

## From the master equation to mean field game limit theory: A central limit theorem

François Delarue, Daniel Lacker, and Kavita Ramanan

### Abstract

Mean field games (MFGs) describe the limit, as  $n$  tends to infinity, of stochastic differential games with  $n$  players interacting with one another through their common empirical distribution. Under suitable smoothness assumptions that guarantee uniqueness of the MFG equilibrium, a form of law of large numbers (LLN), also known as propagation of chaos, has been established to show that the MFG equilibrium arises as the limit of the sequence of empirical measures of the  $n$ -player game Nash equilibria, including the case when player dynamics are driven by both idiosyncratic and common sources of noise. The proof of convergence relies on the so-called master equation for the value function of the MFG, a partial differential equation on the space of probability measures. In this work, under additional assumptions, we establish a functional central limit theorem (CLT) that characterizes the limiting fluctuations around the LLN limit as the unique solution of a linear stochastic PDE. The key idea is to use the solution to the master equation to construct an associated McKean-Vlasov interacting  $n$ -particle system that is sufficiently close to the Nash equilibrium dynamics of the  $n$ -player game for large  $n$ . We then derive the CLT for the latter from the CLT for the former. Along the way, we obtain a new multidimensional CLT for McKean-Vlasov systems. We also illustrate the broader applicability of our methodology by applying it to establish a CLT for a specific linear-quadratic example that does not satisfy our main assumptions, and we explicitly solve the resulting stochastic PDE in this case.

**Keywords:** Mean field games ; master equation ; McKean-Vlasov ; interacting particle systems ; common noise ; fluctuations ; central limit theorem ; systemic risk.

**AMS MSC 2010:** Primary 60F05; 60H10; 60H15; 91A13; 91A15, Secondary 91G80; 60K35.

Submitted to EJP on April 27, 2018, final version accepted on March 18, 2019.

## 1 Introduction

*Finite games and mean field games.* Mean field games (MFG), introduced independently in [35, 33, 34] and [24, 37], are models of competition among a continuum of symmetric agents, each of whom dynamically controls a state variable; see the books [8, 9] for an overview. Equilibria of MFGs were introduced as potentially more tractable approximations of Nash equilibria of large finite systems of agents interacting with one another through their common empirical distribution. While much of the theoretical work of the past decade has focused on questions of existence and uniqueness of MFG equilibria, the probabilistic limit theory is less well understood. A law of large numbers has only recently come into focus [6, 31, 30, 19], clarifying how a MFG arises as the limit

of a suitable sequence of  $n$ -player games as  $n \rightarrow \infty$ , in the presence of both idiosyncratic and common sources of noise. In particular, given for each  $n$  a Nash equilibrium for the  $n$ -player game, the limit of the sequence of empirical distributions of state variables can be effectively characterized in terms of the equilibria of the MFG. The goal of this paper is to complement this law of large numbers with a central limit theorem (CLT). Using similar techniques, a large deviation principle and non-asymptotic concentration bounds are obtained in a companion paper [18].

The main tool in our analysis is the *master equation*, an infinite-dimensional partial differential equation (PDE), which plays a similar role to the Hamilton-Jacobi-Bellman (HJB) equation in classical stochastic control theory. Lions [36] demonstrated how the master equation can be used to construct an equilibrium for the MFG, and this construction was developed further in [4, 5, 7]. To solve the master equation is no simple matter, though some preliminary well-posedness results may be found in [21, 14] in the case without common noise and [6, 9] in the case with common noise.

A major breakthrough came in [6] with the discovery that, for a class of MFG with a unique equilibrium, when the associated master equation has a smooth enough solution, it can be used to prove convergence of  $n$ -player Nash equilibria to the unique MFG equilibrium. The essential idea is that the solution to the master equation can be used to build a system of  $n$  interacting diffusions of McKean-Vlasov type (see [39, 44] for standard references on McKean-Vlasov equations) that is quantitatively “close” to the true  $n$ -player Nash equilibrium system and whose empirical measure converges to the MFG equilibrium. This immediately implies that the empirical measure of the  $n$ -player Nash equilibrium system converges to the MFG equilibrium. In this paper we refine the “closeness” estimates to show that the distance between the McKean-Vlasov and Nash systems decays rapidly enough that the fluctuations are the same.

The results of this paper and the companion [18] mark the first probabilistic limit theorems for MFGs beyond the law of large numbers, at least for diffusion-based models. The very recent simultaneous works [12, 13, 2] carry out a similar program for MFGs with finite state space (and without common noise), using the (finite-dimensional) master equation to connect the  $n$ -player equilibrium to a more classical interacting particle system, and then transferring limit theorems (a law of large numbers, CLT, and LDP) from the latter to the former. Loosely related ideas appeared also in [1], which uses FBSDE methods to study the convergence of MFGs as the common noise parameter vanishes.

*Main results.* Let us explain the idea and main results more clearly. We consider  $n$ -player stochastic differential games, in which agent  $i$  chooses a Markovian control  $\alpha^i = \alpha^i(t, \mathbf{X}_t)$  lying in an action space  $A$  to influence an  $(\mathbb{R}^d)^n$ -valued state process  $\mathbf{X} = (X^1, \dots, X^n)$  given by

$$dX_t^i = b(X_t^i, m_{\mathbf{X}_t}^n, \alpha^i(t, \mathbf{X}_t))dt + \sigma dB_t^i + \sigma_0 dW_t, \quad (1.1)$$

for suitable constant matrices  $\sigma \in \mathbb{R}^{d \times d}$  and  $\sigma_0 \in \mathbb{R}^{d \times d_0}$ , with  $\sigma$  being non-degenerate (see Subsection 2.5 for the complete list of assumptions). Throughout,  $X_0^1, \dots, X_0^n$  are independent and identically distributed initial conditions,  $B^1, \dots, B^n$ , are  $d$ -dimensional Wiener processes,  $W$  is a  $d_0$ -dimensional Wiener process, and  $B^1, \dots, B^n, W, X_0^1, \dots, X_0^n$  are independent, and we write

$$m_{\mathbf{x}}^n = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$$

to denote the empirical measure of a vector  $\mathbf{x} = (x_1, \dots, x_n)$  in  $(\mathbb{R}^d)^n$ . Agent  $i$  seeks to

minimize the cost functional

$$\mathbb{E} \left[ \int_0^T f(X_t^i, m_{\mathbf{X}_t}^n, \alpha^i(t, \mathbf{X}_t)) dt + g(X_T^i, m_{\mathbf{X}_T}^n) \right],$$

where the time horizon  $T \in (0, \infty)$  is fixed. Define the Hamiltonian

$$H(x, m, y) = \inf_{a \in A} [b(x, m, a) \cdot y + f(x, m, a)],$$

and let  $\hat{\alpha}(x, m, y)$  denote a minimizer, which we will always assume to exist. At this stage, we do not require  $\hat{\alpha}(x, m, y)$  to be unique, but it is unique in all the known examples that fit the assumptions we use below, see Subsection 2.3 for a more complete account. Nash equilibria of this game (in closed-loop strategies), defined precisely in Section 2.3, can be studied in terms of solutions  $(v^{n,i})_{i=1}^n$  of the PDE system

$$\begin{aligned} & \partial_t v^{n,i}(t, \mathbf{x}) \\ & + H(x_i, m_{\mathbf{x}}^n, D_{x_i} v^{n,i}(t, \mathbf{x})) + \sum_{j \neq i} D_{x_j} v^{n,i}(t, \mathbf{x}) \cdot b(x_j, m_{\mathbf{x}}^n, \hat{\alpha}(x_j, m_{\mathbf{x}}^n, D_{x_j} v^{n,j}(t, \mathbf{x}))) \quad (1.2) \\ & + \frac{1}{2} \sum_{j=1}^n \text{Tr} [D_{x_j, x_j}^2 v^{n,i}(t, \mathbf{x}) \sigma \sigma^\top] + \frac{1}{2} \sum_{j,k=1}^n \text{Tr} [D_{x_j, x_k}^2 v^{n,i}(t, \mathbf{x}) \sigma_0 \sigma_0^\top] = 0, \end{aligned}$$

for  $(t, \mathbf{x}) \in [0, T] \times (\mathbb{R}^d)^n$ , with terminal condition  $v^{n,i}(T, \mathbf{x}) = g(x_i, m_{\mathbf{x}}^n)$ . The quantity  $v^{n,i}$  describes the value function of the  $i$ th player in the  $n$ -player game. We assume this PDE system has a smooth solution, in which case a (closed-loop) Nash equilibrium is given by

$$\alpha^{n,i}(t, \mathbf{X}_t) = \hat{\alpha}(X_t^i, m_{\mathbf{X}_t}^n, D_{x_i} v^{n,i}(t, \mathbf{X}_t)). \quad (1.3)$$

As for uniqueness of the solution to the Nash system, this is the same as for uniqueness of  $\hat{\alpha}(x, m, y)$ : We do not explicitly require  $(v^{n,i})_{i=1}^n$  to be unique but, in the main examples that are known to fit our assumptions, it is in fact unique, see Subsection 2.5 for details. Noticeably, uniqueness has interesting consequences. Similar to [6, Subsection 6.2], it forces the components of the Nash equilibrium given by (1.3) to be exchangeable. Similar to [9, Proposition 6.27], it also guarantees, at least in some cases where  $(D_{x_i} v^{n,i})_{i=1}^n$  is bounded, that (1.3) is the unique bounded Nash (Markovian) equilibrium.

Substituting (1.3) into the dynamics (1.1), we identify the Nash equilibrium empirical measure  $(m_{\mathbf{X}_t}^n)_{t \in [0, T]}$ . The value function of a typical player in the associated MFG is described by the so-called master equation, whose definition we postpone to Section 2.3. Assuming, among other things, the existence of a sufficiently smooth solution to the master equation, we show the following:

1. As  $n \rightarrow \infty$ ,  $(m_{\mathbf{X}_t}^n)_{t \in [0, T]}$  converges in law to the unique solution  $(\mu_t)_{t \in [0, T]}$  to the MFG, which is itself a stochastic flow of probability measures. This was already shown in [6, Theorem 2.4.9] when the state space is the torus. Our extension to  $\mathbb{R}^d$  is straightforward provided the solution to the master equation together with its derivatives have an appropriate rate of growth and the system (1.2) has a solution. Similar questions are addressed in the reference [9, Chapter 13, Section 6.3].
2. Our main contribution is to show that the fluctuation process  $(\sqrt{n}(m_{\mathbf{X}_t}^n - \mu_t))_{t \in [0, T]}$  converges in law (in a suitable distribution space) to the unique solution to a certain linear Stochastic PDE (SPDE) driven by a space-time Gaussian noise, the coefficients of which depend on the solution to the master equation and on  $(\mu_t)_{t \in [0, T]}$ .

These results are presented precisely in Section 3.

*Strategy of proof.* The idea driving all of the results, both in this work and the companion paper [18], is to use the solution to the master equation,  $U = U(t, x, m)$ , for  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  and  $m$  a probability measure on  $\mathbb{R}^d$ , at least when the master equation has a smooth solution (in which case we may reasonably expect it to be unique, see Subsection 2.5) to construct an interacting diffusion system  $\bar{\mathbf{X}} = (\bar{X}^1, \dots, \bar{X}^n)$  of McKean-Vlasov type:

$$d\bar{X}_t^i = b\left(\bar{X}_t^i, m_{\bar{\mathbf{X}}_t}^n, \hat{\alpha}(\bar{X}_t^i, m_{\bar{\mathbf{X}}_t}^n, D_x U(t, \bar{X}_t^i, m_{\bar{\mathbf{X}}_t}^n))\right) dt + \sigma dB_t^i + \sigma_0 dW_t, \quad (1.4)$$

starting from  $\bar{X}_0^i = X_0^i$ , and driven by the same Wiener processes. By showing that the functions  $(u^{n,i}(t, x) := U(t, x_i, m_x^n))_{i=1}^n$  nearly solve the  $n$ -player PDE system written for  $(v^{n,i})_{i=1}^n$  above (see the crucial Proposition 4.1), we can find good estimates between  $D_{x_i} v^{n,i}$  and  $D_{x_i} u^{n,i}$  (see Theorem 4.2), which enable an estimate of the distance between the empirical measures  $m_{\bar{\mathbf{X}}}^n$  and  $m_{\mathbf{X}}^n$ . This is indeed the strategy of [6], where it is shown that

$$\mathbb{E} [\mathcal{W}_{1, \mathcal{C}^d}(m_{\bar{\mathbf{X}}}^n, m_{\mathbf{X}}^n)] = O(n^{-1}), \quad (1.5)$$

where  $\mathcal{W}_{p, \mathcal{C}^d}$  denotes the  $p$ -Wasserstein distance on the space of probability measures on the path space  $\mathcal{C}^d := C([0, T]; \mathbb{R}^d)$  with finite  $p^{\text{th}}$  moment. Equivalently, (1.4) must be regarded as an approximating  $n$ -particle system of the  $n$ -player Nash equilibrium. From McKean-Vlasov theory (see, e.g., [22, 41, 44] for the case without common noise, or [11, 15, 17, 28] and [9, Chapter 2, Section 2.1] for the common noise case), we conclude that both empirical measure sequences converge in law to the unique solution  $(\mu_t)_{t \in [0, T]}$  of the (conditional) McKean-Vlasov SDE

$$\begin{aligned} dX_t &= b\left(X_t, \mu_t, \hat{\alpha}(X_t, \mu_t, D_x U(t, X_t, \mu_t))\right) dt + \sigma dB_t + \sigma_0 dW_t, \\ \mu_t &= \text{Law}(X_t \mid (W_s)_{s \leq t}), \end{aligned}$$

where  $(X_0, B)$  is a copy of  $(X_0^1, B^1)$ , independent of  $W$ . Moreover, it is known from early work on the master equation that this limit  $(\mu_t)_{t \in [0, T]}$  is the unique equilibrium of the MFG, which is to say that:

- (i)  $(X_t)_{t \in [0, T]}$  solves the following optimization problem in the environment  $(\mu_t)_{t \in [0, T]}$ :

$$\begin{cases} \inf_{\alpha} \mathbb{E} \left[ \int_0^T f(X_t^\alpha, \mu_t, \alpha_t) dt + g(X_T^\alpha, \mu_T) \right], \\ \text{s.t. } dX_t^\alpha = b(X_t^\alpha, \mu_t, \alpha_t) dt + \sigma dB_t + \sigma_0 dW_t, \end{cases}$$

the infimum being taken over  $A$ -valued processes  $\alpha = (\alpha_t)_{t \in [0, T]}$  that are progressively measurable with respect to the filtration generated by  $X_0$ ,  $B$  and  $W$ , and for which the SDE for  $X^\alpha$  is strongly well-posed<sup>1</sup>.

- (ii) any other (pathwise continuous) environment  $\nu = (\nu_t)_{t \in [0, T]}$  that takes values in the space of probability measures on  $\mathbb{R}^d$ , that is adapted to the filtration generated by  $W$  and under which the above control problem (with  $\mu_t$  replaced by  $\nu_t$ ) has an optimizer with  $\nu$  as flow of conditional marginal laws with respect to  $W$  coincides with  $(\mu_t)_{t \in [0, T]}$ .

<sup>1</sup>To simplify, we also use strong solutions when dealing with stochastic optimal controls over open loop controls.

This leads to the convergence claimed in point (1) of the program detailed above.

Next, to prove the CLT described in point (2) above, we study first the McKean-Vlasov fluctuations  $(\sqrt{n}(m_{\mathbf{X}_t}^n - \mu_t))_{t \in [0, T]}$ . The early results of [46, 45, 40] only cover affine dependence on the mean field term, but requiring the interaction in (1.4) to be affine is far too restrictive, except perhaps in linear-quadratic models. On the other hand, the later work [29] treats more general mean field interactions to include the non-affine yet smooth interactions that we obtain from the solution to the master equation. However, the state space in [29] is only one-dimensional, and so we extend their result to our multi-dimensional setting; further, in order to track the influence of the dimension explicitly, we follow some of the arguments introduced in [40]. While  $O(n^{-1/2})$  would not suffice, the estimate (1.5) is sharp enough to show that the fluctuations of the McKean-Vlasov and Nash systems are then the same.

Under additional assumptions, the estimate (1.5) is further refined in the companion paper [18] to establish exponential closeness of the McKean-Vlasov and Nash systems, which is then used to obtain large deviations results and non-asymptotic concentration bounds.

*Required assumptions and examples.* The above results are proved under admittedly very strong hypotheses. The main conditions, Assumptions **A**, **B**, and **B'**, are spelled out in Section 2.5 below, and the CLT requires some additional hypotheses (see Assumption **C**). Most notably, we assume throughout that  $U$  admits bounded derivatives up to order two in the measure argument. For the CLT, additional smoothness is required of the first and second order derivatives with respect to  $m$ , the precise form of which depends on the dimension  $d$  of the state space. In addition, we require either that the Hamiltonian  $H$  is Lipschitz in  $y$  or that  $\{v^{n,i} : n \in \mathbb{N}, i = 1, \dots, n\}$  are uniformly bounded. Although this rules out many natural examples, such as linear-quadratic models (see, however, the discussion of the next paragraph), the main well-posedness results of [6] ensure that our assumptions cover a broad class of coefficients. More importantly, we avoid as much as possible imposing specific assumptions on the data  $(b, f, g)$ ; for instance, in contrast with [6], we do not impose any explicit monotonicity condition on  $f$  and  $g$ . Instead, we aim as much as possible to create a “black box” in the sense that our results should read as follows: If there exists a sufficiently regular solution to the master equation, then the results (1) and (2) described above are valid. We hope that our work will apply not only to the setting of [6] but also to those illuminated by future research on the well-posedness of the master equation.

Despite the restrictive assumptions required for our general results, the same ideas can be successfully adapted to specific examples which fall outside the scope of our main theorems. In Section 6 we discuss the linear-quadratic model of [10], which does not fit the assumptions of our main theorems but for which an explicit solution is known for both the  $n$ -player and MFG equilibria. These explicit solutions allow us to bypass the difficult estimates required in the general setting, and we again connect the  $n$ -player equilibrium and a corresponding McKean-Vlasov system. As before, this connection lets us transfer the CLT for the latter system to the former. In other words, while a single unifying theorem seems out of reach for technical reasons, the strategy of our arguments seems much more broadly applicable. In addition, we compute the explicit solution for the limiting SPDE in this specific model.

*Organization of the paper.* The paper is organized as follows. In Section 2 we carefully define the  $n$ -player game and the Nash system, the MFG and the master equation, and the main sets of assumptions. Section 3 then states precisely the main results, with the LLN in Section 3.1 and the CLT in Section 3.2. The key estimates relating the Nash system and master equation are developed in Section 4, to prepare for the proofs of the LLN in Section 4.4 and the CLT in Section 5. Section 6 treats the specific model

mentioned in the previous paragraph. Relevant properties of derivatives of functionals on the Wasserstein space of probability measures are collected in Appendix A.

## 2 Nash systems and Master equations

### 2.1 Notation and model inputs

For a normed space  $(E, \|\cdot\|)$ , let  $\mathcal{P}(E)$  denote the set of Borel probability measures on  $E$ . Throughout the paper we make use of the standard notation  $\langle \mu, \varphi \rangle := \int_E \varphi d\mu$  for integrable functions  $\varphi$  on  $E$  and measures  $\mu$  on  $E$ . Given  $p \in [1, \infty)$ , we write  $\mathcal{P}^p(E, \|\cdot\|)$ , or simply  $\mathcal{P}^p(E)$  if the norm is understood, for the set of  $\mu \in \mathcal{P}(E)$  satisfying  $\langle \mu, \|\cdot\|^p \rangle < \infty$ . For a separable Banach space  $(E, \|\cdot\|)$ , we always endow  $\mathcal{P}^p(E, \|\cdot\|)$  with the  $p$ -Wasserstein metric  $\mathcal{W}_{p,E}$  defined by

$$\mathcal{W}_{p,E}(\mu, \nu) := \inf_{\pi} \left( \int_{E \times E} \|x - y\|^p \pi(dx, dy) \right)^{1/p}, \quad (2.1)$$

where the infimum is over all probability measures  $\pi$  on  $E \times E$  with marginals  $\mu$  and  $\nu$ . When the space  $E$  is understood, we may omit it from the subscript in  $\mathcal{W}_{p,E}$  by writing simply  $\mathcal{W}_p$ .

For a positive integer  $k$ , we always equip  $\mathbb{R}^k$  with the Euclidean norm, denoted  $|\cdot|$ , unless stated otherwise. For fixed  $T \in (0, \infty)$ , we will make use of the path spaces

$$\mathcal{C}^k := C([0, T]; \mathbb{R}^k),$$

which are always endowed with the supremum norm  $\|x\|_{\infty} = \sup_{t \in [0, T]} |x_t|$ . For  $m \in \mathcal{P}(\mathcal{C}^k)$  and  $t \in [0, T]$ , we write  $m_t$  for the time- $t$  marginal of  $m$ , i.e., the image of  $m$  under the map  $\mathcal{C}^k \ni x \mapsto x_t \in \mathbb{R}^k$ .

For a set  $E$  and  $n \in \mathbb{N}$ , we often use boldface  $\mathbf{x} = (x_1, \dots, x_n)$  for an element of  $E^n$ , and we write

$$m_{\mathbf{x}}^n := \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$$

for the associated empirical measure, which lies in  $\mathcal{P}(E)$ .

### 2.2 Derivatives on Wasserstein space

To define the master equation, we must first introduce a suitable derivative for functions of probability measures. As we will be applying but not solving the master equation, we need very few properties of these derivatives, all of which are outlined here. Following [6], we fix an exponent  $q \in [1, \infty)$ , and we say that a function  $V : \mathcal{P}^q(\mathbb{R}^d) \rightarrow \mathbb{R}$  is  $\mathcal{C}^1$  if there exists a continuous map  $\frac{\delta V}{\delta m} : \mathcal{P}^q(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$  satisfying

- (i) For every  $\mathcal{W}_q$ -compact set  $K \subset \mathcal{P}^q(\mathbb{R}^d)$ , there exists a constant  $c < \infty$  such that  $\sup_{m \in K} |\frac{\delta V}{\delta m}(m, v)| \leq c(1 + |v|^q)$  for all  $v \in \mathbb{R}^d$ .
- (ii) For every  $m, m' \in \mathcal{P}^q(\mathbb{R}^d)$ ,

$$V(m') - V(m) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta V}{\delta m}((1-t)m + tm', v) (m' - m)(dv) dt. \quad (2.2)$$

Note that the condition (i) is designed to make the integral in (ii) well-defined. Only one function  $\frac{\delta V}{\delta m}$  can satisfy (2.2), up to a constant shift; that is, if  $\frac{\delta V}{\delta m}$  satisfies (2.2) then so does  $\frac{\delta V}{\delta m} + c$  for any  $c \in \mathbb{R}$ . For concreteness we always choose the shift to ensure

$$\int_{\mathbb{R}^d} \frac{\delta V}{\delta m}(m, v) m(dv) = 0. \quad (2.3)$$

If  $\frac{\delta V}{\delta m}(m, v)$  is continuously differentiable in  $v$ , we define its *intrinsic derivative*  $D_m V : \mathcal{P}^q(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  by

$$D_m V(m, v) = D_v \left( \frac{\delta V}{\delta m}(m, v) \right), \quad (2.4)$$

where we use the notation  $D_v$  for the gradient in  $v$ . If, for each  $v \in \mathbb{R}^d$ , the map  $m \mapsto \frac{\delta V}{\delta m}(m, v)$  is  $\mathcal{C}^1$ , then we say that  $V$  is  $\mathcal{C}^2$  and let  $\frac{\delta^2 V}{\delta m^2}$  denote its derivative, or more explicitly,

$$\frac{\delta^2 V}{\delta m^2}(m, v, v') = \frac{\delta}{\delta m} \left( \frac{\delta V}{\delta m}(\cdot, v) \right)(m, v').$$

We will also make some use of the derivative

$$D_v D_m V(m, v) = D_v [D_m V(m, v)],$$

when it exists, and we note that  $D_v D_m V$  takes values in  $\mathbb{R}^{d \times d}$ ; for some results, we will also consider higher order derivatives  $D_v^k D_m V(m, v)$  with values in  $\mathbb{R}^{d \times \dots \times d} \cong \mathbb{R}^{d^{k+1}}$  for  $k \in \mathbb{N}$ . Finally, if  $V$  is  $\mathcal{C}^2$  and if  $\frac{\delta^2 V}{\delta m^2}(m, v, v')$  is twice continuously differentiable in  $(v, v')$ , we let

$$D_m^2 V(m, v, v') = D_{v, v'}^2 \frac{\delta^2 V}{\delta m^2}(m, v, v')$$

denote the  $d \times d$  matrix of partial derivatives  $(\partial_{v_i} \partial_{v'_j} [\delta^2 V / \delta m^2](m, v, v'))_{i, j}$ . Equivalently (see [6, Lemma 2.2.4]),

$$D_m^2 V(m, v, v') = D_m (D_m V(\cdot, v))(m, v').$$

Aside from these definitions, the only result we should state here is an important observation on how these derivatives interact with empirical measures, essentially known from [6]. For completeness, the proof is given in Appendix A.

**Proposition 2.1.** *Given  $V : \mathcal{P}^q(\mathbb{R}^d) \rightarrow \mathbb{R}$ , define  $u_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}$  by  $u_n(\mathbf{x}) = V(m_{\mathbf{x}}^n)$  for some fixed  $n \geq 1$ .*

- (i) *If  $V$  is  $\mathcal{C}^1$  and if  $D_m V$  exists and is bounded and jointly continuous, then  $u_n$  is continuously differentiable, and*

$$D_{x_j} u_n(\mathbf{x}) = \frac{1}{n} D_m V(m_{\mathbf{x}}^n, x_j), \text{ for } j = 1, \dots, n. \quad (2.5)$$

- (ii) *If  $V$  is  $\mathcal{C}^2$  and if  $D_m^2 V$  exists and is bounded and jointly continuous, then  $u$  is twice continuously differentiable, and*

$$D_{x_k} D_{x_j} u_n(\mathbf{x}) = \frac{1}{n^2} D_m^2 V(m_{\mathbf{x}}^n, x_j, x_k) + \delta_{j, k} \frac{1}{n} D_v D_m V(m_{\mathbf{x}}^n, x_j),$$

where  $\delta_{j, k} = 1$  if  $j = k$  and  $\delta_{j, k} = 0$  if  $j \neq k$ .

### 2.3 Nash systems and $n$ -player games

We fix throughout the paper a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ , supporting independent  $\mathbb{F}$ -Wiener processes  $W$  of dimension  $d_0$  and  $(B^i)_{i=1}^\infty$  of dimension  $d$  (we choose the dimension of the idiosyncratic noises equal to the dimension of the state space for convenience only), as well as a sequence of i.i.d.  $\mathcal{F}_0$ -measurable  $\mathbb{R}^d$ -valued initial states  $(X_0^i)_{i=1}^\infty$  with distribution  $\mu_0$ .

We describe the  $n$ -player game and PDE systems first, deferring a precise statement of assumptions to Section 2.5. We are given an exponent  $p^*$ , an action space  $A$ , assumed to be a Polish space, and Borel measurable functions

$$\begin{aligned}(b, f) : \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \times A &\rightarrow \mathbb{R}^d \times \mathbb{R}, \\ g : \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) &\rightarrow \mathbb{R},\end{aligned}$$

along with two (constant) matrices  $\sigma \in \mathbb{R}^{d \times d}$  and  $\sigma_0 \in \mathbb{R}^{d \times d_0}$ , where  $\sigma$  is non-degenerate.

In the  $n$ -player game, players  $i = 1, \dots, n$  control the  $(\mathbb{R}^d)^n$ -valued state process  $(\mathbf{X}_t = (X_t^1, \dots, X_t^n))_{t \in [0, T]}$ , given by

$$dX_t^i = b(X_t^i, m_{\mathbf{X}_t}^n, \alpha^i(t, \mathbf{X}_t))dt + \sigma dB_t^i + \sigma_0 dW_t, \quad (2.6)$$

where we recall that  $m_{\mathbf{X}_t}^n$  denotes the empirical measure of the vector  $\mathbf{X}_t$ . Here  $\alpha^i$  is the control chosen by player  $i$  in feedback form. The objective of player  $i$  is to try to choose  $\alpha^i$  to minimize

$$J^{n,i}(\alpha^1, \dots, \alpha^n) = \mathbb{E} \left[ \int_0^T f(X_t^i, m_{\mathbf{X}_t}^n, \alpha^i(t, \mathbf{X}_t))dt + g(X_T^i, m_{\mathbf{X}_T}^n) \right].$$

A (closed-loop) Nash equilibrium is defined in the usual way as a vector of feedback functions  $(\alpha^1, \dots, \alpha^n)$ , where  $\alpha^i : [0, T] \times (\mathbb{R}^d)^n \rightarrow A$  are such that the solution to the SDE (2.6) is unique in law, such that

$$J^{n,i}(\alpha^1, \dots, \alpha^n) \leq J^{n,i}(\alpha^1, \dots, \alpha^{i-1}, \tilde{\alpha}, \alpha^{i+1}, \dots, \alpha^n),$$

for any alternative choice of feedback control  $\tilde{\alpha}$  for which the SDE (2.6), with  $\alpha$  replaced by  $\tilde{\alpha}$ , has a solution that is unique in law. In this regard, it might be worth pointing out that (2.6) is in fact uniquely solvable in the strong sense whenever  $\alpha^i$  is at most of linear growth in  $x$ , uniformly in time. When  $\alpha^i$  is bounded, unique strong solvability is a consequence of [47]. The latter may be shown to remain true whenever  $\alpha^i$  is at most of linear growth in  $x$ , uniformly in time, by a standard localization argument.

From the work of [3], we know that a Nash equilibrium can be built using a system of HJB equations. Define the Hamiltonian  $H : \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$H(x, m, y) = \inf_{a \in A} [b(x, m, a) \cdot y + f(x, m, a)].$$

Assume that this infimum is attained for each  $(x, m, y)$ , and let  $\hat{\alpha}(x, m, y)$  denote a minimizer; we will place assumptions on the function  $\hat{\alpha}$  in the next section. While we do not explicitly require  $\hat{\alpha}$  to be unique, the reader must be aware of the fact that the set of assumptions we use below is rather constraining. For instance, Assumption **A** requires the existence of a smooth solution to the aforementioned master equation, see Subsection 2.4 for details. In all the existing works on the subject, existence of a smooth solution to the master equation is in fact proven under the assumption that  $\hat{\alpha}(x, m, y)$  is unique. The latter is clear in [21, 14, 9]. In [6], the Hamiltonian is required to be smooth and  $b(x, m, \alpha)$  to match  $\alpha$ : Following the proof of [43, Theorem 2], uniqueness of the minimizer follows.

It is now convenient to define the functionals  $\hat{b}$  and  $\hat{f}$  on  $\mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \times \mathbb{R}^d$  by

$$\hat{b}(x, m, y) = b(x, m, \hat{\alpha}(x, m, y)) \quad \text{and} \quad \hat{f}(x, m, y) = f(x, m, \hat{\alpha}(x, m, y)), \quad (2.7)$$

and note that then

$$H(x, m, y) = \hat{b}(x, m, y) \cdot y + \hat{f}(x, m, y). \quad (2.8)$$



As already mentioned in the Introduction, see (1.2), the  $n$ -player Nash system is a PDE system for  $n$  functions,  $(v^{n,i} : [0, T] \times (\mathbb{R}^d)^n \rightarrow \mathbb{R})_{i=1}^n$ , given by

$$\begin{aligned} \partial_t v^{n,i}(t, \mathbf{x}) + H(x_i, m_{\mathbf{x}}^n, D_{x_i} v^{n,i}(t, \mathbf{x})) + \sum_{j=1, j \neq i}^n D_{x_j} v^{n,i}(t, \mathbf{x}) \cdot \widehat{b}(x_j, m_{\mathbf{x}}^n, D_{x_j} v^{n,j}(t, \mathbf{x})) \\ + \frac{1}{2} \sum_{j=1}^n \text{Tr} [D_{x_j, x_j}^2 v^{n,i}(t, \mathbf{x}) \sigma \sigma^\top] + \frac{1}{2} \sum_{j,k=1}^n \text{Tr} [D_{x_j, x_k}^2 v^{n,i}(t, \mathbf{x}) \sigma_0 \sigma_0^\top] = 0, \end{aligned}$$

with terminal condition  $v^{n,i}(T, \mathbf{x}) = g(x_i, m_{\mathbf{x}}^n)$ .

Using (classical) solutions to the  $n$ -player Nash system, we may construct an equilibrium for the  $n$ -player game. The  $i^{\text{th}}$  agent uses the feedback control

$$[0, T] \times (\mathbb{R}^d)^n \ni (t, \mathbf{x}) \mapsto \widehat{\alpha}(\mathbf{x}, m_{\mathbf{x}}^n, D_{x_i} v^{n,i}(t, \mathbf{x})).$$

As a result, the equilibrium state process  $\mathbf{X} = (X^1, \dots, X^n)$  is governed by

$$dX_t^i = \widehat{b}(X_t^i, m_{\mathbf{X}_t}^n, D_{x_i} v^{n,i}(t, \mathbf{X}_t)) dt + \sigma dB_t^i + \sigma_0 dW_t. \quad (2.9)$$

Under Assumption **A** of Section 2.5 below, the SDE (2.9) is uniquely solvable. Indeed, due to assumption **A**(4), the second derivatives of  $v^{n,i}$  exist and are continuous, which ensures that  $D_{x_i} v^{n,i}$  is locally Lipschitz. In light of Assumption **A**(1) and the fact that  $\mathbf{x} \mapsto m_{\mathbf{x}}^n$  is a Lipschitz function from  $(\mathbb{R}^d)^n$  to  $(\mathcal{P}^{p^*}(\mathbb{R}^d), \mathcal{W}_{p^*})$ , this ensures that the SDE system (2.9) has a unique strong solution.

## 2.4 The mean field game and master equation

The master equation is a PDE for a function  $U : [0, T] \times \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \rightarrow \mathbb{R}$ , given by

$$\begin{aligned} 0 = & \partial_t U(t, x, m) + H(x, m, D_x U(t, x, m)) \\ & + \frac{1}{2} \text{Tr} [(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 U(t, x, m)] \\ & + \int_{\mathbb{R}^d} \widehat{b}(v, m, D_x U(t, v, m)) \cdot D_m U(t, x, m, v) dm(v) \\ & + \frac{1}{2} \int_{\mathbb{R}^d} \text{Tr} [(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_v D_m U(t, x, m, v)] dm(v) \\ & + \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \text{Tr} [\sigma_0 \sigma_0^\top D_m^2 U(t, x, m, v, v')] dm(v) dm(v') \\ & + \int_{\mathbb{R}^d} \text{Tr} [\sigma_0 \sigma_0^\top D_x D_m U(t, x, m, v)] dm(v), \end{aligned} \quad (2.10)$$

for  $(t, x, m) \in (0, T) \times \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d)$ , with terminal condition  $U(T, x, m) = g(x, m)$ . The connection between the Nash system and the master equation will be made clear in Proposition 4.1 below; roughly speaking,  $v^{n,i}(t, \mathbf{x})$  is expected to be close to  $U(t, x_i, m_{\mathbf{x}}^n)$  as  $n$  tends to infinity.

Just as the  $n$ -player Nash system was used to build an equilibrium for the  $n$ -player game, we will use the master equation to describe an equilibrium for the mean field game. First, consider the McKean-Vlasov equation

$$d\mathcal{X}_t = \widehat{b}(\mathcal{X}_t, \mu_t, D_x U(t, \mathcal{X}_t, \mu_t)) dt + \sigma dB_t^1 + \sigma_0 dW_t, \quad \mathcal{X}_0 = X_0^1, \quad \mu = \mathcal{L}(\mathcal{X}|W), \quad (2.11)$$

where  $\mathcal{L}(\mathcal{X}|W)$  denotes the conditional law of  $\mathcal{X}$  given (the path)  $W$ , viewed as a random element of  $\mathcal{P}^{p^*}(\mathbb{C}^d)$ . Here, a solution  $\mathcal{X} = (\mathcal{X}_t)_{t \in [0, T]}$  is required to be adapted to the filtration generated by the process  $(X_0^1, W_t, B_t^1)_{t \in [0, T]}$ . Notice that necessarily  $\mu_t =$

$\mathcal{L}(\mathcal{X}_t|W) = \mathcal{L}(\mathcal{X}_t|(W_s)_{s \in [0,t]})$  a.s., for each  $t \in [0, T]$ , because  $(W_s - W_t)_{s \geq t}$  is independent of  $(\mathcal{X}_s, W_s)_{s \leq t}$ . Assumptions **A(1)** and **A(5)**, stated in Section 2.5 below, ensure that there is a unique strong solution to (2.11); this follows from a straightforward adaptation of the arguments of Sznitman [44, Chapter 1] (cf. [11, Section 7] and [9, Chapter 2, Section 2.1]). For the reader who is more familiar with the PDE formulation of mean field games, we emphasize that the process  $(\mu_t)_{t \in [0, T]}$  is a weak solution to the stochastic Fokker-Planck equation

$$d\mu_t = -\operatorname{div}(\widehat{b}(\cdot, \mu_t, D_x U(t, \cdot, \mu_t))\mu_t)dt + \frac{1}{2}\operatorname{Tr}[D_x^2 \mu_t(\sigma\sigma^\top + \sigma_0\sigma_0^\top)]dt - (\sigma_0^\top D_x \mu_t) \cdot dW_t,$$

for  $t \in [0, T]$ , which follows from a straightforward application of Itô's formula to the process  $(\phi(X_t))_{t \in [0, T]}$  for smooth test functions  $\phi$ .

Since  $U$  is a classical solution to the master equation with bounded derivatives (see Assumptions **A(1)** and **A(5)** in Section 2.5 below), it is known that the measure flow  $\mu$  constructed from the McKean-Vlasov equation (2.11) is the unique equilibrium of the mean field game; see for instance [8, Proposition 5.106]. A mean field game equilibrium is usually defined as a fixed point of the map  $\Phi$  which sends a  $W$ -measurable random measure  $\mu$  on  $\mathcal{C}^d$  (such that  $(\mu_t)_{t \in [0, T]}$  is adapted to the filtration generated by  $W$ ) to a new random measure  $\Phi(\mu)$ , defined as follows:

- (i) Solve the stochastic optimal control problem, with  $\mu$  fixed:

$$\begin{cases} \inf_{\alpha} \mathbb{E} \left[ \int_0^T f(X_t^\alpha, \mu_t, \alpha_t) dt + g(X_T^\alpha, \mu_T) \right], \\ \text{s.t. } dX_t^\alpha = b(X_t^\alpha, \mu_t, \alpha_t) dt + \sigma dB_t^1 + \sigma_0 dW_t, \end{cases}$$

where  $\alpha = (\alpha_t)_{t \in [0, T]}$  is an  $A$ -valued progressively measurable process (with respect to the filtration generated by  $X_0$ ,  $B$  and  $W$ ) such that the SDE admits a unique strong solution and the cost functional makes sense (regarding our choice to use here strong solutions, see the previous footnote 1).

- (ii) Letting  $X^*$  denote the optimally controlled state process, set  $\Phi(\mu) = \mathcal{L}(X^*|W)$ .

Note that if the optimization problem in step (i) has multiple solutions, the map  $\Phi$  may be set-valued, and we seek  $\mu$  such that  $\mu \in \Phi(\mu)$ . The original formulation of Lasry and Lions [35] is a forward-backward PDE system, which is essentially equivalent to this fixed point procedure, when  $\sigma_0 = 0$ . When  $\sigma_0 \neq 0$ , the forward-backward PDE becomes stochastic, but the same connection remains. For more details on the connection between the master equation and more common PDE or probabilistic formulations of mean field games, see [4, 5, 7] or [6, Section 1.2.4]. For our purposes, we simply take the McKean-Vlasov equation (2.11) as the definition of  $\mu$ .

## 2.5 Assumptions

The following standing assumption holds throughout the paper. In fact, more will be needed in Section 3 to derive the central limit theorem, but Assumptions **A**, **B** and **B'** below form the foundation behind the key estimates used this paper and as well as the companion one [18].

### Assumption A.

1. The exponent  $p^*$  lies in  $[1, 2]$ . A minimizer  $\widehat{\alpha}(x, m, y) \in \arg \min_{a \in A} [b(x, m, a) \cdot y + f(x, m, a)]$  exists for every  $(x, m, y) \in \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \times \mathbb{R}^d$ , such that the function  $\widehat{b}(x, m, y)$  defined in (2.7) is Lipschitz in all variables. That is, there exists  $C < \infty$  such that, for all  $x, x', y, y' \in \mathbb{R}^d$  and  $m, m' \in \mathcal{P}^{p^*}(\mathbb{R}^d)$ ,

$$|\widehat{b}(x, m, y) - \widehat{b}(x', m', y')| \leq C(|x - x'| + \mathcal{W}_{p^*}(m, m') + |y - y'|),$$

where we recall that  $\mathcal{W}_{p^*}$  is shorthand for  $\mathcal{W}_{p^*,(\mathbb{R}^d,|\cdot|)}$ .

2. The  $d \times d$  matrix  $\sigma$  is non-degenerate.
3. The initial states  $(X_0^i)_{i=1}^\infty$  are i.i.d. with law  $\mu_0 \in \mathcal{P}^{p'}(\mathbb{R}^d)$  for some  $p' > 4$ .
4. For each  $n$ , the  $n$ -player Nash system (2.9) has a classical solution  $(v^{n,i})_{i=1}^n$ , in the sense that each function  $v^{n,i}(t, \mathbf{x})$  is continuously differentiable in  $t$  and twice continuously differentiable in  $\mathbf{x}$ . Moreover,  $D_{x_j} v^{n,i}$  has linear growth and  $v^{n,i}$  has quadratic growth, for each  $n, i, j$ . That is, there exist  $L_{n,i} > 0$  and  $L_{n,i,j} > 0$  such that, for all  $t \in [0, T]$  and  $\mathbf{x} \in (\mathbb{R}^d)^n$ ,

$$\begin{aligned} |D_{x_j} v^{n,i}(t, \mathbf{x})| &\leq L_{n,i,j} (1 + |\mathbf{x}|), \\ |v^{n,i}(t, \mathbf{x})| &\leq L_{n,i} (1 + |\mathbf{x}|^2), \end{aligned}$$

where  $|\mathbf{x}|$  in **A(4)** is the Euclidean norm of  $\mathbf{x} \in (\mathbb{R}^d)^n \cong \mathbb{R}^{dn}$ .

5. The master equation admits a classical solution  $U : [0, T] \times \mathbb{R}^d \times \mathcal{P}^{p^*}(\mathbb{R}^d) \ni (t, x, m) \mapsto U(t, x, m)$ . The derivative  $D_x U(t, x, m)$  exists and is Lipschitz in  $(x, m)$ , uniformly in  $t$  (using  $\mathcal{W}_{p^*}$  for the argument  $m \in \mathcal{P}^{p^*}(\mathbb{R}^d)$ ), and  $U$  admits continuous derivatives  $\partial_t U$ ,  $D_x U$ ,  $D_m U$ ,  $D_x^2 U$ ,  $D_v D_m U$ ,  $D_x D_m U$ , and  $D_m^2 U$ . Moreover,  $D_x U$ ,  $D_m U$ ,  $D_x D_m U$ , and  $D_m^2 U$  are assumed to be bounded.

The lower bound  $p' > 4$  in assumption **A(3)** is exactly the one we need to formulate our first main result, see Theorem 3.1, but a greater lower bound is needed in the statement of the second main result, see Theorem 3.2. Assumptions **A(4-5)** are heavy, owing to the difficulty of solving the master equation and  $n$ -player systems. As for **A(4)**, earlier results on the solvability of systems of quasilinear parabolic equations may be found in [32]. Those results suffice to guarantee the solvability of the Nash system when the Hamiltonian  $H$  is globally Lipschitz, see for instance [6, Section 2.4.4], but a modicum of care is needed when the Hamiltonian is quadratic in the variable  $y$ . We refer to [3] and to the book [9] (see Section 6.3.1 of Chapter 6 therein) for precise solvability results that fit our framework.

Observe also that, in **A(4)**, we do not explicitly require the solution of the Nash system to be unique, but in fact, in the aforementioned references [32, 9], smooth solutions are shown to be unique. The latter fact should not come as a big surprise: Uniqueness is deeply correlated with the existence of a smooth solution. For instance, if, in addition to Assumption **A**, we ask  $(D_{x_j} v^{n,i})_{i,j=1}^n$  to be bounded, then uniqueness to the Nash system is a mere consequence of the so-called four-step scheme for forward-backward systems, see [38]. In that case, uniqueness to the Nash system then reduces, via Itô's formula, to uniqueness for a standard backward SDE with Lipschitz coefficients, the Lipschitz bound being directly inherited from the bound on the gradient of the solution. When the gradient is unbounded, the four-step scheme no longer applies, but our guess is that uniqueness may follow from results for linear BSDEs with unbounded coefficients; obviously, this latter claim goes beyond the scope of the paper.

Regarding the master equation, the main examples of solvability we have in mind may be found in [6, Theorems 2.4.2 and 2.4.5]. Results therein provide tractable assumptions on the data  $(b, f, g)$  that ensure the existence and uniqueness of a classical solution to the master equation satisfying **A(5)**. Yet it must be stressed that the results of [6] are stated on the torus and under the rather demanding assumption that the Hamiltonian  $H$  is globally Lipschitz. Again, we refer to the book [9] (specifically Section 5.4 of Chapter 5 therein) for an extension to the Euclidean setting that includes quadratic Hamiltonians. We also refer to [14, Theorems 5.3 and 5.5] for another result on existence

and uniqueness for the master equation, although the boundedness requirements of the derivatives, as stated in **A(5)**, are not entirely verified there. Notice in this regard that, although **A(5)** just refers to the existence of a classical solution to the master equation, it implicitly implies uniqueness: We show in Remark 4.5 that, under the additional (modest) Assumption **B** spelled out below, the master equation has a unique solution within the class described in Assumption **A(5)**. Lastly, observe that, for some of the arguments below, the assumptions on the second order derivatives of  $U$  in the direction of the measure can be relaxed when  $\sigma_0 = 0$ , i.e., when there is no common noise. This fact is made clear in [6]: Therein, the assumptions that are needed to prove the solvability of the master equation are slightly weaker in the absence of a common noise (compare Theorems 2.4.2 and 2.4.5 in [6]). We refrain from distinguishing the two cases  $\sigma_0 = 0$  and  $\sigma_0 \neq 0$  in our paper, since the second-order derivatives of  $U$  with respect to  $m$  will be needed to establish the CLT, even in the case  $\sigma_0 = 0$ , see Assumption **C** in Section 3.3.

As indicated above, our analysis heavily depends on properties of the Hamiltonian  $H$ . We thus need an additional assumption to control the growth of the function  $\hat{f}$ , which is defined in (2.7), using of course the same function  $\hat{\alpha}$  from Assumption **A(1)**. We provide two alternatives in Assumptions **B** and **B'**. Assumption **B** is more in the spirit of [6], while Assumption **B'** fits the framework addressed in [9, Sections 5.4 and 6.3].

**Assumption B.**  $\hat{f}(x, m, y)$  is Lipschitz in  $y$ , uniformly in  $(x, m)$ . That is, there exists  $C > 0$  such that, for all  $x, y, y' \in \mathbb{R}^d$  and  $m \in \mathcal{P}^{p^*}(\mathbb{R}^d)$ ,

$$|\hat{f}(x, m, y) - \hat{f}(x, m, y')| \leq C|y - y'|.$$

We will make use of one more set of assumptions, which weakens the Lipschitz assumption on  $\hat{f}$  in favor of uniform boundedness of solutions to the  $n$ -player Nash systems. Note that boundedness of  $D_x U$  and  $D_m U$  in Assumption **A(5)** ensures that  $U$  has linear growth, by Lemma A.2. On the other hand, the following assumption will require  $U$  to be bounded.

**Assumption B'.**

1. The solution  $U$  to the master equation is uniformly bounded.
2. The Nash system solutions  $(v^{n,i})_{i=1}^n$  are bounded, uniformly in  $n$  and  $i$ .
3.  $\hat{f}(x, m, y)$  is locally Lipschitz in  $y$  with quadratic growth, uniformly in  $(x, m)$ . That is, there exists  $C > 0$  such that, for all  $x, y, y' \in \mathbb{R}^d$  and  $m \in \mathcal{P}^{p^*}(\mathbb{R}^d)$ ,

$$|\hat{f}(x, m, y) - \hat{f}(x, m, y')| \leq C(1 + |y| + |y'|)|y - y'|.$$

**Remark 2.2.** It is typically difficult to find estimates on the Nash system solutions  $v^{n,i}$  which are uniform in  $n$ , and this is a key challenge of MFG analysis. However, assumption **B'(2)** can be directly verified in many situations. For instance, if  $\hat{f}$  is of the form  $\hat{f}(x, m, y) = f_1(x, m) + f_2(x, m, y) \cdot y$ , where  $f_2$  is of linear growth in  $y$  uniformly in  $(x, m)$ , and if  $f_1$  and  $g$  are uniformly bounded, then **B'(2)** holds. See [9, Chapter 6, Section 6.3.1] for details.

It may appear that we have not imposed any assumptions directly on the terminal cost function  $g$ . In fact, there are implicit requirements coming from Assumption **A(5)** along with the boundary condition  $U(T, x, m) = g(x, m)$ .

### 3 Statements of main results

This section summarizes the main results of the paper on the  $n$ -player equilibrium empirical measures  $(m_{\mathbf{X}}^n)_{n \geq 1}$  and on their marginal flows  $((m_{\mathbf{X}_t}^n)_{t \in [0, T]})_{n \geq 1}$ , defined by the SDE (2.9). Proofs are deferred to later sections.

### 3.1 Law of large numbers

We first state the law of large numbers, regarding the convergence of  $(m_{\mathbf{X}}^n)_{n \geq 1}$  to  $\mu$ , where  $\mu$  is defined by the McKean-Vlasov equation (2.11). This is essentially the result of [6], see also [9, Chapter 6], though we formulate the result in Euclidean space instead of on the torus; we include the proof in Section 4.4.

**Theorem 3.1.** *Suppose Assumption A holds, as well as either B or B'. Then*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \mathcal{W}_{2, \mathbb{C}^d}^2(m_{\mathbf{X}}^n, \mu) \right] = 0.$$

There are several ways to identify the rate of convergence in Theorem 3.1, and the interested reader is referred to the companion paper [18] for a detailed discussion. Notably, the dimension  $d$  appears in the rate, and we explain below the role of the dimension in the central limit theorem.

### 3.2 Diffusively scaled fluctuations

The central limit theorem is stated as a limit theorem for the sequence of rescaled differences  $((S_t^n := \sqrt{n}(m_{\mathbf{X}_t}^n - \mu_t))_{t \in [0, T]})_{n \geq 1}$  between the empirical measures and the solution to the MFG constructed above.

To state a central limit theorem, we will regard each  $(S_t^n)_{t \in [0, T]}$  as a process with values in a distributional space. To do so, we use the same spaces as in [40] or [25]. For an integer  $j \geq 0$ , a real  $\alpha > 0$ , and a smooth function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  with compact support, we let:

$$\|g\|_{j, \alpha}^2 := \sum_{|\mathbf{k}| \leq j} \int_{\mathbb{R}^d} \frac{|D^{\mathbf{k}} g(x)|^2}{1 + |x|^{2\alpha}} dx, \quad (3.1)$$

where  $\mathbf{k}$  denotes a multi-index  $\mathbf{k} = (k_1, \dots, k_d)$  and  $|\mathbf{k}| := k_1 + \dots + k_d$  is its length. We then call  $\mathcal{H}^{j, \alpha}$  the completion of the set of smooth functions on  $\mathbb{R}^d$  with compact support with respect to  $\|\cdot\|_{j, \alpha}$ . It is a Hilbert space when equipped with  $\|\cdot\|_{j, \alpha}$ . We denote by  $\mathcal{H}^{-j, \alpha}$  its dual space, and by  $\|\cdot\|_{-j, \alpha}$  the corresponding norm. Below, we see  $((S_t^n)_{t \in [0, T]})_{n \geq 1}$  as a sequence with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  for

$$\lambda_d := \lfloor d/2 \rfloor + 1.$$

The limiting law of the sequence  $((S_t^n)_{t \in [0, T]})_{n \geq 1}$  will be identified as the solution to a properly stated SPDE. In order to formulate this SPDE, we define, for any  $m \in \mathcal{P}^p(\mathbb{R}^d)$ , the operator  $\mathcal{A}_{t, m}$  by:

$$\begin{aligned} [\mathcal{A}_{t, m} \phi](x) &:= D_x \phi(x) \cdot \widehat{b}(x, m, D_x U(t, x, m)) + \frac{1}{2} \text{Tr}[D_x^2 \phi(x)(\sigma \sigma^\top + \sigma_0 \sigma_0^\top)] \\ &\quad + \int_{\mathbb{R}^d} D_x \phi(v) \cdot \frac{\delta}{\delta m} \left[ \widehat{b}(v, m, D_x U(t, v, m)) \right] (x) m(dv), \quad x \in \mathbb{R}^d, \end{aligned} \quad (3.2)$$

for  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ . Observe if needed that the derivative with respect to  $m$  in the second line may be rewritten as:

$$\begin{aligned} &\frac{\delta}{\delta m} \left[ \widehat{b}(v, m, D_x U(t, v, m)) \right] (x) \\ &= \frac{\delta}{\delta m} \widehat{b}(v, m, D_x U(t, v, m), x) + D_y \widehat{b}(v, m, D_x U(t, v, m)) \frac{\delta}{\delta m} D_x U(t, v, m, x), \end{aligned}$$

where  $D_y$  acts on the third argument of  $\widehat{b}$ . We will see in (5.5) that, for  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$ ,  $D_x \phi(x)$  and  $D_x^2 \phi(x)$  exist pointwise and are bounded by  $C \|\phi\|_{4+2\lambda_d, \lambda_d} (1 + |x|^{\lambda_d})$  for a universal constant  $C$ ; in particular,  $[\mathcal{A}_{t, m} \phi](x)$  exist pointwise. If, as functions of

$(t, x, m)$ ,  $\widehat{b}(x, m, D_x U(t, x, m))$  and  $[\delta/\delta m][\widehat{b}(v, m, D_x U(t, v, m))](x)$  are sufficiently smooth and  $m \in \mathcal{P}^{\lambda_d}(\mathbb{R}^d)$ , then  $\mathcal{A}_{t,m}\phi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}$ .

The limiting SPDE is driven by three inputs: an initial condition and two noises. To state it properly, we assume that the probability space used to construct the mean field game is rich enough to carry a triple  $(W, \theta_0, \xi)$ , where  $W$  is the same common noise as before,  $\theta_0$  is independent of  $(W, \xi)$  and is a centered Gaussian random variable with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  with covariance:

$$\forall \phi_1, \phi_2 \in \mathcal{H}^{2+2\lambda_d, \lambda_d}, \quad \mathbb{E}[\theta_0(\phi_1)\theta_0(\phi_2)] = \langle \mu_0, \phi_1 \phi_2 \rangle,$$

and where  $\xi$  is, conditional on  $W$ , a continuous centered Gaussian process with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  with covariance:

$$\forall \phi_1, \phi_2 \in \mathcal{H}^{2+2\lambda_d, \lambda_d}, \forall s, t \in [0, T], \quad \mathbb{E}[\xi_t(\phi_1)\xi_s(\phi_2) | W] = \int_0^{s \wedge t} \langle \mu_r, \sigma \sigma^\top D_x \phi_1 \cdot D_x \phi_2 \rangle dr.$$

(Recall that  $\mu$  is  $W$ -measurable.) We will see in Section 5 that both covariances are well-defined provided  $p' \geq 2\lambda_d$ , with  $p'$  as in Assumption A. In both of them,  $\langle \cdot, \cdot \rangle$  denotes the duality product.

We will see in Lemma 5.8 that, with probability 1, for all  $t \in (0, T]$ ,  $\mu_t$  has a density. As a consequence,  $(\xi_t)_{t \in [0, T]}$  may be represented as  $(\xi_t(\phi) = \int_0^t \sqrt{\mu_s}(\sigma^\top D_x \phi) \cdot d\beta_s)_{0 \leq t \leq T}$ , where  $\beta = (\beta^1, \dots, \beta^{d_0})$  is a  $d_0$ -tuple of independent cylindrical Wiener processes with values in  $L^2(\mathbb{R}^d)$ ,  $\beta$  being independent of  $(\theta_0, W)$ . Since  $\mu$  is  $W$ -measurable,  $(W, \mu)$ ,  $\theta_0$  and  $\beta$  are independent.

We state the SPDE for the weak limit  $(S_t)_{t \in [0, T]}$  of  $((S_t^n)_{t \in [0, T]})_{n \geq 1}$  as an equation in  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$ :

$$d\langle S_t, \phi \rangle = \langle S_t, \mathcal{A}_{t, \mu_t} \phi \rangle dt + \langle S_t, \sigma_0^\top D_x \phi \rangle \cdot dW_t + d\xi_t(\phi), \quad t \in [0, T], \quad \phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}, \quad (3.3)$$

where  $\langle S_t, \sigma_0^\top D_x \phi \rangle \cdot dW_t$  is understood as  $\sum_{i=1}^d \langle S_t, (\sigma_0^\top D_x \phi)_i \rangle dW_t^i$ . As we already alluded to right above, we here require the test function to be in  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$  in order to guarantee that  $\mathcal{A}_{t, \mu_t} \phi$  indeed belongs to  $\mathcal{H}^{2+2\lambda_d, \lambda_d}$ . This is the minimal regularity we need to give a meaning to the first term of the right-hand side. We refer the reader to Lemma 5.5 for the derivation of the SPDE and to Theorem 5.6 for the proof of uniqueness.

As we will see in the example treated in Section 6, an important feature of this SPDE is the fact that the operator  $\mathcal{A}_{t, \mu_t}$  is nonlocal. To clarify the challenges this presents, suppose for the moment that there is no common noise,  $\sigma_0 = 0$ . Then, if  $\mathcal{A}_{t, \mu_t}$  were local, we could regard (3.3) as a (standard) parabolic partial differential equation driven by the adjoint  $\mathcal{A}_{t, \mu_t}^*$  and forced by an additive noise; then, we could use existing results on the fundamental solution of  $\mathcal{A}_{t, \mu_t}^*$  to represent the solution. In a sense, the same is true when  $\sigma_0$  is non zero, because it suffices to shift  $\phi$  into  $\phi(\cdot - \sigma_0 W_t)$  to recover the case without common noise.

Of course, the nonlocal term in (3.2) arises because of the McKean-Vlasov interaction, and we must be prepared to face it in practice. In the example in Section 6, the coefficients are linear-quadratic, which makes the form of the interaction not too complicated. Ultimately, we manage to derive an explicit expression for the solution to (3.3), by solving first for the interaction term and then constructing the entire solution by proceeding as if the structure were local.

### 3.3 Additional assumptions and statement of the central limit theorem

In order to state our CLT, we need the following additional set of assumptions:

**Assumption C.**

1.  $\hat{b}$  is locally bounded in  $y$ , uniformly in  $(x, m)$ .
2.  $\hat{b}$  and  $D_x U$  have derivatives up to the order  $5 + 2\lambda_d$  with respect to  $(x, y)$  and  $x$  respectively, and these derivatives are jointly continuous and uniformly bounded with respect to all arguments.
3.  $\delta\hat{b}/\delta m$  and  $\delta(D_x U)/\delta m$  exist, are bounded and jointly continuous in all the arguments, and have derivatives in the variable  $v$  up to the order  $4 + 2\lambda_d$  that are jointly continuous and uniformly bounded with respect to all arguments.
4.  $\delta(D_y \hat{b})/\delta m$  exists and is jointly continuous and uniformly bounded in all the arguments.  $\delta(D_x U)/\delta m$ ,  $D_y \hat{b}$  and  $\delta(D_y \hat{b})/\delta m$  are Lipschitz continuous in  $m$  with respect to  $\mathcal{W}_1$  (understood here as  $\mathcal{W}_{1, \mathbb{R}^d}$ ), uniformly in the other arguments.
5.  $\delta^2 \hat{b}/\delta m^2$  and  $\delta^2(D_x U)/\delta m^2$  exist, are bounded and are Lipschitz continuous in  $m$  with respect to  $\mathcal{W}_1$ , uniformly in the other arguments.

**Theorem 3.2.** *Let Assumptions **A** and **C** hold, as well as either Assumption **B** or **B'**, with  $p' > 12\lambda_d$  and  $p^* = 1$ . Then, the sequence  $(S^n)_{n \geq 1}$  converges in law on the space  $C([0, T]; \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$  to the unique solution to the SPDE (3.3) with  $S_0 = \theta_0$  as initial condition.*

The rationale for requiring the coefficients to be so regular may be explained as follows. In the proof of the CLT, we need to regard the coefficients (or their derivatives, depending on the step of the proof) as test functions, namely as elements of the space  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$ . In this respect, the fact that the dimension  $d$  of the state space enters the definition of the space of test functions through the parameter  $\lambda_d$  should not come as a surprise. Indeed, the Wasserstein distance between a probability distribution and the empirical measure formed by  $n$  i.i.d. samples is less and less accurate as the dimension of the space increases (see, e.g., [20] or the refinement of Theorem 3.1 above in our companion paper [18, Theorem 3.1]). Still, we prove in Lemma 5.10 below the rather striking fact (which seems to be a new point) that the rate of convergence of the empirical distribution of a sample of i.i.d. random vectors with values  $\mathbb{R}^d$  can be made dimension-free, provided the distance is measured using smooth enough test functions. Noting, for example, that the 1-Wasserstein and total variation distances between two measures  $m$  and  $m'$  may each be written as a supremum of  $\int \phi d(m - m')$  over a suitable family of test functions  $\phi$ , it is clear that the strength of a probability metric and thus the rate of convergence of empirical measures therein depends heavily on the regularity of the test functions. Hence, to preserve the  $O(\sqrt{n})$  fluctuations, we must be prepared to refine our family of test functions. The two key difficulties are identifying the right level of smoothness and dealing with the supremum appearing in a metric of the aforementioned form. Invoking suitable forms of Sobolev embeddings seems to be the most efficient strategy for addressing both difficulties. As is well known, Sobolev embeddings heavily depend on the dimension and this is precisely what dictates the choice of the space of test functions.

Lastly, observe that requiring  $p^* = 1$  in Assumptions **A**, **B** or **B'** is somewhat redundant with the contents of Assumption **C**. Indeed, the fact that the derivatives  $\delta\hat{b}/\delta m$  and  $\delta(D_x U)/\delta m$  have bounded derivatives in  $v$  implies that both  $\hat{b}$  and  $D_x U$  are Lipschitz continuous in  $m$  with respect to  $\mathcal{W}_1$ .

As for the proof, we will follow the strategy put forth in the Introduction. Namely, we will first prove (1.5) in Section 4 below. Then, extending earlier works on the same subject, we will prove the CLT for the McKean-Vlasov system (1.4) in Section 5. We will conclude by combining the two of them.

## 4 Main estimates

All of the results proved in this paper and in the companion one [18] rely on the crucial estimates developed in this section. In the following results and proofs,  $U$  is the classical solution to the master equation (2.10). The letter  $C$  denotes a generic positive constant, which may change from line to line but is universal in the sense that it never depends on  $i$  or  $n$ , though it may of course depend on model parameters, including, e.g., the bounds on the growth and the regularity of  $U$  and its derivatives, the Lipschitz constants of  $\widehat{b}$  and  $\widehat{f}$ , and the time horizon  $T$ .

### 4.1 A preliminary estimate

For  $(t, \mathbf{x}) \in [0, T] \times (\mathbb{R}^d)^n$ , define

$$u^{n,i}(t, \mathbf{x}) = U(t, x_i, m_{\mathbf{x}}^n).$$

The following result, largely borrowed from Proposition 6.1.3 in [6], shows that the functions  $(u^{n,i})_{i=1}^n$  nearly solve the  $n$ -player Nash system defined in Section 2.3. We provide the proof in Appendix A.

**Proposition 4.1.** *Suppose Assumption A holds as well as either Assumption B or B'. There exist a constant  $C < \infty$  and, for each  $n \in \mathbb{N}$ , continuous functions  $r^{n,i} : [0, T] \times (\mathbb{R}^d)^n \rightarrow \mathbb{R}$ , for  $1 \leq i \leq n$ , with*

$$\|r^{n,i}\|_{\infty} \leq C/n,$$

such that

$$\begin{aligned} & \partial_t u^{n,i}(t, \mathbf{x}) \\ & + H(x_i, m_{\mathbf{x}}^n, D_{x_i} u^{n,i}(t, \mathbf{x})) + \sum_{j=1, j \neq i}^n D_{x_j} u^{n,i}(t, \mathbf{x}) \cdot \widehat{b}(x_j, m_{\mathbf{x}}^n, D_{x_j} u^{n,j}(t, \mathbf{x})) \\ & + \frac{1}{2} \sum_{j=1}^n \text{Tr} \left[ \sigma \sigma^\top D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{x}) \right] + \frac{1}{2} \sum_{j,k=1}^n \text{Tr} \left[ \sigma_0 \sigma_0^\top D_{x_j, x_k}^2 u^{n,i}(t, \mathbf{x}) \right] = -r^{n,i}(t, \mathbf{x}), \end{aligned}$$

for  $(t, \mathbf{x}) \in (0, T) \times (\mathbb{R}^d)^n$ .

To proceed, we define an  $n$ -particle SDE system of McKean-Vlasov type, which we will compare to the true Nash system. Precisely, let  $\overline{\mathbf{X}} = (\overline{X}^1, \dots, \overline{X}^n)$  solve the approximating  $n$ -particle system

$$d\overline{X}_t^i = \widehat{b}(\overline{X}_t^i, m_{\overline{\mathbf{X}}_t}^n, D_{x_i} U(t, \overline{X}_t^i, m_{\overline{\mathbf{X}}_t}^n)) dt + \sigma d\overline{B}_t^i + \sigma_0 dW_t, \quad \overline{X}_0^i = X_0^i. \quad (4.1)$$

Because of Assumptions A(1) and A(5), this SDE system admits a unique strong solution and (by A(3) as well) the processes  $(W, B^1, \overline{X}^1), \dots, (W, B^n, \overline{X}^n)$  are exchangeable. We make the following convenient abbreviations: for  $i, j = 1, \dots, n$ ,

$$Y_t^i = v^{n,i}(t, \mathbf{X}_t), \quad Z_t^{i,j} = D_{x_j} v^{n,i}(t, \mathbf{X}_t), \quad (4.2)$$

$$\mathcal{Y}_t^i = U(t, X_t^i, m_{\mathbf{X}_t}^n) = u^{n,i}(t, \mathbf{X}_t), \quad \mathcal{Z}_t^{i,j} = D_{x_j} u^{n,i}(t, \mathbf{X}_t). \quad (4.3)$$

Recalling the definition of  $u^{n,i}$ , we can apply Proposition 2.1 to get

$$\mathcal{Z}_t^{i,j} = D_{x_j} [U(t, X_t^i, m_{\mathbf{X}_t}^n)] = \begin{cases} \frac{1}{n} D_m U(t, X_t^i, m_{\mathbf{X}_t}^n, X_t^j) & \text{if } j \neq i, \\ D_{x_i} U(t, X_t^i, m_{\mathbf{X}_t}^n) + \frac{1}{n} D_m U(t, X_t^i, m_{\mathbf{X}_t}^n, X_t^i) & \text{if } j = i. \end{cases} \quad (4.4)$$

In particular, recalling that  $D_x U$  and  $D_m U$  are bounded, we have the crucial bounds

$$|\mathcal{Z}_t^{i,j}| \leq \frac{C}{n}, \text{ for } j \neq i, \quad |\mathcal{Z}_t^{i,i}| \leq C. \quad (4.5)$$



Also, in what follows, for  $i = 1, \dots, n$ , define:

$$M_t^i = \int_0^t \left( \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot \sigma dB_s^j + \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot \sigma_0 dW_s \right), \quad (4.6)$$

$$N_t^i = \int_0^t (Y_s^i - \mathcal{Y}_s^i) dM_s^i. \quad (4.7)$$

## 4.2 Main estimate

We state and prove our main estimate first under assumptions **A** and **B**, and then under assumptions **A** and **B'**.

**Theorem 4.2.** *Suppose Assumptions **A** and **B** hold. Then, there exists  $C < \infty$  such that, for each  $n$ ,*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \int_0^T |D_{x_i} v^{n,i}(t, \mathbf{X}_t) - D_x U(t, X_t^i, m_{\mathbf{X}_t}^n)|^2 dt \right] \leq \frac{C}{n^2}, \quad (4.8)$$

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \|X^i - \bar{X}^i\|_\infty^2 \right] \leq \frac{C}{n^2}. \quad (4.9)$$

Moreover, it holds almost surely that

$$\frac{1}{n} \sum_{i=1}^n \|X^i - \bar{X}^i\|_\infty^2 \leq \frac{C}{n} \sum_{i=1}^n \int_0^T |D_{x_i} v^{n,i}(t, \mathbf{X}_t) - D_x U(t, X_t^i, m_{\mathbf{X}_t}^n)|^2 dt, \quad (4.10)$$

$$\frac{1}{n} \sum_{i=1}^n \int_0^T |D_{x_i} v^{n,i}(t, \mathbf{X}_t) - D_x U(t, X_t^i, m_{\mathbf{X}_t}^n)|^2 dt \leq \frac{C}{n} \sum_{i=1}^n \int_0^T |Z_t^{i,i} - \mathcal{Z}_t^{i,i}|^2 dt + \frac{C}{n^2}, \quad (4.11)$$

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

$$\frac{1}{n} \sum_{i=1}^n [N^i]_t \leq \frac{C}{n^3} \sum_{i=1}^n [M^i]_t, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n [M^i]_T \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n |N_T^i|. \quad (4.12)$$

**Remark 4.3.** As explained right after (1.3), the components of the Nash equilibrium  $\mathbf{X} = (X^1, \dots, X^n)$  are exchangeable whenever the Nash system (1.2) has a unique solution satisfying **A**(4). Then, the processes  $(X^1, \bar{X}^1), \dots, (X^n, \bar{X}^n)$  are also exchangeable, and, as a result, the bounds (4.8) and (4.9) are not only true for the mean over the indices  $i = 1, \dots, n$  but are also true for any single index  $i = 1, \dots, n$ .

**Remark 4.4.** In this paper, we make no use of (4.12), which is used only in the companion paper [18]. Since the proofs follow from the same computations, we find it more convenient to prove here all estimates at once. This permits us not to repeat the arguments in [18] and to instead directly invoke Theorem 4.2.

**Remark 4.5.** As we will see in Section 4.4 below, the law of large numbers of Theorem 3.1 follows quickly from Theorem 4.2. Interestingly, these theorems together imply uniqueness for the master equation, within the class described in Assumption **A**, and we sketch the argument here. Let  $\tilde{U}$  denote another smooth solution, and let  $(\tilde{\mathcal{X}}, \tilde{\mu})$  solve the corresponding McKean-Vlasov equation (2.11), driven by the same Wiener processes  $B^1$  and  $W$ , but with  $D_x U$  replaced by  $D_x \tilde{U}$ . Then Theorem 3.1 shows that  $(m_{\mathbf{X}_t}^n)_{t \in [0, T]}$  converges in probability in  $C([0, T]; \mathcal{P}^{p*}(\mathbb{R}^d))$  to both  $(\mu_t)_{t \in [0, T]}$  and  $(\tilde{\mu}_t)_{t \in [0, T]}$ , so  $\mu = \tilde{\mu}$  and  $\mathcal{X} = \tilde{\mathcal{X}}$  a.s.

Combining (4.8) and (4.9) and using in addition the smoothness of  $D_x U$  and  $D_x \tilde{U}$ , we get

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \int_0^T |D_x U(t, \bar{X}_t^i, m_{\bar{X}_t}^n) - D_x \tilde{U}(t, \bar{X}_t^i, m_{\bar{X}_t}^n)|^2 dt \\ &= \mathbb{E} \int_0^T |D_x U(t, \mathcal{X}_t, \mu_t) - D_x \tilde{U}(t, \mathcal{X}_t, \mu_t)|^2 dt. \end{aligned}$$

By expanding  $(U(t, \mathcal{X}_t, \mu_t))_{t \in [0, T]}$  and  $(\tilde{U}(t, \mathcal{X}_t, \mu_t))_{t \in [0, T]}$  using Itô's formula for functions defined on  $\mathcal{P}^2(\mathbb{R}^d)$  (see for instance [14] for the case  $\sigma_0 = 0$  and [9, Chapter 4, Section 4.3] for the general case), and by invoking the fact that both  $U$  and  $\tilde{U}$  solve the master equation, we finally obtain that, for any  $t \in [0, T]$ ,  $U(t, \mathcal{X}_t, \mu_t)$  and  $\tilde{U}(t, \mathcal{X}_t, \mu_t)$  are almost surely equal. Choosing arbitrarily the initial condition of the game (i.e., choosing arbitrarily the initial time and the initial distribution), we deduce that  $U(t, \cdot, m)$  and  $\tilde{U}(t, \cdot, m)$  are equal for any  $t \in [0, T]$  and any  $m$  with full support. By a standard mollification argument, the same holds true for any  $m$ , whether the support of  $m$  is full or not.

### Proof of Theorem 4.2

We begin by proving that (4.9) follows from (4.8). Abbreviate

$$\begin{aligned} \hat{b}^{n,i}(t, \mathbf{x}) &= \hat{b}(x_i, m_{\mathbf{x}}^n, D_{x_i} v^{n,i}(t, \mathbf{x})), \\ \tilde{b}^{n,i}(t, \mathbf{x}) &= \hat{b}(x_i, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)), \end{aligned} \quad (4.13)$$

for  $(t, \mathbf{x}) \in [0, T] \times (\mathbb{R}^d)^n$ . Note first that (2.9), (4.1) and the Lipschitz assumption on  $\hat{b}$  from Assumption A(1) yields (together with the fact that  $\mathcal{W}_{p^*} \leq \mathcal{W}_2$ ):

$$\begin{aligned} &|X_t^i - \bar{X}_t^i|^2 \\ &\leq T \int_0^t |\hat{b}^{n,i}(s, \mathbf{X}_s) - \tilde{b}^{n,i}(s, \bar{\mathbf{X}}_s)|^2 ds \\ &\leq 2T \int_0^t |\hat{b}^{n,i}(s, \mathbf{X}_s) - \tilde{b}^{n,i}(s, \mathbf{X}_s)|^2 ds + 2T \int_0^t |\tilde{b}^{n,i}(s, \mathbf{X}_s) - \tilde{b}^{n,i}(s, \bar{\mathbf{X}}_s)|^2 ds \\ &\leq C \int_0^t |D_{x_i} v^{n,i}(s, \mathbf{X}_s) - D_x U(s, X_s^i, m_{\mathbf{X}_s}^n)|^2 ds \\ &\quad + C \int_0^t (|X_s^i - \bar{X}_s^i|^2 + \mathcal{W}_2^2(m_{\mathbf{X}_s}^n, m_{\bar{\mathbf{X}}_s}^n)) ds. \end{aligned}$$

By Gronwall's inequality,

$$|X_t^i - \bar{X}_t^i|^2 \leq C \int_0^t |D_{x_i} v^{n,i}(s, \mathbf{X}_s) - D_x U(s, X_s^i, m_{\mathbf{X}_s}^n)|^2 ds + C \int_0^t \mathcal{W}_2^2(m_{\mathbf{X}_s}^n, m_{\bar{\mathbf{X}}_s}^n) ds. \quad (4.14)$$

Noting that

$$\mathcal{W}_2^2(m_{\mathbf{X}_s}^n, m_{\bar{\mathbf{X}}_s}^n) \leq \frac{1}{n} \sum_{i=1}^n |X_s^i - \bar{X}_s^i|^2,$$

we may average (4.14) over  $i = 1, \dots, n$  and apply Gronwall's inequality once again to obtain

$$\mathcal{W}_2^2(m_{\mathbf{X}_t}^n, m_{\bar{\mathbf{X}}_t}^n) \leq \frac{C}{n} \sum_{i=1}^n \int_0^t |D_{x_i} v^{n,i}(s, \mathbf{X}_s) - D_x U(s, X_s^i, m_{\mathbf{X}_s}^n)|^2 ds.$$

Using this bound in (4.14) yields (4.10). Also, (4.9) will follow from (4.8), which we now prove.

First, use Itô's formula, (2.8), the fact that  $v^{n,i}$  solves the Nash system, and (4.2) to obtain:

$$\begin{aligned} dY_t^i &= \partial_t v^{n,i}(t, \mathbf{X}_t) dt \\ &+ \sum_{j=1}^n D_{x_j} v^{n,i}(t, \mathbf{X}_t) \cdot \widehat{b}^{n,j}(t, \mathbf{X}_t) dt + \frac{1}{2} \sum_{j=1}^n \text{Tr} \left[ D_{x_j, x_j}^2 v^{n,i}(t, \mathbf{X}_t) \sigma \sigma^\top \right] dt \\ &+ \frac{1}{2} \sum_{j,k=1}^n \text{Tr} \left[ D_{x_j, x_k}^2 v^{n,i}(t, \mathbf{X}_t) \sigma_0 \sigma_0^\top \right] dt + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma dB_t^j + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma_0 dW_t \\ &= -\widehat{f} \left( X_t^i, m_{\mathbf{X}_t}^n, Z_t^{i,i} \right) dt + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma dB_t^j + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma_0 dW_t. \end{aligned}$$

On the other hand, use Itô's formula, (2.8), Proposition 4.1, and (4.3) to get:

$$\begin{aligned} d\mathcal{Y}_t^i &= \partial_t u^{n,i}(t, \mathbf{X}_t) dt + \sum_{j=1}^n D_{x_j} u^{n,i}(t, \mathbf{X}_t) \cdot \widehat{b} \left( X_t^j, m_{\mathbf{X}_t}^n, D_{x_j} v^{n,j}(t, X_t^j, m_{\mathbf{X}_t}^n) \right) dt \\ &+ \frac{1}{2} \sum_{j=1}^n \text{Tr} \left[ D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{X}_t) \sigma \sigma^\top \right] dt + \frac{1}{2} \sum_{j,k=1}^n \text{Tr} \left[ D_{x_j, x_k}^2 u^{n,i}(t, \mathbf{X}_t) \sigma_0 \sigma_0^\top \right] dt \\ &+ \sum_{j=1}^n Z_t^{i,j} \cdot \sigma dB_t^j + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma_0 dW_t \\ &= -\left[ \widehat{f} \left( X_t^i, m_{\mathbf{X}_t}^n, Z_t^{i,i} \right) + r^{n,i}(t, \mathbf{X}_t) \right] dt + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma dB_t^j + \sum_{j=1}^n Z_t^{i,j} \cdot \sigma_0 dW_t \\ &+ \sum_{j=1}^n Z_t^{i,j} \cdot \left( \widehat{b} \left( X_t^j, m_{\mathbf{X}_t}^n, Z_t^{j,j} \right) - \widehat{b} \left( X_t^j, m_{\mathbf{X}_t}^n, Z_t^{j,j} \right) \right) dt. \end{aligned}$$

Use Itô's formula and the terminal conditions for the Nash system and the master equation to get, for  $t \in [0, T]$ ,

$$\begin{aligned} (Y_t^i - \mathcal{Y}_t^i)^2 &= 2 \int_t^T (Y_s^i - \mathcal{Y}_s^i) \left\{ \widehat{f} \left( X_s^i, m_{\mathbf{X}_s}^n, Z_s^{i,i} \right) - \widehat{f} \left( X_s^i, m_{\mathbf{X}_s}^n, Z_s^{i,i} \right) \right. \\ &\quad \left. + \sum_{j=1}^n Z_s^{i,j} \cdot \left( \widehat{b} \left( X_s^j, m_{\mathbf{X}_s}^n, Z_s^{j,j} \right) - \widehat{b} \left( X_s^j, m_{\mathbf{X}_s}^n, Z_s^{j,j} \right) \right) - r^{n,i}(s, \mathbf{X}_s) \right\} ds \\ &\quad - \int_t^T \sum_{j=1}^n \left| \sigma^\top (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \right|^2 ds - \int_t^T \left| \sum_{j=1}^n \sigma_0^\top (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \right|^2 ds \\ &\quad - 2 \int_t^T (Y_s^i - \mathcal{Y}_s^i) \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot (\sigma dB_s^j + \sigma_0 dW_s). \end{aligned} \quad (4.15)$$

Use the Lipschitz assumption on  $\widehat{b}$  along with the estimates (4.5) to get:

$$\begin{aligned} & \left| \sum_{j=1}^n \mathcal{Z}_s^{i,j} \cdot \left( \widehat{b}(X_s^j, m_{\mathbf{X}_s}^n, Z_s^{j,j}) - \widehat{b}(X_s^j, m_{\mathbf{X}_s}^n, \mathcal{Z}_s^{j,j}) \right) \right| \\ & \leq C \sum_{j=1}^n |\mathcal{Z}_s^{i,j}| |Z_s^{j,j} - \mathcal{Z}_s^{j,j}| \\ & \leq C |Z_s^{i,i} - \mathcal{Z}_s^{i,i}| + \frac{C}{n} \sum_{j=1, j \neq i}^n |Z_s^{j,j} - \mathcal{Z}_s^{j,j}|. \end{aligned}$$

Substitute the last inequality back into (4.15), and use the Lipschitz assumption on  $\widehat{f}$  from Assumption **B**, the elementary inequality  $2ab \leq \epsilon^{-1}a^2 + \epsilon b^2$ , and the estimate  $|r^{n,i}| \leq C/n$  from Proposition 4.1 to obtain, for  $\epsilon > 0$ ,

$$\begin{aligned} (Y_t^i - \mathcal{Y}_t^i)^2 & \leq \frac{C}{n^2} + C \int_t^T \left( \epsilon^{-1} (Y_s^i - \mathcal{Y}_s^i)^2 + \epsilon |Z_s^{i,i} - \mathcal{Z}_s^{i,i}|^2 + \frac{\epsilon}{n} \sum_{j \neq i} |Z_s^{j,j} - \mathcal{Z}_s^{j,j}|^2 \right) ds \\ & \quad - \int_t^T \sum_{j=1}^n |\sigma^\top (Z_s^{i,j} - \mathcal{Z}_s^{i,j})|^2 ds - \int_t^T \left| \sum_{j=1}^n \sigma_0^\top (Z_t^{i,j} - \mathcal{Z}_t^{i,j}) \right|^2 ds \\ & \quad - 2 \int_t^T (Y_s^i - \mathcal{Y}_s^i) \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot (\sigma dB_s^j + \sigma_0 dW_s). \end{aligned} \quad (4.16)$$

Here  $C$  does not depend on  $\epsilon$ . Noting that  $\sigma$  is non-degenerate, we may choose  $\epsilon$  small enough and average over  $i = 1, \dots, n$  to get

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left( (Y_t^i - \mathcal{Y}_t^i)^2 + \int_t^T \sum_{j=1}^n |Z_s^{i,j} - \mathcal{Z}_s^{i,j}|^2 ds \right) \\ & \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n \int_t^T (Y_s^i - \mathcal{Y}_s^i)^2 ds - \frac{\widetilde{C}}{n} \sum_{i=1}^n \int_t^T (Y_s^i - \mathcal{Y}_s^i) \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot (\sigma dB_s^j + \sigma_0 dW_s), \end{aligned} \quad (4.17)$$

where the constant  $\widetilde{C} < \infty$  depends only on  $\sigma$ . It follows from Lemma 4.6 below that the stochastic integrals are martingales. Take conditional expectations to get, for  $r \leq t$ ,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(Y_t^i - \mathcal{Y}_t^i)^2 | \mathcal{F}_r] \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n \int_t^T \mathbb{E}[(Y_s^i - \mathcal{Y}_s^i)^2 | \mathcal{F}_r] dt.$$

Using Gronwall's inequality and taking  $r = t$  we conclude that:

$$\frac{1}{n} \sum_{i=1}^n (Y_t^i - \mathcal{Y}_t^i)^2 \leq \frac{C}{n^2}. \quad (4.18)$$

Also, taking conditional expectations in (4.17), and using (4.18), we find:

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[ \int_t^T |Z_s^{i,j} - \mathcal{Z}_s^{i,j}|^2 ds \middle| \mathcal{F}_t \right] \leq \frac{C}{n^2}. \quad (4.19)$$

Then, returning to (4.16), taking the conditional expectation given  $\mathcal{F}_t$ , and applying Gronwall's lemma yields

$$\max_{i=1, \dots, n} (Y_t^i - \mathcal{Y}_t^i)^2 \leq \frac{C}{n^2}. \quad (4.20)$$

This implies that, for  $M$  and  $N$  defined in (4.6)-(4.7),

$$\frac{1}{n} \sum_{i=1}^n [N^i]_T = \frac{1}{n} \sum_{i=1}^n \int_0^T (Y_t^i - \mathcal{Y}_t^i)^2 d[M^i]_t \leq \frac{C}{n^3} \sum_{i=1}^n [M^i]_T,$$

which proves the first part of (4.12). On the other hand, applying (4.18) to (4.17) shows that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (Y_t^i - \mathcal{Y}_t^i)^2 + \frac{1}{n} \sum_{i=1}^n ([M^i]_T - [M^i]_t) \\ & \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n \int_t^T (Y_s^i - \mathcal{Y}_s^i)^2 ds - \frac{\tilde{C}}{n} \sum_{i=1}^n (N_T^i - N_t^i) \\ & \leq \frac{C}{n^2} - \frac{\tilde{C}}{n} \sum_{i=1}^n (N_T^i - N_t^i). \end{aligned}$$

Take  $t = 0$  to derive the second part of (4.12).

Finally, using the definition (4.2) of  $Z$  and recalling from identity (4.4) and inequality (4.3) that  $D_x U(t, X_t^i, m_{\mathbf{X}_t}^n) = Z_t^{i,i} - \frac{1}{n} D_m U(t, X_t^i, m_{\mathbf{X}_t}^n, X_t^i)$  and that  $D_m U$  is bounded, we get:

$$\frac{1}{n} \sum_{i=1}^n \int_0^T |D_{x_i} v^{n,i}(t, \mathbf{X}_t) - D_x U(t, X_t^i, m_{\mathbf{X}_t}^n)|^2 dt \leq \frac{C}{n} \sum_{i=1}^n \int_0^T |Z_t^{i,i} - Z_t^{i,i}|^2 dt + \frac{C}{n^2},$$

which is (4.11). Taking expectations and using (4.19) with  $t = 0$ , we get (4.8).  $\square$

We close the section by proving standard integrability results that were used in the above proof to conclude that a certain local martingale was in fact a martingale.

**Lemma 4.6.** *Suppose Assumption A holds, as well as either Assumption B or B'. Then, for all  $i, j$ ,*

$$\mathbb{E} [\|Y^i\|_\infty^2 + \|Z^{i,j}\|_\infty^2] < \infty.$$

*Proof.* By the SDE (1.1) for  $X^i$ , we have:

$$\begin{aligned} |X_t^i| & \leq |X_0^i| + C' \left( 1 + \int_0^t |\mathbf{X}_s| ds + \int_0^t |D_{x_i} v^{n,i}(t, \mathbf{X}_s)| ds \right) + |\sigma| |B_t^i| + |\sigma_0| |W_t| \\ & \leq |X_0^i| + (C' + L_{n,i,i}) \left( 1 + \int_0^t |\mathbf{X}_s| ds \right) + |\sigma| |B_t^i| + |\sigma_0| |W_t|, \end{aligned}$$

where  $C'$  is taken from Assumption A(1) and is allowed to vary from the first to the second line and  $L_{n,i,i}$  is the constant from Assumption A(4). This holds for all  $i = 1, \dots, n$ , and Gronwall's inequality yields:

$$\|\mathbf{X}\|_\infty \leq C \left( 1 + \sum_{i=1}^n |X_0^i| + \sum_{i=1}^n \|B^i\|_\infty + \|W\|_\infty \right),$$

where we note that the constant  $C < \infty$  may depend on  $n$  here. Recalling from Assumption A(3) that  $X_0^i$  are i.i.d. with finite moments of order  $p' > 4$ , we conclude that  $\mathbb{E}[\|\mathbf{X}\|_\infty^4] < \infty$ . Using the definitions of  $Y^i$  and  $Z^{i,j}$  from (4.2) along with assumption A(4), we find

$$|Y_t^i|^2 = |v^{n,i}(t, \mathbf{X}_t)|^2 \leq 2L_{n,i}^2 (1 + |\mathbf{X}_t|^4), \quad |Z_t^{i,j}|^2 \leq 2L_{n,i,j}^2 (1 + |\mathbf{X}_t|^2).$$

$\square$

### 4.3 An alternative estimate

Under Assumption **B'**, instead of Assumption **B**, we have somewhat different estimates, based on an exponential transformation.

**Theorem 4.7.** *Suppose Assumptions **A** and **B'** hold. Then, (4.10) and (4.11) hold, and, for sufficiently large  $n$ , the estimates (4.8) and (4.9) hold. For  $i = 1, \dots, n$  and a constant  $\eta > 0$ , define  $M^i$  as in (4.6) and  $Q^i$  by*

$$Q_t^i = \int_0^t [2(Y_s^i - \mathcal{Y}_s^i) + \eta \sinh(\eta(Y_s^i - \mathcal{Y}_s^i))] dM_s^i.$$

Then, for sufficiently large  $n$  and  $\eta$ , we have for all  $t \in [0, T]$ ,

$$\frac{1}{n} \sum_{i=1}^n [Q^i]_t \leq \frac{C}{n^3} \sum_{i=1}^n [M^i]_t, \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n [M^i]_T \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n |Q_T^i|. \quad (4.21)$$

*Proof.* The proofs of (4.10) and (4.11) are the same as in Theorem 4.2. In particular, (4.9) follows from (4.8). The proof of (4.8) given below is adapted from [9, Theorem 6.32].

**Step 1:** We first estimate  $(Y_s^i - \mathcal{Y}_s^i)^2$ , for  $s \in [0, T]$ . Our starting point is the equation (4.15) from the proof of Theorem 4.2. The local Lipschitz assumption on  $\hat{f}$  along with the uniform boundedness of  $v^{n,i}$ ,  $u^{n,i}$ , and  $\mathcal{Z}_t^{i,i} = D_{x_i} u^{n,i}(t, \mathbf{X}_t)$  implied by Assumption **B'** shows that

$$\begin{aligned} & \left| (Y_r^i - \mathcal{Y}_r^i) \left( \hat{f}(X_r^i, m_{\mathbf{X}_r}^n, Z_r^{i,i}) - \hat{f}(X_r^i, m_{\mathbf{X}_r}^n, \mathcal{Z}_r^{i,i}) \right) \right| \\ & \leq C (1 + |Z_r^{i,i}| + |\mathcal{Z}_r^{i,i}|) |Y_r^i - \mathcal{Y}_r^i| |Z_r^{i,i} - \mathcal{Z}_r^{i,i}| \\ & \leq C (|Y_r^i - \mathcal{Y}_r^i| |Z_r^{i,i} - \mathcal{Z}_r^{i,i}| + |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2) \\ & \leq C (|Y_r^i - \mathcal{Y}_r^i|^2 + |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2). \end{aligned}$$

Substitute this back into equation (4.15), using the Lipschitz continuity of  $\hat{b}$ , Proposition 4.1, the estimate (4.5), and the elementary inequality  $2ab \leq a^2 + b^2$  to get

$$\begin{aligned} (Y_s^i - \mathcal{Y}_s^i)^2 & \leq \frac{C}{n^2} + C \int_s^T \left( |Y_r^i - \mathcal{Y}_r^i|^2 + |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2 + \frac{1}{n} \sum_{j=1}^n |Z_r^{i,j} - \mathcal{Z}_r^{j,j}|^2 \right) dr \\ & \quad - \int_s^T \sum_{j=1}^n \left| \sigma^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr - \int_s^T \left| \sum_{j=1}^n \sigma_0^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr \\ & \quad - 2 \int_s^T (Y_r^i - \mathcal{Y}_r^i) \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r). \end{aligned}$$

**Step 2:** The term  $|Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2$  poses some problems, as it is not canceled by the quadratic variation term. To deal with this, our next step is to estimate  $\cosh(\eta(Y_s^i - \mathcal{Y}_s^i))$  for a fixed  $\eta > 0$ , to be chosen later. We note that the constant  $C < \infty$  will change from line to line

but will not depend on the value of  $\eta$ . Use Itô's formula to get:

$$\begin{aligned} & d \cosh(\eta(Y_s^i - \mathcal{Y}_s^i)) \\ &= \eta \sinh(\eta(Y_s^i - \mathcal{Y}_s^i)) \left[ - \sum_{j=1}^n \mathcal{Z}_s^{i,j} \cdot \left( \widehat{b}(X_s^j, m_{\mathbf{X}_s}^n, Z_s^{j,j}) - \widehat{b}(X_s^j, m_{\mathbf{X}_s}^n, \mathcal{Z}_s^{j,j}) \right) \right. \\ & \quad \left. + r^{n,i}(s, \mathbf{X}_s) - \widehat{f}(X_s^i, m_{\mathbf{X}_s}^n, Z_s^{i,i}) + \widehat{f}(X_s^i, m_{\mathbf{X}_s}^n, \mathcal{Z}_s^{i,i}) \right] ds \\ & \quad + \frac{\eta^2}{2} \cosh(\eta(Y_s^i - \mathcal{Y}_s^i)) \left( \sum_{j=1}^n \left| \sigma^\top(Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \right|^2 + \left| \sum_{j=1}^n \sigma_0^\top(Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \right|^2 \right) ds \\ & \quad + \eta \sinh(\eta(Y_s^i - \mathcal{Y}_s^i)) \left( \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot \sigma dB_s^j + \sum_{j=1}^n (Z_s^{i,j} - \mathcal{Z}_s^{i,j}) \cdot \sigma_0 dW_s \right). \end{aligned}$$

As before, use the fact from (4.5) that  $\mathcal{Z}_r^{i,i}$  is uniformly bounded and the local Lipschitz continuity of  $\widehat{f}$  to get:

$$\begin{aligned} & \left| \eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \left( \widehat{f}(X_r^i, m_{\mathbf{X}_r}^n, Z_r^{i,i}) - \widehat{f}(X_r^i, m_{\mathbf{X}_r}^n, \mathcal{Z}_r^{i,i}) \right) \right| \\ & \leq C (1 + |Z_r^{i,i}| + |\mathcal{Z}_r^{i,i}|) |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))| |Z_r^{i,i} - \mathcal{Z}_r^{i,i}| \\ & \leq C |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))| |Z_r^{i,i} - \mathcal{Z}_r^{i,i}| + C |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))| |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2 \\ & \leq C |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))|^2 + C (1 + |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))|) |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2 \\ & \leq C |\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i))|^2 + C (1 + \eta \cosh(\eta(Y_r^i - \mathcal{Y}_r^i))) |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2, \end{aligned}$$

where the last line follows from the elementary inequality  $|\sinh(x)| \leq \cosh(x)$ . Thus, use the Lipschitz continuity of  $\widehat{b}$ , the inequality  $2ab \leq a^2 + b^2$ , and (4.5) to obtain

$$\begin{aligned} & \cosh(\eta(Y_s^i - \mathcal{Y}_s^i)) \\ & \leq C \int_s^T \left[ \eta^2 \sinh^2(\eta(Y_r^i - \mathcal{Y}_r^i)) + (1 + \eta \cosh(\eta(Y_r^i - \mathcal{Y}_r^i))) |Z_s^{i,i} - \mathcal{Z}_r^{i,i}|^2 \right. \\ & \quad \left. + \frac{\eta}{n} |\sinh(\eta(Y_r^i - \mathcal{Y}_r^i))| \sum_{j=1}^n |Z_r^{j,j} - \mathcal{Z}_r^{j,j}| + |r^{n,i}(r, \mathbf{X}_r)|^2 \right] dr \\ & \quad - \frac{\eta^2}{2} \int_s^T \cosh(\eta(Y_r^i - \mathcal{Y}_r^i)) \left( \sum_{j=1}^n \left| \sigma^\top(Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 + \left| \sum_{j=1}^n \sigma_0^\top(Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 \right) dr \\ & \quad - \eta \int_s^T \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r). \end{aligned}$$

Finally, since  $\sigma$  is non-degenerate, use the estimate  $\|r^{n,i}\|_\infty \leq C/n$  from Proposition 4.1

to conclude that

$$\begin{aligned} & \frac{\eta^2}{2} \int_s^T \cosh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n |Z_r^{i,j} - \mathcal{Z}_r^{i,j}|^2 dr + \cosh(\eta(Y_s^i - \mathcal{Y}_s^i)) \\ & \leq \frac{C}{n^2} + C \int_s^T \left[ \eta^2 \sinh^2(\eta(Y_r^i - \mathcal{Y}_r^i)) + (1 + \eta \cosh(\eta(Y_r^i - \mathcal{Y}_r^i))) |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2 \right. \\ & \quad \left. + \frac{\eta}{n} |\sinh(\eta(Y_r^i - \mathcal{Y}_r^i))| \sum_{j=1}^n |Z_r^{j,j} - \mathcal{Z}_r^{j,j}| \right] dr \\ & \quad - \tilde{C} \eta \int_s^T \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r). \end{aligned}$$

where the constant  $\tilde{C} < \infty$  depends only on  $\sigma$ .

**Step 3:** We now combine the previous estimates. Use the facts that  $\sinh$  is locally Lipschitz and  $Y_s^i$  and  $\mathcal{Y}_s^i$  are uniformly bounded to find a constant  $c_\eta < \infty$  such that:

$$\begin{aligned} & (Y_s^i - \mathcal{Y}_s^i)^2 + \frac{\eta^2}{2} \int_s^T \cosh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n |Z_r^{i,j} - \mathcal{Z}_r^{i,j}|^2 dr \\ & \leq \frac{C}{n^2} + c_\eta \int_s^T |Y_r^i - \mathcal{Y}_r^i|^2 dr + \frac{C(1+\eta)}{n} \int_s^T \sum_{j=1}^n |Z_r^{j,j} - \mathcal{Z}_r^{j,j}|^2 dr \\ & \quad - \int_s^T \sum_{j=1}^n \left| \sigma^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr - \int_s^T \left| \sum_{j=1}^n \sigma_0^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr \\ & \quad + C \int_s^T (1 + \eta \cosh(\eta(Y_r^i - \mathcal{Y}_r^i))) |Z_r^{i,i} - \mathcal{Z}_r^{i,i}|^2 dr \\ & \quad - \int_s^T \left[ 2(Y_r^i - \mathcal{Y}_r^i) + \tilde{C} \eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \right] \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r). \end{aligned}$$

Now, for sufficiently large  $\eta$ , we have:

$$C(1 + \eta \cosh(\eta x)) \leq 2C\eta \cosh(\eta x) \leq \frac{\eta^2}{4} \cosh(\eta x),$$

for any  $x \in \mathbb{R}$ . Hence, for large  $\eta$ , we find

$$\begin{aligned} & (Y_s^i - \mathcal{Y}_s^i)^2 + \frac{\eta^2}{4} \int_s^T \cosh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n |Z_r^{i,j} - \mathcal{Z}_r^{i,j}|^2 dr \\ & \leq \frac{C}{n^2} + c_\eta \int_s^T |Y_r^i - \mathcal{Y}_r^i|^2 dr + \frac{C\eta}{n} \int_s^T \sum_{j=1}^n |Z_r^{j,j} - \mathcal{Z}_r^{j,j}|^2 dr \\ & \quad - \int_s^T \sum_{j=1}^n \left| \sigma^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr - \int_s^T \left| \sum_{j=1}^n \sigma_0^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr \\ & \quad - \int_s^T \left[ 2(Y_r^i - \mathcal{Y}_r^i) + \tilde{C} \eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \right] \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r). \end{aligned} \tag{4.22}$$



Averaging (4.22) over  $i = 1, \dots, n$  and increasing the value of  $\eta$  if needed, we obtain

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n (Y_s^i - \mathcal{Y}_s^i)^2 + \frac{\eta^2}{8n} \sum_{i=1}^n \int_s^T \cosh(\eta(Y_r^i - \mathcal{Y}_r^i)) \sum_{j=1}^n |Z_r^{i,j} - \mathcal{Z}_r^{i,j}|^2 dr \\ & + \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \int_s^T \left| \sigma^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr + \frac{1}{n} \sum_{i=1}^n \int_s^T \left| \sum_{j=1}^n \sigma_0^\top (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \right|^2 dr \\ & \leq \frac{C}{n^2} + \frac{c_\eta}{n} \sum_{i=1}^n \int_s^T |Y_r^i - \mathcal{Y}_r^i|^2 dr \\ & - \frac{1}{n} \sum_{i=1}^n \int_s^T \left[ 2(Y_r^i - \mathcal{Y}_r^i) + \tilde{C}_\eta \sinh(\eta(Y_r^i - \mathcal{Y}_r^i)) \right] \sum_{j=1}^n (Z_r^{i,j} - \mathcal{Z}_r^{i,j}) \cdot (\sigma dB_r^j + \sigma_0 dW_r) . \end{aligned} \quad (4.23)$$

Noting that the stochastic integral is a martingale thanks to Lemma 4.6 and the boundedness of  $Y^i$  and  $\mathcal{Y}^i$ , we take conditional expectations and apply Gronwall's inequality as in (4.18) to find

$$\frac{1}{n} \sum_{i=1}^n (Y_s^i - \mathcal{Y}_s^i)^2 \leq \frac{C}{n^2}. \quad (4.24)$$

Moreover, substituting (4.24) back into (4.23), using non-degeneracy of  $\sigma$ , and taking conditional expectations, we see that

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[ \int_s^T |Z_r^{i,j} - \mathcal{Z}_r^{i,j}|^2 dr | \mathcal{F}_s \right] \leq \frac{C}{n^2}. \quad (4.25)$$

Returning to (4.22) and following the proof of (4.20), we obtain:

$$\max_{i=1, \dots, n} (Y_s^i - \mathcal{Y}_s^i)^2 \leq \frac{C}{n^2}. \quad (4.26)$$

Recall now that  $\sinh$  is locally Lipschitz and  $Y^i$  and  $\mathcal{Y}^i$  are uniformly bounded. Using (4.26) we conclude that

$$[Q^i]_t \leq \frac{C}{n^2} [M^i]_t,$$

for all  $i \in \{1, \dots, n\}$  and  $t \in [0, T]$ . On the other hand, from (4.23) and (4.26) we get:

$$\frac{1}{n} \sum_{i=1}^n [M^i]_T \leq \frac{C}{n^2} + \frac{C}{n} \sum_{i=1}^n |Q_T^i|.$$

This completes the proof of (4.21).

**Step 4:** Finally, to prove (4.8), it suffices to take expectations in (4.11) and to use (4.25).  $\square$

#### 4.4 Law of large numbers

As in [6], the estimates of Theorems 4.2 and 4.7 imply a law of large numbers for the equilibrium empirical measure  $m_{\mathbf{X}}^n$ . That is, we can now prove Theorem 3.1:

*Proof of Theorem 3.1.* Note first that (4.9) implies

$$\mathbb{E} \left[ \mathcal{W}_{2, \mathcal{C}^d}^2(m_{\mathbf{X}}^n, m_{\mathbf{X}}^n) \right] \leq \frac{C}{n^2}. \quad (4.27)$$

On the other hand, a standard McKean-Vlasov limit theorem (e.g., [11, Section 7] or [9, Chapter 2, Section 2.1]) implies

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ \mathcal{W}_{2, \mathcal{C}^d}^2(m_{\bar{\mathbf{X}}}^n, \mu) \right] = 0.$$

This proves the theorem.  $\square$

## 5 Proof of the central limit theorem

As we already mentioned in the Introduction, the proof of the CLT is split into two parts. The first part is to control the distance between the two particle systems  $\mathbf{X}$  and  $\bar{\mathbf{X}}$ , which mostly follows from Theorem 4.2. The second part consists of deriving a CLT for the flow of empirical distributions  $(m_{\bar{\mathbf{X}}_t}^n)_{0 \leq t \leq T}$ . The first part is addressed in Section 5.3, after we have developed the requisite machinery, and we begin with the second part.

Generally speaking, the proof of the CLT goes along the lines of earlier works on the subject, see for instance [23], [40], [25] and [29]. However, among these four references, only the last one allows for a common noise in the particle system or a nonlinear mean field term. In the approach we implement below, we follow the main steps of [29], except that the space of distributions in which the CLT is proven is taken from [40, 25].

A crucial point throughout the proof is that, under the assumption of Theorem 3.2 and for the same real  $p'$  as therein, one has  $\mathbb{E}[\sup_{t \in [0, T]} \int_{\mathbb{R}^d} |x|^{p'} d\mu_t(x)] < \infty$  and  $\sup_{1 \leq i \leq N} \mathbb{E}[\sup_{t \in [0, T]} |\bar{X}_t^i|^{p'}] < \infty$ , with  $\bar{\mathbf{X}}$  as in (4.1).

### 5.1 Tightness

The first step for proving a CLT for  $\bar{\mathbf{X}}$  is to show that the difference  $(\bar{S}_t^n = \sqrt{n}(m_{\bar{\mathbf{X}}_t}^n - \mu_t))_{t \in [0, T]}$  is tight in a suitable space of distributions. This is precisely the point where we prefer to follow [40, 25] instead of [23, 29]. There are several reasons for that. First, the framework used in [40, 25] allows for weaker integrability conditions of the initial distribution. Second, the analysis therein is explicitly fitted to the higher-dimensional setting, whilst the one performed in [23, 29] is in dimension 1 only; an additional effort would be needed to make the latter adapted to the case  $d \geq 2$ , since, as we shall see below, the dimension plays a crucial role in the definition of the space of distributions used in the statement of the CLT. In this regard, it seems that the space we use below permits to track the dependence on the dimension in a more explicit way.

Here, we use the same space of distributions as in [40, 25], as defined in (3.1). Roughly speaking, we shall prove under the assumptions of Theorem 3.2 that the sequence  $(\bar{S}_t^n)_{t \in [0, T]}$  is tight in the space  $C([0, T], \mathcal{H}^{-(2+2\lambda_d), 2\lambda_d})$ , where  $\lambda_d = \lfloor d/2 \rfloor + 1$ . See Theorem 5.3 below for a more precise statement.

We only provide the main steps of the proof, as it is quite similar to the aforementioned references. There are two key ingredients, described in the following two subsections.

#### 5.1.1 Coupling with the limiting process.

The first step to get tightness is to prove that the particle system  $\bar{\mathbf{X}}$  is at most at distance  $n^{-1/2}$  of the auxiliary particle system  $\widehat{\mathbf{X}}$  defined as the solution to the following SDE system:

$$d\widehat{X}_t^i = \widehat{b}(\widehat{X}_t^i, \mu_t, D_x U(t, \widehat{X}_t^i, \mu_t))dt + \sigma dB_t^i + \sigma_0 dW_t, \quad \widehat{X}_0^i = \bar{X}_0^i, \quad (5.1)$$

for  $i \in \{1, \dots, n\}$ . It is customary to introduce  $\widehat{\mathbf{X}}$  in the analysis of mean field systems, see for instance [44]. A key feature is that, conditional on  $W$ , the particles  $\widehat{X}^1, \dots, \widehat{X}^n$  are independent. Moreover, the processes  $(W, B^1, \widehat{X}^1), \dots, (W, B^n, \widehat{X}^n)$  are exchangeable.

Below, we use the shortened notation  $\tilde{b}(t, x, m) = \hat{b}(x, m, D_x U(t, x, m))$ , which is similar to (4.13). We have the following bound, analogous to [29, Theorem 2.4].

**Lemma 5.1.** *Under the assumptions of Theorem 3.2, we can find a constant  $C$  such that, for any  $n \geq 1$  and any  $i \in \{1, \dots, n\}$ ,*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |\bar{X}_t^i - \hat{X}_t^i|^4 \right]^{1/4} \leq C n^{-1/2},$$

*Proof.*

**Step 1:** The key point is to show the bound

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \mathcal{W}_2(m_{\bar{X}_t}^n, m_{\hat{X}_t}^n)^4 \right]^{1/4} \leq C n^{-1/2}. \quad (5.2)$$

Although the derivation of (5.2) may seem standard, as the rate  $n^{-1/2}$  is the one we expect from the law of large numbers, there is in fact a subtle point. The proof would be fairly straightforward if the mean field interaction was of order 1, in the sense that  $\tilde{b}(t, x, m)$  was of the form of  $\int_{\mathbb{R}^d} \bar{b}(t, x, x') dm(x')$ , for some Lipschitz function  $\bar{b} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  (cf. [40, 25]). Obviously, there is no reason for this to be the case here, as  $U$  is expected to inherit the nonlinear structure of the master equation. As shown in [29] (see in particular the assumption (S3) therein), (5.2) holds true provided that for any  $n \in \mathbb{N}$ , any  $m_0 \in \mathcal{P}^{p'}(\mathbb{R}^d)$ , any i.i.d. samples  $\zeta = (\zeta_1, \dots, \zeta_n)$  with law  $m_0$ , any  $i \in \{1, \dots, n\}$ , and any  $t \in [0, T]$  we have:

$$\mathbb{E} \left[ |\tilde{b}(t, \zeta_i, m_\zeta^n) - \tilde{b}(t, \zeta_i, m_0)|^4 \right]^{1/4} \leq C \left( 1 + \int_{\mathbb{R}^d} |x|^4 m_0(dx) \right) n^{-1/2}. \quad (5.3)$$

The role of (5.3) is well-understood. Indeed, the proof of (5.2) is a straightforward application of Gronwall's lemma, provided we can estimate the term

$$\int_0^T \mathbb{E} \left[ |\tilde{b}(t, \hat{X}_t^i, m_{\hat{X}_t}^n) - \tilde{b}(t, \hat{X}_t^i, \mu_t)|^4 \right]^{1/4} dt.$$

Now, it suffices to invoke (5.3) but under  $\mathbb{E}[\cdot | W]$ , as the particles  $(\hat{X}_t^1, \dots, \hat{X}_t^n)$  are i.i.d. with law  $\mu_t$ , conditional on  $W$ .

**Step 2:** We turn to the proof of (5.3). Using the Lipschitz property of  $\tilde{b}$  in  $m$ , we can easily replace  $m_\zeta^n$  by  $m_\zeta^{n,-i}$ , the empirical measure of  $(\zeta_k)_{k \neq i}$ . Indeed, since  $\tilde{b}$  is Lipschitz, the left-hand side in (5.3) can be bounded by:

$$C n^{-1/2} \mathbb{E} [|\zeta_i|^4]^{1/4} + \mathbb{E} \left[ |\tilde{b}(t, \zeta_i, m_\zeta^{n,-i}) - \tilde{b}(t, \zeta_i, m_0)|^4 \right]^{1/4}.$$

By conditioning on  $\zeta_i$  in the second term and by changing the variable  $n$  into  $n - 1$ , it is in fact sufficient to prove:

$$\mathbb{E} \left[ |\tilde{b}(t, x, m_\zeta^n) - \tilde{b}(t, x, m_0)|^4 \right]^{1/4} \leq C \left( 1 + \int_{\mathbb{R}^d} |x'|^4 m_0(dx') \right) n^{-1/2}, \quad (5.4)$$

for any  $x \in \mathbb{R}^d$ , the constant  $C$  being independent of  $x$ . So, the only point is to find sufficient conditions on  $\tilde{b}$  to guarantee (5.4). This is exactly the purpose of Lemma 5.10 in Section 5.5. It requires that  $m \mapsto \tilde{b}(t, x, m)$  is  $\mathcal{C}^2$  in the sense defined in Subsection 2.2 with bounded derivatives of order 1 and 2 and that the derivative of order 2 is Lipschitz continuous in  $m$  with respect to the 1-Wasserstein distance, uniformly in the other parameters. This holds true under the assumption of Theorem 3.2.  $\square$

### 5.1.2 A preliminary bound toward tightness

A second important point in the proof is that the following (Sobolev) embeddings are continuous:

$$\mathcal{H}^{\ell+j,\alpha} \hookrightarrow \mathfrak{C}^{j,\alpha}, \quad \ell > \frac{d}{2}, \quad j \geq 0, \quad \alpha \geq 0, \quad (5.5)$$

$$\mathcal{H}^{\ell+j,\alpha} \hookrightarrow_{\text{H.S.}} \mathcal{H}^{j,\alpha+\beta}, \quad \ell > \frac{d}{2}, \quad j \geq 0, \quad \alpha \geq 0, \quad \beta > \frac{d}{2}, \quad (5.6)$$

with similar Sobolev embeddings for the dual spaces (with converse order), where  $\mathfrak{C}^{j,\alpha}$  is the space of functions  $g$  with continuous partial derivatives up to order  $j$  such that:

$$\lim_{|x| \rightarrow \infty} \frac{|D^{\mathbf{k}}g(x)|}{1 + |x|^\alpha} = 0, \quad \text{for all multi-indices } \mathbf{k} \text{ such that } |\mathbf{k}| \leq j.$$

In the embedding (5.5), the space  $\mathfrak{C}^{j,\alpha}$  is equipped with the norm:

$$\|g\|_{\mathfrak{C}^{j,\alpha}} = \sum_{|\mathbf{k}| \leq j} \sup_{x \in \mathbb{R}^d} \frac{|D^{\mathbf{k}}g(x)|}{1 + |x|^\alpha}.$$

In (5.6), the symbol H.S. means that the embedding is Hilbert-Schmidt.

As a consequence of Lemma 5.1, we have the following result, the proof of which is a mere adaptation of Proposition 5.3 in [40].

**Proposition 5.2.** *Under the assumptions of Theorem 3.2, for any  $\epsilon \in (0, 1)$  such that  $p' > 8\lambda_d(1 + \epsilon)/(1 - \epsilon)$  with  $p'$  as in Assumption **A**, it holds:*

$$\sup_{n \geq 1} \sup_{t \in [0, T]} \mathbb{E}[\|\bar{S}_t^n\|_{-(1+\lambda_d), 2\lambda_d}^{2(1+\epsilon)}] < \infty.$$

*Proof.* For a test function  $\phi \in \mathcal{H}^{1+\lambda_d, 2\lambda_d}$ , we let:

$$\bar{K}_t^n(\phi) = n^{-1/2} \sum_{i=1}^n (\phi(\bar{X}_t^i) - \phi(\hat{X}_t^i)), \quad \bar{L}_t^n(\phi) = n^{-1/2} \sum_{i=1}^n (\phi(\hat{X}_t^i) - \mathbb{E}[\phi(\hat{X}_t^1) | W]),$$

noting that  $\bar{K}_t^n(\phi) + \bar{L}_t^n(\phi) = \bar{S}_t^n(\phi)$ . As for the first term, we use the fact that

$$\|\bar{K}_t^n\|_{-(1+\lambda_d), 2\lambda_d}^{2(1+\epsilon)} \leq n^{1+\epsilon} n^{-1} \sum_{i=1}^n \|d_{\bar{X}_t^i, \hat{X}_t^i}\|_{-(1+\lambda_d), 2\lambda_d}^{2(1+\epsilon)} = n^\epsilon \sum_{i=1}^n \|d_{\bar{X}_t^i, \hat{X}_t^i}\|_{-(1+\lambda_d), 2\lambda_d}^{2(1+\epsilon)},$$

and we conclude as in the derivation of [40, (5.3)], using Hölder instead of Cauchy-Schwarz inequality with exponents  $2/(1 + \epsilon)$  and  $2/(1 - \epsilon)$ . Here we define  $d_{x,y}(\phi) = \phi(x) - \phi(y)$  and  $\|d_{x,y}\|_{-(1+\lambda_d), 2\lambda_d} \leq C|x - y|(1 + |x|^{2\lambda_d} + |y|^{2\lambda_d})$ , see [40, Lemma 6.1].

Regarding the second term, we apply Rosenthal's inequality for Hilbert-valued independent random variables, see [42, Theorem 5.2]. Indeed, we observe that

$$\bar{L}_t^n = n^{-1/2} \sum_{i=1}^n (d_{\hat{X}_t^i} - \mathbb{E}[d_{\hat{X}_t^1} | W]),$$

with  $d_x(\phi) = \phi(x)$  and  $\|d_x\|_{-(1+\lambda_d), 2\lambda_d} \leq C(1 + |x|^{2\lambda_d})$ , see again [40]. In particular,  $\|d_{\hat{X}_t^i} - \mathbb{E}[d_{\hat{X}_t^1}]\|_{-(1+\lambda_d), 2\lambda_d} \leq C(1 + |\hat{X}_t^i|^{2\lambda_d} + \mathbb{E}[|\hat{X}_t^1|^{2\lambda_d}])$ . The result follows.  $\square$

### 5.1.3 Statement and proof of tightness

The proof of tightness relies on the following expansion, which holds true for any smooth test function  $\phi$  with compact support.

$$\begin{aligned} d\langle \bar{S}_t^n, \phi \rangle &= \sqrt{n} \left[ \langle m_{\bar{X}_t}^n, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, m_{\bar{X}_t}^n) \rangle - \langle \mu_t, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, \mu_t) \rangle \right] dt \\ &\quad + \frac{1}{2} \langle \bar{S}_t^n, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 \phi] \rangle dt \\ &\quad + \langle \bar{S}_t^n, \sigma_0^\top D_x \phi \rangle \cdot dW_t + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma^\top D_x \phi(\bar{X}_t^i)) \cdot dB_t^i. \end{aligned} \quad (5.7)$$

Following [40, (5.6)], we let:

$$M_t^{n,\phi} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t (\sigma^\top D_x \phi(\bar{X}_s^i)) \cdot dB_s^i.$$

If we call  $(d_x D_{x_\ell})(\phi) = (\partial \phi / \partial x_\ell)(x)$  for  $\ell \in \{1, \dots, d\}$ , we have  $\|d_x D_{x_\ell}\|_{-(1+\lambda_d), 2\lambda_d} \leq C(1 + |x|^{2\lambda_d})$ , see the proof of [40, Lemma 6.1]. Hence, if  $(e_k)_{k \geq 1}$  denotes a complete orthonormal system in  $\mathcal{H}^{1+\lambda_d, 2\lambda_d}$ , we use the martingale property of  $M^{n,\phi}$ , the Burkholder-Davis-Gundy inequality, the exchangeability of  $(W, B^1, \bar{X}^1), \dots, (W, B^n, \bar{X}^n)$  and Parseval's identity to get

$$\mathbb{E} \left[ \sum_{k \geq 1} \sup_{t \in [0, T]} |M_t^{n, e_k}|^2 \right] \leq C \sum_{k \geq 1} \int_0^T \mathbb{E}[|D_x e_k(\bar{X}_t^1)|^2] dt \leq C \int_0^T \mathbb{E}[(1 + |\bar{X}_t^1|^{4\lambda_d})] dt < \infty.$$

Therefore, we can define:

$$M_t^n = \sum_{k \geq 1} M_t^{n, e_k} e_k^*, \quad t \in [0, T],$$

where  $e_k^* \in \mathcal{H}^{-(1+\lambda_d), 2\lambda_d}$  is defined by  $e_k^*(\phi) = \langle e_k, \phi \rangle$ . The process  $(M_t^n)_{t \in [0, T]}$  forms a continuous martingale with values in  $\mathcal{H}^{-(1+\lambda_d), 2\lambda_d}$ . Obviously,  $(M^n(\phi))_{t \in [0, T]}$  coincides almost surely with  $M^{n,\phi}$ . Actually, following [40, Proposition 5.7], the above bound implies

$$\sup_{n \geq 1} \mathbb{E} \left[ \sup_{t \in [0, T]} \|M_t^n\|_{-(1+\lambda_d), 2\lambda_d}^2 \right] < \infty. \quad (5.8)$$

Here is now the main claim of this subsection. We first define, for any  $(t, m, m') \in [0, T] \times \mathcal{P}^{p*}(\mathbb{R}^d) \times \mathcal{P}^{p*}(\mathbb{R}^d)$ ,

$$[\mathcal{G}(t, m, m')\phi](v) = \left\langle m', D_x \phi(\cdot) \cdot \left( \int_0^1 \frac{\delta \tilde{b}}{\delta m}(t, \cdot, rm + (1-r)m', v) dr \right) \right\rangle, \quad (5.9)$$

where recall that, for every fixed  $(t, x) \in [0, T] \times \mathbb{R}^d$ , we write  $[\delta \tilde{b} / \delta m](t, x, \cdot, \cdot) : \mathcal{P}^{p*}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$  to represent the  $\delta / \delta m$ -derivative (defined in Section 2.2) of the real-valued function  $m \mapsto \tilde{b}(t, x, m)$  defined on  $\mathcal{P}^{p*}(\mathbb{R}^d)$ .

**Proposition 5.3.** *Under the assumptions of Theorem 3.2, the sequences  $((\bar{S}_t^n)_{t \in [0, T]})_{n \geq 1}$  and  $((M_t^n)_{t \in [0, T]})_{n \geq 1}$  are tight on the space  $C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$ , where again  $\lambda_d := \lfloor d/2 \rfloor + 1$ .*

*Proof.*

**Step 1:** We write the first term in the right hand side of (5.7) in the form:

$$\begin{aligned} &\sqrt{n} \left[ \langle m_{\bar{X}_t}^n, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, m_{\bar{X}_t}^n) \rangle - \langle \mu_t, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, \mu_t) \rangle \right] \\ &= \langle \bar{S}_t^n, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, m_{\bar{X}_t}^n) \rangle + \Delta_t^n(\phi), \end{aligned} \quad (5.10)$$

where

$$\begin{aligned}\Delta_t^n(\phi) &= \sqrt{n} \langle \mu_t, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, m_{\bar{\mathbf{X}}_t}^n) - D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, \mu_t) \rangle \\ &= \int_{\mathbb{R}^d} d\mu_t(x) \int_0^1 dr \int_{\mathbb{R}^d} D_x \phi(x) \cdot \frac{\delta \tilde{b}}{\delta m}(t, x, rm_{\bar{\mathbf{X}}_t}^n + (1-r)\mu_t, v) d\bar{S}_t^n(v) \\ &= \int_{\mathbb{R}^d} d\bar{S}_t^n(v) \left[ \int_{\mathbb{R}^d} D_x \phi(x) \cdot \left( \int_0^1 \frac{\delta \tilde{b}}{\delta m}(t, x, rm_{\bar{\mathbf{X}}_t}^n + (1-r)\mu_t, v) dr \right) d\mu_t(x) \right].\end{aligned}$$

Following [40, (5.8)], we write  $\Delta_t^n(\phi)$  as

$$\Delta_t^n(\phi) = \left\langle \bar{S}_t^n, [\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)\phi](\cdot) \right\rangle, \quad (5.11)$$

with  $\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)\phi$  defined as in (5.9). We then have the analogue of [40, Lemma 5.6], namely

$$\begin{aligned}\|[\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)\phi](\cdot)\|_{(1+\lambda_d), 2\lambda_d} &\leq C \|\phi\|_{2+2\lambda_d, \lambda_d} \left( \int_{\mathbb{R}^d} (1 + |y|^{2\lambda_d}) d\mu_t(y) \right)^{1/2} \\ \|D_{x_\ell}^2 \phi(\cdot)\|_{1+\lambda_d, 2\lambda_d} + \|D_{x_\ell} \phi(\cdot)\|_{1+\lambda_d, 2\lambda_d} &\leq C \|\phi\|_{3+\lambda_d, 2\lambda_d}, \quad \ell, \ell' = 1, \dots, d, \\ \| [D_x \phi \cdot \tilde{b}(t, \cdot, m_{\bar{\mathbf{X}}_t}^n)](\cdot) \|_{1+\lambda_d, 2\lambda_d} &\leq C \|\phi\|_{3+\lambda_d, 2\lambda_d}.\end{aligned} \quad (5.12)$$

Indeed, under assumption **C**,  $\tilde{b}$  and its derivatives in  $x$  up to order  $\lambda_d + 1$  are bounded, and the last two lines follow easily. After noticing that  $\|\phi\|_{3+\lambda_d, 2\lambda_d} \leq C \|\phi\|_{2+2\lambda_d, \lambda_d}$  since  $\lambda_d \geq 1$ , the first line follows from the fact that  $\delta \tilde{b} / \delta m$  and its derivatives in  $v$  up to order  $\lambda_d + 1$  are bounded, see [40, Lemma 5.6]. Notice that the right-hand side in the first line is random, which is a special feature due to the presence of the common noise. It implies in particular that

$$\|\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)^* \bar{S}_t^n\|_{-(2+2\lambda_d), \lambda_d}^2 \leq C \left( 1 + \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\mu_t(y) \right) \|\bar{S}_t^n\|_{-(1+\lambda_d), 2\lambda_d}^2. \quad (5.13)$$

**Step 2:** We now have the analogue of [40, Proposition 5.9]. We claim that, for a complete orthonormal system  $(\psi_p)_{p \geq 1}$  in  $\mathcal{H}^{2+2\lambda_d, \lambda_d}$ , we have

$$\sup_{n \geq 1} \mathbb{E} \left[ \sum_{p \geq 1} \sup_{t \in [0, T]} \langle \bar{S}_t^n, \psi_p \rangle^2 \right] < \infty. \quad (5.14)$$

In comparison with the proof of [40, Proposition 5.9], there are two main differences, concerning the two terms (see (5.7)):

$$\int_0^T \left| \left\langle \bar{S}_s^n, [\mathcal{G}(s, m_{\bar{\mathbf{X}}_s}^n, \mu_s)\psi_p](\cdot) \right\rangle \right|^2 ds, \quad \sup_{t \in [0, T]} \left| \int_0^t \langle \bar{S}_s^n, \sigma_0^\top D_x \psi_p \rangle \cdot dW_s \right|^2. \quad (5.15)$$

As for the first term, the difference with [40] comes from the fact that, in (5.13), the integral with respect to  $\mu_t$  is random. A careful inspection of the proof of [40, Proposition 5.9] combined with the statement of our Proposition 5.2 shows that (with  $\epsilon$  as therein)

$$\begin{aligned}\mathbb{E} \left[ \sum_{p \geq 1} \langle \bar{S}_t^n, [\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)\psi_p](\cdot) \rangle^2 \right] &= \mathbb{E} \left[ \|\mathcal{G}(t, m_{\bar{\mathbf{X}}_t}^n, \mu_t)^* \bar{S}_t^n\|_{-(2+2\lambda_d), \lambda_d}^2 \right] \\ &\leq C \mathbb{E} \left[ \left( 1 + \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\mu_t(y) \right) \|\bar{S}_t^n\|_{-(1+\lambda_d), 2\lambda_d}^2 \right] \\ &\leq C \mathbb{E} \left[ \left( 1 + \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\mu_t(y) \right)^{\frac{1+\epsilon}{\epsilon}} \right]^{\frac{\epsilon}{1+\epsilon}}.\end{aligned}$$

Since  $p' > 12\lambda_d$ , we can choose  $\epsilon = 1/5$ , then  $8\lambda_d(1+\epsilon)/(1-\epsilon) = 12\lambda_d$  and  $2\lambda_d(1+\epsilon)/\epsilon = 12\lambda_d$ , and then Proposition 5.2 applies and the above right hand side is indeed finite, uniformly in  $t$ .

We now handle the second term in (5.15). This term is new. By Doob's inequality,

$$\begin{aligned} \sum_{p \geq 1} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \int_0^t \langle \bar{S}_s^n, \sigma_0^\top D_x \psi_p \rangle \cdot dW_s \right|^2 \right] &\leq C \sum_{p \geq 1} \sum_{i=1}^d \mathbb{E} \left[ \int_0^T \langle \bar{S}_s^n, \sigma_0^\top D_{x_i} \psi_p(\cdot) \rangle^2 ds \right] \\ &\leq C \sum_{i=1}^d \mathbb{E} \left[ \int_0^T \|D_{x_i}^* \bar{S}_s^n\|_{-(2+2\lambda_d), \lambda_d}^2 ds \right] \\ &\leq C \sum_{i=1}^d \mathbb{E} \left[ \int_0^T \|\bar{S}_s^n\|_{-(1+2\lambda_d), \lambda_d}^2 ds \right] \end{aligned}$$

which completes the proof of (5.14) since the space  $\mathcal{H}^{-(1+\lambda_d), 2\lambda_d}$  embeds (continuously) in  $\mathcal{H}^{-(1+2\lambda_d), \lambda_d}$ , see (5.6) and then Proposition 5.2. We then complete the proof with an easy adaptation of [40, Proposition 5.10] and [40, Theorem 5.12] (which relies on (5.6) as well).  $\square$

## 5.2 Identification of limit points

We now identify the (weak) limit points of the sequence  $(\bar{S}^n, M^n)_{n \geq 1}$ .

### 5.2.1 Limit of $(M^n)_{n \geq 1}$

The first step is to identify the limit (in law) of the sequence  $(M^n)_{n \geq 1}$ . We follow the proof of [40, Theorem 5.14]. The only difference with the framework investigated in [40] comes from the fact that the limit process  $(\mu_t)_{t \in [0, T]}$  is random, which prompts us to work conditionally on its realization.

To state the result, we first construct, on the same probability space as the one used to define the mean field game, a random element  $(\theta_0, \xi)$  of the space  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d} \times C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$  which, as in Subsection 3.2, satisfies the following:  $\theta_0$  and  $(W, \mu, \xi)$  are independent,  $\theta_0$  is a Gaussian random variable with covariance

$$\forall \phi_1, \phi_2 \in \mathcal{H}^{2+2\lambda_d, \lambda_d}, \quad \mathbb{E}[\theta_0(\phi_1)\theta_0(\phi_2)] = \langle \mu_0, \phi_1 \phi_2 \rangle, \quad (5.16)$$

and  $\xi$  is, conditional on  $(W, \mu)$ , a continuous Gaussian process with covariance

$$\begin{aligned} \forall \phi_1, \phi_2 \in \mathcal{H}^{2+2\lambda_d, \lambda_d}, \forall s, t \in [0, T], \\ \mathbb{E}[\xi_t(\phi_1)\xi_s(\phi_2) | W, \mu] = \int_0^{s \wedge t} \langle \mu_r, \sigma \sigma^\top D_x \phi_1 \cdot D_x \phi_2 \rangle dr. \end{aligned} \quad (5.17)$$

Notice from (5.5) that, for all  $\phi \in \mathcal{H}^{1+2\lambda_d, \lambda_d}$ , the function  $\mathbb{R}^d \ni x \mapsto \phi(x)/(1 + |x|^{\lambda_d})$  is bounded. Since  $p'$  in Assumption **A** is greater than  $2\lambda_d$ , we have  $\mathbb{E}[\sup_{t \in [0, T]} \langle \mu_t, \phi^2 \rangle] < \infty$ , which shows that the above covariances make sense. Recall also that  $\mu$  is  $W$ -measurable, so that conditioning on  $\mu$  in (5.17) is redundant. We claim:

**Lemma 5.4.** *The sequence  $(W, \mu, \bar{S}_0^n, M^n)_{n \geq 1}$  converges in law to  $(W, \mu, \theta_0, \xi)$  on the space  $C([0, T], \mathbb{R}^{d_0}) \times C([0, T], \mathcal{P}^1(\mathbb{R}^d)) \times \mathcal{H}^{-(2+2\lambda_d), \lambda_d} \times C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$ .*

*Proof.* Because the first two components  $(W, \mu)$  are fixed, it follows from Proposition 5.3 that the sequence in question is tight. Hence, we characterize the limit points.

We first observe that, for any  $n \geq 1$ ,  $(W, \mu)$  is  $W$ -measurable and  $\bar{S}_0^n$  is measurable with respect to  $(X_0^1, \dots, X_0^N)$ . As a consequence,  $(W, \mu)$  and  $\bar{S}_0^n$  are independent and, for

any weak limit  $(\varpi, \nu, \vartheta_0)$  of  $(W, \mu, \bar{S}_0^n)_{n \geq 1}$ ,  $(\varpi, \nu)$  and  $\vartheta_0$  are also independent. For any  $\phi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}$ , the law of  $\vartheta_0(\phi)$  is easily identified by means of the standard version of the central limit theorem. As a byproduct, we easily identify the law of  $\vartheta_0$ , regarded as a random variable with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$ , with the law of  $\theta_0$ .

We finally show that any weak limit  $(\varpi, \nu, \vartheta_0, \zeta)$  of the quadruple  $(W, \mu, \bar{S}_0^n, M^n)_{n \geq 1}$  coincides in law with  $(W, \mu, \theta_0, \xi)$ . The proof is as follows. Fix  $\phi, \psi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}$ . We know that, for any  $n \geq 1$ ,  $M^n(\phi)$  can be written as

$$M_t^n(\phi) = \int_0^t \frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma^\top D_x \phi(\bar{X}_s^i)) \cdot dB_s^i, \quad t \in [0, T].$$

As  $(W, \mu, \bar{S}_0^n)$  is independent of  $B^1, \dots, B^n$ , we easily deduce that, conditional on  $(W, \mu, \bar{S}_0^n)$ , the three processes  $(M_t^n(\phi))_{t \in [0, T]}$ ,  $(M_t^n(\psi))_{t \in [0, T]}$ , and

$$M_t^n(\phi)M_t^n(\psi) - \int_0^t \langle m_{\bar{X}_s}^n, \sigma \sigma^\top D_x \phi \cdot D_x \psi \rangle ds, \quad t \in [0, T], \quad (5.18)$$

are martingales. That is, for any real-valued bounded continuous functions  $h_1$  on  $C([0, T], \mathbb{R}^{d_0}) \times C([0, T], \mathcal{P}^2(\mathbb{R}^d)) \times \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  and  $h_2$  on  $C([0, T], \mathbb{R})$  and for any  $0 \leq s \leq t \leq T$ ,

$$\mathbb{E} \left[ h_1(W, \mu, \bar{S}_0^n) h_2(M_{\cdot \wedge s}^n(\phi)) \left( M_t^n(\phi) - M_s^n(\phi) \right) \right] = 0,$$

and similarly for the other two martingales. Letting  $n$  tend to  $\infty$ , we get

$$\mathbb{E} \left[ h_1(\varpi, \nu, \theta_0) h_2(\zeta_{\cdot \wedge s}(\phi)) \left( \zeta_t(\phi) - \zeta_s(\phi) \right) \right] = 0.$$

Arguing similarly for the other two processes, we deduce that, conditional on  $(\varpi, \nu, \vartheta_0)$ , the processes  $(\zeta_t(\phi))_{t \in [0, T]}$ ,  $(\zeta_t(\psi))_{t \in [0, T]}$ , and (the limit of (5.18))

$$\zeta_t(\phi)\zeta_t(\psi) - \int_0^t \langle \nu_s, \sigma \sigma^\top D_x \phi \cdot D_x \psi \rangle ds, \quad t \in [0, T]$$

are martingales. So, conditional on  $(\varpi, \nu, \vartheta_0)$ ,  $(\zeta_t(\phi))_{t \in [0, T]}$  and  $(\zeta_t(\psi))_{t \in [0, T]}$  are Gaussian with covariation process  $(\int_0^t \langle \nu_s, |\sigma \sigma^\top D_x \phi \cdot D_x \psi|^2 \rangle ds)_{t \in [0, T]}$ . Since this covariation process is  $\nu$ -measurable, and since  $(\varpi, \nu)$  and  $\vartheta_0$  are independent, we deduce in particular that  $(\varpi, \nu, \zeta)$  and  $\vartheta_0$  are independent. Moreover, we have found that, conditionally on  $(\varpi, \nu)$ ,  $\zeta$  is a Gaussian process on  $H^{-(2+2\lambda_d), \lambda_d}$  with covariance

$$\begin{aligned} \forall \phi_1, \phi_2 \in \mathcal{H}^{2+2\lambda_d, \lambda_d}, \quad \forall s, t \in [0, T], \\ \mathbb{E} [\zeta_t(\phi_1) \zeta_s(\phi_2) | \varpi, \nu] = \int_0^{s \wedge t} \langle \nu_r, \sigma \sigma^\top D_x \phi_1 \cdot D_x \phi_2 \rangle dr. \end{aligned} \quad (5.19)$$

Since  $(W, \mu)$  and  $(\varpi, \nu)$  have the same distribution, this shows that  $(\varpi, \nu, \zeta)$  has the same law as  $(W, \mu, \xi)$ .  $\square$

### 5.2.2 Limit of $(m_{(\bar{X}^1, \dots, \bar{X}^n)}^n, W, \bar{S}^n)_{n \geq 1}$

We claim:

**Lemma 5.5.** *Let  $(\nu, \nu', \varpi, \zeta, S)$  be a (weak) limit point of  $(m_{\bar{X}}^n, \mu, W, M^n, \bar{S}^n)_{n \geq 1}$ . Then, with probability 1 (under the limiting probability distribution), we have  $\nu' = \nu$ , and  $S$  solves the equation:*

$$\begin{aligned} d\langle S_t, \phi \rangle = & \left[ \langle S_t, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, \nu_t) \rangle + \langle S_t, [\mathcal{G}(t, \nu_t, \nu_t) \phi](\cdot) \rangle \right] dt \\ & + \frac{1}{2} \langle S_t, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 \phi] \rangle dt + \langle S_t, \sigma_0^\top D_x \phi \rangle \cdot d\varpi_t + d\zeta_t(\phi), \end{aligned} \quad (5.20)$$

for  $t \in [0, T]$  and  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$ , with  $\mathcal{G}$  as in (5.9).



*Proof.* The fact that  $\nu = \nu'$  is a mere consequence of the fact that  $\sup_{t \in [0, T]} W_2(m_{\mathbf{X}_t}^n, \mu_t)$  tends to 0 with probability 1.

In order to handle the second part of the statement, we come back to the expansion (5.7), see also (5.10) and (5.11). We rewrite the expansion under the form

$$\begin{aligned} d\langle \bar{S}_t^n, \phi \rangle &= \left[ \langle \bar{S}_t^n, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, m_{\mathbf{X}_t}^n) \rangle + \langle \bar{S}_t^n, [\mathcal{G}(t, m_{\mathbf{X}_t}^n, \mu_t) \phi](\cdot) \rangle \right] dt \\ &\quad + \frac{1}{2} \langle \bar{S}_t^n, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 \phi] \rangle dt + \langle \bar{S}_t^n, \sigma_0^\top D_x \phi \rangle \cdot dW_t + dM_t^n(\phi), \quad t \in [0, T]. \end{aligned}$$

We take  $\phi$  in the space  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$  as required in the statement. Under the assumption of Theorem 3.2,  $\tilde{b}$  is differentiable in  $x$  up to the order  $2 + 2\lambda_d$ , with bounded and continuous derivatives. Similarly,  $\delta \tilde{b} / \delta m$  is differentiable in the variable  $v$  up to the order  $2 + 2\lambda_d$ , with joint continuous (in all the arguments) and bounded derivatives. In particular,  $\mathcal{G}(t, m_{\mathbf{X}_t}^n, \mu_t) \phi$  has bounded derivatives up to the order  $2 + 2\lambda_d$ . The key point is that a function with derivatives up to the order  $2 + 2\lambda_d$  that are at bounded is in  $\mathcal{H}^{2+2\lambda_d, \lambda_d}$ , which follows from the fact that  $\mathbb{R}^d \ni x \mapsto 1/(1 + |x|^{2\lambda_d})$  is integrable since  $2\lambda_d \geq d + 1$ . Hence,  $\mathcal{G}(t, m_{\mathbf{X}_t}^n, \mu_t) \phi$  converges almost surely in  $\mathcal{H}^{2+2\lambda_d, \lambda_d}$  to  $\mathcal{G}(t, \mu_t, \mu_t) \phi$  and there is no real difficulty for passing to the limit in the above expansion. We get (5.20).

By a standard separability argument, (5.20) holds with probability 1 for all  $\phi$  in the space  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$ .  $\square$

Observe that equation (5.20) may be regarded as an equation in the space  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$ , while  $S$  is known to take values in the smaller space  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$ . In order to prove convergence of the sequence  $(m_{\mathbf{X}_t}^n, W, M^n, \bar{S}^n)_{n \geq 1}$ , one must prove uniqueness of the solution to this equation.

### 5.2.3 Uniqueness to the limiting equation

We shall prove in Subsection 5.4 the following uniqueness result:

**Theorem 5.6.** *Fix a given filtered probability space  $(\Xi, \mathcal{G}, \mathbb{G}, \mathbb{Q})$  equipped with an adapted triple  $(\varpi, \nu, \zeta)$  such that  $\varpi = (\varpi_t)_{t \in [0, T]}$  is a  $d$ -dimensional Brownian motion,  $\nu = (\nu_t)_{t \in [0, T]}$  is a  $\mathcal{P}^{2\lambda_d}(\mathbb{R}^d)$ -valued process, and, conditional on  $(\varpi, \nu)$ ,  $\zeta$  is a Gaussian process with the same covariance structure as in (5.19). Then, for any square integrable initial condition  $S_0$  with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  and independent of  $(\varpi, \nu, \zeta)$ , the equation (5.20) has at most one (adapted) square-integrable solution with paths in  $C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$ .*

Invoking Lemma 5.4 and recalling that strong uniqueness implies weak uniqueness in law, we have the following corollary.

**Corollary 5.7.** *The sequence  $(W, m_{\mathbf{X}_t}^n, \bar{S}^n)_{n \geq 1}$  converges in law on the space  $C([0, T], \mathbb{R}^d) \times C([0, T], \mathcal{P}^1(\mathbb{R}^d)) \times C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$  to the tuple  $(W, \mu, S)$  where  $S$  solves, for any  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$  the equation:*

$$\begin{aligned} d\langle S_t, \phi \rangle &= \left[ \langle S_t, D_x \phi(\cdot) \cdot \tilde{b}(t, \cdot, \mu_t) \rangle + \langle S_t, [\mathcal{G}(t, \mu_t, \mu_t) \phi](\cdot) \rangle \right] dt \\ &\quad + \frac{1}{2} \langle S_t, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 \phi] \rangle dt + \langle S_t, \sigma_0^\top D_x \phi \rangle \cdot dW_t + d\xi_t(\phi), \quad t \in [0, T], \end{aligned} \quad (5.21)$$

where  $\xi$  is given by (5.17) and where  $S_0$  is independent of  $(W, \mu, \xi)$  and has the same law as  $\theta_0$  in (5.16).

### 5.2.4 Reformulation of the limiting equation

In fact, we can provide another representation of the term  $\xi$ , based on the fact that, almost surely, for any  $t > 0$ ,  $\mu_t$  has a density and may be identified with a function, which is the purpose of the following lemma.

**Lemma 5.8.** *With probability 1, for any  $t > 0$ , the probability measure  $\mu_t$  has a density with respect to the Lebesgue measure on  $\mathbb{R}^d$ .*

*Proof.* Let  $Y_t = X_t - \sigma_0 W_t$ . Then

$$dY_t = \tilde{b}(t, Y_t + \sigma_0 W_t, \mu_t)dt + \sigma dB_t.$$

Let  $\tilde{\mu}_t = \mathcal{L}(Y_t | (W_s)_{s \leq t}) = \mathcal{L}(Y_t | W)$  denote the flow of conditional laws of  $Y$  given  $W$ . Given  $W$ , the process  $Y$  solves a deterministic SDE with uniformly bounded drift. By Girsanov's theorem, the law  $\tilde{\mu}_t$  has strictly positive density with respect to Lebesgue measure, almost surely. Hence, so too does  $\mu_t = \mathcal{L}(X_t | (W_s)_{s \leq t})$ , which is simply a translation of  $\tilde{\mu}_t$ .  $\square$

We then have the following result.

**Proposition 5.9.** *Let  $(\beta_t := (\beta_t^1, \dots, \beta_t^{d_0}))_{t \in [0, T]}$  be a  $W$ -independent  $d_0$ -tuple of independent cylindrical Wiener processes with values in  $L^2(\mathbb{R}^d)$ . Then, the collection*

$$\left( \left( \int_0^t (\sqrt{\mu_s} \sigma^\top D_x \phi) \cdot d\beta_s \right)_{\phi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}} \right)_{t \in [0, T]}$$

*defines a process with values in  $C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$  which has the same joint law with  $(W, \mu)$  as  $\xi$  in (5.17).*

In other words, the above statement says that, in (5.20), the term  $\zeta_t(\phi)$  may be written as  $\int_0^t (\sqrt{\mu_s} \sigma^\top D_x \phi) \cdot d\beta_s$ .

*Proof.* The proof follows from a standard computation of covariance. To do so, it suffices to observe from (5.5) that, for any  $\phi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}$ , the function  $\mathbb{R}^d \ni x \mapsto |D_x \phi(x)|/(1 + |x|^{\lambda_d})$  is bounded. As a result, the function  $(t, x) \mapsto \sqrt{\mu_t(x)} D_x \phi(x)$  is square-integrable, since  $p'$  in Assumption **A** is greater than  $2\lambda_d$ .  $\square$

### 5.3 Proof of the fluctuation theorem for MFG

Postponing the proof of the uniqueness Theorem 5.6, we are now in position to the complete the proof of Theorem 3.2.

There are two steps. The first one is to observe that the limiting SPDE (5.21) may be rewritten in terms of  $\mathcal{A}$  in (3.2), simply by expanding the definition of  $\mathcal{G}$ . Hence, the SPDE (5.21) may be easily identified with (3.3).

Then, it remains to show that  $(D_t^n := \bar{S}_t^n - S_t^n)_{t \in [0, T]}$  converges to zero in law on the space  $C([0, T], \mathcal{H}^{-(2+2\lambda_d), \lambda_d})$ . In fact, we will show that  $\mathbb{E}[\sup_{t \in [0, T]} \|D_t^n\|_{-(2+2\lambda_d), \lambda_d}]$  tends to 0 as  $n$  tends to  $\infty$ . Notice first that

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} \|D_t^n\|_{-(2+2\lambda_d), \lambda_d} \right] &= \mathbb{E} \left[ \sup_{t \in [0, T]} \sup_{\|\phi\|_{2+2\lambda_d, \lambda_d} \leq 1} \langle D_t^n, \phi \rangle \right] \\ &\leq \frac{1}{\sqrt{n}} \mathbb{E} \left[ \sup_{t \in [0, T]} \sup_{\|\phi\|_{2+2\lambda_d, \lambda_d} \leq 1} \left| \sum_{i=1}^n (\phi(X_t^i) - \phi(\bar{X}_t^i)) \right| \right]. \end{aligned}$$

For  $\phi$  satisfying  $\|\phi\|_{2+2\lambda_d, \lambda_d} \leq 1$ , we know from Sobolev's embeddings that the function  $\mathbb{R}^d \ni x \mapsto |D_x \phi(x)|/(1 + |x|^{\lambda_d})$  is bounded, independently of  $\phi$ . Therefore, there exists a

constant  $C$ , independent of  $n$ , such that:

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \|D_t^n\|_{-(2+2\lambda_d), \lambda_d} \right] \leq \frac{C}{\sqrt{n}} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \sum_{i=1}^n \left( 1 + |X_t^i|^{\lambda_d} + |\bar{X}_t^i|^{\lambda_d} \right) |X_t^i - \bar{X}_t^i| \right| \right].$$

By the Cauchy-Schwarz inequality, we get:

$$\begin{aligned} & \mathbb{E} \left[ \sup_{t \in [0, T]} \|D_t^n\|_{-(2+2\lambda_d), \lambda_d} \right] \\ & \leq C \sqrt{n} \left( \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \left( 1 + \|X^i\|_{\infty}^{2\lambda_d} + \|\bar{X}^i\|_{\infty}^{2\lambda_d} \right) \right] \right)^{1/2} \left( \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \|X^i - \bar{X}^i\|_{\infty}^2 \right] \right)^{1/2}. \end{aligned}$$

The first expectation is bounded uniformly in  $n$ , and we may use (4.9) to conclude that the right-hand side tends to 0.  $\square$

#### 5.4 Proof of the uniqueness Theorem 5.6

In order to prove uniqueness of the solution to (5.20), we use the same argument as in [29]. To do so, we take two solutions  $S$  and  $S'$  of the limiting equation, with the same initial condition and with square-integrable and continuous paths with values in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  and we consider an orthonormal basis  $(f_k)_{k \geq 1}$  of the space  $\mathcal{H}^{(4+2\lambda_d), \lambda_d}$ . Letting  $(\delta S_t = S_t - S'_t)_{t \in [0, T]}$ , we know that, for any  $k \geq 1$ ,

$$\begin{aligned} d\langle \delta S_t, f_k \rangle &= \left[ \langle \delta S_t, D_x f_k(\cdot) \cdot \tilde{b}(t, \cdot, \nu_t) \rangle + \langle \delta S_t, [\mathcal{G}(t, \nu_t, \nu_t) f_k](\cdot) \rangle \right] dt \\ &\quad + \frac{1}{2} \langle \delta S_t, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 f_k] \rangle dt + \langle \delta S_t, \sigma_0^\top D_x f_k \rangle \cdot d\varpi_t, \quad t \in [0, T], \end{aligned}$$

with  $\delta S_0 = 0$ . Then, recalling that  $\varpi$  has dimension  $d_0$ ,

$$\begin{aligned} d\langle \delta S_t, f_k \rangle^2 &= 2\langle \delta S_t, f_k \rangle \left( \langle \delta S_t, D_x f_k(\cdot) \cdot \tilde{b}(t, \cdot, \nu_t) \rangle + \langle \delta S_t, [\mathcal{G}(t, \nu_t, \nu_t) f_k](\cdot) \rangle \right) dt \\ &\quad + \langle \delta S_t, f_k \rangle \langle \delta S_t, \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2 f_k] \rangle dt + \sum_{i=1}^{d_0} \langle \delta S_t, (\sigma_0^\top D_x f_k)_i \rangle^2 dt \\ &\quad + 2\langle \delta S_t, f_k \rangle \langle \delta S_t, \sigma_0^\top D_x f_k \rangle \cdot d\varpi_t, \quad t \in [0, T]. \end{aligned} \tag{5.22}$$

At this stage, the problem is the same as in [29], namely the right hand side depends on higher derivatives of the basis function  $f_k$ . To handle these higher derivatives, we proceed as follows.

**Step 1:** As a first step, we tackle the second order term in (5.22). Defining the operator  $\mathcal{L}_0 = \text{Tr}[(\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x^2(\cdot)]$  and summing over  $k \geq 1$ , we get

$$\sum_{k \geq 1} \langle \delta S_t, f_k \rangle \langle \delta S_t, \mathcal{L}_0 f_k \rangle = \langle \delta S_t, \mathcal{L}_0^* \delta S_t \rangle_{-(4+2\lambda_d), \lambda_d},$$

where  $\langle \cdot, \cdot \rangle_{-(4+2\lambda_d), \lambda_d}$  stands for the inner product in  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  and  $\mathcal{L}_0^* \delta S_t$  is the element of  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  defined as  $\langle \mathcal{L}_0^* \delta S_t, \phi \rangle = \langle \delta S_t, \mathcal{L}_0 \phi \rangle$  for all  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$  (which makes sense since  $\delta S_t \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$ ). We then make use of Riesz' representation theorem. For an element  $\varrho$  in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d} \subset \mathcal{H}^{-(4+2\lambda_d), \lambda_d}$ , we call  $\tilde{\varrho}$  its representative element in  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$ , namely

$$\forall \phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}, \quad \langle \phi, \varrho \rangle = \langle \phi, \tilde{\varrho} \rangle_{4+2\lambda_d, \lambda_d},$$

where  $\langle \cdot, \cdot \rangle$  denotes the duality bracket while  $\langle \cdot, \cdot \rangle_{4+2\lambda_d, \lambda_d}$  stands for the inner product in  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$ . We then take an element  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  such that  $\tilde{\varrho}$  belongs to  $\mathcal{H}^{6+2\lambda_d, \lambda_d}$ . Since  $\mathcal{L}_0^* \varrho \in \mathcal{H}^{-(4+2\lambda_d), \lambda_d}$ , we get

$$\begin{aligned} & \langle \varrho, \mathcal{L}_0^* \varrho \rangle_{-(4+2\lambda_d), \lambda_d} \\ &= \langle \tilde{\varrho}, \mathcal{L}_0^* \varrho \rangle = \langle \mathcal{L}_0 \tilde{\varrho}, \varrho \rangle = \langle \mathcal{L}_0 \tilde{\varrho}, \tilde{\varrho} \rangle_{4+2\lambda_d, \lambda_d} \\ &= \sum_{i,j=1}^d (\sigma \sigma^\top + \sigma_0 \sigma_0^\top)_{i,j} \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} \frac{D^{\mathbf{k}} \tilde{\varrho}(x) D_{x_i, x_j}^2 D^{\mathbf{k}} \tilde{\varrho}(x)}{1 + |x|^{2\lambda_d}} dx \\ &= - \sum_{i,j=1}^d (\sigma \sigma^\top + \sigma_0 \sigma_0^\top)_{i,j} \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} \frac{D_{x_i} D^{\mathbf{k}} \tilde{\varrho}(x) D_{x_j} D^{\mathbf{k}} \tilde{\varrho}(x)}{1 + |x|^{2\lambda_d}} dx \\ &\quad + \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^\top + \sigma_0 \sigma_0^\top)_{i,j} \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} D_{x_i, x_j}^2 ((1 + |x|^{2\lambda_d})^{-1}) (D^{\mathbf{k}} \tilde{\varrho})^2(x) dx \\ &= -(\|\sigma^\top D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2 + \|\sigma_0^\top D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2) + O(\|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2), \end{aligned}$$

where  $\|\sigma^\top D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2$  is a shortened notation for  $\sum_{i=1}^d \|(\sigma^\top D_x \tilde{\varrho})_i\|_{4+2\lambda_d, \lambda_d}^2$  (and similarly for the term  $\|\sigma_0^\top D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2$ ) and  $O(\cdot)$  is the Landau notation.

**Step 2:** We now handle the quadratic term in (5.22). Proceeding as before, we have

$$\sum_{k \geq 1} \sum_{i=1}^{d_0} \langle \delta S_t, (\sigma_0^\top D_x f_k)_i \rangle^2 = \sum_{i=1}^{d_0} \|(\sigma_0^\top D_x)_i^* \delta S_t\|_{-(4+2\lambda_d), \lambda_d}^2,$$

where  $\langle (\sigma_0^\top D_x)_i^* \delta S_t, \phi \rangle = \langle \delta S_t, (\sigma_0^\top D_x)_i \phi \rangle$  for all  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$ .

Choose now  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  such that  $\tilde{\varrho} \in \mathcal{H}^{6+2\lambda_d, \lambda_d}$ . Then, for any smooth function  $\phi$  with compact support,

$$\begin{aligned} \langle (\sigma_0^\top D_x)_i^* \varrho, \phi \rangle &= \langle \tilde{\varrho}, (\sigma_0^\top D_x)_i \phi \rangle_{4+2\lambda_d, \lambda_d} \\ &= \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} \frac{D^{\mathbf{k}} \tilde{\varrho}(x) (\sigma_0^\top D_x)_i D^{\mathbf{k}} \phi(x)}{1 + |x|^{2\lambda_d}} dx \\ &= - \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} \frac{D^{\mathbf{k}} [(\sigma_0^\top D_x)_i \tilde{\varrho}](x) D^{\mathbf{k}} \phi(x)}{1 + |x|^{2\lambda_d}} dx \\ &\quad - \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} (\sigma_0^\top D_x)_i ((1 + |x|^{2\lambda_d})^{-1}) D^{\mathbf{k}} \tilde{\varrho}(x) D^{\mathbf{k}} \phi(x) dx. \end{aligned}$$

Therefore,

$$\langle (\sigma_0^\top D_x)_i^* \varrho, \phi \rangle = -\langle (\sigma_0^\top D_x)_i \tilde{\varrho}, \phi \rangle_{4+2\lambda_d, \lambda_d} + O(\|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} \|\phi\|_{4+2\lambda_d, \lambda_d}),$$

which shows that  $(\sigma_0^\top D_x)_i^* \varrho$  belongs to  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  and that the above equality holds for any  $\phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}$ . Hence,

$$\begin{aligned} & \|(\sigma_0^\top D_x)_i^* \varrho\|_{-(4+2\lambda_d), \lambda_d} \\ &= \sup_{\|\phi\|_{4+2\lambda_d, \lambda_d} \leq 1} \left[ -\langle (\sigma_0^\top D_x)_i \tilde{\varrho}, \phi \rangle_{4+2\lambda_d, \lambda_d} + O(\|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} \|\phi\|_{4+2\lambda_d, \lambda_d}) \right] \\ &= \|(\sigma_0^\top D_x)_i \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} + O(\|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}). \end{aligned} \tag{5.23}$$

**Step 3:** Lastly, we handle the first term in the right-hand side of (5.22). We have

$$\begin{aligned} \sum_{k \geq 1} \langle \delta S_t, f_k \rangle \langle \delta S_t, D_x f_k(\cdot) \cdot \tilde{b}(t, \cdot, \nu_t) \rangle &= \sum_{i=1}^d \sum_{k \geq 1} \langle \delta S_t, f_k \rangle \langle D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \delta S_t), f_k \rangle \\ &= \sum_{i=1}^d \langle \delta S_t, D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \delta S_t) \rangle, \end{aligned}$$

where, for  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$ ,  $\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho$  is the element of  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  defined by

$$\langle \Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho, \phi \rangle = \langle \varrho, \tilde{b}_i(t, \cdot, \nu_t) \phi \rangle, \quad \text{for } \phi \in \mathcal{H}^{4+2\lambda_d, \lambda_d}.$$

As above, we investigate the above term when  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  satisfies  $\tilde{\varrho} \in \mathcal{H}^{6+2\lambda_d, \lambda_d}$ . Since the derivatives of  $\tilde{b}_i$  are bounded, it is easily verified that  $\tilde{b}_i(t, \cdot, \nu_t) \phi$  belongs to  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$  which guarantees that the above right-hand side is well defined. In fact, when  $\phi \in \mathcal{H}^{5+2\lambda_d, \lambda_d}$ ,

$$\langle \Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho, D_{x_i} \phi \rangle = \langle \tilde{\varrho}, \tilde{b}_i(t, \cdot, \nu_t) D_{x_i} \phi \rangle_{4+2\lambda_d, \lambda_d}.$$

Since the derivatives of  $\tilde{b}_i$  up to the order  $5 + 2\lambda_d$  are bounded, the right-hand side is bounded by  $O(\|D_{x_i} \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} + \|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}) \|\phi\|_{4+2\lambda_d, \lambda_d}$ . In particular,  $D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho)$  belongs to the space  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  and

$$\|D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho)\|_{-(4+2\lambda_d), \lambda_d} = O(\|D_{x_i} \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} + \|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}).$$

Then,

$$\left| \langle \varrho, D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho) \rangle_{\mathcal{H}^{-(4+2\lambda_d), 2\lambda_d}} \right| = O(\|D_{x_i} \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} \|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} + \|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2). \quad (5.24)$$

**Step 4:** We next consolidate the first three steps into an estimate on the  $dt$  term of the right-hand side of (5.22), or rather the sum thereof over  $k \in \mathbb{N}$ , which can be rewritten as  $\mathcal{F}(\delta S_t)$ , where

$$\begin{aligned} \mathcal{F}(\varrho) &= \left[ 2 \sum_{i=1}^d \langle \varrho, D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho) \rangle_{-(4+2\lambda_d), \lambda_d} \right. \\ &\quad \left. + 2 \langle \varrho, \mathcal{G}(t, \nu_t, \nu_t)^* \varrho \rangle_{-(4+2\lambda_d), \lambda_d} + \langle \varrho, \mathcal{L}_0^* \varrho \rangle_{-(4+2\lambda_d), \lambda_d} \right] + \sum_{i=1}^{d_0} \|(\sigma_0^\top D_x)_i^* \varrho\|_{-(4+2\lambda_d), \lambda_d}^2. \end{aligned}$$

Here, following the proof of (5.13),  $\mathcal{G}(t, \nu_t, \nu_t)^* \varrho$  belongs to the space  $\mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  and  $\|\mathcal{G}(t, \nu_t, \nu_t)^* \varrho\|_{-(4+2\lambda_d), \lambda_d} \leq c(1 + \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\nu_t(y)) \|\varrho\|_{-(4+2\lambda_d), \lambda_d}$ . Returning to (5.22) and collecting (5.23) and (5.24), the second point is that  $\mathcal{F}(\varrho)$  satisfies, for  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  such that  $\tilde{\varrho} \in \mathcal{H}^{6+2\lambda_d, \lambda_d}$ ,

$$\begin{aligned} \mathcal{F}(\varrho) &= -\|\sigma^\top D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2 + O(\|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} \|D_x \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d} + \|\tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2) \\ &\quad + 2 \langle \varrho, \mathcal{G}(s, \nu_s, \nu_s)^* \varrho \rangle_{-(4+2\lambda_d), \lambda_d}. \end{aligned}$$

Using the fact that  $\sigma$  is non-degenerate together with the corresponding form of (5.13), this yields

$$\mathcal{F}(\varrho) \leq C \|\varrho\|_{-(4+2\lambda_d), \lambda_d}^2 + C \left[ \|\varrho\|_{-(4+2\lambda_d), \lambda_d}^2 \left( \int_{\mathbb{R}^d} (1 + |y|^{2\lambda_d}) d\nu_s(y) \right)^{1/2} \right]. \quad (5.25)$$

**Step 5:** We now consider the general case when  $\varrho \in \mathcal{H}^{-(2+2\lambda_d), \lambda_d} \subset \mathcal{H}^{-(4+2\lambda_d), \lambda_d}$  and  $\tilde{\varrho}$  just belongs to  $\mathcal{H}^{4+2\lambda_d, \lambda_d}$ . Let  $(\eta_n)_{n \geq 1}$  be a sequence of even mollifiers, and let  $(\varrho_n)_{n \geq 1}$  be the sequence of elements in  $\mathcal{H}^{-(2+2\lambda_d), \lambda_d}$  defined by

$$\langle \varrho_n, \phi \rangle = \langle \varrho, \phi * \eta_n \rangle,$$

where  $*$  denotes the usual convolution product. Let  $\widehat{\varrho}$  and  $\widehat{\varrho}_n$  denote the elements of  $\mathcal{H}^{2+2\lambda_d, \lambda_d}$  such that, for all  $\phi \in \mathcal{H}^{2+2\lambda_d, \lambda_d}$ ,

$$\langle \widehat{\varrho}, \phi \rangle_{\mathcal{H}^{2+2\lambda_d, \lambda_d}} = \langle \varrho, \phi \rangle, \quad \langle \widehat{\varrho}_n, \phi \rangle_{2+2\lambda_d, \lambda_d} = \langle \varrho_n, \phi \rangle.$$

Then,

$$\langle \rho_n, \phi \rangle = \langle \varrho, \phi * \eta_n \rangle = \langle \widehat{\varrho}, \phi * \eta_n \rangle_{2+2\lambda_d, \lambda_d}.$$

We observe that

$$\begin{aligned} \|\phi * \eta_n\|_{2+2\lambda_d, \lambda_d}^2 &= \sum_{|\mathbf{k}| \leq 2+2\lambda_d} \int_{\mathbb{R}^d} \frac{|D^{\mathbf{k}}(\phi * \eta_n)(x)|^2}{1 + |x|^{2\lambda_d}} dx \\ &= \sum_{|\mathbf{k}| \leq 2+2\lambda_d} \int_{\mathbb{R}^d} \frac{|(D^{\mathbf{k}}\phi * \eta_n)(x)|^2}{1 + |x|^{2\lambda_d}} dx \\ &\leq \sum_{|\mathbf{k}| \leq 2+2\lambda_d} \int_{\mathbb{R}^d} \frac{1}{1 + |x|^{2\lambda_d}} \left( \int_{\mathbb{R}^d} |D^{\mathbf{k}}\phi(y)|^2 \eta_n(x-y) dy \right) dx \\ &= \sum_{|\mathbf{k}| \leq 2+2\lambda_d} \int_{\mathbb{R}^d} \frac{|D^{\mathbf{k}}\phi(y)|^2}{1 + |y|^{2\lambda_d}} \left( \int_{\mathbb{R}^d} \frac{\eta_n(x-y)(1 + |y|^{2\lambda_d})}{1 + |x|^{2\lambda_d}} dx \right) dy \\ &\leq C \|\phi\|_{\mathcal{H}^{2+2\lambda_d, \lambda_d}}^2, \end{aligned}$$

for a constant  $C$  independent of  $n$ . For a given  $\epsilon > 0$ , we choose  $\tilde{\varrho}'$  in  $C_c^\infty(\mathbb{R}^d)$  such that  $\|\tilde{\varrho}' - \widehat{\varrho}\|_{\mathcal{H}^{2+2\lambda_d, \lambda_d}} \leq \epsilon$  and we get

$$|\langle \tilde{\varrho}', \phi * \eta_n \rangle_{2+2\lambda_d, \lambda_d} - \langle \widehat{\varrho}, \phi * \eta_n \rangle_{2+2\lambda_d, \lambda_d}| \leq C\epsilon \|\phi\|_{\mathcal{H}^{2+2\lambda_d, \lambda_d}}. \quad (5.26)$$

Now,

$$\begin{aligned} &\langle \tilde{\varrho}', \phi * \eta_n - \phi \rangle \\ &= \sum_{|\mathbf{k}| \leq 4+2\lambda_d} \int_{\mathbb{R}^d} \frac{1}{1 + |y|^{2\lambda_d}} D^{\mathbf{k}}\phi(y) \left( \int_{\mathbb{R}^d} \eta_n(y-x) \left[ \frac{1 + |y|^{2\lambda_d}}{1 + |x|^{2\lambda_d}} D^{\mathbf{k}}\tilde{\varrho}'(x) - D^{\mathbf{k}}\tilde{\varrho}'(y) \right] dx \right) dy. \end{aligned}$$

Since  $\tilde{\varrho}'$  has a compact support, we deduce that there exists a constant  $C_\epsilon$  such that

$$|\langle \tilde{\varrho}', \phi * \eta_n - \phi \rangle| \leq \frac{C_\epsilon}{n} \|\phi\|_{2+2\lambda_d, \lambda_d}. \quad (5.27)$$

By (5.26) and (5.27), we deduce that  $\|\varrho_n - \varrho\|_{-(2+2\lambda_d), \lambda_d}^2 \rightarrow 0$  as  $n \rightarrow \infty$ . As a consequence, we have  $\|\tilde{\varrho}_n - \tilde{\varrho}\|_{4+2\lambda_d, \lambda_d}^2 \rightarrow 0$  as  $n \rightarrow \infty$ . And then,  $\|D_{x_i, x_j}^2 \tilde{\varrho}_n - D_{x_i, x_j}^2 \tilde{\varrho}\|_{2+2\lambda_d, \lambda_d}^2 \rightarrow 0$  as  $n \rightarrow \infty$ , for all  $i, j \in \{1, \dots, d\}$ . Also,  $\|D_{x_i}^* \varrho_n - D_{x_i}^* \varrho\|_{-(4+2\lambda_d), \lambda_d}^2 \rightarrow 0$  and  $\|D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho_n) - D_{x_i}^* (\Pi_{\tilde{b}_i(t, \cdot, \nu_t)} \varrho)\|_{-(4+2\lambda_d), \lambda_d}^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

When  $\varrho$  is replaced by  $\varrho_n$ , (5.25) holds true. Passing to the limit over  $n \rightarrow \infty$ , we deduce that (5.25) remains true for a general  $\varrho$ .

**Conclusion of proof:** We finally return to (5.22), using (5.25) to get

$$\begin{aligned} d[\|\delta S_t\|_{-(4+2\lambda_d), \lambda_d}^2] &\leq C \|\delta S_t\|_{-(4+2\lambda_d), \lambda_d}^2 \left[ 1 + \left( \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\nu_t(y) \right)^{1/2} \right] dt \\ &\quad + \sum_{i=1}^d \langle \delta S_t, (\sigma_0^\top D_x)_i^* \delta S_t \rangle_{-(4+2\lambda_d), \lambda_d} \cdot d\varpi_t^i. \end{aligned}$$

By expanding  $d[\|\delta S_t\|_{-(4+2\lambda_d), \lambda_d}^2 \exp(-C(t + \int_0^t (\int_{\mathbb{R}^d} |y|^{2\lambda_d} d\nu_s(y))^{1/2} ds))]$ , we get

$$\mathbb{E} \left[ \exp \left( -C \left( t + \int_0^t \left( \int_{\mathbb{R}^d} |y|^{2\lambda_d} d\nu_s(y) \right)^{1/2} ds \right) \right) \|\delta S_t\|_{-(4+2\lambda_d), \lambda_d}^2 \right] = 0, \quad t \in [0, T].$$

□

### 5.5 Auxiliary lemma: Rates of convergence for smooth functions of an empirical sample

We prove here the remarkable fact that the rate of convergence of the empirical distribution of a sample of i.i.d. random vectors with values  $\mathbb{R}^d$  can be made dimension-free when it is tested against test functions that are sufficiently smooth. This is in contrast with existing results in the literature that address the rate of convergence in the  $p$ -Wasserstein distance, such as [20]. It is worth mentioning also that the results of Kolokoltsov et al. [26] are substantially different: Therein, the authors state dimension-free  $1/n$  convergence rates in a MFG context, in a norm involving twice differentiable test functions on a space of measures, but only by imposing as an assumption that the empirical measure of initial states converges at rate  $1/n$  in norm in the dual  $(C_b^2(\mathbb{R}^d))^*$  of the space of bounded twice differentiable functions. So, it is a different form of convergence than what we consider below. Also, it looks to be fitted to other settings than the i.i.d. one.

**Lemma 5.10.** *Assume that  $\phi$  is a  $\mathcal{C}^2$  function on  $\mathcal{P}^4(\mathbb{R}^d)$ , as defined in Section 2.2, with bounded derivatives  $\delta\phi/\delta m$  and  $\delta^2\phi/\delta m^2$ , and that  $\delta^2\phi/\delta m^2$  is Lipschitz continuous in  $m$  with respect to  $\mathcal{W}_1$ , uniformly in the other variables. Then, there exists a constant  $c$ , depending on the bounds on the first and second-order derivatives and on the Lipschitz constant of the second-order derivative such that, for any  $m_0 \in \mathcal{P}^4(\mathbb{R}^d)$  and any i.i.d. sample  $\zeta = (\zeta_1, \dots, \zeta_n)$  from  $m_0$ ,*

$$\mathbb{E} \left[ |\phi(m_\zeta^n) - \phi(m_0)|^4 \right]^{1/4} \leq c \left( 1 + \left( \int_{\mathbb{R}^d} |x| m_0(dx) \right)^{1/4} \right) n^{-1/2}.$$

*Proof.* Since  $\phi$  is differentiable, we have the following expansion:

$$\begin{aligned} \phi(m_\zeta^n) - \phi(m_0) &= \int_0^1 d\lambda \int_{\mathbb{R}^d} \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, v) d(m_\zeta^n - m_0)(v) \\ &= S_1 + S_2, \end{aligned}$$

where

$$\begin{aligned} S_1 &= \int_{\mathbb{R}^d} \frac{\delta\phi}{\delta m}(m_0, v) d(m_\zeta^n - m_0)(v), \\ S_2 &= \int_0^1 d\lambda \int_{\mathbb{R}^d} \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, v) - \frac{\delta\phi}{\delta m}(m_0, v) \right] d(m_\zeta^n - m_0)(v). \end{aligned}$$

**Step 1:** We first deal with the term  $S_1$ . This may be rewritten as

$$S_1 = \frac{1}{n} \sum_{i=1}^n \left( \frac{\delta\phi}{\delta m}(m_0, \zeta_i) - \mathbb{E} \left[ \frac{\delta\phi}{\delta m}(m_0, \zeta_1) \right] \right).$$

Since  $\delta\phi/\delta m$  is bounded, we get  $\mathbb{E}[|S_1|^4]^{1/4} \leq cn^{-1/2}$ , for a constant  $c < \infty$  independent of  $n$ . The value of  $c$  is allowed to vary from line to line.

**Step 2:** Now, we turn to the term  $S_2$ , which we first rewrite in the form

$$S_2 = \frac{1}{n} \sum_{i=1}^n \int_0^1 \varphi_\lambda^i d\lambda,$$

where we define, for  $i = 1, \dots, n$  and  $\lambda \in [0, 1]$ ,

$$\begin{aligned} \varphi_\lambda^i := & \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(m_0, \zeta_i) \right] \\ & - \tilde{\mathbb{E}} \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, \tilde{\zeta}) - \frac{\delta\phi}{\delta m}(m_0, \tilde{\zeta}) \right], \end{aligned} \quad (5.28)$$

and where the expectation  $\tilde{\mathbb{E}}$  is taken over the random variable  $\tilde{\zeta}$ , which has the same distribution as  $\zeta_1$  and which is independent of  $\zeta$ . Now, we have

$$|S_2|^4 \leq n^{-4} \int_0^1 \left( \sum_{i=1}^n \varphi_\lambda^i \right)^4 d\lambda,$$

so we focus on estimating the integrand for a fixed  $\lambda$ . We have

$$\begin{aligned} \left( \sum_{i=1}^n \varphi_\lambda^i \right)^4 = & \sum_{i=1}^n (\varphi_\lambda^i)^4 + 3 \sum_{i \neq j} (\varphi_\lambda^i)^2 (\varphi_\lambda^j)^2 + 4 \sum_{i \neq j} \varphi_\lambda^i (\varphi_\lambda^j)^3 \\ & + 6 \sum_{i_1, i_2, i_3 \text{ distinct}} \varphi_\lambda^{i_1} \varphi_\lambda^{i_2} (\varphi_\lambda^{i_3})^2 + \sum_{i_1, i_2, i_3, i_4 \text{ distinct}} \varphi_\lambda^{i_1} \varphi_\lambda^{i_2} \varphi_\lambda^{i_3} \varphi_\lambda^{i_4}. \end{aligned} \quad (5.29)$$

Noting that  $\varphi_\lambda^i$  is bounded, uniformly in  $i$  and  $\lambda$ , we bound the first three terms on the right-hand side by  $cn^2$ . The only difficulty is to upper bound the last two terms.

**Step 3:** We start with the sum over  $i_1, i_2, i_3$  distinct. We let, for any  $(i_1, i_2, i_3) \in \{1, \dots, n\}^3$ ,  $m_\zeta^{n, -(i_1, i_2, i_3)} = \frac{1}{n-3} \sum_{\ell \neq i_1, i_2, i_3} \delta_{\zeta_\ell}$ . We then observe that

$$\begin{aligned} & \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(m_0, \zeta_i) \right] \\ & - \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^{n, -(i_1, i_2, i_3)} + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(m_0, \zeta_i) \right] \\ = & \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(\lambda m_\zeta^{n, -(i_1, i_2, i_3)} + (1-\lambda)m_0, \zeta_i) \right] \\ = & \int_0^1 ds \int_{\mathbb{R}^d} \left[ \frac{\delta^2\phi}{\delta m^2} \left( s\lambda m_\zeta^n + (1-s)\lambda m_\zeta^{n, -(i_1, i_2, i_3)} + (1-\lambda)m_0, \zeta_i, v \right) \right. \\ & \left. d(m_\zeta^n - m_\zeta^{n, -(i_1, i_2, i_3)})(v) \right]. \end{aligned}$$

Now,

$$m_\zeta^n - m_\zeta^{n, -(i_1, i_2, i_3)} = \frac{1}{n} \sum_{j=i_1, i_2, i_3} \delta_{\zeta_j} - \frac{3}{n(n-3)} \sum_{j \neq i_1, i_2, i_3} \delta_{\zeta_j}, \quad (5.30)$$

and then, since  $\delta^2\phi/\delta m^2$  is bounded, we get

$$\begin{aligned} & \left| \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^n + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(m_0, \zeta_i) \right] \right. \\ & \left. - \left[ \frac{\delta\phi}{\delta m}(\lambda m_\zeta^{n, -(i_1, i_2, i_3)} + (1-\lambda)m_0, \zeta_i) - \frac{\delta\phi}{\delta m}(m_0, \zeta_i) \right] \right| \leq \frac{c}{n}. \end{aligned}$$

Proceeding in the same way for the term containing  $\tilde{\zeta}$  in (5.28), we get

$$|\varphi_\lambda^i - \varphi_\lambda^{i, -(i_1, i_2, i_3)}| \leq \frac{c}{n},$$



where

$$\begin{aligned} \varphi_{\lambda}^{i, -(i_1, i_2, i_3)} &= \left[ \frac{\delta \phi}{\delta m} (\lambda m_{\zeta}^{n, -(i_1, i_2, i_3)} + (1 - \lambda) m_0, \zeta_i) - \frac{\delta \phi}{\delta m} (m_0, \zeta_i) \right] \\ &\quad - \tilde{\mathbb{E}} \left[ \left[ \frac{\delta \phi}{\delta m} (\lambda m_{\zeta}^{n, -(i_1, i_2, i_3)} + (1 - \lambda) m_0, \tilde{\zeta}) - \frac{\delta \phi}{\delta m} (m_0, \tilde{\zeta}) \right] \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E} \left[ \sum_{i_1, i_2, i_3 \text{ distinct}} \varphi_{\lambda}^{i_1} \varphi_{\lambda}^{i_2} (\varphi_{\lambda}^{i_3})^2 \right] \\ \leq cn^2 + \mathbb{E} \left[ \sum_{i_1, i_2, i_3 \text{ distinct}} \varphi_{\lambda}^{i_1, -(i_1, i_2, i_3)} \varphi_{\lambda}^{i_2, -(i_1, i_2, i_3)} (\varphi_{\lambda}^{i_3, -(i_1, i_2, i_3)})^2 \right], \end{aligned}$$

and it is clear, by a conditional independence argument, that the last sum is in fact zero. This shows that the left-hand side is less than  $cn^2$ .

**Step 4:** It now remains to study the sum over distinct  $i_1, \dots, i_4$  in (5.29). To do so, we may define in a similar manner  $\varphi_{\lambda}^{i, -(i_1, \dots, i_4)}$  for  $(i_1, \dots, i_4) \in \{1, \dots, n\}^4$ . Notice first that, for any  $(i_1, \dots, i_4)$ ,

$$\begin{aligned} \varphi_{\lambda}^{i_1} \varphi_{\lambda}^{i_2} \varphi_{\lambda}^{i_3} \varphi_{\lambda}^{i_4} &= \varphi_{\lambda}^{i_1, -(i_1, \dots, i_4)} \varphi_{\lambda}^{i_2, -(i_1, \dots, i_4)} \varphi_{\lambda}^{i_3, -(i_1, \dots, i_4)} \varphi_{\lambda}^{i_4, -(i_1, \dots, i_4)} \\ &\quad + \sum_{j=1}^4 (\varphi_{\lambda}^{i_j} - \varphi_{\lambda}^{i_j, -(i_1, \dots, i_4)}) \prod_{\substack{k=1 \\ k \neq j}}^4 \varphi_{\lambda}^{i_k, -(i_1, \dots, i_4)} + S_3, \end{aligned} \quad (5.31)$$

where  $S_3$  is a sum of terms that involve a product of four terms, at least two of which are of the form  $(\varphi_{\lambda}^{i_j} - \varphi_{\lambda}^{i_j, -(i_1, \dots, i_4)})$  for some  $j$ , and the rest are of the form  $\varphi_{\lambda}^{i_j}$  or  $\varphi_{\lambda}^{i_j, -(i_1, \dots, i_4)}$ . But since  $|\varphi_{\lambda}^{i_j} - \varphi_{\lambda}^{i_j, -(i_1, \dots, i_4)}| \leq c/n$ , and since the first term on the right-hand side of (5.31) has mean zero, we find

$$\begin{aligned} \mathbb{E} \left[ \sum_{i_1, i_2, i_3, i_4 \text{ distinct}} \varphi_{\lambda}^{i_1} \varphi_{\lambda}^{i_2} \varphi_{\lambda}^{i_3} \varphi_{\lambda}^{i_4} \right] \\ \leq cn^2 + \mathbb{E} \left[ \sum_{i_1, \dots, i_4} \sum_{\text{distinct } j=1}^4 (\varphi_{\lambda}^{i_j} - \varphi_{\lambda}^{i_j, -(i_1, \dots, i_4)}) \prod_{\substack{k=1 \\ k \neq j}}^4 \varphi_{\lambda}^{i_k, -(i_1, \dots, i_4)} \right]. \end{aligned} \quad (5.32)$$

Returning to the proof of the third step but using in addition the  $\mathcal{W}_1$ -Lipschitz property of  $\delta^2 \phi / \delta m^2$  with respect to the measure argument, we can write:

$$\begin{aligned} \varphi_{\lambda}^i - \varphi_{\lambda}^{i, -(i_1, \dots, i_4)} &= \int_0^1 ds \int_{\mathbb{R}^d} \frac{\delta^2 \phi}{\delta m^2} (\lambda m_{\zeta}^{n, -(i_1, \dots, i_4)} + (1 - \lambda) m_0, \zeta_i, v) d(m_{\zeta}^n - m_{\zeta}^{n, -(i_1, \dots, i_4)})(v) \\ &\quad - \tilde{\mathbb{E}} \int_0^1 ds \int_{\mathbb{R}^d} \frac{\delta^2 \phi}{\delta m^2} (\lambda m_{\zeta}^{n, -(i_1, \dots, i_4)} + (1 - \lambda) m_0, \tilde{\zeta}, v) d(m_{\zeta}^n - m_{\zeta}^{n, -(i_1, \dots, i_4)})(v) \\ &\quad + \varepsilon_n^{i_1, \dots, i_4}, \end{aligned} \quad (5.33)$$

where

$$|\varepsilon_n^{i_1, \dots, i_4}| \leq \frac{c}{n^2} (|\zeta_{i_1}| + \dots + |\zeta_{i_4}|) + \frac{c}{n^3} \sum_{j=1}^n |\zeta_j|.$$

Following (5.30), we can write the difference  $\varphi_\lambda^i - \varphi_\lambda^{i, -(i_1, \dots, i_4)}$  in the form

$$\varphi_\lambda^i - \varphi_\lambda^{i, -(i_1, \dots, i_4)} = \frac{1}{n} \sum_{j=1}^4 F(m_\zeta^{n, -(i_1, \dots, i_4)}, \zeta_i, \zeta_{i_j}) + \frac{1}{n} G(m_\zeta^{n, -(i_1, \dots, i_4)}, \zeta_i) + \varepsilon_n^{i_1, \dots, i_4},$$

for bounded measurable functions  $F$  and  $G$ . Now note that  $F(m_\zeta^{n, -(i_1, \dots, i_4)}, \zeta_{i_k}, \zeta_{i_j})$  is measurable with respect to  $(\zeta_\ell : \ell \in \{i_1, \dots, i_4\}^c \cup \{i_k, i_j\})$ , for  $k, j \in \{1, \dots, 4\}$ , and similarly for the  $G$  term. On the other hand, for  $k = 1, \dots, 4$ ,

$$\mathbb{E}[\varphi_\lambda^{i_k, -(i_1, \dots, i_4)} | (\zeta_\ell : \ell \neq i_k)] = 0.$$

Hence, for each  $j \in \{1, \dots, 4\}$  and each  $i_1, \dots, i_4$  distinct, we find that the corresponding term in (5.32) simplifies in the sense that

$$\mathbb{E} \left[ \sum_{j=1}^4 (\varphi_\lambda^{i_j} - \varphi_\lambda^{i_j, -(i_1, \dots, i_4)} - \varepsilon_n^{i_1, \dots, i_4}) \prod_{\substack{k=1 \\ k \neq j}}^4 \varphi_\lambda^{i_k, -(i_1, \dots, i_4)} \right] = 0.$$

This finally lets us bound the  $(i_1, \dots, i_4)$ -term in (5.32) by  $c\mathbb{E}[|\varepsilon_n^{i_1, \dots, i_4}|] \leq cn^{-2}\mathbb{E}[|\zeta_1|]$ , which completes the proof.  $\square$

## 6 A linear-quadratic example

This section discusses an explicitly solvable model that does not satisfy Assumption A. Nonetheless, we show that our strategy for deriving the central limit theorem by comparison with a more classical McKean-Vlasov system is still successful.

### 6.1 Description of the model

Consider the mean field game model of systemic risk proposed in [10]. Here,  $d = 1$ ,  $\sigma$  and  $\sigma_0$  are positive constants, the action space is  $A = \mathbb{R}$ , and for some  $\bar{g}, \epsilon, \bar{b} > 0$  and  $0 \leq q^2 \leq \epsilon$  we have

$$\begin{aligned} b(x, m, a) &= \bar{b}(\bar{m} - x) + a, \\ f(x, m, a) &= \frac{1}{2}a^2 - qa(\bar{m} - x) + \frac{\epsilon}{2}(\bar{m} - x)^2, \\ g(x, m) &= \frac{\bar{g}}{2}(\bar{m} - x)^2, \end{aligned}$$

where  $\bar{m} = \int_{\mathbb{R}} y dm(y)$ . Both the drift and cost functions induce a herding behavior toward the population average; see [10] for a thorough discussion.

It was shown in [10, (3.24)] that the unique closed loop Nash equilibrium dynamics is given by:

$$\alpha_t^i = \left[ q + \varphi_t^n \left( 1 - \frac{1}{n} \right) \right] (\bar{X}_t - X_t^i), \quad t \in [0, T], \quad (6.1)$$

where  $\bar{X}_t = \frac{1}{n} \sum_{i=1}^n X_t^i$ , and where  $\varphi^n$  is the unique solution to the Riccati equation:

$$\dot{\varphi}_t^n = 2(\bar{b} + q)\varphi_t^n + \left( 1 - \frac{1}{n^2} \right) |\varphi_t^n|^2 - (\epsilon - q^2), \quad \varphi_T^n = \bar{g}.$$

The explicit solution takes the form

$$\varphi_t^n = \frac{-(\epsilon - q^2) \left( e^{(\delta_n^+ - \delta_n^-)(T-t)} - 1 \right) - \bar{g} \left( \delta_n^+ e^{(\delta_n^+ - \delta_n^-)(T-t)} - \delta_n^- \right)}{\left( \delta_n^- e^{(\delta_n^+ - \delta_n^-)(T-t)} - \delta_n^+ \right) - \bar{g} \left( 1 - \frac{1}{n^2} \right) \left( e^{(\delta_n^+ - \delta_n^-)(T-t)} - 1 \right)}, \quad (6.2)$$

where

$$\delta_n^\pm = -(\bar{b} + q) \pm \sqrt{(\bar{b} + q)^2 + \left(1 - \frac{1}{n^2}\right)(\epsilon - q^2)}. \quad (6.3)$$

In particular, the Nash equilibrium state process is given by the solution  $\mathbf{X} = (X^1, \dots, X^n)$  of the SDE system:

$$dX_t^i = \left(\bar{b} + q + \varphi_t^n \left(1 - \frac{1}{n}\right)\right)(\bar{X}_t - X_t^i)dt + \sigma dB_t^i + \sigma_0 dW_t, \quad t \in [0, T]. \quad (6.4)$$

It is straightforward to show that  $\varphi_t^n \rightarrow \varphi_t^\infty$  as  $n \rightarrow \infty$ , uniformly in  $t \in [0, T]$ , where  $\varphi^\infty$  is the unique solution to the Riccati equation

$$\dot{\varphi}_t^\infty = 2(\bar{b} + q)\varphi_t^\infty + |\varphi_t^\infty|^2 - (\epsilon - q^2), \quad \varphi_T^\infty = \bar{g}.$$

The explicit solution is of the same form given by (6.2) and (6.3), with  $n = \infty$ . It follows that  $\mathbf{X} = (X^1, \dots, X^n)$  should be “close” in some sense to the solution  $\mathbf{Y} = (Y^1, \dots, Y^n)$  of the auxiliary SDE system:

$$dY_t^i = (\bar{b} + q + \varphi_t^\infty)(\bar{Y}_t - Y_t^i)dt + \sigma dB_t^i + \sigma_0 dW_t, \quad (6.5)$$

initialized at the same points  $Y_0^i = X_0^i$ . Of course, it should be noted that the process  $\mathbf{Y}$  plays here the same role as the process  $\bar{\mathbf{X}}$  in (4.1), the solution  $U$  to the master equation being given in the current framework by:

$$U(t, x, m) = \frac{\varphi_t^\infty}{2}(\bar{m} - x)^2.$$

In this regard, the fact that  $\mathbf{X}$  and  $\mathbf{Y}$  should be “close” is completely analogous to the statements of Theorems 4.2 and 4.7.

In fact, there are two ways to compare  $\mathbf{X}$  and  $\mathbf{Y}$ . One is to use the explicit expressions for their solutions. Another strategy, which we prefer, is to compare (6.4) and (6.5) and use the fact that  $\mathbf{X}_0 = \mathbf{Y}_0$ , applying Gronwall’s inequality to deduce that there exists a constant  $C < \infty$  such that

$$\frac{1}{n} \sum_{i=1}^n \|X^i - Y^i\|_\infty \leq C \left\| \left(1 - \frac{1}{n}\right) \varphi^n - \varphi^\infty \right\|_\infty \frac{1}{n} \sum_{i=1}^n \|X^i\|_\infty, \quad a.s., \quad (6.6)$$

where, as usual,  $\|\cdot\|_\infty$  denotes the supremum norm on  $[0, T]$ . On the other hand, the equation (6.4) and Gronwall’s inequality together yield

$$\frac{1}{n} \sum_{i=1}^n \|X^i\|_\infty \leq C \left(1 + \frac{1}{n} \sum_{i=1}^n |X_0^i| + \frac{1}{n} \sum_{i=1}^n \|B^i\|_\infty + \|W\|_\infty\right), \quad a.s.$$

As soon as  $X_0^i$  are i.i.d. and subgaussian (e.g.,  $\mathbb{E}[\exp(\kappa|X_0^1|^2)] < \infty$  for some  $\kappa > 0$ ), we find a uniform subgaussian bound on these averages; that is, there exist constants  $C < \infty, \delta > 0$ , independent of  $n$ , such that

$$\mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \|X^i\|_\infty > a\right) \leq \exp(-\delta^2 a^2), \quad \text{for all } a \geq C, \quad n \in \mathbb{N}.$$

Assuming without any loss of generality that the constant  $C$  in the last display coincides with the one in (6.6), and letting  $r_n = C \left\| \left(1 - \frac{1}{n}\right) \varphi^n - \varphi^\infty \right\|_\infty$ , we find that, for  $a \geq Cr_n$ ,

$$\begin{aligned} \mathbb{P}(\mathcal{W}_{1,C^d}(m_{\mathbf{X}}^n, m_{\mathbf{Y}}^n) > a) &\leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \|X^i - Y^i\|_\infty > a\right) \leq \mathbb{P}\left(\frac{r_n}{n} \sum_{i=1}^n \|X^i\|_\infty > a\right) \\ &\leq \exp\left(-\frac{\delta^2 a^2}{r_n^2}\right). \end{aligned} \quad (6.7)$$

It is straightforward to check that  $r_n = O(1/n)$ , which implies in particular (4.9).

## 6.2 Derivation of the CLT

Following our study in Section 5, the key estimate (6.7) is sufficient to transfer a central limit theorem from the classical McKean-Vlasov particle system  $Y$  over to the Nash equilibrium state process  $X$ . Still, the derivation of the CLT suffers from the fact that the drift is unbounded, which prevents us from directly applying the results of Section 5. We explain below how to adapt the arguments.

The first step in the analysis is to check that the tightness property, as stated in Proposition 5.3, remains true. In fact, the main point is to revisit (5.12). For  $\lambda \geq 1$ , we have

$$\begin{aligned} \|\phi''\|_{1+\lambda, 2\lambda} + \|\phi'\|_{1+\lambda, 2\lambda} &\leq C\|\phi\|_{2+2\lambda, \lambda}, \\ \|\phi' \tilde{b}(t, \cdot, \mu)\|_{1+\lambda, 2\lambda} &\leq C\|\phi\|_{2+2\lambda, \lambda}, \end{aligned}$$

where  $\tilde{b}(t, x, \mu)$  is here given by  $(\bar{b} + q + \varphi_t^\infty)(x - \bar{\mu})$ .

Similar to (5.9), we let

$$[\mathcal{G}(t, m, \mu)\phi](v) = -(b + q + \varphi_t^\infty)\langle \mu, \phi' \rangle v.$$

It is then easily checked that the first line in (5.12) remains true (with  $\lambda_d$  replaced by  $\lambda \geq 1$ ). The rest of the proof of tightness, see the statement of Proposition 5.3, is similar. Importantly, it holds true for values of  $\lambda$  greater than or equal to  $\lfloor d/2 \rfloor + 1$  (which is here equal to 1), and not only for  $\lambda = 1$ , provided the constraint  $p' > 12\lambda$  in the statement of Theorem 3.2 is modified accordingly.

The second step, which is in fact more difficult, is the identification of the limit. The first point is to write down the corresponding equation (5.20). The difficulty is twofold: first, the function  $v \mapsto [\mathcal{G}(t, m, \mu)\phi](v)$  is not in  $\mathcal{H}^{2\lambda+4, \lambda}$  when  $\lambda = 1$ ; second, when the test function  $\phi$  is in the space  $\mathcal{H}^{2\lambda+4, \lambda}$ , the function  $\phi' \tilde{b}(t, \cdot, \mu)$  may not belong to  $\mathcal{H}^{2\lambda+2, \lambda}$  because of the term  $x\phi'(x)$  in  $\phi'(x)\tilde{b}(t, x, \mu)$ . The first difficulty may be easily circumvented by choosing  $\lambda > 3/2$ , in which case the identity function  $\text{Id} : v \mapsto v$  indeed belongs to  $\mathcal{H}^{2\lambda+4, \lambda}$ ; as already explained, this mostly requires us to change  $p'$  accordingly.

It is more challenging to deal with the fact that the function  $x \mapsto x\phi'(x)$  does not belong to  $\mathcal{H}^{2\lambda+2, \lambda}$ . The key point is to modify the space of test functions. Indeed, if we assume that  $\phi$  is in the space  $\mathcal{H}^{2\lambda+4, \lambda-1}$ , then the function  $\phi' \tilde{b}(t, \cdot, \mu)$  is in  $\mathcal{H}^{2\lambda+2, \lambda}$ . For this, it suffices to write down the analogue of (5.21), but in a smaller space of test functions. Specifically, this takes the form

$$\begin{aligned} d\langle S_t, \phi \rangle &= (\bar{b} + q + \varphi_t^\infty) \left[ \langle S_t, \phi' \times (\text{Id} - \bar{\mu}_t) \rangle - \langle S_t, \text{Id} \rangle \langle \mu_t, \phi' \rangle \right] dt \\ &\quad + \frac{1}{2} \langle S_t, (\sigma^2 + \sigma_0^2) \phi'' \rangle dt + \sigma_0 \langle S_t, \phi' \rangle dW_t + d\xi_t(\phi), \end{aligned} \quad (6.8)$$

for  $\phi$  in the space  $\mathcal{H}^{2\lambda+4, \lambda-1}$ .

Of course, the difficulty is then to prove uniqueness of the limiting equation. By linearity of the equation, it suffices to prove that any solution  $(S_t)_{t \in [0, T]}$  in  $\mathcal{H}^{-(2\lambda+2), \lambda}$  to the above equation, with a null initial condition and with  $\xi$  set to 0, is null. To do so, we assume that  $\lambda > 5/2$  (say  $\lambda = 3$  to simplify). Choosing  $\phi \equiv 1$  as test function, it is clear that  $\langle S_t, 1 \rangle = 0$ . Then, choosing  $\phi$  as the identity  $\text{Id} : x \mapsto \phi(x) = x$  as test function, we obtain  $d\langle S_t, \text{Id} \rangle = 0$ . The next step is to take advantage of the linear structure of the drift. Indeed, for any  $t \in [0, T]$  and any test function  $\phi \in \mathcal{H}^{2\lambda+4, \lambda-1}$ , we may consider the new test function  $\phi_t : \mathbb{R} \ni x \mapsto \phi(\ell_t x)$ , where

$$\ell_t = \exp\left(-\int_0^t (\bar{b} + q + \varphi_s^\infty) ds\right).$$

We then compute  $d\langle S_t, \phi_t \rangle$  as

$$\begin{aligned} d\langle S_t, \phi_t \rangle &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t \langle S_t, \phi'(\ell_t \cdot) \rangle dt \\ &\quad + \frac{\ell_t^2}{2} \langle S_t, (\sigma^2 + \sigma_0^2) \phi''(\ell_t \cdot) \rangle dt + \sigma_0 \ell_t \langle S_t, \phi'(\ell_t \cdot) \rangle dW_t, \quad t \in [0, T]. \end{aligned} \quad (6.9)$$

In order to conclude, we call  $S'_t$  the element of  $\mathcal{H}^{-(2\lambda+2), \lambda}$  defined by

$$\langle S'_t, \psi \rangle = \langle S_t, \psi(\ell_t \cdot) \rangle.$$

With this notation in hand, we can write

$$\begin{aligned} d\langle S'_t, \phi \rangle &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t \langle S'_t, \phi' \rangle dt \\ &\quad + \frac{\ell_t^2}{2} \langle S'_t, (\sigma^2 + \sigma_0^2) \phi'' \rangle dt + \sigma_0 \ell_t \langle S'_t, \phi' \rangle dW_t, \quad t \in [0, T]. \end{aligned}$$

Obviously,  $S'_0 = 0$ . Once again, the above holds true for any test function  $\phi \in \mathcal{H}^{2\lambda+4, \lambda-1}$ , but, since the drift no longer involves the function  $x \mapsto \phi'(x)x$ , it also holds true for any test function  $\phi \in \mathcal{H}^{2\lambda+4, \lambda}$ . Now, we can use the same argument in the proof of uniqueness in Section 5 to show that, necessarily,  $S'$  is null.

### 6.3 Explicit solution of the SPDE

We now provide an explicit expression for the solution to (6.8). To do so, we first notice that, in the above framework, the limiting McKean-Vlasov equation takes the form

$$d\mathcal{X}_t = (\bar{b} + q + \varphi_t^\infty)(\bar{\mu}_t - \mathcal{X}_t)dt + \sigma dB_t^1 + \sigma_0 dW_t, \quad t \in [0, T], \quad (6.10)$$

where  $\mu_t = \mathcal{L}(\mathcal{X}_t|W)$  and  $\bar{\mu}_t = \mathbb{E}[\mathcal{X}_t|W]$ . Equation (6.10) is the analogue of (2.11). Taking the conditional mean with respect to  $W$  in (6.10), we deduce that  $\bar{\mu}_t = \bar{\mu}_0 + \sigma_0 W_t$ , for  $t \in [0, T]$ . Also, we have

$$d(\mathcal{X}_t - \bar{\mu}_t) = -(\bar{b} + q + \varphi_t^\infty)(\mathcal{X}_t - \bar{\mu}_t)dt + \sigma dB_t^1,$$

that is,

$$\mathcal{X}_t - \bar{\mu}_t = \ell_t \left( \mathcal{X}_0 - \bar{\mu}_0 + \sigma \int_0^t \ell_s^{-1} dB_s^1 \right),$$

which shows that, conditional on  $(\mathcal{X}_0, W)$ ,  $\mathcal{X}_t$  has a Gaussian distribution. We then get

$$\mu_t = \int_{\mathbb{R}} \mathcal{N} \left( \ell_t x + (1 - \ell_t) \bar{\mu}_0 + \sigma_0 W_t, \sigma^2 \ell_t^2 \int_0^t \ell_s^{-2} ds \right) d\mu_0(x).$$

In particular,  $\mu_t$  has a smooth density, with which we identify it.

Recall from Section 3.2 the expression of the Gaussian process  $\xi$  in terms of a cylindrical white noise  $\beta$  with values in  $L^2(\mathbb{R})$ , independent of  $(\theta_0, W)$ . Taking  $\phi = \text{Id}$  in (6.8), we get as in the previous paragraph that  $d\langle S_t, \text{Id} \rangle = \sigma \langle \sqrt{\mu_t}, d\beta_t \rangle$ . That is,

$$\langle S_t, \text{Id} \rangle = \langle \theta_0, \text{Id} \rangle + \sigma \int_0^t \langle \sqrt{\mu_s}, d\beta_s \rangle. \quad (6.11)$$

Observe that, in the last term, the process  $(\int_0^t \langle \sqrt{\mu_s}, d\beta_s \rangle)_{t \in [0, T]}$  is a standard Brownian motion.

Now, performing the same computation as in (6.9), we have for any  $\phi \in \mathcal{H}^{2\lambda+4, \lambda-1}$ ,

$$\begin{aligned} d\langle S_t, \phi_t \rangle &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t \langle S_t, \phi'(\ell_t \cdot) \rangle dt - \ell_t(\bar{b} + q + \varphi_t^\infty) \langle S_t, \text{Id} \rangle \langle \mu_t, \phi'(\ell_t \cdot) \rangle dt \\ &\quad + \frac{\ell_t^2}{2} \langle S_t, (\sigma^2 + \sigma_0^2) \phi''(\ell_t \cdot) \rangle dt + \sigma_0 \ell_t \langle S_t, \phi'(\ell_t \cdot) \rangle dW_t + \sigma \ell_t \langle \sqrt{\mu_t(\cdot)} \phi'(\ell_t \cdot), d\beta_t(\cdot) \rangle, \end{aligned}$$

where  $\phi_t(\cdot) = \phi(\ell_t \cdot)$ . For  $t \in [0, T]$ , let  $\tilde{S}_t$  denote the element of  $\mathcal{H}^{-(2\lambda+2), \lambda}$  given by  $\langle \tilde{S}_t, \phi \rangle = \langle S_t, \phi(\ell_t \cdot) \rangle$ , for  $\phi \in \mathcal{H}^{2\lambda+2, \lambda}$ . Then, for  $\phi \in \mathcal{H}^{2\lambda+4, \lambda-1}$ ,

$$\begin{aligned} d\langle \tilde{S}_t, \phi \rangle &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t \langle \tilde{S}_t, \phi' \rangle dt - \ell_t(\bar{b} + q + \varphi_t^\infty) \langle S_t, \text{Id} \rangle \langle \mu_t, \phi'(\ell_t \cdot) \rangle dt \\ &\quad + \frac{\ell_t^2}{2} \langle \tilde{S}_t, (\sigma^2 + \sigma_0^2) \phi'' \rangle dt + \sigma_0 \ell_t \langle \tilde{S}_t, \phi' \rangle dW_t + \sigma \ell_t \langle \sqrt{\mu_t(\cdot)} \phi'(\ell_t \cdot), d\beta_t(\cdot) \rangle \\ &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t \langle \tilde{S}_t, \phi' \rangle dt - (\bar{b} + q + \varphi_t^\infty) \langle S_t, \text{Id} \rangle \langle \mu_t(\frac{\cdot}{\ell_t}), \phi'(\cdot) \rangle dt \\ &\quad + \frac{\ell_t^2}{2} \langle \tilde{S}_t, (\sigma^2 + \sigma_0^2) \phi'' \rangle dt + \sigma_0 \ell_t \langle \tilde{S}_t, \phi' \rangle dW_t + \sigma \sqrt{\ell_t} \langle \sqrt{\mu_t(\frac{\cdot}{\ell_t})} \phi'(\cdot), d\tilde{\beta}_t(\cdot) \rangle, \end{aligned} \quad (6.12)$$

where  $\tilde{\beta}$  is a cylindrical Brownian motion with values in  $L^2(\mathbb{R})$ , independent of  $(\theta_0, W)$ , given by

$$\langle \phi(\cdot), d\tilde{\beta}_t(\cdot) \rangle = \sqrt{\ell_t} \langle \phi(\ell_t \cdot), d\beta_t(\cdot) \rangle.$$

In particular, the stochastic term  $\langle \sqrt{\mu_t}, d\beta_t \rangle$  in the representation formula (6.11) may be rewritten as  $\langle \sqrt{\mu_t}, d\beta_t \rangle = \sqrt{\ell_t} \langle \sqrt{\mu_t(\cdot/\ell_t)}, d\tilde{\beta}_t(\cdot) \rangle$ .

Letting  $\psi_t = -\int_0^t \ell_s(\bar{b} + q + \varphi_s^\infty) \bar{\mu}_s ds + \int_0^t \sigma_0 \ell_s dW_s$  for  $t \in [0, T]$ , this prompts us to let, as a candidate for solving (6.12),

$$\begin{aligned} \tilde{S}_t(x) &= \langle \theta_0, g_{0,t}(\cdot - x) \rangle - \int_0^t (\bar{b} + q + \varphi_s^\infty) \langle S_s, \text{Id} \rangle \langle g'_{s,t}(\cdot - x), \mu_s(\frac{\cdot}{\ell_s}) \rangle ds \\ &\quad + \int_0^t \sigma \sqrt{\ell_s} \langle g'_{s,t}(\cdot - x), \sqrt{\mu_s(\frac{\cdot}{\ell_s})}, d\tilde{\beta}_s(\cdot) \rangle, \end{aligned}$$

which makes sense in a distributional sense, with

$$g_{s,t}(x) = \frac{1}{\sqrt{\sigma^2 \int_s^t \ell_r^2 dr}} \Phi\left(\frac{x + \psi_t - \psi_s}{\sqrt{\sigma^2 \int_s^t \ell_r^2 dr}}\right)$$

where  $\Phi$  is the standard Gaussian density. To make it clear, for  $\phi \in \mathcal{H}^{2\lambda+2, \lambda}$ ,

$$\begin{aligned} \langle \tilde{S}_t, \phi \rangle &= \langle \theta_0, g_{0,t} * \phi \rangle - \int_0^t (\bar{b} + q + \varphi_s^\infty) \langle S_s, \text{Id} \rangle \langle g_{s,t} * \phi', \mu_s(\frac{\cdot}{\ell_s}) \rangle ds \\ &\quad + \int_0^t \sigma \sqrt{\ell_s} \langle (g_{s,t} * \phi')(\cdot), \sqrt{\mu_s(\frac{\cdot}{\ell_s})}, d\tilde{\beta}_s(\cdot) \rangle, \end{aligned} \quad (6.13)$$

where  $*$  is the usual convolution operator. We prove below that this defines a square-integrable process with values in  $\mathcal{C}([0, T]; \mathcal{H}^{-(2\lambda+2), \lambda})$ . Then, by standard arguments in stochastic calculus,  $((g_{s,t} * \phi)(x))_{t \geq s}$  is a semi-martingale for any  $x \in \mathbb{R}$  and any  $\phi \in \mathcal{H}^{2, \lambda+1}$  and, with probability 1,

$$\begin{aligned} d_t(g_{s,t} * \phi)(x) &= -\ell_t(\bar{b} + q + \varphi_t^\infty) \bar{\mu}_t(g_{s,t} * \phi')(x) \\ &\quad + \frac{\ell_t^2}{2} \sigma^2 (g_{s,t} * \phi'')(x) dt + \sigma_0 \ell_t (g_{s,t} * \phi')(x) dW_t, \quad t \geq s, \end{aligned}$$

with  $\phi(x)$  as initial condition. Following the proof of the Itô-Wentzell formula (see, for instance, [27]), we deduce that (6.12) holds true for  $\phi \in \mathcal{H}^{2\lambda+4, \lambda-1}$ . By (6.9), this suffices to identify the limit in the central limit theorem.

It then remains to check that (6.13) defines a square-integrable process with values in  $\mathcal{C}([0, T]; \mathcal{H}^{-(2\lambda+2), \lambda})$ . This is easily checked for the first two terms in the right-hand

side, but the verification for the last term is more involved. To prove that it generates a square-integrable process with values in  $\mathcal{C}([0, T]; \mathcal{H}^{-(2\lambda+2), \lambda})$ , it suffices to prove that

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \int_0^t \sqrt{\ell_s} \left\langle (g_{s,t} * \phi')(\cdot) \sqrt{\mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle \right|^2 \right] \leq C \|\phi\|_{2, \lambda+1}^2,$$

and then to use the fact the embedding from  $\mathcal{H}^{2\lambda+2, \lambda}$  into  $\mathcal{H}^{2, \lambda+1}$  is Hilbert Schmidt, see (5.6). To do so, we use the *factorization method* for SPDEs, see [16, Chapter 7]. Notice indeed that, for  $\varepsilon > 0$  small enough and for  $\phi \in \mathcal{H}^{2, \lambda+1}$ ,

$$\begin{aligned} & \int_0^t \sqrt{\ell_s} \left\langle (g_{s,t} * \phi')(\cdot) \sqrt{\mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle \\ &= c_\varepsilon \int_0^t (t-r)^{-1+\varepsilon} \left[ \int_{\mathbb{R}} g_{r,t}(x) \left( \int_0^r \left\langle (r-s)^{-\varepsilon} (g_{s,r} * \phi')(\cdot - x) \sqrt{\ell_s \mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle dx \right) dr \right], \end{aligned}$$

where  $c_\varepsilon$  is a normalization constant such that  $c_\varepsilon \int_0^t (t-r)^{-1+\varepsilon} (r-s)^{-\varepsilon} dr = 1$ . By Hölder's inequality, we notice that, for  $q > 1$ , the right-hand side is less than

$$\begin{aligned} & c_\varepsilon \left( \int_0^t \int_{\mathbb{R}} (t-r)^{(-1+\varepsilon)q/(q-1)} (1+|x|^{\lambda+3})^{q/(q-1)} g_{r,t}^{q/(q-1)}(x) dr dx \right)^{(q-1)/q} \\ & \times \left( \int_0^T \int_{\mathbb{R}} (1+|x|^{\lambda+3})^{-q} \left| \int_0^r \left\langle (r-s)^{-\varepsilon} (g_{s,r} * \phi')(\cdot - x) \sqrt{\ell_s \mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle dx \right|^q dr \right)^{1/q}. \end{aligned}$$

Choosing  $q$  large enough, we can find a constant  $C$  (the value of which is allowed to change from line to line), independent of  $t \in [0, T]$ , such that

$$\begin{aligned} & \int_0^t \int_{\mathbb{R}} (t-r)^{(-1+\varepsilon)q/(q-1)} (1+|x|^{\lambda+3})^{q/(q-1)} g_{r,t}^{q/(q-1)}(x) dr dx \\ & \leq C \left( 1 + \sup_{s \in [0, T]} |\psi_s| \right)^{(\lambda+3)q/(q-1)} \int_0^t (t-r)^{-1+\varepsilon/2} \\ & \leq C \left( 1 + \sup_{s \in [0, T]} |\psi_s| \right)^{(\lambda+3)q/(q-1)}. \end{aligned}$$

Hence, it suffices to prove that, for any  $q \geq 2$ ,

$$\sup_{r \in [0, T]} \mathbb{E} \left[ \int_{\mathbb{R}} (1+|x|^{\lambda+3})^{-q} \left| \int_0^r \left\langle (r-s)^{-\varepsilon} (g_{s,r} * \phi')(\cdot - x) \sqrt{\ell_s \mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle dx \right|^q \right] < \infty.$$

For  $r \in [0, T]$ , the Burkholder-Davis-Gundy inequality and boundedness of  $\ell$  yield

$$\begin{aligned} & \mathbb{E} \left[ \int_{\mathbb{R}} (1+|x|^{\lambda+3})^{-q} \left| \int_0^r \left\langle (r-s)^{-\varepsilon} (g_{s,r} * \phi')(\cdot - x) \sqrt{\ell_s \mu_s\left(\frac{\cdot}{\ell_s}\right)}, d\tilde{\beta}_s(\cdot) \right\rangle dx \right|^q \right] \\ & \leq C \mathbb{E} \left[ \int_{\mathbb{R}} (1+|x|^{\lambda+3})^{-q} \left| \int_0^r \int_{\mathbb{R}} (r-s)^{-2\varepsilon} |(g_{s,r} * \phi')(y-x)|^2 \mu_s\left(\frac{y}{\ell_s}\right) ds dy \right|^{q/2} dx \right] \\ & \leq C \mathbb{E} \left[ \int_0^r \int_{\mathbb{R}} \int_{\mathbb{R}} (1+|x|^{\lambda+3})^{-q} (r-s)^{-2\varepsilon} |(g_{s,r} * \phi')(y-x)|^q \mu_s\left(\frac{y}{\ell_s}\right) ds dx dy \right]. \end{aligned}$$

Noticing from (5.5) that  $|(g_{s,r} * \phi')(x)| \leq C \|\phi\|_{2, \lambda+1} (1+|x| + \sup_{t \in [0, T]} |\psi_t|)^{\lambda+1}$ , the proof is easily completed.

## A On derivatives on Wasserstein space

In this section we state two simple but useful technical results on the derivative  $D_m$ , and we prove two results from the body of the paper.

**Lemma A.1.** For any function  $V = V(x, m) : \mathbb{R}^d \times \mathcal{P}^q(\mathbb{R}^d) \rightarrow \mathbb{R}$ , we have

$$D_x D_m V(x, m, v) = D_m D_x V(x, m, v), \quad \text{for all } v \in \mathbb{R}^d,$$

as long as the derivatives on both sides exist and are jointly continuous. More precisely, assume that  $V$  is  $\mathcal{C}^1$  in the sense of Section 2.2, that the derivatives  $D_x(\delta V/\delta m)$ ,  $D_x V$ , and  $\delta(D_x V)/\delta m$  exist, and that for every compact set  $K \subset \mathcal{P}^q(\mathbb{R}^d)$  there exists  $c < \infty$  such that  $\sup_{m \in K} |D_x \frac{\delta V}{\delta m}(x, m, v)| \leq c(1 + |x|^q + |v|^q)$  for all  $x, v \in \mathbb{R}^d$ .

*Proof.* Fix  $x \in \mathbb{R}^d$ ,  $m, m' \in \mathcal{P}^q(\mathbb{R}^d)$ , and for  $t \in [0, 1]$  let  $m^t = (1 - t)m + tm'$ . Then by (2.2),

$$V(x, m') - V(x, m) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta V}{\delta m}(x, m^t, v)(m' - m)(dv)dt.$$

Apply  $D_x$  to both sides and use dominated convergence to interchange the order of  $D_x$  and the integral to get

$$D_x V(x, m') - D_x V(x, m) = \int_0^1 \int_{\mathbb{R}^d} D_x \frac{\delta V}{\delta m}(x, m^t, v)(m' - m)(dv)dt.$$

Another application of (2.2) shows that for every  $v \in \mathbb{R}^d$ ,

$$\frac{\delta D_x V}{\delta m}(x, m, v) = D_x \frac{\delta V}{\delta m}(x, m, v).$$

To conclude the proof, apply  $D_v$  to both sides, commute  $D_v$  and  $D_x$  on the right-hand side, and use the definition of  $D_m$  from (2.4).  $\square$

**Lemma A.2.** Suppose a function  $V : [0, T] \times \mathbb{R}^d \times \mathcal{P}^q(\mathbb{R}^d) \rightarrow \mathbb{R}$  is continuous. Suppose also that  $D_x V$  and  $D_m V$  exist and are continuous and bounded. Then there exists  $C < \infty$  such that, for all  $(t, x, m)$ ,

$$|V(t, x, m)| \leq C(1 + |x| + \mathcal{W}_1(m, \delta_0)),$$

where  $\mathcal{W}_1$  is the Wasserstein distance on  $\mathcal{P}(\mathbb{R}^d)$  defined in Section 2.1.

*Proof.* Begin by writing

$$\begin{aligned} V(t, x, m) - V(t, 0, \delta_0) &= \int_0^1 (x \cdot D_x V(t, ux, (1 - u)\delta_0 + um) \\ &\quad + \int_{\mathbb{R}^d} \frac{\delta V}{\delta m}(t, ux, (1 - u)\delta_0 + um, v)(m - \delta_0)(dv)du. \end{aligned}$$

Since  $|D_m V|$  is bounded, say by a constant  $L < \infty$ , the map  $v \mapsto \frac{\delta V}{\delta m}(t, x, m, v)$  is  $L$ -Lipschitz, for each  $(t, x, m)$ . Hence, Kantorovich duality implies

$$\left| \int_{\mathbb{R}^d} \frac{\delta V}{\delta m}(t, ux, (1 - u)\delta_0 + um, v)(m - \delta_0)(dv) \right| \leq L\mathcal{W}_1(m, \delta_0).$$

If  $C < \infty$  satisfies  $|D_x V| \leq C$  pointwise, then we conclude

$$|V(t, x, m)| \leq |V(t, 0, \delta_0)| + C|x| + L\mathcal{W}_1(m, \delta_0).$$

To complete the proof, note that the continuous function  $V(\cdot, 0, \delta_0)$  is bounded on  $[0, T]$ .  $\square$



### Proof of Proposition 2.1

Let  $m \in \mathcal{P}^q(\mathbb{R}^d)$  and  $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ . By continuity, it suffices to prove the claims assuming the points  $x_1, \dots, x_n \in \mathbb{R}^d$  are distinct. Fix an index  $j \in \{1, \dots, n\}$  and a bounded continuous function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , to be specified later. We claim that, under the assumptions of part (i),

$$\lim_{h \downarrow 0} \frac{V(m \circ (\text{Id} + h\phi)^{-1}) - V(m)}{h} = \int_{\mathbb{R}^d} D_m V(m, v) \cdot \phi(v) m(dv) \quad (\text{A.1})$$

holds, where  $\text{Id}$  denotes the identity map on  $\mathbb{R}^d$ . Once (A.1) is proven, we complete the proof as follows. For a fixed vector  $v \in \mathbb{R}^d$  we may choose  $\phi$  such that  $\phi(x_j) = v$  while  $\phi(x_i) = 0$  for  $i \neq j$ . Let  $\mathbf{v} \in (\mathbb{R}^d)^n$  have  $j^{\text{th}}$  coordinate equal to  $v$  and  $i^{\text{th}}$  coordinate zero for  $i \neq j$ . Then  $u_n(\mathbf{x}) = V(m_{\mathbf{x}}^n)$  satisfies

$$\begin{aligned} \lim_{h \downarrow 0} \frac{u_n(\mathbf{x} + h\mathbf{v}) - u_n(\mathbf{x})}{h} &= \lim_{h \downarrow 0} \frac{V(m_{\mathbf{x}}^n \circ (\text{Id} + h\phi)^{-1}) - V(m_{\mathbf{x}}^n)}{h} \\ &= \frac{1}{n} \sum_{k=1}^n D_m V(m_{\mathbf{x}}^n, x_k) \cdot \phi(x_k) \\ &= \frac{1}{n} D_m V(m_{\mathbf{x}}^n, x_j) \cdot v. \end{aligned}$$

This proves (i). Under the additional assumptions, (ii) follows by applying (i) again.

It remains to prove (A.1). For  $h > 0$ ,  $t \in [0, 1]$ , and  $m \in \mathcal{P}^q(\mathbb{R}^d)$ , let  $m_{h,t} = tm \circ (\text{Id} + h\phi)^{-1} + (1-t)m$ . Then, using (2.2) and (2.4), respectively, in the first and third equalities below, we obtain

$$\begin{aligned} V(m \circ (\text{Id} + h\phi)^{-1}) - V(m) &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta V}{\delta m}(m_{h,t}, v) (m \circ (\text{Id} + h\phi)^{-1} - m)(dv) dt \\ &= \int_0^1 \int_{\mathbb{R}^d} \left( \frac{\delta V}{\delta m}(m_{h,t}, v + h\phi(v)) - \frac{\delta V}{\delta m}(m_{h,t}, v) \right) m(dv) dt \\ &= h \int_0^1 \int_{\mathbb{R}^d} \int_0^1 D_m V(m_{h,t}, v + sh\phi(v)) \cdot \phi(v) ds m(dv) dt. \end{aligned}$$

As  $h \downarrow 0$  we have  $m_{h,t} \rightarrow m$  and  $sh\phi(v) \rightarrow 0$ , and we deduce (A.1) from the bounded convergence theorem and continuity of  $D_m$ .  $\square$

### Proof of Proposition 4.1

We follow the proof of Proposition 6.1.3 in [6]. For fixed  $n \in \mathbb{N}$  and  $\mathbf{x} \in \mathbb{R}^d$ , recall that  $u^{n,i}(t, \mathbf{x}) = U(t, x_i, m_{\mathbf{x}}^n)$ , and use Proposition 2.1 and Lemma A.1 to derive, for distinct  $i, j, k \in \{1, \dots, n\}$

$$D_{x_i} u^{n,i}(t, \mathbf{x}) = D_x U(t, x_i, m_{\mathbf{x}}^n) + \frac{1}{n} D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) \quad (\text{A.2})$$

$$D_{x_j} u^{n,i}(t, \mathbf{x}) = \frac{1}{n} D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) \quad (\text{A.3})$$

$$\begin{aligned} D_{x_i, x_i}^2 u^{n,i}(t, \mathbf{x}) &= D_x^2 U(t, x_i, m_{\mathbf{x}}^n) + \frac{1}{n} D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) \\ &\quad + \frac{1}{n} D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_i) \end{aligned} \quad (\text{A.4})$$

$$D_{x_i, x_j}^2 u^{n,i}(t, \mathbf{x}) = \frac{1}{n} D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_j) \quad (\text{A.5})$$

$$D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{x}) = \frac{1}{n} D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_j) \quad (\text{A.6})$$

$$D_{x_j, x_k}^2 u^{n,i}(t, \mathbf{x}) = \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_k). \quad (\text{A.7})$$

Use the master equation (2.10) to find

$$0 = \partial_t u^{n,i}(t, \mathbf{x}) + \Phi^{n,i}(t, \mathbf{x}) + \Psi^{n,i}(t, \mathbf{x}) + \Xi^{n,i}(t, \mathbf{x}),$$

where we define

$$\begin{aligned} \Psi^{n,i}(t, \mathbf{x}) &= H(x, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)) \\ &\quad + \int_{\mathbb{R}^d} \widehat{b}(v, m_{\mathbf{x}}^n, D_x U(t, v, m_{\mathbf{x}}^n)) \cdot D_m U(t, x_i, m_{\mathbf{x}}^n, v) dm_{\mathbf{x}}^n(v), \\ \Phi^{n,i}(t, \mathbf{x}) &= \frac{1}{2} \int_{\mathbb{R}^d} \text{Tr} [\sigma \sigma^\top D_v D_m U(t, x_i, m_{\mathbf{x}}^n, v)] dm_{\mathbf{x}}^n(v) + \frac{1}{2} \text{Tr} [\sigma \sigma^\top D_x^2 U(t, x_i, m_{\mathbf{x}}^n)], \end{aligned}$$

and

$$\begin{aligned} \Xi^{n,i}(t, \mathbf{x}) &= \frac{1}{2} \int_{\mathbb{R}^d} \text{Tr} [\sigma_0 \sigma_0^\top D_v D_m U(t, x_i, m_{\mathbf{x}}^n, v)] dm_{\mathbf{x}}^n(v) \\ &\quad + \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \text{Tr} [\sigma_0 \sigma_0^\top D_m^2 U(t, x_i, m_{\mathbf{x}}^n, v, v')] dm_{\mathbf{x}}^n(v) dm_{\mathbf{x}}^n(v') \\ &\quad + \int_{\mathbb{R}^d} \text{Tr} [\sigma_0 \sigma_0^\top D_x D_m U(t, x_i, m_{\mathbf{x}}^n, v)] dm_{\mathbf{x}}^n(v) + \frac{1}{2} \text{Tr} [\sigma_0 \sigma_0^\top D_x^2 U(t, x_i, m_{\mathbf{x}}^n)]. \end{aligned}$$

Now use (A.3), (A.2) and (2.8) to get

$$\begin{aligned} \Psi^{n,i}(t, \mathbf{x}) &= \frac{1}{n} \sum_{j=1}^n \widehat{b}(x_j, m_{\mathbf{x}}^n, D_x U(t, x_j, m_{\mathbf{x}}^n)) \cdot D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) \\ &\quad + H(x, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)) \\ &= \sum_{j=1, j \neq i}^n \widehat{b}(x_j, m_{\mathbf{x}}^n, D_x U(t, x_j, m_{\mathbf{x}}^n)) \cdot D_{x_j} u^{n,i}(t, \mathbf{x}) \\ &\quad + \widehat{b}(x_i, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)) \cdot (D_{x_i} u^{n,i}(t, \mathbf{x}) - D_x U(t, x_i, m_{\mathbf{x}}^n)) \\ &\quad + \widehat{b}(x_i, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)) \cdot D_x U(t, x_i, m_{\mathbf{x}}^n) + \widehat{f}(x_i, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)) \\ &= \sum_{j=1}^n \widehat{b}(x_j, m_{\mathbf{x}}^n, D_x U(t, x_j, m_{\mathbf{x}}^n)) \cdot D_{x_j} u^{n,i}(t, \mathbf{x}) + \widehat{f}(x_i, m_{\mathbf{x}}^n, D_x U(t, x_i, m_{\mathbf{x}}^n)). \end{aligned}$$

Now, recall from Assumption **A**(5) that  $D_m U$  and  $D_x U$  are uniformly bounded. Thus, since  $u^{n,j}(t, \mathbf{x}) = U(t, x_j, m_{\mathbf{x}}^n)$ , we use (A.2) and (A.3) to obtain  $\|D_{x_j} u^{n,i}\|_\infty \leq C/n$  for  $j \neq i$  and  $\|D_{x_i} u^{n,i}\|_\infty \leq C$ . Moreover,  $\widehat{b}(x, m, y)$  and  $\widehat{f}(x, m, y)$  are locally Lipschitz in  $y$  by respectively Assumptions **A**(1) and either Assumption **B** or **B'**(3). Hence, using (A.2), we have

$$\begin{aligned} &\left| \Psi^{n,i}(t, \mathbf{x}) - \sum_{j=1}^n \widehat{b}(x_j, m_{\mathbf{x}}^n, D_{x_j} u^{n,j}(t, \mathbf{x})) \cdot D_{x_j} u^{n,i}(t, \mathbf{x}) - \widehat{f}(x_i, m_{\mathbf{x}}^n, D_{x_i} u^{n,i}(t, \mathbf{x})) \right| \\ &\leq C \sum_{j=1}^n |D_{x_j} u^{n,i}(t, \mathbf{x})| |D_x U(t, x_j, m_{\mathbf{x}}^n) - D_{x_j} u^{n,j}(t, \mathbf{x})| \\ &\quad + C (1 + |D_x U(t, x_i, m_{\mathbf{x}}^n)| + |D_{x_i} u^{n,i}(t, \mathbf{x})|) |D_x U(t, x_i, m_{\mathbf{x}}^n) - D_{x_i} u^{n,i}(t, \mathbf{x})| \\ &\leq \frac{C}{n} \sum_{j=1}^n |D_{x_j} u^{n,i}(t, \mathbf{x})| |D_m U(t, x_j, m_{\mathbf{x}}^n, x_j)| + \frac{C}{n} |D_m U(t, x_i, m_{\mathbf{x}}^n, x_i)| \\ &\leq \frac{C}{n}. \end{aligned}$$

The argument for the volatility terms is similar, but somewhat more involved. Expand the integrals and use (A.4) and (A.6) to get

$$\begin{aligned} \Phi^{n,i}(t, \mathbf{x}) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{2} \text{Tr} [\sigma \sigma^\top D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_j)] + \frac{1}{2} \text{Tr} [\sigma \sigma^\top D_x^2 U(t, x_i, m_{\mathbf{x}}^n)] \\ &= \frac{1}{2} \text{Tr} \left[ \sigma \sigma^\top \left( D_x^2 U(t, x_i, m_{\mathbf{x}}^n) + \frac{1}{n} D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) \right) \right] \\ &\quad + \frac{1}{n} \sum_{j=1, j \neq i}^n \frac{1}{2} \text{Tr} [\sigma \sigma^\top D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_j)] \\ &= \frac{1}{2} \text{Tr} \left[ \sigma \sigma^\top \left( D_{x_i, x_i}^2 u^{n,i}(t, \mathbf{x}) - \frac{1}{n} D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) - \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_i) \right) \right] \\ &\quad + \sum_{j=1, j \neq i}^n \frac{1}{2} \text{Tr} \left[ \sigma \sigma^\top \left( D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{x}) - \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_j) \right) \right]. \end{aligned}$$

Hence, using the bounds from Assumption A(5), we obtain

$$\begin{aligned} &\left| \Phi^{n,i}(t, \mathbf{x}) - \sum_{j=1}^n \frac{1}{2} \text{Tr} [\sigma \sigma^\top D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{x})] \right| \\ &\leq \frac{1}{2n} \left| \sigma \sigma^\top \left( D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) + \frac{1}{n} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_i) \right) \right| \\ &\quad + \frac{1}{2n^2} \sum_{j=1, j \neq i}^n |\sigma \sigma^\top D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_j)| \\ &\leq \frac{C}{n}. \end{aligned}$$

Finally, we use (A.4), (A.5), (A.6), and (A.7) to get

$$\begin{aligned} \Xi^{n,i}(t, \mathbf{x}) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{2} \text{Tr} [\sigma_0 \sigma_0^\top D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_j)] \\ &\quad + \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \frac{1}{2} \text{Tr} [\sigma_0 \sigma_0^\top D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_k)] \\ &\quad + \frac{1}{n} \sum_{j=1}^n \text{Tr} [\sigma_0 \sigma_0^\top D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_j)] + \frac{1}{2} \text{Tr} [\sigma_0 \sigma_0^\top D_x^2 U(t, x_i, m_{\mathbf{x}}^n)] \\ &= \frac{1}{2} \text{Tr} \left[ \sigma_0 \sigma_0^\top \left( \frac{1}{n} D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) + \frac{1}{n} D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_i) \right. \right. \\ &\quad \left. \left. + D_x^2 U(t, x_i, m_{\mathbf{x}}^n) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_i) \right) \right] \\ &\quad + \frac{1}{2} \sum_{j=1, j \neq i}^n \text{Tr} \left[ \sigma_0 \sigma_0^\top \left( \frac{1}{n} D_v D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_j) \right) \right] \\ &\quad + \sum_{j=1, j \neq i}^n \text{Tr} \left[ \sigma_0 \sigma_0^\top \left( \frac{1}{n} D_x D_m U(t, x_i, m_{\mathbf{x}}^n, x_j) + \frac{1}{n^2} D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_i, x_j) \right) \right] \\ &\quad + \frac{1}{2n^2} \sum_{j=1, j \neq i}^n \sum_{k=1, k \notin \{i, j\}}^n \text{Tr} [\sigma_0 \sigma_0^\top D_m^2 U(t, x_i, m_{\mathbf{x}}^n, x_j, x_k)], \end{aligned}$$

and then

$$\begin{aligned}\Xi^{n,i}(t, \mathbf{x}) &= \frac{1}{2} \text{Tr} [\sigma_0 \sigma_0^\top D_{x_i, x_i}^2 u^{n,i}(t, \mathbf{x})] + \frac{1}{2} \sum_{j \neq i} \text{Tr} [\sigma_0 \sigma_0^\top D_{x_j, x_j}^2 u^{n,i}(t, \mathbf{x})] \\ &\quad + \sum_{j \neq i} \text{Tr} [\sigma_0 \sigma_0^\top D_{x_i, x_j}^2 u^{n,i}(t, \mathbf{x})] + \frac{1}{2} \sum_{j=1, j \neq i}^n \sum_{k \notin \{i, j\}} \text{Tr} [\sigma_0 \sigma_0^\top D_{x_j, x_k}^2 u^{n,i}(t, \mathbf{x})] \\ &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \text{Tr} [\sigma_0 \sigma_0^\top D_{x_j, x_k}^2 u^{n,i}(t, \mathbf{x})].\end{aligned}$$

Complete the proof by combining the above results.  $\square$

## Acknowledgment

François Delarue is supported by ANR-16-CE40-0015-01 and Institut Universitaire de France. Kavita Ramanan is supported in part by the National Science Foundation under award DMS-1713032.

The authors are very grateful to the anonymous referee for her/his careful reading of the manuscript.

## References

- [1] S. Ahuja, W. Ren, and T.-W. Yang, *Asymptotic analysis of mean field games with small common noise*, arXiv preprint arXiv:1707.09090 (2017).
- [2] E. Bayraktar and A. Cohen, *Analysis of a finite state many player game using its master equation*, arXiv preprint arXiv:1707.02648 (2017).
- [3] A. Bensoussan and J. Frehse, *Nonlinear elliptic systems in stochastic game theory*, Universität Bonn. SFB 72. Approximation und Optimierung, 1983.
- [4] A. Bensoussan, J. Frehse, and S.C.P. Yam, *The master equation in mean field theory*, Journal de Mathématiques Pures et Appliquées **103** (2015), no. 6, 1441–1474.
- [5] ———, *On the interpretation of the master equation*, Stochastic Processes and their Applications **127** (2017), no. 7, 2093–2137.
- [6] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The master equation and the convergence problem in mean field games*, Annals of Maths Studies, Princeton University Press, 2019.
- [7] R. Carmona and F. Delarue, *The master equation for large population equilibria*, Stochastic Analysis and Applications 2014, Springer, 2014, pp. 77–128.
- [8] ———, *Probabilistic Theory of Mean Field Games: Vol. I, Mean Field FBSDEs, Control, and Games*, Stochastic Analysis and Applications, Springer Verlag, 2017.
- [9] ———, *Probabilistic Theory of Mean Field Games: Vol. II, Mean Field Games with Common Noise and Master Equations*, Stochastic Analysis and Applications, Springer Verlag, 2017.
- [10] R. Carmona, J.-P. Fouque, and L.-H. Sun, *Mean field games and systemic risk*, Communications in Mathematical Sciences **13** (2015), no. 4, 911–933.
- [11] R. Carmona and X. Zhu, *A probabilistic approach to mean field games with major and minor players*, The Annals of Applied Probability **26** (2016), no. 3, 1535–1580.
- [12] A. Cecchin and M. Fischer, *Probabilistic approach to finite state mean field games*, arXiv preprint arXiv:1704.00984 (2017).
- [13] A. Cecchin and G. Pelino, *Convergence, fluctuations and large deviations for finite state mean field games via the master equation*, arXiv preprint arXiv:1707.01819 (2017).
- [14] J.-F. Chassagneux, D. Crisan, and F. Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, arXiv preprint arXiv:1411.3009 (2014).
- [15] M. Coghi and F. Flandoli, *Propagation of chaos for interacting particles subject to environmental noise*, Ann. Appl. Probab. **26** (2016), no. 3, 1407–1442.

- [16] Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, second ed., Encyclopedia of Mathematics and its Applications, vol. 152, Cambridge University Press, Cambridge, 2014. MR-3236753
- [17] D. Dawson and J. Vaillancourt, *Stochastic McKean-Vlasov equations*, Nonlinear Differential Equations and Applications NoDEA **2** (1995), no. 2, 199–229.
- [18] F. Delarue, D. Lacker, and K. Ramanan, *From the master equation to mean field game limit theory: Large deviations and concentration of measure*, 2018.
- [19] M. Fischer, *On the connection between symmetric  $n$ -player games and mean field games*, The Annals of Applied Probability **27** (2017), no. 2, 757–810.
- [20] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields **162** (2015), no. 3-4, 707–738.
- [21] W. Gangbo and A. Swiech, *Existence of a solution to an equation arising from the theory of mean field games*, Journal of Differential Equations **259** (2015), no. 11, 6573–6643.
- [22] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions*, Mathematische Nachrichten **137** (1988), no. 1, 197–248.
- [23] M. Hitsuda and I. Mitoma, *Tightness problem and stochastic evolution equation arising from fluctuation phenomena for interacting diffusions*, Journal of Multivariate Analysis **19** (1986), no. 2, 311–328.
- [24] M. Huang, R. Malhamé, and P. Caines, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Communications in Information & Systems **6** (2006), no. 3, 221–252.
- [25] B. Jourdain and S. Méléard, *Propagation of chaos and fluctuations for a moderate model with smooth initial data*, **34** (1998), 727–766.
- [26] V.N. Kolokoltsov, J. Li, and W. Yang, *Mean field games and nonlinear Markov processes*, arXiv preprint arXiv:1112.3744 (2011).
- [27] H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1990.
- [28] T.G. Kurtz and J. Xiong, *Numerical solutions for a class of SPDEs with application to filtering*, Stochastics in finite and infinite dimensions, Trends Math., Birkhäuser Boston, Boston, MA, 2001, pp. 233–258.
- [29] ———, *A stochastic evolution equation arising from the fluctuations of a class of interacting particle systems*, Communications in Mathematical Sciences **2** (2004), no. 3, 325–358.
- [30] D. Lacker, *A general characterization of the mean field limit for stochastic differential games*, Probability Theory and Related Fields **165** (2016), no. 3-4, 581–648.
- [31] ———, *On the convergence of closed-loop Nash equilibria to the mean field game limit*, arXiv preprint arXiv:1808.02745 (2018).
- [32] O.A. Ladyzhenskaia, V.A. Solonnikov, and N.N. Ural'tseva, *Linear and quasi-linear equations of parabolic type*, vol. 23, American Mathematical Soc., 1988.
- [33] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen. i-le cas stationnaire*, Comptes Rendus Mathématique **343** (2006), no. 9, 619–625.
- [34] ———, *Jeux à champ moyen. ii-horizon fini et contrôle optimal*, Comptes Rendus Mathématique **343** (2006), no. 10, 679–684.
- [35] ———, *Mean field games*, Japanese journal of mathematics **2** (2007), no. 1, 229–260.
- [36] P.-L. Lions, *Cours au Collège de France*.
- [37] Huang M, R. Malhamé, and P. Caines, *Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized  $\varepsilon$ -Nash equilibria*, IEEE transactions on automatic control **52** (2007), no. 9, 1560–1571.
- [38] J. Ma, P. Protter, and J. Yong, *Solving forward-backward stochastic differential equations explicitly – a four step scheme*, Probab. Theory Related Fields **98** (1994), 339–359.
- [39] H.P. McKean, *A class of Markov processes associated with nonlinear parabolic equations*, Proceedings of the National Academy of Sciences **56** (1966), no. 6, 1907–1911.

- [40] S. Méléard, *Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models*, Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math., vol. 1627, Springer, 1996, pp. 42–95.
- [41] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, The Annals of Probability (1984), 458–479.
- [42] I. Pinelis, *Optimum bounds for the distributions of martingales in Banach spaces*, Ann. Probab. **22** (1994), no. 4, 1679–1706.
- [43] T. Strömberg, *A note on the differentiability of conjugate functions*, Archiv der Mathematik **93** (2009), no. 5, 481.
- [44] A.-S. Sznitman, *Topics in propagation of chaos*, Ecole d’Eté de Probabilités de Saint-Flour XIX—1989 (1991), 165–251.
- [45] A.S. Sznitman, *A fluctuation result for nonlinear diffusions*, Infinite Dimensional Analysis and Stochastic Processes (1985), 145–160.
- [46] H. Tanaka and M. Hitsuda, *Central limit theorem for a simple diffusion model of interacting particles*, Hiroshima Mathematical Journal **11** (1981), no. 2, 415–423.
- [47] A. J. Veretennikov, *On strong solutions and explicit formulas for solutions of stochastic integral equations*, Mathematics of the USSR-Sbornik **39** (1981), no. 3, 387.