

Extended mean field control problem: a propagation of chaos result*

Mao Fabrice Djete[†]

Abstract

In this paper, we study the *extended* mean field control problem, which is a class of McKean–Vlasov stochastic control problem where the state dynamics and the reward functions depend upon the joint (conditional) distribution of the controlled state and the control process. By considering an appropriate controlled Fokker–Planck equation, we can formulate an optimization problem over a space of measure-valued processes and, under suitable assumptions, prove the equivalence between this optimization problem and the *extended* mean-field control problem. Moreover, with the help of this new optimization problem, we establish the associated limit theory i.e. the *extended* mean field control problem is the limit of a large population control problem where the interactions are achieved via the empirical distribution of state and control processes.

Keywords: Mean-Field control; McKean–Vlasov process; law of control; propagation of chaos.

MSC2020 subject classifications: 60-XX; 60Gxx; 60Fxx.

Submitted to EJP on September 14, 2020, final version accepted on November 25, 2021.

1 Introduction

The aim of this paper is to provide a rigorous connection between two stochastic control problems: the stochastic control problem of large population (or particles) interacting through the empirical distribution of their states and controls on the one hand, and the other hand the problem of control of stochastic dynamics depending upon the joint (conditional) distribution of the controlled state and the control, also called *extended* mean field control problem.

To fix the ideas, let us briefly described the problems. The large population stochastic control problem can be formulated as follows (see Section 2.1 for more details). Consider N -interacting controlled state processes $\mathbf{X} := (\mathbf{X}^1, \dots, \mathbf{X}^N)$ governed by the following

*This work benefited from support of the région Île-de-France.

[†]Université Paris–Dauphine, PSL University, CNRS, CEREMADE, 75016 Paris, France. E-mail: djete@ceremade.dauphine.fr

system of stochastic differential equations: for $t \in [0, T]$,

$$d\mathbf{X}_t^i = b(t, \mathbf{X}_t^i, (\varphi_s^{N,\mathbf{X}})_{s \in [0,t]}, \varphi_t^N, \alpha_t^i)dt + \sigma(t, \mathbf{X}_t^i, (\varphi_s^{N,\mathbf{X}})_{s \in [0,t]}, \varphi_t^N, \alpha_t^i)d\mathbf{W}_t^i + \sigma_0 dB_t,$$

$$\varphi_t^N := \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{x}_t^i, \alpha_t^i) \text{ and } \varphi_t^{N,\mathbf{X}} := \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_t^i}.$$

Here $T > 0$ is a fixed time horizon, $(B, \mathbf{W}^1, \dots, \mathbf{W}^N)$ are independent Brownian motions, B is called the common noise and $(\alpha^1, \dots, \alpha^N)$ are some admissible controls chosen by a global planner. In this stochastic control problem, the global planner aims to maximise the average reward value given by

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\int_0^T L(t, \mathbf{X}_t^i, (\varphi_s^{N,\mathbf{X}})_{s \in [0,t]}, \varphi_t^N, \alpha_t^i)dt + g(\mathbf{X}_T^i, (\varphi_s^{N,\mathbf{X}})_{s \in [0,T]}) \right].$$

When N goes to infinity, the expectation is that this problem “converges” towards the *extended* mean field control problem. Loosely speaking (see Section 2.2 for more details), in the *extended* mean field control problem the objective is to control via α the state process X which follows the stochastic differential equation of McKean–Vlasov type

$$dX_t = b(t, X_t, (\mathcal{L}(X_s|B))_{s \in [0,t]}, \mathcal{L}(X_t, \alpha_t|B), \alpha_t)dt + \sigma(t, X_t, (\mathcal{L}(X_s|B))_{s \in [0,t]}, \mathcal{L}(X_t, \alpha_t|B), \alpha_t)dW_t + \sigma_0 dB_t,$$

in order to maximise the quantity

$$\mathbb{E} \left[\int_0^T L(t, X_t, (\mathcal{L}(X_s|B))_{s \in [0,t]}, \mathcal{L}(X_t, \alpha_t|B), \alpha_t)dt + g(X_T, (\mathcal{L}(X_s|B))_{s \in [0,T]}) \right],$$

where $\mathcal{L}(X_t, \alpha_t|B)$ (resp $\mathcal{L}(X_t|B)$) denote the conditional distribution of the couple (X_t, α_t) (resp the state X_t) given the common noise B .

The connection we are investigating, i.e. that the stochastic control problem of large population *converges* towards the mean field control problem, is often called *limit theory* or (controlled) *propagation of chaos*. In contrast with the classical framework of McKean–Vlasov stochastic control problem which only considers the conditional distribution of X_t , here, there is in addition the presence of the conditional distribution of (X_t, α_t) . Indeed, when there is no law of control i.e. no $\mathcal{L}(X_t, \alpha_t|B)$ but only $\mathcal{L}(X_t|B)$ in (b, σ, L, g) , these problems have been studied in the literature. Let us mention the work of Snitzman [33] which shows, for particular coefficients (b, σ) in the absence of control (and the law of control), via some compactness arguments, a connection of this type. See also the papers of Oelschläger [31] and Gärtner [17], with no control and no law of control as well, which use martingale problem in the sense of Stroock and Varadhan [34] adapted in the context of McKean–Vlasov equation to prove similar results under minimal assumptions.

In the controlled dynamic case but no *extended* type, that is to say when the dynamic depends on the control but not its law, Fischer and Livieri [15] get a connection between the large population stochastic control problem and the (*extended*) mean field control problem for the study of a mean–variance problem arising in finance. Another interesting work is that of Budhiraja, Dupuis, and Fischer [4], where they study the behavior of empirical measures of controlled interacting diffusion in order to prove a large deviation principle in a McKean–Vlasov framework. Still without touching the case with law of control, the first papers that deal with the case with control under general assumptions are Lacker [24] and Djete, Possamaï, and Tan [11]. Thanks to an (extension of) martingale problem of [34], as well as relaxed controls initiated by Fleming and Nisio

[16], and developed by El Karoui, Huu Nguyen, and Jeanblanc-Picqué [12], combined with compactness arguments adapted to the McKean–Vlasov setting, [24] proves the connection between the two problems under general conditions on (b, σ, L, g) without common noise. Indeed, the idea of using relaxed controls, i.e. control seen as probability measure of type $\delta_{\alpha_t}(du)dt$ helps to find some compactness properties necessary for proving these types of results. Following upon these ideas, [11] develops a general overview of McKean–Vlasov or mean field control problem, and treats the case with common noise, which turns out to be a non trivial extension.

In the presence of the law of control, this propagation of chaos result is a natural expectation. In spite of appearances, this is not an easy extension. The aforementioned techniques do not work in this context. Two main reasons can explain the unsuitable aspect of the techniques mentioned above. Firstly, the continuity of the application $t \mapsto \mathcal{L}(X_t|B)$ (or $t \mapsto \varphi_t^{N, X}$) plays a crucial role. Indeed, the classical idea is to put this application in a canonical space, which is here the space $C([0, T]; \mathcal{P}(\mathbb{R}^n))$ of continuous functions from $[0, T]$ into the space of probability measures on \mathbb{R}^n , and via compactness arguments and martingale problem get this connection (see [24], and [11] for the non–Markovian case with common noise). In our situation, this type of continuity is lost because we must take into account the application $t \mapsto \mathcal{L}(X_t, \alpha_t|B)$ (or $t \mapsto \varphi_t^N$) which does not have this property since the presence of control α can generate some discontinuities. Secondly, as highlighted in [11], proving a result of *propagation of chaos* is extremely related to the search of the closure of the set of all probabilities that are the image measure of the controlled state process, the control and the conditional distribution of the controlled state process and control, i.e. $\mathcal{L}(X, \delta_{\alpha_t}(du)dt, \mathcal{L}(X, \delta_{\alpha_t}(du)dt|B))$. Unfortunately, the natural space that one might think to answer this question is not a closed set due to another problem of continuity (see Remark 2.5 for a more thorough discussion).

There are not many papers in the literature which study the mean field control problem with law of control and its connection with a large population stochastic control problem. To the best of our knowledge, only the recent papers of Laurière and Tangpi [28] (with strong assumptions) and Motte and Pham [30] (for mean field Markov decision processes) treat the limit theory question. Most papers focus on the questions of existence and uniqueness of optimal control. Acciaio, Backhoff Veraguas, and Carmona [1], with the help of Pontryagin’s maximum principle, obtain necessary and sufficient conditions to characterize the optimum with strong assumptions on the coefficients in a no common noise framework. Pham and Wei [32] (without common noise, with closed loop controls) and Djete, Possamaï, and Tan [10] establish the Dynamic Programming Principle (DPP for short) and give a Hamilton–Jacobi equation on a space of probability measures verified by the value function (heuristically proved in [10]). Let us also mention Carmona and Lacker [6], Elie, Mastrolia, and Possamaï [13], Cardaliaguet and Lehalle [5], Alasseur, Taher, and Matoussi [2], Casgrain and Jaimungal [8], Lacker and Soret [26], Féron, Tankov, and Tinsi [14] and [28] who study similar problem in the mean field game framework called mean field game of controls or *extended* mean field game, as well as our companion paper Djete [9] adapts the arguments of this paper to the context of mean field game of controls.

In this article, our goal is to give some properties on the *extended* mean field control problem and to show its connection with the large population stochastic control problem under general assumptions on (b, σ, L, g) (see Theorem 3.3 and Theorem 3.1). To bypass the difficulties highlighted above, we follow the idea mentioned in [11] which is to introduce a new optimization problem by considering a suitable set of controls. This set must be the closure of some set of probability measures. In this framework, the appropriate space is the closure of all the probabilities that are the distributions of the

conditional distribution of the state controlled process and the conditional distribution of the state controlled process and the control, i.e. $\mathcal{L}(\mathcal{L}(X_t|B))_{t \in [0, T]}, \delta_{\mathcal{L}(X_t, \alpha_t|B)}(dm)dt$ (for more details see Section 2.3). Taking into account this type of probability turns out to be the key to solve the main difficulties. The characterization of its closure is possible by the appropriate use of (controlled) Fokker–Planck equation. Inspired by the techniques developed in the proofs of Gyöngy [18], especially [18, Lemma 2.1] (an adaptation of Krylov [22]) and [18, Proposition 4.3] which are regularization results, we can determine the desired set thanks to a Fokker–Planck equation. The conditions used on the coefficients are general, except the non-degeneracy of the volatility σ . This assumption is capital to prove our main results. Apart from this assumption, our result appears to be one of the first to establish some general properties on *extended* mean field control problem and to show its connection with the large population stochastic problem. Lacker [25] used similar techniques in the context of convergence of closed loop Nash equilibria, but his analysis focuses mainly on an adequate manipulation of [18, Theorem 4.6], while ours focuses on the techniques used for the proofs. Also, let us mention Lacker, Shkolnikov, and Zhang [27] which establish a correspondence between Fokker–Planck equations and solutions of SDE in a McKean–Vlasov framework with common noise.

The rest of the paper is structured as follows. After introducing the notations and the probabilistic structure to give an adequate definition of the tools that are used throughout the paper, Section 2 states all the main assumptions and carefully formulates first the large population stochastic control problem, then the strong formulation of the extended mean field control problem and finally the stochastic control of measure-valued processes. Next, in Section 3, we present the main results of this paper: the equivalence between the strong formulation of extended mean field control problem and the stochastic control of measure-valued processes, and the propagation of chaos result i.e. the extended mean field control problem is, when N goes to infinity, the limit of the large population stochastic control problem in presence of interactions through the empirical distribution of state and control processes. Finally, Section 4 is devoted to the proof of our main results and Section 5 provides some approximation results related to the Fokker–Planck equation needed in our proofs.

Notations

(i) Given a Polish space (E, Δ) , $p \geq 1$, we denote by $\mathcal{P}(E)$ the collection of all Borel probability measures on E , and by $\mathcal{P}_p(E)$ the subset of Borel probability measures μ such that $\int_E \Delta(e, e_0)^p \mu(de) < \infty$ for some $e_0 \in E$. We equip $\mathcal{P}_p(E)$ with the Wasserstein metric \mathcal{W}_p defined by

$$\mathcal{W}_p(\mu, \mu') := \left(\inf_{\lambda \in \Lambda(\mu, \mu')} \int_{E \times E} \Delta(e, e')^p \lambda(de, de') \right)^{1/p},$$

where $\Lambda(\mu, \mu')$ denotes the collection of all probability measures λ on $E \times E$ such that $\lambda(de, E) = \mu(de)$ and $\lambda(E, de') = \mu'(de')$. Equipped with \mathcal{W}_p , $\mathcal{P}_p(E)$ is a Polish space (see [35, Theorem 6.18]). For any $\mu \in \mathcal{P}(E)$ and μ -integrable function $\varphi : E \rightarrow \mathbb{R}$, we write

$$\langle \varphi, \mu \rangle = \langle \mu, \varphi \rangle := \int_E \varphi(e) \mu(de), \tag{1.1}$$

and for another metric space (E', Δ') , we denote by $\mu \otimes \mu' \in \mathcal{P}(E \times E')$ the product probability of any $(\mu, \mu') \in \mathcal{P}(E) \times \mathcal{P}(E')$.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a sub- σ -algebra $\mathcal{G} \subset \mathcal{F}$ then for a Polish space E and any random variable $\xi : \Omega \rightarrow E$, both the notations $\mathcal{L}^\mathbb{P}(\xi|\mathcal{G})(\omega)$ and $\mathbb{P}_\omega^\mathcal{G} \circ (\xi)^{-1}$ are used to denote the conditional distribution of ξ knowing \mathcal{G} under \mathbb{P} .

(ii) For any (E, Δ) and (E', Δ') two Polish spaces, we shall refer to $C_b(E, E')$ to designate the set of continuous functions f from E into E' such that $\sup_{e \in E} \Delta'(f(e), e'_0) < \infty$ for some $e'_0 \in E'$. Let \mathbb{N} be the set of non-negative integers and \mathbb{N}^* be the notation of the set of positive integers, i.e. $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$. Given non-negative integers m and n , we denote by $\mathbb{S}^{m \times n}$ the collection of all $m \times n$ -dimensional matrices with real entries, equipped with the standard Euclidean norm, which we denote by $|\cdot|$ regardless of the dimensions, for notational simplicity. We also denote $\mathbb{S}^n := \mathbb{S}^{n \times n}$, and denote by $0_{m \times n}$ the element in $\mathbb{S}^{m \times n}$ whose entries are all 0, and by I_n the identity matrix in \mathbb{S}^n . For any matrix $a \in \mathbb{S}^n$ which is symmetric positive semi-definite, we write $a^{1/2}$ the unique symmetric positive semi-definite square root of the matrix a . Let k be a positive integer, we denote by $C_b^k(\mathbb{R}^n; \mathbb{R})$ the set of bounded maps $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with bounded continuous derivatives of order up to and including k . Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be twice differentiable, we denote by ∇f and $\nabla^2 f$ the gradient and Hessian of f .

(iii) Let $T > 0$, and (Σ, ρ) be a Polish space, we denote by $C([0, T], \Sigma)$ the space of all continuous functions on $[0, T]$ taking values in Σ . Then $C([0, T], \Sigma)$ is a Polish space under the uniform convergence topology, and we denote by $\|\cdot\|$ the uniform norm. When $\Sigma = \mathbb{R}^k$ for some $k \in \mathbb{N}$, we simply write $C^k := C([0, T], \mathbb{R}^k)$, also we shall denote by $C_{\mathcal{W}}^k := C([0, T], \mathcal{P}(\mathbb{R}^k))$, and for $p \geq 1$, $C_{\mathcal{W}}^{k,p} := C([0, T], \mathcal{P}_p(\mathbb{R}^k))$.

With a Polish space E , we denote by $\mathbb{M}(E)$ the space of all Borel measures $q(dt, de)$ on $[0, T] \times E$, whose marginal distribution on $[0, T]$ is the Lebesgue measure dt , that is to say $q(dt, de) = q(t, de)dt$ for a family $(q(t, de))_{t \in [0, T]}$ of Borel probability measures on E . We also consider the subset $\mathbb{M}_0(E) \subset \mathbb{M}(E)$ which is the collection of all $q \in \mathbb{M}(E)$ such that $q(dt, de) = \delta_{\psi(t)}(de)dt$ for some Borel measurable function $\psi : [0, T] \rightarrow E$. For any $q \in \mathbb{M}(E)$, we define

$$q_{t \wedge} (ds, de) := q(ds, de)|_{[0, t] \times E} + \delta_{e_0}(de)ds|_{(t, T] \times E}, \text{ for some fixed } e_0 \in E. \tag{1.2}$$

2 Extended mean field control problem

Let $(\ell, n) \in \mathbb{N} \times \mathbb{N}^*$, (U, ρ) be a nonempty Polish space and \mathcal{P}_U^n denote the space of all Borel probability measures on $\mathbb{R}^n \times U$ i.e. $\mathcal{P}_U^n := \mathcal{P}(\mathbb{R}^n \times U)$. We give ourselves the following Borel measurable functions

$$[b, \sigma, L] : [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n \times U \rightarrow \mathbb{R}^n \times \mathbb{S}^{n \times n} \times \mathbb{R} \text{ and } g : \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \rightarrow \mathbb{R}.$$

Assumption 2.1. *The functions $[b, \sigma, L]$ are non-anticipative in the sense that, for all $(t, x, \pi, m, u) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n \times U$*

$$[b, \sigma, L](t, x, \pi, m, u) = [b, \sigma, L](t, x, \pi_{t \wedge}, m, u).$$

Moreover, there exist positive constants C and p such that $p \geq 2$ and

(i) U is a compact space;

(ii) b and σ are continuous bounded functions, and $\sigma_0 \in \mathbb{S}^{n \times \ell}$ is constant;

(iii) one has for all $(t, x, x', \pi, \pi', m, m', u) \in [0, T] \times (\mathbb{R}^n)^2 \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U$

$$|[b, \sigma](t, x, \pi, m, u) - [b, \sigma](t, x', \pi', m', u)| \leq C(|x - x'| + \sup_{s \in [0, T]} \mathcal{W}_p(\pi_s, \pi'_s) + \mathcal{W}_p(m, m'));$$

(iv) for some constant $\theta > 0$, one has, for all $(t, x, \pi, m, u) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n \times U$,

$$\theta I_n \leq \sigma \sigma^\top(t, x, \pi, m, u);$$

(v) the reward functions L and g are continuous, and for all $(t, x, \pi, m, u) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n \times U$, one has

$$|L(t, x, \pi, m, u)| + |g(x, \pi)| \leq C \left[1 + |x|^p + \sup_{s \in [0, T]} \mathcal{W}_p(\pi_s, \delta_0)^p + \int_{\mathbb{R}^n} |x'|^p m(dx', U) \right].$$

Remark 2.2. These assumptions are standard and in the same spirit as those used in [24] and [11], but with some specific modifications adapted to the context of this article. They ensure the well-posedness of the objects used throughout this paper. Due to the technical aspect of our paper, the point (i) is considered essentially to simplify (the presentation of) the proofs. But, using the classical uniform integrability condition as in [24] and [11], it is possible to work with U a non-bounded set of \mathbb{R}^n for instance. The point (iv) is the least classical assumption in the study in this problem. This is an important assumption for the proofs of our results, in particular to deal with the Fokker–Planck equations and the different SDEs considered in the proofs (see Section 5).

2.1 The large population stochastic control problem

In this section, we present the N -agent stochastic control problem or large population control problem. The study of this control problem when N goes to infinity is one of the main objective of this paper.

For a fixed $(\nu^1, \dots, \nu^N) \in \mathcal{P}_p(\mathbb{R}^n)^N$, let

$$\Omega^N := (\mathbb{R}^n)^N \times (\mathcal{C}^n)^N \times \mathcal{C}^\ell$$

be the canonical space, with canonical variable $\mathbf{X}_0 = (\mathbf{X}_0^1, \dots, \mathbf{X}_0^N)$, canonical processes $\mathbf{W} = (\mathbf{W}_s^1, \dots, \mathbf{W}_s^N)_{0 \leq s \leq T}$ and $B = (B_s)_{0 \leq s \leq T}$, and probability measure \mathbb{P}_ν^N under which $\mathbf{X}_0 \sim \nu_N := \nu^1 \otimes \dots \otimes \nu^N$ and (\mathbf{W}, B) are standard Brownian motions independent of \mathbf{X}_0 . Let $\mathbb{F}^N = (\mathcal{F}_s^N)_{0 \leq s \leq T}$ be defined by

$$\mathcal{F}_s^N := \sigma\{\mathbf{X}_0, \mathbf{W}_r, B_r, r \in [0, s]\}, s \in [0, T].$$

Let us denote by $\mathcal{A}_N(\nu_N)$ the collection of all U -valued \mathbb{F}^N -predictable processes. Then, given $\alpha := (\alpha^1, \dots, \alpha^N) \in (\mathcal{A}_N(\nu_N))^N$, denote by $\mathbf{X}^\alpha := (\mathbf{X}^{\alpha,1}, \dots, \mathbf{X}^{\alpha,N})$ the unique strong solution of the following system of SDEs, $\mathbb{E}^{\mathbb{P}_\nu^N} [\|\mathbf{X}^{\alpha,i}\|^p] < \infty$, for each $i \in \{1, \dots, N\}$,

$$\mathbf{X}_t^{\alpha,i} = \mathbf{X}_0^i + \int_0^t b(r, \mathbf{X}_r^{\alpha,i}, \varphi_{r\wedge\cdot}^{N,\mathbf{X}}, \varphi_r^N, \alpha_r^i) dr + \int_0^t \sigma(r, \mathbf{X}_r^{\alpha,i}, \varphi_{r\wedge\cdot}^{N,\mathbf{X}}, \varphi_r^N, \alpha_r^i) d\mathbf{W}_r^i + \sigma_0 B_t, \tag{2.1}$$

for all $t \in [0, T]$, with

$$\varphi_t^{N,\mathbf{X}}(dx) := \frac{1}{N} \sum_{i=1}^N \delta_{(\mathbf{X}_t^{\alpha,i})}(dx) \text{ and } \varphi_t^N(dx, du) := \frac{1}{N} \sum_{i=1}^N \delta_{(\mathbf{X}_t^{\alpha,i}, \alpha_t^i)}(dx, du).$$

The value function $V_S^N(\nu^1, \dots, \nu^N)$ is defined by

$$V_S^N(\nu^1, \dots, \nu^N) := \sup_{(\alpha^1, \dots, \alpha^N)} J^N(\alpha) \tag{2.2}$$

where

$$J^N(\alpha) := \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}_\nu^N} \left[\int_0^T L(t, \mathbf{X}_t^{\alpha,i}, \varphi_{t\wedge\cdot}^{N,\mathbf{X}}, \varphi_t^N, \alpha_t^i) dt + g(\mathbf{X}_T^{\alpha,i}, \varphi_{T\wedge\cdot}^{N,\mathbf{X}}) \right],$$

which is well-posed under Theorem 2.1.

Remark 2.3. (i) Our formulation allows for coefficients depending on the path of the empirical distribution of \mathbf{X}^α , but can only accommodate a Markovian dependence with respect to \mathbf{X}^α itself. In some sense, we work on a non-Markovian framework w.r.t. the empirical distribution of \mathbf{X}^α . Indeed, as we will see in Section 2.3, our point of view is to

write the entire problem as an optimization involving mainly the empirical distribution of \mathbf{X}^α i.e. $\varphi^{N,\mathbf{X}}$. Therefore our key variable is $\varphi^{N,\mathbf{X}}$ (not \mathbf{X}^α) and we can deal with its path, hence the non-Markovian aspect.

(ii) Sometimes, the probability on $\mathcal{C}_W^n \times \mathbb{M}(\mathcal{P}_U^n) \times \mathcal{C}^\ell$

$$\mathbb{P}(\alpha^1, \dots, \alpha^N) := \mathbb{P}_\nu^N \circ \left((\varphi_t^{N,\mathbf{X}})_{t \in [0,T]}, \delta_{(\varphi_s^N)}(dm)ds, (B_t)_{t \in [0,T]} \right)^{-1} \quad (2.3)$$

will be used to refer to $(\alpha^1, \dots, \alpha^N) \in (\mathcal{A}_N(\nu_N))^N$. The notation $\mathcal{P}_S^N(\nu^1, \dots, \nu^N)$ will designate all probabilities of this type. The need for this space will become clearer in the following.

2.2 The extended mean field control problem

On a fixed probability space, we formulate the classical McKean-Vlasov control problem with common noise including the (conditional) law of control.

For a fixed $\nu \in \mathcal{P}_p(\mathbb{R}^n)$, let

$$\Omega := \mathbb{R}^n \times \mathcal{C}^n \times \mathcal{C}^\ell$$

be the canonical space, with canonical variable ξ , canonical processes $W = (W_t)_{0 \leq t \leq T}$ and $B = (B_t)_{0 \leq t \leq T}$, and probability measure \mathbb{P}_ν under which $\xi \sim \nu$ and (W, B) are standard Brownian motions independent of ξ . Let $\mathbb{F} = (\mathcal{F}_s)_{0 \leq s \leq T}$ and $\mathbb{G} = (\mathcal{G}_s)_{0 \leq s \leq T}$ be defined by: for all $s \in [0, T]$,

$$\mathcal{F}_s := \sigma\{\xi, W_r, B_r, r \in [0, s]\} \quad \text{and} \quad \mathcal{G}_s := \sigma\{B_r, r \in [0, s]\}.$$

Let us denote by $\mathcal{A}(\nu)$ the collection of all U -valued processes $\alpha = (\alpha_s)_{0 \leq s \leq T}$ which are \mathbb{F} -predictable. Then, given $\alpha \in \mathcal{A}(\nu)$, let X^α be the unique strong solution of the SDE (see [10, Theorem A.3]): $\mathbb{E}^{\mathbb{P}_\nu}[\|X^\alpha\|^p] < \infty$, $X_0^\alpha = \xi$, and for $t \in [0, T]$,

$$X_t^\alpha = X_0^\alpha + \int_0^t b(r, X_r^\alpha, \mu_{r\wedge\cdot}^\alpha, \bar{\mu}_r^\alpha, \alpha_r)dr + \int_0^t \sigma(r, X_r^\alpha, \mu_{r\wedge\cdot}^\alpha, \bar{\mu}_r^\alpha, \alpha_r)dW_r + \sigma_0 B_t, \quad (2.4)$$

with $\mu_r^\alpha := \mathcal{L}^{\mathbb{P}_\nu}(X_r^\alpha | \mathcal{G}_r)$ and $\bar{\mu}_r^\alpha := \mathcal{L}^{\mathbb{P}_\nu}(X_r^\alpha, \alpha_r | \mathcal{G}_r)$, for all $r \in [0, T]$.

Let us now introduce the following McKean-Vlasov control problem by

$$V_S(\nu) := \sup_{\alpha \in \mathcal{A}(\nu)} \Phi(\alpha) \quad \text{where} \quad \Phi(\alpha) := \mathbb{E}^{\mathbb{P}_\nu} \left[\int_0^T L(t, X_t^\alpha, \mu_{t\wedge\cdot}^\alpha, \bar{\mu}_t^\alpha, \alpha_t)dt + g(X_T^\alpha, \mu^\alpha) \right]. \quad (2.5)$$

Remark 2.4. Similarly to [11], notice that, this formulation takes into account the case without common noise. Indeed, when $\ell = 0$, the space \mathcal{C}^ℓ and $\mathbb{S}^{n \times \ell}$ degenerate and become $\{0\}$. Then, $B = 0$ and, the filtration \mathbb{G} is constant equal to the trivial σ -algebra $\{\emptyset, \Omega\}$. Therefore, there is no conditional distribution anymore.

Remark 2.5 (Discussion on a possible relaxed extended mean field control problem). An adequate way to study the properties of V_S and/or to give a limit theory is to find the closure $\bar{\mathcal{S}}(\nu)$ of some particular space $\mathcal{S}(\nu)$ for the Wasserstein topology. To simplify, let us take $\ell = 0$ (without common noise), according to the classical ideas of relaxed controls, $\mathcal{S}(\nu) := \{\mathbb{P}_\nu \circ (X^\alpha, \delta_{\alpha_t}(du)dt)^{-1}, \alpha \in \mathcal{A}(\nu)\}$ (see discussion Djete, Possamai, and Tan [11] and also Lacker [24]).

Following [24] and [11], let us give an example to see why the “natural” expected relaxed controls is not a “good” set. Let $n = 1$, $U = [1, 2]$, $\nu = \delta_0$, $\sigma(t, x, \pi, m, u) := |\int_U u' m(\mathbb{R}^n, du')|$ and $b = 0$. Notice that $\mathcal{S}(\nu) \subset \mathcal{P}(\mathcal{C}^n \times \mathbb{M}(U))$, then the canonical space

is $\bar{\Omega}_R := \mathcal{C}^n \times \mathbb{M}(U)$. Denote $(X, \Lambda_t(du)dt)$ the canonical process and $\bar{\mathbb{F}} := (\bar{\mathcal{F}}_t)_{t \in [0, T]}$ the canonical filtration. A naive relaxed controls is $\mathcal{P}_R(\nu) \subset \mathcal{P}(\mathcal{C}^n \times \mathbb{M}(U))$ defined by

$$\mathcal{P}_R(\nu) := \left\{ \bar{\mathbb{P}} : \bar{\mathbb{P}}(X_0 = 0) = 1, (M_t^{\bar{\mathbb{P}}, f})_{t \in [0, T]} \text{ is a } (\bar{\mathbb{P}}, \bar{\mathbb{F}})\text{-martingale } \forall f \in C_b^2(\mathbb{R}) \right\},$$

where $M_t^{\bar{\mathbb{P}}, f} := f(X_t) - \frac{1}{2} \int_0^t \nabla^2 f(X_s) \mathbb{E}^{\bar{\mathbb{P}}} \left[\int_U u \Lambda_s(du) \right]^2 ds$.

But, $\mathcal{P}_R(\nu)$ defined in this way is not a closed set. Indeed the map $q \in \mathbb{M}(U) \rightarrow q_t \in \mathcal{P}(U)$ is not continuous for the Wasserstein topology. Therefore $\mathcal{P}_R(\nu)$ can not be the closure of $\mathcal{S}(\nu)$. More generally, as long as the coefficients (b, σ) are non-linear w.r.t m , this kind discontinuity will appear. Due to this type of lack of continuity, this approach cannot work. We need then to change the framework.

2.3 Stochastic control of measure-valued processes

As previously mentioned, the classical approach of relaxed controls is not appropriate. To bypass the difficulty generated by the (conditional) distribution of control in this study, especially to prove the limit theory result or (controlled) propagation of chaos, we introduce a new stochastic control problem. Motivated by the Fokker–Planck equation verified by the couple $(\mu^\alpha, \bar{\mu}^\alpha)$ from (2.4), we give in this part an equivalent formulation of the extended mean field control problem which is less “rigid”.

2.3.1 Measure-valued rules

Recall that $\mathbb{M} := \mathbb{M}(\mathcal{P}_U^n)$ denotes the collection of all finite (Borel) measures $q(dt, dm)$ on $[0, T] \times \mathcal{P}_U^n$, whose marginal distribution on $[0, T]$ is the Lebesgue measure ds , i.e. $q(ds, dm) = q(s, dm)ds$ for a measurable family $(q(s, dm))_{s \in [0, T]}$ of Borel probability measures on \mathcal{P}_U^n . Let Λ be the canonical element on \mathbb{M} . We then introduce a canonical filtration $\mathbb{F}^\Lambda = (\mathcal{F}_t^\Lambda)_{0 \leq t \leq T}$ on \mathbb{M} by

$$\mathcal{F}_t^\Lambda := \sigma \{ \Lambda(C \times [0, s]) : \forall s \leq t, C \in \mathcal{B}(\mathcal{P}_U^n) \}.$$

For each $q \in \mathbb{M}$, one has a disintegration property: $q(dt, dm) = q(t, dm)dt$, and there is a version of disintegration such that $(t, q) \mapsto q(t, dm)$ is \mathbb{F}^Λ -predictable.

We denote by (μ, Λ, B) the canonical element on $\bar{\Omega} := \mathcal{C}_W^n \times \mathbb{M} \times \mathcal{C}^\ell$. The canonical filtration $\bar{\mathbb{F}} = (\bar{\mathcal{F}}_t)_{t \in [0, T]}$ is then defined by: for all $t \in [0, T]$

$$\bar{\mathcal{F}}_t := \sigma \{ \mu_{t \wedge \cdot}, \Lambda_{t \wedge \cdot}, B_{t \wedge \cdot} \},$$

where $\Lambda_{t \wedge \cdot}$ denotes the restriction of Λ on $\mathcal{P}_U^n \times [0, t]$ (see notation 1.2). Notice that, we can choose a version of disintegration $\Lambda(dm, dt) = \Lambda_t(dm)dt$ with $(\Lambda_t)_{t \in [0, T]}$ a $\mathcal{P}(\mathcal{P}_U^n)$ -valued $\bar{\mathbb{F}}$ -predictable process.

Let us consider \mathcal{L} the following generator: for all $(t, x, \pi, m, u) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}_W^n \times \mathcal{P}_U^n \times U$ and any $\varphi \in C^2(\mathbb{R}^n)$

$$\mathcal{L}_t \varphi(x, \pi, m, u) := \frac{1}{2} \text{Tr} [\sigma \sigma^\top(t, x, \pi, m, u) \nabla^2 \varphi(x)] + b(t, x, \pi, m, u)^\top \nabla \varphi(x),$$

also we introduce, for every $f \in C^2(\mathbb{R}^n)$, $N_t(f)$:

$$\begin{aligned} N_t(f) &:= \langle f(\cdot - \sigma_0 B_t), \mu_t \rangle - \langle f, \mu_0 \rangle \\ &\quad - \int_0^t \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} \mathcal{L}_r [f(\cdot - \sigma_0 B_r)](x, \mu, m, u) m(dx, du) \Lambda_r(dm) dr, \end{aligned} \tag{2.6}$$

recall that $\langle \cdot, \cdot \rangle$ is defined in (1.1). Notice that, under Theorem 2.1, the integral in the definition $N(f)$ is well-posedness. For each $\pi \in \mathcal{P}(\mathbb{R}^n)$, one considers the Borel set \mathcal{Z}_π which is the set of probability measures m on $\mathbb{R}^n \times U$ with marginal on \mathbb{R}^n equal to π i.e.

$$\mathcal{Z}_\pi := \left\{ m \in \mathcal{P}_U^n : m(dx, U) = \pi(dx) \right\}.$$

Definition 2.6. For every $\nu \in \mathcal{P}(\mathbb{R}^n)$, $P \in \mathcal{P}(\overline{\Omega})$ is a measure-valued rule if:

- $P(\mu_0 = \nu) = 1$.
- $(B_t)_{t \in [0, T]}$ is a $(P, \overline{\mathbb{F}})$ Wiener process starting at zero and for P -almost every $\omega \in \overline{\Omega}$, $N_t(f) = 0$ for all $f \in C_b^2(\mathbb{R}^n)$ and every $t \in [0, T]$.
- For $dP \otimes dt$ almost every $(t, \omega) \in [0, T] \times \Omega$, $\Lambda_t(\mathcal{Z}_{\mu_t}) = 1$.

We shall denote by $\overline{\mathcal{P}}_V(\nu)$ the set of all measure-valued rules with initial value ν .

2.3.2 Optimization problem

Let us define, for all $(\pi, q) \in \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}(\mathcal{P}_U^n)$,

$$J(\pi, q) := \int_0^T \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} L(t, x, \pi, m, u) m(dx, du) q_t(dm) dt + \int_{\mathbb{R}^n} g(x, \pi) \pi_T(dx).$$

Notice that, under Theorem 2.1, the map $J : \mathcal{C}_{\mathcal{W}}^{n,p} \times \mathbb{M}_p(\mathcal{P}_U^n) \rightarrow \mathbb{R}$ is continuous (see for instance Theorem A.4). We can now define the measure-valued control problem: for each $\nu \in \mathcal{P}(\mathbb{R}^n)$,

$$V_V(\nu) := \sup_{P \in \overline{\mathcal{P}}_V(\nu)} \mathbb{E}^P [J(\mu, \Lambda)]. \tag{2.7}$$

Remark 2.7. (i) Theorem 2.6 is partly inspired by the Fokker-Planck equation verified by $(\mu_t^\alpha, \overline{\mu}_t^\alpha)_{t \in [0, T]}$ (see (2.4) and Theorem 2.9), in particular the last two points characterize this Fokker-Planck aspect. Indeed, (μ, Λ) satisfy: for all (t, f)

$$\begin{aligned} & \langle f(\cdot - \sigma_0 B_t), \mu_t \rangle \\ &= \langle f, \mu_0 \rangle + \int_0^t \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} \mathcal{L}_r[f(\cdot - \sigma_0 B_r)](x, \mu, m, u) m^x(du) \mu_r(dx) \Lambda_r(dm) dr, \end{aligned}$$

where for each $m \in \mathcal{P}_U^n$, the Borel measurable function $\mathbb{R}^n \ni x \rightarrow m^x \in \mathcal{P}(U)$ verifies $m^x(du) m(dx, U) = m(dx, du)$. This kind of control turns out to be less “rigid”. Especially, $\overline{\mathcal{P}}_V(\nu)$ is a compact set for the Wasserstein topology (see Theorem 3.1).

(ii) Working with these variables seems to be the key to better understand the problem and solves the principal difficulties. Mainly, to prove a limit theory result in this context, we make an approximation of the distribution of (μ, Λ) thanks to the distribution of variables of type $(\mu^\alpha, \delta_{\overline{\mu}_t^\alpha}(dm)dt)$ and not thanks to the approximation of the law of X . This approximation is achieved by using Fokker-Planck equations. To the best of our knowledge, looking at this kind of variable or “control” has never been studied in the literature (except in [11], only for technical reasons).

SDE formulation of measure-valued rules Instead of presenting what we call measure-valued rules as solutions of Fokker-Planck equation, it is possible to formulate the measure-valued rules through solution of SDEs. Indeed, using an equivalence between Fokker-Planck equations and SDEs, there is an alternative way to formulate the measure-valued rules. In order to give more insights about the measure-valued

rules, let us describe the SDEs formulation. For this purpose, we introduce the notion of extended relaxed control rules. We say that the tuple

$$(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, B, X, \mu, \Lambda)$$

is an extended relaxed control rule if

- (i) $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a filtered probability space. on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, (W, B) is a $\mathbb{R}^n \times \mathbb{R}^\ell$ -valued \mathbb{F} -Brownian motion, (X, μ) is a $\mathbb{R}^n \times \mathcal{P}(\mathbb{R}^n)$ -valued \mathbb{F} -adapted continuous process and Λ is a $\mathcal{P}(\mathcal{P}_U^n)$ -valued \mathbb{F} -predictable process.
- (ii) X_0, W and (μ, Λ, B) are independent.
- (iii) The process μ verifies $\mu_t = \mathcal{L}^{\mathbb{P}}(X_t | \mu_{t\wedge \cdot}, \Lambda_{t\wedge \cdot}, B_{t\wedge \cdot}) = \mathcal{L}^{\mathbb{P}}(X_t | \mu, \Lambda, B)$ for all $t \in [0, T]$. The process Λ is s.t. $\Lambda_t(\mathbb{Z}_{\mu_t}) = 1 \, d\mathbb{P} \otimes dt$ a.e. and the process X is solution of: $\mathcal{L}^{\mathbb{P}}(X_0) = \nu$ and

$$dX_t = \int_{\mathcal{P}_U^n} \int_U b(t, X_t, \mu, m, u) m^{X_t}(du) \Lambda_t(dm) dt + \left(\int_{\mathcal{P}_U^n} \int_U \sigma \sigma^\top(t, X_t, \mu, m, u) m^{X_t}(du) \Lambda_t(dm) \right)^{1/2} dW_t + \sigma_0 dB_t,$$

where for each $m \in \mathcal{P}_U^n$, the Borel measurable function $\mathbb{R}^n \ni x \rightarrow m^x \in \mathcal{P}(U)$ verifies $m^x(du)m(dx, U) = m(dx, du)$.

Using [27, Theorem 1.3.] or an easy adaptation of Theorem 5.8 or Theorem 5.9, we have the following equivalence result.

Proposition 2.8. (i) For any extended relaxed control rule $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, B, X, \mu, \Lambda)$, $\mathbb{P} \circ (\mu, \Lambda, B)^{-1}$ belongs to $\overline{\mathcal{P}}_V(\nu)$.

(ii) Conversely, for any $\mathbb{P} \in \overline{\mathcal{P}}_V(\nu)$ measure-valued rule, there exists an extended relaxed control rule $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, W, B, X, \mu, \Lambda)$ s.t.

$$\mathbb{P} = \mathbb{P} \circ (\mu, \Lambda, B)^{-1}.$$

As stated in the preamble of this part, the measure-valued control problem is motivated by the Fokker-Planck equation verified by the couple $(\mu^\alpha, \bar{\mu}^\alpha)$ of the strong formulation. Therefore, the strong controls i.e. $(\mu^\alpha, \bar{\mu}^\alpha)_{\alpha \in \mathcal{A}(\nu)}$ can be seen as a special case of measure-valued rules. By taking into account the previous equivalence Proposition or by applying Itô's formula, it is straightforward to deduce the following proposition.

Proposition 2.9. For each $\nu \in \mathcal{P}_p(\mathbb{R}^n)$, let us introduce

$$\overline{\mathcal{P}}_S(\nu) := \left\{ \mathbb{P}_\nu \circ \left((\mu_t^\alpha)_{t \in [0, T]}, \delta_{\bar{\mu}_t^\alpha}(dm) dr, (B_t)_{t \in [0, T]} \right)^{-1}, \alpha \in \mathcal{A}(\nu) \right\}.$$

one has $\overline{\mathcal{P}}_S(\nu) \subset \overline{\mathcal{P}}_V(\nu)$ and

$$V_S(\nu) = \sup_{\mathbb{Q} \in \overline{\mathcal{P}}_S(\nu)} \mathbb{E}^{\mathbb{Q}}[J(\mu, \Lambda)].$$

Proof. Let $f \in C^2(\mathbb{R}^n)$ and $t \in [0, T]$, denote by $N_t(\mu, \Lambda, B)(f) := N_t(f)$. For any $\alpha \in \mathcal{A}(\nu)$, it is obvious that $\mathbb{P}_\nu(\mu_0^\alpha = \nu) = 1$ and $\delta_{\bar{\mu}_t^\alpha}(\mathbb{Z}_{\mu_t^\alpha}) = 1 \, d\mathbb{P}_\nu \otimes dt$ a.e.. After applying Itô's formula with the process $X^\alpha - \sigma_0 B$, and taking the conditional expectation w.r.t. the σ -field \mathcal{G}_T , one has $N_t(\mu^\alpha, \delta_{\bar{\mu}_t^\alpha}(dm) dt, B)(f) = 0$, \mathbb{P}_ν -a.e. for all (t, f) . Then $\mathbb{P}_\nu \circ (\mu^\alpha, \delta_{\bar{\mu}_t^\alpha}(dm) dt, B)^{-1} \in \overline{\mathcal{P}}_V(\nu)$. Therefore $\overline{\mathcal{P}}_S(\nu) \subset \overline{\mathcal{P}}_V(\nu)$. In addition, notice that

$$\Phi(\alpha) = \mathbb{E}^{\mathbb{P}_\nu} \left[\int_0^T \int_{\mathcal{P}_U^n} \langle L(t, \cdot, \mu_{t\wedge \cdot}^\alpha, m, \cdot), m \rangle \delta_{\bar{\mu}_t^\alpha}(dm) dt + \langle g(\cdot, \mu^\alpha), \mu_T^\alpha \rangle \right],$$

consequently $V_S(\nu) = \sup_{\mathbb{Q} \in \overline{\mathcal{P}}_S(\nu)} \mathbb{E}^{\mathbb{Q}}[J(\mu, \Lambda)]$. □

3 Main results

Now, we formulate the main results of this paper.

Theorem 3.1 (Equivalence). *Let Theorem 2.1 hold true and $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, with $p' > p$. Then $\overline{\mathcal{P}}_V(\nu)$ is convex and compact for the Wasserstein metric \mathcal{W}_p . Moreover*

(i) *When $\ell \neq 0$, for \mathcal{W}_p , the set $\overline{\mathcal{P}}_S(\nu)$ is dense in $\overline{\mathcal{P}}_V(\nu)$.*

(ii) *When $\ell = 0$, for any $P \in \overline{\mathcal{P}}_V(\nu)$, there exists a family $(P_z^k)_{(k,z) \in \mathbb{N}^* \times [0,1]} \subset \overline{\mathcal{P}}_S(\nu)$ such that for each $k \in \mathbb{N}^*$, $[0,1] \ni z \rightarrow P_z^k \in \mathcal{P}(\Omega)$ is Borel measurable and one gets*

$$\lim_{k \rightarrow \infty} \mathcal{W}_p \left(\int_0^1 P_z^k dz, P \right) = 0.$$

Consequently

$$V_V(\nu) = V_S(\nu),$$

and there exists $P^* \in \overline{\mathcal{P}}_V(\nu)$ such that $V_S(\nu) = \mathbb{E}^{P^*} [J(\mu, \Lambda)]$.

Remark 3.2. (i) As in [11] (see also [23] and [9] for the mean field game context), there are some specificities when $\ell = 0$. Indeed, when $\ell = 0$, $(\mu^\alpha, \bar{\mu}^\alpha)$ are deterministic, but (μ, Λ) can still be random, therefore, except in particular situation, it is not possible to approximate the non atomic measure P by a sequence of atomic measure of type $\delta_{(\mu^\alpha, \delta_{\bar{\mu}^\alpha}(\text{dm})\text{ds})}$. However, a randomisation is possible as mentioned in (ii) of Theorem 3.1.

(ii) Theorem 3.1 and the following Theorem 3.3 are in the same spirit that Theorem 3.1 and Theorem 3.6 of [11]. The main difference is the presence of the distribution of controlled state and control, and this particularity turns out to be a non trivial extension (see discussion in Section 2.2).

Theorem 3.3 (Propagation of chaos). *Let Theorem 2.1 hold true, $p' > p$ and $(\nu^i)_{i \in \mathbb{N}^*} \subset \mathcal{P}_{p'}(\mathbb{R}^n)$ satisfying $\sup_{N \geq 1} \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^n} |x'|^{p'} \nu^i(dx') < \infty$. Then*

$$\lim_{N \rightarrow \infty} \left| V_S^N(\nu^1, \dots, \nu^N) - V_S \left(\frac{1}{N} \sum_{i=1}^N \nu^i \right) \right| = 0.$$

Finally, we provide some properties of optimal control of our problem. For any $\nu \in \mathcal{P}(\mathbb{R}^n)$, denote by $\overline{\mathcal{P}}_V^*(\nu)$ the set of optimal control i.e. $P^* \in \overline{\mathcal{P}}_V^*(\nu)$ if $P^* \in \overline{\mathcal{P}}_V(\nu)$ and $V_V(\nu) = \mathbb{E}^{P^*} [J(\mu, \Lambda)]$.

Proposition 3.4. *Suppose that the conditions of Theorem 3.3 hold. For some $\nu \in \mathcal{P}_p(\mathbb{R}^n)$, let $\lim_{N \rightarrow \infty} \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \nu^i, \nu \right) = 0$.*

(i) *For any sequence of non negative numbers $(\varepsilon_N)_{N \in \mathbb{N}^*}$ verifying $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, if $(P^N)_{N \in \mathbb{N}^*}$ is the sequence satisfying $P^N := P(\alpha^1, \dots, \alpha^N)$ (see (2.3)) with*

$$\text{for each } N \in \mathbb{N}^*, \alpha^i \in \mathcal{A}_N(\nu_N) \forall i \in \llbracket 1, N \rrbracket \text{ and } V_S^N(\nu^1, \dots, \nu^N) - \varepsilon_N \leq \mathbb{E}^{P^N} [J(\mu, \Lambda)], \tag{3.1}$$

then

$$\lim_{N \rightarrow \infty} \inf_{P^* \in \overline{\mathcal{P}}_V^*(\nu)} \mathcal{W}_p(P^N, P^*) = 0.$$

(ii) *Moreover, for each $P^* \in \overline{\mathcal{P}}_V^*(\nu)$, there exist $(\varepsilon_N)_{N \in \mathbb{N}^*} \subset (0, \infty)$ verifying $\lim_{N \rightarrow \infty} \varepsilon_N = 0$ and a sequence $(P^{*,N})_{N \in \mathbb{N}^*}$ satisfying $P^{*,N} := P(\alpha^{*,1}, \dots, \alpha^{*,N})$ and condition 3.1 s.t. $\lim_{N \rightarrow \infty} \mathcal{W}_p(P^{*,N}, P^*) = 0$.*

Remark 3.5. (i) The previous proposition shows that any ε_N -optimal control of the large population stochastic control problem converges towards an optimal control of the McKean–Vlasov stochastic control problem in distribution sense. In particular when there exists a unique strong optimal control of the McKean–Vlasov control problem,

any ε_N -optimal control of the large population control problem converges towards this control.

(ii) To the best of our knowledge, Theorem 3.3 and Theorem 3.4 seem to be the first result under these general assumptions to provide these types of convergence results. As mentioned in the introduction, other authors treat these questions but in a particular framework. For instance, while dealing with the convergence of Nash equilibria, [28] gives a limit theory result for the extended mean field control problem. The framework of [28] is less general than ours, in particular, they consider a situation without common noise ($\sigma_0 = 0$), with volatility σ constant. Besides, they need assumptions over (b, g, L) via the Hamiltonian which lead to the uniqueness of the optimum and, these assumptions are sometimes quite difficult to verify in practice. However, it should be mentioned that the results of [28] include a rate of convergence that we do not provide. Let us also mention [30] which treats these questions of convergence but for Markov decision processes in discrete time.

The next corollary is just a combination of Theorem 3.3 and [11, Proposition 4.15]. It states that if a strong control is close enough to the optimum value of the mean field control problem, from this control, we can construct N agents which are close to the optimum of the large population stochastic control problem.

Corollary 3.6. *Let Theorem 2.1 hold true. Let $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, with $p' > p$, $(\varepsilon_N)_{N \in \mathbb{N}^*}$ be a sequence of non negative real such that $\lim_{N \rightarrow \infty} \varepsilon_N = 0$. Also, for each $N \in \mathbb{N}^*$, let $\alpha^N \in \mathcal{A}(\nu)$ satisfying $\alpha_t^N = \phi^N(t, \xi, W_{t \wedge \cdot}, B_{t \wedge \cdot}) \mathbb{P}_\nu$ a.e. for all $t \in [0, T]$ with a Borel function $\phi^N : [0, T] \times \mathbb{R}^n \times \mathcal{C}^d \times \mathcal{C}^\ell \rightarrow U$, and*

$$V_S(\nu) - \varepsilon_N \leq \Phi(\alpha^N).$$

Then, there exists $(\delta_N)_{N \in \mathbb{N}^} \subset (0, \infty)$ s.t. $\lim_{N \rightarrow \infty} \delta_N = 0$ and $(\alpha^{1,N}, \dots, \alpha^{N,N}) \in \mathcal{A}_N(\nu_N)^N$ with $\nu_N := \nu \otimes \dots \otimes \nu$ satisfying*

$$\alpha_t^{i,N} = \phi^N(t, \mathbf{X}_0^i, \mathbf{W}_{t \wedge \cdot}^i, B_{t \wedge \cdot}), \mathbb{P}_\nu^N \text{ a.e. and } V_S^N(\nu, \dots, \nu) - \delta_N \leq J^N(\alpha^{1,N}, \dots, \alpha^{N,N}).$$

4 Proofs of the main results

In this part, we will present the proof of the main results of this paper namely Theorem 3.1 and Theorem 3.3. Some proofs use the results from Section 5 which will be proven just after.

4.1 Equivalence result

This section is devoted to the proof of Theorem 3.1. To achieve this proof, we provide an approximation of measure-valued rule by McKean-Vlasov processes. Before starting the proofs, by shifting some probabilities, let us give a reformulation of measure-valued rules. For all $(t, \mathbf{b}, \pi, m) \in [0, T] \times \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n$,

$$\pi_t[\mathbf{b}](dy) := \int_{\mathbb{R}^n} \delta_{(y' + \sigma_0 \mathbf{b}_t)}(dy) \pi_t(dy'), \quad m[\mathbf{b}_t](du, dy) := \int_{\mathbb{R}^n \times U} \delta_{(y' + \sigma_0 \mathbf{b}_t)}(dy) m(du, dy') \tag{4.1}$$

and any $q \in \mathbb{M}$,

$$q_t[\mathbf{b}](dm) := \int_{\mathcal{P}_U^n} \delta_{(m'[\mathbf{b}_t])}(dm) q_t(dm'). \tag{4.2}$$

In the same way, let us consider the “shifted” generator $\widehat{\mathcal{L}}$,

$$\begin{aligned} &\widehat{\mathcal{L}}_t[\varphi](y, \mathbf{b}, \pi, m, u) \\ &:= \frac{1}{2} \text{Tr}[\sigma \sigma^\top(t, y + \sigma_0 \mathbf{b}_t, \pi_t[\mathbf{b}_t], m[\mathbf{b}_t], u) \nabla^2 \varphi(y)] + b(t, y + \sigma_0 \mathbf{b}_t, \pi_t[\mathbf{b}_t], m[\mathbf{b}_t], u)^\top \nabla \varphi(y). \end{aligned} \tag{4.3}$$

Next, on the canonical filtered space $(\overline{\Omega}, \overline{\mathbb{F}})$ (see Section 2.3), let $(\vartheta_t)_{t \in [0, T]}$ be the $\mathcal{P}(\mathbb{R}^n)$ -valued $\overline{\mathbb{F}}$ -adapted continuous process and $(\Theta_t)_{t \in [0, T]}$ be the \mathcal{P}_U^n -valued $\overline{\mathbb{F}}$ -predictable process defined by

$$\vartheta_t(\bar{\omega}) := \mu_t(\bar{\omega})[-B(\bar{\omega})] \text{ and } \Theta_t(\bar{\omega})(dm) := \Lambda_t(\bar{\omega})[-B(\bar{\omega})](dm), \text{ for all } (t, \bar{\omega}) \in [0, T] \times \overline{\Omega}. \tag{4.4}$$

The next result follows immediately, so we omit the proof.

Lemma 4.1. *Let $P \in \overline{\mathcal{P}}_V(\nu)$. Then, $\Theta_t(\mathcal{Z}_{\vartheta_t}) = 1$, $dP \otimes dt$, a.e. $(t, \bar{\omega}) \in [0, T] \times \overline{\Omega}$, and P -a.e. $\bar{\omega} \in \overline{\Omega}$, for all $(f, t) \in C_b^2(\mathbb{R}^n) \times [0, T]$,*

$$N_t(f) = \langle f, \vartheta_t \rangle - \langle f, \nu \rangle - \int_0^t \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} \widehat{\mathcal{L}}_r f(y, B, \vartheta, m, u) m(du, dy) \Theta_r(dm) dr.$$

Next, let us provide some estimates for the different controls. The first result is standard, the second is just an application of Theorem 5.2 (see also Theorem 5.4) combined with Theorem 4.1.

Lemma 4.2 (Estimates). *Under Theorem 2.1, for any $(\nu, \nu^1, \dots, \nu^N) \in \mathcal{P}_{p'}(\mathbb{R}^n)^{N+1}$ with $p' > p$, there exists $K > 0$, depending only of coefficients (b, σ) and p' , such that: for every $(\alpha^1, \dots, \alpha^N) \in (\mathcal{A}_N(\nu_N))^N$ one has*

$$\mathbb{E}^{P^N} \left[\sup_{t \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \mu_t(dx) \right] \leq K \left[1 + \int_{\mathbb{R}^n} |x'|^{p'} \frac{1}{N} \sum_{i=1}^N \nu^i(dx') \right],$$

where $P^N := \mathbb{P}(\alpha^1, \dots, \alpha^N) \in \mathcal{P}(\overline{\Omega})$ (see definition (2.3)), and for each $P \in \overline{\mathcal{P}}_V(\nu)$ or $\alpha \in \mathcal{A}(\nu)$ with $P = \mathbb{P}_\nu \circ (\mu^\alpha, \delta_{\overline{\mu}_t^\alpha}(dm) dt, B)^{-1}$, P -a.e. $\omega \in \overline{\Omega}$,

$$\sup_{t \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \vartheta_t(\omega)(dx) + \mathbb{E}^P \left[\sup_{t \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \mu_t(dx) \right] \leq K \left[1 + \int_{\mathbb{R}^n} |x'|^{p'} \nu(dx') \right].$$

In addition

$$\mathcal{W}_p(\vartheta_s(\omega), \vartheta_t(\omega))^p \leq K|t - s|, \text{ for all } (t, s) \in [0, T] \times [0, T], \text{ P-a.e. } \omega \in \overline{\Omega},$$

where ϑ is the process given in equation (4.4).

4.1.1 Technical lemmas

In this part, from a measure-valued rule, we will build a sequence of processes that approximate the measure-valued rule and that are close enough to strong control rules. This part is the fundamental part for the proof of Theorem 3.1.

let $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, $P \in \overline{\mathcal{P}}_V(\nu)$, and $(\widetilde{\Omega}, \widetilde{\mathbb{F}}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ be a filtered probability space supporting W \mathbb{R}^n -valued $\widetilde{\mathbb{F}}$ -Brownian motion and let ξ be a $\widetilde{\mathcal{F}}_0$ -random variable s.t. $\mathcal{L}^{\widetilde{\mathbb{P}}}(\xi) = \nu$. We define the filtered probability space $(\widehat{\Omega}, \widehat{\mathbb{F}}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$ which is an extension of the canonical space $(\overline{\Omega}, \overline{\mathbb{F}}, P)$: $\widehat{\Omega} := \widetilde{\Omega} \times \overline{\Omega}$, $\widehat{\mathbb{F}} := (\widetilde{\mathcal{F}}_t \otimes \overline{\mathcal{F}}_t)_{t \in [0, T]}$ and $\widehat{\mathbb{P}} := \widetilde{\mathbb{P}} \otimes P$. The variables (ξ, W) of $\widetilde{\Omega}$ and (B, μ, Λ) of $\overline{\Omega}$ are naturally extended on the space $\widehat{\Omega}$ while keeping the same notation $(\xi, W, B, \mu, \Lambda)$ for simplicity. Also, let us consider the filtration $(\widehat{\mathcal{G}}_t)_{t \in [0, T]}$ defined by

$$\widehat{\mathcal{G}}_t := \sigma\{B_{t \wedge \cdot}, \mu_{t \wedge \cdot}, \Lambda_{t \wedge \cdot}\}, \text{ for all } t \in [0, T].$$

Proposition 4.3. *Under Theorem 2.1, if we take any $[0, 1]$ -valued uniform variable Z $\widehat{\mathbb{P}}$ -independent of $(\xi, W, B, \mu, \Lambda)$, there exists a sequence of $\widehat{\mathbb{F}}$ -predictable processes $(\alpha^k)_{k \in \mathbb{N}^*}$ satisfying: for each $k \in \mathbb{N}^*$,*

$$\alpha_t^k := G^k(t, \xi, \mu_{t \wedge \cdot}, \Lambda_{t \wedge \cdot}, W_{t \wedge \cdot}, B_{t \wedge \cdot}, Z), \widehat{\mathbb{P}}\text{-a.e.}, \text{ for all } t \in [0, T],$$

with a Borel function $G^k : [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}(\mathcal{P}_U^n) \times \mathcal{C}^n \times \mathcal{C}^\ell \times [0, 1] \rightarrow U$ such that if we let \widehat{X}^k be the unique strong solution of: $\mathbb{E}^{\widehat{\mathbb{P}}}[\|\widehat{X}^k\|^{p'}] < \infty$, for all $t \in [0, T]$

$$\widehat{X}_t^k = \xi + \int_0^t b(r, \widehat{X}_r^k, \mu^k, \bar{\mu}_r^k, \alpha_r^k) dr + \int_0^t \sigma(r, \widehat{X}_r^k, \mu^k, \bar{\mu}_r^k, \alpha_r^k) dW_r + \sigma_0 B_t, \widehat{\mathbb{P}}\text{-a.e.}$$

where $\mu_t^k := \mathcal{L}^{\mathbb{P}}(\widehat{X}_t^k | \widehat{\mathcal{G}}_t)$ and $\bar{\mu}_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^k, \alpha_t^k | \widehat{\mathcal{G}}_t)$ then

$$\lim_{k \rightarrow \infty} \left[\mathcal{W}_p \left(\delta_{\bar{\mu}_s^k} (dm) ds, \Lambda_s (dm) ds \right) + \sup_{t \in [0, T]} \mathcal{W}_p(\mu_t^k, \mu_t) \right] = 0, \widehat{\mathbb{P}}\text{-a.e.} \quad (4.5)$$

Therefore

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}} \left((\mu_t^k)_{t \in [0, T]}, \delta_{\bar{\mu}_s^k} (dm) ds, (B_t)_{t \in [0, T]} \right) = \mathbb{P}, \text{ for the Wasserstein metric } \mathcal{W}_p.$$

Proof. As $\mathbb{P} \in \overline{\mathcal{P}}_V(\nu)$, by definition, \mathbb{P} a.e. $\omega \in \overline{\Omega}$, $N_t(f) = 0$ for all $f \in C_b^2(\mathbb{R}^n)$ and $t \in [0, T]$. By Theorem 4.1, by taking into account the extension of all variables on $\widehat{\Omega}$, recall that $(\vartheta_t)_{t \in [0, T]}$ and $(\Theta_t)_{t \in [0, T]}$ are defined in (4.4), one has $\Theta_t(\mathcal{Z}_{\vartheta_t}) = 1$, $d\widehat{\mathbb{P}} \otimes dt$ a.e. $(t, \omega) \in [0, T] \times \widehat{\Omega}$, and $\widehat{\mathbb{P}}$ -a.e. $\omega \in \widehat{\Omega}$, for all $(f, t) \in C_b^2(\mathbb{R}^n) \times [0, T]$,

$$N_t(f) = \langle f, \vartheta_t \rangle - \langle f, \nu \rangle - \int_0^t \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} \widehat{\mathcal{L}}_r f(y, B, \vartheta, m, u) m(du, dy) \Theta_r(dm) dr.$$

Define

$$\Gamma := \left\{ m \in \mathcal{P}_U^n : \int_{\mathbb{R}^n} |y|^{p'} m(dy, U) \leq \widehat{K} \right\},$$

where $\widehat{K} > 0$ is such that $\widehat{K} > K \left[1 + \int_{\mathbb{R}^n} |x'|^{p'} \nu(dx') \right]$, with K is a constant used in Lemma 4.2. Notice that Γ is a compact set of $\mathcal{P}_p(\mathbb{R}^n \times U)$ and by Theorem 4.2, one has that $\Theta_t(\Gamma) = 1$, $d\widehat{\mathbb{P}} \otimes dt$, a.e. $(t, \omega) \in [0, T] \times \widehat{\Omega}$. As Γ is a compact set of $\mathcal{P}_p(\mathbb{R}^n \times U)$, there exists a family of measurable functions $(h^k)_{k \in \mathbb{N}^*}$ with $h^k : [0, T] \times \mathbb{M} \rightarrow \mathcal{P}_U^n$, s.t.

$$\lim_{k \rightarrow \infty} \delta_{h^k(t, \Theta_{t \wedge \cdot})} (dm) dt = \Theta_t (dm) dt, \widehat{\mathbb{P}}\text{-a.e.}$$

then

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}} (\vartheta, \delta_{h^k(t, \Theta_{t \wedge \cdot})} (dm) dt, B) = \mathcal{L}^{\widehat{\mathbb{P}}} (\vartheta, \Theta, B), \text{ in } \mathcal{W}_p.$$

In the same spirit of notations (4.3), we introduce

$$[\widehat{b}, \widehat{\sigma}](t, y, \mathbf{b}, \pi, m, u) := [b, \sigma](t, y + \sigma_0 \mathbf{b}_t, \pi[\mathbf{b}], m[\mathbf{b}_t], u), \quad (4.6)$$

notice that $[\widehat{b}, \widehat{\sigma}] : [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{P}_U^n \times U \rightarrow \mathbb{R}^n \times \mathbb{S}^{n \times n}$ is continuous and for $\mathbf{b} \in \mathcal{C}^\ell$, $[\widehat{b}, \widehat{\sigma}](\cdot, \cdot, \mathbf{b}, \cdot, \cdot, \cdot)$ verify the Assumption 2.1 with constant C and θ independent of \mathbf{b} (see Assumption 2.1).

Extended mean field control problem: a propagation of chaos result

Now, let us apply Theorem 5.8 (see also Theorem 5.6). As $(\vartheta, \delta_{h^k(t, \Theta_{t\wedge\cdot})}(dm)dt, B)_{k \in \mathbb{N}^*}$ is $\widehat{\mathbb{P}}$ independent of (ξ, W) and

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}} \left(\vartheta, \delta_{h^k(s, \Theta_{s\wedge\cdot})}(dm)ds, B \right) = \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta, \Theta_s(dm)ds, B), \text{ in } \mathcal{W}_p,$$

by Theorem 5.8, there exists $G^k : [0, T] \times \mathbb{R}^n \times \mathbb{M} \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}^n \times \mathcal{C}^\ell \times [0, 1] \rightarrow U$ a Borel function such that if X^k is the unique strong solution of: for all $t \in [0, T]$

$$X_t^k = \xi + \int_0^t \hat{b}(r, X_r^k, B, \vartheta^k, \bar{\vartheta}_r^k, \alpha_r^k)dr + \int_0^t \hat{\sigma}(r, X_r^k, B, \vartheta^k, \bar{\vartheta}_r^k, \alpha_r^k)dW_r, \widehat{\mathbb{P}}\text{-a.e.}, \quad (4.7)$$

where

$$\alpha_t^k := G^k(t, \xi, \Theta_{t\wedge\cdot}^k, \vartheta_{t\wedge\cdot}, W_{t\wedge\cdot}, B_{t\wedge\cdot}, Z), \quad \bar{\vartheta}_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^k, \alpha_t^k | \mathcal{G}_t^k) \text{ and } \vartheta_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^k | \mathcal{G}_t^k),$$

with $\Theta_t^k(dm)dt := \delta_{(h^k(t, \Theta_{t\wedge\cdot}))}(dm)dt$, and $\mathbb{G}^k := (\mathcal{G}_s^k)_{s \in [0, T]} := (\sigma\{\vartheta_{s\wedge\cdot}, \Theta_{s\wedge\cdot}^k, B_{s\wedge\cdot}\})_{s \in [0, T]}$, then

$$\lim_{j \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\bar{\vartheta}_t^{k_j}, \mathbf{m}_t^{k_j})^p dt + \sup_{t \in [0, T]} \mathcal{W}_p(\vartheta_t^{k_j}, \vartheta_t) \right] = 0$$

and

$$\lim_{j \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta^{k_j}, \Theta^{k_j}, B) = \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta, \Theta, B), \text{ in } \mathcal{W}_p,$$

where $\mathbf{m}_t^k := h^k(t, \Theta_{t\wedge\cdot})$ and $(k_j)_{j \in \mathbb{N}^*} \subset \mathbb{N}^*$ is a sub-sequence. Notice that, as $\mathbb{G}^k \subset \widehat{\mathbb{G}}$, and (ξ, W, Z) are $\widehat{\mathbb{P}}$ independent of $\widehat{\mathbb{G}}$, one has $\mathcal{L}^{\widehat{\mathbb{P}}}(X_t^k, \alpha_t^k | \mathcal{G}_t^k) = \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^k, \alpha_t^k | \widehat{\mathcal{G}}_t)$, $\widehat{\mathbb{P}}$ -a.e. for all $t \in [0, T]$. Using equation (4.6), we rewrite X^k by: for all $t \in [0, T]$, $\widehat{\mathbb{P}}$ -a.e.

$$X_t^k = \xi + \int_0^t b(r, X_r^k + \sigma_0 B_r, (\mathcal{L}^{\widehat{\mathbb{P}}}(X_s^k + \sigma_0 B_s | \widehat{\mathcal{G}}_s))_{s \in [0, T]}, \mathcal{L}^{\widehat{\mathbb{P}}}(X_r^k + \sigma_0 B_r, \alpha_r^k | \widehat{\mathcal{G}}_r), \alpha_r^k)dr + \int_0^t \sigma(r, X_r^k + \sigma_0 B_r, (\mathcal{L}^{\widehat{\mathbb{P}}}(X_s^k + \sigma_0 B_s | \widehat{\mathcal{G}}_s))_{s \in [0, T]}, \mathcal{L}^{\widehat{\mathbb{P}}}(X_r^k + \sigma_0 B_r, \alpha_r^k | \widehat{\mathcal{G}}_r), \alpha_r^k)dW_r.$$

Denote by $\widehat{X}^k := X^k + \sigma_0 B$, one finds $\widehat{\mathbb{P}}$ -a.e., for all $t \in [0, T]$,

$$\widehat{X}_t^k = \xi + \int_0^t b(r, \widehat{X}_r^k, (\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_s^k | \widehat{\mathcal{G}}_s))_{s \in [0, T]}, \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_r^k, \alpha_r^k | \widehat{\mathcal{G}}_r), \alpha_r^k)dr + \int_0^t \sigma(r, \widehat{X}_r^k, (\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_s^k | \widehat{\mathcal{G}}_s))_{s \in [0, T]}, \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_r^k, \alpha_r^k | \widehat{\mathcal{G}}_r), \alpha_r^k)dW_r + \sigma_0 B_t,$$

With the notation introduced in (4.1) and (4.2), it is straightforward to check that the map

$$(\pi, q, \mathbf{b}) \in \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M} \times \mathcal{C}^\ell \rightarrow (\pi[\mathbf{b}], q_t[\mathbf{b}](dm)dt, \mathbf{b}) \in \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M} \times \mathcal{C}^\ell$$

is continuous. Consequently, one has

$$\lim_{j \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\bar{\vartheta}_t^{k_j} [B_t], \mathbf{m}_t^{k_j} [B_t])^p dt + \sup_{t \in [0, T]} \mathcal{W}_p(\vartheta_t^{k_j} [B], \vartheta_t [B]) \right] = 0,$$

therefore, in \mathcal{W}_p ,

$$\begin{aligned} & \lim_{j \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}} \left((\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^{k_j} | \widehat{\mathcal{G}}_t))_{t \in [0, T]}, \delta_{(\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_s^{k_l}, \alpha_s^{k_l} | \widehat{\mathcal{G}}_s))}(dm)ds, B \right) \\ &= \lim_{j \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta^{k_j} [B], \Theta_t^{k_j} [B](dm)dt, B) = \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta[B], \Theta_t[B](dm)dt, B). \end{aligned}$$

After simple calculations, $(\vartheta[B], \Theta_t[B](dm)dt, B) = (\mu, \Lambda, B)$, $\widehat{\mathbb{P}}$ -a.e. Then

$$\lim_{j \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^{k_j}, \alpha_t^{k_j} | \widehat{\mathcal{G}}_t), \mathbf{m}_t^{k_j}[B_t] \right)^p dt + \sup_{t \in [0, T]} \mathcal{W}_p \left(\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^{k_j} | \widehat{\mathcal{G}}_t), \mu_t \right) \right] = 0,$$

and hence

$$\lim_j \mathcal{L}^{\widehat{\mathbb{P}}} \left((\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^{k_j} | \widehat{\mathcal{G}}_t))_{t \in [0, T]}, \delta_{(\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_s^{k_j}, \alpha_s^{k_j} | \widehat{\mathcal{G}}_s))} (dm)ds, B \right) = \mathcal{L}^{\widehat{\mathbb{P}}}(\mu, \Lambda, B) = P, \text{ in } \mathcal{W}_p.$$

After extraction from $(\widehat{X}^{k_j}, \alpha^{k_j})_{j \in \mathbb{N}^*}$, one has also the $\widehat{\mathbb{P}}$ -a.e. convergence (4.5). \square

4.1.2 Proof of Theorem 3.1

First, for $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, under Theorem 2.1, let us prove that $\overline{\mathcal{P}}_V(\nu)$ is a compact set for the Wasserstein topology \mathcal{W}_p . Let $(P_k)_{k \in \mathbb{N}^*} \subset \overline{\mathcal{P}}_V(\nu)$, by Theorem 4.4, $(P_k)_{k \in \mathbb{N}^*}$ is relatively compact for the Wasserstein topology \mathcal{W}_p and any limit P_∞ of any sub-sequence belongs to $\overline{\mathcal{P}}_V(\nu)$. Therefore $\overline{\mathcal{P}}_V(\nu)$ is compact. By similar techniques used in [11, Theorem 3.1], it is straightforward to show that $\overline{\mathcal{P}}_V(\nu)$ is convex.

Next, we prove the items (i) and (ii) of Theorem 3.1. By applying Theorem 4.3, with the same notations, for any $[0, 1]$ -valued uniform variable Z $\widehat{\mathbb{P}}$ -independent of $(\xi, W, B, \mu, \Lambda)$, there exists a sequence of $\widehat{\mathbb{F}}$ -predictable processes $(\alpha^k)_{k \in \mathbb{N}^*}$ satisfying: for each $k \in \mathbb{N}^*$,

$$\alpha_t^k := G^k(t, \xi, \mu_{t^\wedge}, \Lambda_{t^\wedge}, W_{t^\wedge}, B_{t^\wedge}, Z), \widehat{\mathbb{P}}\text{-a.e.}, \text{ for all } t \in [0, T],$$

with $G^k : [0, T] \times \mathbb{R}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}(\mathcal{P}_U^n) \times \mathcal{C}^n \times \mathcal{C}^\ell \times [0, 1] \rightarrow U$ is a Borel function such that if \widehat{X}^k is the unique strong solution of: for all $t \in [0, T]$

$$\widehat{X}_t^k = \xi + \int_0^t b(r, \widehat{X}_r^k, \mu^k, \bar{\mu}_r^k, \alpha_r^k) dr + \int_0^t \sigma(r, \widehat{X}_r^k, \mu^k, \bar{\mu}_r^k, \alpha_r^k) dW_r + \sigma_0 B_t, \widehat{\mathbb{P}}\text{-a.e.}$$

where $\mu_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^k | \widehat{\mathcal{G}}_t)$ and $\bar{\mu}_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^k, \alpha_t^k | \widehat{\mathcal{G}}_t)$ then

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}} \left((\mu_t^k)_{t \in [0, T]}, \delta_{\bar{\mu}_s^k} (dm)ds, (B_t)_{t \in [0, T]} \right) = P, \text{ for the Wasserstein metric } \mathcal{W}_p.$$

For each $k \in \mathbb{N}^*$, $\widehat{X}_t^k = H_t^k(\xi, W_{t^\wedge}, \mu_{t^\wedge}, \Lambda_{t^\wedge}, B_{t^\wedge}, Z)$, for all $t \in [0, T]$, $\widehat{\mathbb{P}}$ -a.e. with $H^k : \mathbb{R}^n \times \mathcal{C}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M} \times \mathcal{C}^\ell \times [0, 1] \rightarrow \mathcal{C}^n$ a Borel function. Then, as (ξ, W, Z) are $\widehat{\mathbb{P}}$ -independent of (μ, Λ, B) , one gets that for all $t \in [0, T]$, $\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \alpha_t^k | \widehat{\mathcal{G}}_t) = \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \alpha_t^k | \widehat{\mathcal{G}}_T)$, $\widehat{\mathbb{P}}$ -a.e.. Let us introduce the process $(\widehat{\mu}_t^k)_{t \in [0, T]}$,

$$\widehat{\mu}_t^k := \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \widehat{X}_{t^\wedge}^k - \sigma_0 B_{t^\wedge}, W, \Lambda_{t^\wedge}^k | \widehat{\mathcal{G}}_t), \text{ for all } t \in [0, T] \text{ with } \Lambda_t^k(du)dt := \delta_{\alpha_t^k}(du)dt.$$

For each $k \in \mathbb{N}^*$, $\widehat{\mu}_t^k \in \mathcal{P}(\mathcal{C}^n \times \mathcal{C}^n \times \mathcal{C}^n \times \mathbb{M}(U))$, for all $t \in [0, T]$ and if $(\widetilde{X}, \widetilde{Y}, \widetilde{W}, \widetilde{\Lambda})$ is the canonical process on $\mathcal{C}^n \times \mathcal{C}^n \times \mathcal{C}^n \times \mathbb{M}(U)$, one has $\mu_t^k = \mathcal{L}^{\widehat{\mu}_t^k}(\widetilde{X}_t)$, $\widehat{\mathbb{P}}$ -a.e., and $\mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_t^k, \alpha_t^k | \widehat{\mathcal{G}}_t)(dx, du) = \mathbb{E}^{\widehat{\mu}_t^k}[\delta_{\widetilde{X}_t}(dx) \widetilde{\Lambda}_t(du)]$, $\widehat{\mathbb{P}}$ -a.e. for all $t \in [0, T]$. It is straightforward to see that $\widehat{\mu}_t^k = \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \widehat{X}_{t^\wedge}^k - \sigma_0 B_{t^\wedge}, W, \Lambda_{t^\wedge}^k | \widehat{\mathcal{G}}_T)$, for each $k \in \mathbb{N}^*$, then for all $t \in [0, T]$, $\widehat{\mathbb{P}}$ -a.e.,

$$\widehat{\mu}_t^k = \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \widehat{X}_{t^\wedge}^k - \sigma_0 B_{t^\wedge}, W, \Lambda_{t^\wedge}^k | B_{t^\wedge}, \widehat{\mu}_{t^\wedge}^k) = \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{X}_{t^\wedge}^k, \widehat{X}_{t^\wedge}^k - \sigma_0 B_{t^\wedge}, W, \Lambda_{t^\wedge}^k | B, \widehat{\mu}^k)$$

and $(B, \widehat{\mu}^k)$ are $\widehat{\mathbb{P}}$ -independent of (ξ, W) . For all $k \in \mathbb{N}^*$, denote

$$\overline{\mathbb{Q}}^k := \widehat{\mathbb{P}} \circ (\widehat{X}^k, \widehat{X}^k - \sigma_0 B, \Lambda^k, W, B, \widehat{\mu}^k)^{-1}$$

\bar{Q}^k belongs to

$$\mathcal{P}\left(\mathcal{C}^n \times \mathcal{C}^n \times \mathbf{M}(U) \times \mathcal{C}^n \times \mathcal{C}^\ell \times \mathcal{P}(\mathcal{C}^n \times \mathcal{C}^n \times \mathcal{C}^n \times \mathbf{M}(U))\right),$$

then \bar{Q}^k is a weak control according to [11, Definition 2.9]. Then by (a slight extension of) [11, Proposition 4.5],

(1) when $\ell \neq 0$, there exists $\alpha^{j,k} \in \mathcal{A}(\nu)$, and $X^{\alpha^{j,k}}$ the strong solution of (2.4) with control $\alpha^{j,k}$ such that

$$\begin{aligned} & \lim_{j \rightarrow \infty} \mathbb{P}_\nu \circ \left(X^{\alpha^{j,k}}, W, B, \delta_{(\bar{\mu}_s^{\alpha^{j,k}}, \alpha_s^{j,k})} (dm, du) ds \right)^{-1} \\ &= \widehat{\mathbb{P}} \circ \left(\widehat{X}^k, W, B, \delta_{(\bar{\mu}_s^k, \alpha_s^k)} (dm, du) ds \right)^{-1}, \text{ in } \mathcal{W}_p. \end{aligned}$$

(2) When $\ell = 0$, there exists a family of Borel functions $(\kappa_j^k)_{k,j}$ with $\kappa_j^k : [0, T] \times \mathbb{R}^n \times \mathcal{C}^n \times [0, 1] \rightarrow U$, such that if $\alpha_t^{j,k}[z] := \kappa_j^k(t, \xi, W_{t \wedge \cdot}, z)$, for $z \in [0, 1]$, one gets $(\alpha_t^{j,k}[z])_{t \in [0, T]} \in \mathcal{A}(\nu)$ and

$$\begin{aligned} & \lim_{j \rightarrow \infty} \int_0^1 \mathbb{P}_\nu \circ \left(X^{\alpha^{j,k}[z]}, W, B, \delta_{(\bar{\mu}_s^{\alpha^{j,k}[z]}, \alpha_s^{j,k}[z])} (dm, du) ds \right)^{-1} dz \\ &= \widehat{\mathbb{P}} \circ \left(\widehat{X}^k, W, B, \delta_{(\bar{\mu}_s^k, \alpha_s^k)} (dm, du) ds \right)^{-1}, \text{ in } \mathcal{W}_p. \end{aligned}$$

All these results are enough to deduce the items (i) and (ii) of Theorem 3.1, and conclude that: for $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, $V_S(\nu) = V_V(\nu)$ and there exists $\mathbb{P}^* \in \overline{\mathcal{P}}_V(\nu)$ such that $V_V(\nu) = \mathbb{E}^{\mathbb{P}^*} [J(\mu, \Lambda)]$.

4.2 Propagation of chaos

With the help of Theorem 3.1, in this section we provide one of the main objective of this paper, which is to prove the limit theory result or (controlled) propagation of chaos.

4.2.1 Technical results: study of the behavior of processes when N goes to infinity

In this part, the properties of some sequences of probability measures on the canonical space $\bar{\Omega}$ are given. Mainly, the behavior when N goes to infinity of sequences of type $(\mathbb{P}(\alpha^1, \dots, \alpha^N))_{N \in \mathbb{N}^*}$ construct from the formulation of large population stochastic control problem are studied. (see Section 2.1 and Theorem 2.3).

Proposition 4.4. *Let Theorem 2.1 hold true and $(\nu^i)_{i \in \mathbb{N}^*} \subset \mathcal{P}_{p'}(\mathbb{R}^n)$. Recall that $\nu_N := \nu^1 \otimes \dots \otimes \nu^N$, for each $N \in \mathbb{N}^*$.*

(i) *Let $(\mathbb{P}^N)_{N \in \mathbb{N}^*}$ be the sequence satisfying $\mathbb{P}^N := \mathbb{P}(\alpha^{1,N}, \dots, \alpha^{N,N})$ (see definition (2.3)) with $\alpha^{i,N} \in \mathcal{A}_N(\nu_N) \forall i \in \llbracket 1, N \rrbracket$, for each $N \in \mathbb{N}^*$. If*

$$\sup_{N \geq 1} \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^n} |x'|^{p'} \nu^i(dx') < \infty$$

then $(\mathbb{P}^N)_{N \in \mathbb{N}^}$ is precompact in $\mathcal{P}_p(\bar{\Omega})$ for the metric \mathcal{W}_p and for every $\mathbb{P}^\infty \in \overline{\mathcal{P}}(\bar{\Omega})$ the limit of any sub-sequence $(\mathbb{P}^{N_j})_{j \in \mathbb{N}}$, $\mathbb{P}^\infty \in \overline{\mathcal{P}}_V(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i)$.*

(ii) *Let us consider the sequence $(\mathbb{P}_k)_{k \in \mathbb{N}^*}$ of probability measures such that $\mathbb{P}_k \in \overline{\mathcal{P}}_V(\nu^k)$ for each $k \in \mathbb{N}^*$. If*

$$\sup_{k \geq 1} \int_{\mathbb{R}^n} |x'|^{p'} \nu^k(dx') < \infty$$

then $(P_k)_{k \in \mathbb{N}^*}$ is precompact in $\mathcal{P}_p(\bar{\Omega})$ for the metric \mathcal{W}_p and for every $P_\infty \in \mathcal{P}(\bar{\Omega})$ the limit of any sub-sequence $(P_{k_j})_{j \in \mathbb{N}^*}$, $P_\infty \in \bar{\mathcal{P}}_V(\lim_{j \rightarrow \infty} \nu^{k_j})$.

Proof. (i) Thanks to Proposition A.2 or/and Proposition-B.1 of [7], as U is compact, it is easy to check that $(P^N)_{N \in \mathbb{N}^*}$ is pre-compact on $\mathcal{P}_p(\bar{\Omega})$ for the metric \mathcal{W}_p . Let P^∞ be a limit of a sub-sequence $(P^{N_j})_{j \in \mathbb{N}^*}$. For sake of simplicity, we denote $(P^{N_j})_{j \in \mathbb{N}^*} = (P^N)_{N \in \mathbb{N}^*}$ and $\nu := \lim_j \frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i$.

Now, let us show $P^\infty \in \bar{\mathcal{P}}_V(\nu)$. Let $f \in C_b^2(\mathbb{R}^n)$. For each $t \in [0, T]$, we shall denote $N_t(B_{t\wedge\cdot}, \Lambda_{t\wedge\cdot}, \mu_{t\wedge\cdot})(f) = N_t(f)$ to specify the dependence w.r.t. (B, μ, Λ) (see definition (2.6)). Notice that the function $(t, \mathbf{b}, \pi, q) \in [0, T] \times \mathcal{C}^\ell \times \mathcal{C}_W^n \times \mathbb{M} \rightarrow N_t(\mathbf{b}_{t\wedge\cdot}, q_{t\wedge\cdot}, \pi_{t\wedge\cdot})(f) \in \mathbb{R}$ is continuous and bounded. It is straightforward to check that: for all $t \in [0, T]$

$$\begin{aligned} & N_t(B_{t\wedge\cdot}, (\delta_{\varphi_s^N}(dm)ds)_{t\wedge\cdot}, \varphi_{t\wedge\cdot}^{N, \mathbf{X}})(f) \\ &= \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(\mathbf{X}_r^{\alpha, i} - \sigma_0 B_r) \sigma(r, \mathbf{X}_r^{\alpha, i}, \varphi_r^{N, \mathbf{X}}, \varphi_r^N, \alpha_r^i) d\mathbf{W}_r^i, \mathbb{P}_\nu^N\text{-a.e.} \end{aligned}$$

With the same techniques used in the proof of [24, Proposition 5.1] or [11, Proposition 4.17], one has

$$\begin{aligned} & \mathbb{E}^{\mathbb{P}^\infty} \left[|(N_t(f))|^2 \right] \\ &= \mathbb{E}^{\mathbb{P}^\infty} \left[|(N_t(B_{t\wedge\cdot}, \Lambda_{t\wedge\cdot}, \mu_{t\wedge\cdot})(f))|^2 \right] = \lim_N \mathbb{E}^{\mathbb{P}^N} \left[|(N_t(B_{t\wedge\cdot}, \Lambda_{t\wedge\cdot}, \mu_{t\wedge\cdot})(f))|^2 \right] \\ &= \lim_N \mathbb{E}^{\mathbb{P}_\nu^N} \left[|(N_t(B_{t\wedge\cdot}, (\delta_{\varphi_s^N}(dm)ds)_{t\wedge\cdot}, \varphi_{t\wedge\cdot}^{N, \mathbf{X}})(f))|^2 \right] \\ &= \lim_N \mathbb{E}^{\mathbb{P}_\nu^N} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(\mathbf{X}_r^{\alpha, i} - \sigma_0 B_r) \sigma(r, \mathbf{X}_r^{\alpha, i}, \varphi_r^{N, \mathbf{X}}, \varphi_r^N, \alpha_r^i) d\mathbf{W}_r^i \right|^2 \right] \\ &= \lim_N \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}_\nu^N} \left[\int_0^t \left| \nabla f(\mathbf{X}_r^{\alpha, i} - \sigma_0 B_r) \sigma(r, \mathbf{X}_r^{\alpha, i}, \varphi_r^{N, \mathbf{X}}, \varphi_r^N, \alpha_r^i) \right|^2 dr \right] = 0. \end{aligned}$$

By taking (t, f) under a countable set of $[0, T] \times C_b^2(\mathbb{R}^n)$ then \mathbb{P}^∞ a.e. $\omega \in \bar{\Omega}$, $N_t(f) = 0$ for all $(t, f) \in [0, T] \times C_b^2(\mathbb{R}^n)$.

For any bounded continuous functions $h \in C_b(\mathbb{R}^n)$, the map $(q, \pi) \in \mathbb{M} \times \mathcal{C}_W^n \rightarrow \int_0^T \int_{\mathcal{P}_V^n} |\langle h, m(dz, U) \rangle - \langle h, \mu_t(dz) \rangle|^2 q_t(dm)dt \in \mathbb{R}$ is bounded and continuous (see for instance Theorem A.4), one finds that

$$\begin{aligned} & \mathbb{E}^{\mathbb{P}^\infty} \left[\int_0^T \int_{\mathcal{P}_V^n} |\langle h, m(dz, U) \rangle - \langle h, \mu_t(dz) \rangle|^2 \Lambda_t(dm)dt \right] \\ &= \lim_N \mathbb{E}^{\mathbb{P}_\nu^N} \left[\int_0^T \int_{\mathcal{P}_V^n} |\langle h, m(dz, U) \rangle - \langle h, \varphi_t^{N, \mathbf{X}}(dz) \rangle|^2 \delta_{\varphi_t^N}(dm)dt \right] \\ &= \lim_N \mathbb{E}^{\mathbb{P}_\nu^N} \left[\int_0^T \left| \frac{1}{N} \sum_{i=1}^N [h(\mathbf{X}_t^{\alpha, i}) - h(\mathbf{X}_t^{\alpha, i})] \right|^2 dt \right] = 0, \end{aligned}$$

by taking h under a countable set of $C_b(\mathbb{R}^n)$, one concludes $\Lambda_t(\mathbb{Z}_{\mu_t}) = 1 \mathbb{P}^\infty \otimes dt$ a.e.. It is obvious that $(B_t)_{t \in [0, T]}$ is a $(\mathbb{P}^\infty, \bar{\mathbb{F}})$ Wiener process. Let $Q \in \mathbb{N}^*$, and $(h^q)_{q \in \{1, \dots, Q\}} : \mathbb{R}^n \rightarrow \mathbb{R}^Q$ be bounded functions, one has

$$\mathbb{E}^{\mathbb{P}^\infty} \left[\prod_{q=1}^Q \langle h^q, \mu_0 \rangle \right] = \prod_{q=1}^Q \langle h^q, \nu \rangle.$$

Let us show this result when $Q = 2$, when $Q \in \mathbb{N}^*$, the proof is similar.

$$\begin{aligned} \mathbb{E}^{P^\infty} [\langle h^1, \mu_0 \rangle \langle h^2, \mu_0 \rangle] &= \lim_N \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E}^{P^\nu} [h^1(\mathbf{X}_0^{\alpha,i}) h^2(\mathbf{X}_0^{\alpha,j})] \\ &= \lim_N \frac{1}{N^2} \sum_{i=1}^N \langle h^1, \nu^i \rangle \langle h^2, \nu^i \rangle + \frac{1}{N^2} \sum_{i \neq j}^N \langle h^1, \nu^i \rangle \langle h^2, \nu^j \rangle \\ &= \lim_N \langle h^1, \frac{1}{N} \sum_{i=1}^N \nu^i \rangle \langle h^2, \frac{1}{N} \sum_{i=1}^N \nu^i \rangle = \langle h^1, \nu \rangle \langle h^2, \nu \rangle, \end{aligned}$$

by [11, Proposition A.3], $P^\infty \circ (\mu_0)^{-1} = \delta_\nu$, then $\mu_0 = \nu$, P^∞ -a.e.. All these results allow to deduce the first statement of this proposition.

(ii) For the second part of this proposition, notice that, thanks to Lemma 4.2,

$$\sup_{k \in \mathbb{N}^*} \mathbb{E}^{P^k} \left[\sup_{t \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \vartheta_t(dx) \right] \leq K \left[1 + \sup_{k \in \mathbb{N}^*} \int_{\mathbb{R}^n} |x'|^{p'} \nu^k(dx') \right] < \infty$$

and

$$\limsup_{\delta \rightarrow 0} \sup_{k \in \mathbb{N}^*} \sup_{\tau} \mathbb{E}^{P^k} [\mathcal{W}_p(\vartheta_{(\tau+\delta) \wedge T}, \vartheta_\tau)] = 0,$$

where τ is a $[0, T]$ -valued $\bar{\mathbb{F}}$ -stopping time, and recall that $(\vartheta)_{t \in [0, T]}$ is the $\mathcal{P}(\mathbb{R}^n)$ -valued $\bar{\mathbb{F}}$ -adapted continuous process defined in equation (4.4). Then by Aldous' criterion [20, Lemma 16.12] (see also proof of [7, Proposition-B.1]), $(P_k \circ ((\vartheta_t)_{t \in [0, T]})^{-1})_{k \in \mathbb{N}^*}$ is relatively compact for the metric \mathcal{W}_p . Then, using the fact that $P_k \in \bar{\mathcal{P}}_V(\nu^k)$ for each $k \in \mathbb{N}^*$ and the relation between (ϑ, Θ) and the canonical processes (μ, Λ) (see equation (4.4)), we deduce that $(P_k)_{k \in \mathbb{N}^*} = (P_k \circ (\mu, \Lambda, B)^{-1})_{k \in \mathbb{N}^*}$ is relatively compact in \mathcal{W}_p . The rest of the proof is similar to the previous proof. \square

Proposition 4.5. *Let Theorem 2.1 hold true, $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$ with $p' > p$ and $(\nu^i)_{i \in \mathbb{N}} \subset \mathcal{P}_{p'}(\mathbb{R}^n)$ such that*

$$\sup_{i \in \mathbb{N}} \int_{\mathbb{R}^n} |x'|^{p'} \nu^i(dx') < \infty \text{ and } \nu^i \xrightarrow{i \rightarrow \infty} \nu, \text{ then } \lim_{i \rightarrow \infty} V_S(\nu^i) = V_S(\nu).$$

In particular, the map $V_S : \mathcal{P}_{p'}(\mathbb{R}^n) \rightarrow \mathbb{R}$ is continuous.

Proof. By Theorem 3.1, one has $V_S(\nu) = V_V(\nu)$, thanks to this result, the proof is similar to the proof of [11, Proposition 3.7.]. Let $(\delta^k)_{k \in \mathbb{N}^*} \subset \mathbb{N}^*$ with $\lim_{k \rightarrow \infty} \delta^k = 0$ and $(P^k)_{k \in \mathbb{N}^*}$ be a sequence such that $P^k \in \bar{\mathcal{P}}_V(\nu^k)$ and $V_V(\nu^k) - \delta^k \leq \mathbb{E}^{P^k} [J(\mu, \Lambda)]$. By Proposition 4.4, $(P^k)_{k \in \mathbb{N}^*}$ is relatively compact on $(\mathcal{P}_p(\bar{\Omega}), \mathcal{W}_p)$ and if $P \in \mathcal{P}(\bar{\Omega})$ is the limit of a subsequence $(P^{k_j})_{j \in \mathbb{N}^*}$ then $P \in \bar{\mathcal{P}}_V(\nu)$. Using Theorem 2.1, by convergence of $(P^{k_j})_{j \in \mathbb{N}^*}$, one has $\lim_j |\mathbb{E}^{P^{k_j}} [J(\mu, \Lambda)] - \mathbb{E}^P [J(\mu, \Lambda)]| = 0$. Therefore, one gets

$$\limsup_k V_V(\nu^k) \leq \lim_j \mathbb{E}^{P^{k_j}} [J(\mu, \Lambda)] = \mathbb{E}^P [J(\mu, \Lambda)] \leq V_V(\nu) = V_S(\nu).$$

By [11, Proposition 4.15], $V_S(\nu) \leq \liminf_j V_S(\nu^{k_j})$, this is enough to conclude that $\lim_k V_S(\nu^k) = V_S(\nu)$, and deduce the result. \square

4.2.2 Proof of Theorem 3.3

By combining Theorem 3.1, Theorem 4.4 and Theorem 4.5, this proof turns to be the same used in the proof of [11, Theorem 3.6]. For the sake of completeness, we repeat the proof.

(i) By Theorem 4.4 (with the same notations), if the sequence $(P^N)_{N \in \mathbb{N}^*}$ is such that: $V_S^N(\nu^1, \dots, \nu^N) - \varepsilon_N \leq \mathbb{E}^{P^N}[J(\mu, \Lambda)]$, where $(\varepsilon_N)_{N \in \mathbb{N}^*}$ is sequence with $\lim_{N \rightarrow \infty} \varepsilon_N = 0$, then $(P^N)_{N \in \mathbb{N}^*}$ is relatively compact on $(\mathcal{P}_p(\bar{\Omega}), \mathcal{W}_p)$ and for every $P^\infty \in \mathcal{P}(\bar{\Omega})$ the limit of the sub-sequence $(P^{N_j})_{j \in \mathbb{N}^*}$, $P^\infty \in \bar{\mathcal{P}}_V(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i)$, therefore

$$\limsup_{N \rightarrow \infty} V_S^N(\nu^1, \dots, \nu^N) \leq \lim_{j \rightarrow \infty} \mathbb{E}^{P^{N_j}}[J(\mu, \Lambda)] = \mathbb{E}^P[J(\mu, \Lambda)] \leq V_V \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right).$$

Then, as $\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \in \mathcal{P}_{p'}(\mathbb{R}^n)$ and Theorem 2.1 holds true one can deduce that $V_V \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right) = V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right)$. By [11, Proposition 4.15],

$$V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right) \leq \liminf_{j \rightarrow \infty} V_S^{N_j}(\nu^1, \dots, \nu^{N_j}).$$

To recap

$$\begin{aligned} V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right) &\leq \liminf_{j \rightarrow \infty} V_S^{N_j}(\nu^1, \dots, \nu^{N_j}) \\ &\leq \limsup_{j \rightarrow \infty} V_S^{N_j}(\nu^1, \dots, \nu^{N_j}) \leq V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right). \end{aligned}$$

(ii) Let $(N_j)_{j \in \mathbb{N}}$ be the sequence corresponding to:

$$\limsup_{N \rightarrow \infty} \left| V_S^N(\nu^1, \dots, \nu^N) - V_S \left(\frac{1}{N} \sum_{i=1}^N \nu^i \right) \right| = \lim_{j \rightarrow \infty} \left| V_S^{N_j}(\nu^1, \dots, \nu^{N_j}) - V_S \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i \right) \right|.$$

By the previous proof, $\lim_{j \rightarrow \infty} V_S^{N_j}(\nu^1, \dots, \nu^{N_j}) = V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i \right)$, as the sequence $(\frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i)_{j \in \mathbb{N}^*}$ is bounded in $(\mathcal{P}_{p'}(\mathbb{R}^n), \mathcal{W}_{p'})$ and converges in $(\mathcal{P}_p(\mathbb{R}^n), \mathcal{W}_p)$, by Theorem 4.5,

$$\lim_{j \rightarrow \infty} V_S \left(\frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i \right) = V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu^i \right),$$

this is enough to conclude the proof. □

4.3 Proof of Theorem 3.4

Notice that, for $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, by Theorem 3.1, $\bar{\mathcal{P}}_V^*(\nu)$ is nonempty. Let us define the distance function to the set $\bar{\mathcal{P}}_V^*(\nu)$, for each $Q \in \mathcal{P}(\bar{\Omega})$, $\Psi^*(Q) := \inf_{P^* \in \bar{\mathcal{P}}_V^*(\nu)} \mathcal{W}_p(Q, P^*)$. It is well know that, as $\bar{\mathcal{P}}_V^*(\nu)$ is nonempty, the function $\Psi^* : Q \in \mathcal{P}_p(\bar{\Omega}) \rightarrow \mathbb{R}$ is continuous. Then by Theorem 4.4, $(P^N)_{N \in \mathbb{N}^*}$ is precompact in $\mathcal{P}_p(\bar{\Omega})$ for the metric \mathcal{W}_p and if $P \in \mathcal{P}(\bar{\Omega})$ is the limit of a sub-sequence $(P^{N_j})_{j \in \mathbb{N}^*}$, one have $P \in \bar{\mathcal{P}}_V(\nu)$. Under Theorem 2.1, $\lim_{j \rightarrow \infty} \mathbb{E}^{P^{N_j}}[J(\mu, \Lambda)] = \mathbb{E}^P[J(\mu, \Lambda)]$. Combining Theorem 3.3 and Theorem 4.5, one has that

$$\lim_{j \rightarrow \infty} V_S^{N_j}(\nu^1, \dots, \nu^{N_j}) = V_S \left(\lim_{j \rightarrow \infty} \frac{1}{N_j} \sum_{i=1}^{N_j} \nu_i \right) = V_S(\nu) = V_V(\nu) \leq \mathbb{E}^P[J(\mu, \Lambda)],$$

then $P \in \overline{\mathcal{P}}_V^*(\nu)$. Hence each limit of any sub-sequence of $(P^N)_{N \in \mathbb{N}^*}$ belongs to $\overline{\mathcal{P}}_V^*(\nu)$. Consequently, if $(P^{N_j})_{j \in \mathbb{N}}$ is the sub-sequence corresponding to $\limsup_{N \rightarrow \infty} \Psi^*(P^N) = \lim_{j \rightarrow \infty} \Psi^*(P^{N_j})$, by continuity of Ψ^* and the fact that any limit is an optimal control, $\limsup_{N \rightarrow \infty} \Psi^*(P^N) = 0$. The second part of this proposition is just a combination of Theorem 3.1, [11, Proposition 4.15] and Theorem 3.3. This is enough to conclude the result.

5 Approximation of Fokker-Planck equations

In this section, we give an approximation of a particular Fokker-Planck equation via a sequence of measure-valued processes constructed from classical SDE processes interacting through the empirical distribution of their states and controls. This result is a crucial part for the proof of Theorem 3.1 and Theorem 3.3.

5.1 Main ideas leading the proof

Because of the technical aspect of this part, before going into details, let us first explain in a simple situation the main goal of this part and the ideas for the proof. As we said earlier, from a Fokker-Planck equation satisfied by a measure-valued solution P (see Theorem 2.6), we want to construct a sequence of “weak” McKean-Vlasov processes s.t. the limit, in a certain sense, of this sequence will be P . Let us be more precise. For simplification, we assume that $n = \ell = 1$, $U = [1, 2]$, $b = 0$, $\sigma(t, x, \pi, m, u) = \sigma(m, u) := \sigma(m)u$. Let $P \in \overline{\mathcal{P}}_V$, (μ, Λ, B) satisfy: $\Lambda_t(\mathbb{Z}_{\mu_t}) d\tilde{P} \otimes dt$ a.e. and for all (t, f)

$$d\langle f(\cdot - \sigma_0 B_t), \mu_t \rangle = \int_{\mathcal{P}_V^n} \int_{\mathbb{R}^n \times U} f''(x - \sigma_0 B_t) \sigma(m)^2 u^2 m^x (du) \mu_t(dx) \Lambda_t(dm) dt. \quad (5.1)$$

Using the SDEs formulation, on an extension $(\tilde{\Omega}, \tilde{\mathbb{F}}, \tilde{P})$ of $(\overline{\Omega}, \overline{\mathbb{F}}, P)$, we can find X satisfying

$$dX_t = \left(\int_{\mathcal{P}_V^n} \int_U \sigma(m)^2 u^2 m^{X_t} (du) \Lambda_t(dm) \right)^{1/2} dW_t + \sigma_0 dB_t, \quad (5.2)$$

$X_0 = \xi$ with $\mu_t = \mathcal{L}^{\tilde{P}}(X_t | \tilde{\mathcal{G}}_t) = \mathcal{L}^{\tilde{P}}(X_t | \tilde{\mathcal{G}}_T)$, where W is a $\tilde{\mathbb{F}}$ -Brownian motion, ξ a $\tilde{\mathcal{F}}_0$ -random variable s.t. $\mathcal{L}(\xi) = \nu$ and (W, ξ) is independent of $\tilde{\mathcal{G}}_T$. The process $(\Lambda_t)_{t \in [0, T]}$ can be seen as a control of the process X or μ . The goal is to construct a sequence of \mathbb{F} -predictable processes $(\alpha^k)_{k \in \mathbb{N}^*}$ s.t. if X^k is the solution of

$$dX_t^k = \sigma(\overline{m}_t^k) \alpha_t^k dW_t + \sigma_0 dB_t, \quad X_0^k = \xi, \quad m_t^k := \mathcal{L}(X_t^k | \tilde{\mathcal{G}}_t) \text{ and } \overline{m}_t^k := \mathcal{L}(X_t^k, \alpha_t^k | \tilde{\mathcal{G}}_t),$$

one has that

$$\lim_{k \rightarrow \infty} \tilde{P} \circ (m^k, \delta_{\overline{m}_t^k}(dm)dt, B)^{-1} = \tilde{P} \circ (\mu, \Lambda, B)^{-1} \text{ in } \mathcal{W}_p.$$

If it was possible for Equation (5.1) or Equation (5.2) to satisfied an appropriate uniqueness result (in law), this kind of approximation would become much simpler to perform. Unfortunately, for a general Λ , a uniqueness result can not be expected for this type of equation. Therefore, find the sequence $(\alpha^k)_{k \in \mathbb{N}^*}$ becomes a challenging problem.

Strategy of proof: 1-regularization This part is realized in Section 5.2. The main idea here is to regularize Equation (5.1) or Equation (5.2) in order to recover some uniqueness result. Indeed, in Section 5.2, we show that: X^ε solution of

$$dX_t^\varepsilon = \sigma^\varepsilon(t, \Lambda_t, X_t^\varepsilon) dW_t + \sigma_0 dB_t, \quad X_0^\varepsilon = \xi, \quad \mu_t^\varepsilon := \mathcal{L}^{\tilde{P}}(X_t^\varepsilon | \tilde{\mathcal{G}}_t) \quad (5.3)$$

satisfies

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} \mathcal{W}_p(\mu_t, \mu_t^\varepsilon) = 0, \tilde{\mathbb{P}} \text{ a.e.}$$

where for each $\varepsilon > 0$, we define $G_\varepsilon(x) := \varepsilon^{-1}G(\varepsilon^{-1}x)$, where $G \in C^\infty(\mathbb{R}^n; \mathbb{R})$ with compact support satisfying $G \geq 0$, $G(x) = G(-x)$ for $x \in \mathbb{R}^n$, and $\int_{\mathbb{R}^n} G(y)dy = 1$, and (recall that $\Lambda_t(\mathbb{Z}_{\mu_t}) = 1$)

$$\begin{aligned} \sigma^\varepsilon(t, \Lambda_t, x)^2 &:= \int_{\mathcal{P}_U^n} \int_U \sigma(m)^2 u^2 m^y(du) \frac{G_\varepsilon(x-y)}{\int_{\mathbb{R}^n} G_\varepsilon(x-z)\mu_t(dz)} \mu_t(dy) \Lambda_t(dm) \\ &= \int_{\mathcal{P}_U^n} \int_U \sigma(m)^2 u^2 m(du, dy) \frac{G_\varepsilon(x-y)}{\int_{\mathbb{R}^n} G_\varepsilon(x-z)m(dz, U)} \Lambda_t(dm). \end{aligned}$$

Notice that, now, when Λ is given, Equation (5.3) or its associated Fokker-Planck equation satisfies a uniqueness result. Indeed, as σ^ε is smooth in x , Equation (5.3) is uniquely solvable.

Next, we are able to find a sequence of \mathcal{P}_U^n -valued $(\sigma\{\Lambda_{t \wedge \cdot}\})_{t \in [0, T]}$ -predictable processes $(\bar{\nu}^k)_{k \in \mathbb{N}^*}$ s.t. $\lim_{k \rightarrow \infty} \delta_{\bar{\nu}^k}(dm)dt = \Lambda \tilde{\mathbb{P}}$ -a.e. If $\mu_t^{\varepsilon, k} = \mathcal{L}^{\tilde{\mathbb{P}}}(X_t^{\varepsilon, k} | \bar{\mathcal{G}}_t)$ is the solution of

$$d\langle f(\cdot - \sigma_0 B_t), \mu_t^{\varepsilon, k} \rangle = \int_{\mathbb{R}^n} f''(x - \sigma_0 B_t) \sigma^\varepsilon(t, \delta_{\bar{\nu}^k}(dm), x)^2 \mu_t^{\varepsilon, k}(dx)dt, \quad (5.4)$$

one has, when $\varepsilon > 0$ is fixed, by passing to the limit in Equation (5.4) and using uniqueness of Equation (5.3), we find that $\lim_{k \rightarrow \infty} \mu_t^{\varepsilon, k} = \mu_t^\varepsilon$ a.e. Consequently, we can set k and ε as fixed, and focus on the approximation of Equation (5.4) or equivalently of

$$dX_t^{\varepsilon, k} = \sigma^\varepsilon(t, \delta_{\bar{\nu}^k}, X_t^{\varepsilon, k})dW_t + \sigma_0 dB_t, \quad X_0^{\varepsilon, k} = \xi \quad (5.5)$$

Strategy of proof: 2-construction of control and discretization Recall that the map $\sigma^\varepsilon(t, \delta_{\bar{\nu}^k}, x)$ satisfies

$$\sigma^\varepsilon(t, \delta_{\bar{\nu}^k}, x)^2 = \int_U \sigma(\bar{\nu}_t^k)^2 u^2 \bar{\nu}_t^k(du, dy) \frac{G_\varepsilon(x-y)}{\int_{\mathbb{R}^n} G_\varepsilon(x-z)\bar{\nu}_t^k(dz, U)}.$$

Let us assume that it is possible to construct a Borel function $\alpha^{\varepsilon, k} : [0, T] \times U \times \mathbb{R}^n \rightarrow U$, a \mathbb{R}^n -valued \mathbb{F} -adapted continuous process $\tilde{X}^{\varepsilon, k}$ and a $[0, 1]$ -valued \mathbb{F} -predictable process F satisfying: F_t and $\tilde{X}_t^{\varepsilon, k}$ are conditionally independent given $\bar{\mathcal{G}}_t$,

$$\mathcal{L}^{\bar{\mathbb{P}}^{\bar{\mathcal{G}}_t}}(\alpha^{\varepsilon, k}(t, F_t, \tilde{X}_t^{\varepsilon, k}) | \tilde{X}_t^{\varepsilon, k} = x) = \int_U \bar{\nu}_t^k(du, dy) \frac{G_\varepsilon(x-y)}{\int_{\mathbb{R}^n} G_\varepsilon(x-z)\bar{\nu}_t^k(dz, U)},$$

and $\tilde{X}^{\varepsilon, k}$ satisfies

$$d\tilde{X}_t^{\varepsilon, k} = \sigma(\bar{\nu}_t^k)\alpha^{\varepsilon, k}(t, F_t, \tilde{X}_t^{\varepsilon, k})dW_t + \sigma_0 dB_t, \quad \tilde{X}_0^{\varepsilon, k} = \xi.$$

Notice that, by uniqueness of Equation (5.4), $\mathcal{L}(\tilde{X}_t^{\varepsilon, k} | \bar{\mathcal{G}}_t) = \mu_t^{\varepsilon, k}$ a.e. for all $t \in [0, T]$. Given $(\alpha^{\varepsilon, k}, \tilde{X}^{\varepsilon, k}, F)$, our last sequence is then given by: $Y^{\varepsilon, k}$ solution of

$$dY_t^{\varepsilon, k} = \sigma(\bar{m}_t^{\varepsilon, k})\alpha^{\varepsilon, k}(t, F_t, \tilde{X}_t^{\varepsilon, k})dW_t + \sigma_0 dB_t, \text{ with } \bar{m}_t^{\varepsilon, k} := \mathcal{L}(Y_t^{\varepsilon, k} | \bar{\mathcal{G}}_t)$$

and $\bar{m}_t^{\varepsilon, k} := \mathcal{L}(Y_t^{\varepsilon, k}, \alpha^{\varepsilon, k}(t, F_t, \tilde{X}_t^{\varepsilon, k}) | \bar{\mathcal{G}}_t)$. By using some technical results, proving in Theorem A.2 and Theorem A.3, we deduce that

$$\lim_{k \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \tilde{\mathbb{P}} \circ (m^{\varepsilon, k}, \delta_{\bar{m}_t^{\varepsilon, k}}(dm)dt, B)^{-1} = \tilde{\mathbb{P}} \circ (\mu, \Lambda, B)^{-1} \text{ in } \mathcal{W}_p.$$

The fact is we are not able to construct the tuple $(\alpha^{\varepsilon,k}, \tilde{X}^{\varepsilon,k}, F)$ as presented below. This construction will be done through approximation by discretization in time in Section 5.3. Moreover, the framework that we will consider in the next part will be more general than the presentation we have chosen for the main results. The reason is that the techniques we use can be applied to both mean field game and mean field control problem (see our companion paper [9]). Therefore, we made the choice to have a presentation that allows the results to be used in both contexts.

5.2 Regularization of the Fokker-Planck equation

In this part, with the help of a regularization by convolution, we show that it can be possible to approximate a particular solution of a Fokker-Planck equation with “non-smooth” coefficients by a sequence of solutions of Fokker-Planck equations with “smooth” coefficients, this part is largely inspired by the proof of [18, Lemma 2.1].

Let $\mathbf{b} \in \mathcal{C}^\ell$, $(\mathbf{n}_t)_{t \in [0, T]}$ and $(\mathbf{z}_t)_{t \in [0, T]}$ belong to $\mathcal{C}_{\mathcal{W}}^n$ and also $\hat{\mathbf{q}}_t(dm, dm') dt \in \mathbb{M}((\mathcal{P}_U^n)^2)$. Moreover, $(\mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}, \mathbf{b})$ satisfy the following equation: $\mathbf{n}_0 = \nu$ and

$$d\langle f(t, \cdot), \mathbf{n}_t \rangle = \left[\langle \partial_t f(t, \cdot), \mathbf{n}_t \rangle + \int_{(\mathcal{P}_U^n)^2} \langle \mathcal{A}_t[f(t, \cdot)](\cdot, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, \cdot), m \rangle \hat{\mathbf{q}}_t(dm, d\bar{\nu}) \right] dt,$$

for all $(t, f) \in [0, T] \times C_b^{1,2}([0, T] \times \mathbb{R}^n)$, where the generator \mathcal{A} is defined by

$$\begin{aligned} & \mathcal{A}_t \varphi(x, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) \\ & := \frac{1}{2} \text{Tr}[\hat{\sigma} \hat{\sigma}^\top(t, x, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) \nabla^2 \varphi(x)] + \hat{b}(t, x, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u)^\top \nabla \varphi(x), \end{aligned} \quad (5.6)$$

with $(\hat{b}, \hat{\sigma}) : [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U \rightarrow \mathbb{R}^n \times \mathbb{S}^n$ is bounded and continuous function in all arguments, and for each $\bar{\nu} \in \mathcal{P}_U^n$, the map $(\hat{b}, \hat{\sigma})(\cdot, \cdot, \mathbf{b}, \cdot, \mathbf{z}, \cdot, \bar{\nu}, \cdot)$ satisfies Assumption 2.1 with constant θ independent of $\bar{\nu}$.

Remark 5.1. As said in the end of Section 5.1, we consider this type of general Fokker-Planck equation because we want to have a formulation useful both in mean field game and mean field control. Here, the mean field game aspect appears in the integration over $d\bar{\nu}$ in $\hat{\mathbf{q}}$ and \mathbf{z} . The integration over $d\bar{\nu}$ in $\hat{\mathbf{q}}$ and \mathbf{z} play the role of fixed measures as it can happen in mean field game.

Let $G \in C^\infty(\mathbb{R}^n; \mathbb{R})$ with compact support satisfying $G \geq 0$, $G(x) = G(-x)$ for $x \in \mathbb{R}^n$, and $\int_{\mathbb{R}^n} G(y) dy = 1$, and define $G_\varepsilon(x) := \varepsilon^{-n} G(\varepsilon^{-1}x)$ and for all $\pi \in \mathcal{P}(\mathbb{R}^n)$, $\pi^{(\varepsilon)}(x) := \int_{\mathbb{R}^n} G_\varepsilon(x - y) \pi(dy)$ for all $x \in \mathbb{R}^n$. Now, for each $\varepsilon > 0$, let us introduce the generator of the regularized Fokker-Planck equation \mathcal{A}^ε : for all $(t, \hat{\mathbf{q}}, x) \in [0, T] \times \mathcal{P}((\mathcal{P}_U^n)^2) \times \mathbb{R}^n$

$$\mathcal{A}_t^\varepsilon \varphi[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}](x) := \frac{1}{2} \text{Tr}[\hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}](t, x) \nabla^2 \varphi(x)] + \hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}](t, x)^\top \nabla \varphi(x), \quad (5.7)$$

where for $(t, x, \gamma, \pi, \beta, m, \bar{\nu}, u) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U$, $\hat{a}(t, x, \gamma, \pi, \beta, m, \bar{\nu}, u) := \hat{\sigma} \hat{\sigma}^\top(t, x, \gamma, \pi, \beta, m, \bar{\nu}, u)$ and $(\hat{a}^\varepsilon, \hat{b}^\varepsilon)$ are defined by:

$$(\hat{a}^\varepsilon, \hat{b}^\varepsilon)[\mathbf{b}, \pi, \beta, q](t, x) \quad (5.8)$$

$$:= \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n} \int_U (a, b)(t, y, \mathbf{b}_{t \wedge \cdot}, \pi_{t \wedge \cdot}, \beta_{t \wedge \cdot}, m, \bar{\nu}, u) \frac{G_\varepsilon(x - y)}{(m(dz, U))^{(\varepsilon)}(x)} m(du, dy) q(dm, d\bar{\nu}). \quad (5.9)$$

We are now ready to formulate our regularization/approximation result of Fokker-Planck equation. The following proposition is proved in Appendix A.1.

Proposition 5.2 (Regularization of Fokker-Planck equation). *Let $\nu \in \mathcal{P}_p(\mathbb{R}^n)$, for each $\varepsilon > 0$, there exists a unique solution $(\mathbf{n}_t^\varepsilon)_{t \in [0, T]} \in C_{\mathcal{W}}^{n, p}$ of: $\mathbf{n}_0^\varepsilon = \nu$ and for all $f \in C_b^{1, 2}([0, T] \times \mathbb{R}^n)$ and*

$$d\langle f(t, \cdot), \mathbf{n}_t^\varepsilon \rangle = \left[\int_{\mathbb{R}^n} \partial_t f(t, y) \mathbf{n}_t^\varepsilon(dy) + \int_{\mathbb{R}^n} \mathcal{A}_t^\varepsilon f(t, \cdot) [\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](t, y) \mathbf{n}_t^\varepsilon(dy) \right] dt. \quad (5.10)$$

Moreover, if $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$ and $\hat{\mathbf{q}}_t(\mathbb{Z}_{\mathbf{n}_t} \times \mathcal{P}_U^n) = 1$ dt-for almost every $t \in [0, T]$, then

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} \mathcal{W}_p(\mathbf{n}_t^\varepsilon, \mathbf{n}_t) = 0. \quad (5.11)$$

Remark 5.3. (i) Let $(\hat{\Omega}, \hat{\mathbb{F}}, \hat{\mathcal{F}}, \mathbb{P})$ be a probability space supporting W a $\hat{\mathbb{F}}$ -Wiener process of dimension \mathbb{R}^n and ξ a \mathcal{F}_0 -random variable such that $\mathcal{L}^{\mathbb{P}}(\xi)(dy) = \nu(dy)$. Given $\varepsilon > 0$, let Y^ε be the unique strong solution (well defined, see Appendix A.1 (more precisely the Proof of Theorem 5.2))

$$dY_t^\varepsilon = \hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t](t, Y_t^\varepsilon)dt + (\hat{a}^\varepsilon)^{1/2}[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t](t, Y_t^\varepsilon)dW_t, Y_0^\varepsilon = \xi, \quad (5.12)$$

one has, by uniqueness of (5.10), $\mathcal{L}^{\mathbb{P}}(Y_t^\varepsilon) = \mathbf{n}_t^\varepsilon$ for all $t \in [0, T]$ where \mathbf{n}^ε is the solution of (5.10).

(ii) We will sometimes use the previous lemma with Theorem A.2, in which \mathbf{n}^ε must be obtainable through a diffusion process that has a volatility term which verifies $\hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](t, Y_t^\varepsilon) \geq \theta I_{n \times n}$. The SDE (5.12) allows to say that \mathbf{n}^ε satisfies these conditions. Also, from Theorem 5.2 and the SDE representation (5.12), it is straightforward to see that the measure $\mathbf{n}_t(dx)dt$ is equivalent to the Lebesgue measure on $\mathbb{R}^n \times [0, T]$ (see for instance Theorem A.1).

Remark 5.4. Combining Theorem 5.3 (diffusion form (5.12) of \mathbf{n}^ε) with Theorem 5.2 (convergence result (5.11)), as (b, σ) are bounded, there exists a constant $C > 0$, depending only of coefficients (b, σ) , p and p' , such that

$$\sup_{r \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \mathbf{n}_r(dx) \leq C \left(1 + \int_{\mathbb{R}^n} |x|^{p'} \nu(dx) \right)$$

and

$$\mathcal{W}_p(\mathbf{n}_s, \mathbf{n}_t)^p \leq C|t - s|, \text{ for all } (t, s) \in [0, T] \times [0, T].$$

5.3 Approximation by N -agents

Now, let us formulate the approximation result of Fokker-Planck equation by N -interacting SDE equations. In order to achieve this, we first describe the associated framework.

Let $(\Omega^{\mathbf{q}}, \mathcal{F}^{\mathbf{q}}, \mathbb{F}^{\mathbf{q}}, \mathbb{Q})$ be a filtered probability space supporting $(B_t)_{t \in [0, T]}$ a \mathbb{R}^ℓ -valued $\mathbb{F}^{\mathbf{q}}$ -adapted continuous process, μ and ζ two $\mathcal{P}(\mathbb{R}^n)$ -valued $\mathbb{F}^{\mathbf{q}}$ -continuous processes, $\bar{\Lambda}$ a $\mathbb{M}((\mathcal{P}_U^n)^2)$ -valued variable such that $(\bar{\Lambda}_t)_{t \in [0, T]}$ is $\mathbb{F}^{\mathbf{q}}$ -predictable. Besides, $(\mu, B, \zeta, \bar{\Lambda})$ satisfy: $\bar{\Lambda}_t(\mathbb{Z}_{\mu_t} \times \mathcal{P}_U^n) = 1$, for $d\mathbb{Q} \otimes dt$ -almost surely, and \mathbb{Q} -a.e.

$$d\langle f, \mu_t \rangle = \int_{\mathcal{P}_U^n \times \mathcal{P}_U^n} \int_{\mathbb{R}^n \times U} \mathcal{A}_t f(y, B, \phi(\mu), \zeta, m, \bar{\nu}, u) m(dy, du) \bar{\Lambda}_t(dm, d\bar{\nu}) dt, \quad \mu_0 = \nu, \quad (5.13)$$

for all $t \in [0, T]$ and $f \in C_b^2(\mathbb{R}^n)$, where

$$\begin{aligned} & \mathcal{A}_t \varphi(x, \mathbf{b}, \pi, \beta, m, \bar{\nu}, u) \\ & := \frac{1}{2} \text{Tr}[\hat{\sigma} \hat{\sigma}^\top(t, x, \mathbf{b}, \pi, \beta, m, \bar{\nu}, u) \nabla^2 \varphi(x)] + \hat{b}(t, x, \mathbf{b}, \pi, \beta, m, \bar{\nu}, u)^\top \nabla \varphi(x), \end{aligned} \quad (5.14)$$

with, as in (5.6), $(\hat{b}, \hat{\sigma})$ is continuous in all arguments and bounded, and the map $(\hat{b}, \hat{\sigma})(\cdot, \cdot, \mathbf{b}, \cdot, \beta, \cdot, \bar{\nu}, \cdot)$ satisfies Assumption 2.1 with constant C and θ independent of $(\mathbf{b}, \beta, \bar{\nu})$ (see Assumption 2.1). Besides, $\phi : \mathcal{C}_{\mathcal{W}}^n \rightarrow \mathcal{C}_{\mathcal{W}}^n$ is a Lipschitz function s.t. for all $t \in [0, T]$, $\phi_t(\pi) = \phi_t(\pi_{t \wedge \cdot})$.

Remark 5.5. (i) Notice that, (5.13) is an equation over μ in the sense that with the condition $\bar{\Lambda}_t(\mathbb{Z}_{\mu_t} \times \mathcal{P}_U^n) = 1$, for $d\mathbb{Q} \otimes dt$ -almost surely, the process μ appears on both sides on the equality. Under general Theorem 2.1, it is not difficult to show that there are processes $(\mu, \bar{\Lambda})$ verifying equation (5.13) (see for instance [11, Theorem A.2]). However, without additional assumptions, a uniqueness result cannot be expected.

(ii) This type of Fokker-Planck equation appears especially in the study of optimal control of McKean-Vlasov equation (see Section 4 above) and mean field game (see [9]). One the most important variable is $\bar{\Lambda}$. It can play the role of control in optimal control of McKean-Vlasov equation, but also of external parameter as it is the case in the mean field game.

Let $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{F}}, \hat{\mathbb{P}})$ be another filtered probability space supporting:

- $(W^i)_{i \in \mathbb{N}^*}$ a sequence of \mathbb{R}^n -valued independent $\hat{\mathbb{F}}$ -Brownian motions and $(\xi^i)_{i \in \mathbb{N}^*}$ a sequence of independent $\hat{\mathcal{F}}_0$ -random variables s.t. $\mathcal{L}^{\hat{\mathbb{P}}}(\xi^i) = \nu_i \in \mathcal{P}_{p'}(\mathbb{R}^n)$,
- $(\mu^N)_{N \in \mathbb{N}^*}$ and $(\zeta^N)_{N \in \mathbb{N}^*}$ two sequences of $\mathcal{P}(\mathbb{R}^n)$ -valued $\hat{\mathbb{F}}$ -adapted continuous processes, and $(B^N)_{N \in \mathbb{N}^*}$ a sequence of \mathbb{R}^ℓ -valued $\hat{\mathbb{F}}$ -adapted continuous processes,
- $(m^N)_{N \in \mathbb{N}^*}$ and $(\bar{\nu}^N)_{N \in \mathbb{N}^*}$ two sequences of \mathcal{P}_U^n -valued $\hat{\mathbb{F}}$ -predictable processes,

satisfying:

$$\lim_{N \rightarrow \infty} \mathcal{W}_{p'} \left(\frac{1}{N} \sum_{i=1}^N \nu^i, \nu \right) = 0 \text{ and } \lim_{N \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}} \left(\phi(\mu^N), \zeta^N, \bar{\Lambda}^N, B^N \right) = \mathcal{L}^{\mathbb{Q}}(\phi(\mu), \zeta, \bar{\Lambda}, B), \tag{5.15}$$

where $\bar{\Lambda}_t^N(dm, d\bar{\nu})dt := \delta_{(m_t^N, \bar{\nu}_t^N)}(dm, d\bar{\nu})dt$. The convergence takes place in \mathcal{W}_p .

Furthermore, let $(Z^i)_{i \in \mathbb{N}^*}$ be a sequence of independent $[0, 1]$ -valued $\hat{\mathcal{F}}$ -measurable uniform variables independent of other variables, and for each $(i, N) \in \mathbb{N}^* \times \mathbb{N}^*$, denote by $\hat{\mathbb{F}}^{i,N} := (\hat{\mathcal{F}}_t^{i,N})_{t \in [0, T]}$ the filtration defined by:

$$\hat{\mathcal{F}}_t^{i,N} := \sigma \left\{ \xi^i, \bar{\Lambda}_{t \wedge \cdot}^N, \phi_{t \wedge \cdot}(\mu^N), \zeta_{t \wedge \cdot}^N, W_{t \wedge \cdot}^i, B_{t \wedge \cdot}^N, Z^i \right\}, \text{ for each } t \in [0, T]. \tag{5.16}$$

The next proposition describes an approximation by a sequence of N -interacting processes of the Fokker-Planck equation (5.13).

Proposition 5.6. *There exists a sequence of processes $(\alpha^{i,N})_{(i,N) \in \mathbb{N}^* \times \mathbb{N}^*}$ satisfying for each $(i, N) \in \mathbb{N}^* \times \mathbb{N}^*$, $\alpha^{i,N}$ is $\hat{\mathbb{F}}^{i,N}$ -predictable, s.t. if we let $(\hat{X}_t^1, \dots, \hat{X}_t^N)_{t \in [0, T]}$ be the continuous processes unique strong solution of: for each $i \in \{1, \dots, N\}$, $\mathbb{E}^{\hat{\mathbb{P}}}[\|\hat{X}^i\|^{p'}] < \infty$, for all $t \in [0, T]$*

$$\begin{aligned} \hat{X}_t^i &= \xi^i + \int_0^t \hat{b}(r, \hat{X}_r^i, B^N, \phi(\hat{\mu}^N), \zeta^N, \hat{m}_r^N, \bar{\nu}_r^N, \alpha_r^{i,N}) dr \\ &\quad + \int_0^t \hat{\sigma}(r, \hat{X}_r^i, B^N, \phi(\hat{\mu}^N), \zeta^N, \hat{m}_r^N, \bar{\nu}_r^N, \alpha_r^{i,N}) dW_r^i, \hat{\mathbb{P}}\text{-a.e.} \end{aligned} \tag{5.17}$$

where $\hat{m}_t^N(dx, du) := \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{X}_t^i, \alpha_t^{i,N})}(dx, du)$, $\hat{\mu}_t^N(dx) := \hat{m}_t^N(dx, U)$, then, one has, for a sub-sequence $(N_k)_{k \in \mathbb{N}^*} \subset \mathbb{N}^*$,

$$\lim_{k \rightarrow \infty} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\hat{m}_t^{N_k}, m_t^{N_k})^p dt + \sup_{t \in [0, T]} \mathcal{W}_p(\phi_t(\hat{\mu}^{N_k}), \phi_t(\mu^{N_k})) \right] = 0$$

and

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{\mu}^{N_k}, \zeta^{N_k}, \widehat{\Lambda}^{N_k}, B^{N_k}) = \mathcal{L}^{\mathbb{Q}}(\mu, \zeta, \bar{\Lambda}, B), \text{ in } \mathcal{W}_p \tag{5.18}$$

with $\widehat{\Lambda}_s^{N_k}(dm, d\bar{\nu})ds := \delta_{(\widehat{m}_s^{N_k}, \widehat{\nu}_s^{N_k})}(dm, d\bar{\nu})ds$.

Remark 5.7. (i) Theorem 5.6 as well as Theorem 5.8 (see below) can be considered as a general characterization of Fokker–Planck equation of type (5.13) via a sequence of SDE processes interacting through the empirical distribution of the states and “controls”. These results are very useful both in the study of extended mean field control problem (see Theorem 4.3) and in mean field game of controls (see our companion paper [9]).

(ii) Because of non–uniqueness of Fokker–Planck equation (5.13), the condition (5.15) is a crucial and essential assumption. Furthermore, notice that, the condition (5.15) does not require any equation verified by the sequence $(\phi(\mu^N), \zeta^N, \bar{\Lambda}^N, B^N)_{N \in \mathbb{N}^*}$. Only the convergence result (5.15) is necessary.

(iii) Observe that, the sequence $(\Lambda^N)_{N \in \mathbb{N}^*}$ is a subset of $\mathbb{M}_0((\mathcal{P}_U^n)^2)$ and not a general subset of $\mathbb{M}((\mathcal{P}_U^n)^2)$. For an understandable and easy presentation, we consider this type of sequence, but a general subset of $\mathbb{M}((\mathcal{P}_U^n)^2)$ is possible (see Proposition 5.9 below).

(iv) The presence of the map ϕ , notably in (5.15), specifies the condition needed on μ for the result. In particular, if ϕ is null, it means that no assumption of convergence towards μ is necessary to find a sequence of SDE processes converging to μ .

Proof of Proposition 5.6. The proof is divided in three steps for a better understanding.

Step 1 : Approximation by regularization of F-P equation: Let $\varepsilon > 0$ and recall that \mathcal{A}^ε is defined in (5.7). For all $\omega \in \Omega^{\mathfrak{q}}$, by Theorem 5.2, there exists a continuous process $(\mu_t^\varepsilon(\omega))_{t \in [0, T]}$ verifying

$$d\langle f, \mu_t^\varepsilon(\omega) \rangle = \int_{\mathbb{R}^n} \mathcal{A}_t^\varepsilon f [B(\omega), \phi(\mu(\omega)), \zeta(\omega), \bar{\Lambda}_t(\omega)](x) \mu_t^\varepsilon(\omega)(dx) dt, \quad \mu_0^\varepsilon = \nu, \tag{5.19}$$

for all $f \in C_b^2(\mathbb{R}^n; \mathbb{R})$ and for \mathbb{Q} -a.e. $\omega \in \Omega^{\mathfrak{q}}$, $\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} \mathcal{W}_p(\mu_t^\varepsilon(\omega), \mu_t(\omega))$. Also, by Theorem A.6, there is a function $\Phi^\varepsilon : \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2) \rightarrow \mathcal{C}_{\mathcal{W}}^n$ such that \mathbb{Q} -a.e. $\omega \in \Omega^{\mathfrak{q}}$

$$\mu_t^\varepsilon(\omega) = \Phi_t^\varepsilon(B_{t \wedge \cdot}(\omega), \phi_{t \wedge \cdot}(\mu(\omega)), \zeta_{t \wedge \cdot}(\omega), \bar{\Lambda}_{t \wedge \cdot}(\omega)), \text{ for all } t \in [0, T]. \tag{5.20}$$

Step 2 : Approximation by discretization: Now, let us define for all $(x, m) \in \mathbb{R}^n \times \mathcal{P}_U^n$, the probability

$$H^\varepsilon(x, m)(du) := \int_{\mathbb{R}^n} m(du, dy) \frac{G_\varepsilon(x - y)}{(m(U, dz))^{(\varepsilon)}(x)}.$$

Recall that $G \in C^\infty(\mathbb{R}^n; \mathbb{R})$ with compact support satisfying $G \geq 0$, $G(x) = G(-x)$ for $x \in \mathbb{R}^n$, and $\int_{\mathbb{R}^n} G(y)dy = 1$. We denoted $G_\varepsilon(x) := \varepsilon^{-n}G(\varepsilon^{-1}x)$ and for all $\pi \in \mathcal{P}(\mathbb{R}^n)$, $\pi^{(\varepsilon)}(x) := \int_{\mathbb{R}^n} G_\varepsilon(x - y)\pi(dy)$ for all $x \in \mathbb{R}^n$. By Blackwell and Dubins [3], there exists a Borel application $N^\varepsilon : (x, m, v) \in \mathbb{R}^n \times \mathcal{P}_U^n \times [0, 1] \rightarrow N^\varepsilon(x, m)(v) \in U$ s.t. for all $(x, m) \in \mathbb{R}^n \times \mathcal{P}_U^n$ and any $[0, 1]$ -valued uniform random variable F ,

$$\widehat{\mathbb{P}} \circ (N^\varepsilon(x, m)(F))^{-1}(du) = H^\varepsilon(x, m)(du).$$

Step 2.1 : Construction of scheme of discretization: Let us consider the partition $(t_k^N)_{1 \leq k \leq 2^N}$ with $t_k^N = \frac{kT}{2^N}$, and take a sequence of \mathbb{R}^n -valued independent Brownian motions $(Z^i)_{i \in \mathbb{N}^*}$, independent of all of other variables. Let $\varphi : [0, T] \times \mathbb{R}^n \rightarrow [0, 1]$ be a Borel function such that, for all $t \in [0, T]$, $\mathcal{L}^{\widehat{\mathbb{P}}}(\varphi(t - t_k^N, Z_t^i - Z_{t_k^N}^i))$ is the uniform law when

Extended mean field control problem: a propagation of chaos result

$t > t_k^N$. For all $i \in \{1, \dots, N\}$, denote by $V_t^{i,N} := \varphi(t - t_k^N, Z_t^i - Z_{t_k^N}^i)$, when $t \in [t_k^N, t_{k+1}^N)$, and given $\varepsilon > 0$, we define on $(\widehat{\Omega}, \widehat{\mathbb{F}}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$, by Euler scheme, $X^{\varepsilon,i,N} := X^i$ as follows: $X_0^i := \xi^i$ and for all $t \in [0, T]$, $i \in \{1, \dots, N\}$,

$$X_t^i = X_0^i + \int_0^t \widehat{B}\left(s, X_{[s]^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{[s]^N}^i, m_s^N)(V_s^{i,N})\right) ds + \int_0^t \widehat{\Sigma}\left(s, X_{[s]^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{[s]^N}^i, m_s^N)(V_s^{i,N})\right) dW_s^i, \quad (5.21)$$

where $[s]^N = t_k^N$ if $t_k^N \leq s < t_{k+1}^N$, and, for $s \in [t_k^N, t_{k+1}^N)$,

$$\begin{aligned} & \widehat{B}\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{t_k^N}^i, m_s^N)(V_s^{i,N})\right) \\ & := \widehat{b}\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{t_k^N}^i, m_s^N)(V_s^{i,N})\right) \\ & \quad + B\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N\right), \end{aligned}$$

and

$$\begin{aligned} & \widehat{\Sigma}\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{t_k^N}^i, m_s^N)(V_s^{i,N})\right) \\ & := \Sigma\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N\right) \\ & \quad \times \widehat{\sigma}\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, N^\varepsilon(X_{t_k^N}^i, m_s^N)(V_s^{i,N})\right), \end{aligned}$$

with

$$\begin{aligned} & B\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N\right) \\ & := \left[\widehat{b}^\varepsilon[B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_s^N](s, X_{t_k^N}^i) \right. \\ & \quad \left. - \int_U \widehat{b}(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, u) H^\varepsilon(X_{t_k^N}^i, m_s^N)(du) \right] \end{aligned}$$

and

$$\begin{aligned} & \Sigma\left(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N\right) \\ & := \left[\widehat{a}^\varepsilon[B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_s^N](s, X_{t_k^N}^i)^{1/2} \right. \\ & \quad \left. \times \left(\int_U \widehat{a}(s, X_{t_k^N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N, u) H^\varepsilon(X_{t_k^N}^i, m_s^N)(du) \right)^{-1/2} \right], \quad (5.22) \end{aligned}$$

recall that $\bar{\Lambda}_s^N(dm, d\bar{\nu})ds := \delta_{m_s^N}(dm)\delta_{\bar{\nu}_s^N}(d\bar{\nu})ds$.

Notice that, there exists a Borel function $F^N : \mathbb{R}^n \times \mathbb{M}((\mathcal{P}_U^n)^2) \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}^n \times \mathcal{C}^n \times \mathcal{C}^\ell \rightarrow \mathcal{C}^n$ s.t. for each $i \in \{1, \dots, N\}$,

$$X_t^i = F_t^N(\xi^i, \bar{\Lambda}_{t\wedge\cdot}^N, \phi_{t\wedge\cdot}(\mu^N), \zeta_{t\wedge\cdot}^N, W_{t\wedge\cdot}^i, Z_{t\wedge\cdot}^i, B_{t\wedge\cdot}^N), \text{ for all } t \in [0, T], \widehat{\mathbb{P}}\text{-a.e.} \quad (5.23)$$

Step 2.2 : Compactness and identification of the limit: At this stage, we want to show a compactness result and identify the limit of a certain sequence of probability measures constructed from the SDE process (X^1, \dots, X^N) .

Using the assumptions imposed on coefficients $(\hat{b}, \hat{\sigma})$ (see the definition of the generator \mathcal{A} in (5.14)), especially the fact that $\hat{\sigma}\hat{\sigma}^\top \geq \theta I_n$ and $(\hat{b}, \hat{\sigma})$ are bounded, one has that $[\hat{B}, \hat{\Sigma}]$ are bounded and there exists a constant $D > 0$ such that for all ε and N

$$\sup_{i \in \{1, \dots, N\}} \mathbb{E}^{\hat{\mathbb{P}}} \left[|X_t^{\varepsilon, i, N} - X_s^{\varepsilon, i, N}|^p \right] \leq D|t - s|, \text{ for all } (t, s) \in [0, T] \times [0, T]. \quad (5.24)$$

Moreover, by using the fact that $\sup_{N \geq 1} \frac{1}{N} \sum_{i=1}^N \int_{\mathbb{R}^n} |x|^{p'} \nu^i(dx) < \infty$ (see condition (5.15)), it is straightforward to verify that: $\sup_{N \geq 1} \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\hat{\mathbb{P}}} \left[\sup_{t \in [0, T]} |X_t^{\varepsilon, i, N}|^{p'} \right] < \infty$. Then, by [7, Proposition A.2] or/and [7, Proposition-B.1], for each $\varepsilon > 0$, the sequence $(P^N)_{N \in \mathbb{N}^*}$ is relatively compact in \mathcal{W}_p , where

$$P^N := \hat{\mathbb{P}} \circ \left(\vartheta^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}^N, B^N \right)^{-1} \in \mathcal{P} \left(\mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2) \times \mathcal{C}^\ell \right)$$

with $\vartheta_t^N(dx) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{\varepsilon, i, N}}(dx)$.

Let us identify the limit of any convergent sub-sequence of $(P^N)_{N \in \mathbb{N}^*}$. For sake of clarity, we use the notation X^i instead of $X^{\varepsilon, i, N}$. Recall that for the time being $\varepsilon > 0$ is considered as fixed.

For each $N \in \mathbb{N}^*$, $i \in \{1, \dots, N\}$, and $(s, u) \in [0, T] \times U$, let

$$[\hat{b}_s^{\varepsilon, i, N}, \hat{a}_s^{\varepsilon, i, N}] := [\hat{b}^\varepsilon, \hat{a}^\varepsilon][B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_s^N](s, X_{[s]N}^i)$$

and

$$[\hat{b}_s^{i, N}, \hat{a}_s^{i, N}, \hat{B}_s^{i, N}, \hat{\Sigma}_s^{i, N}, \hat{A}_s^{i, N}](u) := [\hat{b}, \hat{a}, \hat{B}, \hat{\Sigma}, \hat{\Sigma}\hat{\Sigma}^\top](s, X_{[s]N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{v}_s^N, u).$$

By Itô's formula, for all $f \in C_b^\infty(\mathbb{R}^n)$ and $t \in [0, T]$

$$\begin{aligned} & \langle f, \vartheta_t^N \rangle \\ &= \langle f, \vartheta_0^N \rangle + \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(X_s^i) \hat{\Sigma}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) dW_s^i \\ & \quad + \frac{1}{N} \sum_{i=1}^N \int_0^t \left[\nabla f(X_s^i) \hat{B}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) \right. \\ & \quad \quad \quad \left. + \frac{1}{2} \text{Tr} \left[\hat{A}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) \nabla^2 f(X_s^i) \right] \right] ds \\ &= \langle f, \vartheta_0^N \rangle + \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(X_s^i) \hat{\Sigma}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) dW_s^i \\ & \quad + \frac{1}{N} \sum_{i=1}^N \int_0^t \left[\nabla f(X_{[s]N}^i) \hat{B}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) \right. \\ & \quad \quad \quad \left. + \frac{1}{2} \text{Tr} \left[\hat{A}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) \nabla^2 f(X_{[s]N}^i) \right] \right] ds \\ & \quad + \frac{1}{N} \sum_{i=1}^N \int_0^t [\nabla f(X_s^i) - \nabla f(X_{[s]N}^i)] \hat{B}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) ds \\ & \quad \quad + \frac{1}{N} \sum_{i=1}^N \int_0^t \frac{1}{2} \text{Tr} \left[\hat{A}_s^{i, N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i, N})) [\nabla^2 f(X_s^i) - \nabla^2 f(X_{[s]N}^i)] \right] ds. \end{aligned}$$

Observe that, for $s \in (t_k^N, t_{k+1}^N)$, for each $i \neq j$, $[\widehat{B}]_s^{i,j} = [\widehat{A}]_s^{i,j} = 0$, where

$$[\widehat{B}]_s^{i,j} := \mathbb{E}^{\widehat{\mathbb{P}}} \left[\nabla f(X_{[s]N}^i) \left\{ \widehat{B}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) - \widehat{b}_s^{\varepsilon,i,N} \right\} \right. \\ \left. \nabla f(X_{[s]N}^j) \left\{ \widehat{B}_s^{j,N} (N^\varepsilon(X_{[s]N}^j, m_s^N)(V_s^{j,N})) - \widehat{b}_s^{\varepsilon,j,N} \right\} \right]$$

and

$$[\widehat{A}]_s^{i,j} := \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left\{ \widehat{A}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) - \widehat{a}_s^{\varepsilon,i,N} \right\} \nabla^2 f(X_{[s]N}^i) \right. \\ \left. \left\{ \widehat{A}_s^{j,N} (N^\varepsilon(X_{[s]N}^j, m_s^N)(V_s^{j,N})) - \widehat{a}_s^{\varepsilon,j,N} \right\} \nabla^2 f(X_{[s]N}^j) \right].$$

Indeed, by using the fact that: for all $(x, m, e) \in \mathbb{R}^n \times \mathcal{P}_U^n \times \{1, \dots, N\}$, one has $\widehat{\mathbb{P}} \circ (N^\varepsilon(x, m)(V_s^{e,K}))^{-1}(du) = H^\varepsilon(x, m)(du)$, and (V_s^i, V_s^j) are independent and independent of other variables, one has

$$[\widehat{B}]_s^{i,j} = \mathbb{E}^{\widehat{\mathbb{P}}} \left[\nabla f(X_{[s]N}^i) \left\{ \widehat{b}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) - \int_U \widehat{b}_s^{i,N}(u) H^\varepsilon(X_{[s]N}^i, m_s^N)(du) \right\} \right. \\ \left. \nabla f(X_{[s]N}^j) \left\{ \widehat{b}_s^{j,N} (N^\varepsilon(X_{[s]N}^j, m_s^N)(V_s^{j,N})) - \int_U \widehat{b}_s^{j,N}(u) H^\varepsilon(X_{[s]N}^j, m_s^N)(du) \right\} \right] = 0. \quad (5.25)$$

By similar way, if we denote by $\Sigma_s^{i,N} := \Sigma(s, X_{[s]N}^i, B^N, \phi(\mu^N), \zeta^N, m_s^N, \bar{\nu}_s^N)$, one finds

$$[\widehat{A}]_s^{i,j} = \mathbb{E}^{\widehat{\mathbb{P}}} \left[\nabla^2 f(X_{[s]N}^i) \left\{ \Sigma_s^{i,N} \widehat{a}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) (\Sigma_s^{i,N})^\top - \widehat{a}_s^{\varepsilon,i,N} \right\} \right. \\ \left. \nabla^2 f(X_{[s]N}^j) \left\{ \Sigma_s^{j,N} \widehat{a}_s^{j,N} (N^\varepsilon(X_{[s]N}^j, m_s^N)(V_s^{j,N})) (\Sigma_s^{j,N})^\top - \widehat{a}_s^{\varepsilon,j,N} \right\} \right] \\ = \mathbb{E}^{\widehat{\mathbb{P}}} \left[\nabla^2 f(X_{[s]N}^i) \left\{ \Sigma_s^{i,N} \int_U \widehat{a}_s^{i,N}(u) H^\varepsilon(X_{[s]N}^i, m_s^N)(du) (\Sigma_s^{i,N})^\top - \widehat{a}_s^{\varepsilon,i,N} \right\} \right. \\ \left. \nabla^2 f(X_{[s]N}^j) \left\{ \Sigma_s^{j,N} \int_U \widehat{a}_s^{j,N}(u) H^\varepsilon(X_{[s]N}^j, m_s^N)(du) (\Sigma_s^{j,N})^\top - \widehat{a}_s^{\varepsilon,j,N} \right\} \right] \\ = \mathbb{E}^{\widehat{\mathbb{P}}} \left[\nabla^2 f(X_{[s]N}^i) \left\{ \widehat{a}_s^{\varepsilon,i,N} - \widehat{a}_s^{\varepsilon,i,N} \right\} \nabla^2 f(X_{[s]N}^j) \left\{ \widehat{a}_s^{\varepsilon,j,N} - \widehat{a}_s^{\varepsilon,j,N} \right\} \right] = 0. \quad (5.26)$$

By simple calculations,

$$\langle f, \vartheta_t^N \rangle - \langle f, \vartheta_0^N \rangle - \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f[B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_r^N](x) \vartheta_{[r]N}^N(dx) dr \\ = \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(X_{[s]N}^i) \widehat{\Sigma}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) dW_s^i \\ + \int_0^t \left[\nabla f(X_{[s]N}^i) \left\{ \widehat{B}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) - \widehat{b}_s^{\varepsilon,i,N} \right\} \right. \\ \left. + \frac{1}{2} \text{Tr} \left[\left\{ \widehat{A}_s^{i,N} (N^\varepsilon(X_{[s]N}^i, m_s^N)(V_s^{i,N})) - \widehat{a}_s^{\varepsilon,i,N} \right\} \nabla^2 f(s, X_{[s]N}^i) \right] \right]$$

$$\begin{aligned}
 & + [\nabla f(X_s^i) - \nabla f(X_{[s]_N}^i)] \widehat{B}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) \\
 & + \frac{1}{2} \text{Tr} \left[\widehat{A}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) [\nabla^2 f(X_s^i) - \nabla^2 f(X_{[s]_N}^i)] \right] ds,
 \end{aligned}$$

consequently, there exists a constant $C > 0$ (independent of N) such that

$$\begin{aligned}
 & \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \langle f, \vartheta_t^N \rangle - \langle f, \vartheta_0^N \rangle - \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f [B^N, \phi(\mu^N), \zeta^N, \overline{\Lambda}_r^N](x) \vartheta_{[r]_N}^N(dx) dr \right|^2 \right] \\
 & \leq C \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(X_{[s]_N}^i) \widehat{\Sigma}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) dW_s^i \right|^2 \right] \right. \\
 & \quad \left. + \int_0^t \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \frac{1}{N} \sum_{i=1}^N \nabla f(X_{[s]_N}^i) \left\{ \widehat{B}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) - \widehat{b}_s^{\varepsilon,i,N} \right\} \right|^2 \right] ds \right. \\
 & \quad \left. + \int_0^t \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \text{Tr} \left[\left\{ \widehat{A}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) - \widehat{a}_s^{\varepsilon,i,N} \right\} \nabla^2 f(X_{[s]_N}^i) \right] \right|^2 \right] ds \right. \\
 & \quad \left. + \int_0^t \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| [\nabla f(X_s^i) - \nabla f(X_{[s]_N}^i)] \right|^2 + \left| \frac{1}{2} [\nabla^2 f(X_s^i) - \nabla^2 f(X_{[s]_N}^i)] \right|^2 \right] ds \right).
 \end{aligned}$$

By successively applying the results (5.25) and (5.26), and inequality (5.24), one gets a constant $M > 0$ depending on (f, b, σ) (which changes from line to line) s.t.

$$\begin{aligned}
 & \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \langle f, \vartheta_t^N \rangle - \langle f, \vartheta_0^N \rangle - \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f [B^N, \phi(\mu^N), \zeta^N, \overline{\Lambda}_r^N](x) \vartheta_{[r]_N}^N(dx) dr \right|^2 \right] \\
 & \leq M \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^t \nabla f(X_s^i) \widehat{\Sigma}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) dW_s^i \right|^2 \right] + \frac{1}{2N} + \frac{1}{N} \right) \\
 & \leq M \left(\frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| \nabla f(X_s^i) \widehat{\Sigma}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) \right|^2 ds \right] + \frac{1}{2N} + \frac{1}{N} \right). \tag{5.27}
 \end{aligned}$$

Remark that as ∇f and $\widehat{\Sigma}$ are bounded,

$$\frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| \nabla f(X_s^i) \widehat{\Sigma}_s^{i,N} (N^\varepsilon(X_{[s]_N}^i, m_s^N)(V_s^{i,N})) \right|^2 ds \right] \leq M \frac{1}{N}. \tag{5.28}$$

Thanks to inequality (5.24), it is straightforward to verify that

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} \mathcal{W}_p \left(\mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta^N, \vartheta^N, \phi(\mu^N), \zeta^N, \overline{\Lambda}^N, B^N), \mathcal{L}^{\widehat{\mathbb{P}}}(\vartheta^N, (\vartheta_{[t]_N}^N)_{t \in [0, T]}, \phi(\mu^N), \zeta^N, \overline{\Lambda}^N, B^N) \right) \\
 & = 0. \tag{5.29}
 \end{aligned}$$

Let $\mathbb{P}^\infty \in \mathcal{P}(\mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2) \times \mathcal{C}^\ell)$ be the limit of any sub-sequence $(\mathbb{P}^{N_k})_{k \in \mathbb{N}^*}$ of $(\mathbb{P}^N)_{N \in \mathbb{N}^*}$, and denote by $(\beta^\vartheta, \beta^\mu, \beta^\zeta, \overline{\beta}, B)$ the canonical process on $\mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2) \times \mathcal{C}^\ell$. By combining inequalities (5.27) and (5.28) with the result (5.29), by passing to the limit, using continuity of coefficients, given $\varepsilon > 0$: for all $(t, f) \in [0, T] \times C_b^\infty(\mathbb{R}^n)$

$$\begin{aligned}
 & \lim_k \mathbb{E}^{\widehat{\mathbb{P}}} \left[\left| \langle f, \vartheta_t^{N_k} \rangle - \langle f, \vartheta_0^{N_k} \rangle - \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f [B^{N_k}, \phi(\mu^{N_k}), \zeta^{N_k}, \overline{\Lambda}_r^{N_k}](x) \vartheta_{[r]^{N_k}}^{N_k}(dx) dr \right|^2 \right] \\
 & = \mathbb{E}^{\mathbb{P}^\infty} \left[\left| \langle f, \beta_t^\vartheta \rangle - \langle f, \nu \rangle - \int_0^t \left[\int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f [B, \beta^\mu, \beta^\zeta, \overline{\beta}_r](x) \beta_r^\vartheta(dx) \right] dr \right|^2 \right] = 0.
 \end{aligned}$$

Therefore, after taking a countable family of (f, t) , one gets: for all $(t, f) \in [0, T] \times C_b^\infty(\mathbb{R}^n)$

$$\langle f, \beta_t^\vartheta \rangle = \langle f, \nu \rangle + \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f[B, \beta^\mu, \beta^\zeta, \bar{\beta}_r](x) \beta_r^\vartheta(dx) dr, \text{ for all } t \in [0, T], \text{ P}^\infty\text{-a.e.}$$

from this equality, we can show the previous equality holds true for all $f \in C_b^2(\mathbb{R}^n)$. For each $\varepsilon > 0$, by uniqueness $\beta^\vartheta := \Phi^\varepsilon(B, \beta^\mu, \beta^\zeta, \bar{\beta})$ with $\Phi^\varepsilon : \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbf{M}((\mathcal{P}_U^n)^2) \rightarrow \mathcal{C}_{\mathcal{W}}^n$ a Borel function used in (5.20). Notice that, by assumptions (5.15),

$$\text{P}^\infty \circ (\beta^\mu, \beta^\zeta, \bar{\beta}, B)^{-1} = \lim_k \widehat{\text{P}} \circ (\phi(\mu^{N_k}), \zeta^{N_k}, \bar{\Lambda}^{N_k}, B^{N_k})^{-1} = \mathbb{Q} \circ (\phi(\mu), \zeta, \bar{\Lambda}, B)^{-1} \text{ in } \mathcal{W}_p.$$

This result is enough to deduce that $\text{P}^\infty = \mathbb{Q} \circ (\mu^\varepsilon, \phi(\mu), \zeta, \bar{\Lambda}, B)^{-1}$. This is true for any limit P^∞ for any sub-sequence of $(\text{P}^N)_{N \in \mathbb{N}^*}$, therefore

$$\lim_{N \rightarrow \infty} \widehat{\text{P}} \circ (\vartheta^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}^N, B^N)^{-1} = \mathbb{Q} \circ (\mu^\varepsilon, \phi(\mu), \zeta, \bar{\Lambda}, B)^{-1} \text{ in } \mathcal{W}_p. \quad (5.30)$$

Step 3 : Last approximation: To finish, now, let us define $\widehat{X}^{\varepsilon, i, N} := \widehat{X}^i$ the strong solution of: $\widehat{X}_0^i = \xi^i$, for all $t \in [0, T]$,

$$d\widehat{X}_t^i = \hat{b}(t, \widehat{X}_t^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_t^N, \bar{v}_t^N, \alpha_t^i) dt + \hat{\sigma}(t, \widehat{X}_t^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_t^N, \bar{v}_t^N, \alpha_t^i) dW_t^i,$$

where $\alpha_t^i := N^\varepsilon(X_{t_k^N}^i, m_{t_k^N}^N)(V_t^{i, N})$ for all $t \in [t_k^N, t_{k+1}^N]$, we denote also $\widehat{m}_t^N(dx, du) := \frac{1}{N} \sum_{i=1}^N \delta_{(\widehat{X}_t^i, \alpha_t^i)}(dx, du)$ and $\widehat{\mu}_t^N(dx) := \widehat{m}_t^N(dx, U)$, recall that (X^1, \dots, X^N) are defined in (5.21). It is straightforward to check that: there exists a constant $D > 0$ (independent of ε and N)

$$\sup_{i \in \{1, \dots, N\}} \mathbb{E}^{\widehat{\text{P}}} \left[|\widehat{X}_t^i - \widehat{X}_s^i|^p \right] \leq D|t - s|, \text{ for all } (t, s) \in [0, T] \times [0, T]. \quad (5.31)$$

By Bukholder–Davis–Gundy inequality, lipschitz property of coefficients and previous inequality (5.31),

$$\begin{aligned} & \mathbb{E}^{\widehat{\text{P}}} \left[\sup_{s \in [0, t]} |\widehat{X}_s^i - X_s^i|^p \right] \\ & \leq \hat{D} \left(\mathbb{E}^{\widehat{\text{P}}} \left[\int_0^t \left| \hat{b}(r, \widehat{X}_r^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \bar{v}_r^N, \alpha_r^i) \right. \right. \right. \\ & \quad \left. \left. \left. - \widehat{B}(r, X_{[r]N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \bar{v}_r^N, N^\varepsilon(X_{[r]N}^i, m_r^N)(V_r^{i, N})) \right|^p dr \right] \right. \\ & \quad \left. + \mathbb{E}^{\widehat{\text{P}}} \left[\int_0^t \left| \hat{\sigma}(r, \widehat{X}_r^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \bar{v}_r^N, \alpha_r^i) \right. \right. \right. \\ & \quad \left. \left. \left. - \widehat{\Sigma}(r, X_{[r]N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \bar{v}_r^N, N^\varepsilon(X_{[r]N}^i, m_r^N)(V_r^{i, N})) \right|^p dr \right] \right) \\ & \leq \hat{D} \left(\mathbb{E}^{\widehat{\text{P}}} \left[\int_0^t \left| \hat{b}(r, \widehat{X}_r^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \bar{v}_r^N, \alpha_r^i) \right. \right. \right. \\ & \quad \left. \left. \left. - \hat{b}(r, X_{[r]N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \bar{v}_r^N, \alpha_r^i) \right|^p \right] \right. \\ & \quad \left. + \mathbb{E}^{\widehat{\text{P}}} \left[\int_0^t \left| \hat{\sigma}(r, \widehat{X}_r^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \bar{v}_r^N, \alpha_r^i) \right. \right. \right. \\ & \quad \left. \left. \left. - \widehat{\Sigma}(r, X_{[r]N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \bar{v}_r^N, N^\varepsilon(X_{[r]N}^i, m_r^N)(V_r^{i, N})) \right|^p dr \right] \right) \end{aligned}$$

$$\begin{aligned}
 & + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| \widehat{\sigma}(r, \widehat{X}_r^i, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \widehat{\nu}_r^N, \alpha_r^i) \right. \right. \\
 & \quad \left. \left. - \widehat{\sigma}(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N, \alpha_r^i) \right|^p dr \right] \\
 & + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| \widehat{\sigma}(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N, \alpha_r^i) \right. \right. \\
 & \quad \left. \left. - \widehat{\Sigma}(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N, N^\varepsilon(X_{[r]_N}^i, m_r^N)(V_r^{i,N})) \right|^p dr \right] \Bigg) \\
 & \leq \widehat{D} \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| \widehat{b}^\varepsilon[B^N, \phi(\mu^N), \zeta^N, \overline{\Lambda}_r^N](r, X_{[r]_N}^i) \right. \right. \right. \\
 & \quad \left. \left. - \int_U \widehat{b}(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N, u) H^\varepsilon(m^N, X_{[s]_N}^i)(du) \right|^p dr \right] \right. \\
 & \quad + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^t \left| 1 - \Sigma(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N) \right|^p dr \right. \\
 & \quad \left. + \int_0^t \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[r]_N}^i, \alpha_r^i)}(dx, du), m_r^N(dx, du) \right)^p dr \right] \\
 & \quad \left. + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\vartheta^N), \phi_{e'}(\mu^N)) + \int_0^t \sup_{e \in [0, r]} |\widehat{X}_e^i - X_e^i|^p dr + \frac{1}{2^N} \right], \right.
 \end{aligned}$$

then by Gronwall lemma

$$\mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\widehat{X}_t^i - X_t^i|^p \right] \leq \widehat{D} \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\vartheta^N), \phi_{e'}(\mu^N)) \right] \right) + \frac{1}{2^N} + E^{\varepsilon, i, N} + C^{\varepsilon, N}$$

where $C^{\varepsilon, N} := \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[r]_N}^i, \alpha_r^i)}(dx, du), m_r^N(dx, du) \right)^p dr \right]$, and

$$\begin{aligned}
 & E^{\varepsilon, i, N} \\
 & := \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \left| \widehat{b}^\varepsilon, \widehat{a}^\varepsilon[B^N, \phi(\mu^N), \zeta^N, \overline{\Lambda}_r^N](r, X_{[r]_N}^i) \right. \right. \\
 & \quad \left. \left. - \int_U \widehat{b}, \widehat{a}(r, X_{[r]_N}^i, B^N, \phi(\mu^N), \zeta^N, m_r^N, \nu_r^N, u) H^\varepsilon(X_{[r]_N}^i, m_r^N)(du) \right|^p dr \right].
 \end{aligned}$$

Firstly, thanks to results (5.30) and the approximation realized in (5.19), one gets

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\vartheta^N), \phi_{e'}(\mu^N)) \right] = \lim_{\varepsilon \rightarrow 0} \mathbb{E}^{\mathbb{Q}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\mu^\varepsilon), \phi_{e'}(\mu)) \right] = 0. \tag{5.32}$$

Secondly, after calculations, it is straightforward to deduce that

$$\begin{aligned}
 & \frac{1}{N} \sum_{i=1}^N E^{\varepsilon, i, N} = \\
 & \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \int_{\mathbb{R}^n} \int_{(\mathcal{P}_V^n)^2} \right. \\
 & \quad \left. \left| \int_{U \times \mathbb{R}^n} \widehat{b}, \widehat{a}(r, y, B^N, \phi(\mu^N), \zeta^N, m, \bar{\nu}, u) \frac{G_\varepsilon(x - y)}{(m(U, dz))^{(\varepsilon)}(x)} m(du, dy) \right|^p \right]
 \end{aligned}$$

$$- \int_{U \times \mathbb{R}^n} [\hat{b}, \hat{a}](r, x, B^N, \phi(\mu^N), \zeta^N, m, \bar{\nu}, u) \frac{G_\varepsilon(x-y)}{(m(U, dz))^{(\varepsilon)}(x)} m(du, dy) \Big|^p \bar{\Lambda}_r^N(dm, d\bar{\nu}) \vartheta_{[r]^N}^N(dx) dr \Big].$$

By regularity of coefficients (Assumption 2.1 and $(\hat{b}, \hat{\sigma})$ bounded), the results (5.30) and (5.29) allow to get

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E^{\varepsilon, i, N} \\ & \leq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \int_{\mathbb{R}^n} \int_{(\mathcal{P}_U^n)^2} \int_{U \times \mathbb{R}^n} \left| [\hat{b}, \hat{a}](r, y, B, \phi(\mu), \zeta, m, \bar{\nu}, u) \right. \right. \\ & \quad \left. \left. - [\hat{b}, \hat{a}](r, x, B, \phi(\mu), \zeta, m, \bar{\nu}, u) \right|^p \frac{G_\varepsilon(x-y)}{(m(U, dz))^{(\varepsilon)}(x)} m(du, dy) \bar{\Lambda}_r(dm, d\bar{\nu}) \mu_r^\varepsilon(dx) dr \right], \end{aligned}$$

then, by Theorem A.3, $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E^{\varepsilon, i, N} = 0$.

Next, let us define the variable

$$\Upsilon_r^N(de', de) dr := \mathbb{E}^{\hat{\mathbb{P}}} \left[\delta_{(\bar{m}_r^N, m_r^N)}(de', de) dr \right] \in \mathbb{M}((\mathcal{P}_U^n)^2),$$

where $\bar{m}_r^N(dx, du) := \frac{1}{N} \sum_{i=1}^N \delta_{(X_{[r]^N}^i, \alpha_r^i)}(dx, du)$. It is easy to check that the sequence $(\Upsilon^N)_{N \in \mathbb{N}^*}$ is relatively compact for the Wasserstein metric \mathcal{W}_p . Denote by Υ^∞ the limit of a sub-sequence $(\Upsilon^{N_k})_{k \in \mathbb{N}^*}$. Let $Q \in \mathbb{N}^*$, $(f^q)_{q \in \{1, \dots, Q\}} : \mathbb{R}^n \times U \rightarrow \mathbb{R}^Q$ be bounded continuous functions and $g : [0, T] \times \mathcal{P}_U^n \rightarrow \mathbb{R}$. One has

$$\begin{aligned} & \int_0^T \int_{(\mathcal{P}_U^n)^2} \prod_{q=1}^Q \langle f^q, e' \rangle g(t, e) \Upsilon_t^\infty(de', de) dt \\ & = \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \int_{\mathcal{P}_U^n} \prod_{q=1}^Q \int_{\mathbb{R}^n} \langle f^q(x, \cdot), H^\varepsilon(x, m) \rangle \mu_t^\varepsilon(dx) g(t, m) \bar{\Lambda}_t(dm, \mathcal{P}_U^n) dt \right]. \end{aligned}$$

We prove this equality when $Q = 2$, the case $Q \in \mathbb{N}^*$ follows immediately. Indeed,

$$\begin{aligned} & \int_0^T \int_{(\mathcal{P}_U^n)^2} \langle f^1, e' \rangle \langle f^2, e \rangle g(t, e) \Upsilon_t^\infty(de', de) dt \\ & = \lim_k \frac{1}{N_k} \frac{1}{N_k} \sum_{i,j=1}^{N_k} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f^1(X_{[t]^{N_k}}^i, \alpha_t^i) f^2(X_{[t]^{N_k}}^j, \alpha_t^j) g(t, m_t^{N_k}) dt \right] \\ & = \lim_k \left(\frac{1}{N_k} \frac{1}{N_k} \sum_{i \neq j} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \int_U f^1(X_{[t]^{N_k}}^i, u) H^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(du) \right. \right. \\ & \quad \left. \left. \int_U f^2(X_{[t]^{N_k}}^j, u) H^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(du) g(t, m_t^{N_k}) dt \right] \right. \\ & \quad \left. + \frac{1}{N_k} \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f^1(X_{[t]^{N_k}}^i, N^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(V_t^{i, N_k})) \right. \right. \\ & \quad \left. \left. f^2(X_{[t]^{N_k}}^j, N^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(V_t^{i, N_k})) g(t, m_t^{N_k}) dt \right] \right) \end{aligned}$$

$$\begin{aligned}
 &= \lim_k \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \int_{\mathbb{R}^n} \int_U f^1(x, u) H^\varepsilon(x, m_t^{N_k})(du) \vartheta_{[t]^{N_k}}^{N_k}(dx) \right. \right. \\
 &\quad \left. \int_{\mathbb{R}^n} \int_U f^2(y, u) H^\varepsilon(y, m_t^{N_k})(du) \vartheta_{[t]^{N_k}}^{N_k}(dy) g(t, m_t^{N_k}) dt \right] \\
 &\quad - \frac{1}{N_k} \frac{1}{N_k} \sum_{i=1} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \int_U f^1(X_{[t]^{N_k}}^i, u) H^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(du) \right. \\
 &\quad \left. \int_U f^2(X_{[t]^{N_k}}^i, u) H^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(du) g(t, m_t^{N_k}) dt \right] \\
 &\quad + \frac{1}{N_k} \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T f^1(X_{[t]^{N_k}}^i, N^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(V_t^{i, N_k})) \right. \\
 &\quad \left. f^2(X_{[t]^{N_k}}^i, N^\varepsilon(X_{[t]^{N_k}}^i, m_t^{N_k})(V_t^{i, N_k})) g(t, m_t^{N_k}) dt \right] \Big) \\
 &= \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \int_{\mathcal{P}_U^n} \int_{\mathbb{R}^n} \int_U f^1(x, u) H^\varepsilon(x, m)(du) \mu_t^\varepsilon(dx) \right. \\
 &\quad \left. \int_{\mathbb{R}^n} \int_U f^2(y, u) H^\varepsilon(y, m)(du) \mu_t^\varepsilon(dy) g(t, m) \bar{\Lambda}_t(dm, \mathcal{P}_U^n) dt \right],
 \end{aligned}$$

where the fourth equality is true because of the same argument used in (5.25) and (5.26), i.e. for all $(s, v) \in (t_k^{N_i}, t_{k+1}^{N_i}) \times \{1, \dots, N_i\}$, $\widehat{\mathbb{P}} \circ (N^\varepsilon(x, m)(V_s^{v, N_i}))^{-1}(du) = H^\varepsilon(x, m)(du)$, and for $i \neq j$ (V_s^i, V_s^j) are independent and independent of other variables, and the last equality follows from (5.30) and (5.29), and the terms starting with $\frac{1}{(N_i)^2} \sum_{i=1}^{N_i}$ go to zero because (f^1, f^2, g) are bounded. Hence,

$$\Upsilon_t^\infty(de', de)dt = \widehat{\Upsilon}_t(de', de)dt,$$

where $\widehat{\Upsilon}_t(de', de)dt := \mathbb{E}^{\mathbb{Q}} \left[\delta_{(H^\varepsilon(x, e)(du) \mu_t^\varepsilon(dx))} (de') \bar{\Lambda}_t(de, \mathcal{P}_U^n) dt \right]$, this is true for any limit Υ^∞ of any sub-sequence. Therefore, the sequence $(\Upsilon^N)_{N \in \mathbb{N}^*}$ converges towards $\widehat{\Upsilon}$ for the wasserstein metric \mathcal{W}_p . Then, to finish, by Theorem A.3,

$$\begin{aligned}
 \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} C^{\varepsilon, N} &= \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[r]^{N_i}}^i, \alpha_r^i)}(dx, du), m_r^N \right)^p dr \right] \\
 &= \lim_{\varepsilon \rightarrow 0} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \int_{\mathcal{P}_U^n} \mathcal{W}_p \left(H^\varepsilon(x, m)(du) \mu_t^\varepsilon(dx), m \right)^p \bar{\Lambda}_t(dm, \mathcal{P}_U^n) dt \right] = 0.
 \end{aligned}$$

All these results allow to deduce that $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\widehat{X}_t^{\varepsilon, i, N} - X_t^{\varepsilon, i, N}|^p \right] = 0$.

As

$$\begin{aligned}
 &\mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\widehat{m}_t^N, m_t^N)^p dr \right] \\
 &\leq \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\widehat{m}_t^N(dx, du), \frac{1}{N} \sum_{i=1}^N \delta_{(X_{[t]^{N_i}}^i, \alpha_t^i)}(dx, du))^p dr \right] \\
 &\quad + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[t]^{N_i}}^i, \alpha_t^i)}(dx, du), m_t^N(dx, du) \right)^p dr \right] \\
 &\leq \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T |\widehat{X}_t^{\varepsilon, i, N, K} - X_{[t]^{N_i}}^{\varepsilon, i, N}|^p dt \right]
 \end{aligned}$$

$$\begin{aligned}
 & + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[t]N}^i, \alpha_t^i)}(dx, du), m_t^N(dx, du) \right)^p dr \right] \\
 \leq & \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T |\widehat{X}_t^{\varepsilon, i, N} - X_t^{\varepsilon, i, N}|^p dt \right] + \frac{1}{2^N} \\
 & + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{(X_{[t]N}^i, \alpha_t^i)}(dx, du), m_t^N(dx, du) \right)^p dr \right]
 \end{aligned}$$

then $\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\widehat{m}_t^N, m_t^N)^p dr \right] = 0$, similarly, using (5.32),

$$\begin{aligned}
 & \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\widehat{\mu}^N), \phi_{e'}(\mu^N)) \right] \\
 \leq & \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \left(\mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\widehat{\mu}^N), \phi_{e'}(\vartheta^N)) \right] + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\vartheta^N), \phi_{e'}(\mu^N)) \right] \right) \\
 \leq & K \lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\widehat{X}_t^{\varepsilon, i, N} - X_t^{\varepsilon, i, N}|^p \right] + \frac{1}{2^N} \right. \\
 & \left. + \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{e' \in [0, T]} \mathcal{W}_p(\phi_{e'}(\vartheta^N), \phi_{e'}(\mu^N)) \right] \right) = 0.
 \end{aligned}$$

All previous result combined with measurability property (5.23) allowed to say $(\alpha^1, \dots, \alpha^N)$ and $(\widehat{X}^1, \dots, \widehat{X}^N)$ are the controls and the processes we are looking for. \square

In fact, in Theorem 5.6, instead of interaction processes of type (5.17), it is possible to use a sequence of *weak* McKean–Vlasov processes and obtain similar result. Let us assume conditions and inputs previously mentioned for Theorem 5.6 are satisfied. Let W be a $(\widehat{\mathbb{P}}, \widehat{\mathbb{F}})$ -Brownian motion, ξ be a $\widehat{\mathcal{F}}_0$ -random variable with $\mathcal{L}^{\widehat{\mathbb{P}}}(\xi) = \nu$, and Z be a uniform variable independent of (ξ, W) . In addition,

$$(\psi(\mu^N), \zeta^N, \overline{\Lambda}^N, B^N)_{N \in \mathbb{N}^*} \text{ are } \widehat{\mathbb{P}}\text{-independent of } (W, \xi, Z). \tag{5.33}$$

For each $N \in \mathbb{N}^*$, define the filtrations $\widehat{\mathbb{F}}^N := (\widehat{\mathcal{F}}_t^N)_{t \in [0, T]}$ and $\widehat{\mathbb{G}} := (\widehat{\mathcal{G}}_t^N)_{t \in [0, T]}$ by

$$\widehat{\mathcal{F}}_t^N := \sigma \left\{ \xi, \overline{\Lambda}_{t \wedge \cdot}^N, \phi_{t \wedge \cdot}(\mu^N), \zeta_{t \wedge \cdot}^N, W_{t \wedge \cdot}, B_{t \wedge \cdot}^N, Z \right\}$$

and

$$\widehat{\mathcal{G}}_t^N := \sigma \left\{ \psi_{t \wedge \cdot}(\mu^N), \zeta_{t \wedge \cdot}^N, \overline{\Lambda}_{t \wedge \cdot}^N, B_{t \wedge \cdot}^N \right\},$$

for all $t \in [0, T]$. $\widehat{\mathbb{G}}$ will play the role of the common noise filtration. We now provide approximations by *weak* McKean–Vlasov processes. The proofs of the next Theorem 5.8 and Theorem 5.9 are left in Appendix A.1.

Proposition 5.8. *There exists a sequence of processes $(\alpha^N)_{N \in \mathbb{N}^*}$ satisfying: for each $N \in \mathbb{N}^*$, α^N is $\widehat{\mathbb{F}}^N$ -predictable, such that if X^N is the unique strong solution of: $\mathbb{E}^{\widehat{\mathbb{P}}}[\|X^N\|^{p'}] < \infty$, $X_0^N = \xi$, for all $t \in [0, T]$,*

$$\begin{aligned}
 & dX_t^N \\
 = & \hat{b}(t, X_t^N, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_r^N, \bar{v}_r^N, \alpha_r^N) dt + \hat{\sigma}(t, X_t^N, B^N, \phi(\widehat{\mu}^N), \zeta^N, \widehat{m}_t^N, \bar{v}_t^N, \alpha_t^N) dW_t,
 \end{aligned} \tag{5.34}$$

where $\widehat{m}_t^N := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^N, \alpha_t^N | \widehat{\mathcal{G}}_t^N)$ and $\widehat{\mu}_t^N := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^N | \widehat{\mathcal{G}}_t^N)$, then for the sub-sequence $(N_k)_{k \in \mathbb{N}^*}$ given in Proposition 5.6,

$$\lim_{k \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\widehat{m}_t^{N_k}, m_t^{N_k})^p dt + \sup_{t \in [0, T]} \mathcal{W}_p(\phi_t(\widehat{\mu}^{N_k}), \phi_t(\mu^{N_k})) \right] = 0,$$

and if $\widehat{\Lambda}_s(dm, d\bar{\nu})ds := \delta_{(\widehat{m}_s^{N_k}, \bar{\nu}_s^{N_k})}(dm, d\bar{\nu})ds$,

$$\lim_{k \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{\mu}^{N_k}, \zeta^{N_k}, \widehat{\Lambda}, B^{N_k}) = \mathcal{L}^{\mathbb{Q}}(\mu, \zeta, \bar{\Lambda}, B), \text{ in } \mathcal{W}_p. \tag{5.35}$$

Another useful approximation Using roughly the same arguments as those used in the proof of the Theorem 5.6, another approximation result can be provided. This can be seen as another version of Theorem 5.8 where the sequence $(\bar{\Lambda}^N)_{N \in \mathbb{N}^*}$ is not necessarily a subset of $\mathbb{M}_0((\mathcal{P}_U^n)^2)$ and the controls that achieve the approximation are probability measures.

Proposition 5.9. *Let us stay in the context of Theorem 5.8 with $(\bar{\Lambda}^N)_{N \in \mathbb{N}^*}$ not necessarily a subset of $\mathbb{M}_0((\mathcal{P}_U^n)^2)$. There exists $(\beta^N)_{N \in \mathbb{N}^*}$ a sequence of $\mathcal{P}(U)$ -valued $(\widehat{\mathcal{F}}_t \otimes \mathcal{B}(\mathcal{P}_U^n))_{t \in [0, T]}$ -predictable processes such that if $(X_t^N)_{t \in [0, T]} := (X_t)_{t \in [0, T]}$ is the unique strong solution of: $\mathbb{E}^{\widehat{\mathbb{P}}}[\|X^N\|^{p'}] < \infty$, $X_0 = \xi$, for all $t \in [0, T]$, $\widehat{\mathbb{P}}$ -a.e.,*

$$\begin{aligned} dX_t &= \int_{(\mathcal{P}_U^n)^2} \int_U \hat{b}(t, X_t, B^N, \phi(\eta^N), \zeta^N, \widehat{\mathbf{m}}_t^N[m], \bar{\nu}, u) \beta_t^N(m)(du) \bar{\Lambda}_t^N(dm, d\bar{\nu})dt \\ &\quad + \left(\int_{(\mathcal{P}_U^n)^2} \int_U \hat{\sigma} \hat{\sigma}^\top(t, X_t, B^N, \phi(\eta^N), \zeta^N, \widehat{\mathbf{m}}_t^N[m], \bar{\nu}, u) \beta_t^N(m)(du) \bar{\Lambda}_t^N(dm, d\bar{\nu}) \right)^{1/2} dW_t \end{aligned}$$

where

$$\widehat{\mathbf{m}}_t^N[m](dx, du) := \mathbb{E}^{\widehat{\mathbb{P}}}[\beta_t^N(m)(du) \delta_{X_t^N}(dx) | \widehat{\mathcal{G}}_t^N] \text{ and } \widehat{\mu}_t^N := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^N | \widehat{\mathcal{G}}_t^N) \text{ for all } t \in [0, T],$$

then, one has, for a sub-sequence $(N_j)_{j \in \mathbb{N}^*} \subset \mathbb{N}^*$,

$$\lim_{j \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T \int_{\mathcal{P}_U^n} \mathcal{W}_p(\widehat{\mathbf{m}}_r^{N_j}[m], m) \bar{\Lambda}_r^{N_j}(dm, \mathcal{P}_U^n) dr \right] = 0$$

and

$$\lim_{j \rightarrow \infty} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\sup_{s \in [0, T]} \mathcal{W}_p(\phi_s(\widehat{\mu}^{N_j}), \phi_s(\mu^{N_j})) \right] = 0,$$

in addition if $\widehat{\Lambda}_s^N(dm, d\bar{\nu})ds := \int_{\mathcal{P}_U^n} \delta_{\widehat{\mathbf{m}}_s^N[e]}(dm) \bar{\Lambda}_s^N(de, d\bar{\nu})ds$,

$$\lim_{j \rightarrow \infty} \mathcal{L}^{\widehat{\mathbb{P}}}(\widehat{\mu}^{N_j}, \zeta^{N_j}, \widehat{\Lambda}^{N_j}, B^{N_j}) = \mathcal{L}^{\mathbb{Q}}(\mu, \phi(\mu), \zeta, \bar{\Lambda}, B), \text{ in } \mathcal{W}_p. \tag{5.36}$$

Remark 5.10. With exactly the same proof, an important observation is the following: if the coefficients functions $(\hat{b}, \hat{\sigma})$ are of the form of type

$$(\hat{b}, \hat{\sigma}^\top)(t, x, \mathbf{b}, \pi, \beta, m, \bar{\nu}, u) := (\hat{b}^*, \hat{a}^*)(t, \mathbf{b}, \pi, \beta, \bar{\nu}) + (\hat{b}^\circ, \hat{a}^\circ)(t, x, \mathbf{b}, \pi, \beta, m, u),$$

where $(\hat{b}^*, \hat{a}^*, \hat{b}^\circ, \hat{a}^\circ)$ are bounded continuous functions, we can replace the convergence assumptions (5.15) by

$$\lim_{N \rightarrow \infty} \mathcal{W}_{p'} \left(\frac{1}{N} \sum_{i=1}^N \nu^i, \nu \right) = 0$$

and

$$\lim_{N \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}} \left(\phi(\mu^N), \zeta^N, \Lambda^{\circ, N}, \Lambda^{*, N}, B^N \right) = \mathcal{L}^{\mathbb{Q}} \left(\phi(\mu), \zeta, \Lambda^{\circ}, \Lambda^*, B \right), \quad \text{in } \mathcal{W}_p, \quad (5.37)$$

with $\Lambda^{\circ, N} := \bar{\Lambda}_t^N(dm, \mathcal{P}_U^n)dt$, $\Lambda^{*, N} := \bar{\Lambda}_t^N(\mathcal{P}_U^n, d\bar{\nu})dt$, $\Lambda^{\circ} := \bar{\Lambda}_t(dm, \mathcal{P}_U^n)dt$, and $\Lambda^* := \bar{\Lambda}_t(\mathcal{P}_U^n, d\bar{\nu})dt$. And then, in Theorem 5.6, Theorem 5.8 and Theorem 5.9, the convergence results (5.18), (5.35) and (5.36) are replaced by

$$\lim_{j \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}} \left(\hat{\mu}^{N_j}, \zeta^{N_j}, \hat{\Lambda}_t^{N_j}(dm, \mathcal{P}_U^n)dt, \hat{\Lambda}_t^{N_j}(\mathcal{P}_U^n, d\bar{\nu})dt, B^{N_j} \right) = \mathcal{L}^{\mathbb{Q}} \left(\mu, \phi(\mu), \zeta, \Lambda^{\circ}, \Lambda^*, B \right),$$

in \mathcal{W}_p . In other words, when the variables $(m, \bar{\nu})$ of $(\hat{b}, \hat{\sigma}\hat{\sigma}^\top)$ are “separated”, we just need separated condition on $(\bar{\Lambda}^N)_{N \in \mathbb{N}^*}$ of type (5.37), i.e. $\bar{\Lambda}^N$ “separated”.

References

- [1] B. Acciaio, J. Backhoff Veraguas, and R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM journal on Control and Optimization*, 57(6), 2019. MR4029806
- [2] C. Alasseur, I. B. Taher, and A. Matoussi. An extended mean field game for storage in smart grids. *Journal of Optimization Theory and Applications*, 184:644–670, 2020. MR4054298
- [3] D. Blackwell and L. E. Dubins. An extension of skorohod’s almost sure representation theorem. *Proceedings of the American Mathematical Society*, 89(4), 1983. MR0718998
- [4] A. Budhiraja, P. Dupuis, and M. Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *The Annals of Probability*, 40(1):74–102, 2012. MR2917767
- [5] P. Cardaliaguet and C.-A. Lehalle. Mean field game of controls and an application to trade crowding. *Mathematics and Financial Economics*, 12(3):335–363, 2018. MR3805247
- [6] R. Carmona and D. Lacker. A probabilistic weak formulation of mean field games and applications. *The Annals of Applied Probability*, 25(3):1189–1231, 2015. MR3325272
- [7] R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. *The Annals of Probability*, 44(6):3740–3803, 2016. MR3572323
- [8] P. Casgrain and S. Jaimungal. Mean-field games with differing beliefs for algorithmic trading. *arXiv preprint arXiv:1810.06101*, 2018. MR4116456
- [9] M. F. Djete. Mean field games of controls: on the convergence of nash equilibria. *arXiv preprint arXiv:2006.12993*, 2020.
- [10] M. F. Djete, D. Possamaï, and X. Tan. McKean–Vlasov optimal control: the dynamic programming principle. *arXiv preprint arXiv:1907.08860*, 2019.
- [11] M. F. Djete, D. Possamaï, and X. Tan. McKean–Vlasov optimal control: limit theory and equivalence between different formulations. *arXiv preprint arXiv:2001.00925*, 2020.
- [12] N. El Karoui, D. Huu Nguyen, and M. Jeanblanc-Picqué. Compactification methods in the control of degenerate diffusions: existence of an optimal control. *Stochastics*, 20(3):169–219, 1987. MR0878312
- [13] R. Elie, T. Mastrolia, and D. Possamaï. A tale of a principal and many many agents. *Mathematics of Operations Research*, 2018. MR3959080
- [14] O. Féron, P. Tankov, and L. Tinsi. Price formation and optimal trading in intraday electricity markets. *arXiv preprint arXiv:2009.04786*, 2020.
- [15] M. Fischer and G. Livieri. Continuous time mean-variance portfolio optimization through the mean field approach. *ESAIM: Probability and Statistics*, 20:30–44, 2016. MR3528616
- [16] W. Fleming and M. Nisio. On stochastic relaxed control for partially observed diffusions. *Nagoya Mathematical Journal*, 93:71–108, 1984. MR0738919
- [17] J. Gärtner. On the McKean–Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, 137(1):197–248, 1988. MR0968996

- [18] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986. MR0833267
- [19] N. J. Higham. *Functions of matrices: theory and computation*. Society for Industrial & Applied Mathematics, 2008. MR2396439
- [20] O. Kallenberg. *Foundations of modern probability*. Probability and its applications. Springer-Verlag New York, 2nd edition, 2002. MR1876169
- [21] N. V. Krylov. *Controlled Diffusion Processes*. Springer, 1980. MR0601776
- [22] N. V. Krylov. Once more about the connection between elliptic operators and Itô’s stochastic equations. *Statistics and Control of Stochastic Processes (Moscow 1984)*, pages 214–229, 1984. MR0808203
- [23] D. Lacker. A general characterization of the mean field limit for stochastic differential games. *Probability Theory and Related Fields*, 165(3-4):581–648, 2016. MR3520014
- [24] D. Lacker. Limit theory for controlled McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017. MR3654119
- [25] D. Lacker. On the convergence of closed-loop nash equilibria to the mean field game limit. *arXiv preprint arXiv:1808.02745*, 2018. MR4133381
- [26] D. Lacker and A. Soret. Many-player games of optimal consumption and investment under relative performance criteria. *arXiv preprint arXiv:1905.11782*, 2019. MR4076185
- [27] D. Lacker, M. Shkolnikov, and J. Zhang. Superposition and mimicking theorems for conditional McKean–Vlasov equations. *arXiv preprint arXiv:2004.00099*, 2020.
- [28] M. Laurière and L. Tangpi. Convergence of large population games to mean field games with interaction through controls. *arXiv preprint arXiv:2004.08351*, 2020.
- [29] R. Liptser and A. Shiryaev. *Statistics of Random Processes*. Springer-Verlag, 1977. MR0474486
- [30] M. Motte and H. Pham. Mean-field markov decision processes with common noise and open-loop controls. *arXiv preprint arXiv:1912.07883*, 2020.
- [31] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *The Annals of Probability*, 12(2):458–479, 1984. MR0735849
- [32] H. Pham and X. Wei. Dynamic programming for optimal control of stochastic McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(2):1069–1101, 2017. MR3631380
- [33] A.-S. Snitzman. Topics in propagation of chaos. In P. Hennequin, editor, *École d’été de probabilités de Saint-Flour XIX – 1989*, number 1464 in Lecture notes in mathematics, pages 165–251. Springer Berlin Heidelberg, 1991. MR1108185
- [34] D. Stroock and S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1977. MR0532498
- [35] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 2008. MR2459454

Acknowledgments. The author is grateful to Dylan Possamaï and Xiaolu Tan for helpful comments and suggestions.

A Some technical results

A.1 Technical proofs

We will give here successively the proofs of Theorem 5.2, Theorem 5.8 and Theorem 5.9.

Proof of Theorem 5.2. Let $\delta > 0$ and define

$$\mathbf{q}_t^\delta(dm, dm') := \frac{1}{\delta} \int_{(t-\delta) \vee 0}^t \hat{\mathbf{q}}_s^\delta(dm, dm') ds, \text{ for all } t \in [0, T].$$

By using similar approach to [29, Lemma 4.4], the sequence $(\hat{\mathbf{q}}^\delta)_{\delta>0}$ satisfying: for each $\delta > 0$, $\hat{\mathbf{q}}_t^\delta(dm, dm')dt \in \mathbb{M}((\mathcal{P}_U^n)^2)$, $\hat{\mathbf{q}}^\delta : t \in [0, T] \rightarrow \hat{\mathbf{q}}_t^\delta(dm, dm') \in (\mathcal{P}_U^n)^2$ is continuous, and $\lim_{\delta \rightarrow 0} \hat{\mathbf{q}}_t^\delta = \hat{\mathbf{q}}_t$, in weakly sense for ds almost every $t \in [0, T]$.

Let us fix $t_0 \in (0, T]$, $\phi \in C_b^2(\mathbb{R}^n)$, by [21, Chapter 2 Section 9 Theorem 10], there exists $v^{\varepsilon, \delta} \in C_b^{1,2}([0, t_0] \times \mathbb{R}^n)$ satisfying: $v^{\varepsilon, \delta}(t_0, x) = \phi(x)$ and

$$\partial_t v^{\varepsilon, \delta}(t, x) + \mathcal{A}_t^\varepsilon[v^{\varepsilon, \delta}(t, \cdot)][\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta](x) = 0 \text{ for all } (t, x) \in [0, t_0] \times \mathbb{R}^n. \quad (\text{A.1})$$

Notice that, under Theorem 2.1, for each $\varepsilon > 0$, $\hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \kappa](t, x) \geq \theta \mathbf{I}_{n \times n}$ for all $(t, x, \kappa) \in [0, T] \times \mathbb{R}^n \times \mathcal{P}((\mathcal{P}_U^n)^2)$. By Proposition A.5, for all $t \in [0, T]$, $x \in \mathbb{R}^n \rightarrow (\hat{a}^\varepsilon)^{1/2}[\mathbf{b}, \mathbf{n}, \mathbf{z}, \kappa](t, x) \in \mathbb{S}^{n \times n}$ is Lipschitz (with Lipschitz constant independent of $(t, \mathbf{n}, \mathbf{z}, \kappa)$).

Let $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ be a probability space supporting W a \mathbb{R}^n -valued (\mathbb{P}, \mathbb{F}) -Brownian motion, and ξ a \mathcal{F}_0 -random variable such that $\mathcal{L}^\mathbb{P}(\xi) \in \mathcal{P}_p(\mathbb{R}^n)$. Now, for every $t \in [0, t_0]$, denote by $X^{\varepsilon, \delta, t, \xi} := X$ the continuous process unique strong solution of: \mathbb{P} -a.e.,

$$X_s = \xi + \int_t^s \hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta](r, X_r)dr + \int_t^s (\hat{a}^\varepsilon)^{1/2}[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta](r, X_r)dW_r \text{ for all } s \in [t, T].$$

By applying Itô's formula, one has that (Feynman Kac's formula)

$$v^{\varepsilon, \delta}(t, x) = \mathbb{E}^\mathbb{P} \left[\phi(X_{t_0}^{\varepsilon, \delta, t, \xi}) | \xi = x \right] = \mathbb{E}^\mathbb{P} \left[\phi(X_{t_0}^{\varepsilon, \delta, t, x}) \right] \text{ for all } (t, x) \in [0, t_0] \times \mathbb{R}^n. \quad (\text{A.2})$$

By definition of \hat{a}^ε and \hat{b}^ε (see (5.8)), and by using the fact that $\hat{\mathbf{q}}^\delta \in \mathbb{M}((\mathcal{P}_U^n)^2)$, there exists a constant C_ε (independent of $\delta > 0$) such that: for all $(t, x) \in [0, T] \times \mathbb{R}^n$,

$$|\nabla^2(\hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta], \hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta])(t, x)| + |\nabla(\hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta], \hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta])(t, x)| \leq C_\varepsilon.$$

Then, by [21, Chapter 2 Section 8 Theorem 8, Theorem 7], for two unit vectors $(w^1, w^2) \in \mathbb{R}^n \times \mathbb{R}^n$, there exist two \mathbb{R}^n -valued \mathbb{F} -adapted continuous processes $Y^{\varepsilon, \delta, t, x, w^1} := Y$ and $Z^{\varepsilon, \delta, t, x, w^1, w^2} := Z$ such that

$$\lim_{h \rightarrow 0} \mathbb{E}^\mathbb{P} \left[\sup_{s \in [t, t_0]} \left| \frac{X_s^{\varepsilon, \delta, t, x + hw^1} - X_s^{\varepsilon, \delta, t, x}}{h} - Y_s \right| \right] = 0$$

and

$$\lim_{h \rightarrow 0} \mathbb{E}^\mathbb{P} \left[\sup_{s \in [t, t_0]} \left| \frac{Y_s^{\varepsilon, \delta, t, x + hw^2, w^1} - Y_s^{\varepsilon, \delta, t, x, w^1}}{h} - Z_s \right| \right] = 0,$$

formally speaking, Y can be seen as the “derivative” (given a direction w^1) of $x \rightarrow X^x$, and Z the “derivative” (given w^1 and another direction w^2) of Y . In addition $\mathbb{E}^\mathbb{P} [\sup_{s \in [t, t_0]} |Y_s| + |Z_s|] \leq K_\varepsilon$, with K_ε depending on ε but not of δ . As $\phi \in C_b^2(\mathbb{R}^n)$, by using the previous results and equation (A.2), there exists $\hat{K}_\varepsilon > 0$ (independent of δ) satisfying: for all $(t, x) \in [0, T] \times \mathbb{R}^n$

$$|\nabla^2 v^{\varepsilon, \delta}(t, x)| + |\nabla v^{\varepsilon, \delta}(t, x)| + |v^{\varepsilon, \delta}(t, x)| \leq \hat{K}_\varepsilon. \quad (\text{A.3})$$

Therefore, for all $\varepsilon > 0$,

$$\begin{aligned} & |\mathcal{A}_t^\varepsilon v^{\varepsilon, \delta}(t, \cdot)[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t](x) - \mathcal{A}_t^\varepsilon v^{\varepsilon, \delta}(t, \cdot)[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta](x)| \\ & \leq \hat{K}_\varepsilon (|\hat{b}^\varepsilon, \hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t](t, x) - \hat{b}^\varepsilon, \hat{a}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta](t, x)|), \end{aligned}$$

by definition (5.8), as $\lim_{\delta \rightarrow 0} \hat{\mathbf{q}}_t^\delta = \hat{\mathbf{q}}_t$, for ds almost every $t \in [0, T]$, one gets:

$$\lim_{\delta \rightarrow 0} |\mathcal{A}_t^\varepsilon v^{\varepsilon, \delta}(t, \cdot)[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t](x) - \mathcal{A}_t^\varepsilon v^{\varepsilon, \delta}(t, \cdot)[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_t^\delta](x)| = 0, \quad (\text{A.4})$$

for each $\varepsilon > 0$ and $x \in \mathbb{R}^n$, for ds almost every $t \in [0, T]$.

Uniqueness: For each $\varepsilon > 0$ fixed, let us prove the uniqueness of $(\mathbf{n}_t^\varepsilon)_{t \in [0, T]}$ solution of equation (5.10). Let $\mathbf{n}^{1, \varepsilon}$ and $\mathbf{n}^{2, \varepsilon}$ be two solutions of the Fokker-Planck equation (5.10) mentioned in the Lemma, for any $t_0 \in [0, T]$ and $\phi \in C_b^2(\mathbb{R}^n)$, denote by $v := v^{\varepsilon, \delta, \phi, t_0}$ solution of (A.1) associated to (t_0, ϕ) . One finds

$$\begin{aligned} & \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{1, \varepsilon}(dy) - \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{2, \varepsilon}(dy) \\ &= \int_0^{t_0} \langle \partial_t v(r, \cdot), \mathbf{n}_r^{1, \varepsilon} \rangle - \langle \partial_t v(r, \cdot), \mathbf{n}_r^{2, \varepsilon} \rangle \\ & \quad + \langle \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](\cdot), \mathbf{n}_r^{1, \varepsilon} \rangle - \langle \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](\cdot), \mathbf{n}_r^{2, \varepsilon} \rangle dr \\ &= \int_0^{t_0} \langle \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](\cdot) - \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta](\cdot), \mathbf{n}_r^{1, \varepsilon} \rangle \\ & \quad + \langle \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](\cdot) - \mathcal{A}_r^\varepsilon v[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta](\cdot), \mathbf{n}_r^{2, \varepsilon} \rangle dr, \end{aligned}$$

by (A.4), given $\varepsilon > 0$, after taking $\delta \rightarrow 0$, by Lebesgue's dominated convergence theorem, $\int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{1, \varepsilon}(dy) = \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{2, \varepsilon}(dy)$, this is true for all $(t_0, \phi) \in [0, T] \times C_b^2(\mathbb{R}^n)$, then $\mathbf{n}^{1, \varepsilon} = \mathbf{n}^{2, \varepsilon}$.

Convergence of \mathbf{n}^ε : Now, we show the second assertion of our Lemma. Using the fact that $\hat{\mathbf{q}}_t(\mathbb{Z}_{\mathbf{n}_t} \times \mathcal{P}_U^n) = 1$ dt -almost surely $t \in [0, T]$, one gets for all $t \in [0, T]$,

$$\begin{aligned} & \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(t, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_t(dz) dy = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(t, z - y) \mathbf{n}_t(dz) G_\varepsilon(y) dy \\ &= \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy + \int_0^t \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} \partial_t v^{\varepsilon, \delta}(s, z - y) \mathbf{n}_s(dz) \right. \\ & \quad \left. + \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \mathcal{A}_s[v^{\varepsilon, \delta}(s, \cdot - y)](z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) m(dz, du) \hat{\mathbf{q}}_s(dm, d\bar{\nu}) \right] G_\varepsilon(y) dy ds \\ &= \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy + \int_0^t \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left[\partial_t v^{\varepsilon, \delta}(s, z - y) \right. \\ & \quad \left. + \int_{(\mathcal{P}_U^n)^2} \int_U \mathcal{A}_s[v^{\varepsilon, \delta}(s, \cdot - y)](z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) m^z(du) \hat{\mathbf{q}}_s(dm, d\bar{\nu}) \right] G_\varepsilon(y) \mathbf{n}_s(dz) dy ds \\ &= \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy + \int_0^t \int_{\mathbb{R}^n} \left[\partial_t v^{\varepsilon, \delta}(s, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_s(dz) \right. \\ & \quad \left. + \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \hat{b}(s, z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) \nabla v^{\varepsilon, \delta}(s, y) G_\varepsilon(z - y) m^z(du) \mathbf{n}_s(dz) \hat{\mathbf{q}}_s(dm, d\bar{\nu}) \right. \\ & \quad \left. + \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \frac{1}{2} \text{Tr}[\hat{a}(s, z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) \nabla^2 v^{\varepsilon, \delta}(s, y)] G_\varepsilon(z - y) m^z(du) \mathbf{n}_s(dz) \hat{\mathbf{q}}_s(dm, d\bar{\nu}) \right] dy ds \\ &= \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy + \int_0^t \int_{\mathbb{R}^n} \left[\partial_t v^{\varepsilon, \delta}(s, y) (\mathbf{n}_s)^{(\varepsilon)}(y) \right. \\ & \quad \left. + \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \hat{b}(s, z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{\nu}, u) \frac{G_\varepsilon(z - y)}{(m(dz', U))^{(\varepsilon)}(y)} m(dz, du) \hat{\mathbf{q}}_s(dm, d\bar{\nu}) \nabla v^{\varepsilon, \delta}(s, y) (\mathbf{n}_s)^{(\varepsilon)}(y) \right. \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} \text{Tr} \left[\int_{(\mathcal{P}_{\bar{v}}^n)^2} \int_{\mathbb{R}^n \times U} \hat{a}(s, z, \mathbf{b}, \mathbf{n}, \mathbf{z}, m, \bar{v}, u) \frac{G_\varepsilon(z - y)}{(m(dz', U))^{(\varepsilon)}(y)} m(dz, du) \hat{\mathbf{q}}_s(dm, d\bar{v}) \nabla^2 v^{\varepsilon, \delta}(s, y) \right. \\
 & \qquad \qquad \qquad \left. (\mathbf{n}_s)^{(\varepsilon)}(y) \right] dy ds \\
 & = \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy \\
 & \qquad \qquad \qquad + \int_0^t \int_{\mathbb{R}^n} [\partial_t v^{\varepsilon, \delta}(r, y) + \mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r)(r, y)] (\mathbf{n}_r)^{(\varepsilon)}(y) dy dr,
 \end{aligned}$$

where for each $\pi \in \mathcal{P}(\mathbb{R}^n)$, we write $\pi^{(\varepsilon)}(x) := \int_{\mathbb{R}^n} G_\varepsilon(x - z) \pi(dz)$, for all $x \in \mathbb{R}^n$.

Then by (A.1)

$$\begin{aligned}
 & \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \nu^{(\varepsilon)}(y) dy \\
 & = \int_{\mathbb{R}^n} \phi(y) (\mathbf{n}_{t_0})^{(\varepsilon)}(y) dy \\
 & \qquad \qquad \qquad + \int_0^{t_0} \int_{\mathbb{R}^n} [\mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta)(y) - \mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r)(y)] (\mathbf{n}_r)^{(\varepsilon)}(y) dy dr.
 \end{aligned}$$

By equation (A.2), one has

$$\begin{aligned}
 & \int_{\mathbb{R}^n} v^{\varepsilon, \delta}(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy \\
 & = \int_{\mathbb{R}^n} \mathbb{E} \left[\phi(X_{t_0}^{\varepsilon, \delta, 0, \xi}) | \xi = y \right] \nu^{(\varepsilon)}(y) dy = \int_{\mathbb{R}^n} \phi(x) \mathbf{n}_{t_0}^{\varepsilon, \delta}(dx),
 \end{aligned}$$

where $\mathbf{n}_t^{\varepsilon, \delta} := \mathcal{L}^{\mathbb{P}}(X_t^{\varepsilon, \delta, 0, \xi^\varepsilon})$ for $t \in [0, T]$, with $\mathcal{L}^{\mathbb{P}}(\xi^\varepsilon)(dy) = \nu^{(\varepsilon)}(y) dy$. Combining the previous equality,

$$\begin{aligned}
 & \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) \\
 & = \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_{t_0}(dz) dy \\
 & \qquad \qquad \qquad + \int_{\mathbb{R}^n} \phi(y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_{t_0}(dz) dy \\
 & = \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_{t_0}(dz) dy \\
 & \qquad \qquad \qquad + \int_{\mathbb{R}^n} v^\varepsilon(0, y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) \nu(dz) dy + \\
 & \int_0^{t_0} \\
 & \qquad \int_{\mathbb{R}^n} \left[\mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r)(y) - \mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta)(y) \right] \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_r(dz) dy dr \\
 & = \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{\varepsilon, \delta}(dy) + \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) dy \mathbf{n}_{t_0}(dz) + \\
 & \int_0^{t_0} \\
 & \qquad \int_{\mathbb{R}^n} \left[\mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r)(y) - \mathcal{A}_r^\varepsilon[v^{\varepsilon, \delta}(r, \cdot)](\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r^\delta)(y) \right] \int_{\mathbb{R}^n} G_\varepsilon(z - y) \mathbf{n}_r(dz) dy dr.
 \end{aligned}$$

Consequently, for each $\varepsilon > 0$,

$$\begin{aligned} & \limsup_{\delta \rightarrow 0} \left| \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{\varepsilon, \delta}(dy) \right| \\ & \leq \left| \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \int_{\mathbb{R}^n} G_\varepsilon(z - y) dy \mathbf{n}_{t_0}(dz) \right|. \end{aligned}$$

Finally

$$\lim_{\varepsilon \rightarrow 0} \limsup_{\delta \rightarrow 0} \left| \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}(dy) - \int_{\mathbb{R}^n} \phi(y) \mathbf{n}_{t_0}^{\varepsilon, \delta}(dy) \right| = 0, \tag{A.5}$$

for any $\phi \in C_b^2(\mathbb{R}^n)$ and $t_0 \in [0, T]$, where we used that $\lim_{\varepsilon \rightarrow 0} \left| \int_{\mathbb{R}^n} \phi(y) G_\varepsilon(z - y) dy - \phi(z) \right| = 0$, for all $z \in \mathbb{R}^n$.

Notice that $\nu^{(\varepsilon)}(y)(dy)$ converges weakly to $\nu(dy)$. By Skorokhod’s representation theorem, one can find a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ supporting $(\xi^\varepsilon)_{\varepsilon > 0}$ and ξ such that $\mathcal{L}^{\tilde{\mathbb{P}}}(\xi^\varepsilon) = \nu^{(\varepsilon)}(y)(dy)$ and $\mathcal{L}^{\tilde{\mathbb{P}}}(\xi) = \nu(dy)$, and $\lim_{\varepsilon \rightarrow 0} \xi^\varepsilon = \xi$ $\tilde{\mathbb{P}}$ a.e.. And when $\mathcal{L}^{\tilde{\mathbb{P}}}(\xi) = \nu \in \mathcal{P}_{p'}(\nu)$, one has $\sup_{\varepsilon > 0} \mathbb{E}^{\tilde{\mathbb{P}}} [|\xi^\varepsilon|^{p'}] = \sup_{\varepsilon > 0} \int_{\mathbb{R}^n} |y|^{p'} \nu^{(\varepsilon)}(y)(dy) \leq C(1 + \int_{\mathbb{R}^n} |y|^{p'} \nu(dy)) < \infty$, by using standard techniques of uniform integrability, $\lim_{\varepsilon \rightarrow 0} \mathbb{E}^{\tilde{\mathbb{P}}} [|\xi^\varepsilon - \xi|^p] = 0$, recall that $p' > p$. If necessary, it is possible to enlarge the initial space, for sake of clarity and without technical problems, let us assume $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is equal to the initial space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $\varepsilon > 0$, let X^ε be the continuous process unique strong solution of: \mathbb{P} -a.e.

$$X_s^\varepsilon = \xi + \int_0^s \hat{b}^\varepsilon[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](r, X_r^\varepsilon) dr + \int_0^s (\hat{a}^\varepsilon)^{1/2}[\mathbf{b}, \mathbf{n}, \mathbf{z}, \hat{\mathbf{q}}_r](r, X_r^\varepsilon) dW_r \text{ for all } s \in [0, T].$$

By using the regularity of $(\hat{b}^\varepsilon, \hat{a}^\varepsilon)$ for ε fixed, it is straightforward to find that

$$\lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \mathbb{E}^{\mathbb{P}} \left[\sup_{t \in [0, T]} |X_t^\varepsilon - X_t^{\varepsilon, \delta, 0, \xi^\varepsilon}|^p \right] = 0.$$

By Itô’s formula and uniqueness of the Fokker–Planck equation (5.10), $\mathbf{n}_t^\varepsilon = \mathcal{L}^{\mathbb{P}}(X_t^\varepsilon)$ for each $t \in [0, T]$. Thanks to (A.5) and the previous result, one gets that, in weakly convergence sense, $\lim_{\varepsilon} \mathbf{n}_t^\varepsilon = \lim_{\varepsilon \rightarrow 0} \lim_{\delta \rightarrow 0} \mathbf{n}_t^{\varepsilon, \delta} = \mathbf{n}_t$ for each $t \in [0, T]$. Therefore, we proved that: for each $t \in [0, T]$, \mathbf{n}_t^ε converges weakly to \mathbf{n}_t . To deduce the Wasserstein convergence \mathcal{W}_p , notice that: $\sup_{\varepsilon > 0} \sup_{t \in [0, T]} \int_{\mathbb{R}^n} |x|^{p'} \mathbf{n}_t^\varepsilon(dx) \leq C(1 + \int_{\mathbb{R}^n} |y|^{p'} \nu(dy)) < \infty$, and

$$\begin{aligned} & \limsup_{\delta' \rightarrow 0} \sup_{\varepsilon > 0} \sup_{s \in [0, T]} \mathcal{W}_p(\mathbf{n}_{(s+\delta') \wedge T}^\varepsilon, \mathbf{n}_s^\varepsilon)^p \\ & = \limsup_{\delta' \rightarrow 0} \sup_{\varepsilon > 0} \sup_{s \in [0, T]} \mathcal{W}_p(\mathcal{L}^{\mathbb{P}}(X_{(s+\delta') \wedge T}^\varepsilon), \mathcal{L}^{\mathbb{P}}(X_s^\varepsilon))^p \\ & \leq \limsup_{\delta' \rightarrow 0} \sup_{\varepsilon > 0} \sup_{s \in [0, T]} \mathbb{E}^{\mathbb{P}} [|X_{(s+\delta') \wedge T}^\varepsilon - X_s^\varepsilon|^p] \leq \hat{C} \limsup_{\delta' \rightarrow 0} \delta' = 0, \end{aligned}$$

where the last equality follows from the Holder’s property of trajectories of X^ε with a constant independent of ε (essentially because $(\hat{b}, \hat{\sigma})$ are bounded). By Aldous’ criterion [20, Lemma 16.12] (see also proof of [7, Proposition-B.1]), $(\mathbf{n}^\varepsilon)_{\varepsilon > 0}$ is relatively compact in $C([0, T]; \mathcal{P}_p(\mathbb{R}^n))$ with the metric $\Delta(\nu, \nu') := \sup_{t \in [0, T]} \mathcal{W}_p(\nu_t, \nu'_t)$ for all $(\nu, \nu') \in C([0, T]; \mathcal{P}_p(\mathbb{R}^n)) \times C([0, T]; \mathcal{P}_p(\mathbb{R}^n))$. As for each $t \in [0, T]$, \mathbf{n}_t^ε converges weakly to \mathbf{n}_t , then the limit of each sub-sequence of $(\mathbf{n}^\varepsilon)_{\varepsilon > 0}$ is \mathbf{n} , consequently $\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} \mathcal{W}_p(\mathbf{n}_t^\varepsilon, \mathbf{n}_t) = 0$. □

Proof of Theorem 5.8. Before starting, let us mention that many parts of this proof use Theorem 5.6 and its associated proof.

Let us take the sequence of processes $(\alpha^{i,N})_{(i,N) \in \mathbb{N}^* \times \mathbb{N}^*}$ given in Theorem 5.6 with $\mathcal{L}^{\widehat{\mathbb{P}}}(\xi^i) = \nu^i = \nu$ for each i , and define the unique strong solution $X^{i,N}$ of: $X_0^{i,N} = \xi$ and

$$dX_t^{i,N} = \hat{b}(t, X_t^{i,N}, B^N, \phi(\widehat{\mu}^{i,N}), \zeta^N, \widehat{m}_t^{i,N}, \bar{\nu}_t^N, \alpha_t^{i,N}) dt + \hat{\sigma}(t, X_t^{i,N}, B^N, \phi(\widehat{\mu}^{i,N}), \zeta^N, \widehat{m}_t^{i,N}, \bar{\nu}_t^N, \alpha_t^{i,N}) dW_t^i,$$

with $\widehat{m}_t^{i,N} := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^{i,N}, \alpha_t^{i,N} | \widehat{\mathcal{G}}_t^N)$ and $\widehat{\mu}_t^{i,N} := \mathcal{L}^{\widehat{\mathbb{P}}}(X_t^{i,N} | \widehat{\mathcal{G}}_t^N)$. As $\alpha^{i,N}$ is $\widehat{\mathbb{F}}^{i,N}$ -predictable ($\widehat{\mathbb{F}}^{i,N}$ is defined in (5.16)), there exists a Borel function $G : [0, T] \times \mathbb{R}^n \times \mathbb{M}((\mathcal{P}_U^N)^2) \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}^n \times \mathcal{C}^\ell \times [0, 1] \rightarrow U$ satisfying $\alpha_t^{i,N} = G(t, \xi^i, \bar{\Lambda}_{t \wedge \cdot}^N, \phi_{t \wedge \cdot}(\mu^N), \zeta_{t \wedge \cdot}^N, W_{t \wedge \cdot}^i, B_{t \wedge \cdot}^N, Z^i)$, $dt \otimes d\widehat{\mathbb{P}}$ -a.e.. Define $\alpha_t^N := G(t, \xi, \bar{\Lambda}_{t \wedge \cdot}^N, \phi_{t \wedge \cdot}(\mu^N), \zeta_{t \wedge \cdot}^N, W_{t \wedge \cdot}^N, B_{t \wedge \cdot}^N, Z)$. Let X^N be the unique strong solution of equation (5.34) (associated to α^N). By independence condition in Assumption (5.33), recall that \widehat{m}^N is given in equation (5.34),

$$\widehat{m}_t^{i,N} = \widehat{m}_t^N, \widehat{\mathbb{P}}\text{-a.e.}$$

and

$$\text{given the } \sigma\text{-field } \widehat{\mathcal{G}}_t^N, \text{ for } i \neq j, (X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N}) \text{ are independent of } (X_{t \wedge \cdot}^{j,N}, \alpha_t^{j,N}) \quad (\text{A.6})$$

and $\mathcal{L}^{\widehat{\mathbb{P}}}(X^{i,N}, \xi^i, \bar{\Lambda}^N, \phi(\mu^N), \zeta^N, W^i, B^N, Z^i) = \mathcal{L}^{\widehat{\mathbb{P}}}(X^N, \xi, \bar{\Lambda}^N, \phi(\mu^N), \zeta^N, W, B^N, Z)$ for each i .

Let us introduce for each $N \in \mathbb{N}^*$, the measure on $[0, T] \times \mathcal{P}(\mathcal{C}^n \times U) \times \mathcal{P}(\mathcal{C}^n \times U)$

$$\Gamma_t^N(de, de') dt := \mathbb{E}^{\widehat{\mathbb{P}}} \left[\delta(\bar{\beta}_t^N, \mathcal{L}^{\mathbb{P}}(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N} | \widehat{\mathcal{G}}_t^N))(de, de') dt \right],$$

with $\bar{\beta}_t^N(dx, du) := \frac{1}{N} \sum_{i=1}^N \delta_{(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N})}(dx, du)$. As $(\hat{b}, \hat{\sigma})$ are bounded and $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, it is straightforward to check that $\sup_{N \geq 1} \sup_{i \in \{1, \dots, N\}} \mathbb{E}^{\widehat{\mathbb{P}}} [\sup_{t \in [0, T]} |X_t^{i,N}|^{p'}] < \infty$, and hence $(\Gamma^N)_{N \in \mathbb{N}^*}$ is relatively compact for the Wasserstein metric \mathcal{W}_p . Denote by Γ^∞ the limit of a sub-sequence of $(\Gamma^N)_{N \in \mathbb{N}^*}$. For simplicity, we will use the same notation for the sequence and the sub-sequence. One gets

$$\Gamma_t^\infty(de, de') dt = \delta_e(de') \Gamma_t^\infty(de, \mathcal{P}(\mathcal{C}^n \times U)) dt. \quad (\text{A.7})$$

It is enough to show that: for all $Q \in \mathbb{N}^*$, any bounded functions $(f^q)_{q \in \{1, \dots, Q\}} : \mathcal{C}^n \times U \rightarrow \mathbb{R}^Q$ and $g : [0, T] \times \mathcal{P}(\mathcal{C}^n \times U) \rightarrow \mathbb{R}$

$$\begin{aligned} & \int_0^T \int_{\mathcal{P}(\mathcal{C}^n \times U)^2} \prod_{q=1}^Q \langle f^q, e \rangle g(t, e') \Gamma_t^\infty(de, de') dt \\ &= \int_0^T \int_{\mathcal{P}(\mathcal{C}^n \times U)} \prod_{q=1}^Q \langle f^q, e \rangle g(t, e) \Gamma_t^\infty(de, \mathcal{P}(\mathcal{C}^n \times U)) dt. \end{aligned}$$

Let us prove this result when $Q = 2$, the case $Q \in \mathbb{N}^*$ is true by similar way.

$$\begin{aligned} & \int_0^T \int_{\mathcal{P}(\mathcal{C}^n \times U)^2} \prod_{q=1}^Q \langle f^q, e \rangle g(t, e') \Gamma_t^\infty(de, de') dt \\ &= \lim_N \frac{1}{N} \sum_{i,j=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T f^1(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N}) f^2(X_{t \wedge \cdot}^{j,N}, \alpha_t^{j,N}) g(t, \widehat{m}_t^N) dt \right] \\ &= \lim_N \frac{1}{N^2} \left(\sum_{i \neq j} \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T f^1(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N}) f^2(X_{t \wedge \cdot}^{j,N}, \alpha_t^{j,N}) g(t, \widehat{m}_t^N) dt \right] \right. \\ & \quad \left. + \sum_{i=1}^N \mathbb{E}^{\widehat{\mathbb{P}}} \left[\int_0^T f^1(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N}) f^2(X_{t \wedge \cdot}^{i,N}, \alpha_t^{i,N}) g(t, \widehat{m}_t^N) dt \right] \right) \end{aligned}$$

$$\begin{aligned}
 &= \lim_N \left(\frac{1}{N^2} \sum_{i \neq j} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathbb{E}^{\hat{\mathbb{P}}} [f^1(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) | \hat{\mathcal{G}}_t^N] \mathbb{E}^{\hat{\mathbb{P}}} [f^2(X_{t\wedge\cdot}^{j,N}, \alpha_t^{j,N}) | \hat{\mathcal{G}}_t^N] g(t, \hat{m}_t^N) dt \right] \right. \\
 &\quad \left. + \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f^1(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) f^2(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) g(t, \hat{m}_t^N) dt \right] \right) \\
 &= \lim_N \left(\mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \langle f^1, \hat{m}_t^N \rangle \langle f^2, \hat{m}_t^N \rangle g(t, \hat{m}_t^N) dt \right] \right. \\
 &\quad - \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathbb{E}^{\hat{\mathbb{P}}} [f^1(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) | \hat{\mathcal{G}}_t^N] \mathbb{E}^{\hat{\mathbb{P}}} [f^2(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) | \hat{\mathcal{G}}_t^N] g(t, \hat{m}_t^N) dt \right] \\
 &\quad \left. + \frac{1}{N^2} \sum_{i=1}^N \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f^1(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) f^2(X_{t\wedge\cdot}^{i,N}, \alpha_t^{i,N}) g(t, \hat{m}_t^N) dt \right] \right) \\
 &= \int_0^T \int_{\mathcal{P}(\mathbb{R}^n \times U)} \langle f^1, e \rangle \langle f^2, e \rangle g(t, e) \Gamma_t^\infty (de, \mathcal{P}(\mathbb{C}^n \times U)) dt,
 \end{aligned}$$

where we used result (A.6) and the fact that the terms starting with $\frac{1}{(N_i)^2} \sum_{i=1}^{N_i}$ go to zero because (f^1, f^2, g) are bounded.

Next, for all $t \in [0, T]$, using Lipschitz property, there exists a constant $C > 0$ (which changes from line to line)

$$\begin{aligned}
 &\mathbb{E}^{\hat{\mathbb{P}}} \left[\sup_{s \in [0, t]} |X_s^{i,N} - \hat{X}_s^i|^p \right] \\
 &\leq C \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^t \sup_{r \in [0, s]} |X_r^{i,N} - \hat{X}_r^i|^p \right. \\
 &\quad \left. + \sup_{r \in [0, s]} \mathcal{W}_p(\hat{\mu}_r^N, \frac{1}{N} \sum_{i=1}^N \delta_{X_r^{i,N}})^p + \mathcal{W}_p(\hat{m}_s^N, \frac{1}{N} \sum_{i=1}^N \delta_{(X_s^{i,N}, \alpha_s^{i,N})})^p ds \right] \\
 &\leq C \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^t \sup_{r \in [0, s]} |X_r^{i,N} - \hat{X}_r^i|^p + \mathcal{W}_p(\mathcal{L}^{\hat{\mathbb{P}}}((X_{s\wedge\cdot}^{i,N}, \alpha_s^{i,N}) | \hat{\mathcal{G}}_s^N), \bar{\beta}_s^N)^p ds \right],
 \end{aligned}$$

recall that $(\hat{X}^1, \dots, \hat{X}^N)$ are defined in equation (5.17) (in Theorem 5.6), and $\hat{m}_t^N := \mathcal{L}^{\hat{\mathbb{P}}}(X_t^N, \alpha_t^N | \hat{\mathcal{G}}_t^N)$ and $\hat{\mu}_t^N := \mathcal{L}^{\hat{\mathbb{P}}}(X_t^N | \hat{\mathcal{G}}_t^N)$.

Then by Gronwall Lemma, we find

$$\mathbb{E}^{\hat{\mathbb{P}}} \left[\sup_{s \in [0, T]} |X_s^{i,N} - \hat{X}_s^i|^p \right] \leq C \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\mathcal{L}^{\hat{\mathbb{P}}}((X_{s\wedge\cdot}^{i,N}, \alpha_s^{i,N}) | \hat{\mathcal{G}}_s^N), \bar{\beta}_s^N)^p ds \right].$$

As,

$$\begin{aligned}
 &\mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\hat{m}_s^N, \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{X}_s^{i,N}, \alpha_s^{i,N})})^p ds \right] \\
 &\leq \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\hat{m}_s^N, \frac{1}{N} \sum_{i=1}^N \delta_{(X_s^{i,N}, \alpha_s^{i,N})})^p ds \right] \\
 &\quad + \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\frac{1}{N} \sum_{i=1}^N \delta_{(X_s^{i,N}, \alpha_s^{i,N})}, \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{X}_s^{i,N}, \alpha_s^{i,N})})^p ds \right]
 \end{aligned}$$

$$\begin{aligned} &\leq C \left(\mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\mathcal{L}^{\hat{\mathbb{P}}}((X_s^{i,N}, \alpha_s^{i,N} | \hat{\mathcal{G}}_s^N), \bar{\beta}_s^N)^p ds) \right] + \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T |X_s^{i,N} - \hat{X}_s^i|^p ds \right] \right) \\ &\leq C \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\mathcal{L}^{\hat{\mathbb{P}}}((X_s^{i,N}, \alpha_s^{i,N} | \hat{\mathcal{G}}_s^N), \bar{\beta}_s^N)^p ds) \right], \end{aligned}$$

therefore, by taking the sub-sequence corresponding to the lim sup, by result (A.7),

$$\limsup_{l \rightarrow \infty} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \mathcal{W}_p(\hat{m}_s^{N_l}, \frac{1}{N_l} \sum_{i=1}^{N_l} \delta_{(\hat{X}_s^{i,N}, \alpha_s^{i,N})})^p ds + \sup_{t \in [0, T]} \mathcal{W}_p(\phi_t(\hat{\mu}^{N_l}), \phi_t(\mu^{N_l})) \right] = 0.$$

From all previous results, it is straightforward to check that

$$\begin{aligned} &\lim_{N \rightarrow \infty} \mathcal{W}_p(\mathcal{L}^{\hat{\mathbb{P}}}(\hat{\mu}^N, \zeta^N, \delta_{(\hat{m}_s^N, \bar{\nu}_s^N)})(dm, d\bar{\nu}) ds, B^N), \mathcal{L}^{\hat{\mathbb{P}}}(\hat{\gamma}^N, \zeta^N, \delta_{(\hat{\theta}_s^N, \bar{\nu}_s^N)})(dm, d\bar{\nu}) ds, B^N) \\ &= 0, \end{aligned}$$

where $\hat{\gamma}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{X}_t^{i,N})}$ and $\hat{\theta}_t := \frac{1}{N} \sum_{i=1}^N \delta_{(\hat{X}_t^{i,N}, \alpha_t^{i,N})}$. Consequently, by Theorem 5.6

$$\begin{aligned} &\lim_{k \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}}(\hat{\mu}^{N_k}, \zeta^{N_k}, \delta_{(\hat{m}_s^{N_k}, \bar{\nu}_s^{N_k})})(dm, d\bar{\nu}) ds, B^{N_k}) \\ &= \lim_{k \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}}(\hat{\gamma}^{N_k}, \zeta^{N_k}, \delta_{(\hat{\theta}_s^{N_k}, \bar{\nu}_s^{N_k})})(dm, d\bar{\nu}) ds, B^{N_k}) = \mathcal{L}^{\mathbb{Q}}(\mu, \zeta, \bar{\Lambda}, B), \end{aligned}$$

recall that $\hat{m}_t^N := \mathcal{L}^{\hat{\mathbb{P}}}(X_t^N, \alpha_t^N | \hat{\mathcal{G}}_t^N)$ and $\hat{\mu}_t^N := \mathcal{L}^{\hat{\mathbb{P}}}(X_t^N | \hat{\mathcal{G}}_t^N)$. □

Proof of Theorem 5.9. The proof of this Proposition is exactly the same as Proposition 5.6, we essentially recall the main step.

Approximation by SDE : Tightness and identification of the limit process:

We denote by $Z^{\varepsilon, N}$ the unique strong solution of:

$$\begin{aligned} Z_t^{\varepsilon, N} &= \xi + \int_0^t \hat{b}^\varepsilon[B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_r^N](r, Z_r^{\varepsilon, N}) dr \\ &\quad + \int_0^t (\hat{a}^\varepsilon)^{1/2}[B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}_r^N](r, Z_r^{\varepsilon, N}) dW_r, \quad t \in [0, T], \quad \hat{\mathbb{P}}\text{-a.e.} \end{aligned}$$

And for all $(t, \omega) \in [0, T] \times \Omega$, denote $\vartheta_t^{\varepsilon, N}(\omega) := \mathcal{L}^{\hat{\mathbb{P}}}(Z_t^{\varepsilon, N} | \hat{\mathcal{G}}_t^N)(\omega)$, and

$$\mathbb{P}^{\varepsilon, N} := \mathcal{L}^{\hat{\mathbb{P}}}(\vartheta^{\varepsilon, N}, B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}^N) \in \mathcal{P}(\mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2)).$$

As $[\hat{b}^\varepsilon, \hat{a}^\varepsilon]$ are bounded, again it is straightforward to check that $(\mathbb{P}^{\varepsilon, N})_{N \in \mathbb{N}^*}$ is relatively compact for the Wasserstein metric \mathcal{W}_p . Denote by $\mathbb{P}^{\varepsilon, \infty}$ the limit of a sub-sequence of $(\mathbb{P}^{\varepsilon, N})_{N \in \mathbb{N}^*}$. Therefore, under Theorem 2.1, by applying similar techniques to those used in step 2.2 of proof of Theorem 5.6, one gets for all $(f, t) \in C_b^2(\mathbb{R}^n; \mathbb{R}) \times [0, T]$, one gets

$$\langle f, \beta_t \rangle = \int_{\mathbb{R}^n} f(y) \nu(dy) + \int_0^t \int_{\mathbb{R}^n} \mathcal{A}_r^\varepsilon f[B, \beta^\mu, \beta^\zeta, \bar{\beta}](x) \beta_r(dx) dr, \quad \mathbb{P}^{\varepsilon, \infty}\text{-a.e.}, \quad (\text{A.8})$$

where $(\beta, B, \beta^\mu, \beta^\zeta, \bar{\beta})$ is the canonical element on $\mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}^\ell \times \mathcal{C}_{\mathcal{W}}^n \times \mathcal{C}_{\mathcal{W}}^n \times \mathbb{M}((\mathcal{P}_U^n)^2)$. Using a countable family of (f, t) , we can deduce $\mathbb{P}^{\varepsilon, \infty}$ -a.e. equation (A.8) is true for all $(f, t) \in C_b^2(\mathbb{R}^n; \mathbb{R}) \times [0, T]$. By Theorem A.6, one has $\beta = \Phi^\varepsilon(B, \beta^\mu, \beta^\zeta, \bar{\beta})$ where Φ^ε is the function used in equation (5.20). Also

$$\begin{aligned} &\mathcal{L}^{\mathbb{P}^{\varepsilon, \infty}}(B, \beta^\mu, \beta^\zeta, \bar{\beta}) = \lim_{N \rightarrow \infty} \mathcal{L}^{\mathbb{P}^{\varepsilon, N}}(B, \beta^\mu, \beta^\zeta, \bar{\beta}) \\ &= \lim_{N \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}}(B, \phi(\mu), \zeta^N, \bar{\Lambda}^N) = \mathcal{L}^{\mathbb{Q}}(B, \phi(\mu), \zeta, \bar{\Lambda}). \end{aligned}$$

Then $\mathcal{L}^{\mathbb{P}^{\varepsilon, \infty}}(\beta, B, \beta^\mu, \beta^\zeta, \bar{\beta}) = \mathcal{L}^{\mathbb{Q}}(\mu^\varepsilon, B, \phi(\mu), \zeta, \bar{\Lambda})$. This result is true for any limit of any sub-sequence of $(\mathbb{P}^{\varepsilon, N})_{N \in \mathbb{N}^*}$, consequently $(\mathbb{P}^{\varepsilon, N})_{N \in \mathbb{N}^*}$ converges and

$$\lim_{N \rightarrow \infty} \mathcal{L}^{\hat{\mathbb{P}}}(\vartheta^{\varepsilon, N}, B^N, \phi(\mu^N), \zeta^N, \bar{\Lambda}^N) = \mathcal{L}^{\mathbb{Q}}(\mu^\varepsilon, B, \phi(\mu), \zeta, \bar{\Lambda}).$$

Last approximation: Let us consider for all $(\varepsilon, N) \in (0, \infty) \times \mathbb{N}^*$, the $\hat{\mathbb{F}}$ -adapted \mathbb{R}^n -valued continuous process $X^{\varepsilon, N} := X$ strong solution of: $X_0 = \xi$, for all $s \in [0, T]$

$$dX_s =$$

$$\int_{(\mathcal{P}_U^n)^2} \hat{b}(s, X_s, B^N, \phi(\hat{\mu}^{\varepsilon, N}), \zeta^N, \hat{\mathbf{m}}_s^{\varepsilon, N}[m], \bar{\nu}, u) H^\varepsilon(Z_s^{\varepsilon, N}, m)(du) \bar{\Lambda}_s^N(dm, d\bar{\nu}) ds + R_s dW_s$$

with

$$R_s :=$$

$$\left(\int_{(\mathcal{P}_U^n)^2} \hat{\sigma} \hat{\sigma}^\top(s, X_s, B^N, \phi(\hat{\mu}^{\varepsilon, N}), \zeta^N, \hat{\mathbf{m}}_s^{\varepsilon, N}[m], \bar{\nu}, u) H^\varepsilon(Z_s^{\varepsilon, N}, m)(du) \bar{\Lambda}_s^N(dm, d\bar{\nu}) \right)^{1/2}$$

where recall that $H^\varepsilon(x, m)(du) := \int_{\mathbb{R}^n} m(du, dy) \frac{G_\varepsilon(x-y)}{(m(U, dz))^{\varepsilon}(x)}$ and

$$\hat{\mathbf{m}}_r^{\varepsilon, N}[m](dz, du) := \mathbb{E}^{\hat{\mathbb{P}}} \left[H^\varepsilon(Z_r^{\varepsilon, N}, m)(du) \delta_{X_r^{\varepsilon, N}}(dz) \middle| \hat{\mathcal{G}}_r^N \right] \text{ and } \hat{\mu}_r^{\varepsilon, N} := \mathcal{L}^{\hat{\mathbb{P}}}(X_r^{\varepsilon, N} | \hat{\mathcal{G}}_r^N).$$

Combining Proposition A.5 and the techniques applied in *step 3* of Proof of Proposition 5.6, one gets

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\hat{\mathbb{P}}} \left[\sup_{t \in [0, T]} |X_t^{\varepsilon, N} - Z_t^{\varepsilon, N}|^p \right] = 0$$

and

$$\lim_{\varepsilon \rightarrow 0} \lim_{N \rightarrow \infty} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T \int_{\mathcal{P}_U^n} \mathcal{W}_p(\hat{\mathbf{m}}_r^{\varepsilon, N}[m], m) \bar{\Lambda}_r^N(dm, \mathcal{P}_U^n) dr \right] = 0.$$

Similarly, $\limsup_{\varepsilon \rightarrow 0} \limsup_{N \rightarrow \infty} \mathbb{E}^{\hat{\mathbb{P}}} \left[\sup_{s \in [0, T]} \mathcal{W}_p(\phi(\hat{\mu}^{\varepsilon, N}), \phi_s(\mu^N)) \right] = 0$. $X^{\varepsilon, N}$ is the process we are looking for. □

A.2 Regularization by convolution and consequence

This part presents results about the approximation of Borel measurable functions through a sequence of “smooth” functions. The main point is that this approximation is achieved via a convolution. The convolution is realized by a probability measure constructed by an SDE process. Before presenting the main results, we start by recalling an equivalence result coming from [18, Proposition 4.2].

Let $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ be a filtered probability space supporting W a \mathbb{R}^n -valued \mathbb{F} -Brownian motion and ξ a \mathcal{F}_0 -random variable verifying $\mathbb{E}^{\mathbb{P}}[|\xi|^p] < \infty$, $(b_t, \sigma_t)_{t \in [0, T]} \mathbb{R}^n \times \mathbb{S}^n$ bounded predictable process such that there exists $\theta > 0$ satisfying $[\sigma_t][\sigma_t]^\top \geq \theta I_{n \times n}$. For all $t \in [0, T]$, denote by

$$X_t = \xi + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad \mathbb{P}\text{-a.e.}$$

the following proposition is just an application of [18, Proposition 4.2] (see also [21])

Proposition A.1 (equivalence of measures). *With the previous considerations, the measure \mathbf{n} on $\mathbb{R}^n \times [0, T]$ defined by*

$$\mathbf{n}(dx, dt) := \mathbb{P} \circ (X_t)^{-1}(dx) dt$$

is equivalent to the Lebesgue measure on $\mathbb{R}^n \times [0, T]$.

Approximation by convolution We set $G \in C^\infty(\mathbb{R}^n; \mathbb{R})$ with compact support satisfying $G \geq 0$, $G(x) = G(-x)$ for $x \in \mathbb{R}^n$, and $\int_{\mathbb{R}^n} G(y)dy = 1$. Let us introduce $G_\varepsilon(x) := \varepsilon^{-n}G(\varepsilon^{-1}x)$ for all $x \in \mathbb{R}^n$.

Let X^k be the process defined by

$$X_t^k = \xi + \int_0^t b_r^k dr + \int_0^t \sigma_r^k dW_r \text{ for all } t \in [0, T], \text{ P-a.e.,}$$

where there exists $D > 0$ s.t. for all k and t , $|\sigma_t^k| + |b_t^k| \leq D$, P-a.e., $[\sigma_t^k][\sigma_t^k]^\top \geq \theta I_{n \times n}$, P-a.e. In addition $\mathbb{E}^\mathbb{P}[|\xi|^p] < \infty$ where $p \geq 1$. Also, we take $(\mathbf{n}_t)_{t \in [0, T]} \in \mathcal{C}_W^n$ such that $\mathbf{n}_t(dx)dt$ is equivalent to the Lebesgue measure on $[0, T] \times \mathbb{R}^n$, and for the weak topology,

$$\lim_{k \rightarrow \infty} \mathcal{L}^\mathbb{P}(X_t^k) = \mathbf{n}_t \text{ for each } t \in [0, T].$$

The following proposition shows that it is possible to approach some bounded measurable functions via smooth functions (bounded derivative functions) by using the marginal distributions of X^k . We consider $(\varepsilon_k)_{k \in \mathbb{N}^*} \subset (0, \infty)$ such that $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. We pose $G_k(x) = G_{\varepsilon_k}(x)$ and for $\pi \in \mathcal{P}(\mathbb{R}^n)$, $\pi^{(k)}(x) := \int_{\mathbb{R}^n} G_k(x - y)\pi(dy)$, for all $x \in \mathbb{R}^n$.

Proposition A.2 (regularization by convolution). *For any bounded Borel measurable function $\varphi : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^q$, such that for all $(t, z) \in [0, T] \times \mathbb{R}^n$, $\varphi(t, \cdot, z) : y \in \mathbb{R}^n \rightarrow \varphi(t, y, z) \in \mathbb{R}^q$ is continuous, one has*

$$\lim_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_k(t, x - y)}{(\mathbf{n}_t)^{(k)}(x)} \mathbf{n}_t(dy) - \varphi(t, x, x) \right| \mathbf{n}_t(dx)dt = 0 \tag{A.9}$$

and

$$\lim_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \mathbb{E}^\mathbb{P} \left[\int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_k(t, X_t^k - y)}{(\mathbf{n}_t)^{(k)}(X_t^k)} \mathbf{n}_t(dy) \right] - \int_{\mathbb{R}^n} \varphi(t, x, x) \mathbf{n}_t(dx) \right| dt = 0.$$

Proof. Mention that, as $\mathbf{n}_t(dx)dt$ is equivalent to the Lebesgue measure on $[0, T] \times \mathbb{R}^n$, there exists Borel measurable function $c : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that $c(s, z) > 0$ $dt \otimes dx$ a.e. $(s, z) \in [0, T] \times \mathbb{R}^n$, and $\mathbf{n}_t(dx)dt = c(t, x)dxdt$.

First, let us prove the result (A.9). If

$$A_k := \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_k(x - y)}{(\mathbf{n}_t)^{(k)}(x)} \mathbf{n}_t(dy) - \varphi(t, x, x) \right| \mathbf{n}_t(dx)dt,$$

one finds that

$$\begin{aligned} & \left| A_k - \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{\varphi(t, x, y) - \varphi(t, x, x)\} G_k(x - y) c(t, y) dy \right| dxdt \right| \\ &= \left| \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{\varphi(t, x, y) - \varphi(t, x, x)\} G_k(x - y) c(t, y) dy \right| \left(\frac{c(t, x)}{(\mathbf{n}_t)^{(k)}(x)} - 1 \right) dxdt \right| \\ &\leq K \left| \int_0^T \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} G_k(x - y) c(t, y) dy \left| \frac{c(t, x)}{(\mathbf{n}_t)^{(k)}(x)} - 1 \right| dxdt \right| \\ &= K \left| \int_0^T \int_{\mathbb{R}^n} (\mathbf{n}_t)^{(k)}(x) \left| \frac{c(t, x)}{(\mathbf{n}_t)^{(k)}(x)} - 1 \right| dxdt \right| \\ &\leq K \left| \int_0^T \int_{\mathbb{R}^n} |c(t, x) - (\mathbf{n}_t)^{(k)}(x)| dxdt \right| \\ &= K \left| \int_0^T \int_{\mathbb{R}^n} \left| c(t, x) - \int_{\mathbb{R}^n} G_k(x - y) c(t, y) dy \right| dxdt \right| \rightarrow_{k \rightarrow \infty} 0, \tag{A.10} \end{aligned}$$

where the first inequality is true because φ is bounded and the last result is obtained by the classical result of approximation by convolution.

Now, for all $(t, y, \delta) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}_+^*$, $v(t, y, \delta) := \sup_{z \mid |y-z| \leq \delta} |\varphi(t, y, y) - \varphi(t, z, y)|$, notice that $\lim_{\delta \rightarrow 0} v(t, y, \delta) = 0$. Observe that

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{ \varphi(t, x, y) - \varphi(t, y, y) \} G_k(x - y) c(t, y) dy \right| dx dt \\ &= \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{ \varphi(t, x, y) - \varphi(t, y, y) \} (1_{|x-y| \leq \delta} + 1_{|x-y| > \delta}) G_k(x - y) c(t, y) dy \right| dx dt \\ &\leq \int_0^T \int_{\mathbb{R}^n} v(t, y, \delta) \int_{\mathbb{R}^n} 1_{|x-y| \leq \delta} G_k(x - y) c(t, y) dy dx dt \\ &\quad + K \int_0^T \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} 1_{|x-y| > \delta} G_k(x - y) c(t, y) dy dx dt \\ &\leq \int_0^T \int_{\mathbb{R}^n} v(t, y, \delta) \int_{\mathbb{R}^n} G_k(x - y) c(t, y) dy dx dt + K T \int_{\mathbb{R}^n} 1_{|z| > \delta} G_k(z) dz \\ &\leq \int_0^T \int_{\mathbb{R}^n} v(t, y, \delta) c(t, y) dy dt + K T \int_{\mathbb{R}^n} 1_{|z| > \delta} G_k(z) dz, \end{aligned}$$

it is well known, for each $\delta > 0$, $\lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} 1_{|z| > \delta} G_k(z) dz = 0$, one gets that

$$\begin{aligned} & \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{ \varphi(t, x, y) - \varphi(t, y, y) \} G_k(x - y) c(t, y) dy \right| dx dt \\ & \leq \lim_{\delta \rightarrow 0} \int_0^T \int_{\mathbb{R}^n} v(t, x, \delta) c(t, x) dx dt = 0, \end{aligned} \tag{A.11}$$

the last inequality is true because of Lebesgue's dominated convergence theorem. Finally, one has that

$$\begin{aligned} & \limsup_{k \rightarrow \infty} A_k \\ &= \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \{ \varphi(t, x, y) - \varphi(t, x, x) \} G_k(x - y) c(t, y) dy \right| dx dt \\ &= \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, x, y) c(t, y) G_k(x - y) dy - \int_{\mathbb{R}^n} \varphi(t, x, x) G_k(x - y) c(t, y) dy \right| dx dt \\ &= \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, y, y) c(t, y) G_k(x - y) dy - \int_{\mathbb{R}^n} \varphi(t, x, x) G_k(x - y) c(t, y) dy \right| dx dt \\ &\leq \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, y, y) c(t, y) G_k(x - y) dy - \varphi(t, x, x) c(t, x) \right| dx dt \\ &\quad + \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \varphi(t, x, x) c(t, x) - \int_{\mathbb{R}^n} \varphi(t, x, x) G_k(x - y) c(t, y) dy \right| dx dt \\ &\leq \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, y, y) c(t, y) G_k(x - y) dy - \varphi(t, x, x) c(t, x) \right| dx dt \\ &\quad + \limsup_{k \rightarrow \infty} K \int_0^T \int_{\mathbb{R}^n} \left| c(t, x) - \int_{\mathbb{R}^n} G_k(x - y) c(t, y) dy \right| dx dt = 0, \end{aligned}$$

where the first equality derived from (A.10), the third equality follows from (A.11) and we find 0 because of approximation by convolution result. Therefore $\lim_{k \rightarrow \infty} A_k = 0$, then the first assertion is proved.

For the second point, let $k_0 \in \mathbb{N}^*$, one has that

$$\begin{aligned} S^k(\varphi) &:= \int_0^T \left| \mathbb{E}^{\mathbb{P}} \left[\int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_k(t, X_t^k - y)}{(\mathbf{n}_t)^{(k)}(X_t^k)} \mathbf{n}_t(dy) \right] - \int_{\mathbb{R}^n} \varphi(t, x, x) \mathbf{n}_t(dx) \right| dt \\ &\leq \\ &\int_0^T \left| \mathbb{E}^{\mathbb{P}} \left[\int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_k(t, X_t^k - y)}{(\mathbf{n}_t)^{(k)}(X_t^k)} \mathbf{n}_t(dy) - \int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_{k_0}(t, X_t^k - y)}{(\mathbf{n}_t)^{(k_0)}(X_t^k)} \mathbf{n}_t(dy) \right] \right| dt \\ &\quad + \int_0^T \left| \mathbb{E}^{\mathbb{P}} \left[\int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_{k_0}(t, X_t^k - y)}{(\mathbf{n}_t)^{(k_0)}(X_t^k)} \mathbf{n}_t(dy) \right] - \int_{\mathbb{R}^n} \varphi(t, x, x) \mathbf{n}_t(dx) \right| dt. \end{aligned}$$

By [21, Chapter 2 Section 3 Theorem 4] and Markov inequality, for each $R > 0$, there exists a constant $C > 0$ depending only on (D, θ, T, R) satisfying

$$\begin{aligned} S^k(\varphi) &\leq \\ &C \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_k(t, x - y)}{(\mathbf{n}_t)^{(k)}(x)} \mathbf{n}_t(dy) \right. \\ &\quad \left. - \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_{k_0}(t, x - y)}{(\mathbf{n}_t)^{(k_0)}(x)} \mathbf{n}_t(dy) \right|^n 1_{|x| \leq R} dx dt \\ &\quad + T \frac{\mathbb{E}^{\mathbb{P}}[\sup_{t \in [0, T]} |X_t^k|^p]}{R^p} \\ &\quad + \int_0^T \left| \mathbb{E}^{\mathbb{P}} \left[\int_{\mathbb{R}^n} \varphi(t, X_t^k, y) \frac{G_{k_0}(t, X_t^k - y)}{(\mathbf{n}_t)^{(k_0)}(X_t^k)} \mathbf{n}_t(dy) \right] - \int_{\mathbb{R}^n} \varphi(t, x, x) \mathbf{n}_t(dx) \right| dt. \end{aligned}$$

By using the first statement of the proposition (see proof above), there exists a subsequence $(k_j)_{j \in \mathbb{N}^*} \subset \mathbb{N}^*$ such that: a.e. $(s, z) \in [0, T] \times \mathbb{R}^n$,

$$\lim_{j \rightarrow \infty} \left| \int_{\mathbb{R}^n} \varphi(s, z, y) \frac{G_{k_j}(t, x - y)}{(\mathbf{n}_t)^{(k_j)}(x)} \mathbf{n}_t(dy) - \varphi(s, z, z) \right| = 0, \mathbf{n}_t(dx) dt.$$

As $\mathbf{n}_t(dx) dt$ is equivalent to the Lebesgue measure on $[0, T] \times \mathbb{R}^n$,

$$\lim_{j \rightarrow \infty} \left| \int_{\mathbb{R}^n} \varphi(s, z, y) \frac{G_{k_j}(t, x - y)}{(\mathbf{n}_t)^{(k_j)}(x)} \mathbf{n}_t(dy) - \varphi(s, z, z) \right| = 0, dt \otimes dx \text{ a.e. } (s, z) \in [0, T] \times \mathbb{R}^n.$$

All these observations allow us to say, by Lebesgue's dominated convergence theorem

$$\begin{aligned} \limsup_{k_0 \rightarrow \infty} \limsup_{k \rightarrow \infty} \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_k(t, x - y)}{(\mathbf{n}_t)^{(k)}(x)} \mathbf{n}_t(dy) \right. \\ \left. - \int_{\mathbb{R}^n} \varphi(t, x, y) \frac{G_{k_0}(t, x - y)}{(\mathbf{n}_t)^{(k_0)}(x)} \mathbf{n}_t(dy) \right|^n 1_{|x| \leq R} dx dt = 0. \end{aligned}$$

Finally, combining the previous result with the weak convergence, $\lim_{k \rightarrow \infty} \mathcal{L}^{\mathbb{P}}(X_t^k) = \mathbf{n}_t$ for each $t \in [0, T]$, and an obvious application of the first statement of the proposition,

one gets

$$\begin{aligned} & \limsup_{k \rightarrow \infty} S^k(\varphi) \\ & \leq \limsup_{k_0, k \rightarrow \infty} C \int_0^T \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} h(t, x, y) \frac{G_k(t, x - y)}{(\mathbf{n}_t)^{(k)}(x)} \mathbf{n}_t(dy) \right. \\ & \quad \left. - \int_{\mathbb{R}^n} h(t, x, y) \frac{G_{k_0}(t, x - y)}{(\mathbf{n}_t)^{(k_0)}(x)} \mathbf{n}_t(dy) \right|^n 1_{|x| \leq R} dx dt \\ & + \limsup_{l_0 \rightarrow \infty} \int_0^T \left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} h(t, x, y) \frac{G_{l_0}(t, x - y)}{(\mathbf{n}_t)^{(l_0)}(x)} \mathbf{n}_t(dy) \mathbf{n}_t(dx) dt \right. \\ & \quad \left. - \int_0^T \int_{\mathbb{R}^n} h(t, x, x) \mathbf{n}_t(dx) dt \right| \\ & + T \frac{\sup_{k > 0} \mathbb{E}^{\mathbb{P}}[\sup_{t \in [0, T]} |X_t^k|^p]}{R^p} \leq T \frac{\sup_{k > 0} \mathbb{E}^{\mathbb{P}}[\sup_{t \in [0, T]} |X_t^k|^p]}{R^p}, \end{aligned}$$

as $\sup_{k > 0} \mathbb{E}^{\mathbb{P}}[\sup_{t \in [0, T]} |X_t^k|^p] < \infty$, by taking $R \rightarrow \infty$, we deduce the result. \square

The next result is essentially an application of Theorem A.2. It states the result of Theorem A.2 under a form usually used in the paper. Let us consider the map $(\hat{b}, \hat{\sigma}) : [0, T] \times \mathbb{R}^n \times \mathcal{P}_U^n \times \mathcal{P}_U^n \times U \rightarrow \mathbb{R}^n \times \mathbb{S}^n$ and $\hat{\mathbf{q}} \in \mathbb{M}((\mathcal{P}_U^n)^2)$ s.t. $\hat{\mathbf{q}}_t(\mathbb{Z}_{\mathbf{n}_t} \times \mathcal{P}_U^n) = 1$ dt-for almost every $t \in [0, T]$. Recall that $\lim_{k \rightarrow \infty} \mathcal{L}(X_t^k) = \mathbf{n}_t$ in \mathcal{W}_p for all $t \in [0, T]$. We pose $\mathbf{n}_t^k := \mathcal{L}(X_t^k)$.

Corollary A.3. *One has that*

$$\begin{aligned} & \lim_{k \rightarrow \infty} \left[\int_0^T \int_{(\mathcal{P}_U^n)^2} \left[\int_{\mathbb{R}^n} |K^k(r, x, m, m')|^p \mathbf{n}_r^k(dx) \right. \right. \\ & \quad \left. \left. + \mathcal{W}_p \left(H^k(z, m)(du) \mathbf{n}_r^k(dz), m(du, dz) \right)^p \right] \hat{\mathbf{q}}_r(dm, dm') dr \right] = 0, \end{aligned}$$

where

$$\begin{aligned} & K^k(s, x, m, \bar{\nu}) \\ & := \\ & \int_{\mathbb{R}^n \times U} [\hat{b}, \hat{\sigma} \hat{\sigma}^\top](s, y, m, \bar{\nu}, u) \bar{H}^k(x, m)(dy, du) - \int_U [\hat{b}, \hat{\sigma} \hat{\sigma}^\top](s, x, m, \bar{\nu}, u) H^k(x, m)(du), \end{aligned}$$

with

$$\bar{H}^k(x, m)(dy, du) := m(dy, du) \frac{G_k(x - y)}{(m(U, dz))^{(k)}(x)}$$

and

$$H^k(x, m)(du) := \int_{\mathbb{R}^n} \bar{H}^k(x, m)(du, dy).$$

Proof. As $\hat{\mathbf{q}}_t(\mathbb{Z}_{\mathbf{n}_t} \times \mathcal{P}_U^n) = 1$ dt-almost surely $t \in [0, T]$, using convex inequality and Proposition A.2,

$$\lim_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n} |K^k(r, x, m, \bar{\nu})|^p \mathbf{n}_r^k(dx) \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr$$

$$\begin{aligned} &\leq \limsup_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n \times U} \left| [\hat{b}, \hat{\sigma} \hat{\sigma}^\top](r, x, m, \bar{\nu}, u) \right. \\ &\quad \left. - [\hat{b}, \hat{\sigma} \hat{\sigma}^\top](r, y, m, \bar{\nu}, u) \right|^p \frac{G_k(x-y)}{(\mathbf{n}_r)^{(k)}(x)} m^y(du) \mathbf{n}_r(dy) \mathbf{n}_r^k(dx) \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr \\ &= 0. \end{aligned}$$

For all bounded continuous function $h : \mathbb{R}^n \times U \rightarrow \mathbb{R}$, using Proposition A.2 again,

$$\begin{aligned} &\lim_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \left| \int_{\mathbb{R}^n \times U} h(x, u) H^k(x, m)(du) \mathbf{n}_r^k(dx) \right. \\ &\quad \left. - \int_{\mathbb{R}^n \times U} h(z, u) m(dz, du) \right| \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr \\ &\leq \lim_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n \times U} h(x, u) m_r^y(du) \frac{G_k(y-x)}{(\mathbf{n}_r)^{(k)}(x)} \mathbf{n}_r(dy) \right. \\ &\quad \left. - \int_{\mathbb{R}^n \times U} h(z, u) m(dz, du) \right| \mathbf{n}_r^k(dx) \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr = 0, \end{aligned}$$

similarly to [34, Theorem 1.1.2.], one finds a countable family of bounded continuous functions $(h^k)_{k \in \mathbb{N}^*}$ characterizing the weak convergence, therefore by Lebesgue's dominated convergence,

$$\begin{aligned} &\lim_{k \rightarrow \infty} \sum_{q \geq 0} \int_0^T \int_{(\mathcal{P}_U^n)^2} \frac{1}{2^q} \left| \int_{\mathbb{R}^n \times U} h^q(x, u) H^k(x, m)(du) \mathbf{n}_r^k(dx) \right. \\ &\quad \left. - \int_{\mathbb{R}^n \times U} h^q(z, u) m(dz, du) \right| \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr = 0, \end{aligned}$$

then $\lim_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \Delta \left(H^k(z, m)(du) \mathbf{n}_r^k(dz), m(du, dz) \right) \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr = 0$, where Δ is the metric characterizing the weak convergence on \mathcal{P}_U^n . As $[\hat{b}, \hat{\sigma}]$ are bounded and $\nu \in \mathcal{P}_{p'}(\mathbb{R}^n)$, for $(r, m) \in [0, T] \times \mathcal{P}_U^n$,

$$\lim_{K \rightarrow \infty} \sup_{k \in \mathbb{N}^*} \int_{|z| + \rho(u_0, u) \geq K} |z|^p + \rho(u_0, u)^p H^k(z, m)(du) \mathbf{n}_r^k(dz) = 0.$$

This is enough to conclude that

$$\lim_{k \rightarrow \infty} \int_0^T \int_{(\mathcal{P}_U^n)^2} \mathcal{W}_p \left(H^k(z, m)(du) \mathbf{n}_r^k(dz), m(du, dz) \right) \hat{\mathbf{q}}_r(dm, d\bar{\nu}) dr = 0. \quad \square$$

Consequence of the regularization: a continuity property Now, we want to provide some properties satisfying by a regularized map. Let $\psi : [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U \rightarrow \mathbb{R}^j$ be a Borel function, with $j \in \mathbb{N}^*$. For each $\varepsilon > 0$, one defines the function $\psi^\varepsilon : \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times \mathcal{P}((\mathcal{P}_U^n)^2) \times [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^j$ as follows: for every $(t, x, \mathbf{b}, \pi, \beta, q) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times \mathcal{P}((\mathcal{P}_U^n)^2)$

$$\begin{aligned} &\psi^\varepsilon[\mathbf{b}, \pi, \beta, q](t, x) \\ &:= \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n} \int_U \psi(t, y, \mathbf{b}_{t \wedge \cdot}, \pi_{t \wedge \cdot}, \beta_{t \wedge \cdot}, m, \bar{\nu}, u) \frac{G_\varepsilon(x-y)}{(m(dz, U))^{(\varepsilon)}(x)} m(du, dy) q(dm, d\bar{\nu}), \end{aligned}$$

where for every $m \in \mathcal{P}_U^n$, $(m(dz, U))^{(\varepsilon)}(x) := \int_{\mathbb{R}^n} G_\varepsilon(x-y) m(dy, U)$.

Notice that

$$|\psi^\varepsilon[\mathbf{b}, \pi, \beta, q](t, x)| \leq \sup_{z', \mathbf{b}', \zeta', m', \nu', u'} |\psi(t, z', \mathbf{b}', \zeta', m', \bar{\nu}', u')|,$$

for all $(\mathbf{b}, \pi, \beta, q, t, x)$. Then if ψ is bounded, ψ^ε is bounded uniformly in $\varepsilon > 0$. Also, given $(t, \mathbf{b}, \pi, \beta, q)$, for each $\varepsilon > 0$, the function $\mathbb{R}^n \ni x \rightarrow \psi^\varepsilon[\mathbf{b}, \pi, \beta, q](t, x) \in \mathbb{R}^j$ belongs to $C_b^\infty(\mathbb{R}^n)$, hence the name of *regularization*.

Under additional conditions, we have shown in the previous Theorem A.2, in some sense, “ $\lim_{\varepsilon \rightarrow 0} \psi^\varepsilon = \psi$ ” (see Proposition A.2 for more details). The next result checks that given $\varepsilon > 0$, the map ψ^ε satisfies a general continuity property.

Proposition A.4. *For any $\psi : [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U \rightarrow \mathbb{R}$ and $\phi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ two bounded continuous functions. For each $\varepsilon > 0$, the function*

$$(\mathbf{b}, \vartheta, \pi, \beta, q) \in \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^3 \times \mathbb{M}((\mathcal{P}_U^n)^2) \longrightarrow \int_0^T \int_{\mathbb{R}^n} \psi^\varepsilon[\mathbf{b}, \pi, \beta, q_t](t, x) \phi(t, x) \vartheta_t(dx) dt \in \mathbb{R}$$

is continuous.

Proof. Let $(\mathbf{b}^k, \vartheta^k, \pi^k, \beta^k, q^k)_{k \in \mathbb{N}} \subset \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^3 \times \mathbb{M}((\mathcal{P}_U^n)^2)$ and $(\mathbf{b}, \vartheta, \pi, \beta, q) \in \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^3 \times \mathbb{M}((\mathcal{P}_U^n)^2)$ verifying $\lim_k (\mathbf{b}^k, \vartheta^k, \pi^k, \beta^k, q^k) = (\mathbf{b}, \vartheta, \pi, \beta, q)$. Notice that,

$$\begin{aligned} & \int_0^T \int_{\mathbb{R}^n} \psi^\varepsilon[\mathbf{b}, \pi, \beta, q_t](t, x) \phi(t, x) \vartheta_t(dx) dt \\ &= \int_0^T \int_{\mathbb{R}^n} \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \\ & \quad \psi(t, y, \mathbf{b}_{t\wedge\cdot}, \pi_{t\wedge\cdot}, \beta_{t\wedge\cdot}, m, \bar{v}, u) \frac{G_\varepsilon(x - y)}{(m(dz, U))^{(\varepsilon)}(x)} m(du, dy) q_t(dm, d\bar{v}) \phi(t, x) \vartheta_t(dx) dt \\ &= \int_0^T \int_{\mathbb{R}^n} \int_{\mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2} \int_{(\mathcal{P}_U^n)^2} \int_{\mathbb{R}^n \times U} \\ & \quad \psi(t, y, g, e, e', m, \bar{v}, u) \phi(t, x) H^\varepsilon(x, m)(du, dy) q_t(dm, d\bar{v}) \vartheta_t(dx) \Psi_t(dg, de, de') dt, \end{aligned}$$

where

$$H^\varepsilon(x, m)(du, dy) := \frac{G_\varepsilon(x - y)}{(m(dz, U))^{(\varepsilon)}(x)} m(du, dy)$$

and

$$\Psi_t(dg, de, de') dt := \delta_{(\mathbf{b}_{t\wedge\cdot}, \pi_{t\wedge\cdot}, \beta_{t\wedge\cdot})} (dg, de, de') dt.$$

Next, we define

$$\begin{aligned} & Z^k(du, dy, dm, d\bar{v}, dg, de, de', dx, dt) \\ &:= \frac{1}{T} H^\varepsilon(x, m)(du, dy) q_t^k(dm, d\bar{v}) \vartheta_t^k(dx) \delta_{(\mathbf{b}_{t\wedge\cdot}^k, \pi_{t\wedge\cdot}^k, \beta_{t\wedge\cdot}^k)} (dg, de, de') dt \end{aligned}$$

and

$$\begin{aligned} & Z(du, dy, dm, d\bar{v}, dg, de, de', dx, dt) \\ &:= \frac{1}{T} H^\varepsilon(x, m)(du, dy) q_t(dm, d\bar{v}) \vartheta_t(dx) \Psi_t(dg, de, de') dt. \end{aligned}$$

Then $(Z^k)_{k \in \mathbb{N}}$ is a sequence of probability measures belonging to $\mathcal{P}(U \times \mathbb{R}^n \times (\mathcal{P}_U^n)^2 \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times \mathbb{R}^n \times [0, T])$. As $\lim_k (\mathbf{b}^k, \vartheta^k, \pi^k, \beta^k, q^k) = (\mathbf{b}, \vartheta, \pi, \beta, q)$, it is straightforward to see that $(Z^k)_{k \in \mathbb{N}}$ is relatively compact in $\mathcal{P}(U \times \mathbb{R}^n \times (\mathcal{P}_U^n)^2 \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times \mathbb{R}^n \times [0, T])$ and each sub-sequence converges to Z , therefore $(Z^k)_{k \in \mathbb{N}}$ converges to Z in a weak sense. As the function $(t, y, \mathbf{b}, e, e', m, \bar{v}, u, x) \in [0, T] \times \mathbb{R}^n \times \mathcal{C}^\ell \times (\mathcal{C}_{\mathcal{W}}^n)^2 \times (\mathcal{P}_U^n)^2 \times U \times \mathbb{R}^n \rightarrow \psi(t, y, \mathbf{b}_{t\wedge\cdot}, e, e', m, \bar{v}, u) \phi(t, x) \in \mathbb{R}^n$ is bounded continuous, we can conclude. \square

A.3 Some properties of Fokker-Planck equation

Let us recall a useful result on square root of matrices. Denote by \mathbb{S}_n^+ the set of symmetric positive definite matrices of dimension $n \in \mathbb{N}^*$. The principal square root function is denoted by: $f : Q \in \mathbb{S}_n^+ \mapsto f(Q) := Q^{1/2} \in \mathbb{S}_n^+$.

Proposition A.5. [19, Theorem 6.2] *There exists a constant $C(n)$ depending only of the dimension $n \in \mathbb{N}^*$ such that for any $(A, B) \in \mathbb{S}_n^+ \times \mathbb{S}_n^+$*

$$|f(A) - f(B)| \leq C(n) [\lambda_{\min}(A)^{1/2} + \lambda_{\min}(B)^{1/2}]^{-1} |A - B|,$$

where $\lambda_{\min}(\cdot)$ is the smallest eigenvalue.

Let E and E' be two Polish spaces and $[\bar{b}, \bar{a}] : [0, T] \times \mathbb{R}^n \times C([0, T]; E) \times \mathbb{M}(E') \rightarrow \mathbb{R}^n \times \mathbb{S}^{n \times n}$ be a bounded Borel functions s.t.: for all $(t, \pi, \hat{q}) \in [0, T] \times C([0, T]; E) \times \mathbb{M}(E')$, the function $x \in \mathbb{R}^n \rightarrow [\bar{b}, \bar{a}](t, x, \pi_{t\wedge\cdot}, \hat{q}_{t\wedge\cdot}) \in \mathbb{R}^n \times \mathbb{S}^{n \times n}$ belongs to $C_b^2(\mathbb{R}^n)$ and $\bar{a} \geq \rho I_n$, (A.12)

for a certain $\rho > 0$.

Also, let us introduce, for all $\varphi \in C^2(\mathbb{R}^n)$, $\bar{\mathcal{L}}_t \varphi[\pi, \hat{q}](x) := \frac{1}{2} \text{Tr}[\bar{a}(t, x, \pi, \hat{q}_{t\wedge\cdot}) \nabla^2 \varphi(x)] + \bar{b}(t, x, \pi, \hat{q}_{t\wedge\cdot})^\top \nabla \varphi(x)$.

Lemma A.6. *Let $\nu \in \mathcal{P}_p(\mathbb{R}^n)$. There exists a Borel function $Z : C([0, T]; E) \times \mathbb{M}(E') \rightarrow \mathcal{C}_{\mathbb{W}}^n$, s.t. if $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ is a filtered probability space supporting $(\mu_t)_{t \in [0, T]}$ a E -valued \mathbb{F} -adapted continuous process and $(\hat{\Lambda}_t)_{t \in [0, T]}$ a $\mathcal{P}(E')$ -valued \mathbb{F} -predictable process, then, the unique $\mathcal{P}(\mathbb{R}^n)$ -valued $(\sigma\{\mu_{t\wedge\cdot}, \hat{\Lambda}_{t\wedge\cdot}\})_{t \in [0, T]}$ -adapted continuous process $(\vartheta_t)_{t \in [0, T]}$ solution of: $\vartheta \in \mathcal{C}_{\mathbb{W}}^{n,p}$, and for all $(t, f) \in [0, T] \times C_b^2(\mathbb{R}^n)$,*

$$\langle f, \vartheta_t \rangle = \int_{\mathbb{R}^n} f(y) \nu(dy) + \int_0^t \int_{\mathbb{R}^n} \bar{\mathcal{L}}_r f[\mu, \hat{\Lambda}](x) \vartheta_r(dx) dr, \mathbb{P}\text{-a.e.} \quad (\text{A.13})$$

satisfies

$$\vartheta_t = Z_t(\mu_{t\wedge\cdot}, \hat{\Lambda}_{t\wedge\cdot}), \text{ for all } t \in [0, T], \mathbb{P}\text{-a.e.}$$

Proof. For the uniqueness of (A.13), as the coefficients $[\bar{b}, \bar{a}]$ verify (A.12), by a slight extension of (proof of) Theorem 5.2, one gets that equation (A.13) has at most one solution.

Let W be a \mathbb{R}^n -valued (\mathbb{P}, \mathbb{F}) Brownian motion and ξ be a \mathcal{F}_0 -random variable of law ν , in addition, (ξ, W) are \mathbb{P} -independent of $(\mu, \hat{\Lambda})$. Next, let us show the existence and find the function Z . Combining (A.12) and Theorem A.5, for any (t, π, \hat{q}) , the application $x \in \mathbb{R}^n \rightarrow (\bar{a}(t, x, \pi_{t\wedge\cdot}, \hat{q}_{t\wedge\cdot}))^{1/2} \in \mathbb{S}^{n \times n}$ is Lipschitz, with a Lipschitz constant depends only on \bar{a} . Therefore, there exists the \mathbb{R}^n -valued \mathbb{F} -adapted process X unique strong solution of

$$X_s = \xi + \int_0^s \bar{b}(r, X_r, \mu, \hat{\Lambda}) dr + \int_0^s (\bar{a}(r, X_r, \mu, \hat{\Lambda}))^{1/2} dW_r \text{ for all } s \in [0, T].$$

It is well known that $X_t = H_t(\xi, W_{t\wedge\cdot}, \mu_{t\wedge\cdot}, \hat{\Lambda}_{t\wedge\cdot})$, for all $t \in [0, T]$, \mathbb{P} -a.e. where $H : \mathbb{R}^n \times \mathcal{C}^n \times C([0, T]; E) \times \mathbb{M}(E') \rightarrow \mathcal{C}^n$ is a Borel function (independent of \mathbb{P}).

Denote by $\mathbb{G} := (\mathcal{G}_t)_{t \in [0, T]}$ the filtration defined by $\mathcal{G}_t := \sigma\{\mu_{t\wedge\cdot}, \hat{\Lambda}_{t\wedge\cdot}\}$, for all $t \in [0, T]$. As (ξ, W) are \mathbb{P} -independent of $(\mu, \hat{\Lambda})$, one has that: for all $t \in [0, T]$, $\mathcal{L}^{\mathbb{P}}(X_{t\wedge\cdot} | \mathcal{G}_t) = \mathcal{L}^{\mathbb{P}}(X_{t\wedge\cdot} | \mathcal{G}_T)$, \mathbb{P} -a.e. then by [10, Lemma A.1], the process $(\beta_t)_{t \in [0, T]}$ is a $\mathcal{P}(\mathbb{R}^n)$ -valued \mathbb{G} -adapted continuous process where $\beta : (t, \omega) \in [0, T] \times \Omega \rightarrow \mathcal{L}^{\mathbb{P}}(X_t | \mathcal{G}_t)(\omega) \in \mathcal{P}(\mathbb{R}^n)$, and by Itô's formula $(\beta_t)_{t \in [0, T]}$ is solution of equation (A.13). In addition, there exists a Borel function (independent of \mathbb{P}) $Z : C([0, T]; E) \times \mathbb{M}(E') \rightarrow \mathcal{C}_{\mathbb{W}}^n$ such that: \mathbb{P} -a.e., for all $t \in [0, T]$, $\beta_t = Z_t(\mu_{t\wedge\cdot}, \hat{\Lambda}_{t\wedge\cdot})$. \square