Mean Field Games

Tiago Miguel Saldanha Salvador

Math 581 Project

April 2013

Contents

1 Introduction **2**

2 Analysis of second order MFG **3**

2.1 On the Fokker-Plank equation 4

2.2 Existence of solutions to a 2^{nd} MFG 6

2.3 Uniqueness of solutions of a 2^{nd} order MFG 9

3 Analysis of first order MFG **11**

3.1 On the Hamilton-Jacobi equation 11

3.2 On the continuity equation 19

3.3 Existence of solutions to a 1^{st} order MFG 24

A Stochastic Calculus **25**

A.1 Brownian Motion and filtration 25

A.2 Stochastic integral and Itô's formula 26

A.3 Stochastic differential equations 27

B Auxiliary results **27**

1 Introduction

Mean Field Games (MFG) is a class of systems of partial differential equations that are used to understand the behaviour of multiples agents each individually trying to optimize their position in space and time, but with their preferences being partly determined by the choices of all other agents, in the asymptotic limit when the number of agents goes to infinity. This theory has been recently developed by J. M. Lasry and P. L. Lions in a series of papers [6, 7, 8, 9] and presented through several lectures of P. L. Lions at the Collège de France.

The typical model for MFG is the following:

$$\begin{cases} -\partial_t u - \nu\Delta u + H(x, m, D_x u) = F(x, m) \quad \text{in } \mathbb{R}^d \times [0, T], \\ \partial_t m - \nu\Delta m - div(D_p H(x, m, D_x u)m) = 0 \quad \text{in } \mathbb{R}^d \times [0, T], \\ m(0) = m_0, u(x, T) = G(x, m(T)) \quad \text{in } \mathbb{R}^d. \end{cases} \tag{MFG}$$

where ν is a non-negative parameter. The first equation is an Hamilton-Jacobi equation evolving backward in time whose solution is the value function of each agent. Indeed, the interpretation is the following: an average agent moves accordingly to the stochastic differential equation

$$dX_t = \alpha_t dt + \sqrt{2\nu} dW_t$$

where $W = \{W_t : t \in \mathbb{R}_+\}$ is a standard Brownian motion and α is the control to be chosen by the agent. He then wishes to minimize

$$\mathbb{E}\left[\int_0^T L(X_s, m(s), \alpha(s)) + F(X_s, m(s))ds + G(X_T, m(T))\right]$$

where L is the Legendre transform of H with respect to the last variable. The second equation is a Fokker-Planck type equation evolving forward in time that governs the evolution of the density function m of the agents.

In this report we will focus on studying the existence and uniqueness of solutions of MFG. In Section 2 we consider (MFG) with $\nu = 1$ and the Hamiltonian $H(p) = \frac{1}{2}|p|^2$, proving the existence and uniqueness of classical solutions. In Section 3 we consider the same Hamiltonian but with $\nu = 0$ and prove existence and uniqueness of (weak) solutions. For both sections we follow closely [3], trying to provide more detail in the proofs where it felt needed. Finally in the Appendix we review some basic definitions and results of stochastic calculus, as well as some results from measure theory that are used.

2 Analysis of second order MFG

Our goal in this Section is to prove the existence of classical solutions for the following MFG:

$$\begin{cases} -\partial_t u - \Delta u + \frac{1}{2}|D_x u|^2 = F(x,m) \quad \text{in } \mathbb{R}^d \times (0,T) \\ \partial_t m - \Delta m - div(m D_x u) = 0 \quad \text{in } \mathbb{R}^d \times (0,T) \\ m(0) = m_0, \; u(x,T) = G(x, m(T)) \quad \text{in } \mathbb{R}^d \end{cases} \tag{1}$$

Here $D_x u$ denotes the partial gradient with respect to x. We need to introduce some definitions.

Definition 2.1. *A pair (u,m) is a classical solution to (1) if $u, m \in C^{2,1}(\mathbb{R}^d \times (0,T)) \cap C(\mathbb{R}^d \times [0,T])$ and (u,m) satisfies (1) in the classical sense.*

Definition 2.2. *$\mathcal{P}$ is the set of Borel probability measures m on $\mathbb{R}^d$ with finite first order moment, i.e., $\int_{\mathbb{R}^d} |x| dm(x) < \infty$.*

We endow $\mathcal{P}$ with the following (Kantorovich-Rubistein) distance

$$\mathbf{d}(\mu,\nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^{2d}} |x-y| d\gamma(x,y)$$

where $\Pi(\mu,\nu)$ is the set of Borel probability measures on $\mathbb{R}^{2d}$ such that

$$\gamma(A \times \mathbb{R}^d) = \mu(A) \text{ and } \gamma(\mathbb{R}^d \times A) = \nu(A)$$

for any Borel set $A \subseteq \mathbb{R}^d$.

We can now state the main theorem of this Section:

Theorem 2.3. *Suppose there is some constant C_0 such that*

- *(Bounds on F and G) F and G are uniformly bounded by C_0 over $\mathbb{R}^d \times \mathcal{P}$,*
- *(Lipschitz continuity of F and G) For all $(x_1,m_1),(x_2,m_2) \in \mathbb{R}^d \times \mathcal{P}$, we have*

$$|F(x_1,m_1) - F(x_2,m_2)| \le C_0 \left(|x_1 - x_2| + \boldsymbol{d}(m_1,m_2)\right)$$

and

$$|G(x_1,m_1) - G(x_2,m_2)| \le C_0 \left(|x_1 - x_2| + \boldsymbol{d}(m_1,m_2)\right),$$

- *The probability measure m_0 is absolutely continuous with respect to the Lebesgue measure, denoted by $\mathcal{L}^d$ and has a Hölder continuous density, still denote by m_0, which satisfies*

$$\int_{\mathbb{R}^d} |x|^2 m_0(x)dx \le C_0.$$

Then there is at least one classical solution to (1)*.*

We will first treat two PDE's in (1) separately: we obtain some estimates on the Fokker-Planck equation and recall some known facts of the heat equation.

2.1 On the Fokker-Plank equation

In this Section we will derive some results on the following Fokker-Planck equation

$$\begin{cases} \partial_t m - \Delta m - div(mb) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ m(0) = m_0 \end{cases} \tag{2}$$

where $b : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ is a given vector field. We can look at it as an evolution equation on the space of probability measures. We will assume that the vector field b is continuous, uniformly Lipschitz in space and bounded. The reason for this is that in the proof of Theorem 2.3 we will take $b = -D_x u$.

Definition 2.4. *We say that m is a weak solution to* (2) *if $m \in L^1([0,T],\mathcal{P})$ is such that for any test function $\varphi \in \mathcal{D}(\mathbb{R}^d \times [0,T))$ we have*

$$\int_{\mathbb{R}^d} \varphi(x,0)dm_0(x) + \int_0^T \int_{\mathbb{R}^d} \left(\partial_t\varphi(x,t) + \Delta\varphi(x,t) - D_x\varphi(x,t)\cdot b(x,t)\right) dm(t)(x).$$

Consider the following stochastic differential equation (SDE)

$$\begin{cases} dX_t = -b(X_t,t)dt + \sqrt{2}dW_t & t \in [0,T] \\ X_0 = Z_0 \end{cases} \tag{3}$$

where W_t is a standard d-dimensional Brownian motion and the initial condition $Z_0 \in \mathbb{L}_1$ is random variable independent of W_t. Under the assumption on b by Theorem A.11 there is a unique solution to (3). The next Lemma shows that the solution of (3) is closely related to the solution of (2).

Lemma 2.5. *If $\mathcal{L}(Z_0) = m_0$, then $m(t) := \mathcal{L}(X_t)$ is a weak solution of* (2)*, where X_t is the solution of* (3)*. Here $\mathcal{L}(X)$ denotes the law (density function) of the random variable X.*

Proof. This is a straightforward consequence of Itô's formula. Indeed, let $\varphi \in C^{2,1}(\mathbb{R}^d \times [0,T])$. Then by Itô's formula (Theorem A.9)

$$\varphi(X_t,t) = \varphi(Z_0,0) + \int_0^T \left(\partial_t\varphi(X_s,s) - D_x\varphi(X_s,s)\cdot b(X_s,s) + \Delta\varphi(X_s,s)\right) ds + \int_0^T D_x\varphi(X_s,s)\cdot dW_s.$$

We know that

$$\mathbb{E}\left[\int_0^T D_x\varphi(X_s,s)\cdot dW_s\right] = 0.$$

Hence taking the expectation on the above equality leads to

$$\mathbb{E}\left[\varphi(X_t,t)\right] = \mathbb{E}\left[\varphi(Z_0,0) + \int_0^t \left(\partial_t\varphi(X_s,s) - D_x\varphi(X_s,s)\cdot b(X_s,s) + \Delta\varphi(X_s,s)\right) ds\right].$$

So by the definition of m we have

$$\int_{\mathbb{R}^d} \varphi(x,t)dm(t)(x) = \int_{\mathbb{R}^d} \varphi(x,0)dm_0(x) + \int_0^t \int_{\mathbb{R}^d} \left(\partial_t\varphi(x,s) - D\varphi(x,s)\cdot b(x,s) + \Delta\varphi(x,s)\right) dm(s)(x)ds.$$

Therefore for $\varphi \in \mathcal{D}\left(\mathbb{R}^d \times [0,T)\right)$ and taking $t = T$ we have

$$\int_{\mathbb{R}^d} \varphi(x,0)dm_0(x) + \int_0^T \int_{\mathbb{R}^d} \left(\partial_t\varphi(x,t) - D\varphi(x,t)\cdot b(x,t) + \Delta\varphi(x,t)\right) dm(t)(x)ds = 0,$$

i.e., m is a weak solution of (2). □

The above interpretation of m_0 as the probability density of the solution of (3) allows us to show that the map $t \mapsto m(t)$ is Hölder continuous.

Lemma 2.6. *Let $m(t) := \mathcal{L}(X_t)$ where X_t is the solution of* (3)*. Then there is a constant $c_0 = c_0(T)$ (i.e., depending only on T), such that for all $s,t \in [0,T]$*

$$\boldsymbol{d}(m(t),m(s)) \leq c_0(1 + \|b\|_\infty)|t-s|^{1/2}.$$

Proof. We start by observing that the probability measure γ of the pair (X_t, X_s) belongs to $\Pi(m(t), m(s))$. Therefore

$$\mathbf{d}(m(t),m(s)) \leq \int_{\mathbb{R}^{2d}} |x-y|d\gamma(x,y) = \mathbb{E}\left[|X_t - X_s|\right].$$

Without loss of generality suppose $s < t$. Then

$$\begin{aligned}
\mathbb{E}\left[|X_t - X_s|\right] &= \mathbb{E}\left[\left|\int_s^t b(X_\tau,\tau)d\tau + \sqrt{2}(W_t - W_s)\right|\right] \\
&\leq \mathbb{E}\left[\int_s^t |b(X_\tau,\tau)|d\tau + \sqrt{2}|W_t - W_s|\right] \\
&\leq \|b\|_\infty(t-s) + \frac{2}{\sqrt{\pi}}\sqrt{t-s} \\
&\leq \sqrt{t-s}\left(\|b\|_\infty\sqrt{T} + \frac{2}{\sqrt{\pi}}\right) \\
&\leq \sqrt{t-s}\left(\|b\|_\infty + 1\right)\max\left\{\sqrt{T}, \frac{2}{\sqrt{\pi}}\right\}
\end{aligned}$$

So by taking $c_0 = \max\left\{\sqrt{T}, \frac{2}{\sqrt{\pi}}\right\}$ we are done. □

We can also obtain easily an estimate on the second order moment of m

Lemma 2.7. *Let $m(t) := \mathcal{L}(X_t)$ where X_t is the solution of* (3)*. Then there is a constant $c_0 = c_0(T)$ such that for all $t \in [0,T]$*

$$\int_{\mathbb{R}^d} |x|^2 dm(t)(x) \leq c_0\left(\int_{\mathbb{R}^d} |x|^2 dm_0(x) + 1 + \|b\|_\infty^2\right).$$

Proof. By definition of m we have

$$\int_{\mathbb{R}^d} |x|^2 dm(t)(x) = \mathbb{E}\left[|X_t|^2\right]$$

Hence

$$\begin{aligned}\int_{\mathbb{R}^d} |x|^2 dm(t)(x) &\le 3\mathbb{E}\left[|X_0|^2 + \left|\int_0^t b(X_s,s)ds\right|^2 + 2|W_t|^2\right] \\ &\le 3\left(\int_{\mathbb{R}^d} |x|^2 dm_0(x) + \|b\|_\infty^2 t^2 + 2t\right) \\ &\le c_0\left(\int_{\mathbb{R}^d} |x|^2 dm_0(x) + \|b\|_\infty^2 + 1\right)\end{aligned}$$

where $c_0 = \max\{3, 3T^2, 6T\}$. □

2.2 Existence of solutions to a 2^{nd} MFG

In this Section we prove Theorem 2.3. In order to do that we need first to recall some existence and uniqueness results for the following heat equation

$$\begin{cases}\partial_t w - \Delta w + a(x,t)\cdot Dw + b(x,t)w = f(x,t) & \text{in } \mathbb{R}^d \times [0,T] \\ w(x,0) = w_0(x) & \text{in } \mathbb{R}^d\end{cases} \tag{4}$$

where $a, b, f : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $w_0 : \mathbb{R}^d \to \mathbb{R}$. For this we introduce some notation.

Definition 2.8. *Let $s \ge 0$ be an integer and $\alpha \in (0,1)$. We denote by $\mathcal{C}^{s,\alpha}(\mathbb{R}^d \times [0,T])$ the set of functions $f : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ such that for any pair (k,l) with $2k + l \le s$, $\partial_t^k D_x^l f$ exists and such that these derivatives are bounded, α-Hölder continous in space and $\alpha/2$-Hölder continous in time.*

We then have the following theorem whose proof can be found in [5]:

Theorem 2.9. *Suppose that $a, b, f \in \mathcal{C}^{0,\alpha}(\mathbb{R}^d \times [0,T])$ and that $w_0 \in C^{0,\alpha}(\mathbb{R}^d)$ (the classical Hölder space). Then (4) has a unique weak solution $u \in \mathcal{C}^{2,\alpha}(\mathbb{R}^d \times [0,T])$.*

We also have the following interior estimate:

Theorem 2.10. *Suppose $a \equiv b \equiv 0$ and that $f \in C(\mathbb{R}^d \times [0,T])$ is bounded. Then any classical bounded solution w of (4) satisfies, for any compact set $K \subseteq \mathbb{R}^d \times (0,T)$*

$$\sup_{(x,t),(y,s)\in K} \frac{|D_x w(x,t) - D_x w(y,s)|}{|x-y|^\beta + |t-s|^{\beta/2}} \le C\|f\|_\infty$$

where $\beta \in (0,1)$ depends only on the dimension d and $C = C(K, \|w\|_\infty, d)$.

The idea of the proof is to construct a map Ψ such that a fixed point of Ψ is a solution of the system (1). Then we use the Schauder fixed point theorem to prove the existence of the fixed point.

Theorem 2.11 (Schauder fixed point)**.** *Let K be a convex, closed and compact subspace of a topological vector space V and $\Psi : K \to K$ a continuous map. Then Ψ has a fixed point.*

Proof of Theorem 2.3. Let C be a large constant to be fixed later and let $\mathcal{M}$ be the set of maps $\mu \in C([0,T],\mathcal{P})$ such that

$$\sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{\mathbf{d}(\mu(t),\mu(s))}{|t-s|^{1/2}} \leq C$$

and

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d\mu(t)(x) \leq C.$$

To any $\mu \in \mathcal{M}$ we associate an $m = \Psi(\mu) \in \mathcal{M}$ in the following way: let u be the unique solution to

$$\begin{cases} -\partial_t u - \Delta u + \frac{1}{2}|D_x u|^2 = F(x,\mu(t)) & \text{in } \mathbb{R}^d \times [0,T] \\ u(x,T) = G(x,\mu(T) & \text{in } \mathbb{R}^d \end{cases} \tag{5}$$

Then we define $m = \psi(\mu) \in \mathcal{M}$ as the unique solution of the Fokker-Plank equation

$$\begin{cases} \partial_t m - \Delta m - div(mD_x u) = 0 & \text{in } \mathbb{R}^d \times [0,T] \\ m(0) = m_0(x) & \text{in } \mathbb{R}^d \end{cases} \tag{6}$$

In order to apply the Schauder fixed point theorem, we need to show that: $\mathcal{M}$ is a convex closed and compact subset of $C([0,T],\mathcal{P})$, Ψ is well defined and Ψ is continuous.

1) $\mathcal{M}$ is a convex closed and compact subset of $C([0,T],\mathcal{P})$.

Let $\lambda \in [0,1]$, $\mu_1,\mu_2 \in \mathcal{M}$, $\gamma_1 \in \Pi(\mu_1(t),\mu_1(s))$ and $\gamma_2 \in \Pi(\mu_2(t),\mu_2(s))$. We have that

$$\lambda\gamma_1 + (1-\lambda)\gamma_2 \in \Pi(\lambda\mu_1(t) + (1-\lambda)\mu_2(t), \lambda\mu_1(s) + (1-\lambda)\mu_2(s))$$

and therefore

$$\begin{aligned} &\mathbf{d}(\lambda\mu_1(t) + (1-\lambda)\mu_2(t), \lambda\mu_1(s) + (1-\lambda)\mu_2(s)) \\ &\leq \int_{\mathbb{R}^{2d}} |x-y| d(\lambda\gamma_1(x,y) + (1-\lambda)\gamma_2(x,y)) \\ &= \lambda \int_{\mathbb{R}^{2d}} |x-y| d\gamma_1(x,y) + (1-\lambda)\int_{\mathbb{R}^{2d}} |x-y| d\gamma_2(x,y)). \end{aligned}$$

Then taking the infimum over $\gamma_1 \in \Pi(\mu_1(t),\mu_1(s))$ and $\gamma_2 \in \Pi(\mu_2(t),\mu_2(s))$ shows that

$$\mathbf{d}(\lambda\mu_1(t) + (1-\lambda)\mu_2(t), \lambda\mu_1(s) + (1-\lambda)\mu_2(s)) \leq \lambda\mathbf{d}(\mu_1(t),\mu_1(s)) + (1-\lambda)\mathbf{d}(\mu_2(t),\mu_2(s)).$$

We also have

$$\int_{\mathbb{R}^d} |x|^2 d(\lambda\mu_1 + (1-\lambda)\mu_2)(t)(x) = \lambda\int_{\mathbb{R}^d} |x|^2 d\mu_1(t)(x) + (1-\lambda)\int_{\mathbb{R}^d} |x|^2 d\mu_2(t)(x).$$

From the last two equalities it's now easy to see that, indeed, $\lambda\mu_1 + (1-\lambda)\mu_2 \in \mathcal{M}$ and so $\mathcal{M}$ is convex.

Now let $\mu_n \in \mathcal{M}$ such that $\mu_n \to \mu$ in $C([0,T],\mathcal{P})$. To prove that $\mathcal{M}$ is closed we need to show that $\mu \in \mathcal{M}$.

It's easy to show that

$$\mathbf{d}(\mu(t),\mu(s)) \leq \mathbf{d}(\mu(t)-\mu_n(t),\mu(s)-\mu_n(s)) + \mathbf{d}(\mu_n(t),\mu_n(s))$$

and from this it follows easily that

$$\sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{\mathbf{d}(\mu(t),\mu(s))}{|t-s|^{1/2}} \leq C.$$

As for the second order moment estimate we note that

$$\int_{\mathbb{R}^d} |x|^2 d\mu(t)(x) = \int_{\mathbb{R}^d} |x|^2 d(\mu(t)-\mu_n(t))(x) + \int_{\mathbb{R}^d} |x|^2 d\mu_n(t)(x)$$

Taking the supremum for $t \in [0,T]$ we get

$$\begin{aligned}\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d\mu(t)(x) &\leq \sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d(\mu(t)-\mu_n(t))(x) + \sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d\mu_n(t)(x) \\ &\leq \sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d(\mu(t)-\mu_n(t))(x) + C\end{aligned}$$

Now since $\mu_n \to \mu$ in $C([0,T],\mathcal{P})$, by taking $n\to\infty$ we get

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^d} |x|^2 d\mu(t)(x) \leq C$$

as desired.

For the proof that $\mathcal{M}$ is compact we refer the reader to Lemma 5.7 of [3].

2) ψ is well-defined.

First we need to see that a solution of (5) exists and is unique. Consider then the Hopf-Cole transformation given by $w = e^{u/2}$. Then it is easy to check that u is a solution of (5) if and only if w is a solution of

$$\begin{cases} -\partial_t w - \Delta w = wF(x,\mu(t)) & \text{in } \mathbb{R}^d \times [0,T] \\ w(x,T) = e^{G(x,\mu(T))/2} & \text{in } \mathbb{R}^d \end{cases} \tag{7}$$

The map $(x,t) \mapsto F(x,m(t))$ belongs to $\mathcal{C}^{0,1/2}$ since F is Lipschitz in both variables, uniformly bounded over $\mathbb{R}^d \times \mathcal{P}$ and

$$\sup_{\substack{s,t\in[0,T]\\ s\neq t}} \frac{\mathbf{d}(\mu(t),\mu(s))}{|t-s|^{1/2}} \leq C.$$

because $\mu \in \mathcal{M}$. The map $x \mapsto e^{G(x,\mu(T))/2}$ is in $C^{0,1/2}(\mathbb{R}^d)$ since G is Lipschitz in x and uniformly bounded over $\mathbb{R}^d \times \mathcal{P}$. Then appealing to Theorem 2.9 there is a unique solution in $\mathcal{C}^{2,1/2}$ to (7) which implies that there is a unique solution in $\mathcal{C}^{2,1/2}$ to (5). Recall that, by assumption, the maps $(x,t) \mapsto F(x,m(t))$ and $x \mapsto G(x,\mu(T)$ are bounded by C_0. Hence a straightforward application of the comparison principle implies that u is bounded by $(1+T)C_0$. Similarly the maps $x \mapsto F(x,m(t))$ and $x \mapsto G(x,\mu(T)$ are C_0-Lipschitz continuous (again by our assumptions on F and G) and so u is also C_0-Lipschitz continuous. Hence $D_x u$ is bounded by C_0.

Now we look at the Fokker-Planck equation (6). By expanding the divergence term, we can write it into the form

$$\begin{cases} \partial_t m - \Delta m - D_x m \cdot D_x u(x,t) - m\Delta u(x,t) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ m(0) = m_0 \end{cases}$$

Since $u \in \mathcal{C}^{2,1/2}$, the maps $(x,t) \mapsto D_x u(x,t)$ and $(x,t) \mapsto \Delta u(x,t)$ belong to $\mathcal{C}^{0,1/2}$. Also by assumption $m_0 \in C^{0,\alpha}(\mathbb{R}^d)$. Hence by Theorem (2.9) there is a unique solution $m \in \mathcal{C}^{2,1/2}$ to (6). Moreover, by Lemma 2.6, for all $s,t \in [0,T]$

$$\mathbf{d}(m(t), m(s)) \le c_0(1 + C_0)|t-s|^{\frac{1}{2}}$$

and by Lemma 2.7 for all $t \in [0,T]$

$$\int_{\mathbb{R}^d} |x|^2 dm(t)(x) \le c_0(C_0 + 1 + C_0^2)$$

where c_0 depends only on T. So if we choose $C = max\{c_0(1+C_0), c_0(C_0+1+C_0^2)\}$, $m \in \mathcal{M}$ and Ψ is then well-defined.

3) Ψ is continuous.

Let $\mu_n \in \mathcal{M}$ converge to some μ. Let (u_n, m_n) and (u,m) be the corresponding solutions. Note that $(x,t) \mapsto F(x, \mu_n(t))$ and $x \mapsto G(x, \mu_n(T))$ converge locally uniformly to $(x,t) \mapsto F(x,\mu(t))$ and $x \mapsto G(x, \mu(T)$ respectively. Hence we can conclude that (u_n) converges locally uniformly to u by a standard argument with viscosity solutions. Since the $(D_x u_n)$ are uniformly bounded (by C_0), the (u_n) solve an equation of the form

$$\partial_t u_n - \Delta u_n = f_n$$

where $f_n = \frac{1}{2}|D_x u_n|^2 - F(x, m_n)$ is uniformly bounded in x and n. Then by Theorem 2.10 $(D_x u_n)$ is locally uniform Hölder continuous and therefore converge locally uniform to $D_x u$. This implies that any converging subsequence of the relatively compact sequence (m_n) is a weak solution of (6). But m is the unique solution of (6), which proves that (m_n) converges to m. Hence Ψ is continuous.

Finally, by the Schauder fixed point theorem, the continuous map $\mu \mapsto m = \Psi(\mu)$ has a fixed point in $\mathcal{M}$. To this fixed point $m \in \mathcal{M}$ corresponds a pair (u,m) that is a classical solution to (1) and so we are done. □

2.3 Uniqueness of solutions of a 2^{nd} order MFG

In this Section we prove a uniqueness result to the system (1).

Theorem 2.12. *Besides the assumptions of Theorem 2.3, assume that*

- *For all $m_1, m_2 \in \mathcal{P}$ with $m_1 \neq m_2$ we have*

$$\int_{\mathbb{R}^d} (F(x,m_1) - F(x,m_2)) d(m_1 - m_2)(x) > 0,$$

- *For all $m_1, m_2 \in \mathcal{P}$*
$$\int_{\mathbb{R}^d} (G(x, m_1) - G(x, m_2)) d(m_1 - m_2)(x) \geq 0.$$

Then there is a unique solution to (1).

Proof. Let (u_1, m_1) and (u_2, m_2) be two classical solutions of (1). We set $u = u_1 - u_2$ and $m = m_1 - m_2$. Then

$$\begin{cases} -\partial_t u - \Delta u + \frac{1}{2}(|D_x u_1|^2 - |D_x u_2|^2) - (F(x, m_1) - F(x, m_2)) = 0 \\ \partial_t m - \Delta m - div(m_1 D_x u_1 - m_2 D_x u_2) = 0 \end{cases} \tag{8}$$

Since $u \in C^{2,1}(\mathbb{R}^d \times (0, T))$, we can multiply the second equation by u, integrate over $\mathbb{R}^d \times [0, T]$, followed by part integration to get

$$-\int_{\mathbb{R}^d} m(T) u(x, T) dx + \int_{\mathbb{R}^d} m_0(x) u(x, 0) dx + \int_0^T \int_{\mathbb{R}^d} (\partial_t u + \Delta u) m - Du \cdot (m_1 D_x u_1 - m_2 D_x u_2) dx dt.$$

Multiplying now the first equation by m, integrating over $\mathbb{R}^d \times [0, T]$ and adding to the previous equality, leads to

$$\begin{aligned} &-\int_{\mathbb{R}^d} m(T)(G(x, m_1(T)) - G(x, m_2(T)) dx \\ &+ \int_0^T \int_{\mathbb{R}^d} \Big(-\frac{m}{2} |D_x u_1 - D_x u_2|^2 - m(F(x, m_1) - F(x, m_2)))\Big) dx dt = 0 \end{aligned}$$

where we used the fact that $m(0) = 0$ and that

$$\frac{m}{2}(|D_x u_1|^2 - |D u_2|^2) - D_x u \cdot (m_1 D_x u_1 - m_2 D_x u_2) = -\frac{m}{2} |D_x u_1 - D_x u_2|^2.$$

By assumption

$$\int_{\mathbb{R}^d} m(T)(G(x, m_1(T)) - G(x, m_2(T)) dx \geq 0$$

and therefore

$$\int_0^T \int_{\mathbb{R}^d} m(F(x, m_1) - F(x, m_2)) dx dt \leq 0.$$

Hence, by our assumption on F, this implies that $m = 0$ and therefore $u = 0$ since u_1 and u_2 (now) solve the same equation. □

We finish this Section by mentioning that the existence of solutions for second order MFG hold under more general assumptions. Indeed, in [7, 8] the authors consider equations of the form

$$\begin{cases} -\partial_t u(x, t) - \Delta u + H(x, Du) = F(x, m)) & \text{in } Q \times (0, T) \\ \partial_t m(x, t) - \Delta m - div(m \frac{\partial H}{\partial p}(x, D_x u)) = 0 & \text{in } Q \times (0, T) \\ m(0) = m_0, \ u(x, T) = G(x, m(T)) & \text{in } Q \end{cases}$$

where $Q = [0, 1]^d$ (with periodic boundary conditions), $H : \mathbb{R}^d \times \mathbb{R}^d$ is Lipschitz continuous with respect to x and uniformly bounded in p, convex and of class C^1 with respect to p. The conditions on F and G are one of the following:

- F and G are regularizing, i.e., satisfy the same conditions as in Theorem 2.3.
- $F(x, m) = f(x, m(x))$ and $G(x, m) = g(x, m(x))$, where $f = f(x, \lambda)$ and $g = g(x, \lambda)$ satisfy suitable growth conditions with respect to λ and H is sufficiently strictly convex.

3 Analysis of first order MFG

In this Section we will prove the existence of solutions to the following first order MFG:

$$\begin{cases} -\partial_t u(x,t) + \frac{1}{2}|Du(x,t)|^2 = F(x, m(t)) & \text{in } \mathbb{R}^d \times (0,T) \\ \partial_t m(x,t) - div(Du(x,t)m(x,t) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ m(0) = m_0, u(x,T) = G(x, m(T)) & \text{in } \mathbb{R}^d \end{cases} \tag{9}$$

We consider the following definition of weak solutions.

Definition 3.1. *We call the pair (u, m) a weak solution of (9) if $u \in W^{1,\infty}_{loc}(\mathbb{R}^d \times [0,T])$, $m \in L_1(\mathbb{R}^d \times (0,T))$ such that the first equation of (9) is satisfied in the viscosity sense and the second in satisfied in the sense of distributions.*

Note that here we don't look any more for classical solutions mainly because we no longer have the smoothing terms Δu and Δm. Our goal is then to prove the following.

Theorem 3.2. *Suppose that*

1. *F and G are continuous over $\mathbb{R}^d \times \mathcal{P}$,*
2. *There is a constant C such that for any $m \in \mathcal{P}$, $F(\cdot, m), G(\cdot, m) \in C^2(\mathbb{R}^d)$ and*
$$\|F(\cdot, m)\|_{C^2(\mathbb{R}^d)} \leq C \quad \|G(\cdot, m)\|_{C^2(\mathbb{R}^d)} \leq C$$
where for $f \in C^2(\mathbb{R}^d)$ we denote $\|\cdot\|_{C^2(\mathbb{R}^d)}$ by
$$\|f\|_{C^2(\mathbb{R}^d)} = \sup_{x \in \mathbb{R}^d} \{|f(x)| + |Df(x)| + |D^2 f(x)|\},$$
3. *m_0 is absolutely continuous with respect to the Lebesgue measure and has a density, still denoted by m_0, which is bounded and has a compact support.*

Then there is at least one weak solution of (9).

Remark 3.3. *Under the assumptions of Theorem 2.12 we can show that the solution is unique. The proof is the same with the only difference being that now we use the Lipschitz continuous map u as a test function because the density m is bounded and has compact support.*

As in Section 2, we will study the two equations separately.

3.1 On the Hamilton-Jacobi equation

In this Section we study the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u + \frac{1}{2}|D_x u|^2 = f(x,t) & \text{in } \mathbb{R}^d \times (0,T) \\ u(x,T) = g(x) & \text{in } \mathbb{R}^d \end{cases} \tag{10}$$

We will start by recalling some basic facts about the notion of semi-concavity which will play a role here. The proofs for these results can be found in [2].

Definition 3.4. *A map $w : \mathbb{R}^d \to \mathbb{R}$ is semi-concave if there is some constant $C > 0$ such that one of the following equivalent conditions is satisfied:*

1. *the map $x \mapsto w(x) - \frac{C}{2}|x|^2$ is concave in $\mathbb{R}^d$.*
2. *$w(\lambda x + (1-\lambda)y) \geq \lambda w(x) + (1-\lambda)w(y) - C\lambda(1-\lambda)|x-y|^2$ for any $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$.*
3. *$D^2 w \leq CI_d$ in the sense of distributions.*
4. *$(p-q)\cdot(x-y) \leq C|x-y|^2$ for any $x, y \in \mathbb{R}^d$, $t \in [0,T]$, $p \in D_x^+ w(x)$ and $q \in D_x^+ w(y)$, where $D_x^+ w$ denotes the super-differential of w with respect to the x variable, namely*

$$D_x^+ w(x) = \{p \in \mathbb{R}^d : \limsup_{y \to x} \frac{w(y) - w(x) - p\cdot(y-x)}{|y-x|} \leq 0\}.$$

Lemma 3.5. *Let $w : \mathbb{R}^d \to \mathbb{R}$ be semi-concave. Then w is locally Lipschitz continuous in $\mathbb{R}^d$. Moreover $D_x^+ w(x)$ is the closed convex hull of the set $D_x^* w(x)$ of reachable gradients defined by*

$$D_x^* w(x) = \{p \in \mathbb{R}^d : \exists\,(x_n) \text{ with } x_n \to x \text{ such that } D_x w(x_n) \text{ exists and converges to } p\}$$

In particular, $D_x^+ w(x)$ is compact, convex and non empty subset of $\mathbb{R}^d$ for any $x \in \mathbb{R}^d$. Finally w is differentiable at x if and only if $D^+ w(x)$ is a singleton.

Lemma 3.6. *Let (w_n) be a sequence of uniformly semi-concave maps on $\mathbb{R}^d$ which converge pointwiseto a map $w : \mathbb{R}^d \to \mathbb{R}$. Then the convergence is locally uniform and w is semi-concave. Moreover, for any $x_n \to x$ and any $p_n \in D^+ w_n(x_n)$, the set of accumulation points of (p_n) is contained in $D^+ w(x)$. Finally, $Dw_n(x)$ converges to $Dw(x)$ for a.e. $x \in \mathbb{R}^d$.*

Definition 3.7. *Let $(x,t) \in \mathbb{R}^d \times [0,T]$. We denote by $\mathcal{A}(x,t)$ the nonempty set of optimal controls to $u(x,t)$, i.e., $\alpha \in L^2([t,T], \mathbb{R}^d)$ such that*

$$u(x,t) = \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\right) ds + g(x(T))$$

where $x(s) = x + \int_t^s \alpha(\tau) d\tau$. We call $x(\cdot)$ the associated trajectory to the control α.

Lemma 3.8. *If $(x_n, t_n) \to (x,t)$ with $\alpha_n \in \mathcal{A}(x_n, t_n)$, then, up to a subsequence, (α_n) weakly converges in L^2 to some $\alpha \in \mathcal{A}(x,t)$.*

We can now study equation (10).

Lemma 3.9. *Let $f : \mathbb{R}^d \times [0,T] \to \mathbb{R}$ and $g : \mathbb{R}^d \to \mathbb{R}$ be continuous functions. For any $t \in [0,T]$, $f(\cdot,t), g \in C^2(\mathbb{R}^d)$ with*

$$\|f(\cdot,t)\|_{C^2} \leq C, \quad \|g\|_{C^2} \leq C \tag{11}$$

for some constant C. Then equation (10) has a unique bounded uniformly continuous viscosity solution which is given by the representation formula

$$u(x,t) = \inf_{\alpha \in L^2([t,T],\mathbb{R}^d)} \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x(s),s)\right) ds + g(x(T)),$$

where $x(s) = x + \int_s^t \alpha(\tau)d\tau$. Moreover u is Lipschitz continuous and satisfies

$$\|D_{x,t}u\|_\infty \le C_1, \quad D^2_{xx}u \le C_1 I_d$$

where the last inequality holds in the sense of distributions.

Proof. From the theory of Hamilton-Jacobi equations we already know that (10) has a unique bounded uniformly continuous viscosity solution given by

$$u(x,t) = \int_{\alpha \in L^2([t,T],\mathbb{R}^d)} \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x(s),s)\right) ds + g(x(T)).$$

Hence we only need to check that u is Lipschitz continuous with $\|D_{x,t}u\|_\infty \le C_1$ and $D^2_{xx}u \le C_1 I_d$ in the sense of distributions for some constant $C_1 = C_1(T)$.

1) u is Lipschitz continuous with respect to x.

Let $x_1, x_2 \in \mathbb{R}^d$, $t \in [0,T]$ and $\alpha \in \mathcal{A}(x,t)$. We then have

$$\begin{aligned}
u(x_2,t) &\le \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x(s)+x_2-x_1,s)\right) ds + g(x(T)+x_2-x_1) \\
&\le \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x(s),s) + C|x_2-x_1|\right) ds + g(x(T)) + C|x_2-x_1| \\
&\le u(x_1,t) + C(T+1)|x_2-x_1|
\end{aligned}$$

Thus u is Lipschitz continuous with respect to x with Lipschitz constant $C(T+1)$.

2) u is Lipschitz continuous with respect to t.

Fix $x \in \mathbb{R}^d$ and $t \in [0,T]$. From the dynamic programming principle we have for any $t < s \le T$

$$u(x,t) = \int_t^s \frac{1}{2}|\alpha(\tau)|^2 + f(x(\tau),\tau)d\tau + u(x(s),s)$$

where $\alpha \in \mathcal{A}(x,t)$ and $x(\cdot)$ is its associated trajectory. We have

$$\begin{aligned}
|u(x,t) - u(x,s)| &\le |u(x,t) - u(x(s),s)| + |u(x(s),s) - u(x,s)| \\
&\le \int_t^s \left(\frac{1}{2}|\alpha(\tau)|^2 + |f(x(\tau),\tau)|\right) d\tau + C(T+1)|x(s)-x| \\
&\le (s-t)\left(\frac{1}{2}\|\alpha\|^2_\infty + \|f\|_\infty + C(T+1)\right)
\end{aligned}$$

where in the second inequality we used the fact u is $C(T+1)$-Lipschitz continuous with respect to x. In Lemma 3.10, we show that α is bounded by a constant $C_2 = C_2(T)$. Hence the inequality above proves that u is Lipschitz continuous with respect to t.

3) $\|D_{x,t}u\|_\infty \le C_1$.

It follows easily from 1) and 2).

4) $D^2_{xx}u \leq C_1 I_d$ in the sense of distributions.

Let $x, y \in \mathbb{R}^d$, $t \in [0,T]$, $\lambda \in [0,1]$ and set $x_\lambda = \lambda x + (1-\lambda)y$. By Definition 3.4 it's enough to show that

$$\lambda u(x,t) + (1-\lambda)u(y,t) \leq u(x_\lambda, t) + \tilde{C}\lambda(1-\lambda)|x-y|^2$$

where $\tilde{C} = \tilde{C}(T)$ is a constant. Let $\alpha \in \mathcal{A}(x_\lambda, t)$ and $x_\lambda(\cdot)$ its associated trajectory. Then

$$\begin{aligned}
\lambda u(x,t) + (1-\lambda)u(y,t) &\leq \lambda \left[\int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x_\lambda(s) + x - x_\lambda, s) \right) ds + g(x_\lambda(T) + x - x_\lambda) \right] \\
&\quad + (1-\lambda) \left[\int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x_\lambda(s) + y - x_\lambda, s) \right) ds + g(x_\lambda(T) + y - x_\lambda) \right] \\
&\leq \int_t^T \left(\frac{1}{2}|\alpha(s)|^2 + f(x_\lambda(s), s) \right) ds + g(x_\lambda(T) + C(T+1)\lambda(1-\lambda)|x-y|^2 \\
&= u(x_\lambda, t) + \|\alpha\|_\infty (T+1)\lambda(1-\lambda)|x-y|^2.
\end{aligned}$$

Hence u is semi-concave. ☐

Lemma 3.10 (Euler-Lagrange optimality condition)**.** *If $\alpha \in \mathcal{A}(x,t)$, then α is of class $C^1([t,T])$ with*

$$\begin{cases} \alpha'(s) = Df(x(s), s) & \text{in } [t,T] \\ \alpha(T) = -Dg(x(T)) \end{cases}$$

In particular, there is a constant $C_1 = C_1(C)$ such that for $(x,t) \in \mathbb{R}^d \times [0,T)$ and any $\alpha \in \mathcal{A}(x,t)$ we have $\|\alpha\|_\infty \leq C_1$, where C satisfies (11).

Lemma 3.11 (Regularity of u along optimal solution)**.** *Let $(x,t) \in \mathbb{R}^d \times [0,T]$, $\alpha \in \mathcal{A}(x,t)$ and let us set $x(s) = x + \int_s^t \alpha(\tau)d\tau$. Then*

1. *(Uniqueness of the optimal control along optimal trajectories) for any $s \in (t,T]$, the restriction of α to $[s,T]$ is the unique element of $\mathcal{A}(x(s), s)$.*

2. *(Uniqueness of the optimal trajectories) $D_x u(x,t)$ exists if and only if $\mathcal{A}(x,t)$ is a reduced to singleton. In this case, $D_x u(x,t) = -\alpha(t)$ where $\mathcal{A}(x,t) = \{\alpha\}$.*

Remark 3.12. *In particular, if we combine the above statements, we see that $u(\cdot, s)$ is always differentiable at $x(s)$ for $s \in (t,T)$ with $D_x u(x(s), s) = -\alpha(s)$.*

Proof. Let $\alpha_1 \in \mathcal{A}(x(s), s)$ and let $x_1(\cdot)$ be its associated trajectory. For any $h > 0$ sufficiently small we define $\alpha_h \in L^2([t,T], \mathbb{R}^d)$ in the following way

$$\alpha_h(\tau) = \begin{cases} \alpha(\tau) & \text{if } \tau \in [t, s-h) \\ \frac{x_1(s+h) - x(s-h)}{2h} & \text{if } \tau \in [s-h, s+h) \\ \alpha_1(\tau) & \text{if } \tau \in [s+h, T] \end{cases}$$

Then one easily checks that

$$x_h(\tau) = \begin{cases} x(\tau) & \text{if } \tau \in [t, s-h) \\ x(s-h) + (\tau - (s-h))\frac{x_1(s+h)-x(s-h)}{2h} & \text{if } \tau \in [s-h, s+h) \\ x_1(\tau) & \text{if } \tau \in [s+h, T] \end{cases}$$

Since both $\alpha|_{[s,T]}$ and α_1 are optimal for $u(x(s), s)$, α_0, which is nothing but the concatenation of $\alpha|_{[t,s]}$ and α_1, is also optimal for $u(x,t)$. Also observe that $x_0(\tau) = x + \int_t^\tau \alpha_0(\sigma)d\sigma$ is given by $x(\tau)$ on $[t, s]$ and $x_1(\tau)$ on $[s, T]$. Hence

$$u(x,t) = \int_t^s \frac{1}{2}|\alpha(\tau)|^2 + f(x(\tau),\tau)d\tau + \int_s^T \left(\frac{1}{2}|\alpha_1(\tau)|^2 + f(x_1(\tau),\tau)\right)d\tau + g(x_1(T))$$

and

$$u(x,t) \le \int_s^T \left(\frac{1}{2}|\alpha_h(\tau)|^2 + f(x_h(\tau),\tau)\right)d\tau + g(x_h(T)).$$

Using the definitions of α_h and x_h we can write the above inequality as

$$\int_{s-h}^s \left(\frac{1}{2}|\alpha(\tau)|^2 + f(x(\tau),\tau)\right)d\tau + \int_s^{s+h} \left(\frac{1}{2}|\alpha_1(\tau)|^2 + f(x_1(\tau),\tau)\right)d\tau$$
$$- \int_{s-h}^{s+h} \left(\frac{1}{2}\left|\frac{x_1(s+h)-x(s-h)}{2h}\right|^2 + f(x_h(\tau),\tau)\right)d\tau \le 0$$

Now dividing h and taking $h \to 0^+$ shows that

$$\frac{1}{2}|\alpha(s)|^2 + \frac{1}{2}|\alpha_1(s)|^2 - \frac{1}{4}|\alpha(s) + \alpha_1(s)|^2 \le 0$$

since $\lim_{h\to 0} x_h(s) = x(s) = x_1(s)$. Therefore $|\alpha(s) - \alpha_1(s)|^2 \le 0$, i.e., $\alpha(s) = \alpha_1(s)$. In particular $x(\cdot)$ and $x_1(\cdot)$ satisfy the same second order differential equation

$$\begin{cases} y''(\tau) = D_x f(y(\tau),\tau) \\ y'(s) = x'(s) = \alpha(s) = \alpha_1(s) = x_1'(s) \\ y(s) = x(s) = x_1(s) \end{cases}$$

Hence $x(\cdot) = x_1(\cdot)$ and $\alpha = \alpha_1$ on $[s,T]$. This means that the optimal solution for $u(x(s), s)$ is unique, thus proving point 1.

We now show that if $D_x u(x,t)$ exists, then $\mathcal{A}(x,t)$ is reduced to singleton and $D_x u(x,t) = -\alpha(t)$ where $\mathcal{A}(x,t) = \{\alpha\}$. Indeed, let $\alpha \in \mathcal{A}(x,t)$ and $x(\cdot)$ be the associated trajectory. Then for any $v \in \mathbb{R}^d$

$$u(x+v,t) \le \int_t^T \frac{1}{2}|\alpha(s)|^2 ds + f(x(s)+v,s)ds + g(x(T)+v).$$

Since equality holds for $v = 0$ and since both sides of the inequality are differentiable with respect to v at $v = 0$ we get

$$D_x u(x,t) = \int_t^T D_x f(x(s),s)ds + D_x g(x(T)).$$

Then by Lemma 3.10 we have $D_x u(x,t) = -\alpha(t)$. Therefore $x(\cdot)$ has to be the unique solution of the second order differential equation

$$\begin{cases} x''(s) = D_x f(x(s), s) \\ x'(t) = -D_x u(x,t) \\ x(t) = x \end{cases}$$

which in turn implies that $\alpha = x'$ is unique.

Conversely, suppose that $\mathcal{A}(x,t)$ is a singleton. We want to show that $u(\cdot,t)$ is differentiable at x. For this we note that, if p belongs to $D_x^* u(x,t)$ (the set of reachable gradients of the map $u(\cdot,t)$), then the solution

$$\begin{cases} x''(s) = D_x f(x(s), s) \\ x'(t) = -p \\ x(t) = x \end{cases}$$

is optimal. Indeed, by definition of p there is a sequence $x_n \to x$ such that $u(\cdot,t)$ is differentiable at x_n and $D_x u(x_n,t) \to p$. Now since $u(\cdot,t)$ is differentiable at x_n, we know that the unique solution $x_n(\cdot)$ of

$$\begin{cases} x_n''(s) = D_x f(x_n(s), s) \\ x_n(t) = x \\ x_n'(t) = -Du(x_n,t) \end{cases}$$

is optimal. Passing to the limit as $n \to \infty$ implies by Lemma 3.8 that $x(\cdot)$, which is the uniform limit of the $x_n(\cdot)$, is also optimal. But from our assumptions, there is a unique optimal solution in $\mathcal{A}(x,t)$. Hence $D_x^* u(x,t)$ has to be reduced to a singleton and since $u(\cdot,t)$ is semi-concave by Lemma 3.9, we have that $u(\cdot,t)$ is differentiable at x by Lemma 3.5. □

Let us consider again $(x,t) \in \mathbb{R}^d \times [0,T)$, $\alpha \in \mathcal{A}(x,t)$ and $x(\cdot)$. Then we have just proved that $u(\cdot,s)$ is differentiable at $x(s)$ for any $s \in (t,T)$ with

$$x'(s) = \alpha(s) = -D_x u(x(s), s).$$

So given α optimal, its associated trajectory $x(\cdot)$ is a solution of the differential equation

$$\begin{cases} x'(s) = -D_x u(x(s), s) \quad \text{on } [t,T] \\ x(t) = x \end{cases}$$

The following Lemma, states that the reverse also holds. This is an optimal synthesis result since it says the optimal control can be obtained at each position y and at each time s as by the synthesis $\alpha^*(y,s) = -D_x u(y,s)$.

Lemma 3.13 (Optimal synthesis)**.** *Let $(x,t) \in \mathbb{R}^d \times [0,T)$ and $x(\cdot)$ be an absolutely continuous solution to the differential equation*

$$\begin{cases} x'(s) = -D_x u(x(s),s), & a.e.\ in\ [t,T] \\ x(t) = x \end{cases} \tag{12}$$

Then the control $\alpha := x'$ is optimal for $u(x,t)$, i.e., $\alpha \in \mathcal{A}(x,t)$. In particular, if $u(\cdot,t)$ is differentiable at x, then equation (12) has a unique solution, corresponding to the optimal trajectory.

Proof. We start by observing that $x(\cdot)$ is Lipschitz continuous because u is. Let $s \in (t,T)$ be such that equation (12) holds. Hence u is differentiable with respect to x at $(x(s),s)$ and the Lipschitz continuous map $s \mapsto u(x(s),s)$ has a derivative at s. Since u is Lipschitz continuous, Lebourg's mean value theorem ([4], Th. 2.3.7), states that, for any $h > 0$ small enough there is some $(y_h, s_h) \in [(x(s),s),(x(s+h),s+h)]$ and some $(\xi_x^h, \xi_t^h) \in CoD_x^* u(y_h,s_h)$ with

$$u(x(s+h),s+h) - u(x(s),s) = \xi_x^h \cdot (x(s+h)-x(s)) + \xi_t^h h, \tag{13}$$

where $CoD_{x,t}^* u(y,s)$ denotes the closure of the convex hull of the set of reachable gradients $D_{x,t}^* u(y,s)$. Now from Carathéodory Theorem, there are $(\lambda^{h,i}, \xi_x^{h,i}, \xi_t^{h,i})_{i=1,\dots,d+2}$ such that

$$\lambda^{h,i} \geq 0,\ \sum_{i=1}^{d+2} \lambda^{h,i} = 1,\ (\xi_x^{h,i}, \xi_t^{h,i}) \in D_x^* u(y_h,s_h) \text{ and } (\xi_x^h, \xi_t^h) = \sum_{i=1}^{d+2} \lambda^{h,i} (\xi_x^{h,i}, \xi_t^{h,i}).$$

For each $i = 1, \dots, d+2$, the $\xi_x^{h,i}$ converges to $D_x u(x(s),s)$ as $h \to 0$ because, from Lemma 3.6, any accumulation point of $(\xi_x^{h,i})_h$ must belong to $D_x^+ u(x(s),s)$ which is reduced $D_x u(x(s),s)$ since $u(\cdot,s)$ is differentiable at $x(s)$. Therefore

$$\xi_{x,h} = \sum_i \lambda^{h,i} \xi_x^{h,i} \to D_x u(x(s),s)$$

as $h \to 0$. Since u is a viscosity solution to (10) and $(\xi_x^{h,i}, \xi_t^{h,i}) \in D_{x,t}^* u(x(s),s)$ we have

$$-\xi_t^{h,i} + \frac{1}{2} |\xi_x^{h,i}|^2 = f(y_h, s_h).$$

Therefore

$$\xi_t^h = \sum_{i=1}^{d+2} \lambda^{h,i} \xi_t^{h,i} = \frac{1}{2} \sum_{i=1}^{d+2} \lambda^{h,i} |\xi_x^{h,i}|^2 - f(y_h,s_h) \to \frac{1}{2} |D_x u(x(s),s)|^2 - f(x(s),s)$$

as $h \to 0$. Then dividing (13) by h and letting $h \to 0$ we get

$$\frac{d}{ds} u(x(s),s) = D_x u(x(s),s) \cdot x'(s) + \frac{1}{2} |D_x u(x(s),s)|^2 - f(x(s),s).$$

and, since $x'(s) = -D_x u(x(s),s)$, we have

$$\frac{d}{ds} u(x(s),s) = -\frac{1}{2} |x'(s)|^2 - f(x(s),s) \quad \text{a.e. in } (t,T).$$

Integrating the above inequality over $[t,T]$ we finally obtain

$$u(x,t) = \int_t^T \frac{1}{2} |x'(s)|^2 + f(x(s),s) ds + g(x(T))$$

where we used the fact that $u(y,T) = g(y)$ for $y \in \mathbb{R}^d$. Therefore $\alpha := x'$ is optimal.

The last statement of the Lemma is a just direct consequence of point 2. of Lemma 3.11. ☐

From the stability of optimal solutions, the graph map $(x,t) \mapsto \mathcal{A}(x,t)$ is closed when the set $L^2([0,T],\mathbb{R}^d)$ is endowed with the weak topology. This implies that the map $(x,t) \mapsto \mathcal{A}(x,t)$ is measurable with nonempty closed values, so that it has a Borel measurable selection $\bar{\alpha}$: namely $\bar{\alpha}(x,t) \in \mathcal{A}(x,t)$ for any (x,t) (see [1]).

Fix $(x,t) \in \mathbb{R}^d \times (0,T)$. We define the flow

$$\Phi(x,t,s) = x + \int_t^s \bar{\alpha}(x,t)(\tau)d\tau$$

for all $s \in [t,T]$. We will use it in the next Section to construct a solution to the Fokker-Planck equation.

Lemma 3.14. *The flow Φ has the semi-group property*

$$\Phi(x,t,s') = \Phi(\Phi(x,t,s),s,s') \tag{14}$$

for all $t \le s \le s' \le T$. Moreover for any $x \in \mathbb{R}^d$ and $s,s' \in (t,T)$

$$\partial_s \Phi(x,t,s) = -D_x u(\Phi(x,t,s),s)$$

and

$$|\Phi(x,t,s') - \Phi(x,t,s)| \le \|D_x u\|_\infty |s'-s|.$$

Proof. For any $s \in (t,T)$ we know from Lemma 3.11 that $\mathcal{A}(\Phi(x,t,s),s) = \{\bar{\alpha}(x,t)|_{[s,T]}\}$ and so (14) holds. Moreover, Lemma 3.11 also states that $u(\cdot,s)$ is differentiable at $\Phi(x,t,s)$ with $D_x u(\Phi(x,t,s),s) = -\bar{\alpha}(x,t)(s)$. But by definition $\partial_s \Phi(x,t,s) = \bar{\alpha}(x,t)(s)$ and so $\partial_s \Phi(x,t,s) = -D_x u(\Phi(x,t,s),s)$. Finally this last equality also implies the $\|D_x u\|_\infty$-Lipschitz continuity of $\Phi(x,t,\cdot)$ on (t,T). □

We finish this Section with the following contraction property of the flow Φ.

Lemma 3.15. *If C satisfies* (11)*, then there is some constant $C_2 = C_2(C)$ such that, if u is a solution of* (10)*, then for all $0 \le t \le s \le T$ and $x,y \in \mathbb{R}^d$*

$$|x-y| \le C_2 |\Phi(x,t,s) - \Phi(y,t,s)|.$$

In particular, the map $x \mapsto \Phi(x,t,s)$ has a Lipshitz continuous inverse on the set $\Phi(\mathbb{R}^d,t,s)$.

Proof. Let u be the solution of (10). Then by Lemma 3.9 $D^2_{xx}u \le C_1 I_d$ on $\mathbb{R}^d \times (0,T)$ in the sense of distributions. Let $x(\tau) = \Phi(x,t,s-\tau)$ and $y(\tau) = \Phi(y,t,s-\tau)$ for $\tau \in [0,s-t]$. Then from Lemma 3.14, $x(\cdot)$ and $y(\cdot)$ satisfy respectively

$$\begin{cases} x'(\tau) = D_x u(x(\tau, s-\tau) & \tau \in [0,s-t) \\ x(0) = \Phi(x,t,s) \end{cases} \quad \text{and} \quad \begin{cases} y'(\tau) = D_x u(y(\tau, s-\tau) & \tau \in [0,s-t) \\ y(0) = \Phi(y,t,s) \end{cases} \tag{15}$$

We observe that for almost all $\tau \in [0,s-t]$ we have

$$\frac{d}{d\tau}\left(\frac{1}{2}|(x-y)(\tau)|^2\right) = ((x'-y')(\tau)) \cdot ((x-y)(\tau)) \le C_1 |(x-y)(\tau)|^2$$

where the last inequality comes from (15) and the fact that $D^2_{xx}u \le C_1 I_d$ (see Definition 3.4). Hence by Grownwall's inequality

$$|(x-y)(\tau)| \le e^{C_1/2\tau}|x(0)-y(0)|$$

for all $\tau \in [0, s-t]$. In particular for $\tau = s-t$ we get

$$|x-y| \le e^{C_1/2\tau T}|\Phi(x,t,s)-\Phi(y,t,s)|$$

thus proving the claim. □

3.2 On the continuity equation

Our aim is now to show that, given a solution (10) and under assumption (11), the continuity equation

$$\begin{cases} \partial_t \mu(x,s) - div(D_x u(x,s)\mu(x,s)) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ \mu(x,0) = m_0(x) & \text{in } \mathbb{R}^d \end{cases} \tag{16}$$

has a unique solution which is the density of the measure $\mu(s) = \Phi(\cdot,0,s)_* m_0$ for $s \in [0,T]$, where $\Phi(\cdot,0,s)_* m_0$ denotes the push-forward of the measure m_0 by the map $\Phi(\cdot,0,s)$, i.e., the measure defined by $\Phi(\cdot,0,s)_* m_0(A) = m_0(\Phi(\cdot,0,s)^{-1}(A))$ for any Borel set $A \subseteq \mathbb{R}^d$.

We start by observing that the measure $\Phi(\cdot,0,s)_* m_0$ is absolutely continuous with respect to the Lebesgue measure.

Lemma 3.16. *Let C be a constant such that* (11) *holds and such that m_0 is absolutely continuous, has a compact support contained in the ball $B(0,C)$ and satisfies $\|m_0\|_{L^\infty} \le C$. Let us set $\mu(s) := \Phi(\cdot,0,s)_* m_0$ for $s \in [0,T]$.*

Then there is a constant $C_3 = C_3(C)$ such that, for any $s \in [0,T]$, $\mu(s)$ is absolutely continuous, has a compact support contained in the ball $B(0,C_3)$ and satisfies $\|\mu(s)\|_{L^\infty} \le C_3$. Moreover

$$\boldsymbol{d}(\mu(s'),\mu(s)) \le \|D_x u\|_\infty |s'-s|$$

for all $t \le s \le s' \le T$.

Proof. By definition μ satisfies

$$\mathbf{d}(\mu(s'),\mu(s)) \le \int_{\mathbb{R}^d} |\Phi(x,0,s') - \Phi(x,0,s)| dm_0(x) \le \|D_x u\|_\infty (s'-s).$$

Recall that Φ is given by

$$\Phi(x,0,s) = x + \int_0^s \bar{\alpha}(x,0)(\tau)d\tau$$

where $\bar{\alpha}(x,0)(\tau) = D_x u(\Phi(x,0,\tau),\tau)$. Also since u is a solution of (10), $\|D_x u\|_\infty \le C_1$. Additionally, m_0 has compact support contained in $B(0,C)$. Hence the $(\mu(s))$ have a compact support contained in $B(0,R)$ where $R = C + TC_1$.

Let us now fix $t \in [0,T]$. From Lemma 3.15, we know that there is some $C_2 = C_2(T)$ such that the map $x \mapsto \Phi(x,0,t)$ has a C_2-Lipschitz continuous inverse on the set $\Phi(\mathbb{R}^d,0,t)$. Let us denote this inverse by Ψ. Then, if A is a Borel subset of $\mathbb{R}^d$ we have

$$\mu(s)(A) = m_0(\Phi^{-1}(\cdot,0,t)(A)) = m_0(\Psi(A)) \le \|m_0\|_\infty \mathcal{L}^d(\Psi(A)) \le \|m_0\|_\infty C_2 \mathcal{L}^d(A).$$

Therefore $\mu(s)$ is absolutely continuous with a density (still denoted by $\mu(s)$) which satisfies

$$\|\mu(s)\|_\infty \le \|m_0\|_\infty C_2$$

for all $s \in [0,T]$. □

Our goal is to show that the map $s \mapsto \mu(s) := \Phi(\cdot,0,s)_* m_0$ is the unique weak solution of (16). We first prove that μ is a weak solution of (16).

Lemma 3.17. *The map $s \mapsto \mu(s) := \Phi(\cdot,0,s)_* m_0$ is the weak solution of* (16).

Proof. Let $\varphi \in \mathcal{D}(\mathbb{R}^d \times [0,T))$. Then, since by Lemma 3.16 $s \mapsto \mu(s)$ is Lipschitz continuous in $\mathcal{P}$, the map

$$s \mapsto \int_{\mathbb{R}^d} \varphi(x,s)\mu(x,s)dx$$

is absolutely continuous. Then using Lemma 3.14 we have

$$\begin{aligned}
\frac{d}{ds}\int_{\mathbb{R}^d} \varphi(x,s)\mu(x,s)dx &= \frac{d}{ds}\int_{\mathbb{R}^d} \varphi(\Phi(x,0,s),s)m_0(x)dx \\
&= \int_{\mathbb{R}^d} \left(\partial_s\varphi(\Phi(x,0,s),s) + D_x\varphi(\Phi(x,0,s),t)\cdot\partial_s\Phi(x,0,s)\right)m_0(x)dx \\
&= \int_{\mathbb{R}^d} \left(\partial_s\varphi(\Phi(x,0,s),s) - D_x\varphi(\Phi(x,0,s),t)\cdot D_x u(\Phi(x,0,s),s)\right)m_0(x)dx \\
&= \int_{\mathbb{R}^d} \left(\partial_s\varphi(y,s) - D_x\varphi(y,s)\cdot D_x u(y,s)\right)\mu(y,s)dy
\end{aligned}$$

Integrating the above inequality over $[0,T]$ we get, since $\mu(0) = m_0$

$$\int_{\mathbb{R}^d} \varphi(y,0)m_0(y)dy + \int_0^T\int_{\mathbb{R}^d} \left(\partial_s\varphi(y,s) - D_x\varphi(y,s)\cdot D_x u(y,s)\right)\mu(y,s)dy = 0$$

which means that m is a weak solution of (16). □

We now focus on proving the uniqueness property. The difficulty arises since $-D_x u(x,t)$ may be discontinuous. In fact if $-D_x u(x,t)$ had some Lipschitz regularity property, then the uniqueness would follow easily as we show in the next Lemma.

Lemma 3.18. *Let $b \in L^\infty(\mathbb{R}^d \times (0,T), \mathbb{R}^d)$ be such that, for any $R > 0$ and for almost all $t \in [0,T]$, there is a constant $L = L(R)$ such that $b(\cdot,t)$ is L-Lipschitz continuous on $B(0,R)$. Then the continuity equation*

$$\begin{cases} \partial_t\mu(x,s) + div(b(x,s)\mu(x,s)) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ \mu(x,0) = m_0(x) & \text{in } \mathbb{R}^d \end{cases} \tag{17}$$

has a unique solution, given by $\mu(t) = \Phi(\cdot,t)_* m_0$ where Φ is the flow of the differential equation

$$\begin{cases} \partial_s\Phi(x,s) = b(\Phi(x,s),s) \\ \Phi(x,0) = x \end{cases}$$

Proof. It's easy to see by mimicking the proof of Lemma 3.17 that the map $t \mapsto \Phi(\cdot, t)_* m_0$ is a solution to (17).

We know that that the map $x \mapsto \Phi(x, t)$ is locally Lipschitz continuous, with a locally Lipschitz continuous inverse denoted by $\Psi(x, t)$. Note also that Ψ is actually locally Lipschitz continuous in space-time. Let $\varphi \in \mathcal{D}(\mathbb{R}^d)$ and let us consider the map w defined by $w(x, t) = \varphi(\Psi(x, t))$. Then w is Lipschitz continuous with compact support and satisfies

$$0 = \frac{d}{dt}\varphi(x) = \frac{d}{dt}w(\Phi(x,t),t) = \partial_t w(\Phi(x,t),t) + D_x w(\Phi(x,t),t) \cdot b(\Phi(x,t),t) \quad \text{a.e.,}$$

and therefore w is a solution to

$$\partial_t w(y,t) + D_x w(y,t) \cdot b(y,t) = 0 \quad a.e. \text{ in } \mathbb{R}^d \times (0,T).$$

Using now w as a test function for μ we have

$$\frac{d}{dt}\int_{\mathbb{R}^d} w(y,t)\mu(y,t)dy = \int_{\mathbb{R}^d} \left(\partial_t w(y,t) + Dw(y,t) \cdot b(y,t)\right) \mu(y,t)dy = 0$$

and therefore

$$\int_{\mathbb{R}^d} \varphi(\Psi(y,t))\mu(y,t)dy = \int_{\mathbb{R}^d} \varphi(y)m_0(y)dy.$$

Changing the test function shows that

$$\int_{\mathbb{R}^d} \psi(y)\mu(y,t)dy = \int_{\mathbb{R}^d} \psi(\Phi(y,s))m_0(y)dy,$$

for any $\psi \in \mathcal{D}(\mathbb{R}^d)$, thus proving that $\mu(t) = \Phi(\cdot, t)_* m_0$ as desired. □

We now return to equation (16) and prove that it has a unique solution.

Theorem 3.19. *Given a solution u to* (10) *and under assumption* (11)*, the map $s \mapsto \mu(s) := \Phi(\cdot, 0, s)_* m_0$ is the unique weak solution of* (16).

Proof. Due to Lemma 3.17 we only need to show that if μ is a solution of (3.17), then μ is given by $\Phi(\cdot, 0, t)_* m_0$. Let then μ be any solution. The idea is to regularize μ to a get a sequence of solutions to which we can apply Lemma 3.18. Let $\rho_\varepsilon \in \mathcal{D}(B(0, \varepsilon))$ denote the standard mollifier. In particular $\int_{\mathbb{R}^d} \rho_\varepsilon(x)dx = 1$ and $\rho_\varepsilon(x) \geq 0$ in $\mathbb{R}^d$ Define

$$\mu^\varepsilon(x,t) := \mu * \rho_\varepsilon \text{ and } b^\varepsilon(x,t) := -\frac{(D_x u\, \mu) * \rho_\varepsilon(x,t)}{\mu^\varepsilon(x,t)}.$$

Then $\|b^\varepsilon\|_\infty \leq \|D_x u\|_\infty$ and b^ε is locally Lipschitz continuous in the sense of Lemma 3.18. Moreover μ^ε satisfies the continuity equation for b^ε because

$$\partial_t \mu^\varepsilon + div(b^\varepsilon \mu^\varepsilon) = (\partial_t \mu) * \rho_\varepsilon - div((D_x u \mu) * \rho_\varepsilon) = [\partial_t \mu - div(D_x u \mu)] * \rho_\varepsilon = 0.$$

Then by Lemma 3.18, $\mu^\varepsilon(t) = \Phi^\varepsilon(\cdot, t)_* m_\varepsilon$, where $m_\varepsilon = m_0 * \rho_\varepsilon$ and Φ^ε is the flow associated to b^ε:

$$\begin{cases} \partial_s \Phi^\varepsilon(x,s) = b^\varepsilon(\Phi(x,s),s) \\ \Phi^\varepsilon(x,0) = x \end{cases}$$

The difficulty now boils down to passing the limit in the equality $\mu^\varepsilon(t) = \Phi^\varepsilon(\cdot,t)_* m_\varepsilon$.

Let us set, to simplify notations, $\Gamma_T := C([0,T],\mathbb{R}^d)$ and associate with μ^ε the measure η^ε on $\mathbb{R}^d \times \Gamma_T$ defined by

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(x,\gamma)d\eta^\varepsilon(x,\gamma) = \int_{\mathbb{R}^d} \varphi(x,\Phi^\varepsilon(x,\cdot))m_\varepsilon(x)dx$$

for all $\varphi \in C(\mathbb{R}^d \times \Gamma_T)$. Also for $t \in [0,T]$ we denote by e_t the evaluation map at t, i.e., $e_t(\gamma) = \gamma(t)$ for $\gamma \in \Gamma_T$. Then for any $\varphi \in C^0_b(\mathbb{R}^d,\mathbb{R})$ we have

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_t(\gamma))d\eta^\varepsilon(x,\gamma) = \int_{\mathbb{R}^d} \varphi(\Phi^\varepsilon(x,t))m^\varepsilon(x)dx = \int_{\mathbb{R}^d} \varphi(x)\mu^\varepsilon(x,t)dx. \tag{18}$$

Let us now prove that (η^ε) is tight in $\mathbb{R}^d\times\Gamma_T$. Indeed, since m_ε converges to m_0 as $\varepsilon \to 0$, we can find for any $\delta > 0$ some compact set $K_\delta \subseteq \mathbb{R}^d$ such that $m_\varepsilon(K_\delta) \geq 1-\delta$ for any ε small enough. Let $\mathcal{K}_\delta$ be the subset of $K_\delta \times \Gamma_T$ consisting in pairs (x,γ) where $x \in K_\delta$, $\gamma(0) = x$, γ is Lipschitz continuous with $\|\gamma'\|_\infty \leq \|D_x u\|_\infty$. Then $\mathcal{K}_\delta$ is compact and by definition of η^ε,

$$\eta^\varepsilon(\mathcal{K}_\delta) = m_\varepsilon(K_\delta) \geq 1-\delta$$

for all $t \in [0,T]$. Therefore (η^ε) is tight and from Prokhorov compactness theorem one can find a subsequence, still denoted (η^ε), which converges weakly to some probability measure η on $\mathbb{R}^d \times \Gamma_T$. Then letting $\varepsilon \to 0$ in (18) gives

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(e_t(\gamma))d\eta(x,\gamma) = \int_{\mathbb{R}^d} \varphi(x)\mu(x,t)dx \tag{19}$$

for all $t \in [0,T]$ and for any $\varphi \in C_b(\mathbb{R}^d)$, and therefore for any Borel bounded measurable map $\varphi : \mathbb{R}^d \to \mathbb{R}$. Moreover, since, by definition of η^ε, we have

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(x)d\eta^\varepsilon(x,\gamma) = \int_{\mathbb{R}^d} \varphi(x)m_\varepsilon(x)dx$$

for all $\varphi \in C(\mathbb{R}^d,\mathbb{R})$, we also have that

$$\int_{\mathbb{R}^d\times\Gamma_T} \varphi(x)d\eta(x,\gamma) = \int_{\mathbb{R}^d} \varphi(x)m_0(x)dx \tag{20}$$

for all $\varphi \in C(\mathbb{R}^d)$, i.e., the first marginal of η is m_0. The key step of the proof consists now in showing that η is concentrated on solutions of the differential equation for $-D_x u$. More precisely, we want to show that for any $t \in [0,T]$

$$\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x + \int_0^t D_x u(\gamma(s),s)ds\right| d\eta(x,\gamma) = 0. \tag{21}$$

For this we have to regularize a bit the vector field $-D_x u$. Let $c : \mathbb{R}^d \times [0,T] \to \mathbb{R}^d$ be a continuous vector field with compact support. We claim

$$\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x - \int_0^t c(\gamma(s),s)ds\right| d\eta(x,\gamma) \leq \int_0^T \int_{\mathbb{R}^d} |c(x,t) + D_x u(x,t)|\mu(x,t)dxdt. \tag{22}$$

Indeed, we have for any $\varepsilon > 0$ small

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\Gamma_T} \left|\gamma(t) - x - \int_0^t c(\gamma(s), s)ds\right| d\eta^\varepsilon(x,\gamma) \\
&= \int_{\mathbb{R}^d} \left|\Phi^\varepsilon(x,t) - x - \int_0^t c(\Phi^\varepsilon(x,s), s)ds\right| m_\varepsilon(x)dx \\
&= \int_{\mathbb{R}^d} \left|\int_0^t (b^\varepsilon(\Phi^\varepsilon(x,t), s) - c(\Phi^\varepsilon(x,s), s)ds\right| m_\varepsilon(x)dx \\
&\leq \int_0^t \int_{\mathbb{R}^d} |b^\varepsilon(\Phi^\varepsilon(x,t), s) - c(\Phi^\varepsilon(x,s), s)| m_\varepsilon(x)dxds \\
&= \int_0^t \int_{\mathbb{R}^d} |b^\varepsilon(x,s) - c(x,s)|\mu^\varepsilon(x,t)dxds
\end{aligned}$$

where, setting $c^\varepsilon = \frac{(c\mu)*\rho^\varepsilon}{\mu^\varepsilon}$, we have

$$\begin{aligned}
&\int_0^t \int_{\mathbb{R}^d} |b^\varepsilon(x,s) - c(x,s)|\mu^\varepsilon(x,t)dxds \\
&\leq \int_0^t \int_{\mathbb{R}^d} |b^\varepsilon(x,s) - c^\varepsilon(x,s)|\mu^\varepsilon(x,t)dxds + \int_0^t \int_{\mathbb{R}^d} |c^\varepsilon(x,s) - c(x,s)|\mu^\varepsilon(x,t)dxds \\
&\leq \int_0^t \int_{\mathbb{R}^d} |(D_x u + c)\mu * \rho_\varepsilon(x,t)|dxds + \int_0^t \int_{\mathbb{R}^d} |c^\varepsilon(x,s) - c(x,s)|\mu^\varepsilon(x,t)dxds
\end{aligned}$$

Note that the integrand of the first integral converges to $|c(x,t) + D_x u(x,t)|\mu(x,t)$ as $\varepsilon \to 0$ and the last term converges to 0 as $\varepsilon \to 0$ due to the continuity of c. This gives (22). As for (21) it's now enough to take a sequence of uniformly bounded continuous maps c_n with compact support which converges a.e. to $-D_x u$. Replacing then c by c_n in (22) gives the desired result since, from (19),

$$\int_0^t \int_{\mathbb{R}^d\times\Gamma_T} |D_x u(\gamma(s), s) + c_n(\gamma(s), s)| d\eta(x,\gamma)ds = \int_0^t \int_{\mathbb{R}^d} |D_x u(x,s) + c_n(x,s)|\mu(x,s)ds.$$

Let us now desintegrate η with respect to its first marginal, which according to (20), is m_0 (see the desintegration theorem in Appendix B). We get $d\eta(x,\gamma) = d\eta_x(\gamma)dm_0(x)$. Then (21) implies that, for m_0-a.e. $x \in \mathbb{R}^d$, η_x-a.e. γ is a solution of the differential equation

$$\begin{cases} \gamma'(s) = -D_x u(\gamma(s), s) & s \in [t, T] \\ \gamma(t) = x \end{cases}$$

But for almost all $x \in \mathbb{R}^d$, $u(\cdot, 0)$ is differentiable at x and Lemma 3.14 then says that the above differential equation has a unique solution given by $\Phi(x, 0, \cdot)$. Since m_0 is absolutely continuous, this implies that, for m_0-a.e. $x \in \mathbb{R}^d$, η_x-a.e. γ given by $\Phi(x, 0, \cdot)$. Then equality (19) becomes

$$\int_{\mathbb{R}^d} \varphi(x)\mu(x,t)dx = \int_{\mathbb{R}^d} \int_{\Gamma_T} \varphi(e_t(\gamma))m_0(x)d\eta_x(\gamma)dx = \int_{\mathbb{R}^d} \int_{\Gamma_T} \varphi(\Phi(x,0,t))m_0(x)dx$$

for any $\varphi \in C_b(\mathbb{R}^d)$ and $t \in [0, T]$. This proves that $\mu(t)$ is given by $\Phi(\cdot, 0, t)_* m_0$ as desired. □

3.3 Existence of solutions to a 1^{st} order MFG

Before starting the proof of Theorem 3.2, we need to show that the system (9) is stable. Let (m_n) be a sequence of $C([0,T],\mathcal{P})$ which uniformly converges to $m \in C([0,T],\mathcal{P})$. Let u_n be the solution to

$$\begin{cases} -\partial_t u_n + \frac{1}{2}|Du_n|^2 = F(x, m_n(t)) & \text{in } \mathbb{R}^d \times (0,T) \\ u_n(x,T) = G(x, m_n(T)) & \text{in } \mathbb{R}^d \end{cases}$$

and u be the solution to

$$\begin{cases} -\partial_t u + \frac{1}{2}|Du|^2 = F(x, m(t) & \text{in } \mathbb{R}^d \times (0,T) \\ u(x,T) = G(x, m(T)) & \text{in } \mathbb{R}^d \end{cases}$$

Let us denote by Φ_n (respectively Φ) the flow associated to u_n (respectively u) as above and let us set $\mu_n(s) = \Phi_n(\cdot, 0, s)_* m_0$ and $\mu(s) = \Phi(\cdot, 0, s)_* m_0$.

Lemma 3.20 (Stability)**.** *The solution (u_n) locally uniformly converges to u in $\mathbb{R}^d \times [0,T]$ and (μ_n) converges to μ in $C([0,T],\mathcal{P})$.*

Proof. From our assumptions on F and G, the sequences of maps $(x,t) \mapsto F(x, m_n(t))$ and $x \mapsto G(x, m_n(T))$ locally uniform converge to the maps $(x,t) \mapsto F(x, m(t))$ and $x \mapsto G(x, m(T))$ respectively. Hence the local uniform convergence of (u_n) to u is just a consequence of the standard stability of viscosity solutions.

From Lemma 3.9 there is a constant C_1 such that $D^2_{xx} u_n \leq C_1 I_d$ for all n. hence the local uniform convergence of (u_n) to u implies by Lemma 3.6 that $D_x u_n$ converges almost everywhere in $\mathbb{R}^d \times (0,T)$ to $D_x u$. From Lemma 3.16, we know that the (μ_n) are absolutely continuous with support contained in $K := B(0, C_3)$ and $\|\mu_n\|_\infty \leq C_3$. Moreover Lemma 3.16 also states that

$$\mathbf{d}(\mu_n(s'), \mu_n(s)) \leq C_1 |s' - s|$$

for all $t \leq s \leq s' \leq T$. Since $\mathcal{P}(K)$, the set of probability measures on K is compact, the Arzela-Ascoli theorem states that the sequence (μ_n) is precompact in $C([0,T],\mathcal{P}(K))$. Therefore a subsequence (still denoted (μ_n)) of the (μ_n) converges in $C([0,T],\mathcal{P}(K))$ and in $L^\infty - weak - *$ to some m which has a support in $K \times [0,T]$, belongs to $L^\infty(\mathbb{R}^d \times [0,T])$ and to $C([0,T],\mathcal{P}(K))$. Since the (μ_n) solve the continuity equation for (u_n), one easily gets by passing to the limit that m satisfies the continuity equation for u. By uniqueness this implies that $m = \mu$ and the the proof is complete. □

Finally we prove Theorem 3.2

Proof of Theorem 3.2. Let $\mathcal{M}$ be the closed convex subset of maps $m \in C([0,T],\mathcal{P})$ such that $m(0) = m_0$. To any $m \in \mathcal{M}$ one associates the unique solution u to

$$\begin{cases} -\partial_t u + \frac{1}{2}|D_x u|^2 = F(x, m(t)) & \text{in } \mathbb{R}^d \times (0,T) \\ u(x,T) = G(x, m(T)) & \text{in } \mathbb{R}^d \end{cases}$$

and to this solution one associates the unique solution of the continuity equation

$$\begin{cases} \partial_t \mu - div(Du(x,s)\mu(x,s)) = 0 \quad \text{in } \mathbb{R}^d \times (0,T) \\ \mu(0) = m_0 \quad \text{in } \mathbb{R}^d \end{cases}$$

Then $\mu \in \mathcal{M}$ and, from Lemma 3.20, the mapping $m \mapsto \mu$ is continuous. From Lemma 3.16 there is a constant C_3 independent of m, such that, for any $s \in [0,T]$, $\mu(s)$ has a support in $B(0,C_3)$ and satisfies

$$\mathbf{d}(\mu(s'), \mu(s)) \leq C|s' - s| \quad \forall s, s' \in [0,T].$$

This implies that the mapping $m \mapsto \mu$ is compact because $s \mapsto \mu(s)$ is uniformly Lipschitz continuous with values in the compact set of $\mathcal{P}(B(0,C_3))$. As in the second order MFG, we now appeal to the Schauder fixed point theorem to complete the proof. □

A Stochastic Calculus

In this Appendix we review some of the basic definitions of stochastic calculus and related results used along this report.

A.1 Brownian Motion and filtration

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure.

Definition A.1. *Let $W = \{W_t : t \in \mathbb{R}_+\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. W is a Brownian motion if*

- *$W_0 = 0$ and the sample paths $t \mapsto W_t(\omega)$ are continuous for a.e. $\omega \in \Omega$;*
- *W has independent increments, i.e., $W_{t_4} - W_{t_3} \perp\!\!\!\perp W_{t_2} - W_{t_1}$ where $0 \leq t_1 \leq t_2 \leq t_3 \leq t_4$;*
- *W has increments which are normally distributed, i. e., $W_{t_2} - W_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$ where $0 \leq t_1 < t_2$.*

We can extend the definition of the Brownian motion to the vector case.

Definition A.2. *Let $W = \{W_t : t \in \mathbb{R}_+\}$ be an $\mathbb{R}^N$-stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. W is a N-dimensional Brownian motion if the components W^i, $i = 1, \ldots, N$ are independent Brownian motions and the distribution of $W_{t_2} - W_{t_1}$ is $\mathcal{N}(0, (t_2 - t_1)I_N)$ for all $t_2 > t_1 \geq 0$, where I_N denotes the identity matrix in $\mathbb{R}^N$.*

We now introduce the concept of filtration.

Definition A.3. *A filtration is an increasing collection of σ-algebras $\{\mathcal{F}_t\}_{t \geq 0}$, i.e., if $s < t$ then $\mathcal{F}_s \subseteq \mathcal{F}_t$.*

Remark A.4. *The increasing feature of the filtration means that information can only increase as time goes on.*

We consider in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a filtration $\mathbb{F} = \{\mathcal{F}_t : 0 \leq t \leq T\}$. The filtration we consider is the one induced by W augmented with the $\mathbb{P}$-null sets, that is,

$$\mathcal{F}_t = \sigma\{W_s : s \leq t\} \vee \mathcal{N}_{\mathbb{P}}.$$

Definition A.5. *A stochastic process X is $\mathbb{F}$-adapted if for all $t \in \mathbb{R}_+$ X_t is $\mathcal{F}_t$-measurable.*

A.2 Stochastic integral and Itô's formula

The stochastic integral

$$\int_0^t \psi_s dW_s$$

is defined for processes $\psi : \Omega \times [0,T] \to \mathcal{M}_{\mathbb{R}}(d,N)$ in

$$\mathbb{H}^2_{loc} = \left\{ \psi : \mathbb{F} - \text{adapted processes with } \int_0^T |\psi_s|^2 ds < \infty \, a.s. \right\}.$$

However in the smaller space

$$\mathbb{H}^2 = \left\{ \psi : \mathbb{F} - \text{adapted processes with } \mathbb{E}\left[\int_0^T |\psi_s|^2 ds \right] < \infty \right\},$$

we can prove additional results since this one is a Hilbert space when equipped with the norm

$$\|\psi\|_{\mathbb{H}^2} = \sqrt{\mathbb{E}\left[\int_0^T |\psi_s|^2 ds \right]}.$$

Remark A.6. *We recall that W is a N-dimensional Brownian motion. Thus $I := \int_0^T \psi_s dW_s$ is an abbreviation for the vector $(I_i)_{i=1,\ldots,N}$ such that*

$$I_i := \sum_{j=1}^N \int_0^t \psi_s^{i,j} dW_s^j,$$

where ψ is a $(d \times N)$-dimensional process. The norm of ψ is the Fröbenius norm.

Definition A.7. *An Itô process, X, is a continuous-time process defined by*

$$X_t := X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, t \geq 0,$$

where μ, σ are $\mathbb{F}$-adapted processes satisfying $\int_0^t |\mu_s| + |\sigma_s|^2 ds < \infty$. Note that μ and σ take values in $\mathbb{R}^d$ and $\mathcal{M}_{\mathbb{R}}(d,N)$, respectively.

Remark A.8. *Itô's processes are frequently written in differential notation as*

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

Itô's formula can be seen as the chain rule of stochastic calculus. It tells us how stochastic differentials change under composition. Recall that given a smooth function $f(x,t)$ we will denote by $D_x f$ and $D^2_{xx} f$ the partial gradient with respect to x and the partial Hessian with respect to x.

Theorem A.9 (Itô formula). *Let $f \in C^{2,1}(\mathbb{R}^N \times [0,T])$ and X an Itô process given by*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s.$$

Then, with probability 1,

$$f(X_t,t) = f(X_0,0) + \int_0^t \partial_t f(X_s,s) + D_x f(X_s,s)\cdot\mu_s + \frac{1}{2}Tr(\sigma_s\sigma_s^T D^2_{xx} f(X_s,s)ds + \int_0^t D_x f(X_s,s)\cdot\sigma_s dW_s$$

Moreover

$$\mathbb{E} = \left[\int_0^t D_x f(X_s,s)\cdot\sigma_s dW_s\right].$$

A.3 Stochastic differential equations

In this section we give meaning to the stochastic differential equation

$$\begin{cases} dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dW_t \\ X_0 = Z_0 \end{cases} \tag{23}$$

Definition A.10. *A strong solution to* (23) *is an $\mathbb{F}$-adapted process X with continuous samples paths such that*

- $\int_0^T |\mu(X_t,t)| + |\sigma(X_t,t)|^2 dt < \infty, \mathbb{P} - a.s.$
- $X_t = Z_0 + \int_0^t \mu(X_s,s)ds + \int_0^t \sigma(X_s,s)dW_s$ *for all* $t \in [0,T]$.

The next Theorem gives sufficient conditions for the existence and uniqueness of strong solutions for (23).

Theorem A.11. *Let $Z_0 \in \mathbb{L}^2$ be a random variable independent of W. Suppose that the functions $|\mu(0,\cdot)|$, $|\sigma(0,\cdot)| \in L^2(\mathbb{R}_+)$ and that for some $K > 0$*

$$|\mu(x,t) - \mu(y,t)| + |\sigma(x,t) - \sigma(y,t)| \leq K|x-y| \text{ for all } t \in [0,T] \text{ and } x,y \in \mathbb{R}^d.$$

Then, for all $T > 0$, there exists a unique strong solution $X \in \mathbb{H}^2$ to (23).

B Auxiliary results

Definition B.1. *Let (X,T) be a topological space and let Σ be a σ-algebra on X that contains T. Let M be a collection of (possibly signed or complex) measures defined on Σ. The collection M is called tight if, for any $\varepsilon > 0$, there is a compact subset K_ε of X such that, for all measures μ in M,*

$$|\mu|(X \setminus K_\varepsilon) < \varepsilon.$$

where $|\mu|$ is the total variation of μ. Very often, the measures in question are probability measures, so the last part can be written as

$$\mu(K_\varepsilon) > 1 - \varepsilon.$$

Theorem B.2 (Prokhorov theorem)**.** *Let (S, ρ) be a metric space. Let $\mathcal{P}(S)$ denote the collection of all probability measures defined on S (with its Borel σ-algebra). Then*

1. *A collection $K \subseteq \mathcal{P}(S)$ of probability measures is tight if and only if the closure of K is sequentially compact in the space $\mathcal{P}(S)$ equipped with the topology of weak convergence.*

2. *The space $\mathcal{P}(S)$ with the topology of weak convergence is metrizable.*

3. *Suppose that in addition, (S, ρ) is a complete metric. There is a complete metric d_0 on $\mathcal{P}(S)$ equivalent to the weak topology convergence. Moreover $K \subseteq \mathcal{P}(S)$ is tight if and only if the closure of K in $(\mathcal{P}(S), d_0)$ is compact.*

Theorem B.3 (Desintegration of a measure)**.** *Let X and Y be two Polish spaces and λ be a Borel probability measure on $X \times Y$. Let us set $\mu = \pi_{X*}\lambda$, where ϕ_X is the standard projection from $X \times Y$ onto X. Then there exists a μ-almost everywhere uniquely determined family of Borel probability measures (λ_x) on Y such that*

1. *the function $x \mapsto \lambda_x$ is Borel measurable, in the sense that $x \mapsto \lambda_x(B)$ is Borel measurable function for each Borel measurable set $B \subseteq Y$,*

2. *for every Borel measurable function $f : X \times Y \to [0, \infty]$,*

$$\int_{X\times Y} f(x,y)d\lambda(x,y) = \int_X \int_Y f(x,y)d\lambda_x(y)d\mu(x).$$

References

[1] J. P. Aubin and H. Frankowska. Set-valued analysis. Birkhäuser, Boston (1990).

[2] P. Cannarsa and C. Sinestrari. Semiconcave functions, Hamilton-Jacobi equations, and optimal control. Progress in Nonlinear Differential Equations and their Applications, 58. Birkhäuser Boston Inc., Boston, MA, 2004.

[3] Pierre Cardaliaguet. *Notes on Mean Field Games.* January 15, 2012.

[4] F. H. Clarke. Optimization and nonsmooth analysis. Second edition. Classics in Applied Mathematics, 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[5] O. A. Ladyženskaja, V. A. Solonnikov and N.N. Ural'ceva. Linear and quasilinear equations of parabolic type. Translations of Mathematical Monographs, Vol. 23 American Mathematical Society, Providence, R.I. 1967.

[6] J. M. Lasry and P. L. Lions. *Large investor trading impacts on volatility.* Ann. Inst. H. Poincaré Anal. Non Linéaire 24 (2007), no. 2, 311-323.

[7] J. M. Lasry and P. L. Lions. *Mean Field Games.* Jpn. J. Math. 2 (2007), no. 1, 229-260.

[8] J. M. Lasry and P. L. Lions. *Jeux à champ moyen. II. Horizon fini et contrôle optimal.* C. R. Math. Acad. Sci. Paris 343 (2006), no. 10, 679-684.

[9] J. M. Lasry and P. L. Lions. *Jeux à champ moyen. I. Le cas stationnaire*. C. R. Math. Acad. Sci. Paris 343 (2006), no. 9, 619-625.

[10] P. Tankov and N. Touzi. *Calcul Stochastique en Finance (polycopié)*. École Polytechnique, 2010.