Electron. Commun. Probab. **18** (2013), no. 68, 1–15.
DOI: 10.1214/ECP.v18-2446
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Mean field forward-backward stochastic differential equations

René Carmona* Francois Delarue†

Abstract

The purpose of this note is to provide an existence result for the solution of fully coupled Forward Backward Stochastic Differential Equations (FBSDEs) of the mean field type. These equations occur in the study of mean field games and the optimal control of dynamics of the McKean Vlasov type.

Keywords: FBSDEs; Mean Field Interactions.
AMS MSC 2010: Primary: 93E20, 60H30; Secondary: 60H10, 60F99.
Submitted to ECP on November 12, 2012, final version accepted on July 2, 2013.

1 Introduction

Stochastic differential equations of the McKean - Vlasov type are Itô's stochastic differential equations where the coefficients depend upon the marginal distribution of the solution. In their partial differential form, they were introduced by Mark Kac's in his analysis of the Boltzmann equation for the density of particules in kinetic theory of dilute monatomic gases, and a toy model for the Vlasov kinetic equation of plasma (see [11, 12]).

The purpose of this note is to provide an existence result for the solution of Forward Backward Stochastic Differential Equations (FBSDEs) of the McKean-Vlasov type. Following the wave of interest created by the pathbreaking work of Lasry and Lions on mean field games [14, 15, 16], simple forms of Backward Stochastic Differential Equations (BSDEs) of McKean Vlasov type have been introduced and called of *mean field type*. Fully coupled FBSDEs are typically more involved and more difficult to solve than BSDEs. FBSDEs of mean field type occur naturally in the probabilistic analysis of mean field games and the optimal control of dynamics of the McKean Vlasov type as considered in [4, 3]. See also [1, 5, 18] for the particular case of Linear Quadratic (LQ) models. Detailed explanations on how these FBSDEs occur in these contexts and the particular models which were solved are given in Section (3) below.

The existence proofs given in [4] and [3] depend heavily on the fact that the problems at hand are in fact stochastic control problems and FBSDEs are derived from an application of a version of the stochastic maximum principle, and the compactness estimates are derived from the linear nature of the forward dynamics and strong convexity properties of the cost functions of the stochastic optimization problems. The purpose of

*ORFE, Bendheim Center for Finance, Princeton University, Princeton, NJ 08544, USA. Partially supported by NSF: DMS-0806591 E-mail: rcarmona@princeton.edu
†Laboratoire Jean-Alexandre Dieudonné Université de Nice Sophia-Antipolis, Parc Valrose 06108 Cedex 02, Nice, FRANCE. E-mail: Francois.Delarue@unice.fr

this note is to provide a general existence result which does not depend upon strong linearity and convexity assumptions. Such an existence result is proven in Section (2). The proof relies on Schauder's fixed point theorem used in appropriate spaces of functions and measures. A short Section (3) concludes with a short discussion of applications to mean field games and control of McKean-Vlasov dynamics studied in [4] and [3].

2 Solvability of Forward-Backward Systems of McKean-Vlasov Type

2.1 First Notation

All the processes considered in this note are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which an m-dimensional Wiener process $\underline{W} = (W_t)_{0\le t\le T}$ is defined. For each random variable/vector or stochastic process X, we denote by $\mathbb{P}_X$ the law (alternatively called the distribution) of X. We shall denote by $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le T}$ the filtration of $\underline{W}$ and by $\mathbb{H}^{2,n}$ the Hilbert space

$$\mathbb{H}^{2,n} := \Big\{ Z \in \mathbb{H}^{0,n};\ \mathbb{E}\int_0^T |Z_s|^2 ds < \infty \Big\}$$

where $\mathbb{H}^{0,n}$ stands for the collection of all $\mathbb{R}^n$-valued progressively measurable processes on $[0,T]$. We shall also denote by $\mathbb{S}^{2,n}$ for the collection of all continuous processes $(U_t)_{0\le t\le T}$ in $\mathbb{H}^{0,n}$ such that $\mathbb{E}[\sup_{0\le t\le T}|U_t|^2] < +\infty$. For any measurable space $(E,\mathcal{E})$, we denote by $\mathcal{P}(E)$ the space of probability measures on $(E,\mathcal{E})$ assuming that the σ-field $\mathcal{E}$ is understood. When E is a normed space (most often $E = \mathbb{R}^d$ in what follows), we denote by $\mathcal{P}_p(E)$ the subspace of $\mathcal{P}(E)$ of the probability measures of order p, namely those elements of $\mathcal{P}(E)$ which integrate the p-th power of the distance to a fixed point (whose choice is irrelevant in the definition of $\mathcal{P}_p(E)$). For each $p \ge 1$, if μ and μ' are probability measures of order p, $W_p(\mu,\mu')$ denotes the p-Wasserstein's distance defined as

$$W_p(\mu,\mu') = \inf\left\{ \left[\int |x-y|_E^p \pi(dx,dy)\right]^{1/p};\ \pi \in \mathcal{P}_p(E\times E) \text{ with marginals } \mu \text{ and } \mu' \right\}.$$

Notice that if X and X' are random variables of order 2 with values in E, then by definition we have

$$W_2(\mathbb{P}_X, \mathbb{P}_{X'}) \le [\mathbb{E}|X-X'|_E^2]^{1/2}.$$

2.2 Assumptions and Statement of the Main Existence Result

Our goal is to solve fully coupled McKean-Vlasov forward-backward systems of the general form:

$$\begin{aligned} dX_t &= B\big(t, X_t, Y_t, Z_t, \mathbb{P}_{(X_t,Y_t)}\big)dt + \Sigma\big(t, X_t, Y_t, \mathbb{P}_{(X_t,Y_t)}\big)dW_t \\ dY_t &= -F\big(t, X_t, Y_t, Z_t, \mathbb{P}_{(X_t,Y_t)}\big)dt + Z_t dW_t, \quad 0 \le t \le T, \end{aligned} \tag{2.1}$$

with initial condition $X_0 = x_0$ for a given deterministic point $x_0 \in \mathbb{R}^d$, and terminal condition $Y_T = G(X_T, \mathbb{P}_{X_T})$. Here, the unknown processes $(\underline{X}, \underline{Y}, \underline{Z})$ are of dimensions d, p and $p\times m$ respectively, the coefficients B and F map $[0,T]\times\mathbb{R}^d\times\mathbb{R}^p\times\mathbb{R}^{p\times m}\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p)$ into $\mathbb{R}^d$ and $\mathbb{R}^p$ respectively, while the coefficient Σ maps $[0,T]\times\mathbb{R}^d\times\mathbb{R}^p\times\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p)$ into $\mathbb{R}^{d\times m}$ and the function G giving the terminal condition maps $\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)$ into $\mathbb{R}^p$, all these functions being assumed to be Borel-measurable. Recall that the spaces $\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p)$ and $\mathcal{P}_2(\mathbb{R}^d)$ are assumed to be endowed with the topology of the 2-Wasserstein distance W_2 defined earlier.

We now state the standing assumptions of the paper.

(A1). There exists a constant $L \geq 1$, such that for any $t \in [0,T]$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $x, x' \in \mathbb{R}^d$, $y, y' \in \mathbb{R}^p$ and $z, z' \in \mathbb{R}^{p\times m}$,

$$\begin{aligned}&|(B,F)(t,x',y',z',\mu) - (B,F)(t,x,y,z,\mu)| + |\Sigma(t,x',y',\mu) - \Sigma(t,x,y,\mu)|\\&\quad + |G(x',\mu) - G(x,\mu)| \leq L|(x,y,z) - (x',y',z')|.\end{aligned}$$

Moreover, for any $(t,x,y,z) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^p\times\mathbb{R}^{p\times m}$, the coefficients $B(t,x,y,z,\cdot)$, $F(t,x,y,z,\cdot)$, $\Sigma(t,x,y,\cdot)$ and $G(x,\cdot)$ are continuous in the measure argument with respect to the 2-Wasserstein distance.

(A2) The functions Σ and G are bounded, the common bound being also denoted by L. Moreover, for any $t \in [0,T]$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^p$, $z \in \mathbb{R}^{p\times m}$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^p)$,

$$\begin{aligned}|B(t,x,y,z,\mu)| &\leq L\left[1 + |x| + |y| + |z| + \left(\int_{\mathbb{R}^d\times\mathbb{R}^p} |(x',y')|^2 d\mu(x',y')\right)^{1/2}\right],\\ |F(t,x,y,z,\mu)| &\leq L\left[1 + |y| + \left(\int_{\mathbb{R}^d\times\mathbb{R}^p} |y'|^2 d\mu(x',y')\right)^{1/2}\right].\end{aligned}$$

(A3) The function Σ is uniformly elliptic in the sense that for any $t \in [0,T]$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^p$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p)$ the following inequality holds

$$\Sigma(t,x,y,\mu)\Sigma(t,x,y,\mu)^\dagger \geq L^{-1} I_d$$

in the sense of symmetric matrices, where I_d is the d-dimensional identity matrix. Here and throughout the paper, we use the exponent † to denote the transpose of a matrix. Moreover, the function $[0,T] \ni t \hookrightarrow \Sigma(t,0,0,\delta_{(0,0)})$ is also assumed to be continuous.

We can now state the main result of the paper.

Theorem 2.1. *Under (A1–3), the FBSDE* (2.1) *has a solution* $(X,Y,Z) \in \mathbb{S}^{2,d}\times\mathbb{S}^{2,p}\times\mathbb{H}^{2,p\times m}$.

The strategy of the proof was sketched in a simpler setting in [5]. We review this strategy before giving the details. Because of the Markovian nature of the set-up, we expect Y_t and X_t to be connected by a deterministic relationship of the form $Y_t = \varphi(t,X_t)$, φ being a function from $[0,T]\times\mathbb{R}^d$ into $\mathbb{R}^p$ usually called the FBSDE value function. If this is the case, the law of the pair (X_t,Y_t) is entirely determined by the law of X_t since the distribution $\mathbb{P}_{(X_t,Y_t)}$ of (X_t,Y_t) is equal to $(I_d,\varphi(t,\cdot))(\mathbb{P}_{X_t})$. For a random variable X with values in $\mathbb{R}^d$ and for a measurable mapping ψ from $\mathbb{R}^d$ into $\mathbb{R}^p$, we shall denote by $\psi\diamond\mathbb{P}_X$ the image of the distribution $\mathbb{P}_X$ of X under the map $(I_d,\psi) : \mathbb{R}^d \ni x \hookrightarrow (x,\psi(x)) \in \mathbb{R}^d\times\mathbb{R}^p$, that is $\psi\diamond\mathbb{P}_X = (I_d,\psi)(\mathbb{P}_X)$. With this notation in hand, it is natural to look for a function $\varphi : [0,T]\times\mathbb{R}^d \to \mathbb{R}^p$ such that

$$\begin{aligned}dX_t &= B\big(t,X_t,Y_t,Z_t,\varphi(t,\cdot)\diamond\mathbb{P}_{X_t}\big)dt + \Sigma\big(t,X_t,Y_t,\varphi(t,\cdot)\diamond\mathbb{P}_{X_t}\big)dW_t\\ dY_t &= -F\big(t,X_t,Y_t,Z_t,\varphi(t,\cdot)\diamond\mathbb{P}_{X_t}\big)dt + Z_t dW_t, \quad 0 \leq t \leq T,\end{aligned} \tag{2.2}$$

under the constraint that $Y_t = \varphi(t,X_t)$ for $t \in [0,T]$. Translating the above into a nonlinear PDE, φ appears as the solution to a nonlinear PDE of the McKean-Vlasov type. The strategy we use below consists in recasting the stochastic system (2.2) into a well-posed fixed point problem over the arguments $(\varphi,(\mathbb{P}_{X_t})_{0\leq t\leq T})$. The first step is to use $\varphi(t,\cdot)\diamond\mathbb{P}_{X_t}$ as an input and then to solve (2.2) as a standard FBSDE. In order to do so, we use known existence results for standard FBSDEs.

Remark 2.2. *The proof of Theorem* (2.1) *given in this note can be used to derive existence of a solution of* (2.1) *when the law* $\mathbb{P}_{(X_t,Y_t)}$ *is replaced by* $\mathbb{P}_{(X_t,Y_t,Z_t)}$. *Indeed,*

Z_t is also given by a function of X_t in the same way as Y_t is, since $Z_t = v(t, X_t)$ with $v(t,x) = \partial_x u(t,x)\Sigma(t,x,u(t,x),u(t,\cdot)\diamond\mathbb{P}_{X_t})$ whenever $Y_t = u(t,X_t)$ (we prove below that $\partial_x u$ makes sense). We refrain from giving the details to keep the technicalities to a minimum, and because we do not know of a practical application of such a generalization.

We also emphasize that Σ is here assumed to be independent of $(Z_t)_{0\le t\le T}$. The reason is that, even in the classical case when the coefficients do not depend upon the law of the solution, solvability of the equation becomes really challenging when $\Sigma(t,\cdot)$ depends on Z_t. The standard framework for handling the equation then consists of specific monotonicity assumptions, which are intended to mimic the role played by convexity in the analysis of so-called adjoint *FBSDEs arising in optimal stochastic control. We refer the reader to Subsection* (3.2) *for the form of the adjoint FBSDE in the McKean-Vlasov setting and to the paper [3] for a much more complete overview of such a strategy.*

We finally mention that the 2-Wasserstein distance we use here could be replaced by another p-Wasserstein distance, for some $p \ge 1$. Indeed, when $p < 2$, continuity w.r.t. the p-Wasserstein distance implies continuity w.r.t. the 2-Wasserstein distance. For $p > 2$, continuity w.r.t. the p-Wasserstein distance implies continuity w.r.t. the 2-Wasserstein distance along sequences that are uniformly integrable in L^p, which turns out to be the case in all the continuity arguments applied below. Similarly, the 2-Wasserstein distance could be replaced by the Lévy-Prohorov distance, which generates the topology of weak convergence of probability measures, as continuity with respect to the Lévy-Prohorov distance implies continuity for the Wasserstein distance.

2.3 Preliminary

Our fixed point argument relies on the following lemma which puts together the existence and uniqueness result contained in Theorem 2.6 of Delarue [6] and the control of the FBSDE value function provided by Corollary 2.8 of [6]:

Lemma 2.3. *On the top of (A1-3), let us also assume that B and F are bounded by L and that there exists a family of moduli of uniform continuity in the argument measure, that is a measurable function $[0,T]\times\mathbb{R}\ni(t,r)\mapsto w_t(r)$, bounded by L and satisfying $\lim_{r\searrow 0} w(t,r) = 0$ for all $t\in[0,T]$, such that, for all $t\in[0,T]$, $x\in\mathbb{R}^d$, $y\in\mathbb{R}^p$ and $z\in\mathbb{R}^{p\times m}$,*

$$
\begin{aligned}
&\forall \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p),\ \ |(B,F)(t,x,y,z,\mu') - (B,F)(t,x,y,z,\mu)| \le w_t\big(W_2(\mu',\mu)\big),\\
&\forall \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p),\ \ |\Sigma(t,x,y,\mu') - \Sigma(t,x,y,\mu)| \le w_t\big(W_2(\mu',\mu)\big),\\
&\forall \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d),\ \ |G(x,\mu') - G(x,\mu)| \le w_T\big(W_2(\mu,\mu')\big).
\end{aligned}
\tag{2.3}
$$

Then, given a probability measure $\nu'\in\mathcal{P}_2(\mathbb{R}^d)$, a deterministic continuous function $\nu : [0,T]\ni t\hookrightarrow \nu_t\in\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^p)$, and an initial condition $(t,x)\in[0,T]\times\mathbb{R}^d$, the forward-backward system

$$
\begin{aligned}
dX_s &= B\big(s,X_s,Y_s,Z_s,\nu_s\big)ds + \Sigma(s,X_s,Y_s,\nu_s)dW_s\\
dY_s &= -F\big(s,X_s,Y_s,Z_s,\nu_s\big)ds + Z_s dW_s, \quad t\le s\le T,
\end{aligned}
\tag{2.4}
$$

with $X_t = x$ as initial condition and $Y_T = G(X_T,\nu')$ as terminal condition, has a unique solution, denoted by $(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})_{t\le s\le T}$. Moreover, the FBSDE value function $u : [0,T]\times\mathbb{R}^d\ni(t,x)\hookrightarrow Y_t^{t,x}\in\mathbb{R}^p$ is bounded by a constant γ depending only upon T and L, and is $1/2$-Hölder continuous in time and Lipschitz continuous in space in the sense that:

$$
|u(t,x) - u(t',x')| \le \Gamma\big(|t-t'|^{1/2} + |x-x'|\big),
$$

for some constant Γ only depending upon T and L. In particular, both γ and Γ are independent of ν' and ν. Finally, it holds $Y_s^{t,x} = u(s, X_s^{t,x})$ for any $t\le s\le T$.

The boundedness assumption on B and F is stronger than what is necessary for the result of Lemma (2.3) to hold. For instance, the result of [6] only requires that the bound (A2) holds without $|x|$ in the right hand side. We will not need this extra level of generality. We use this existence result in the following way. We start with a bounded continuous function φ from $[0,T]\times\mathbb{R}^d$ into $\mathbb{R}^p$ and a probability measure μ on $C([0,T],\mathbb{R}^d)$ which we want to think of as the law $\mathbb{P}_X$ of the solution, we denote by μ_t its marginal distributions on $\mathbb{R}^d$, and we apply the above existence result for (2.4) to $\nu'=\mu_T$ and $\nu_t=\varphi(t,\cdot)\diamond\mu_t$ for $t\in[0,T]$ and solve:

$$\begin{aligned} dX_t &= B\big(t,X_t,Y_t,Z_t,\varphi(t,\cdot)\diamond\mu_t\big)dt+\Sigma(t,X_t,Y_t,\varphi(t,\cdot)\diamond\mu_t)dW_t\\ dY_t &= -F\big(t,X_t,Y_t,Z_t,\varphi(t,\cdot)\diamond\mu_t\big)dt+Z_t dW_t,\quad 0\le t\le T, \end{aligned}\tag{2.5}$$

with the terminal condition $Y_T=G(X_T,\mu_T)$ and a prescribed initial condition $X_0=x_0\in\mathbb{R}^d$. The following estimate will be instrumental in the proof of the main result.

Lemma 2.4. *On the top of (A1-3), let us also assume that B and F are bounded by L and that there exists a family of moduli of uniform continuity such that* (2.3) *holds true. Then, there exists a positive constant Γ, depending on T and L only, such that, for any inputs (φ,μ) and (φ',μ') as above, the processes (X,Y,Z) and (X',Y',Z') obtained by solving* (2.5) *with (φ,μ) and (φ',μ') respectively, satisfy*

$$\begin{aligned} &\mathbb{E}\big[\sup_{0\le t\le T}|X_t-X'_t|^2\big]+\mathbb{E}\big[\sup_{0\le t\le T}|Y_t-Y'_t|^2\big]+\mathbb{E}\int_0^T|Z_t-Z'_t|^2dt\\ &\qquad\le\Gamma\bigg(w_T^2\big(W_2(\mu_T,\mu'_T)\big)+\int_0^T w_t^2\big(W_2(\varphi(t,\cdot)\diamond\mu_t,\varphi'(t,\cdot)\diamond\mu'_t)\big)dt\bigg). \end{aligned}\tag{2.6}$$

Proof. For small time $T>0$, this estimate follows immediately from the main estimate Theorem 1.3 p. 218 of [6], the Lipschitz assumption (A1) and the uniform continuity property (2.3). We only need to show that one can extend it to arbitrarily large values of T. Notice that Lemma (2.3) gives the existence of the FBSDE value functions u and u' such that $Y_t=u(t,X_t)$ and $Y'_t=u(t,X'_t)$ for all $t\in[0,T]$.

As in Corollary 2.8 of [6], we choose a regular subdivision $0=T_0<T_1<\cdots<T_{N-1}<T_N=T$ so that the common length of the intervals $[T_i,T_{i+1}]$ is small enough in order to apply the main estimate Theorem 1.3 p. 218 of [6]. For any $i\in\{0,\cdots,N-1\}$ we have:

$$\begin{aligned} &\mathbb{E}\big[\sup_{T_i\le t\le T_{i+1}}|X_t-X'_t|^2\big]+\mathbb{E}\big[\sup_{T_i\le t\le T_{i+1}}|Y_t-Y'_t|^2\big]+\mathbb{E}\int_{T_i}^{T_{i+1}}|Z_t-Z'_t|^2dt\\ &\qquad\le\Gamma\bigg(\mathbb{E}[|X_{T_i}-X'_{T_i}|^2]+\mathbb{E}[|u(T_{i+1},X_{T_{i+1}})-u'(T_{i+1},X_{T_{i+1}})|^2]\\ &\qquad\qquad+\int_{T_i}^{T_{i+1}}w_t^2\big(W_2(\varphi'(t,\cdot)\diamond\mu'_t,\varphi(t,\cdot)\diamond\mu_t)\big)dt\bigg). \end{aligned}\tag{2.7}$$

For simplicity, we denote the left-hand side by $\Theta(T_i,T_{i+1})$ and we let $\delta_t=w_t^2(W_2(\varphi'(t,\cdot)\diamond\mu'_t,\varphi(t,\cdot)\diamond\mu_t))$.

We first consider the last interval $[T_{N-1},T_N]$ corresponding to the case $i=N-1$. Since $T_N=T$ we have $u(T,\cdot)=G(\cdot,\mu_T)$ and $u'(T,\cdot)=G(\cdot,\mu'_T)$ so that using the Lipschitz property of G in x and the uniform continuity in μ, as stated in (2.3), we get:

$$\Theta(T_{N-1},T)\le\Gamma\bigg(\mathbb{E}[|X_{T_{N-1}}-X'_{T_{N-1}}|^2]+\delta_T+\int_{T_{N-1}}^T\delta_t dt\bigg),$$

this estimate being true for all φ, φ', μ, μ' and all possible initial conditions for the processes $\underline{X}$ and $\underline{X}'$. Note that we can assume $\Gamma>1$ without any loss of generality, and allow the value of Γ to change from line to line as long as this new value depends only upon T and L. Since the FBSDE value function u (resp. u') depends only upon φ and μ (resp. φ' and μ'), we can choose to keep φ, φ', μ, μ', and set $X_{T_{N-1}} = X'_{T_{N-1}} = x$ for an arbitrary $x \in \mathbb{R}^d$. Then the above inequality implies

$$\sup_{x\in\mathbb{R}^d} |u(T_{N-1}, x) - u'(T_{N-1}, x)|^2 \leq \Gamma\bigg(\delta_T + \int_{T_{N-1}}^T \delta_t dt\bigg).$$

We can now plug this estimate into inequality (2.7) with $i = N-2$ to get:

$$\Theta(T_{N-2}, T_{N-1}) \leq \Gamma\bigg(\mathbb{E}[|X_{T_{N-2}} - X'_{T_{N-2}}|^2] + \delta_T + \int_{T_{N-2}}^T \delta_t dt\bigg).$$

As before, we can write what this estimate gives if we keep φ, φ', μ, μ', and set $X_{T_{N-2}} = X'_{T_{N-2}} = x$ for an arbitrary $x \in \mathbb{R}^d$:

$$\sup_{x\in\mathbb{R}^d} |u(T_{N-2}, x) - u'(T_{N-2}, x)|^2 \leq \Gamma\bigg(\delta_T + \int_{T_{N-2}}^T \delta_t dt\bigg).$$

Plugging this estimate into inequality (2.7) with $i = N-3$ we get:

$$\Theta(T_{N-3}, T_{N-2}) \leq \Gamma\bigg(\mathbb{E}[|X_{T_{N-3}} - X'_{T_{N-3}}|^2] + \delta_T + \int_{T_{N-3}}^T \delta_t dt\bigg).$$

Iterating and summing up these estimates we get (as before the value of the constants can change from line to line)

$$\sum_{i=0}^{N-1} \Theta(T_i, T_{i+1}) \leq \Gamma \sum_{i=0}^{N-1} \bigg(\mathbb{E}[|X_{T_i} - X'_{T_i}|^2] + \delta_T + \int_{T_i}^T \delta_t dt\bigg),$$

from which we get the desired estimate (2.6) after noticing that for each $i \geq 1$, we have:

$$\mathbb{E}[|X_{T_i} - X'_{T_i}|^2] \leq \mathbb{E}[\sup_{T_{i-1}\leq t\leq T_i} |X_t - X'_t|^2] \leq \Gamma\bigg(\mathbb{E}[|X_{T_{i-1}} - X'_{T_{i-1}}|^2] + \delta_T + \int_{T_{i-1}}^T \delta_t dt\bigg).$$

from which we easily conclude. □

We shall estimate the integral in the right hand side of (2.6) from the remark:

$$\begin{aligned} W_2(\varphi(t,\cdot) \diamond \mu_t, \varphi'(t,\cdot) \diamond \mu'_t) \leq C \bigg[& W_2(\mu_t, \mu'_t) + W_2(\varphi(t,\cdot)(\mu_t), \varphi(t,\cdot)(\mu'_t)) \\ & + \bigg(\int_{\mathbb{R}^d} |(\varphi - \varphi')(t,x)|^2 d\mu'_t(x)\bigg)^{1/2}\bigg]. \end{aligned} \tag{2.8}$$

2.4 Fixed Point Argument in the Bounded Case

In this subsection we still assume that the coefficients B and F are bounded by the constant L and that there exists a family of moduli of uniform continuity satisfying (2.3). For any bounded continuous function $\varphi : [0,T] \times \mathbb{R}^d \to \mathbb{R}^p$ and for any probability measure $\mu \in \mathcal{P}_2(\mathcal{C}([0,T];\mathbb{R}^d))$, if we denote by μ_t the time t marginal of μ, the map $[0,T] \ni t \hookrightarrow \varphi(t,\cdot) \diamond \mu_t \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^p)$ is continuous. So by Lemma (2.3), there exists a unique triplet $(X_t, Y_t, Z_t)_{0\leq t\leq T}$ satisfying (2.5). Moreover, there exists a bounded and continuous mapping u from $[0,T] \times \mathbb{R}^d$ into $\mathbb{R}^p$ such that $Y_t = u(t, X_t)$. This maps the input (φ, μ) into the output $(u, \mathbb{P}_X)$ and our goal is to find a fixed point for this map. We shall take advantage of the a-priori L^∞ bound on u to restrict the choice of the functions φ to the set:

$$E_1 = \big\{\varphi \in \mathcal{C}([0,T]\times\mathbb{R}^d;\mathbb{R}^p);\quad \forall (t,x)\in[0,T]\times\mathbb{R}^d,\quad |\varphi(t,x)|\le\gamma\big\}. \tag{2.9}$$

Similarly, since the drift B and the volatility Σ are uniformly bounded, the fourth moment of the supremum $\sup_{0\le t\le T}|X_t|$ is bounded by a constant depending only upon the bounds of B and Σ. Consequently, we shall choose the input measure μ in the set:

$$E_2 = \bigg\{\mu\in\mathcal{P}_4\big(\mathcal{C}([0,T];\mathbb{R}^d)\big);\int_{\mathcal{C}([0,T];\mathbb{R}^d)}\sup_{0\le t\le T}|w_t|^4\,d\mu(w)\le\gamma'\bigg\} \tag{2.10}$$

for γ' appropriately chosen. We then denote by E the Cartesian product $E=E_1\times E_2$. We view E as a subset of the product vector space $V=V_1\times V_2$, where $V_1=\mathcal{C}_b([0,T]\times\mathbb{R}^d;\mathbb{R}^p)$ stands for the space of bounded continuous functions from $[0,T]\times\mathbb{R}^d$ into $\mathbb{R}^p$, and $V_2=\mathcal{M}_b(\mathcal{C}([0,T];\mathbb{R}^d))$ for the space of finite signed measures on the space $\mathcal{C}([0,T];\mathbb{R}^d)$ endowed with the Borel σ-field generated by the topology of uniform convergence. On V_1, we use the exponentially weighted supremum-norm

$$\|h\|_1 = \sup_{(t,x)\in[0,T]\times\mathbb{R}^d} e^{-|x|}|h(t,x)|,$$

and on V_2 the Kantorovitch-Rubinstein norm

$$\|\mu\|_2 = \sup\bigg\{\int_{\mathcal{C}([0,T];\mathbb{R}^d)} F d\mu;\quad F\in\mathrm{Lip}_1(\mathcal{C}([0,T];\mathbb{R}^d)),\ \sup_{h\in\mathcal{C}([0,T];\mathbb{R}^d)}|F(h)|\le 1\bigg\}.$$

Here, $\mathrm{Lip}_1(\mathcal{C}([0,T];\mathbb{R}^d))$ stands for the Lip-1 functions on $\mathcal{C}([0,T];\mathbb{R}^d)$ equipped with the metric of uniform convergence on compact subsets.

We emphasize that E_1 is a convex closed bounded subset of V_1. Moreover, we notice that the convergence for the norm $\|\cdot\|_1$ of a sequence of functions in E_1 is equivalent to the uniform convergence on compact subsets of $[0,T]\times\mathbb{R}^d$. Similarly, E_2 is a convex closed bounded subset of V_2 as the convergence of non-negative measures on a metric space for the Kantorovitch-Rubinstein norm implies weak convergence of measures. We now claim:

Lemma 2.5. *Assume that, in addition to (A1–3), the coefficients B and F are also bounded by L and that there exists a family of moduli of uniform continuity such that* (2.3) *holds true. Then, the mapping $\Phi : E\ni(\varphi,\mu)\hookrightarrow(u,\mathbb{P}_X)\in E$ defined above is continuous and has a relatively compact range.*

Proof. We first check the continuity of Φ. Given a sequence (φ^n,μ^n) in E converging towards $(\varphi,\mu)\in E$ with respect to the product norm on $V_1\times V_2$, and given the corresponding solutions (X^n,Y^n,Z^n) and (X,Y,Z) obtained by solving (2.5) with (φ^n,μ^n) and (φ,μ) respectively, we have (compare with (2.8)): (i) for any $t\in[0,T]$, $W_2(\mu_t,\mu^n_t)\to 0$ as $n\to+\infty$ since $(\mu^n_t)_{n\ge1}$ converges weakly towards μ_t and the moments of order 4 of the measures $(\mu^n_t)_{n\ge1}$ are uniformly bounded by γ'; (ii) by continuity and boundedness of φ, and by a similar argument, $W_2(\varphi(t,\cdot)(\mu_t),\varphi(t,\cdot)(\mu^n_t))$ converges toward 0 as $n\to+\infty$; (iii) since the sup-norms of all the φ^n are not greater than γ, the tightness of the measures $(\mu^n)_{n\ge1}$ together with the uniform convergence of $(\varphi^n)_{n\ge1}$ towards φ on compact sets can be used to prove that

$$\forall t\in[0,T],\quad \lim_{n\to+\infty}\int_{\mathbb{R}^d}|(\varphi-\varphi^n)(t,y)|^2 d\mu^n_t(y)=0.$$

Therefore, by (2.8),

$$\forall t \in [0,T], \quad \lim_{n\to+\infty} W_2(\varphi(t,\cdot) \diamond \mu_t, \varphi^n(t,\cdot) \diamond \mu_t^n) = 0,$$

so that, by Lebesgue dominated theorem,

$$\lim_{n\to+\infty} \int_0^T w_t^2\big(W_2(\varphi(t,\cdot) \diamond \mu_t, \varphi^n(t,\cdot) \diamond \mu_t^n)\big) dt = 0.$$

Similarly, $W_2(\mu_T, \mu_T^n) \to 0$ as n tends to $+\infty$. From (2.7) and (2.6), we obtain

$$\lim_{n\to+\infty} \left[\mathbb{E} \sup_{0\le t\le T} |X_t - X_t^n|^2 + \mathbb{E} \sup_{0\le t\le T} |Y_t - Y_t^n|^2 + \mathbb{E} \int_0^T |Z_t - Z_t^n|^2 dt \right] = 0,$$

from which we deduce that $\mathbb{P}_{X^n}$ converges towards $\mathbb{P}_X$ as n tend to $+\infty$ for the topology of weak convergence of measures and thus for $\|\cdot\|_2$. Denoting by u^n the FBSDE value function which is a function from $[0,T] \times \mathbb{R}^d$ into $\mathbb{R}^p$ such that $Y_t^n = u^n(t, X_t^n)$, and by u the FBSDE value function for which $Y_t = u(t, X_t)$, we deduce that

$$\lim_{n\to+\infty} \sup_{0\le t\le T} \mathbb{E}\big[|u^n(t, X_t^n) - u(t, X_t)|^2\big] = 0.$$

By Lemma (2.3), we know that all the mappings $(u^n)_{n\ge 1}$ are Lipschitz continuous with respect to x, uniformly with respect to n. Therefore

$$\lim_{n\to+\infty} \sup_{0\le t\le T} \mathbb{E}\big[|u^n(t, X_t) - u(t, X_t)|^2\big] = 0.$$

Moreover, by Arzèla-Ascoli's theorem, the sequence $(u^n)_{n\ge1}$ is relatively compact for the uniform convergence on compact sets, so denoting by $\hat{u}$ the limit of a subsequence converging for the norm $\|\cdot\|_1$, we deduce that, for any $t \in [0,T]$, $\hat{u}(t,\cdot) = u(t,\cdot)$ $\mathbb{P}_{X_t}$-a.s.. By Girsanov Theorem, $\mathbb{P}_{X_t}$ is equivalent to Lebesgue measure for any $t \in (0,T]$, so that $\hat{u}(t,\cdot) = u(t,\cdot)$ for any $t \in (0,T]$. By continuity of u and $\hat{u}$ on the whole $[0,T] \times \mathbb{R}^d$, equality holds at $t = 0$ as well. This shows that $(u^n)_{n\ge1}$ converges towards u for $\|\cdot\|_1$ and completes the proof of the continuity of Φ.

We now prove that $\Phi(E)$ is relatively compact for the product norm of $V_1 \times V_2$. Given $(u,\nu) = \Phi(\varphi,\mu)$ for some $(\varphi,\mu) \in E$, we know from Lemma (2.3) that u is bounded by γ and $(1/2,1)$-Hölder continuous with respect to (t,x), the Hölder constant being bounded by Γ. In particular, u remains in a compact subset of $\mathcal{C}([0,T] \times \mathbb{R}^d; \mathbb{R}^p)$ for the topology of uniform convergence on compact sets as (φ,μ) varies over E. Similarly, ν remains in a compact set when (φ,μ) varies over E. Indeed, if $\mathbb{P}_X = \nu$ is associated to (φ,μ), the modulus of continuity of X is controlled by the fact that B and Σ are bounded by constants independent of φ and μ. □

We have completed all the steps needed to get a quick proof of the main result of this subsection.

Proposition 2.6. *Assume that, in additition to (A1–3), the coefficients B, Σ, F and G are bounded by L and that there exists a family of moduli of uniform continuity satisfying* (2.3). *Then equation* (2.1) *has a solution* $(X,Y,Z) \in \mathbb{S}^{2,d} \times \mathbb{S}^{2,p} \times \mathbb{H}^{2,p\times m}$.

Proof. By Schauder's fixed point theorem, Φ has a fixed point (φ,μ). As explained in our description of the strategy of proof, solving (2.5) with this (φ,μ) as input, and denoting by $(X_t,Y_t,Z_t)_{0\le t\le T}$ the resulting solution, by definition of a fixed point, we have $Y_t = \varphi(t,X_t)$ for any $t \in [0,T]$, a.s., and $\mathbb{P}_X = \mu$ so that $\mathbb{P}_{X_t} = \mu_t$. In particular, $\varphi(t,\cdot) \diamond \mu_t$ coincides with $\mathbb{P}_{(X_t,Y_t)}$. We conclude that $(X_t,Y_t,Z_t)_{0\le t\le T}$ satisfies (2.1). □

2.5 Relaxing the Boundedness Condition.

We now complete the proof of Theorem (2.1) when the coefficients only satisfy (A1–3). The proof consists in approximating B, F, Σ and G by sequences of coefficients $(B^n)_{n\geq 1}$, $(F^n)_{n\geq 1}$, $(\Sigma^n)_{n\geq 1}$ and $(G^n)_{n\geq 1}$ satisfying the assumption of Proposition (2.6).

Proof. (Theorem 1) For any $n \geq 1$, $t \in [0,T]$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^p$, $z \in \mathbb{R}^{p\times m}$, $\mu \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^p)$ and $\mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we set:

$$
\begin{aligned}
&(B^n, F^n)(t,x,y,z,\mu) = (B,F)\big(t, \Pi_n^{(d)}(x), \Pi_n^{(p)}(y), \Pi_n^{(p\times m)}(z), \Pi_n^{(d+p)}(\mu)\big),\\
&\Sigma^n(t,x,y,\mu) = \Sigma\big(t, \Pi_n^{(d)}(x), \Pi_n^{(p)}(y), \Pi_n^{(d\times p)}(\mu)\big), \quad G^n(x,\mu') = G\big(x, \Pi_n^{(d)}(\mu')\big),
\end{aligned}
$$

where, for any integer k, $\Pi_n^{(k)}$ is the orthogonal projection from $\mathbb{R}^k$ onto the k-dimensional ball of center 0 and of radius n and, for any probability measure ν on $\mathbb{R}^k$, $\Pi_n^{(k)}(\nu)$ is the push-forward of ν by $\Pi_n^{(k)}$.

For each n, assumptions (A1-3) are satisfied with $(B^n, F^n, \Sigma^n, G^n)$ instead of (B, F, Σ, G). Moreover, noticing that, for any integers $k, \ell \geq 1$, the set of pairs (χ, ν), χ belonging to the k-dimensional ball of center 0 and of radius n and ν being a probability measure supported by the ℓ-dimensional ball of center 0 and radius n, is a compact subset of $\mathbb{R}^k \times \mathcal{P}_2(\mathbb{R}^\ell)$ (endowed with the product topology generated by the Euclidean distance and the 2-Wasserstein distance), we deduce that, for any $t \in [0,T]$, the coefficients $B^n(t,\cdot)$, $F^n(t,\cdot)$, $\Sigma^n(t,\cdot)$ and G^n are uniformly continuous in the measure argument, as required in the statement of Proposition (2.6). Therefore, we can denote by (X^n, Y^n, Z^n) the solution of (2.1) given by Proposition (2.6) when the system (2.1) is driven by B^n, F^n, Σ^n and G^n. As explained in the previous subsection, the process Y^n satisfies $Y_t^n = u^n(t, X_t^n)$, for any $t \in [0,T]$, for some deterministic function u^n. The first step of the proof is to establish the relative compactness of the families $(u^n)_{n\geq 1}$ and $(\mathbb{P}_{X^n})_{n\geq 1}$.

We notice first that the processes $(Y^n)_{n\geq 1}$ are uniformly bounded by a constant that depends upon L only. Indeed, applying Itô's formula and using the specific growth condition (A2), we get:

$$
\forall t \in [0,T], \quad \mathbb{E}\big[|Y_t^n|^2\big] \leq C + C\int_t^T \mathbb{E}\big[|Y_s^n|^2\big]ds,
$$

for some constant C depending on T and L only. As usual, the value of C may vary from line to line. By Gronwall's lemma, we deduce that the quantity $\sup_{0\leq t\leq T}\mathbb{E}[|Y_t^n|^2]$ can be bounded in terms of T and L only. Injecting this estimate into (A2) shows that the driver $(-F^n(t, X_t^n, Y_t^n, Z_t^n, \mathbb{P}_{(X_t^n, Y_t^n)}))_{0\leq t\leq T}$ is bounded by $(C(1+|Y_t^n|))_{0\leq t\leq T}$, for a possibly new value of C. Using now the uniform boundedness of the function G giving the terminal condition Y_T, we conclude that the processes $(Y_t^n)_{0\leq t\leq T}$ are bounded, uniformly in $n \geq 1$. Indeed, for any $n \geq 1$, the process $(|Y_t^n|^2)_{0\leq t\leq T}$ can be seen as a solution of a BSDE with random coefficients; thus, it can be compared with the deterministic solution of a deterministic BSDE with a constant terminal condition. In particular, there exists a positive constant, say γ, such that, for any $n \geq 1$ for any $t \in [0,T]$, $|u^n(t, X_t^n)| \leq \gamma$. Since for each $t \in [0,T]$, X_t^n has a density with respect to Lebesgue's measure on $\mathbb{R}^d$ (recall that B^n is bounded and $\Sigma^n(\Sigma^n)\dagger$ is Lipschitz and bounded from above and from below away from 0), we deduce that u^n is bounded by γ on the whole $[0,T]\times\mathbb{R}^d$, and for any $n \geq 1$. As a by-product, we get

$$
\mathbb{E}\bigg[\bigg(\int_0^T |Z_s^n|^2 ds\bigg)^2\bigg] \leq C. \tag{2.11}
$$

When $p = 1$, Theorem 2.1 in [8] says that the value function u^n is a continuous solution with Sobolev derivatives of order 1 in time and of order 2 in space of a quasi-linear PDE. As mentioned in the conclusion of [8], the result remains true when $p \geq 2$, the value function u^n then solving a system of quasilinear PDEs. Using the fact that the proofs of Theorems 1.1 and 1.3 of [7] are local (even though they are stated under global assumptions), we deduce that the functions $(u^n)_{n\geq 1}$ are continuous in (t,x), uniformly in $n \geq 1$, on any compact subsets of $[0,T] \times \mathbb{R}^d$. In a similar way, by Theorem 2.7 of [7], the proof of which is also local, the functions $((u^n(t,\cdot))_{n\geq 1})_{0\leq t\leq T}$ are locally Lipschitz-continuous, uniformly in $t \in [0,T]$ and in $n \geq 1$.

Applying Itô's formula to $|X^n|^2$, using the growth conditions (A2), the uniform boundedness of Y^n_t, the boundedness of Σ^n and the bound (2.11), we can use Gronwall's lemma and get the existence of a finite constant C such that, for any $n \geq 1$,

$$\mathbb{E}[\sup_{0\leq t\leq T} |X^n_t|^4] \leq C. \tag{2.12}$$

Using the bound (A2) for B^n with the same constant L, together with the uniform boundedness of the paths $(Y^n_t)_{0\leq t\leq T}$, (2.11) and (2.12), it is easy to check that

$$\mathbb{E}[|X^n_t - X^n_s|^4] \leq C|t-s|^2 \tag{2.13}$$

for all $s,t \in [0,T]$ for a constant C independent of n, s and t. Consequently, Kolmogorov's criterion shows that the family $(\mathbb{P}_{X^n})_{n\geq 1}$ of probability measures on $\mathcal{C}([0,T];\mathbb{R}^d)$ is tight. Replacing Y^n_t by $u^n(t,X^n_t)$ in $\Sigma(t,X^n_t,Y^n_t,\mathbb{P}_{(X^n_t,Y^n_t)})$, we see that each component of u^n satisfies a PDE, and since $\Sigma^n(t,x,u^n(t,x),\mathbb{P}_{(X^n_t,Y^n_t)})$ is locally Hölder continuous uniformly in $n \geq 1$, we can use Theorem 4, Chapter 7 Section 2 of [9] and conclude that the gradients of these components are locally Hölder continuous in (t,x) on compact subsets of $[0,T)\times\mathbb{R}^d$, uniformly in $n \geq 1$. Consequently, one can thus extract a subsequence $(n_k)_{k\geq 1}$ such that u^{n_k} and $\partial_x u^{n_k}$ converge uniformly on compact subsets of $[0,T] \times \mathbb{R}^d$ and of $[0,T) \times \mathbb{R}^d$ respectively, and $\mathbb{P}_{X^{n_k}}$ converges towards a probability measure μ on $\mathcal{C}([0,T];\mathbb{R}^d)$. If we denote by u the limit of u^{n_k}, the function u is continuously differentiable with respect to x on $[0,T) \times \mathbb{R}^d$, and $\partial_x u^{n_k}$ converges towards $\partial_x u$ uniformly on compact subsets of $[0,T) \times \mathbb{R}^d$. As before, we can use the uniform boundedness of the functions u^n, the bound (2.12) and the uniform convergence of u^{n_k} towards u on any compact subset of $[0,T) \times \mathbb{R}^d$ to conclude that for any $t \in [0,T]$,

$$\lim_{p\to+\infty} W_2\big(u^{n_k}(t,\cdot)\diamond\mathbb{P}_{X^{n_k}_t}, u(t,\cdot)\diamond\mu_t\big) = 0, \tag{2.14}$$

where μ_t stands for the marginal law of μ of time index t. This proves the convergence of the measure argument (up to a subsequence).

We now prove the convergence of the forward components $(X^{n_k})_{k\geq 1}$. Recalling that the functions $(u^n, \partial_x u^n)_{n\geq 1}$ are locally Hölder continuous in (t,x) on compact subsets of $[0,T) \times \mathbb{R}^d$, uniformly in $n \geq 1$, we deduce from [13] that the functions $(\partial_x u^n)_{n\geq 1}$ are in fact locally Lipschitz continuous in x on compact subsets of $[0,T) \times \mathbb{R}^d$, uniformly in $n \geq 1$. For each $R \geq 1$, we then denote by τ^n_R the first time the process $\underline{X}^n$ leaves the cylinder of center 0, radius R and length $T(1-1/R)$, in other words $\tau^n_R = \inf\{t \geq 0 : |X^n_t| \geq R\} \wedge (T(1-1/R))$. Using the uniform convergence of $(u^{n_k}, \partial_x u^{n_k})$ to $(u, \partial_x u)$ on compact subsets of $[0,T) \times \mathbb{R}^d$, we have for each fixed $R > 0$:

$$\lim_{k\to+\infty} \sup_{\ell,\ell'\geq 0} \mathbb{E}\Big[\sup_{0\leq t\leq \tau^{n_{k+\ell}}_R \wedge \tau^{n_{k+\ell'}}_R} |X^{n_{k+\ell}}_t - X^{n_{k+\ell'}}_t|^2\Big] = 0,$$

since for each $n \geq 1$ we have:

$$\begin{aligned} dX^n_t &= B^n\big(t, X^n_t, u^n(t,X^n_t), v^n(t,X^n_t), u^n(t,\cdot)\diamond\mathbb{P}_{X^n_t}\big)dt \\ &\quad + \Sigma^n(t, X^n_t, u^n(t,X^n_t), u^n(t,\cdot)\diamond\mathbb{P}_{X^n_t})dW_t, \end{aligned} \tag{2.15}$$

for $0 \le t \le T$, with $v^n(t,x) = \partial_x u^n(t,x)\Sigma^n(t,x,u^n(t,x),u^n(t,\cdot)\diamond \mathbb{P}_{X_t^n})$.

Since (2.12) implies that, for any $\varepsilon \in (0,T)$,

$$\lim_{R\to+\infty} \sup_{\ell,\ell'\ge 1} \mathbb{E}\Big[\sup_{0\le t\le T-\varepsilon} |X_t^{n_{k+\ell'}} - X_t^{n_{k+\ell}}|^2 \mathbf{1}_{\{T-\varepsilon > \tau_R^{n_{k+\ell}} \wedge \tau_R^{n_{k+\ell'}}\}}\Big] = 0,$$

we must have, for any $\varepsilon \in (0,T)$,

$$\lim_{k\to+\infty} \sup_{\ell,\ell'\ge 0} \mathbb{E}\Big[\sup_{0\le t\le T-\varepsilon} |X_t^{n_{k+\ell}} - X_t^{n_{k+\ell'}}|^2\Big] = 0.$$

By (2.13), the above limit also holds when $\varepsilon = 0$. This shows that the sequence $(\underline{X}^{n_k})_{k\ge1}$ is a Cauchy sequence. We denote by $nderlineX$ the limit.

Now, we can take the limit along the subsequence $(n_k)_{k\ge1}$ in (2.15). Since μ is the law of $\underline{X}$, the local uniform convergence of u^{n_k} and $\partial_x u^{n_k}$ towards u and $\partial_x u$, together with (2.14) imply that

$$dX_t = B\big(t,X_t,u(t,X_t),v(t,X_t),u(t,\cdot)\diamond\mathbb{P}_{X_t}\big)dt + \Sigma(t,X_t,u(t,X_t),u(t,\cdot)\diamond\mathbb{P}_{X_t})dW_t$$

for $0 \le t \le T$, with $v(t,x) = \partial_x u(t,x)\Sigma(t,x,u(t,x),u(t,\cdot)\diamond\mathbb{P}_{X_t})$, which is exactly the forward component of the McKean-Vlasov FBSDE (2.1) provided we set $Y_t = u(t,X_t)$ and $Z_t = v(t,X_t)$ for $t\in[0,T)$. It is plain to deduce that the sequences $(\underline{Y}^{n_k})_{k\ge1}$ and $(\underline{Z}^{n_k})_{k\ge1}$ are Cauchy sequences for the norms $\mathbb{E}[\sup_{0\le s\le T}|\cdot_s|^2]^{1/2}$ and $\mathbb{E}[\int_0^T|\cdot_s|^2 ds]^{1/2}$ respectively. Denoting the respective limits by $\underline{Y}$ and $\underline{Z}$, it holds, $\mathbb{P}$-a.s., for any $t\in[0,T]$, $Y_t = u(t,X_t)$, and, $\mathbb{P}\otimes dt$ a.e., $Z_t = v(t,X_t)$. Passing to the limit in (2.1) with (B,F,Σ,G) therein replaced by $(B^{n_k},F^{n_k},\Sigma^{n_k},G^{n_k})$, we deduce that $(\underline{X},\underline{Y},\underline{Z})$ satisfies (2.1). □

2.6 Counter-Example to Uniqueness

We close this section with a counter-example showing that uniqueness cannot hold in general under assumptions (A1–3) and additional Cauchy-Lipschitz property in the measure argument (with respect to the 2-Wasserstein distance), even in the case $d = m = p = 1$. Indeed, let us consider the forward-backward system

$$\begin{aligned} &dX_t = B(\mathbb{E}(Y_t))dt + dW_t, \qquad X_0 = x_0,\\ &dY_t = -F(\mathbb{E}(X_t))dt + Z_t dW_t, \qquad Y_T = G(\mathbb{E}(X_T)), \end{aligned} \tag{2.16}$$

where B, F and G are real valued bounded and Lipschitz-continuous functions on the real line. Let us also assume that they coincide with the identity on $[-R,R]$ for R large enough. In other words, we assume that $B(x) = F(x) = G(x) = x$ for $|x| \le R$. For $T = \pi/4$ and for any $A\in\mathbb{R}$, the pair

$$x_t = A\sin(t), \quad y_t = A\cos(t), \quad 0\le t\le T = \frac{\pi}{4},$$

satisfies $\dot{x}_t = y_t$, $\dot{y}_t = -x_t$, for $0\le t\le T$, with $y_T = x_T$ as terminal condition and $x_0 = 0$ as initial condition. Therefore, for $|A|\le R$, $(A\sin(t),A\cos(t))_{0\le t\le T}$ is a solution to the deterministic forward-backward system

$$\begin{aligned} &\dot{x}_t = B(y_t), \quad \dot{y}_t = -F(x_t), \quad 0\le t\le T,\\ &y_T = G(x_T), \end{aligned}$$

with $x_0 = 0$ as initial condition. For such a value of A, set now:

$$X_t = x_t + W_t, \quad Y_t = y_t, \quad 0\le t\le T.$$

Then, $(X,Y,0)$ solves

$$\begin{aligned} &dX_t = B(\mathbb{E}(Y_t))dt + dW_t\\ &dY_t = -F(\mathbb{E}(X_t))dt + 0\, dW_t, \end{aligned}$$

with $X_0 = 0$ and $Y_T = G(\mathbb{E}(X_T))$, so that uniqueness fails.

Remark 2.7. *The reason for the failure of uniqueness can be explained as follows. In the standard framework, as explained in [6], uniqueness holds because of the smoothing effect of the diffusion operator in the spatial direction. However, in the McKean-Vlasov setting, the smoothing effect of the diffusion operator is ineffective in the direction of the measure variable.*

3 Applications

We now describe the set-up of the two applications mentioned in the introduction.

We define the set $\mathbb{A}$ of admissible controls as the set of progressively measurable processes $\underline{\alpha} = (\alpha_t)_{0\le t\le T}$ with values in a measurable space $(A, \mathcal{A})$. Typically, A is a Borel subset of a Euclidean space $\mathbb{R}^k$, and $\mathcal{A}$ the σ-field induced by the Borel σ-field of this Euclidean space. For each admissible control process $\underline{\alpha}$ we consider the solution $\underline{X} = (X_t)_{0\le t\le T}$ of the (nonlinear) stochastic differential equation of McKean-Vlasov type

$$dX_t = b(t, X_t, \mathbb{P}_{X_t}, \alpha_t)dt + \sigma(t, X_t, \mathbb{P}_{X_t}, \alpha_t)dW_t \qquad 0 \le t \le T, \quad X_0 = x_0, \tag{3.1}$$

where the drift b and the volatility σ are deterministic measurable functions $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times A \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times A \to \mathbb{R}^{d\times m}$ satisfying regularity conditions to be specified below. The stochastic optimization problem associated to the optimal control of McKean-Vlasov stochastic dynamics is to minimize the objective function

$$J(\underline{\alpha}) = \mathbb{E}\left\{\int_0^T f(t, X_t, \mathbb{P}_{X_t}, \alpha_t)dt + g(X_T, \mathbb{P}_{X_T})\right\}, \tag{3.2}$$

over a set $\underline{\alpha} \in \mathbb{A}$ where the running cost function f is a real valued deterministic measurable function defined on $[0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times A$ and the terminal cost function g is an $\mathbb{R}^d$–valued deterministic measurable function defined on $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. Though different in its goal, the mean field problem as stated by Lasry and Lions is similar in many ways. The main difference comes from the fact that, because the ultimate goal is to identify approximate Nash equilibriums for large games, the stochastic optimization problem is to minimize the objective function

$$J(\underline{\alpha}) = \mathbb{E}\left\{\int_0^T f(t, X_t, \mu_t, \alpha_t)dt + g(X_T, \mu_T)\right\}, \tag{3.3}$$

over the same set $\mathbb{A}$ of admissible control processes when the measure argument in the dynamics of the state, the running cost function f and the terminal cost function g are fixed – one often say *frozen* to emphasize this difference – and given by a deterministic flow $\underline{\mu} = (\mu_t)_{0\le t\le T}$ of probability measures on $\mathbb{R}^d$. The similarity comes from the fact that, once the optimization for $\underline{\mu}$ fixed is accomplished, this flow of measures is determined in such a way that the marginal distribution of the solution of the stochastic differential equation

$$dX_t = b(t, X_t, \mu_t, \alpha_t)dt + \sigma(t, X_t, \mu_t, \alpha_t)dW_t \qquad 0 \le t \le T, \quad X_0 = x_0, \tag{3.4}$$

is μ_t, in other words in such a way that $\mu_t = \mathbb{P}_{X_t}$ for all $0 \le t \le T$, forcing the optimally controlled state process $\underline{X}$ to actually solve the McKean - Vlasov equation (3.1). If an appropriate stochastic version of the Pontryagin maximum can be proven, the probabilistic approach to these stochastic optimization problems is to introduce adjoint processes $(\underline{Y}, \underline{Z}) = (Y_t, Z_t)_{0\le t\le T}$ solving the *adjoint* BSDE, and to couple these backward equations with the forward equations (3.4) and (3.1) by introducing a control $\underline{\alpha}$ minimizing a specific Hamiltonian function. This coupling of the forward and backward equations creates a FBSDE of the McKean-Vlasov type in which the marginal distributions of the solutions appear in the coefficients, and which needs to be solved in order to find a solution of the stochastic optimization problems. This was the motivation behind the statement of the problem considered in Section (2).

For the sake of completeness we assume that, in both applications, the drift coefficient b and the volatility matrix σ satisfy the following assumptions.

(S1) The functions $(b(t,0,\delta_0,0))_{0\le t\le T}$ and $(\sigma(t,0,\delta_0,0))_{0\le t\le T}$ are continuous;

(S2) $\exists c>0$, $\forall t\in[0,T]$, $\forall \alpha\in A$, $\forall x,x'\in\mathbb{R}^d$, $\forall \mu,\mu'\in\mathcal{P}_2(\mathbb{R}^d)$,

$$|b(t,x,\mu,\alpha)-b(t,x',\mu',\alpha)|+|\sigma(t,x,\mu,\alpha)-\sigma(t,x',\mu',\alpha)|\le c(|x-x'|+W_2(\mu,\mu')).$$

We further restrict the set $\mathbb{A}$ to the set of progressively measurable processes $\underline{\alpha}$ in $\mathbb{H}^{2,k}$.

Together with the Lipschitz assumption (S2), this definition guarantees that, for any $\underline{\alpha}\in\mathbb{A}$, there exists a unique solution $\underline{X}=\underline{X}^{\underline{\alpha}}$ of both (3.4) and (3.1) and that moreover, this solution satisfies

$$\mathbb{E}\big[\sup_{0\le t\le T}|X_t|^q\big]<\infty$$

for every $q\ge 1$. See for example [17, 10] for a proof for the McKean-Vlasov equations.

3.1 Solvability of FBSDE for Mean-Field Games

In this subsection, we apply the abstract existence result of Section (2) to the Mean Field Game (MFG) problem described in the introduction. As in [4], we assume that the volatility function is a constant matrix σ (of size $d\times m$), but here we assume in addition that $\det(\sigma\sigma^\dagger)>0$. Since the stochastic optimization problem is solved after the flow of measures is frozen, for each fixed μ, the Hamiltonian of the system reads:

$$H^\mu(t,x,y,\alpha)=b(t,x,\mu,\alpha)\cdot y+f(t,x,\mu,\alpha),\quad x,y\in\mathbb{R}^d,\ \alpha\in A,\ \mu\in\mathcal{P}_2(\mathbb{R}^k),\tag{3.5}$$

'$\cdot$' standing for the inner product. For each μ, we assume the existence of a function $(t,x,y)\hookrightarrow\hat{\alpha}^\mu(t,x,y)\in A$ which is Lipschitz-continuous with respect to (x,y,μ), uniformly in $t\in[0,T]$, the Lipschitz property holding true with respect to the product distance of the Euclidean distances on $\mathbb{R}^d\times\mathbb{R}^d$ and the 2-Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$ and such that:

$$\hat{\alpha}^\mu(t,x,y)\in\text{argmin}_{\alpha\in\mathbb{R}^d}H^\mu(t,x,y,\alpha),\quad t\in[0,T],\ x,y\in\mathbb{R}^d,\ \mu\in\mathcal{P}_2(\mathbb{R}^d).\tag{3.6}$$

The existence of such a function was proven in [4] under specific assumptions on the drift b and the running cost function f used there. Using a standard version of the stochastic maximum principle, and coupling the forward dynamics and the adjoint BSDE by plugging the optimizer (3.6) lead to the solution, for each frozen flow $\underline{\mu}=(\mu_t)_{0\le t\le T}$ of measures, of a standard FBSDE. Now, if we add the requirement that μ_t should coincide for each t with the marginal distribution of the optimally controlled state, the solution of the MFG stochastic optimization problem reduces to the solution of the FBSDE of McKean - Vlasov type

$$\begin{aligned}&dX_t=b\big(t,X_t,\mathbb{P}_{X_t},\hat{\alpha}(t,X_t,Y_t,\mathbb{P}_{X_t})\big)dt+\sigma dW_t,\\&dY_t=-\partial_x f\big(t,X_t,\mathbb{P}_{X_t},\hat{\alpha}(t,X_t,Y_t,\mathbb{P}_{X_t})\big)dt-\partial_x b\big(t,X_t,\mathbb{P}_{X_t},\hat{\alpha}(t,X_t,Y_t,\mathbb{P}_{X_t})\big)Y_tdt+Z_tdW_t,\end{aligned}\tag{3.7}$$

with initial condition $X_0=x_0$ for a given deterministic point $x_0\in\mathbb{R}^d$, and terminal condition $Y_T=\partial_x g(X_T,\mathbb{P}_{X_T})$. Existence of a solution for this McKean-Vlasov FBSDE follows by applying Theorem (2.1) if we assume that $\partial_x f$ and $\partial_x g$ are bounded and Lipschitz-continuous and set:

$$\begin{aligned} B\big(t, X_t, Y_t, Z_t, \mathbb{P}_{(X_t,Y_t)}\big) &= b(t, X_t, \mathbb{P}_{X_t}, \alpha_t), \\ F\big(t, X_t, Y_t, Z_t, \mathbb{P}_{(X_t,Y_t)}\big) &= \partial_x f\big(t, X_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, X_t, Y_t, \mathbb{P}_{X_t})\big) \\ &\qquad + \partial_x b\big(t, X_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, X_t, Y_t, \mathbb{P}_{X_t})\big)Y_t, \\ G(X_T, \mathbb{P}_{X_T}) &= \partial_x g(X_T, \mathbb{P}_{X_T}). \end{aligned}$$

While the result of the present note provides existence in quite a general set up, [4] also allows for running costs with at most linear growth in x (and provides much more in terms of the identification of approximate Nash equilibriums). However, the drift b needs to be of a very specific *affine form*, namely $b(t,x,y,\alpha,\mu) = b_0(t,\mu) + b_1(t)x + b_2(t)\alpha$ for some deterministic functions b_0, b_1 and b_2, and the running cost function f has to satisfy a strong convexity assumption, the latter allowing σ to be degenerate, providing a more efficient approximation procedure to reduce the result to the bounded case and ensuring the validity of the converse of the stochastic maximum principle.

3.2 Optimal Control of McKean-Vlasov Stochastic Dynamics

Finally, we explain how the existence result of this paper complements the existence result of [3] where a solution of the optimal control of stochastic differential equations of the McKean-Vlasov type is given.

As before, we assume that the volatility function is a constant matrix σ such that $\det(\sigma\sigma^\dagger) > 0$, so the Hamiltonian has the form

$$H(t, x, y, \mu, \alpha) = b(t, x, \mu, \alpha) \cdot y + f(t, x, \mu, \alpha),$$

for $t \in [0,T]$, $x, y \in \mathbb{R}^d$, $\alpha \in A$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Again, we assume the existence of a function $(t,x,y,\mu) \hookrightarrow \hat{\alpha}(t,x,y,\mu) \in A$ which is Lipschitz-continuous with respect to (x,y,μ), uniformly in $t \in [0,T]$, the Lipschitz property holding true with respect to the product distance of the Euclidean distances on $\mathbb{R}^d \times \mathbb{R}^d$ and the 2-Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$ and such that:

$$\hat{\alpha}(t,x,y,\mu) \in \text{argmin}_{\alpha \in \mathbb{R}^d} H(t,x,y,\mu,\alpha), \quad t \in [0,T],\ x,y \in \mathbb{R}^d,\ \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{3.8}$$

The existence of such a function was proven in [3] under specific assumptions on the drift b and the running cost function f used there. However, the major difference with the mean field game problem comes from the form of the adjoint equation which now involves differentiation of the Hamiltonian with respect to the measure parameter. A special form of adjoint equation was introduced, and a new stochastic maximum principle was proven in [3]. Once this new form of adjoint equation is coupled with the forward dynamical equation through the plugged-in optimal control feedback $\hat{\alpha}$ defined in (3.8), the associated McKean-Vlasov FBSDE takes the form

$$\begin{aligned} dX_t &= b\big(t, X_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, X_t, Y_t, \mathbb{P}_{X_t})\big)dt + \sigma dW_t, \\ dY_t &= -\partial_x f\big(t, X_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, X_t, Y_t, \mathbb{P}_{X_t})\big)dt \\ &\qquad - \partial_x b\big(t, X_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, X_t, Y_t, \mathbb{P}_{X_t})Y_t\big)dt + Z_t dW_t \\ &\qquad - \tilde{\mathbb{E}}\big[\partial_\mu f\big(t, \tilde{X}_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, \tilde{X}_t, \tilde{Y}_t, \mathbb{P}_{X_t})\big)(X_t)\big]dt \\ &\qquad - \tilde{\mathbb{E}}\big[\partial_\mu b\big(t, \tilde{X}_t, \mathbb{P}_{X_t}, \hat{\alpha}(t, \tilde{X}_t, \tilde{Y}_t, \mathbb{P}_{X_t})\big)(X_t) \cdot Y_t\big]dt, \end{aligned} \tag{3.9}$$

with initial condition $X_0 = x_0$ for a given deterministic point $x_0 \in \mathbb{R}^d$, and terminal condition $Y_T = \partial_x g(X_T, \mathbb{P}_{X_T}) + \tilde{\mathbb{E}}[\partial_\mu g(\tilde{X}_T, \mathbb{P}_{X_T})(X_T)]$. Above, the tilde refers to an independent copy of (X_t, Y_t, Z_t). The notations $\partial_\mu b(t,x,\mathbb{P}_\xi,\alpha)(\xi)$ and $\partial_\mu f(t,x,\mathbb{P}_\xi,\alpha)(\xi)$, for an $\mathbb{R}^d$–valued random variable ξ of order 2, represent the derivatives of b and f with respect to the measure argument at $(t, x, \mathbb{P}_\xi, \alpha)$ according to the rule of differentiation introduced by P.L. Lions as explained in [2]: the derivatives are represented by random variables, obtained by plugging the current value of ξ into some functions $\partial_\mu b(t, x, \mathbb{P}_\xi, \alpha)(\cdot)$ and $\partial_\mu f(t, x, \mathbb{P}_\xi, \alpha)(\cdot)$. We chose the order in which the various terms appear in the right hand side of the backward equation in order to emphasize the last two terms which are not present in standard forms of adjoint BSDEs. Because of these two last terms, the coefficients depend on the joint law of (X_t, Y_t), which is not the case in (3.7). Despite this non-standard form of the FBSDE, the existence result of this note applies (under appropriate conditions on the coefficients). Under very restrictive conditions on the drift b and a strict convexity assumption on f, [3] provides uniqueness, allows for at most linear growth in the variable x and guarantees the validity of the converse of the maximum principle.

References

[1] A. Bensoussan, K.C.J. Sung, S.C.P. Yam, and S.P. Yung. Linear quadratic mean field games. Technical report, 2011.

[2] P. Cardaliaguet. Notes on mean field games. Technical report, 2010.

[3] R. Carmona and F. Delarue. Optimal control of McKean-Vlasov stochastic dynamics. Technical report, 2012. arXiv:1211.4186

[4] R. Carmona and F. Delarue. Probabilistic analysis of mean field games. *SIAM Journal on Control and Optimization*, 51:2705–2734, 2013. MR-3072222

[5] R. Carmona, F. Delarue, and A. Lachapelle. Control of McKean-Vlasov versus Mean Field Games. *Mathematics and Financial Economics*, 7:131–166, 2013. MR-3045029

[6] F. Delarue. On the existence and uniqueness of solutions to FBSDEs in a non-degenerate case. *Stochastic Processes and Applications*, 99:209–286, 2002. MR-1901154

[7] F. Delarue. Estimates of the solutions of a system of quasi-linear PDEs. a probabilistic scheme. Séminaire de Probabilités XXXVII, pages 290–332. Springer Verlag, 2003. MR-2053051

[8] F. Delarue and G. Guatteri. Weak existence and uniqueness for fbsdes. *Stochastic Processes and Applications*, 116:1712–1742, 2006. MR-2307056

[9] Avner Friedman. *Partial differential equations of parabolic type.* Prentice-Hall, Englewood Cliffs, N.J., first edition, 1964. MR-0181836

[10] B. Jourdain, S. Meleard, and W. Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *ALEA, Latin American Journal of Probability*, 4:1–29, 2008. MR-2383731

[11] M. Kac. Foundations of kinetic theory. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, pages 171 – 197, 1956. MR-0084985

[12] M. Kac. *Probability and Related Topics in the Physical Sciences.* Interscience, 1958. MR-0102849

[13] B. F. Knerr. Parabolic interior schauder estimates by the maximum principle. *Archive for Rational Mechanics and Analysis.*

[14] J.M. Lasry and P.L.Lions. Jeux à champ moyen i. le cas stationnaire. *Comptes Rendus de l'Académie des Sciences de Paris, ser. A*, 343(9), 2006. MR-2269875

[15] J.M. Lasry and P.L.Lions. Jeux à champ moyen ii. horizon fini et contrôle optimal. *Comptes Rendus de l'Académie des Sciences de Paris, ser. A*, 343(10), 2006. MR-2271747

[16] J.M. Lasry and P.L.Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1), Mar. 2007. MR-2295621

[17] A.S. Sznitman. Topics in propagation of chaos. In *D. L. Burkholder et al. , Ecole de Probabilités de Saint Flour, XIX-1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251, 1989. MR-1108185

[18] J. Yong. A linear-quadratic optimal control problem for mean field stochastic differential equations. Technical report, 2011.