 + \left(1 - \left(1 - \frac{1}{T}\right)^n\right) m_y^N,$$

and so

$$\begin{aligned} m_{n+1} &= \left(1 - \frac{1}{N}\right)^{n+1} m_0 + \left(1 - \left(1 - \frac{1}{T}\right)^{n+1}\right) m_y^N \\ &= \left(1 - \frac{1}{T}\right)[\left(1 - \frac{1}{T}\right)^n m_0 - \left(1 - \frac{1}{T}\right)^n m_y^N] + m_y^n \pm \frac{1}{T} m_y^N \\ &= \left(1 - \frac{1}{T}\right) m_n + \frac{1}{T} m_y^N. \end{aligned}$$

So, by (3.26)

$$\begin{aligned} &|\mathcal{U}(m_T) - \mathcal{U}(m_0) - T^{-1}\sum_{n=0}^{T-1}\int_{\mathbb{R}^d}\frac{d}{ds}\hat{\mathcal{U}}(0,m_n,y)m_y^N(dy)| \\ &\leq \sum_{n=0}^{T-1}|\mathcal{U}\left(\left(1-\frac{1}{T}\right)m_n + \frac{1}{T}m_y^N\right) - \mathcal{U}(m_n) - \frac{1}{T}\int_{\mathbb{R}^d}\frac{d}{ds}\hat{\mathcal{U}}(0,m_n,y)m_y^N(dy)| \\ &\leq C\sum_{n=0}^{T-1}\left(\frac{\epsilon}{T} + \left(\frac{1}{T}\right)^2\right) \leq C\left(\epsilon + \frac{1}{T}\right). \end{aligned}$$

Recalling that $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$, letting $T \to \infty$ and then $\epsilon \to 0$ by continuity of $\mathcal{U}$ and of $\frac{d}{ds}\hat{\mathcal{U}}$ we can conclude that

$$\begin{aligned} \mathcal{U}(e^{-1}m_0 + (1-e^{-1})m_y^N) - \mathcal{U}(m_0) &= \int_0^1\int_{\mathbb{R}^d}\frac{d}{ds}\hat{\mathcal{U}}(0,e^{-s}m_0 + (1-e^{-s})m_y^N,y)m_y^N(dy)ds \\ &= \int_0^{e^{-1}}\int_{\mathbb{R}^d}\frac{d}{ds}\hat{\mathcal{U}}(0,(1-\tau)m_0 + \tau m_y^N,y)m_y^N(dy)\frac{d\tau}{1-\tau}. \end{aligned} \quad (3.27)$$

By the continuity of $\mathcal{U}$ and of $\frac{d}{ds}\hat{\mathcal{U}}$ and by the density of the empirical measures, from (3.27), $\forall m_0, m_1 \in \mathcal{P}_2(\mathbb{R}^d)$, we obtain

$$\mathcal{U}(e^{-1}m_0 + (1-e^{-1}m_1) - \mathcal{U}(m_0) = \int_0^{e^{-1}}\int_{\mathbb{R}^d}\frac{d}{ds}\hat{\mathcal{U}}(0,(1-\tau)m_0 + \tau m_1,y)m_1(dy)\frac{d\tau}{1-\tau}. \quad (3.28)$$

For $m_1 = m_0$, we have

$$\int_{\mathbb{R}^d} \frac{d}{ds}\hat{\mathcal{U}}(0, m_0, y)m_0(dy) = 0 \quad \text{for any } m_0 \in \mathcal{P}_2(\mathbb{R}^d).$$

In particular, this yields

$$\int_{\mathbb{R}^d} \frac{d}{ds}\hat{\mathcal{U}}(0, (1-\tau)m_0+\tau m_1, y)m_1(dy) = (1-\tau)\int_{\mathbb{R}^d} \frac{d}{ds}\hat{\mathcal{U}}(0, (1-\tau)m_0+\tau m_1, y)(m_1-m_0)(dy).$$

Using this relation in (3.28) we get

$$\mathcal{U}(e^{-1}m_0+(1-e^{-1})m_1)-\mathcal{U}(m_0) = \int_0^{e^{-1}}\int_{\mathbb{R}^d} \frac{d}{ds}\hat{\mathcal{U}}(0, (1-\tau)m_0+\tau m_1, y)(m_1-m_0)(dy)d\tau.$$

Using again the continuity of $\mathcal{U}$ and of $\frac{d}{ds}\hat{\mathcal{U}}$, one can deduce from this the desired equality.

□

Theorem 3.13. *Assume (1.1).*
The value function $\mathcal{U}$ is globally Lipschitz continuous on $[0,T]\times\mathcal{P}_1(\mathbb{R}^d)$ and of class $\mathcal{C}^1$ in the set $\mathcal{O}$. In addition, $\mathcal{U}$ is a classical solution in $\mathcal{O}$ of the master Hamilton-Jacobi equation

$$-\partial_t\mathcal{U}(t,m)-\int_{\mathbb{R}^d}\operatorname{div}(D_m\mathcal{U}(t,m,y))m(dy)+\int_{\mathbb{R}^d}H(y,D_m\mathcal{U}(t,m,y))m(dy) = \mathcal{F}(m). \tag{3.29}$$

Moreover, $\forall (t_0,m_0)\in\mathcal{O}$, $\exists\ \epsilon>0$ and a constant $C=C(t_0,m_0)>0$ so that, $\forall t\in[0,T]$, x,$y\in\mathbb{R}^d$ and m^1,$m^2\in\mathcal{P}_2(\mathbb{R}^d)$ with $|t-t_0|<\epsilon$, $d_2(m_0,m^1)<\epsilon$ and $d_2(m_0,m^2)<\epsilon$

$$|D_m\mathcal{U}(t,m^1,x)-D_m\mathcal{U}(t,m^2,y)| \le C(|x-y|+d_2(m^1,m^2)). \tag{3.30}$$

Proof. Let us divide the proof in three parts.

First part: The regularity of $\mathcal{U}$

By Lemma 2.2, $\mathcal{U}$ is Lipschitz continuous. We want to show that $\mathcal{U}$ is differentiable at any $(t_0,m_0)\in\mathcal{O}$. Let us fix $(t_0,m_0)\in\mathcal{O}$ and be u the associated multiplier to (m,α) unique minimizer for $\mathcal{U}(t_0,m_0)$.
Let $\delta>0$ and $0<\delta'<\delta$ be such that the δ-neighborhood of the $[0,T]\times\mathcal{P}_2(\mathbb{R}^d)$-compact set $\{(t,m(t)) : t\in[t_0,T]\}\subset\mathcal{O}$, and, for any $m_0^1\in B(m_0,\delta')$, $\sup_{t\in[t_0,T]} d_1(m(t),m^1(t))<\delta$, where u^1 is the associated minimizer to (m^1,α^1) minimizer for $\mathcal{U}(t_0,m_0^1)$.

Let (z,μ) the solution to the linearized system (3.9) with initial condition $\mu(0) = m_0^1 - m_0$.
Set

$$(w,\rho) = (u^1 - u - z, m^1 - m - \mu), \tag{3.31}$$

and it is the solution of the system (3.2) with

$$\xi = 0$$

$$\begin{aligned}
R^1(t,x) =& -\left(H(x,Du^1) - H(x,Du) - H_p(x,Du)\cdot(Du^1 - Du)\right)+\\
&+ F(x,m^1) - F(x,m) - \frac{\delta F}{\delta m}(x,m(t))(m^1(t) - m(t)),\\
R^2(t,x) =& H_p(x,Du^1)m^1 - H_p(x,Du)m - H_p(x,Du)(m^1 - m)+\\
&- H_{pp}(x,Du)\cdot(Du^1 - Du)m,\\
R^3(x) =& G(x,m^1(T)) - G(x,m(T)) - \frac{\delta G}{\delta m}(x,m(T))(m^1(T) - m(T)).
\end{aligned}$$

We can rewrite $R^1(t,x)$ in the following way $R^1(t,x) = r_1^1(t,x) + r_2^1(t,x)$ with

$$r_1^1(t,x) = -\int_0^1 (H_p(x, sDu^1 + (1-s)Du) - H_p(x,Du))\cdot(Du^1 - Du)ds,$$

and

$$\begin{aligned}
r_2^1(t,x) =& \int_0^1\int_{\mathbb{R}^d}\left(\frac{\delta F}{\delta m}(x, sm^1(t) + (1-s)m(t), y\right) d(m^1(t) - m(t))(y)ds\\
&- \int_0^1\int_{\mathbb{R}^d}\left(\frac{\delta F}{\delta m}(x,m(t),y)\right) d(m^1(t) - m(t))(y)ds.
\end{aligned}$$

We have

$$||R^1||_{\mathcal{C}^{\delta/2,\delta}} \le ||r_1^1||_{\mathcal{C}^{\delta/2,\delta}} + ||r_2^1||_{\mathcal{C}^{\delta/2,\delta}}.$$

Since

$$||r_1^1||_{\mathcal{C}^{\delta/2,\delta}} \le C||u^1 - u||^2_{\mathcal{C}^{\delta/2,\delta}}$$

and

$$||r_2^1||_{\mathcal{C}^{\delta/2,\delta}} \le ||D_y\frac{\delta F}{\delta m}(\cdot,m^1,\cdot) - D_y\frac{\delta F}{\delta m}(\cdot,m,\cdot)||_{\mathcal{C}^{\delta/2,\delta}} d(m^1,m) \le Cd_1^2(m^1,m)$$

by Lemma 3.10 and the fact that $d_1 < d_2$

$$||R^1||_{\mathcal{C}^{\delta/2,\delta}} \le Cd_2^2(m_0^1, m_0).$$

Noting that

$$H_p(x, Du^1)m^1 - H_p(x, Du)m = [H_p(x, Du^1) - H_p(x, Du)]m^1 + H_p(x, Du^1)(m^1 - m),$$

and

$$[H_p(x, Du^1) - H_p(x, Du)]m^1 + H_{pp}(x, Du)(Du^1 - Du)m =$$
$$m^1 \int_0^1 H_{pp}(x, sDu^1 + (1-s)Du) \cdot (Du^1 - Du)ds - H_{pp}(x, Du)(Du^1 - Du)m.$$

We can rewrite

$$\begin{aligned} R^2(t,x) =& m^1 \int_0^1 \big(H_{pp}(x, sDu^1 + (1-s)Du) - H_{pp}(x, Du)\big) \cdot (Du^1 - Du)ds \\ &+ H_{pp}(x, Du)(Du^1 - Du)(m^1(t) - m(t)). \end{aligned}$$

By definition

$$\sup_{t\in[t_0,T]} ||R^2(t)||_{(W^{1,\infty})'} \le \sup_{t\in[t_0,T]} \left(\sup_{||\xi||_{W^{1,\infty}}\le 1} \langle \xi, R^2(t)\rangle_{W^{1,\infty},(W^{1,\infty})'} \right)$$

and

$$\begin{aligned} &\langle \xi, R^2(t)\rangle_{W^{1,\infty},(W^{1,\infty})'} = \int_{\mathbb{R}^d} \langle \xi, (m^1 - m)H_{pp}(\cdot, Du(t,\cdot))(Du^1 - Du)(t,\cdot) \\ &+ m^1(t) \int_0^1 \big(H_{pp}(\cdot, [sDu^1 + (1-s)Du](t,\cdot)\big)(Du^1 - Du)(t,\cdot)ds \\ &- m^1(t) \int_0^1 \left(H_{pp}(\cdot, Du(t,\cdot))\right)(Du^1 - Du)(t,\cdot)ds\rangle_{W^{1,\infty},(W^{1,\infty})'} \\ &\le C(||\xi||_{W^{1,\infty}}||u^1 - u||_2 d_1(m^1, m) + ||\xi||_{W^{1,\infty}}||u^1 - u||_1^2). \end{aligned}$$

Finally, using Lemma 3.10 and the previous estimates

$$\sup_{t\in[t_0,T]} ||R^2(t)||_{(W^{1,\infty})'} \le Cd_2^2(m_0^1, m_0).$$

Lastly, rewriting R^3 as

$$\begin{aligned} R^3(x) = &\int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta G}{\delta m}(x, sm^1(T) + (1-s)m(T), y)\right) d(m^1(T) - m(T))(y)ds \\ &- \int_0^1 \int_{\mathbb{R}^d} \left(\frac{\delta G}{\delta m}(x, m(T), y)\right) d(m^1(T) - m(T))(y)ds, \end{aligned}$$

and noting that

$$\begin{aligned}||R^3||_{\mathcal{C}^{2+\delta}} &\le ||D_y \frac{\delta G}{\delta m}(\cdot, m^1(T), \cdot) - D_y \frac{\delta G}{\delta m}(\cdot, m(T), \cdot)||_{\mathcal{C}^{2+\delta}} d_1^2(m^1(T), m(T)) \\ &\le C d_1^2(m^1, m),\end{aligned}$$

using again Lemma 3.10, we obtain that

$$||R^1||_{\mathcal{C}^{\delta/2,\delta}} + \sup_{t\in[t_0,T]} ||R^2(t)||_{(W^{1,\infty})'} + ||R^3||_{\mathcal{C}^{2+\delta}} \le C d_2^2(m_0^1, m_0).$$

Then, for Lemma 3.2 and noting that $z = u^1 - u - w$ and $\rho = m^1 - m - \mu$

$$||u^1 - u - w||_{\mathcal{C}^{(2+\delta)/2,2+\delta}} + \sup_{t\in[t_0,T]} ||m^1(t) - m(t) - \mu(t)||_{(\mathcal{C}^{2+\delta})'} \le C d_2^2(m_0^1, m_0).$$

Recall that $\alpha^1 = -H_p(x, Du^1)$. Thus

$$\alpha^1 = \alpha - H_{pp}(x, Du) \cdot Dw + o(d_1(m_0^1, m_0)),$$

where $o(\cdot)$ is small in uniform norm.

It follows that

$$\begin{aligned}\mathcal{U}(t_0, m_0^1) &= \int_{t_0}^T \left(\int_{\mathbb{R}^d} L(x, \alpha^1) m^1 + \mathcal{F}(m^1) \right) dt + \mathcal{G}(m^1(T)) \\ &= \mathcal{U}(t_0, m_0) + \int_{t_0}^T \int_{\mathbb{R}^d} (L_\alpha(x,\alpha) \cdot [-H_{pp}(x, Du)] Dwm + L(x,\alpha)\mu(t,x) + \\ &+ F(x, m(t))\mu(t,x)) dx dt + \int_{\mathbb{R}^d} G(x, m(T))\mu(T,x) dx + o(d_2(m_0^1, m_0)).\end{aligned}$$

Using duality on the equations satisfied by u and μ

$$\begin{aligned}&\int_{\mathbb{R}^d} G(x, m(T))\mu(T,x) dx - \int_{\mathbb{R}^d} u(t_0, x)(m_0^1 - m_0)(dx) = \\ &= \int_{t_0}^T \int_{\mathbb{R}^d} ((H(x, Du) - F(x, m(t)) - H_p(x, Du) \cdot Du)\mu - H_{pp}(x, Du) Du \cdot Dwm).\end{aligned}$$

Thus,

$$\begin{aligned}&\mathcal{U}(t_0, m_0^1) = \mathcal{U}(t_0, m_0) + \int_{\mathbb{R}^d} u(t_0, x)(m_0^1 - m_0)(dx) \\ &+ \int_{t_0}^T \int_{\mathbb{R}^d} ((H(x, Du) - H_p(x, Du) \cdot Du + L(x,\alpha))\mu - (H_{pp}(x, Du) Du \cdot Dw \\ &+ L_\alpha(x,\alpha) \cdot (H_{pp}(x, Du) Dw)) m) dx dt + o(d_2(m_0^1, m_0)).\end{aligned}$$

Since $\alpha = -H_p(x, Du)$, in view of the definition of H and L we have

$$H(x, Du) - H_p(x, Du) \cdot Du + L(x, \alpha) = 0,$$

and knowing that $L_\alpha(x, \alpha) = -Du$ we get

$$H_{pp}(x, Du)Du \cdot Dw + L_\alpha(x, \alpha) \cdot (H_{pp}(x, Du)Dw) = 0.$$

Combining all these results, we obtain

$$\mathcal{U}(t_0, m_0^1) = \mathcal{U}(t_0, m_0) + \int_{\mathbb{R}^d} u(t_0, x)(m_0^1 - m_0)(dx) + o(d_2(m_0^1, m_0)).$$

Rewriting m_0^1 as $m_0^1 = (1-s)m_0 + s\delta_y$ for some $y \in \mathbb{R}^d$ and $s \in (0,1)$ and using it in the above equality we get that the following limit exists:

$$u(t_0, y) = \lim_{s \to 0^+} \frac{1}{s}(\mathcal{U}(t_0, (1-s)m_0 + s\delta_y) - \mathcal{U}(t_0, m_0)).$$

This limit depends in a continuous way on (m_0, y) since the stability of the map $m_0 \to (u, m)$ proved in Lemma 1.10.
We have that

$$D_m U(t_0, m_0, x) = Du(t_0, x),$$

since, in view of Lemma 3.11, $\mathcal{U}(t_0, \cdot)$ has a linear derivative in a neighborhood of m_0 given by $u(t_0, \cdot)$.
Using again the stability of the map $m_0 \to (u, m)$, we actually have that $(t_0, m_0) \to D_m U(t_0, m_0, \cdot)$ is continuous in $\mathcal{O}$ with respect to the d_2-distance for the measure variable into $\mathcal{C}^2$.

Second part: The Hamilton-Jacobi equation

Here we shows that $\mathcal{U}$ is a classical solution to (3.29).
Let $h > 0$ small. The dynamic programming principle gives

$$\mathcal{U}(t_0, m_0) = \int_{t_0}^{t_0+h} \left(\int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t,x) dx + \mathcal{F}(m(t)) \right) dt + \mathcal{U}(t_0+h, m(t_0+h)),$$

and, since $\mathcal{U}$ is $\mathcal{C}^1$,

$$\mathcal{U}(t_0+h, m(t_0+h)) - \mathcal{U}(t_0+h, m_0) = \int_{t_0}^{t_0+h} \int_{\mathbb{R}^d} (Tr D^2_{ym} \mathcal{U}(t_0+h, m(t), y) + \\ + D_m \mathcal{U}(t_0+h, m(t), y) \cdot H_p(t, Du(t,y))) m(t, dy) dy dt.$$

It follows that $\partial_t \mathcal{U}(t_0, m_0)$ exists and is given by

$$\partial_t \mathcal{U}(t_0, m_0) = -\int_{\mathbb{R}^d} L(x, \alpha(t_0, x)) m_0(dx) - \mathcal{F}(m_0) +$$
$$- \int_{\mathbb{R}^d} (Tr D^2_{ym} \mathcal{U}(t_0, m_0, y) + D_m \mathcal{U}(t_0, m_0, y) \cdot H_p(t_0, Du(t_0, y))) m_0(dy).$$

Since we proved $D_m U(t_0, m_0, x) = Du(t_0, x)$ and we used $\alpha(t_0, x) = -H_p(x, Du(t_0, x))$, by the definition of H with the Legendre transform and knowing that the divergence of $D_m \mathcal{U}$ is the trace of $D^2_{ym}\mathcal{U}$, (3.29) is satisfied.

Third part: The regularity of $D_m \mathcal{U}$

Let $\delta > 0$ and $C > 0$ be such that for any t, m_0^1, m_0^2 satisfying $(t, m_0^i) \in \mathcal{O}$, $|t - t_0| < \delta$, $d_2(m_0, m_0^i) < \delta$ and for any $x^1, x^2 \in \mathbb{R}^d$, let (m^i, α^i) the unique minimizer for $\mathcal{U}(t, m_0^i)$ with associated multiplier u^i for $i = 1, 2$ and

$$||u^1 - u^2||_{\mathcal{C}^{(\delta+2)/2, \delta+2}} + \sup_{t \in [t_0', T]} d_2(m^1(t), m^2(t)) \leq C d_2(m_0^1, m_0^2).$$

Since $D_m U(t, m_0^1, x^1) = Du^1(t, x^1)$ and $D_m U(t, m_0^2, x^2) = Du^2(t, x^2)$

$$\begin{aligned} &|D_m U(t, m_0^1, x^1) - D_m U(t, m_0^2, x^2)| = |Du^1(t, x^1) - Du^2(t, x^2)| \\ &\leq |Du^1(t, x^2) - Du^2(t, x^2)| + |Du^1(t, x^1) - Du^1(t, x^2)| \\ &\leq |Du^1(t, x^2) - Du^2(t, x^2)| + ||D^2 u^1||_\infty |x^1 - x^2| \leq C(d_2(m_0^1, m_0^2) + |x^1 - x^2|), \end{aligned}$$

where we used lemmas 1.9 and 3.10.

□

Chapter 4

The propagation of chaos

The aim of this chapter is to show the following **propagation of chaos** property.

Theorem 4.1. *Assume* (1.1).
There exists a constant $\gamma = \gamma(d) \in (0,1)$ *such that,* $\forall (t_0, m_0) \in \mathcal{O}$ *with* $M_{d+5}(m_0) < +\infty$*, there is a constant* $C = C(t_0, m_0) > 0$*, such that, if* $\mathbf{Z} = (Z^k)_{k=1,\dots,N}$ *is a sequence of independent random variables with law* m_0*,* $\mathbf{B} = (B^k)_{k=1,\dots,N}$ *is a sequence of independent Brownian motions independent of* $\mathbf{Z}$*, and* $\mathbf{Y}^N = (Y^{N,k})_{k=1,\dots,N}$ *is the optimal trajectory of* $\mathcal{V}^N(t_0, (Z^k)_{k=1,\dots,N})$*, that is, for each* $k = 1, \dots, N$ *and* $t \in [t_0, T]$

$$Y_t^{N,k} = Z^k - \int_{t_0}^t H_p(Y_s^k, D\mathcal{V}^N(s, Y_s^N))ds + \sqrt{2}(B_t^k - B_{t_0}^k). \tag{4.1}$$

Then

$$\mathbb{E}[\sup_{t\in[t_0,T]} d_1(m^N_{Y_t^N}, m(t))] \leq CN^{-\gamma}.$$

Remark 4.2. *We note that this theorem exploit the regularity of the function* $\mathcal{U}$ *in* $\mathcal{O}$ *obtained in the results of chapter 3 to get a convergence of the optimal trajectories of the* N *particle system to the one of the limit problem.*
In chapter 2 we show a different kind of convergence, one more variational: the convergence of the two value functions that holds not only in $\mathcal{O}$*, but in* $[0,T) \times \mathcal{P}_2(\mathbb{R}^d)$.

4.1 Introduction

For large particle systems, the notion of propagation of chaos was initially introduced by Boltzmann in statistical physics: the idea is that the correlations between two (or more) given particles for large systems, which are due to the interactions become negligible. It means that only an averaged behaviour can be observed instead of the detailed correlated trajectories of each particle.

The mathematical formalisation is due to Kac and McKean and dates back to the 20^{th} century. It has recently spread out in many areas of mathematics.

The propagation of chaos property describes the limit behaviour of the particle system when the number of particles grows to infinity: any subsystem of the N-particle system asymptotically behaves as a system of i.i.d processes with common law m.

Definition 4.3. *Let E a separable metric space, u_N a sequence of symmetric probabilities on E^N. We say that u_N is u-*chaotic*, u probability on E, if for $\phi_1, \ldots, \phi_k \in \mathcal{C}_b(E)$, $k \geq 1$,*

$$\lim_{N\to\infty} \langle u_N, \phi_1 \otimes \cdots \otimes \phi_k \otimes 1 \otimes \cdots \otimes 1\rangle = \prod_{i=1}^{k} \langle u, \phi_i \rangle, \tag{4.2}$$

where we use $\otimes$ for the product measure.

The notation of u-chaotic means that the empirical measures of the coordinate variables of E^N, under u_N tend to concentrate near u, as the next proposition shows.
Condition (4.2) can also be restated as the convergence of the projection of u_N as E^k to $u^{\otimes k}$ when N goes to infinity. In the next proposition, we suppose u_N symmetric.

Proposition 4.4. *u_N is u-chaotic is equivalent to $\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$ converges in law to the constant random variable u.*

4.2 The propagation of chaos property

Let $\mathbf{Z} = (Z^k)_{k=1,\ldots,N}$ be a sequence of independent random variables with law m_0, $\mathbf{B} = (B^k)_{k=1,\ldots,N}$ a sequence of independent Brownian motions independent of $\mathbf{Z}$, and $\mathbf{Y}^N = (N^{N,k})_{k=1,\ldots,N}$ the optimal trajectory of $\mathcal{V}^N(t_0, (Z^k)_{k=1,\ldots,N})$, that is, for each $k = 1, \ldots, N$ and $t \in [t_0, T]$

$$Y_t^{N,k} = Z^k - \int_{t_0}^{t} H_p(Y_s^k, D\mathcal{V}^N(s, Y_s^N))ds + \sqrt{2}(B_t^k - B_{t_0}^k). \tag{4.3}$$

It follows from the Theorem 3.13 and the compactness of the curve $\{(t, m(t)) : t \in [t_0, T]\}$ that there exists δ, $C > 0$ such that, for any $t_1 \in [t_0, T]$, $|t - t_1| < \delta$

and m_0^1, $m_0^2 \in \mathcal{P}_2(\mathbb{R}^d)$ with $d_2(m(t_1), m_0^1) < \delta$, $d_2(m(t_1), m_0^2) < \delta$, $(t_1, m_0^1) \in \mathcal{O}$ and $(t_1, m_0^2) \in \mathcal{O}$, and x^1, $x^2 \in \mathbb{R}^d$

$$|D_m\mathcal{U}(t_1, m_0^1, x^1) - D_m\mathcal{U}(t_1, m_0^2, x^2)| \leq C(|x^1 - x^2| + d_2(m_0^1, m_0^2)) \tag{4.4}$$

Definition 4.5. *Let $\sigma \in (0, \delta)$. We define*

$$V_\sigma = \{(t, m') \in [t_0, T] \times \mathcal{P}_2(\mathbb{R}^d) : d_2(m', m(t)) < \sigma\}$$

and

$$V_\sigma^N = \{(t, x) \in [0, T] \times \mathbb{R}^d : (t, m_X^N) \in V_\sigma\}.$$

Definition 4.6. *The stopping time τ^N is defined by*

$$\tau^N = \begin{cases} \inf_{t\in[t_0,T]}\{(t, X_t^N) \notin V_{\delta/2}^N\} \\ T \text{ if there is no such a } t, \end{cases}$$

and

$$\widetilde{\tau}^N = \begin{cases} \inf_{t\in[t_0,\tau^N]}\{(t, Y_t^N) \notin V_\delta^N\} \\ \tau^N \text{ if there is no such a } t. \end{cases}$$

We consider the solution $(\mathbf{X}_t^N)_{t\in[t_0,T]} = (X_t^{N,1}, \dots, X_t^{N,N})_{t\in[t_0,T]}$ to

$$dX_t^{N,j} = Z^j - \int_{t_0}^t H_p(X_s^{N,j}, D_m\mathcal{U}(s, m_{X_s^N}^N, X_s^{N,j}))ds + \sqrt{2}(B_t^j - B_{t_0}^j) \text{ on } [t_0, \tau^N]. \tag{4.5}$$

Theorem 4.7. *Let $M_{d+5}(m_0) < \infty$, then there exists a constant $C > 0$ such that*

$$\mathbb{E}[\sup_{t\in[0,T]} d_2^2(m_{X_t^N}^N, m(t))] \leq CN^{-\frac{2}{d+8}}.$$

Proof. Let us construct i.i.d. copies of the unique solutions to

$$d\widetilde{X}_t^{N,i} = -H_p(\widetilde{X}_t^{N,i}, D_m\mathcal{U}(t, m(t), \widetilde{X}_t^{N,i}))dt + \sqrt{2}dB_t^i \quad \text{and} \quad \widetilde{X}_{t_0}^{N,i} = Z^i,$$

with $\widetilde{\mu}(t)$ law of $\widetilde{X}_t^{N,i}$. So,

$$\begin{aligned} |X_t^{N,i} - \widetilde{X}_t^{N,i}| &\leq \int_{t_0}^T |H_p(s, X_s^{N,i}, m_{X_s}) - H_p(s, \widetilde{X}_t^{N,i}, \widetilde{\mu}_s)|ds \\ &\leq \int_{t_0}^T (|X_s^{N,i} - \widetilde{X}_s^{N,i}| + d_2(m_{X_s}^N, \widetilde{\mu}_s))ds, \end{aligned}$$

since H_p is Lipschitz.
By Gronwall's inequality

$$|X_t^{N,i} - \widetilde{X}_t^{N,i}| \leq C \int_{t_0}^{T} d_2(m_{X_s}^N, \widetilde{\mu}_s) ds.$$

Taking the power 2 and averaging the left-hand side

$$d_2^2(m_{X_t}^N, m_{\widetilde{X}_t}^N) \leq \frac{1}{N} \sum_i |X_t^{N,i} - \widetilde{X}_t^{N,i}|^2 \leq C \int_{t_0}^{T} d_2^2(m_{X_s}^N, \widetilde{\mu}_s)) ds.$$

Using the triangle and the Gronwall's inequality

$$d_2^2(m_{X_t}^N, m_{\widetilde{X}_t}^N) \leq C \int_{t_0}^{T} d_2^2(m_{\widetilde{X}_s}^N, \widetilde{\mu}_s) ds,$$

using again the triangle inequality

$$d_2^2(m_{X_t}^N, \widetilde{\mu}_t) \leq C d_2^2(m_{\widetilde{X}_t}^N, \widetilde{\mu}_t) + C \int_{t_0}^{T} d_2^2(m_{\widetilde{X}_s}^N, \widetilde{\mu}_s) ds.$$

By Theorem 1.3 in [16], we get the thesis. □

Lemma 4.8. *Assume* (1.1).
There is a constant $C = C(t_0, m_0) > 0$ *such that*

$$\mathbb{E}[\sup_{t \in [t_0, \tau^N]} d_1(m_{X_t^N}^N, m(t))] \leq C N^{-\frac{1}{d+8}} \tag{4.6}$$

and

$$\mathbb{P}[\tau^n < T] \leq C N^{-\frac{1}{d+8}}. \tag{4.7}$$

Proof. Using Theorem 4.7, we have that

$$\mathbb{E}[\sup_{t \in [0,T]} d_2^2(m_{X_t^N}, m(t))] \leq C N^{-\frac{2}{d+8}}.$$

Since $d_1 \leq d_2$ we can conclude that

$$\mathbb{E}[\sup_{t \in [t_0, \tau^N]} d_1(m_{X_t^N}^N, m(t))] \leq C N^{-\frac{1}{d+8}}$$

Then

$$\mathbb{P}[\tau^N < T] \leq \mathbb{P}\left[\sup_{t \in [t_0, \tau^N]} d_1(m_{X_t^N}^N, m(t)) \geq \frac{\delta}{2}\right] \leq C \delta^{-1} N^{-\frac{1}{d+8}}.$$

□

Another way to understand the derivatives in the space of measures is to project the map $\mathcal{U}$ to the finite dimensional space $(\mathbb{R}^d)^N$ via the empirical measure $m_x^N = \frac{1}{N}\sum_{j=1}^N \delta_{x_j}$ with $x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$.
Let us call

$$\mathcal{U}^N(t,x) = \mathcal{U}(t, m_x^N) \quad \text{for} \quad (t,x) \in V_\delta^N.$$

From preposition 6.2 in [6]

$$D_{x_j}\mathcal{U}^N(t,x) = \frac{1}{N} D_m \mathcal{U}(t, m_x^N, x_j)$$

and

$$D^2_{x_j,x_j}\mathcal{U}^N(t,x) = \frac{1}{N} D^2_{ym}\mathcal{U}(t, m_x^N, x_j) + \frac{1}{N^2} D^2_{mm}\mathcal{U}(t, m_x^N, x_j).$$

By the Theorem 3.13, we have the regularity of $\mathcal{U}$ and we can easily conclude that, on V_δ^N, $\mathcal{U}^N$ is $\mathcal{C}^1$ in time-space with $D_{x_j}\mathcal{U}^N$ Lipschitz continuous in space.
We get

$$|D^2_{x_j x_j}\mathcal{U}^N(t,x) - \frac{1}{N} D^2_{ym}\mathcal{U}(t, m_x^N, x_j)| \leq \frac{C}{N^2} \text{ a.e. in } V_\delta^N.$$

Let $O^N(t,x)$ be such that

$$|O^N(t,x)| \leq \frac{C}{T} \text{ a.e. in } V_\delta^N. \tag{4.8}$$

Then $\mathcal{U}^N$ satisfies

$$\begin{cases} -\partial_t \mathcal{U}^N(t,x) - \sum_{j=1}^N \Delta_{x_j}\mathcal{U}^N(t,x) + O^N(t,x) \\ +\frac{1}{N}\sum_{j=1}^N H(x_j, N D_{x_j}\mathcal{U}^N(t,x)) = \mathcal{F}(m_x^N) \text{ a.e. in } V_\delta^N, \\ \mathcal{U}^N(T,x) = \mathcal{G}(m_x^N) \text{ on } (\mathbb{R}^d)^N. \end{cases} \tag{4.9}$$

Lemma 4.9. *Let $\mathbf{Y}^N = (Y^{N,i})_{i=1,\dots,N}$ defined by (4.3).*
Calling $R^N = ||\mathcal{U}^N - \mathcal{V}^N||_\infty$, then

$$\mathbb{E}[\int_{t_0}^{\widetilde{\tau}^N} N^{-1} \sum_j |H_p(Y_t^{N,j}, N D_{x_j}\mathcal{U}^N) - H_p(Y_t^{N,j}, N D_{x_j}\mathcal{V}^N)|^2 dt] \leq C(N^{-1} + R^N). \tag{4.10}$$

Proof. Let $t \in [t_0, \widetilde{\tau}^N]$ then

$$\begin{aligned}
d\mathcal{U}^N(t, Y_t^N) &= (\partial_t \mathcal{U}^N + \sum_j \Delta_{x_j} \mathcal{U}^N - \sum_j H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N) \cdot D_{x_j}\mathcal{U}^N)dt \\
&+ \sqrt{2} \sum_j D_{x_j}\mathcal{U}^N \cdot dB_t^j \\
&= (\frac{1}{N} \sum_j H(Y^{N,j}, ND_{x_j}\mathcal{U}^N(t, Y_t^N)) - \sum_j H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N) \cdot D_{x_j}\mathcal{U}^N \\
&+ O^N - \mathcal{F}(m_{Y_t^N}^N))dt + \sqrt{2} \sum_j D_{x_j}\mathcal{U}^N \cdot dB_t^j \\
&\geq (\frac{1}{N} \sum_j (-L(Y^{N,j}, -H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)) + C^{-1}|H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N) \\
&- H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)|^2) - CN^{-1} - \mathcal{F}(m_{Y_t^N}^N))dt + \sqrt{2} \sum_j D_{x_j}\mathcal{U}^N \cdot dB_t^j,
\end{aligned}$$

where we used that $|O^N(t, x)| \leq CN^{-1}$ and the uniform convexity of H in bounded sets:

$$L(-H_p(p^*)) = p^* \cdot H_p(p^*) - H(p^*).$$

We take expectations and integrate between t_0 and $\widetilde{\tau}^N$ above and get

$$\begin{aligned}
&\mathbb{E}[\mathcal{U}^N(\widetilde{\tau}^N, Y_{\widetilde{\tau}^n}^N)] - \mathbb{E}[\mathcal{U}^N(t_0, Z^N)] \\
&\geq \mathbb{E}[\int_{t_0}^{\widetilde{\tau}^N} (\frac{1}{N} \sum_j (-L(Y^{N,j}, -H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)) + C^{-1}|H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N) \\
&- H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)|^2) - CN^{-1} - \mathcal{F}(m_{Y_t^N}^N))dt].
\end{aligned}$$

Rearranging, using the definition of R^N, the dynamic programming principle and the optimality of Y^N for $\mathcal{V}^N(t_0, Z^N)$ we find

$$\begin{aligned}
&\mathbb{E}[\mathcal{U}^N(t_0, Z^N)] + \mathbb{E}[\int_{t_0}^{\widetilde{\tau}^N} (\frac{1}{CN} \sum_j |H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N) - H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)|^2 dt] \\
&\leq \mathbb{E}[\int_{t_0}^{\widetilde{\tau}^N} (\frac{1}{N} \sum_j (L(Y^{N,j}, -H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)) + CN^{-1} + \mathcal{F}(m_{Y_t^N}^N))dt + \\
&+ \mathcal{V}^N(\widetilde{\tau}^N, Y_{\widetilde{\tau}^N}^N)] + R^N \\
&\leq \mathbb{E}[\mathcal{V}^N(t_0, Z^N)] + CN^{-1} + R^N,
\end{aligned}$$

and using once more the definition of R^N, we get

$$\mathbb{E}[\int_{t_0}^{\widetilde{\tau}^N} (\frac{1}{CN} \sum_j |H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N) - H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N)|^2 dt] \leq CN^{-1} + 2R^N.$$

□

Lemma 4.10. *For* $\mathbf{X}^N = (X^{N,i})$ *and* $\mathbf{Y}^N = (Y^{N,i})$ *defined by* (4.5) *and* (4.3) *respectively, we have*

$$\mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} N^{-1}\sum_j |X_s^{N,j} - Y_s^{N,j}|] \leq C(N^{-1} + R^N)^{1/2} \tag{4.11}$$

and

$$\mathbb{P}[\widetilde{\tau}^N < T] \leq C(N^{-\frac{1}{d+8}} + (R^N)^{1/2}). \tag{4.12}$$

Proof. From Lemma 4.9, the regularity of $\mathcal{U}^N$ in (4.4)

$$\begin{aligned}
&\mathbb{E}[\sup_{s\in[t_0,t\wedge\widetilde{\tau}^N]} N^{-1}\sum_j |X_s^{N,j} - Y_s^{N,j}|] \\
&\leq \mathbb{E}[\int_{t_0}^{t\wedge\widetilde{\tau}^N} N^{-1}\sum_j |H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N(t,Y_t^N)) - H_p(Y_t^{N,j}, ND_{x_j}\mathcal{V}^N(t,Y_t^N))|dt]+ \\
&\mathbb{E}[\int_{t_0}^{t\wedge\widetilde{\tau}^N} N^{-1}\sum_j |H_p(Y_t^{N,j}, ND_{x_j}\mathcal{U}^N(t,Y_t^N)) - H_p(X_t^{N,j}, ND_{x_j}\mathcal{U}^N(t,X_t^N))|dt] \\
&\leq C(N^{-1} + R^N)^{1/2} + CN^{-1}\sum_j \mathbb{E}[\int_{t_0}^{t\wedge\widetilde{\tau}^N} |X_s^{N,j} - Y_s^{N,j}|ds],
\end{aligned}$$

and Gronwall's inequality gives

$$\mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} N^{-1}\sum_j |X_s^{N,j} - Y_s^{N,j}|] \leq C(N^{-1} + R^N)^{1/2}.$$

Then by definition of $\widetilde{\tau}^N$, Lemma 4.8 and 4.10

$$\begin{aligned}
\mathbb{P}[\widetilde{\tau}^N < T] &\leq \mathbb{P}[\tau^N < T] + \mathbb{P}[\sup_{s\in[t_0,\widetilde{\tau}^N]} N^{-1}\sum_j |X_s^{N,j} - Y_s^{N,j}| > \frac{\delta}{2}] \\
&\leq CN^{-\frac{1}{d+8}} + C\delta^{-1}(N^{-1} + R^N)^{1/2}.
\end{aligned}$$

□

We are ready to prove Theorem 4.1.

Proof of Theorem 4.1. We can notice that using the triangular inequality of the distance and the additive interval property

$$\begin{aligned}
&\mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{Y^N_s}, m(s))] \\
&\leq \mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{Y^N_s}, m^N_{X^N_s})] + +\mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{X^N_s}, m(s))] \\
&\leq \mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} N^{-1}\sum_j |X^{N,j}_s - Y^{N,j}_s] + \mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{X^N_s}, m(s))] \\
&\leq C(N^{-1} + R^N)^{1/2} + CN^{-\frac{1}{d+8}},
\end{aligned} \tag{4.13}$$

where in the second inequality we used that $m^N_{X^N_s}$ and $m^N_{Y^N_s}$ are the laws respectively of X^N_s and Y^N_s and in the last one we used lemmas 4.10 and 4.8.

Thanks to Theorem 2.1 we can estimate R^N:

$$R^N = ||\mathcal{U}^N - \mathcal{V}^N||_\infty \leq CN^{-\beta} \text{ such that } \exists\beta > 0. \tag{4.14}$$

Thus, $\exists\gamma' = \gamma'(d)$ and $C = C(t_0, m_0)$

$$\mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{Y^N_s}, m(s))] \leq CN^{-\gamma}. \tag{4.15}$$

Finally, we have

$$\begin{aligned}
&\mathbb{E}[\sup_{s\in[t_0,T]} d_1(m^N_{Y^N_s}, m(s))] \\
&\leq \mathbb{E}[\sup_{s\in[t_0,\widetilde{\tau}^N]} d_1(m^N_{Y^N_s}, m(s)] + \mathbb{E}[\sup_{s\in[t_0,T]} d_1^2(m^N_{Y^N_s}, m(s))]^{1/2}\mathbb{P}[\widetilde{\tau}^N < T]^{1/2} \\
&\leq CN^{-\gamma'} + C(N^{-\frac{1}{d+8}} + (R^N)^{1/2})^{1/2},
\end{aligned} \tag{4.16}$$

where we used the Holder inequality, Lemma 4.10 and that $m(s)$ has a uniformly bounded second order moment and by Lemma 2.2 the drift of the process Y^N is also uniformly bounded:

$$\mathbb{E}[\sup_{s\in[t_0,T]} d_1^2(m^N_{Y^N_s}, m(s))] \leq C.$$

Thanks to Theorem 2.1, there exists a constant $\gamma'' = \gamma''(d) \in (0,1)$ and a new constant $C' = C'(t_0, m_0) > 0$ so that we can conclude

$$\mathbb{E}[\sup_{s\in[t_0,T]} d_1(m^N_{Y^N_s}, m(s))] \leq C'N^{-\gamma''}.$$

□

Appendix A

The Wasserstein distance

Theorem A.1. *Let, for* $i = 1, 2$

$$\begin{cases} \partial_t m^i - \Delta m^i - \operatorname{div}(b^i m^i) = 0 & \text{in } \mathbb{R}^d \times (0,T), \\ m^i(0) = m_0^i & \text{in } \mathbb{R}^d \end{cases} \tag{A.1}$$

with

$$b^i(x) = H_p(x, Du^i) \quad \text{and} \quad m^i(t,x) = \int_{\mathbb{R}^d} \mu(t,x,y)dy.$$

Then,

$$\sup_{t\in[0,T]} d_2(m^1(t), m^2(t)) \leq C \left(||Du^1 - Du^2||_\infty + d_2(m_0^1, m_0^2) \right). \tag{A.2}$$

Proof. We write

$$\partial_t\mu - \Delta_x\mu - \Delta_y\mu - 2\nabla^2_{xy}\mu - \operatorname{div}_x(b^1(x)\mu(t,x,y)) - \operatorname{div}_y(b^2(y)\mu(t,x,y)) = 0 \text{ in } \mathbb{R}^{2d}\times(0,T). \tag{A.3}$$

Multiplying (A.3) by $|x-y|^2$ and integrating over $\mathbb{R}^{2d}$ we obtain

$$\int_{\mathbb{R}^{2d}} \partial_t\mu \cdot |x-y|^2 dxdy - \int_{\mathbb{R}^{2d}} \Delta_x\mu \cdot |x-y|^2 + \Delta_y\mu \cdot |x-y|^2 - 2\nabla^2_{xy}\mu \cdot |x-y|^2 dxdy$$
$$- \int_{\mathbb{R}^{2d}} \operatorname{div}_x(b^1(x)\mu(x,y)) \cdot |x-y|^2 dxdy - \int_{\mathbb{R}^{2d}} \operatorname{div}_y(b^2(y)\mu(x,y)) \cdot |x-y|^2 dxdy = 0.$$

Notice that, using the divergence theorem,

$$\int_{\mathbb{R}^{2d}} \Delta_x\mu \cdot |x-y|^2 + \Delta_y\mu \cdot |x-y|^2 - 2\nabla^2_{xy}\mu \cdot |x-y|^2 dxdy = 0.$$

Using the divergence theorem, we obtain

$$\int_{\mathbb{R}^{2d}} \partial_t\mu \cdot |x-y|^2 dxdy + \int_{\mathbb{R}^{2d}} 2[b^1(x) - b^2(y)]\mu(x,y)|x-y| dxdy = 0,$$

and so

$$\int_{\mathbb{R}^{2d}} \partial_t \mu \cdot |x-y|^2 dxdy = 2\int_{\mathbb{R}^{2d}} [b^2(y) - b^1(x)]\mu(x,y)|x-y|dxdy,$$

and, now, recalling that $b^1(x) = H_p(x, Du^1)$ and $b^2(y) = H_p(y, Du^2)$ and using that H_p is Lipschitz continuous

$$\begin{aligned}
&\int_{\mathbb{R}^{2d}} \partial_t \mu \cdot |x-y|^2 dxdy = 2\int_{\mathbb{R}^{2d}} [b^2(y) - b^1(x)]\mu(x,y)|x-y|dxdy \\
&\leq 2\int_{\mathbb{R}^{2d}} |b^2(y) - b^1(x)|\mu(x,y)|x-y|dxdy \\
&\leq 2\int_{\mathbb{R}^{2d}} C(|x-y| + |Du^1(t,x) - Du^2(t,y)|)\mu(x,y)|x-y|dxdy \\
&\leq C\int_{\mathbb{R}^{2d}} (|x-y| + |Du^1(t,x) - Du^2(t,x)| + |Du^2(t,x) - Du^2(t,y)|)\mu(x,y)|x-y|dxdy \\
&\leq C\int_{\mathbb{R}^{2d}} (|x-y| + ||Du^1 - Du^2||_\infty + ||D^2u^2||_\infty |x-y|)\mu(x,y)|x-y|dxdy \\
&\leq C\left(\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y) + ||Du^1 - Du^2||_\infty \int_{\mathbb{R}^{2d}} \mu(x,y)|x-y|\right).
\end{aligned} \tag{A.4}$$

Using Holder's inequality and the assumptions on μ

$$\begin{aligned}
\int_{\mathbb{R}^{2d}} \mu(x,y)|x-y|dxdy &\leq \left(\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y)dxdy\right)^{1/2} \left(\int_{\mathbb{R}^{2d}} \mu(x,y)dxdy\right)^{1/2} \\
&\leq C\left(\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y)dxdy\right)^{1/2},
\end{aligned}$$

and by Young's inequality

$$\begin{aligned}
&||Du^1 - Du^2||_\infty \left(\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y,t)dxdy\right)^{1/2} \\
&\leq \frac{1}{2}||Du^1 - Du^2||_\infty^2 + \frac{1}{2}\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y,t)dxdy.
\end{aligned}$$

Going back to (A.4) and using the previous estimates we get

$$\partial_t \int_{\mathbb{R}^{2d}} \mu(x,y,t) \cdot |x-y|^2 dxdy \leq C||Du^1 - Du^2||_\infty^2 + C\int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y,t)dxdy, \tag{A.5}$$

and, now, applying the Gronwall's lemma we obtain

$$\int_{\mathbb{R}^{2d}} \mu(t,x,y)\cdot|x-y|^2 dxdy \leq C\left(||Du^1 - Du^2||_\infty^2 + \int_{\mathbb{R}^{2d}} \mu(0,x,y)\cdot|x-y|^2 dxdy\right). \tag{A.6}$$

Knowing that $d_2^2(m_0^1, m_0^2) = \inf_{\pi\in\Pi(m_0^1,m_0^2)} \int_{\mathbb{R}^{2d}} |x-y|^2 d\pi(x,y)$ and so $\exists\pi^* \in \Pi(m_0^1, m_0^2)$ such that $d_2^2(m_0^1, m_0^2) = \int_{\mathbb{R}^{2d}} |x-y|^2 d\pi^*(x,y)$ and so

$$\int_{\mathbb{R}^{2d}} \mu(t,x,y)\cdot|x-y|^2 dxdy \leq C\left(||Du^1 - Du^2||_\infty^2 + d_2^2(m_0^1, m_0^2)\right). \tag{A.7}$$

Recalling that by definition

$$d_2^2(m^1, m^2) = \inf_{\pi\in\Pi(m^1,m^2)} \int_{\mathbb{R}^{2d}} |x-y|^2 d\pi(x,y) \leq \int_{\mathbb{R}^{2d}} |x-y|^2 \mu(x,y) dxdy.$$

As all factors involved are positive and taking the supremum on the left-hand side, we get

$$\sup_t d_2(m^1(t), m^2(t)) \leq C\left(||Du^1 - Du^2||_\infty + d_2(m_0^1, m_0^2)\right). \tag{A.8}$$

□

Appendix B

Kantorovich duality theorem

In this section, we investigate a powerful duality formula due to Kantorovich. It will used to prove the Kantorovich-Rubistein Theorem 1.1 state in the first Chapter 1.

We need some definitions and a theorem to prove the Kantorovich duality theorem.

Definition B.1. *Let E be a normed vector space, E^* its the topological dual and $\Theta : E \to \mathbb{R} \cup \{+\infty\}$ a convex function. The* Legendre-Fenchel transform *of Θ is given by the function Θ^* defined on E^* with the formula*

$$\Theta^*(z^*) = \sup_{z \in E}[\langle z^*, z\rangle - \Theta(z)].$$

Definition B.2. *A* Polish space *is a separable completely metrizable topological space.*

Theorem B.3 (Fenchel-Rockafellar duality theorem)**.** *Let E be a normed vector space, E^* its topological dual space and two convex functions $\Theta, \Xi : E \to \mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be the Legendre-Fenchel transform of Θ, Ξ respectively.*
Assume that there exists $z_0 \in E$, such that Θ continuous at z_0, $\Theta(z_0) < +\infty$ and $\Xi(z_0) < +\infty$. Then,

$$\inf_E[\Theta + \Xi] = \max_{z^* \in E^*}[-\Theta^*(-z^*) - \Xi^*(z^*)].$$

For the proof, see Theorem 1.9 in [20].

Finally, we state and prove a dual formulation for linear minimization problem that was introduced by Kantorovich in 1942.

Theorem B.4 (Kantorovich duality theorem)**.** *Let X and Y be Polish spaces and μ, ν two probability measures such that $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$.*
Let $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ be a lower semicontinuous cost function.
Call

$$I[\pi] = \int_{X\times Y} c(x, y)\, d\pi(x, y),$$

and

$$J(\phi, \psi) = \int_X \phi\, d\mu + \int_Y \psi\, d\nu,$$

with $\pi \in \mathcal{P}(X \times Y)$ and $(\phi, \psi) \in L^1(d\mu) \times L^1(d\nu)$.
Let us define the set $\Pi(\mu, \nu)$ of Borel probability measures on $X \times Y$ with marginals μ and ν, and Φ_c the set of measurable functions $(\phi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ such that

$$\phi(x) + \psi(y) \le c(x, y) \tag{B.1}$$

for $d\mu$-almost all $x \in X$ and $d\nu$-almost all $y \in Y$.
Then .

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_{(\phi,\psi)\in\Phi_c} J(\phi, \psi)$$

Sketch of the proof. ($\le$) Let $(\phi, \psi) \in \Phi_c$ and $\pi \in \Pi(\mu, \nu)$. By definition of $\Pi(\mu, \nu)$

$$J(\phi, \psi) = \int_X \phi\, d\mu + \int_Y \psi\, d\mu = \int_{X\times Y} [\phi(x) + \psi(y)] d\pi(x, y).$$

(B.1) holds only almost everywhere and so, let N_X and N_Y such that $\mu(N_X) = 0$ and $\nu(N_Y) = 0$. Then (B.1) holds for any $(x, y) \in N_X^C \times N_Y^C$. Since $\pi[N_X \times Y] = \mu(N_X) = 0$ and $\pi[X \times N_Y] = \nu(N_Y) = 0$, we have that $\pi[(N_X^C \times N_Y^C)^C] = 0$. As a consequence

$$\int_{X\times Y} [\phi(x) + \psi(y)] d\pi(x, y) \le \int_{X\times Y} c(x, y) d\pi(x, y) = I[\pi].$$

Taking the supremum on the left-hand side and the infimum on the right-hand side in the last inequality we get the thesis.

($\ge$) Let us divide the proof in three parts.

(1) *X and Y compact sets and c continuous function on $X \times Y$.*
Set E the set of bounded continuous functions on $X \times Y$ equipped with the supremum norm $||\cdot||_\infty$. Using the Riesz' theorem, the dual of E is $E^* = M(X \times Y)$, the set on Radon measures. Let us define

$$\Theta : u \in E \mapsto \begin{cases} 0 \ \text{ if } \ u(x, y) \ge -c(x, y) \\ +\infty \ \text{ else,} \end{cases}$$

and

$$\Xi : u \in E \mapsto \begin{cases} \int_x \phi \, d\mu + \int_Y \psi \, d\nu & \text{if } \; u(x,y) = \phi(x) + \psi(y) \\ +\infty \; \text{ else.} \end{cases}$$

The assumption of Theorem B.3 are satisfied with $z_0 = 1$.
Note that

$$\inf \left\{ \int_X \phi \, d\mu + \int_Y \psi \, d\nu : \phi(x) + \psi(y) \geq -c(x,y) \right\} = -\sup\{J(\phi, \psi) : (\phi, \psi) \in \Phi_c\}.$$

Let $\pi \in M(X \times Y)$. We compute the Legendre-Fenchel transform of Θ and Ξ.

$$\begin{aligned} \Theta^*(-\pi) &= \sup_{u \in E} \left\{ -\int u(x,y) \, d\pi(x,y) : u(x,y) \geq -c(x,y) \right\} \\ &= \sup_{u \in E} \left\{ \int u(x,y) \, d\pi(x,y) : u(x,y) \leq c(x,y) \right\}. \end{aligned}$$

Thus

$$\Theta^*(-\pi) = \begin{cases} \int_{X \times Y} c(x,y) \, d\pi(x,y) & \text{if } \pi \text{ is non negative} \\ +\infty \text{ else.} \end{cases}$$

Similarly, calling $C_b(A)$ the set of bounded continuous function on A and setting $D = C_b(X) \times C_b(Y)$,

$$\Xi^*(\pi) = \begin{cases} 0 \; \text{ if } \; \forall (\phi, \psi) \in D \; : \; \int_{X \times Y} (\phi(x) + \psi(y)) d\pi(x,y) = \int_x \phi d\mu + \int_y \psi d\nu \\ +\infty \text{ else.} \end{cases}$$

Noting that $C_b(X) \times C_b(Y) \subset L^1(d\mu) \cap L^1(d\nu)$, putting everything together and changing signs, we recover

$$\inf_{\Pi(\mu,\nu)} I(\pi) = \sup_{\Phi_c} J(\phi, \psi).$$

(2) *c is bounded and uniformly continuous.*
We define

$$||c||_\infty = \sup_{X \times Y} c(x,y).$$

Let $\pi_* \in \Pi(\mu, \nu)$ such that $I[\pi_*] = \inf_{\pi \in \Pi(\mu,\nu)} I[\pi]$. Let $\delta > 0$ small. We have that $X \times Y$ is a Polish space since X and Y are so. π_* is tight and so there exist $X_0 \subset X$ and $Y_0 \subset Y$ so that $\mu(X \setminus X_0) \leq \delta$ and $\nu(Y \setminus Y_0) \leq \delta$. It can be proved that $\pi_*[(X \times Y) \setminus (X_0 \times Y_0)] \leq 2\delta$.
Define

$$\pi_0 * = \frac{I_{X_0 \times Y_0}}{\pi_*(X_0 \times Y_0)} \pi_*,$$

and let μ_0 and ν_0 be the marginals of π_0* onto X_0 and Y_0 respectively.
We define $\Pi_0(\mu_0, \nu_0)$ the set of probability with marginals μ_0 and ν_0.
Let us define

$$I_0[\pi_0] = \int_{X_0 \times Y_0} c(x, y) d\pi_0(x, y).$$

Let $\widetilde{\pi_0} \in \Pi(\mu_0, \nu_0)$ be such that

$$I[\widetilde{\pi_0}] = \inf_{\Pi_0(\mu_0,\nu_0)} I_0[\pi_0].$$

We construct from $\widetilde{\pi_0}$,

$$\widetilde{\pi} = \pi_*(X_0 \times Y_0)\widetilde{\pi_0} + \mathbb{I}_{(X_0 \times Y_0)^C}\pi_*.$$

We calculate $I(\widetilde{\pi})$

$$\begin{aligned} I(\widetilde{\pi}) &= \pi_*(X_0 \times Y_0)I_0(\widetilde{\pi_0}) + \int_{(X_0\times Y_0)^C} c(x, y)d\pi_*(x, y) \leq I_0(\widetilde{\pi_0}) + 2||c||_\infty\delta \\ &= \inf I_0 + 2||c||_\infty\delta. \end{aligned}$$

Thus

$$\inf_{\Pi(\mu,\nu)} I[\pi] \leq \inf I_0 + 2||c||_\infty\delta.$$

Let

$$J_o(\phi_0, \psi_0) = \int_{X_0} \phi_0 \, d\mu_0 + \int_{Y_0} \psi_0 \, d\nu_0 \;\text{ defined on }\; L^1(d\mu_0) \times L^1(d\nu_0).$$

By (1) we have $\inf I_0 = \sup J_0$.
There exists a couple of function $(\widetilde{\phi_0}, \widetilde{\psi_0})$ such that $J_0(\widetilde{\phi_0}, \widetilde{\psi_0}) \geq \sup J_0 - \delta$.
We ensure that $\widetilde{\phi_0}(x)+\widetilde{\psi_0}(y) \leq c(x, y)$ for all x and y allowing that $\widetilde{\phi_0}$ and $\widetilde{\psi_0}$ take values in $\mathbb{R}\cup\{-\infty\}$. Without loss of generality, we assume $\delta \leq 1$. Since $J_0(0, 0) = 0$ then $\sup J_0 \geq 0$ and so $J_0(\widetilde{\phi_0}, \widetilde{\psi_0}) \geq -\delta \geq -1$. So exists $(x_0, y_0) \in X_0 \times Y_0$ such that $\widetilde{\phi_0}(x_0) + \widetilde{\psi_0}(y_0) \geq -1$. We have $\widetilde{\phi_0}(x_0) \geq -\frac{1}{2}$ and $\widetilde{\psi_0}(y_0) \geq -\frac{1}{2}$. We get that, $\forall (x, y) \in X_0 \times Y_0$,

$$\widetilde{\phi_0}(x) \leq c(x, y_0) - \widetilde{\psi_0}(y_0) \leq c(x, y_0) + \frac{1}{2},$$

$$\widetilde{\psi_0}(y) \leq c(x_0, y) - \widetilde{\phi_0}(x_0) \leq c(x_0, y) + \frac{1}{2}.$$

Let us define

$$\overline{\phi_0}(x) = \inf_{y\in Y_0} [c(x, y) - \widetilde{\psi_0}(y)].$$

We have $\widetilde{\phi_0} \leq \overline{\phi_0}$ on X_0 and so $J_0(\overline{\phi_0}, \widetilde{\psi_0}) \geq J_0(\widetilde{\phi_0}, \widetilde{\psi_0})$ and

$$\overline{\phi_0}(x) \geq \inf_y [c(x,y) - c(x_0,y)] - \frac{1}{2},$$

and

$$\overline{\phi_0}(x) \leq c(x,y_0) - \widetilde{\psi_0}(y_0) \leq c(x,y_0) + \frac{1}{2}.$$

We define

$$\overline{\psi_0}(y) = \inf_{x \in X_0} [c(x,y) - \overline{\phi_0}(x)],$$

and we still have $(\overline{\phi_0}, \overline{\psi_0}) \in \Phi_c$ and $J_0(\overline{\phi_0}, \overline{\psi_0}) \geq J_0(\overline{\phi_0}, \widetilde{\psi_0}) \geq J_0(\widetilde{\phi_0}, \widetilde{\psi_0})$.
We have

$$\overline{\psi_0}(y) \geq \inf_x [c(x,y) - c(x,y_0)] - \frac{1}{2},$$

and

$$\overline{\psi_0}(y) \leq c(x_0,y) - \overline{\phi_0}(x_0) \leq c(x_0,y) - \widetilde{\phi_0}(x_0) \leq c(x_0,y) + \frac{1}{2}.$$

In particular

$$\overline{\phi_0}(x) \geq -||c||_\infty - \frac{1}{2},$$

and

$$\overline{\psi_0}(y) \geq -||c||_\infty - \frac{1}{2}.$$

With some simple inequality we get

$$J(\overline{\phi_0}, \overline{\psi_0}) \geq (1-2\delta)[\inf I - 2(2||c||_\infty + 1)\delta] - 2(2||c||_\infty + 1)\delta.$$

Since δ is arbitrary we can conclude that $\sup J(\phi, \psi) \geq \inf I$.

(3) *General case*
We write $c = \sup_n c_n$, where c_n is a non-decreasing sequence of non-negative uniformly continuous cost functions. Replacing c_n by $\inf_n(c_n, n)$ one can assume that c_n is bounded.
Let

$$I_n[\pi] = \int_{X \times Y} c_n \, d\pi \quad \text{for} \quad \pi \in \Pi(\mu, \nu).$$

By (2) we have

$$\inf_{\pi \in \Pi(\mu,\nu)} I_n[\pi] = \sup_{(\phi,\psi) \in \Phi_{c_n}} J(\phi, \psi). \tag{B.2}$$

Since $c_n \leq c$ by construction, it follows that $\Phi_{c_n} \subset \Phi_c$, then

$$\sup_{(\phi,\psi) \in \Phi_{c_n}} J(\phi, \psi) \leq \sup_{(\phi,\psi) \in \Phi_c} J(\phi, \psi). \tag{B.3}$$

Since I_n is a non-decreasing sequence of functional, then I_n is a non-decreasing sequence bounded above by $\inf I$. $\Pi(\mu,\nu)$ is tight and relatively compact for the weak topology.
In particular, if $(\pi_n^k)_{k\in\mathbb{N}}$ is any minimizing sequence for the problem $\inf I_n[\pi]$ then, up to a subsequence, $\pi_n^k \xrightarrow{k\to\infty} \pi_n \in \mathcal{P}(X\times Y)$ weakly and so $\forall\theta$ bounded and continuous function

$$\int \theta(x,y)\, d\pi_n^k(x,y) \xrightarrow{k\to\infty} \int \theta(x,y)\, d\pi_n(x,y),$$

and we see that $\pi_n \in \Pi(\mu,\nu)$ and

$$\inf I_n = \lim_{k\to\infty} \int c_n\, d\pi_n^k = \int c_n\, d\pi_n.$$

By compactness of $\Pi(\mu,\nu)$, the sequence $(\pi_n)_n$ admits a cluster point π_*. Whenever $n \geq m$

$$I_n[\pi_n] \geq I_m[\pi_n].$$

By continuity

$$\lim_{n\to\infty} I_n[\pi_n] \geq \limsup_{n\to\infty} I_m[\pi_n] \geq I_m[\pi_*].$$

By monotone convergence $I_m[\pi_*] \xrightarrow{m\to\infty} I[\pi_k]$ and so

$$\lim_{n\to\infty} I_n[\pi_n] \geq \lim_{m\to\infty} I_m[\pi_*] = I[\pi_*] \geq \inf_{\pi\in Pi(\mu,\nu)} I[\pi],$$

which proves

$$\lim_{n\to\infty} \inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi] \geq \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

By invoking the monotone convergence theorem for the increasing sequence $(c_n)_n$, we have

$$I[\pi_*] = \lim_{n\to\infty} I_n[\pi_*] \leq \lim_{n\to\infty} \limsup_{k\to\infty} I_n[\pi_k] \leq \limsup_{k\to\infty} I[\pi_k] = \inf I.$$

Combining all these results, we get the thesis.

□

Definition B.5. *Let $c : X\times Y \to \mathbb{R}\cup\{+\infty\}$ be a bounded lower semicontinuous function. Let $\phi : X \to \mathbb{R}$ be a bounded function. We define*

$$\phi^c(y) = \inf_x [c(x,y) - \phi(x)]$$

and

$$\phi^{cc}(x) = \inf_y [c(x,y) - \phi^c(y)].$$

The couple (ϕ^{cc},ϕ^c) is called conjugate c-concave functions.

Now we are ready to prove Theorem 1.1.

Proof of theorem 1.1. Let us call $d(x,y)$ the metric on $X \times Y$ and be $d_n := \frac{d}{1+n^{-1}d}$. Since $1+dn^{-1} \geq 1$, we note that $d_n \leq d$ and for all $x, y \in \mathbb{R}^d$, $d_n(x,y) \xrightarrow{n\to\infty} d(x,y)$ monotonically. Finally, let us point out that the set of 1-Lipschitz functions for d_n is included in the set of 1-Lipschitz functions for d since $d_n \leq d$ and so

$$|f(x) - f(y)| \leq d_n(x,y) \leq d(x,y).$$

So, by that, we can use Theorem B.4 with d bounded (otherwise we can reason with d_n) and we only have to check that

$$\sup_{(\phi,\psi)\in\Phi_c} J(\phi,\psi) = \sup_{\phi:1-Lip} \left\{ \int_{\mathbb{R}^d} \phi \, d(m-m') \right\},$$

recalling that $J(\phi,\psi) = \int_X \phi d\,\mu + \int_Y \psi \, d\nu$.
Let (ϕ^{dd}, ϕ^d) be the d-concave conjugate functions. We have that ϕ^d is 1-Lipschitz since is the infimum of ϕ that is a 1-Lipschitz function. It follows from the proof of Theorem B.4 (with d bounded) that

$$\sup_{(\phi,\psi)\in\Phi_c} J(\phi,\psi) = \sup_{\phi\in L^1(dm)} J(\phi^{dd},\phi^d),$$

and noting that, using the lipschitzianity,

$$-\phi^d(x) \leq \inf_y [d(x,y) - \phi^d(y)] \leq -\phi^d(x),$$

where in the last inequality we put $x = y$. We conclude that

$$\phi^{dd} = -\phi^d.$$

It can also proved that $(\phi^{dd})^d = \phi^d$. Then

$$\begin{aligned}\sup_{(\phi,\psi)\in\Phi_c} J(\phi,\psi) &= \sup_{\phi\in L^1(dm)} J(\phi^{dd},\phi^d) = \sup_{\phi\in L^1(dm)} J(-\phi^d,\phi^d) \\ &\leq \sup_{\phi:1-Lip} J(-\phi,\phi) \leq \sup_{\Phi_c} J(\phi,\psi),\end{aligned}$$

and we get the thesis.

□

Appendix C

Lions-Malgrange-type argument

In this part we show the *Lions-Malgrange-type argument*: we prove the uniqueness of solution a general linear forward-backward system with given initial data. This result is used several times in the thesis.

Let the pair $((z^k)_{k=1,\dots,d}, \mu) = ((\partial_{x_k}(u - \widetilde{u}))_{k=1,\dots,d}, m - \widetilde{m})$ solves the system

$$\begin{cases} -\partial_t z^k - \Delta z^k + g^k(t,x) = 0 \text{ in } (t_1, T) \times \mathbb{R}^d, \\ \partial_t \mu - \Delta \mu + \operatorname{div}(h) = 0 \text{ in } (t_1, T) \times \mathbb{R}^d \\ \mu(t_1) = 0 \ \textit{and} \ z^k(t_1, \cdot) = 0 \text{ on } \mathbb{R}^d. \end{cases} \tag{C.1}$$

Assuming that

$$\begin{aligned} \sum_{k=1}^{d} |g^k(t,x)|^2 \leq C(|z(t,x)|^2 + |Dz(t,x)|^2 + ||\mu(t)||^2_{L^2}), \\ |h(t,x)|^2 \leq C(|z(t,x)|^2 + |\mu(t,x)|^2), \\ |\operatorname{div}(h(t,x))|^2 \leq C(|z(t,x)|^2 + |Dz(t,x)|^2 + |\mu(t,x)|^2 + |D\mu(t,x)|^2). \end{aligned} \tag{C.2}$$

Let us point out that the system in (C.1) is different to the classical MFG and MFC systems since the data are both given in an initial point t_1 and the two equations are one forward and the other one backward.

Theorem C.1. *Under the above assumptions, $(z^k, \mu) = (0, 0)$ on $[t_1, T] \times \mathbb{R}^d$*

Proof. Without loss of generality, let us assume that $t_1 = 0$ and it is sufficient to prove that, for $T > 0$ sufficiently small, $((z^k)_{k=1,\dots,d}, \mu) = (0, 0)$ on $[0, T/2]$.
Let $\theta : [0, T] \to [0, 1]$ be a smooth non decreasing function so that $||\theta'(t)||_\infty \leq C/T$ and

$$\theta(t) = \begin{cases} 1 \text{ on } [0, T/2], \\ \theta(t) \text{ on } (T/2, 2T/3), \\ 0 \text{ on } [2T/3, T]. \end{cases}$$

For $c \geq 1$, we set

$$\widetilde{z^k}(x,t) = e^{c(t-T)^2/2}\theta(t)z^k(x,t) \quad \text{and} \quad \widetilde{\mu}(x,t) = e^{c(t-T)^2/2}\theta(t)\mu(x,t).$$

Then, $((\widetilde{z^k})_{k=1,\dots,d}, \widetilde{\mu})$ satisfies

$$\begin{cases} (i) & -\partial_t \widetilde{z^k} - \Delta \widetilde{z^k} + c(t-T)\widetilde{z^k} + e^{c(t-T)^2/2}\theta' z^k + e^{c(t-T)^2/2}\theta g_k = 0, \\ (ii) & \partial_t \widetilde{\mu} - \Delta \widetilde{\mu} - c(t-T)\widetilde{\mu} - e^{c(t-T)^2/2}\theta'\mu + e^{c(t-T)^2/2}\theta \operatorname{div}(h) = 0, \\ (iii) & \widetilde{z^k}(x,0) = \widetilde{\mu}(x,0) = \widetilde{z^k}(x,T) = \widetilde{\mu}(x,T) = 0. \end{cases} \tag{C.3}$$

Multiplying (C.3)-(ii) by $\partial_t \widetilde{\mu}$ and integrating in time-space, we obtain

$$\int_0^T \int_{\mathbb{R}^d} \partial_t\widetilde{\mu}\cdot\partial_t\widetilde{\mu} - \Delta\widetilde{\mu}\cdot\partial_t\widetilde{\mu} - c(t-T)\widetilde{\mu}\cdot\partial_t\widetilde{\mu} - e^{c(t-T)^2/2}\theta'\mu\cdot\partial_t\widetilde{\mu} + e^{c(t-T)^2/2}\theta\operatorname{div}(h)\cdot\partial_t\widetilde{\mu} = 0,$$

and after integrating by parts the second term and using that $\partial_t(\widetilde{\mu})^2 = 2\widetilde{\mu}\cdot\partial_t\widetilde{\mu}$

$$\int_0^T \int_{\mathbb{R}^d} (\partial_t\widetilde{\mu})^2 - \frac{1}{2}\partial_t|D\widetilde{\mu}|^2 - \frac{c}{2}(t-T)\partial_t(\widetilde{\mu})^2 - (e^{c(t-T)^2/2}\theta'\mu - e^{c(t-T)^2/2}\theta\operatorname{div}(h))\partial_t\widetilde{\mu} = 0.$$

Integrating the second term in time and by parts in time the third one (using the boundary conditions)

$$\int_0^T \int_{\mathbb{R}^d} (\partial_t\widetilde{\mu})^2 + \frac{c}{2}(\widetilde{\mu})^2 - (e^{c(t-T)^2/2}\theta'\mu - e^{c(t-T)^2/2}\theta\operatorname{div}(h))\partial_t\widetilde{\mu} = 0.$$

Using the Young's inequality on the third term

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} (\partial_t\widetilde{\mu})^2 + \frac{c}{2}(\widetilde{\mu})^2 \leq C \int_0^T \int_{\mathbb{R}^d} (e^{c(t-T)^2}(\theta')^2\mu^2 + e^{c(t-T)^2}\theta^2|\operatorname{div}(h)|^2) \\ + \frac{1}{2}\int_0^T \int_{\mathbb{R}^d} (\partial_t\widetilde{\mu})^2, \end{aligned} \tag{C.4}$$

and so

$$\int_0^T \int_{\mathbb{R}^d} \frac{1}{2}(\partial_t\widetilde{\mu})^2 + \frac{c}{2}(\widetilde{\mu})^2 \leq C \int_0^T \int_{\mathbb{R}^d} (e^{c(t-T)^2}(\theta')^2\mu^2 + e^{c(t-T)^2}\theta^2|\operatorname{div}(h)|^2). \tag{C.5}$$

Repeating the same steps for (C.3)-(i) multiply by $\partial_t\widetilde{z^k}$ and integrating in time-space

$$\int_0^T \int_{\mathbb{R}^d} \frac{1}{2}(\partial_t\widetilde{z^k})^2 + \frac{c}{2}(\widetilde{z^k})^2 \leq C \int_0^T \int_{\mathbb{R}^d} (e^{c(t-T)^2}(\theta')^2(z^k)^2 + e^{c(t-T)^2}\theta^2 g_k^2). \tag{C.6}$$

Consider (C.3)-(ii) multiply by $\widetilde{\mu}$ and integrating in time-space

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \widetilde{\mu} \cdot \widetilde{\mu} - \Delta \widetilde{\mu} \cdot \widetilde{\mu} - c(t-T)\widetilde{\mu} \cdot \widetilde{\mu} - e^{c(t-T)^2/2}\theta' \mu \cdot \widetilde{\mu} + e^{c(t-T)^2/2}\theta \operatorname{div}(h) \cdot \widetilde{\mu} = 0,$$

and considering the bounding conditions, integrating by parts the first and the second term and using the divergence theorem on the last one we get

$$\int_0^T \int_{\mathbb{R}^d} |D\widetilde{\mu}|^2 - c(t-T)(\widetilde{\mu})^2 + (-e^{c(t-T)^2/2}\theta'\mu) \cdot \widetilde{\mu} - (e^{c(t-T)^2/2}\theta h) \cdot D\widetilde{\mu} = 0.$$

Hence, by Young's inequality

$$\int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|D\widetilde{\mu}|^2 \leq \int_0^T \int_{\mathbb{R}^d} (\widetilde{\mu})^2 + C(e^{c(t-T)^2}(\theta')^2\mu^2 + e^{c(t-T)^2}\theta^2 h^2). \tag{C.7}$$

Repeating the same steps multiplying (C.3)-(i) by z^k and integrating in time-space

$$\int_0^T \int_{\mathbb{R}^d} |D\widetilde{z^k}|^2 + c(t-T)(\widetilde{z^k})^2 + (e^{c(t-T)^2/2}\theta' z^k + e^{c(t-T)^2/2}\theta g_k) \cdot \widetilde{z^k} = 0.$$

Then, for $\epsilon \in (0,1)$ to be chosen later

$$\int_0^T \int_{\mathbb{R}^d} |D\widetilde{z^k}|^2 \leq \int_0^T \int_{\mathbb{R}^d} (cT + \epsilon^{-1})(\widetilde{z^k})^2 + C\epsilon(e^{c(t-T)^2}(\theta')^2(z^k)^2 + e^{c(t-T)^2}\theta^2 g_k^2). \tag{C.8}$$

Since the assumption on h, g_k and $|\operatorname{div}(h)|^2$, the result obtain in (C.5) becomes

$$\frac{c}{2}\int_0^T ||\widetilde{\mu}(t)||_2^2 \leq C \int_0^T \int_{\mathbb{R}^d} (e^{c(t-T)^2}(\theta')^2||\mu(t)||_2^2 \\ + e^{c(t-T)^2}\theta^2(||z(t)||_2^2 + ||Dz(t)||_2^2 + ||\mu(t)||_2^2 + ||D\mu(t)||_2^2)).$$

Rearranging, noting that $e^{c(t-T)^2}\theta^2 \leq 1$ and using that $||v(t)||_{H^1}^2 := ||v(t)||_2^2 + ||Dv(t)||_2^2$,

$$c\int_0^T ||\widetilde{\mu}(t)||_2^2 \leq C\int_0^T (e^{c(t-T)^2}(\theta')^2||\mu(t)||_2^2 + ||\widetilde{z}(t)||_{H^1}^2 + ||\widetilde{\mu}(t)||_{H^1}^2). \tag{C.9}$$

The same argument for (C.6) yields to

$$c\int_0^T ||\widetilde{z^k}(t)||_2^2 + \leq C\int_0^T e^{c(t-T)^2}(\theta')^2||z^k(t)||_2^2 + ||\widetilde{z}(t)||_{H^1}^2 + ||\widetilde{\mu}(t)||_{H^1}^2. \tag{C.10}$$

Then (C.7) becomes

$$\int_0^T ||D\widetilde{\mu}(t)||_2^2 \leq C \int_0^T e^{c(t-T)^2}(\theta')^2||\mu(t)||_2^2 + ||\widetilde{z}(t)||_2^2 + ||\mu(t)||_2^2, \tag{C.11}$$

and (C.8)

$$\begin{aligned}\int_0^T ||D\widetilde{z^k}(t)||_2^2 \leq \int_0^T (cT+\epsilon^{-1})||\widetilde{z^k}(t)||_2^2 \\ + C\epsilon(e^{c(t-T)^2}(\theta')^2||\widetilde{z^k}(t)||_2^2 + ||\widetilde{z}(t)||_{H^1}^2 + ||\widetilde{\mu}(t)||_{H^1}^2).\end{aligned} \tag{C.12}$$

Summing (C.12) over k, for $\epsilon > 0$ small enough

$$\int_0^T ||D\widetilde{z}(t)||_2^2 \leq C\int_0^T (cT+1)||\widetilde{z}(t)||_2^2 + e^{c(t-T)^2}(\theta')^2||z(t)||_2^2 + ||\widetilde{\mu}(t)||_{H^1}^2.$$

Plugging (C.11) into the above inequality gives

$$\int_0^T ||D\widetilde{z}(t)||_2^2 \leq C\int_0^T (cT+1)||\widetilde{z}(t)||_2^2 + e^{c(t-T)^2}(\theta')^2(||z(t)||_2^2 + ||\mu(t)||_2^2) + ||\widetilde{\mu}(t)||_2^2. \tag{C.13}$$

Sum over k the inequality in (C.10) and collecting (C.9), (C.11) and (C.13)

$$\begin{aligned}c\int_0^T ||\widetilde{\mu}(t)||_2^2 + ||\widetilde{z}(t)||_2^2 &\leq C\int_0^T e^{c(t-T)^2}(\theta')^2(||\mu(t)||_2^2 + ||z(t)||_2^2) + ||\widetilde{z}(t)||_{H^1}^2 + ||\widetilde{\mu}(t)||_{H^1}^2 \\ &\leq C\int_0^T e^{c(t-T)^2}(\theta')^2(||\mu(t)||_2^2 + ||z(t)||_2^2) + (cT+1)||\widetilde{z}(t)||_2^2 + ||\widetilde{\mu}(t)||_2^2\end{aligned}$$

.

We can now fix $T > 0$ small enough, so that, for any c large enough

$$\frac{c}{2}\int_0^T ||\widetilde{\mu}(t)||_2^2 + ||\widetilde{z}(t)||_2^2 \leq C\int_0^T e^{c(t-T)^2}(\theta')^2(||\mu(t)||_2^2 + ||z(t)||_2^2).$$

By the definition of θ

$$\frac{c}{2}\int_0^{T/2} e^{c(t-T)^2}(||\mu(t)||_2^2 + ||z(t)||_2^2) \leq \frac{C}{T}\int_{T/2}^T e^{c(t-T)^2}(||\mu(t)||_2^2 + ||z(t)||_2^2).$$

Hence,

$$\frac{c}{2}e^{c(T/2)^2}\int_0^{T/2}(||\mu(t)||_2^2 + ||z(t)||_2^2) \leq \frac{C}{T}e^{c(T/2)^2}\int_{T/2}^T(||\mu(t)||_2^2 + ||z(t)||_2^2).$$

Dividing both terms in the last inequality by $e^{c(T/2)^2}$ and letting $c \to \infty$ yields to $(z, \mu) = (0, 0)$ on $[0, T/2]$.

$\square$

Bibliography

[1] Bogachev, V.I., Da Prato, G., Röckner, M.: *On parabolic equations for measures*, Commun. Partial Differ. Equ., 2008

[2] Briani, A., Cardaliaguet, P.: *Stable solutions in potential mean field game systems*, Nonlinear Differ. Equ. Appl. 25(1), 2018

[3] Cannarsa, P., Tessitore, M. E.: *Optimality conditions for boundary control problems of parabolic type. Control and estimation of distributed parameter systems: nonlinear phenomena*, (Vorau, 1993), 79-96, Internat. Ser. Numer. Math., 118, Birkäuser, Basel, 1994.

[4] Cardaliaguet, P., Cirant, M., Porretta, A.: *Splitting methods and short time existence for the master equations in mean field games*, To appear in JEMS, 2020

[5] Cardaliaguet, P., Daudin S., Jackson J., Souganidis P.: *An algebraic convergence rate for the optimal control of McKean-Vlasov dynamics*, arXiv preprint, arXiv:2203.14554

[6] Cardaliaguet, P., Delarue, F., Lasry, J. M., Lions, P.-L.: *The Master Equation and the Convergence Problem in Mean Field Games* (AMS-201) (Vol. 381). Princeton University Press, Princeton, 2019

[7] Cardaliaguet, P., Graber, J., Porretta, A., Tonon D.: *Second order mean field games with degenerate diffusion and local coupling*, Nonlinear Differ. Equ. Appl. 22, 5, (2015), 1287-1317.

[8] Cardaliaguet, P., Masoero, M.: *Weak KAM theory for potential MFG*, Journal of Differential Equations, 268(7), 3255-3298, 2020

[9] Cardaliaguet, P., Porretta, A.: *An introduction to Mean Field Game theory*, 2013

[10] Cardaliaguet, P., Souganidis, P. E.: *Regularity of the value function and quantitative propagation of chaos for mean field control problems*, Nonlinear Differential Equations and Applications NoDEA, 2023

[11] Carmona, R., Delarue, F.: *Probabilistic Theory of Mean Field Games with Applications I-II.* Springer, Berlin, 2018

[12] Chaintron, L. P., Diez, A.: *Propagation of chaos: a review of models, methods and applications. II. Applications*, 2022

[13] Daudin, S.: *Optimal control of the Fokker-Planck equation under state constraints in the Wasserstein space*, arXiv preprint. arXiv: 2109.14978

[14] Delarue, F., Lacker, D., Ramanan, K.: *From the master equation to mean field game limit theory: large deviations and concentration of measure*, Ann. Probab. 48(1), 2020

[15] Fournier, N., Guillin, A.: *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields, 162(3), 707-738, 2015

[16] Horowitz, J., Karandikar, R. L.: *Mean rates of convergence of empirical measures in the Wasserstein metric*, J. Comput. Appl. Math. 55, 1994

[17] Lacker, D.: *Limit theory for controlled McKean-Vlasov dynamics*, SIAM J. Control. Optim. 55(3), 2017

[18] Lasry, J.-M., Lions, P.-L.: *Mean field games*, Jpn. J. Math, 2007

[19] Lieberman, G.: *Second Order Parabolic Differential Equations*, World Scientific Publishing Co. Pte. Ltd., 1996

[20] Villani, C.: *Topics in Optimal Transportation*, American Mathematical Society, 2003