Backward propagation of chaos[*]

Mathieu Laurière[†]   Ludovic Tangpi[‡]

Abstract

This paper develops a theory of propagation of chaos for a system of weakly interacting particles whose *terminal* configuration is fixed as opposed to the initial configuration as customary. Such systems are modeled by backward stochastic differential equations. Under standard assumptions on the coefficients of the equations, we prove propagation of chaos results and quantitative estimates on the rate of convergence in Wasserstein distance of the empirical measure of the interacting system to the law of a McKean-Vlasov type equation. These results are accompanied by non-asymptotic concentration inequalities. As an application, we derive rate of convergence results for solutions of second order semilinear partial differential equations to the solution of a partial differential written on an infinite dimensional space.

1 Introduction

The theory of propagation of chaos takes its origin in the work of M. Kac [39] whose initial aim was to investigate particle system approximations of some nonlocal partial differential equations (PDE) arising in thermodynamics. The intuitive idea is the following: Consider a large number n of (random) particles starting from n given independent and identically distributed random variables and whose respective dynamics interact. Because there is no deterministic pattern for the starting position of the particles, one says that the *initial configuration is chaotic*. Kac's insight was that if the interaction between the particles is "sufficiently weak" and the particles are "symmetric", then as the size of the system increases, there is less and less interaction and in the limit the particles "become independent". That is, the initial chaotic configuration *propagates* over time. This intuition was put into firm mathematical ground notably by [48], [53]

[†]NYU Shanghai, China. E-mail: mathieu.lauriere@nyu.edu
[‡]Princeton University, USA. E-mail: ludovic.tangpi@princeton.edu

and [26] and has generated a rich literature with a variety of fundamental applications. We refer for instance to [47, 41, 7, 52, 35, 37, 36] for a few recent developments and applications. In particular, the theory of propagation of chaos has undoubtedly motivated (and benefited from) the more recent and very active theory of mean-field games introduced by [43] and [33].

The basic question motivating the present work is to ask whether Kac's intuition carries over to systems of particles with chaotic *terminal* configurations. There are numerous such examples, for instance in quantitative finance where different parties independently set investment goals which need to be met at a prescribed future date, but with inter-temporal trading decisions that are correlated. More precisely, we ask *whether a chaotic terminal configuration will propagate to past configurations* as the size of the system becomes large. As mentioned above, an important application at the origin of the theory of propagation of chaos is the particle system approximation of some nonlocal PDEs. We also analyze such an application in the present setting and use the backward propagation of chaos viewpoint to derive a particle system approximation of a semilinear PDE written on an infinite dimensional space (akin to the master equation in the theory of mean-field games). The interest here lies in the fact that, being written on a finite dimensional space, the approximating PDEs are much easier to handle analytically. For instance, well-developed theories of weak solutions and interior estimates for the gradients are available for such equations. The main idea leading to this approximation result is the probabilistic representation of solutions of some parabolic PDEs, especially due to [20], which allows us to transform the problem of approximating PDE solutions into a purely probabilistic question.

In the present paper, we model *backward particles* by solutions of backward stochastic differential equations (BSDEs) as introduced by [50]:

$$Y_t^{i,n} = G^i + \int_t^T F_u(Y_u^{i,n}, Z_u^{i,i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{Y_u^{j,n}})\, du - \sum_{k=1}^n \int_t^T Z_u^{i,k,n}\, dW_u^k, \quad i = 1,\ldots,n, \quad (1.1)$$

where $W^1,\ldots,W^n$ are independent Brownian motions and δ_x is the Dirac mass at a vector x. That is, the interaction is through the empirical distribution of the system. In the main contributions of the paper we derive various convergence results of the n-particle system to solutions of McKean-Vlasov BSDEs of the form

$$Y_t^i = G^i + \int_t^T F_u(Y_u^i, Z_u^i, \mathcal{L}(Y_u^i))\, du - \int_t^T Z_u^i\, dW_u^i, \quad \mathcal{L}(Y_t^i) = \mathrm{law}(Y_t^i) \qquad (1.2)$$

under classical Lipschitz continuity conditions on the generator and integrability conditions on the terminal value. The focus is put on deriving explicit, non-asymptotic convergence rates for the empirical measure as well as the processes. We strengthen our convergence results by deriving concentration inequalities, some of which dimension-free. All our results are gathered in the next section. The main result relies on an adaptation of the coupling technique of [53] and BSDE estimates on the one hand, and arguments from the theory of measure concentration on the other hand, notably results from [5].

To the best of our knowledge, only the papers of [10], [11] and [30] touch upon limit results for interacting backward particles. The first paper considers mean reflected BSDEs, with mean field interactions only through the reflection constraint. The authors proved propagation of chaos with convergence rates for the particle approximation. We shall come back later to the connection with some of their results. The other two papers consider a particular type of interaction, see Remark 2.10 and Example 2.14 for details. In [11], a convergence rate for the interacting processes to the McKean-Vlasov

equation is derived; we recover their result by a different argument based on functional inequalities for BSDEs. In [30] (where the term "backward propagation of chaos" is first coined) a convergence result for the empirical measure of the interacting particles is obtained. However, nothing is said concerning the rate of convergence. Another somewhat related article is the work by [8] on the approximation of BSDEs with normal constraints in law.

The ideas and results of the present paper are also connected to the theory of mean field games, which has recently attracted a surge of interest. In fact, BSDEs of mean-field type arise naturally in optimality conditions for mean field games (MFG) with interactions through the controls, which are sometimes referred to as "extended MFG" or "MFG of controls" and have been introduced by [27]. Such models are particularly relevant in economics and finance, cf. e.g. [19, 15]. The connection with mean-field BSDEs stems from Pontryagin's maximum principle and has been stressed by [16, Section 4.7.1] and more recently by [1]. We provide a more extensive discussion on the applications of our results to large population games and mean-field games in [44]. Another application of the limit theorems investigated here concern the study of particle systems with known terminal positions. We refer for instance to works on crowd motion with tagged pedestrians by Aurell and Djehiche in [4, 3]. Hereby, $Y^{i,n}$ represents the position of a given pedestrian (in a population of size n) who must be at position G^i at time T. Our results show that the path followed in the infinite population limit is given by (1.2).

Concerning the approximation problem of PDEs on the Wasserstein space by PDEs on finite dimensional Euclidean spaces, let us mention that a similar question was first analyzed by [14] (see also [42, 12]) based on PDE estimations they derive for the finite dimensional system. Their results concern the *quasilinear* form of the master equation. Our contribution here is mainly methodological, as we obtain a convergence result by purely probabilistic techniques. However, our setting also differs from that of [14, 12] in a number of ways, the most important difference being the type of nonlinearities in the measure argument that we consider.

In the rest of the paper, we dedicate Section 2 to the presentation of the precise setting of the work and its main results. The proofs are postponed to Section 3.

2 Setting and main results

2.1 Setting and notation

Let $d, m \in \mathbb{N}$ be fixed. Unless otherwise specified, $\mathbb{R}^d, \mathbb{R}^m$ and $\mathbb{R}^{d \times m}$ are endowed with the Euclidean norm denoted by $|\cdot|$ in all cases. Let us denote by $(\Omega, \mathcal{F}, P)$ a probability space carrying a sequence $(W^i)_{i \geq 1}$ of d-dimensional Brownian motions. As usual, equalities and inequalities between random variables will be understood to hold up to null sets of the Wiener measure P. Denote by $\mathcal{F}^n := (\mathcal{F}_t^n)_{t \in [0,T]}$ the completion of the raw filtration of $W^1, \ldots, W^n$. Let us equip Ω with the filtration $\mathcal{F}^n$. We will always use the identification

$$W \equiv W^1 \quad \text{and} \quad \mathcal{F} \equiv \mathcal{F}^1.$$

Given a vector $\boldsymbol{x} := (x_1, \ldots, x_n) \in (\mathbb{R}^m)^n$, denote by

$$L^n(\boldsymbol{x}) := \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$$

the empirical measure associated to $\boldsymbol{x}$. Then, $L^n(\boldsymbol{x}) \in \mathcal{P}_p(\mathbb{R}^m)$, the set of probability measures on $\mathbb{R}^m$ with finite p^{th} moment. Let us be given a function $F : [0, T] \times \Omega \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \times \mathcal{P}_2(\mathbb{R}^m) \to \mathbb{R}^m$, and a family of $\mathcal{F}_T$-measurable i.i.d. random variables $G^1, \ldots, G^n$. We are interested in the asymptotic behavior (as n becomes large) of a family of weakly

interacting processes $(Y^{1,n}, \cdots, Y^{n,n})$ evolving backward in time and given by

$$Y_t^{i,n} = G^i + \int_t^T F_u(Y_u^{i,n}, Z_u^{i,i,n}, L^n(\mathbf{Y}_u))\, du - \sum_{k=1}^n \int_t^T Z_u^{i,k,n}\, dW_u^k, \quad i = 1, \ldots, n, \quad (2.1)$$

where we used the notation $\mathbf{Y} := (Y^{1,n}, \ldots, Y^{n,n})$. Here as well as in the remainder of the article, we assume that for every $(y, z, \mu) \in \mathbb{R}^m \times \mathbb{R}^{m \times d} \times \mathcal{P}_2(\mathbb{R}^m)$ the stochastic process $F(\cdot, \cdot, y, z, \mu) : (t, \omega) \mapsto F(t, \omega, y, z, \mu)$ is progressively measurable. In analogy to weakly interacting particles evolving forward in time, in the limit, the above family will be intrinsically linked to the so-called McKean-Vlasov BSDE

$$Y_t^i = G^i + \int_t^T F_u(Y_u^i, Z_u^i, \mathcal{L}(Y_u^i))\, du - \int_t^T Z_u^i\, dW_u^i. \quad (2.2)$$

Hereby (and henceforth) $\mathcal{L}(X)$ denotes the law of the random variable X with respect to the probability measure P. Since under our assumptions on F and G^i the processes $(Y^i)_i$ will be i.i.d., we will often omit the superscript i and simply write $\mathcal{L}(Y)$ for the law of Y^i.

We equip the space $\mathcal{P}_p(\mathbb{R}^m)$ with the pth order Wasserstein distance denoted by $\mathcal{W}_p$ and defined as

$$\mathcal{W}_p(\mu, \nu) := \inf \left\{ \int_{\mathbb{R}^m \times \mathbb{R}^m} |x - y|^p\, d\pi \right\}^{1/p}$$

where the infimum is over probability measures π on $\mathbb{R}^m \times \mathbb{R}^m$ with first and second marginals μ and ν, respectively. Given $p \in [1, 2]$, we will often consider the condition

(Lip_p) The function F is L_F-Lipschitz continuous and of linear growth in the sense that there is a constant $L_F \geq 0$ such that,

$$|F_t(y, z, \mu) - F_t(y', z', \mu')| \leq L_F \left(|y - y'| + |z - z'| + \mathcal{W}_p(\mu, \mu')\right)$$

and

$$|F_t(y, z, \mu)| \leq L_F \left(1 + |y| + |z| + \left(\int_{\mathbb{R}^d} |x|^p\, d\mu\right)^{1/p}\right)$$

for all $t \in [0, T]$, $y, y' \in \mathbb{R}^m$, $z, z' \in \mathbb{R}^{m \times d}$ and $\mu, \mu' \in \mathcal{P}_p(\mathbb{R}^m)$.

Remark 2.1. Note at once that under condition (Lip_p), and if G^i has a finite second moment, i.e. $E[|G^i|^2] < \infty$, then the equations (2.1) and (2.2) admit unique, square integrable solutions. See the beginning of Section 3 for details.

Throughout, we denote by $\mathbf{Y}$ the value process of the solution of (2.1) and by Y that of (2.2), say with $i = 1$.

Having made precise the probabilistic setting governing the paper, let us now presents its main results. Most of them pertain to the limiting behavior of $Y^{i,n}$. As explained in the introduction, we also deduce approximation of parabolic PDEs on the Wasserstein space. The focus is put on quantitative (i.e. non-asymptotic) estimations of convergence rates. All proofs are postponed to Section 3.

2.2 Convergence of empirical distributions

We start by showing that the empirical distribution $L^n(\mathbf{Y}_t)$ of the system converges to the law $\mathcal{L}(Y_t)$ of the McKean-Vlasov BSDE.

Theorem 2.2. Let $p \in [1, 2]$. Assume that $E[|G^i|^k] < \infty$ for some k such that $k > p$ and $k \geq 2$, and that F satisfies (Lip_p). Then it holds that

$$E\left[\mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \leq C r_{n,m,q,p}, \qquad \forall t \in [0, T], \quad (2.3)$$

with

$$r_{n,m,q,p} := \begin{cases} n^{-1/2} + n^{-(q-p)/q}, & \text{if } p > m/2 \text{ and } q \neq 2p \\ n^{-1/2}\log(1+n) + n^{-(q-p)/q}, & \text{if } p = m/2 \text{ and } q \neq 2p \\ n^{-p/m} + n^{-(q-p)/q}, & \text{if } p \in (0, m/2) \text{ and } q \neq d/(m-p) \end{cases} \qquad (2.4)$$

for all $p < q < k$, and for some constant C depending on T, m, L_F, p, k and $E[|G^i|^k]$.

It is well-known that the Wasserstein topology is much stronger than the weak topology. Thus, Theorem 2.2 shows, in particular, that the sequence of (random) measures $(L^n(\mathbf{Y}_t))_n$ converges to the (deterministic) measure $\mathcal{L}(Y_t)$ in the weak topology. This can be seen as a type of quantitative law of large numbers. As a direct application we obtain the following strong law of large numbers for the sequence $Y^{i,n}$.

Remark 2.3. The reader may wonder whether the Lipschitz–continuity condition (Lip$_p$) on the generator F can be generalized to allow quadratic generators. The main issue here is that (2.1) is actually a system of coupled BSDEs and, as is well-known, the well-posedness of multi–dimensional quandratic BSDEs cannot be guaranteed in general. More delicate structural conditions need to be imposed. We refer for instance to [40, 21, 31, 38, 55, 46] for ample discussions on this issue.

However, if the condition (Lip$_p$) is replaced by

$$|F_t(y, z, \mu) - F_t(y', z', \mu')| \leq L_F \left(|y - y'| + (|z| + |z'|)|z - z'| + \mathcal{W}_p(\mu, \mu')\right), \quad p < 2$$

and the equations (2.1) and (2.2) admit unique, square integrable solutions, then the conclusion of Theorem 2.2 remains true, that is, we have

$$E\left[\mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \leq C r_{n,m,2,p}, \qquad \forall t \in [0, T].$$

The argument of proof is the same, it will be explained after the proof of Theorem 2.2, see Section 3.1.

Corollary 2.4. *Let $p \in [1, 2)$. Assume that $E[|G^i|^2] < \infty$ and that F satisfies (Lip$_p$). Then we have the $L^1(\Omega, P)$-limit*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_t^{i,n} = E[Y_t] \quad \text{for every } t.$$

Proof. By the Kantorovich-Rubinstein duality, we have

$$E\left[|\int_{\mathbb{R}^m} f(y) d L^n(\mathbf{Y}_t)(y) - \int_{\mathbb{R}^m} f(y) d\mathcal{L}(Y_t)(y)|\right] \leq E\left[\mathcal{W}_1(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \leq C r_{n,m,q,p}$$

for some $p < q < 2$ and for every 1-Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}$. In particular, taking $f(x) = x$ yields the result. $\square$

Remark 2.5. Under a stronger integrability condition, namely that $E[|G^i|^k] < \infty$ for some $k > m + 5$, the argument of the above theorem allows to obtain the bound

$$E\left[\mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \leq C n^{-p/(m+4)} \quad \text{for all } (t, n) \in [0, T] \times \mathbb{N} \qquad (2.5)$$

for some constant C depending only on T, m, L_F, G and $E[|G^i|^k]$.

The estimates (2.3) and (2.5) are uniform in time in the sense that the convergence rate is time-independent, but the supremum (in t) can be taken only outside the expectation on the left hand side. A stronger uniform estimate can be obtained at the cost of also stronger integrability conditions and a worse convergence rate.

Proposition 2.6. *Assume that $E[|G^i|^k] < \infty$ for some $k > m + 5$, and that F satisfies (Lip_p) for some $p \in [1,2]$. Further assume that the solution (Y, Z) of (2.2) is such that $\sup_{t \in [0,T]} E[|Z_t|^{2k}] < \infty$. Then it holds*

$$E\left[\sup_{t \in [0,T]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) \right] \leq C n^{-p/(m+8)} \quad \text{for all } n \in \mathbb{N}, \text{ and } p \in [1,2] \tag{2.6}$$

for some constant C depending on $T, L_F, k, E[|G^i|^k]$ and $\sup_{t \in [0,T]} E[|Z_t|^{2k}]$.

Remark 2.7. Compared with Theorem 2.2, the above statement relies on the assumption $\sup_t E[|Z_t|^{2k}] < \infty$. However, this is by no means a restrictive one, since it has been shown to hold in many classical cases. For instance, when $G^i = G(W_T^i)$ for a bounded and Lipschitz continuous function G, and $F_t(\cdot, \cdot, \mu)$ is differentiable for all (t, μ), then it holds $E[\sup_{t \in [0,T]} |Z_t|^{2k}] < \infty$ for all $k \geq 1$, see [34, Theorem 5.3]. Alternatively, under conditions on the Malliavin differentiability of G and F, it can be shown that Z is even bounded, see [22, 40] for details. The results of these papers apply for instance when G is Lipschitz continuous on the path space equipped with the supremum norm and F is deterministic. In this case, the integrability condition on G also follows. Furthermore, a similar condition is also used in [10, Theorem 4.3] to obtain rates of convergence, and the authors provided a sufficient condition as well as an example in [10, Lemma 4.4, and Example 4.5]. Lastly, in some cases, Z can also be bounded as a random variable, see [9].

2.3 Concentration estimates

Given two probability measures Q_1 and Q_2 on Ω, let us denote the pth order Wasserstein distance on Ω equipped with the supremum norm by

$$\mathcal{W}_{p,||\cdot||_\infty}(Q_1, Q_2) := \inf \left\{ \int_{\Omega \times \Omega} \sup_{t \in [0,T]} |\omega_1(t) - \omega_2(t)|^p \, d\pi(\omega_1, \omega_2) \right\}^{1/p}$$

where the infimum is over probability measures π on $\Omega \times \Omega$ with first and second marginals Q_1 and Q_2, respectively. The following result gives concentration estimates for the interacting family $\mathbf{Y}$. We consider concentration for the time t marginal as well as for the law of the entire process.

Theorem 2.8. *Let $p \in [1,2]$. Assume that $E[|G^i|^k] < \infty$ for some $k > 2p$, and that F satisfies (Lip_p). Then it holds that, for all $\varepsilon \in (0, \infty)$ and $\varepsilon_{F,T} := \varepsilon / \exp(Te^{L_F T})$,*

$$P\left(\mathcal{W}_p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) \geq \varepsilon \right) \leq C\left(a_{n, \varepsilon_{F,T}} 1_{\{\varepsilon_{F,T} \leq 1\}} + b_{n,k,\varepsilon_{F,T}} \right) \tag{2.7}$$

with $b_{n,k,\varepsilon} := n(n\varepsilon)^{-(k-\delta)/p}$ and

$$a_{n,\varepsilon} := \begin{cases} \exp(-cn\varepsilon^2), & \text{if } p > m/2 \\ \exp(-cn(\varepsilon/\log(2+1/x))^2), & \text{if } p = m/2 \\ \exp(-cn\varepsilon^{m/p}), & \text{if } p \in (0, m/2) \end{cases}$$

for three positive constants $\delta \in (0, k)$, C and c depending on p, m, k, T, L_F and $E[|G^i|^k]$.

Moreover, if the functions F_t and G^i are also Lipschitz continuous as functions on $(\Omega, ||\cdot||_\infty)$, that is,

$$|F_t(\omega, y, z, \mu) - F_t(\omega', y', z', \mu')| \leq L_F \left(||\omega - \omega'||_\infty + |y - y'| + |z - z'| + \mathcal{W}_p(\mu, \mu') \right)$$
$$\text{and} \quad |G^i(\omega) - G^i(\omega')| \leq L_G ||\omega - \omega'||_\infty,$$

then it holds that

$$P\left(\left|\mathcal{W}_{2,\|\cdot\|_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y))-E[\mathcal{W}_{2,\|\cdot\|_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y))]\right|\geq\varepsilon\right)\leq 2e^{-C\varepsilon^2 n} \tag{2.8}$$

for a constant C depending only on L_F, L_G and T.

If in addition F does not depend on z, then there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have

$$P\left(\mathcal{W}_{2,\|\cdot\|_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y))\geq\varepsilon\right)\leq e^{-C\varepsilon^2 n} \tag{2.9}$$

for some constant C depending on T, L_F, L_G, m and k.

The proof of Theorem 2.8 relies on quadratic transportation inequalities for BSDEs investigated in [5] and on standard results from the theory of concentration of measure, see Section 3.2.

2.4 Interacting particles approximation of McKean-Vlasov BSDE

This section is concerned with convergence of the sequence of stochastic processes $(Y^{i,n}, Z^{i,n})$ to the solutions of the McKean-Vlasov equation. These results will easily yield quantitative propagation of chaos results and have interesting applications in terms of PDEs.

Theorem 2.9. *Assume that $E[|G^1|^k] < \infty$ for some $k > 2$, and that F satisfies (Lip$_p$). Then it holds that*

$$E\left[\sup_{t\in[0,T]}|Y_t^{1,n}-Y_t^1|^2\right]+E\left[\int_0^T|Z_t^{1,1,n}-Z_t^1|^2\,dt\right]\leq Cr_{n,m,q,2} \quad \text{for all } (t,n)\in[0,T]\times\mathbb{N} \tag{2.10}$$

for all $q \in (2,k)$ and for some constant C depending on T, m, L_F, L_G and $E[|Y_t^1|^k]$ and $r_{n,m,q,2}$ is defined by (2.4).

Remark 2.10. The above result shows that in general, the sequences $(Y^{i,n})$ and $(Z^{i,i,n})$ converge at the same rate as $L^n(\mathbf{Y}_t)$. In the special case of particles in "linear" interaction, such a convergence result has been analyzed in [11]. More precisely, [11] considers the case when $\mathbf{Y} = (Y^{1,n}, \dots, Y^{n,n})$ solves the system

$$Y_t^{i,n}=G^i+\int_t^T\frac{1}{n}\sum_{j=1}^n f_u(Y_u^{i,n},Y_u^{j,n},Z_u^{i,i,n})\,du-\int_t^T Z_u^{i,n}\,dW_u \tag{2.11}$$

where W is a given Brownian motion, and $(G^1, \dots, G^n)$ are functions of the terminal values of a system of interacting (forward) particles. In this case, the rate of convergence of the n-particle system to the McKean-Vlasov equation can be improved and does not depend on the dimension. Interestingly, we can slightly generalize the result of [11] using different arguments. We consider the system

$$Y_t^{i,n}=G^i+\int_t^T F_u\left(Y_u^{i,n},Z_u^{i,i,n},\frac{1}{n}\sum_{j=1}^n f_u(Y_u^{i,n},Y_u^{j,n},Z_u^{i,n})\right)du-\int_t^T\sum_{j=1}^n Z_u^{i,j,n}\,dW_u^j \tag{2.12}$$

that often appears in applications, see e.g. [32, 6, 18] for linear-quadratic mean-field models and [30] for a contract theory problem. We obtain the usually optimal rate $1/\sqrt{n}$ for this more general system.

In fact, consider the McKean-Vlasov equation

$$Y_t^i=G^i+\int_t^T F_u\left(Y_u^i,Z_u^i,\int_{\mathbb{R}^m}f_u(Y_u^i,y,Z_u^i)\,d\mathcal{L}(Y_u^i)(y)\right)du-\int_t^T Z_u^i\,dW_u^i \tag{2.13}$$

and the following Lipschitz continuity and linear growth conditions

(Lip) The functions $F : [0,T] \times \Omega \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \times \mathbb{R}^m \to \mathbb{R}^m$ and $f : [0,T] \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \to \mathbb{R}^m$ are respectively L_F-Lipschitz and L_f-Lipschitz continuous and of linear growth in (y, z, a) and (y_1, y_2, z) uniformly with respect to (t, ω) and t respectively. That is, there are constants $L_F, L_f \geq 0$ such that

$$|F_t(\omega, y, z, a) - F_t(\omega', y', z', a')| \leq L_F \left(\|\omega - \omega'\|_\infty + |y - y'| + |z - z'| + |a - a'| \right)$$

$$|F_t(y, z, a)| \leq L_F \left(1 + |y| + |z| + |a| \right)$$

and

$$|f_t(y_1, y_2, z) - f_t(y_1', y_2', z')| \leq L_f \left(|y_1 - y_1'| + |y_2 - y_2'| + |z - z'| \right)$$

$$|f_t(y, y_1, z)| \leq L_f \left(1 + |y| + |y_1| + |z| \right)$$

for all $t \in [0,T]$, $a, a', y, y', y_1, y_2, y_1', y_2' \in \mathbb{R}^m$, $z, z' \in \mathbb{R}^{m \times d}$.

Remark 2.11. Note that under (Lip), if G^i has a second moment, then both equations (2.12) and (2.13) admits unique, square integrable solutions.

Proposition 2.12. *Assume that $E[|G^i|^2] < \infty$ and is Lipschitz continuous with respect to the uniform norm on Ω, and that the functions F and f satisfy (Lip). The respective solutions $(Y^{i,n}, Z^{i,j,n})$ and (Y^i, Z^i) of the equations (2.12) and (2.13) satisfy*

$$E\left[|Y_t^{1,n} - Y_t^1|^2 \right] + E\left[\int_0^T |Z_t^{1,n} - Z_t^1|^2 \, dt \right] \leq C n^{-1} \quad \text{for all } (t, n) \in [0,T] \times \mathbb{N} \quad (2.14)$$

for some constant C depending only on T, L_F, L_f and L_G.

Direct consequences of Theorem 2.9 and Proposition 2.12 are the following quantitative propagations of chaos.

Corollary 2.13. *Put $\theta^{k,n} := Law(Y^{1,n}, \ldots, Y^{k,n})$ and let $\mathcal{L}(Y)^{\otimes k}$ be the k-fold product of the law $\mathcal{L}(Y^1)$ of Y^1, solution of the McKean-Vlasov BSDE (2.2). If $(Y^{i,n}, Z^{i,j,n})$ and (Y^i, Z^i) solve (2.1) and (2.2) respectively, then under the conditions of Theorem 2.2, we have, for all $n \in \mathbb{N}$ and all $k \leq n$, that*

$$\begin{cases} \mathcal{W}_{2,||\cdot||_\infty}^2 (\mathcal{L}(Y^{1,n}), \mathcal{L}(Y^1)) \leq C r_{n,m,q,p} \\ \mathcal{W}_{2,||\cdot||_\infty}^2 (\theta^{k,n}, \mathcal{L}(Y)^{\otimes k}) \leq k C r_{n,m,q,p} \end{cases} \quad (2.15)$$

for some constant C depending on T, L_F, L_G and m.

If $(Y^{i,n}, Z^{i,j,n})$ and (Y^i, Z^i) solve (2.12) and (2.13) respectively, then under the conditions of Proposition 2.12 we have, for all $(t, n) \in [0,T] \times \mathbb{N}$ and all $k \leq n$, that

$$\begin{cases} \mathcal{W}_2^2 (\mathcal{L}(Y_t^{1,n}), \mathcal{L}(Y_t^1)) \leq C n^{-1} \\ \mathcal{W}_2^2 (\theta_t^{k,n}, \mathcal{L}(Y_t)^{\otimes k}) \leq k C n^{-1} \end{cases} \quad (2.16)$$

for some constant C depending on T, L_F, L_G, L_f and m.

Proof. Since $\mathcal{L}(Y_t^1) = \mathcal{L}(Y_t^i)$, it follows by definition that

$$\mathcal{W}_{2,||\cdot||_\infty}^2 (\mathcal{L}(Y^{1,n}), \mathcal{L}(Y^1)) \leq E\left[\sup_{t \in [0,T]} |Y_t^{1,n} - Y_t^1|^2 \right]$$

and

$$\mathcal{W}_{2,||\cdot||_\infty}^2 (\theta^{k,n}, \mathcal{L}(Y)^{\otimes k}) \leq E\left[\sum_{i=1}^k \sup_{t \in [0,T]} |Y_t^{i,n} - Y_t^i|^2 \right].$$

Thus, (2.15) follows by (2.10). Similarly (2.16) follows by (2.14). $\square$

Example 2.14 (Convolution interaction). In relation to a principal-agent problem of mean-field type, [30] investigated the case of the generator $F_t(y, z, \mu) := \varphi * \mu(y)$ for some function $\varphi : \mathbb{R}^m \to \mathbb{R}^m$, where the convolution $\varphi * \mu$ is defined as $\varphi * \mu(x) := \int_{\mathbb{R}^m} \varphi(x - y) \, d\mu(y)$. This case falls within the scope of Proposition 2.12 (with $F(t, y, z, a) = a$ and $f(y, y', z) = \varphi(y - y')$), and Corollary 2.13 additionally gives a sharp convergence rate.

2.5 Finite dimensional approximation of parabolic PDEs on the Wasserstein space

In this subsection, we assume that F does not depend on (t, ω). Given four functions $B : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d \times d}$, $G : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^m$ and $F : \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m \times d} \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^m) \to \mathbb{R}^m$, we consider the PDE

$$
\begin{cases}
\partial_t V(t, x, \mu) + B(x, \mu) \partial_x V(t, x, \mu) + \frac{1}{2} \mathrm{tr}(\partial_{xx} V(t, x, \mu) a(x, \mu)) \\
\quad + F(x, V(t, x, \mu), \sigma'(x, \mu) \partial_x V(t, x, \mu), \mu, \nu) \\
\quad + \int_{\mathbb{R}^d} \partial_\mu V(t, x, \mu)(y) \cdot B(y, \mu) d\mu(y) + \int_{\mathbb{R}^d} \frac{1}{2} \mathrm{tr} \left(\partial_y \partial_\mu V(t, x, \mu)(y) a(y, \mu) \right) d\mu(y) = 0 \\
V(T, x, \mu) = G(x, \mu)
\end{cases}
$$

(2.17)

with $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, $a := \sigma\sigma'$, and ν the law of $V(t, \xi, \mu)$ when $\mathcal{L}(\xi) = \mu$. The derivative

$$\partial_\mu V(t, x, \mu)(y)$$

denotes the so-called Wasserstein derivative (also called L-derivative) of the function V in the direction of the probability measure μ, see e.g. [2, 45] or [16, Chapter 5] for details. The goal of this section is to show that the solution V of the PDE (2.17), written on the infinite dimensional space $[0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ can be approximated by a sequence of solutions of PDEs written on the finite dimensional space $[0, T] \times (\mathbb{R}^d)^n$. More precisely, we will be interested in the system of PDEs

$$
\begin{cases}
\partial_t v^{i,n}(t, \boldsymbol{x}) + B(x_i, L^n(\boldsymbol{x})) \partial_{x_i} v^{i,n}(t, \boldsymbol{x}) + \frac{1}{2} \mathrm{tr} \left(\partial_{x_i x_i} v^{i,n}(t, \boldsymbol{x}) a(x_i, L^n(\boldsymbol{x})) \right) \\
\quad + F \left(x_i, v^{i,n}(t, \boldsymbol{x}), \partial_{x_i} v^{i,n}(t, \boldsymbol{x}), L^n(\boldsymbol{x}) \sigma(x_i, L^n(\boldsymbol{x})), \frac{1}{n} \sum_{j=1}^n \delta_{v^{j,n}(t, \boldsymbol{x})} \right) = 0 \\
\quad \text{with } (t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^n \\
v^{i,n}(T, \boldsymbol{x}) = G(x_i, L^n(\boldsymbol{x})), \quad \boldsymbol{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n \\
i = 1, \dots, n.
\end{cases}
$$

(2.18)

The following condition is copied almost verbatim from [20]. It guarantees the existence of a unique classical solution V to (2.17).

(PDE) The functions σ, B, F and G satisfy the following:

   (PDE1) The function σ is bounded, and the functions B, σ, F and G are three times continuously differentiable in $w = (x, y, z)$ and μ, with bounded and Lipschitz-continuous first and second derivatives (with common bound and Lipschitz constant denoted L_F).

   (PDE2) There exist a constant $\alpha \geq 0$ and a function $\Phi_\alpha : (L^2(\Omega, \mathcal{F}_T, P; \mathbb{R}^{d+m}))^2 \ni (\chi, \chi') \mapsto \Phi_\alpha(\chi, \chi') \in \mathbb{R}_+$ continuous at any point (χ, χ) of the diagonal and such that

$$\Phi_\alpha(\chi, \chi') \leq E \left[\left(1 + |\chi|^{2\alpha} + |\chi'|^{2\alpha} + ||\chi||_2^{2\alpha} \right) |\chi - \chi'| \right]^{1/2}$$

for all $\chi, \chi \in L^2(\Omega, \mathcal{F}_T, P; \mathbb{R}^{d+m})$ satisfying $\mathcal{L}(\chi) = \mathcal{L}(\chi')$. Moreover, letting $h = B, \sigma, F$ or G, it holds

$$\begin{cases} |\partial_w h(w, \mathcal{L}(\chi)) - \partial_w h(w', \mathcal{L}(\chi'))| \le L_F\Big(|w - w'| + \Phi_\alpha(\chi, \chi')\Big) \\ |\partial_\mu h(w, \chi) - \partial_\mu h(w', \chi')| \le L_F\Big(|w - w'| + \Phi_\alpha(\chi, \chi')\Big) \end{cases}$$

for all $w = (x, y, z), w' = (x', y', z')$ and $\chi, \chi' \in L^2(\Omega, \mathcal{F}_T, P; \mathbb{R}^{d+m})$. Furthermore, for every $\chi \in L^2(\Omega, \mathcal{F}_T, P; \mathbb{R}^{d+m})$, the family $(\partial_\mu h(w, \chi))_w$ is uniformly integrable.

(PDE3) Letting $h = B, \sigma, F$ or G, the mapping $v \mapsto \partial_\mu h(w, \mu)(v)$ is L_F-Lipschitz continuous, the mapping $(w, v) \mapsto \partial_\mu h(w, \mu)(v)$ is continuously differentiable at any point (w, v) such that v is in the support of μ, $(w, v) \mapsto \partial_v[\partial_\mu h(w, \mu)](v)$ and $(w, v) \mapsto \partial_w[\partial_\mu h(w, \mu)](v)$ are continuous and it holds

$$E\big[|\partial_w[\partial_\mu h(w, \mathcal{L}(\chi))](\chi) - \partial_w[\partial_\mu h(w', \mathcal{L}(\chi'))](\chi')|^2\big]^{1/2}$$
$$+ E\big[|\partial_v[\partial_\mu h(w, \mathcal{L}(\chi))](\chi) - \partial_v[\partial_\mu h(w', \mathcal{L}(\chi'))](\chi')|^2\big]^{1/2}$$
$$\le L_F\Big(|w - w'| + \Phi_\alpha(\chi, \chi')\Big).$$

Under the condition (PDE), we then have the announced convergence of $v^{1,n}$ to V. More precisely, we have:

Theorem 2.15. *Assume that F does not depend on (t, ω) and that the condition (PDE) is satisfied. Then the sequence $(v^{1,n})_n$ converges to V in the sense that for every i.i.d. sequence $(\xi_i)_{i \in \mathbb{N}}$ in $L^k(\Omega, \mathcal{F}_t, P; \mathbb{R}^d)$ for some $k > 4$ and every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mathcal{L}(\xi_1) = \mu$, it holds that*

$$E\Big[\sup_{t \in [0,T]} |v^{1,n}(t, \xi_1, \ldots, \xi_n) - V(t, \xi_1, \mu)|^2\Big] \le C_{L_F, T, k}\, \varepsilon_n \qquad (2.19)$$

where ε_n is defined as

$$\varepsilon_n = \begin{cases} n^{-1/2}, & \text{if } d < 4, \\ n^{-1/2}\log(n), & \text{if } d = 4, \\ n^{-2/d}, & \text{if } d > 4 \end{cases} \qquad (2.20)$$

and $C_{L_F, T, k}$ is a constant depending on L_F, T and $E[|\xi_1|^k]$.

Moreover, for every $n \in \mathbb{N}$ and every $t \in [0, T]$ it holds that

$$E\Big[|v^{i,n}(t, \xi_1, \ldots, \xi_n) - V(t, \xi_i, L^n(\boldsymbol{\xi}))|^2\Big] \le C_{L_F, T}(\varepsilon_n + r_{n,d,k,2}) \qquad (2.21)$$

with $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_n)$, where $C_{L_F, T}$ depends on the Lipschitz constant L_F of B, F and G and on T, and $r_{n,d,k,2}$ is defined by (2.4).

3 Proofs

In this final section we give detailed proofs of the results presented above. We start by justifying well-posedness of (2.2) and (2.1). For simplicity of notation, we will put

$$Z^{i,n} := Z^{i,i,n}$$

whenever this does not cause confusions.

Proof of Remark 2.1. By [16, Theorem 4.23], the equation (2.2) admits a unique solution (Y, Z) solution in the space $\mathcal{S}^2(\mathbb{R}^m) \times \mathcal{H}^2(\mathbb{R}^{m \times d})$, where we use the notation: for each integer $k \ge 1$,

$$\mathcal{S}^2(\mathbb{R}^k) := \Big\{ Y \in \mathbb{H}^0(\mathbb{R}^k) \,\Big|\, E \sup_{0 \le t \le T} |Y_t|^2 < +\infty \Big\},$$

and

$$\mathcal{H}^2(\mathbb{R}^k) := \left\{ Z \in \mathbb{H}^0(\mathbb{R}^k) \,\Big|\, E \int_0^T |Z_t|^2 dt < +\infty \right\},$$

with $\mathcal{H}^0(\mathbb{R}^k)$ being the space of all $\mathbb{R}^k$-valued progressively measurable processes.

Moreover, we can apply [24, Theorem 2.1] to the system (2.1) to justify the existence and uniqueness of a solution. To that end, it helps to write it in the more compact form

$$\mathbf{Y}_t = \mathbf{G} + \int_t^T \mathbf{F}_u(\mathbf{Y}_u, \mathbf{Z}_u)\, du - \int_t^T \mathbf{Z}_u\, d\mathbf{W}_u, \tag{3.1}$$

where $\mathbf{G}(\boldsymbol{\omega}) := (G^1(\omega^1), \dots, G^n(\omega^n))$, $\boldsymbol{\omega} := (\omega^1, \dots, \omega^n)$,

$$\mathbf{Y} := (Y^{i,n})_{i=1,\dots,n}, \qquad \mathbf{Z}^{i,n} := (Z^{i,1,n}, \dots, Z^{i,n,n}), \qquad \mathbf{W} = (W^i)_{i=1,\dots,n}$$

$\mathbf{Z} := \operatorname{diag}(\mathbf{Z}^{i,n})_{i=1,\dots,n}$ and $\mathbf{F} : [0,T] \times \Omega^n \times (\mathbb{R}^m)^n \times (\mathbb{R}^{m\times d})^n \to (\mathbb{R}^m)^n$ is defined by

$$\mathbf{F}(t, \boldsymbol{\omega}, \boldsymbol{y}, \boldsymbol{z}) = (F_t(\omega^i, y^i, z^{i,i}, L^n(\boldsymbol{y})))_{i=1,\dots,n}$$

for $t \in [0,T], \boldsymbol{y} = (y^1, \dots, y^n) \in (\mathbb{R}^m)^n, \boldsymbol{z} = (z^1, \dots, z^n) \in (\mathbb{R}^{m\times d})^n$ and $\boldsymbol{\omega} := (\omega^1, \dots, \omega^n) \in \Omega^n$. Now it suffices to check that the function $\mathbf{F}$ is $(8L_F)$-Lipschitz continuous (the Lipschitz constant does not depend on n). We do this for the reader's convenience. For every $t \in [0,T], \boldsymbol{y}_1, \boldsymbol{y}_2 \in (\mathbb{R}^m)^n, \boldsymbol{z}_1, \boldsymbol{z}_2 \in (\mathbb{R}^{m\times d})^n$ it holds that

$$
\begin{aligned}
|\mathbf{F}(t, \boldsymbol{y}_1, \boldsymbol{z}_1) - \mathbf{F}(t, \boldsymbol{y}_2, \boldsymbol{z}_2)|^2 &= \sum_{i=1}^n |F_t(y_1^i, z_1^{i,i}, L^n(\boldsymbol{y}_1)) - F_t(y_2^i, z_2^{i,i}, L^n(\boldsymbol{y}_2))|^2 \\
&\leq L_F \sum_{i=1}^n \left(|y_1^i - y_2^i| + |z_1^{i,i} - z_2^{i,i}| + \mathcal{W}_p(L^n(\boldsymbol{y}_1), L^n(\boldsymbol{y}_2)) \right)^2 \\
&\leq L_F \sum_{i=1}^n \left(|y_1^i - y_2^i| + |z_1^{i,i} - z_2^{i,i}| + \mathcal{W}_2(L^n(\boldsymbol{y}_1), L^n(\boldsymbol{y}_2)) \right)^2 \\
&\leq 4L_F \sum_{i=1}^n \left(|y_1^i - y_2^i|^2 + |z_1^{i,i} - z_2^{i,i}|^2 + \frac{1}{n} \sum_{j=1}^n |y_1^j - y_2^j|^2 \right) \\
&\leq 8L_F \left(|\boldsymbol{y}_1 - \boldsymbol{y}_2|^2 + |\boldsymbol{z}_1 - \boldsymbol{z}_2|^2 \right).
\end{aligned}
$$

To derive these inequalities, we successively used assumption (Lip_p), the fact that $\mathcal{W}_p(\mu,\nu) \leq \mathcal{W}_2(\mu,\nu)$ for any $p \in [1,2]$, and the fact that $\mathcal{W}_2$, for two n-sample empirical distributions $L^n(\boldsymbol{y}_1), L^n(\boldsymbol{y}_2)$ is given by

$$\mathcal{W}_2(L^n(\boldsymbol{y}_1), L^n(\boldsymbol{y}_2)) = \min_\sigma \left(\frac{1}{n} \sum_{i=1}^n |y_1^{\sigma(i)} - y_2^i|^2 \right)^{1/2},$$

where the minimum is over permutations σ of $\{1, \dots, n\}$, see e.g. [13, Lemma 5.1.7]. $\qquad\square$

3.1 Proofs for Subsection 2.2

We begin with two moment estimates for the solution of the McKean-Vlasov BSDE. Given a square integrable progressive process q, we will denote by

$$\mathcal{E}_{s,t}(q \cdot W) := \exp\left(\int_s^t q_u\, dW_u - \frac{1}{2} \int_s^t |q_u|^2\, du \right)$$

the stochastic exponential of q, and for every measure $\mu \in \mathcal{P}_p(\mathbb{R}^m)$, we let

$$M_p(\mu) := \int_{\mathbb{R}^m} |x|^p\, d\mu$$

be the pth-moment of μ.

Lemma 3.1. *Let $k \geq 2$. Assume $E[|G|^k] < \infty$ and that F satisfies (Lip$_p$) for some $p \in [1, 2]$. Then, for all $q \geq 0$, if $q = 2$ or $q < k$, the solution (Y, Z) of (2.2) satisfies*

$$\sup_t E[|Y_t|^q] < \infty.$$

Proof. If $q = 2$, there is nothing to prove because the result follows from Remark 2.1. Let us assume $q < k$. Since F is Lipschitz continuous in the z-variable, it is almost everywhere differentiable. Thus, it follows from the mean-value theorem that

$$Y_t = G + \int_t^T \left(F_u(Y_u, Z_u, \mathcal{L}(Y_u)) - F_u(Y_u, 0, \mathcal{L}(Y_u)) + F_u(Y_u, 0, \mathcal{L}(Y_u)) \right) du - \int_t^T Z_u \, dW_u$$

$$= G + \int_t^T \left(\int_0^1 \partial_z F_u(Y_u, \lambda Z_u, \mathcal{L}(Y_u)) \, d\lambda Z_u + F_u(Y_u, 0, \mathcal{L}(Y_u)) \right) du - \int_t^T Z_u \, dW_u$$

$$= E_Q \left[G + \int_t^T F_u(Y_u, 0, \mathcal{L}(Y_u)) \, du \mid \mathcal{F}_t \right]$$

where we used Girsanov's theorem, with Q being the probability measure with density $dQ/dP := \mathcal{E}_{0,T}(\gamma \cdot W)$ and

$$\gamma_u := \int_0^1 \partial_z F_u(Y_u, \lambda Z_u, \mathcal{L}(Y_u)) \, d\lambda.$$

Hence, using the linear growth of F we obtain

$$|Y_t|^q \leq C_{F,k,q} E_Q \left[|G|^q + \int_t^T \left(1 + |Y_u|^q + M_2^{q/2}(\mathcal{L}(Y_u)) \right) du \mid \mathcal{F}_t \right]$$

for a constant $C_{F,k,q}$ depending only on k and F. Thus, by Gronwall's inequality we have

$$|Y_t|^q \leq e^{C_{F,k,q}T} C_{F,k,q} E_Q \left[|G|^q + \sup_{u \in [0,T]} M_2(\mathcal{L}(Y_u))^{q/2} + T \mid \mathcal{F}_t \right]$$

$$\leq e^{C_{F,k,q}T} C_{F,k,q} E \left[\mathcal{E}_{t,T}(\gamma \cdot W)^{k/(k-q)} \mid \mathcal{F}_t \right]^{(k-q)/k} E \left[|G|^k \mid \mathcal{F}_t \right]^{q/k}$$

$$+ C_{F,k,q}(T + \sup_{u \in [0,T]} M_2(\mathcal{L}(Y_u))^{q/2}).$$

Note that, since γ is a bounded process, the random variable $\mathcal{E}_{t,T}(\gamma \cdot W)$ has moments of all orders. Furthermore, $\sup_{u \in [0,T]} M_2(\mathcal{L}(Y_u)) < \infty$ since $Y \in \mathcal{S}^2(\mathbb{R}^m)$, see Remark 2.1. Therefore, taking expectation on both sides and applying again Hölder's inequality concludes the argument. □

Lemma 3.2. *Let $k \geq 3$. Assume that $E[|G|^{k'}] < \infty$ for some $k' > 2k$ and that F satisfies (Lip$_p$) for some $p \in [1, 2]$. Assume that the solution (Y, Z) of (2.2) is such that $\sup_t E[|Z_t|^{2k}] < \infty$. Let $0 \leq t_1 \leq t_2 \leq T$ be such that $t_2 - t_1 \leq 1$. Then we have*

(i) $E[|Y_t - Y_s|^k |Y_s - Y_r|^k] \leq C|t - r|^2$ *for some $C \geq 0$ and for all $t_1 \leq r < s < t \leq t_2$.*

(ii) $E[|Y_t - Y_s|^k] \leq C|t - s|$ *for some $C \geq 0$ and for every $t_1 \leq s \leq t \leq t_2$.*

Proof. Let us start with the proof of (i). A direct estimation and repeated applications of

Hölder's inequality yield

$$E\left[|Y_t - Y_s|^k |Y_s - Y_r|^k\right]$$

$$\leq 2^{2k-2}(s-r)^{(k-1)}(t-s)^{(k-1)}E\left[\int_r^s |F_u(Y_u, Z_u, \mathcal{L}(Y_u))|^k\, du \int_s^t |F_u(Y_u, Z_u, \mathcal{L}(Y_u))|^k\, du\right]$$

$$+ 2^{2k-2}(s-r)^{(k-1)}E\left[\int_r^s |F_u(Y_u, Z_u, \mathcal{L}(Y_u))|^k\, du \left|\int_s^t Z_u\, dW_u\right|^k\right]$$

$$+ 2^{2k-2}(t-s)^{(k-1)}E\left[\int_s^t |F_u(Y_u, Z_u, \mathcal{L}(Y_u))|^k\, du \left|\int_r^s Z_u\, dW_u\right|^k\right]$$

$$+ 2^{2k-2}E\left[\left|\int_r^s Z_u\, dW_u\right|^{2k}\right]^{1/2} E\left[\left|\int_s^t Z_u\, dW_u\right|^{2k}\right]^{1/2}.$$

Now, recall that by Lemma 3.1, it holds $\sup_{t\in[0,T]} E[|Y_t|^{2k}] < \infty$, $\sup_{t\in[0,T]} M_2(\mathcal{L}(Y_t)) < \infty$ and by assumption, that $\sup_t E[|Z_t|^{2k}] < \infty$. Thus, by the linear growth condition on F and Burkholder-Davis-Gundy inequality, we have

$$E\left[|Y_t - Y_s|^k |Y_s - Y_r|^k\right]$$

$$\leq 2^{2k} L_F T(t-r)^{2(k-1)} E\left[\int_0^T \left(1 + |Y_u|^{2k} + |Z_u|^{2k} + M_2^{2k}(\mathcal{L}(Y_u))\right)\, du\right]$$

$$+ 2^{2k} L_F(t-r)^{(k-1)} E\left[\int_r^s \left(1 + |Y_u|^{2k} + |Z_u|^{2k} + M_2^{2k}(\mathcal{L}(Y_u))\, du\right)\right]^{1/2} E\left[\left(\int_s^t |Z_u|^2\, du\right)^k\right]^{1/2}$$

$$+ 2^{2k} L_F(t-r)^{(k-1)} E\left[\int_s^t \left(1 + |Y_u|^{2k} + |Z_u|^{2k} + M_2^{2k}(\mathcal{L}(Y_u))\, du\right)\right]^{1/2} E\left[\left(\int_r^s |Z_u|^2\, du\right)^k\right]^{1/2}$$

$$+ 2^{2k} E\left[\left(\int_r^s |Z_u|^2\, du\right)^k\right]^{1/2} E\left[\left(\int_s^t |Z_u|^2\, du\right)^k\right]^{1/2}$$

$$\leq 2^{2k} C_{F,T}(t-r)^{2(k-1)} + 2^{2k} C_{F,T}(t-r)^{(k-1)}(s-t)^{k-1} + 2^{2k} C_{F,T}(t-r)^{(k-1)}(t-s)^{k-1}$$

$$+ C_{F,T} 2^{2k} \left[(s-r)^{k-1}(t-s)^{k-1}\right]^{1/2}$$

$$\leq C_{F,T}(t-r)^{k-1}.$$

Since $k \geq 3$ and $t - r \leq 1$, we can conclude from the above that

$$E\left[|Y_t - Y_s|^k |Y_s - Y_r|^k\right] \leq C_{k,T,F}|t-r|^2,$$

where $C_{k,T,F}$ is a constant depending on k, T, L_F and the $2k^{th}$ moments of Y and Z. This proves the first claim.

Let us turn to the proof of the second claim, which is similar (and simpler). In fact, arguing as above we get

$$E\left[|Y_t - Y_s|^k\right] \leq 2^{k-1} E\left[\left|\int_s^t F_u(Y_u, Z_u, \mathcal{L}(Y_u))\, du\right|^k + \left|\int_s^t Z_u\, dW_u\right|^k\right]$$

$$\leq 2^k(t-s)^{(k-1)} E\left[\int_s^t |F_u(Y_u, Z_u, \mathcal{L}(Y_u))|^k\, du\right] + 2^k E\left[\left(\int_s^t |Z_u|^2\, du\right)^{k/2}\right]$$

$$\leq 2^k L_F(t-s)^{(k-1)} + 2^k C(t-s)^{(k-1)/2} \leq C_{F,T}|t-s|,$$

where we used the facts that $t - s \leq 1$ and $k \geq 3$.

In the case that F is bounded in z the argument is exactly the same. Since the terms $|Z_t|^{2k}$ will not appear in the estimates we can conclude without assumptions on the moments of Z. $\qquad\square$

Next, we will adapt to BSDEs a well-known coupling technique that will allow to use some known quantitative bounds for i.i.d. samples in our interacting particles case. This coupling technique, which probably originated from the work of [53], is by now standard in SDE theory, see e.g. [23, 14] for recent references. Hence, let (Y, Z) be the solution of the McKean-Vlasov BSDE (2.2). Let $(\tilde{Y}^1, \tilde{Z}^1), \dots (\tilde{Y}^n, \tilde{Z}^n)$ be i.i.d. copies of (Y, Z) such that for each i, $(\tilde{Y}^i, \tilde{Z}^i)$ solves the equation

$$\tilde{Y}_t^i = G^i + \int_t^T F_u(\tilde{Y}_u^i, \tilde{Z}_u^i, \mathcal{L}(Y_u)) \, du - \int_t^T \tilde{Z}_u^i \, dW_u^i. \tag{3.2}$$

Such copies can be found because the McKean-Vlasov BSDE has a unique solution, and thus we have uniqueness in law. We let $\tilde{\mathbf{Y}} = (\tilde{Y}^1, \dots, \tilde{Y}^n)$.

Lemma 3.3. *Let $p \in [1, 2]$. Assume that $E[|G^1|^2] < \infty$ and that F satisfies (Lip_p). Then it holds that*

$$\mathcal{W}_p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) \leq \exp(T e^{L_F T}) \mathcal{W}_p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t)) \quad P\text{-a.s.} \tag{3.3}$$

for all $(t, n) \in [0, T] \times \mathbb{N}$.

Proof. Let $i \in \{1, \dots, n\}$ be fixed. It follows by the mean-value theorem that

$$
\begin{aligned}
&\tilde{Y}_t^i - Y_t^{i,n} \\
&= \int_t^T \left[F_u(\tilde{Y}_u^i, \tilde{Z}_u^i, \mathcal{L}(Y_u)) - F_u(Y_u^{i,n}, Z_u^{i,i,n}, L^n(\mathbf{Y}_u)) \right] du \\
&\quad - \sum_{j=1}^n \int_t^T \left(\delta_{ij} \tilde{Z}_u^i - Z_u^{i,j,n} \right) dW_u^j \\
&= \int_t^T \left[\alpha_u^i + \beta_u^i + \int_0^1 \partial_z F_u(Y_u^{i,n}, Z_u^{i,i,n} + \lambda(\tilde{Z}_u^i - Z_u^{i,i,n}), \mathcal{L}(Y_u)) \, d\lambda(\tilde{Z}_u^i - Z_u^{i,i,n}) \right] du \\
&\quad - \sum_{j=1}^n \int_t^T \left(\delta_{ij} \tilde{Z}_u^i - Z_u^{i,j,n} \right) dW_u^j
\end{aligned}
$$

with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, $\alpha_u^i := F_u(\tilde{Y}_u^i, \tilde{Z}_u^i, \mathcal{L}(Y_u)) - F_u(Y_u^{i,n}, \tilde{Z}_u^i, \mathcal{L}(Y_u))$ and $\beta_u^i := F_u(Y_u^{i,n}, Z_u^{i,i,n}, \mathcal{L}(Y_u)) - F_u(Y_u^{i,n}, Z_u^{i,i,n}, L^n(\mathbf{Y}_u))$. Note that since F is Lipschitz continuous, the derivative $\partial_z F$ can be defined almost everywhere, and is bounded. Thus, the density process

$$\mathcal{E}_{0,t}(\gamma \cdot W^i) \quad \text{with} \quad \gamma_u := \int_0^1 \partial_z F_u(Y_u^{i,n}, Z_u^{i,i,n} + \lambda(\tilde{Z}_u^i - Z_u^{i,n}), \mathcal{L}(Y_u)) \, d\lambda$$

defines an equivalent probability measure Q. Due to Girsanov's theorem and square integrability of $Z^{i,j,n}$ and $\tilde{Z}^i$, taking the expectation above with respect to Q yields

$$\tilde{Y}_t^i - Y_t^{i,n} = E_Q \left[\int_t^T (\alpha_u^i + \beta_u^i) \, du \mid \mathcal{F}_t^n \right]. \tag{3.4}$$

Again by Lipschitz continuity of F, it holds that

$$|\alpha_u^i| + |\beta_u^i| \leq L_F \left(|\tilde{Y}_u^i - Y_u^{i,n}| + \mathcal{W}_p(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u)) \right)$$

so that by Gronwall's inequality, we have

$$|\tilde{Y}_t^i - Y_t^{i,n}| \le e^{L_F T} E_Q\left[\int_0^T \mathcal{W}_p(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u))\, du \mid \mathcal{F}_t^n\right].$$

Hence, using the definition of the pth-Wasserstein distance, we obtain the estimate

$$\mathcal{W}_p(L^n(\mathbf{Y}_t), L^n(\tilde{\mathbf{Y}}_t)) \le \left(\frac{1}{n}\sum_{i=1}^n |\tilde{Y}_t^i - Y_t^i|^p\right)^{1/p}$$

$$\le e^{L_F T} E_Q\left[\int_0^T \mathcal{W}_p(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u))\, du \mid \mathcal{F}_t^n\right].$$

Now, combine this with the triangle inequality to obtain

$$\mathcal{W}_p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) \le \mathcal{W}_p(L^n(\mathbf{Y}_t), L^n(\tilde{\mathbf{Y}}_t)) + \mathcal{W}_p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))$$

$$\le e^{L_F T} E_Q\left[\int_0^T \mathcal{W}_p(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u))\, du \mid \mathcal{F}_t^n\right] + \mathcal{W}_p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t)).$$

$$(3.5)$$

Applying again Gronwall's inequality yields the desired result. $\qquad\square$

With the proofs of the above lemmas aside, we are ready to prove quantitative estimations for the convergence of the empirical measure of $\mathbf{Y}_t$ to the law $\mathcal{L}(Y_t)$ of the McKean-Vlasov BSDE.

Proof of Theorem 2.2 and Proposition 2.6. The proofs begin with Lemma 3.3. In fact this lemma implies that

$$E\left[\mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \le \exp(Te^{L_F T})E\left[\mathcal{W}_p^p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))\right] \quad \text{for all } (t, n) \in [0, T] \times \mathbb{N}.$$

$$(3.6)$$

Since $E[|G^i|^k] < \infty$ with $k > p$, we have by Lemma 3.1 that $\sup_{t \in [0,T]} E[|Y_t|^q] < \infty$ for $q \in (p, k)$. Thus, it follows by [25, Theorem 1] that

$$E\left[\mathcal{W}_p^p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))\right] \le C r_{n,m,k,p}$$

for a constant C depending on L_F, T, m, p and k. Therefore, the estimate (2.3) is obtained due to (3.6).

To get the estimate (2.6), let $0 = t_0 \le t_1 \le \cdots \le t_N = T$ be a partition of $[0, T]$ in N intervals of length $t_{j+1} - t_j \le 1$ (if $T \le 1$ we simply take $N = 1$). Considering the decomposition: for all $t \in [0, T]$,

$$\mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) = \mathcal{W}_p^p(L^n(\mathbf{Y}_0), \mathcal{L}(Y_0))\mathbf{1}_{\{0\}}(t) + \sum_{j=0}^{N-1} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))1_{(t_j, t_{j+1}]}(t),$$

we have

$$E\left[\sup_{t \in [0,T]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \le \sum_{j=0}^{N-1} E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right]. \qquad (3.7)$$

From Lemma 3.3 we have: for all $n \in \mathbb{N}$, $j \le N - 1$,

$$E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \le \exp(pTe^{L_F T})E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_p^p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))\right].$$

Since $p \in [1,2]$, it follows by Jensen's inequality and the inequality $\mathcal{W}_p \leq \mathcal{W}_2$ that for all $n \in \mathbb{N}$,

$$
E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right]^{2/p} \leq e^{2L_F T} E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_2^2(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))\right] \qquad (3.8)
$$

Since $t_{j+1} - t_j \leq 1$, Lemma 3.2 applies, in view of which it follows from Lemma 3.1 and [29, Theorem 1.3] that

$$
E\left[\sup_{t \in [t_j, t_{j+1}]} \mathcal{W}_2^2(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t))\right] \leq C_{G,F,k,m} n^{-2/(m+8)}.
$$

Therefore, we deduce from (3.8) and (3.7) that

$$
E\left[\sup_{t \in [0,T]} \mathcal{W}_p^p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t))\right] \leq N e^{pL_F T} C_{G,F,k,m} n^{-p/(m+8)}.
$$

This concludes the proof, since N can be chosen less than $T + 1$. $\qquad \square$

Proof of Remark 2.3. In the quadratic case, the estimate of Lemma 3.1 for $q = 2$ is guaranteed by assumption. The proof of Lemma 3.3 remains the same, up to replacing the process γ_u therein by the process

$$
\gamma_u := \frac{F_u(Y_u^{i,n}, Z_u^{i,i,n}, \mathcal{L}(Y_u)) - F_u(Y_u^{i,n}, \tilde{Z}_u^i, \mathcal{L}(Y_u))}{Z_u^{i,i,n} - \tilde{Z}_u^i} 1_{\{Z_u^{i,i,n} \neq \tilde{Z}_u^i\}}.
$$

By the quadratic growth assumption in z and square integrability assumption on $Z^{i,i,n}$ and $\tilde{Z}^i$, the process $(\gamma_t)_t$ is square integrable. The rest of the proof is the same. $\qquad \square$

Proof of Remark 2.5. The proof is the same as that of the estimate (2.3) with application of [29, Theorem 1.2] instead of [25, Theorem 1]. $\qquad \square$

3.2 Proof of Theorem 2.8

The proofs of Theorem 2.8 and Proposition 2.12 partially rely on functional inequalities that we now recall for the reader's convenience. See however [54, Chapters 21 & 22] for further details.

Let $\mathcal{W}_{2,\delta}$ denote the Wasserstein distance of order 2 with respect to a distance δ on a Polish space E. A probability measure $\mu \in \mathcal{P}(E)$ is said to satisfy Talagrand's T_2 inequality with constant C if

$$
\mathcal{W}_{2,\delta}(\mu, \nu) \leq \sqrt{CH(\nu|\mu)} \quad \text{for every probability measure } \nu,
$$

where H is the Kullback-Leibler divergence defined as

$$
H(\nu|\mu) := \begin{cases} \int \log(\frac{d\nu}{d\mu}) d\nu, & \text{if } \nu \ll \mu \\ +\infty, & \text{otherwise,} \end{cases}
$$

with the convention $E[X] := \infty$ whenever $E[X^+] = \infty$. Below, we will exploit the efficiency of Talagrand's inequality in deriving concentration inequalities, but also the fact that it implies other functional inequalities, notably the T_1 inequality

$$
\mathcal{W}_{1,\delta}(\mu, \nu) \leq \sqrt{CH(\nu|\mu)} \quad \text{for every probability measure } \nu,
$$

and Poincaré's inequality

$$
\text{Var}(f) \leq C \int_E |\nabla f|^2 \, d\mu
$$

for every (weakly) differentiable function $f : E \to \mathbb{R}$ and where $\text{Var}(f)$ is the variance with respect to the probability measure μ.

Proof of Theorem 2.8. The proof of the first concentration bound also uses Lemma 3.3. For simplicity, let $C_{F,T} = \exp(Te^{L_F T})$ denote the constant factor appearing in the right hand side of (3.3). It follows by (3.3) that

$$P\left(\mathcal{W}_p(L^n(\mathbf{Y}_t), \mathcal{L}(Y_t)) \geq \varepsilon\right) \leq P\left(\mathcal{W}_p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t)) \geq \varepsilon/C_{F,T}\right).$$

Fix $k > 2p$ such that $E[|G|^k] < \infty$, let $q \in (2p, k)$. By Lemma 3.1, we have $E[|Y_t|^q] < \infty$ and thus we can apply [25, Theorem 2], to obtain the inequality

$$P\left(\mathcal{W}_p(L^n(\tilde{\mathbf{Y}}_t), \mathcal{L}(Y_t)) \geq \varepsilon_{F,T}\right) \leq C(a_{n,\varepsilon_{F,T}} 1_{\{\varepsilon_{F,T} \leq 1\}} + b_{n,\varepsilon_{F,T}}),$$

where $\varepsilon_{F,T} = \varepsilon/C_{F,T}$. This concludes the proof of (2.7).

As for the proof of (2.8), consider the representation of the system (2.1) given in (3.1) on the probability space $(\Omega^n, \mathcal{F}^n, P)$. Since G^i is Lipschitz continuous for each i, it is easily checked that the random variable $\mathbf{G}$ is again L_G-Lipschitz continuous. In fact, given $\boldsymbol{\omega}, \boldsymbol{\theta} \in \Omega^n$, we have

$$|\mathbf{G}(\boldsymbol{\omega}) - \mathbf{G}(\boldsymbol{\theta})|^2 = \sum_{i=1}^n |G^i(\omega^i) - G^i(\theta^i)|^2 \leq L_G \sum_{i=1}^n ||\omega^i - \theta^i||_\infty^2 = L_G ||\boldsymbol{\omega} - \boldsymbol{\theta}||_\infty^2,$$

where for every $\boldsymbol{\omega} \in \Omega^n$, $||\boldsymbol{\omega}||_\infty = \sup_{t \in [0,T]} \left(\sum_{i=1}^n |\omega^i(t)|^2\right)^{1/2}$. Similarly, one shows that the function $\mathbf{F}$ is L_F-Lipschitz continuous. In particular, the Lipschitz constants of $\mathbf{F}$ and $\mathbf{G}$ do not depend on n. By [5, Theorem 1.2], the law of $\mathbf{Y}_t$ satisfies Talagrand's T_2-inequality with the constant

$$C_{F,G,T} := 2(L_G + TL_F)^2 e^{2TL_F}.$$

Thus, it follows by [28, Theorem 1.3] that there is a constant $C > 0$ such that for every 1-Lipschitz continuous functions $f : C([0,T], \mathbb{R}^m)^n \to \mathbb{R}$ we have

$$P\left(f(\mathbf{Y}) - E[f(\mathbf{Y})] \geq \varepsilon\right) \leq e^{-\varepsilon^2 C}.$$

The function $\boldsymbol{\omega} := (\omega^1, \dots \omega^n) \mapsto \sqrt{n}\mathcal{W}_{2,||\cdot||_\infty}(L^n(\boldsymbol{\omega}), \mathcal{L}(Y))$ is 1-Lipschitz continuous on Ω^n. Thus, we have

$$P\left(\mathcal{W}_{2,||\cdot||_\infty}(L^n(\mathbf{Y}), \mathcal{L}(Y)) - E[\mathcal{W}_{2,||\cdot||_\infty}(L^n(\mathbf{Y}), \mathcal{L}(Y))] \geq \varepsilon\right) \leq e^{-C\varepsilon^2 n}, \tag{3.9}$$

from which we deduce (2.8).

Lastly, we turn to the proof of (2.9). If F does not depend on z, we do not need the change of measure to get (3.4). In fact, a direct estimation yields

$$\tilde{Y}_t^i - Y_t^{i,n} = E\left[\int_t^T \left(F_u(\tilde{Y}_u^i, \mathcal{L}(Y_u)) - F_u(Y_u^{i,n}, L^n(\mathbf{Y}_u))\right) du \mid \mathcal{F}_t^n\right].$$

By Lipschitz continuity of F and Gronwall's inequality we have

$$|\tilde{Y}_t^i - Y_t^{i,n}|^2 \leq e^{2TL_F} E\left[\int_t^T \mathcal{W}_2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u)) du \mid \mathcal{F}_t^n\right]^2. \tag{3.10}$$

Thus, it follows by triangle inequality and definition of Wasserstein distance with respect to the supremum norm that

$$\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\mathbf{Y}), \mathcal{L}(Y)) \leq \mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\mathbf{Y}), L^n(\tilde{\mathbf{Y}})) + \mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}), \mathcal{L}(Y))$$

$$\leq \sup_{t \in [0,T]} \frac{1}{n} \sum_{i=1}^n |\tilde{Y}_t^i - Y_t^{i,n}|^2 + \mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}), \mathcal{L}(Y))$$

$$\leq \sup_{t \in [0,T]} e^{2TL_F} E\left[\int_0^T \mathcal{W}_2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u)) du \mid \mathcal{F}_t^n\right]^2$$

$$+ \mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}), \mathcal{L}(Y))$$

where the last inequality follows from (3.10). Hence, it follows by Doob's maximal inequality that

$$
\begin{aligned}
E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\mathbf{Y}),\mathcal{L}(Y))\Big] &\leq e^{2TL_F}TE\Big[\int_0^T \mathcal{W}_2^2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u))\,du\Big] \\
&\quad + E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \\
&\leq Cr_{n,m,k,2} + E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big],
\end{aligned}
$$

for some $q \in (p,k)$ where the last inequality follows by Fubini's theorem and Theorem 2.2. Since $\tilde{\mathbf{Y}} = (\tilde{Y}^1,\ldots,\tilde{Y}^n)$, are i.i.d., it holds $E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \to 0$ as n goes to infinity. Thus, there is and integer $\tilde{n}_0$ large enough such that for all $n \geq \tilde{n}_0$ we have $E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \leq Cr_{n,m,k,2}$. Hence, we have

$$
E\Big[\mathcal{W}_{2,||\cdot||_\infty}^2(L^n(\mathbf{Y}),\mathcal{L}(Y))\Big] \leq Cr_{n,m,k,2} \tag{3.11}
$$

for $n \geq \tilde{n}_0$. Now, by (3.9), it holds that

$$
\begin{aligned}
&P\left(\mathcal{W}_{2,||\cdot||_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y)) \geq \varepsilon\right) \\
&\leq P\left(\mathcal{W}_{2,\|\cdot\|_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y)) - E[\mathcal{W}_{2,\|\cdot\|_\infty}(L^n(\mathbf{Y}),\mathcal{L}(Y))] \geq \varepsilon/2\right) \\
&\quad + P\left(E\Big[\mathcal{W}_{2,||\cdot||_\infty}(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \geq \varepsilon/2\right) \\
&\leq e^{-C\varepsilon^2 n} + P\left(E\Big[\mathcal{W}_{2,||\cdot||_\infty}(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \geq \varepsilon/2\right).
\end{aligned}
$$

In view of (3.11) and the fact that $r_{n,m,k,p} \downarrow 0$ as n goes to infinity, we can choose $n_0 \geq \tilde{n}_0$ large enough such that for all $n \geq n_0$

$$
P\left(E\Big[\mathcal{W}_{2,||\cdot||_\infty}(L^n(\tilde{\mathbf{Y}}),\mathcal{L}(Y))\Big] \geq \varepsilon/2\right) = 0.
$$

This concludes the proof of (2.9). □

3.3 Proofs for Subsection 2.4

We begin with the proof of Theorem 2.9. As we will see below, this result is obtained as a consequence of Theorem 2.2.

Proof of Theorem 2.9. Since $Y^{1,n}$ and Y^1 satisfy (2.1) and (2.2) respectively, we have

$$
Y_t^{1,n} - Y_t^1 = E\left[\int_t^T \left(F_u(Y_u^{1,n}, Z_u^{1,n}, L^n(\mathbf{Y}_u)) - F_u(Y_u^1, Z_u^1, \mathcal{L}(Y_u))\right)\,du \mid \mathcal{F}_t^n\right]
$$

so that by Lipschitz continuity of F, and Gronwall's inequality it holds

$$
|Y_t^{1,n} - Y_t^1| \leq e^{L_F T}E\left[\int_t^T L_F\left(\mathcal{W}_2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u)) + |Z_u^{1,n} - Z_u^1|\right)\,du \mid \mathcal{F}_t^n\right].
$$

Therefore, it follows by Doob's maximal inequality that

$$
E\left[\sup_{t\in[0,T]}|Y_t^{1,n}-Y_t^1|^2\right]
$$
$$
\leq L_F^2 e^{2L_F T} E\left[\sup_{t\in[0,T]} E\left[\int_0^T \left(\mathcal{W}_2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u))+|Z_u^{1,n}-Z_u^1|\right)du \mid \mathcal{F}_t^n\right]^2\right]
$$
$$
\leq L_F^2 e^{2L_F T} E\left[\left(\int_0^T \left(\mathcal{W}_2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u))+|Z_u^{1,n}-Z_u^1|\right)du\right)^2\right]
$$
$$
\leq 2L_F^2 T e^{2L_F T}\left(\int_0^T E\left[\mathcal{W}_2^2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u))\right]du + E\left[\int_0^T |Z_u^{1,n}-Z_u^1|^2\,du\right]\right). \quad (3.12)
$$

On the other hand, applying Itô's formula to the process $|Y_t^{1,n}-Y_t^1|^2$, we have

$$
|Y_t^{1,n}-Y_t^1|^2 = -2\sum_{j=1}^n \int_t^T (Y_s^{1,n}-Y_s^1)(Z_s^{1,j,n}-\delta_{1j}Z_s^1)dW_s^j
$$
$$
-\sum_{j,l=1}^n \int_t^T (Z_s^{1,j,n}-\delta_{1j}Z_s^1)(Z_s^{1,l,n}-\delta_{1l}Z_s^1)d\langle W^j,W^l\rangle_s
$$
$$
+2\int_t^T (Y_s^{1,n}-Y_s^1)\left\{F_s(Y_s^{1,n},Z_s^{1,1,n},L^n(\mathbf{Y}_s))-F_s(Y_s^1,Z_s^1,\mathcal{L}(Y_s^1))\right\}ds.
$$

By Lipschitz continuity of F we then have

$$
|Y_t^{1,n}-Y_t^1|^2 + \sum_{j,l=1}^n \int_t^T (Z_s^{1,j,n}-\delta_{1j}Z_s^1)(Z_s^{1,l,n}-\delta_{1l}Z_s^1)d\langle W^j,W^l\rangle_s
$$
$$
\leq -2\sum_{j=1}^n \int_t^T (Y_s^{1,n}-Y_s^1)(Z_s^{1,j,n}-\delta_{1j}Z_s^1)dW_s^j
$$
$$
+2\int_t^T L_F|Y_s^{1,n}-Y_s^1||Z_s^{1,1,n}-Z_s^1|ds + 2\int_t^T L_F|Y_s^{1,n}-Y_s^1|^2 ds
$$
$$
+2\int_t^T L_F|Y_s^{1,n}-Y_s^1|\mathcal{W}_2(L^n(\mathbf{Y}_s),\mathcal{L}(Y_s^1))\,ds
$$
$$
\leq -2\sum_{j=1}^n \int_t^T (Y_s^{1,n}-Y_s^1)(Z_s^{1,j,n}-\delta_{1j}Z_s^1)dW_s^j + \int_t^T \frac{L_F}{\alpha}|Z_s^{1,1,n}-Z_s^1|^2 ds
$$
$$
+\int_t^T (3\alpha+L_F)|Y_s^{1,n}-Y_s^1|^2 ds + \int_t^T L_F\mathcal{W}_2^2(L^n(\mathbf{Y}_s),\mathcal{L}(Y_s^1))\,ds, \quad (3.13)
$$

where the last inequality follows by Young's inequality with some constant $\alpha > 0$. Choosing $\alpha = L_F+1$, we obtain

$$
|Y_t^{1,n}-Y_t^1|^2 \leq E\left[\int_t^T (4+L_F)L_F|Y_s^{1,n}-Y_s^1|^2 ds + \int_t^T L_F\mathcal{W}_2^2(L^n(\mathbf{Y}_s),\mathcal{L}(Y_s^1))\,ds \mid \mathcal{F}_t^n\right]
$$

so that by Gronwall's inequality we have that

$$
|Y_t^{1,n}-Y_t^1|^2 \leq e^{(4+L_F)L_F T} E\left[\int_t^T L_F\mathcal{W}_2^2(L^n(\mathbf{Y}_u),\mathcal{L}(Y_u^1))\,du \mid \mathcal{F}_t^n\right]. \quad (3.14)
$$

Now, integrating on both sides yields

$$E\left[|Y_t^{1,n} - Y_t^1|^2\right] \le e^{(4+L_F)L_F T} E\left[\int_0^T L_F \mathcal{W}_2^2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u^1))\, du\right]$$

$$\le e^{(4+L_F)L_F T} L_F \int_0^T E\left[\mathcal{W}_2^2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u^1))\right]\, du \qquad (3.15)$$

from which we derive, due to Theorem 2.2, that

$$E\left[|Y_t^{1,n} - Y_t^1|^2\right] \le e^{(4+L_F)L_F T} L_F T C\, r_{n,m,q,2} \qquad (3.16)$$

with $q \in (2, k)$.

To get the convergence estimate for the control processes $Z^{1,n}$, notice that by (3.13) (with the choice $\alpha = L_F + 1$), we have

$$E\left[\int_0^T |Z_u^{1,n} - Z_u^1|^2\, du\right]$$

$$\le (L_F + 1)E\left[(4+L_F)L_F \int_0^T |Y_u^{1,n} - Y_u^1|^2\, du + L_F \int_0^T \mathcal{W}_2^2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u^1))\, du\right]$$

$$\le (L_F + 1)\Big((4+L_F)L_F \int_0^T E\left[|Y_u^{1,n} - Y_u^1|^2\right]\, du$$

$$+ L_F \int_0^T E\left[\mathcal{W}_2^2(L^n(\mathbf{Y}_u), \mathcal{L}(Y_u^1))\right]\, du\Big). \qquad (3.17)$$

Now combine this with the inequalities (3.12), (3.16) and Theorem 2.2 to conclude. □

We now turn to the particular case of systems with linear interaction. Unlike Theorem 2.9, the proof of Proposition 2.12 does not follow from the convergence of the empirical measure of $\mathbf{Y}$, but by a direct argument which seems tailor-made for "linear interaction functions". Before presenting the proof, let us justify the well-posedness of the system.

Proof of Remark 2.11. The existence and uniqueness of square integrable solutions follows, as in the non-linear interaction case, from [24] for the system (2.12) and [16, Theorem 4.23] for the McKean-Vlasov BSDE (2.13). It suffices to show that the generators are Lipschitz continuous (with Lipschitz constant independent of n). We give only the argument for the McKean-Vlasov equation. Let $y, y' \in \mathbb{R}^m$, $z, z' \in \mathbb{R}^{m \times d}$ and $\mu, \mu' \in \mathcal{P}(\mathbb{R}^m)$. By (Lip) we have

$$|F(y, z, \mu) - F(y', z', \mu')|$$

$$\le L_F\Big(|y - y'| + |z - z'| + |\int_{\mathbb{R}^m} f(y, a, z)\, \mu(da) - \int_{\mathbb{R}^m} f(y', a, z')\, \mu'(da)|\Big)$$

$$\le L_F\Big(|y - y'| + |z - z'| + L_f \Big|\int_{\mathbb{R}^m} \frac{1}{L_f} f(y, a, z)\, \mu(da) - \int_{\mathbb{R}^m} \frac{1}{L_f} f(y, a, z)\, \mu'(da)\Big|$$

$$+ \Big|\int_{\mathbb{R}^m} f(y, a, z)\, \mu'(da) - \int_{\mathbb{R}^m} f(y', a, z')\, \mu'(da)\Big|\Big)$$

$$\le L_F\Big(|y - y'| + |z - z'| + L_f \mathcal{W}_1(\mu, \mu') + L_f \int_{\mathbb{R}^m} (|y - y'| + |z - z'|)\, \mu(da)\Big)$$

$$\le \max(L_F, L_F L_f, L_f)\Big(|y - y'| + |z - z'| + \mathcal{W}_2(\mu, \mu')\Big).$$

To derive the penultimate inequality, we used Kantorovich-Rubinstein duality formula. □

Proof of Proposition 2.12. It follows by application of Itô's formula that

$$|Y_t^{1,n} - Y_t^1|^2 + \sum_{j,l=1}^{n} \int_t^T (Z_s^{1,j,n} - \delta_{1j} Z_s^1)(Z_s^{1,l,n} - \delta_{1l} Z_s^1) d\langle W^j, W^l \rangle_u$$

$$\leq -2 \sum_{j=1}^{n} \int_t^T (Y_u^{1,n} - Y_u^1)(Z_{u}^{1,j,n} - \delta_{1,j} Z_u^1) dW_u^j$$

$$+ 2 \int_t^T (Y_u^{1,n} - Y_u^1) \left(F_u\Big(Y_u^{1,n}, Z_u^{1,n}, \frac{1}{n} \sum_{j=1}^{n} f_u(Y_u^{1,n}, Y_u^{j,n}, Z_u^{1,n})\Big) \right.$$

$$\left. - F_u\Big(Y_u^1, Z_u^1, \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1) \, d\mathcal{L}(Y_u^1)(y)\Big) \right) du$$

$$\leq -2 \sum_{j=1}^{n} \int_t^T (Y_u^{1,n} - Y_u^1)(Z_u^{1,j,n} - \delta_{1j} Z_u^1) dW_u^j + L_F \int_t^T \frac{1}{\alpha} |Z_u^{1,n} - Z_u^1|^2 \, du$$

$$+ L_F \int_t^T (2\alpha + 2)|Y_u^{1,n} - Y_u^1|^2 \, du$$

$$+ L_F \int_t^T \frac{1}{\alpha} \left| \frac{1}{n} \sum_{j=1}^{n} f_u(Y_u^{1,n}, Y_u^{j,n}, Z^{1,n}) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1) \, d\mathcal{L}(Y_u^1)(y) \right|^2 du \qquad (3.18)$$

where the last inequality follows by Young's inequality with some constant $\alpha > 0$. Let us analyze the last term above. It follows by triangle inequality that

$$\left| \frac{1}{n} \sum_{j=1}^{n} f_u(Y_u^{1,n}, Y_u^{j,n}, Z^{1,n}) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1) \, d\mathcal{L}(Y_u^1)(y) \right|^2$$

$$\leq 2 \left| \frac{1}{n} \sum_{j=1}^{n} \left\{ f_u(Y_u^{1,n}, Y_u^{j,n}, Z_u^{1,n}) - f_u(Y_u^1, Y_u^{j,n}, Z_u^{1,n}) \right\} \right|^2$$

$$+ 4 \left| \frac{1}{n} \sum_{j=1}^{n} \left(f_u(Y_u^1, Y_u^{j,n}, Z_u^{1,n}) - f_u(Y_u^1, Y_u^j, Z_u^{1,n}) \right) \right|^2$$

$$+ 8 \left| \frac{1}{n} \sum_{j=1}^{n} \left(f_u(Y_u^1, Y_u^j, Z_u^{1,n}) - f_u(Y_u^1, Y_u^j, Z_u^1) \right) \right|^2$$

$$+ 8 \left| \frac{1}{n} \sum_{j=1}^{n} \left(f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1) \, d\mathcal{L}(Y_u^1)(y) \right) \right|^2$$

$$\leq L_f \left\{ 2|Y_u^{1,n} - Y_u^1|^2 + 8|Z_u^{1,n} - Z_u^1|^2 \right\} + 4 L_f \frac{1}{n} \sum_{j=1}^{n} |Y_u^{j,n} - Y_u^j|^2$$

$$+ \frac{8}{n^2} \left| \sum_{j=1}^{n} \left[f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1) \, d\mathcal{L}(Y_u^1)(y) \right] \right|^2.$$

Coming back to (3.18), this allows to obtain, after taking conditional expectation with

respect to the sigma algebra $\mathcal{F}_t^n$,

$$|Y_t^{1,n} - Y_t^1|^2 + E\left[\int_t^T |Z_s^{1,n} - Z_s^1|^2 ds \mid \mathcal{F}_t^n\right]$$

$$\leq E\left[\int_t^T \left\{(L_F(2\alpha+2) + \frac{2L_f}{\alpha})|Y_u^{1,n} - Y_u^1|^2 + \frac{8L_f + L_F}{\alpha}|Z_u^{1,n} - Z_u^1|^2\right\} du \mid \mathcal{F}_t^n\right]$$

$$+ \frac{4L_f}{\alpha} E\left[\frac{1}{n}\sum_{j=1}^n \int_t^T |Y_u^{j,n} - Y_u^j|^2 du \mid \mathcal{F}_t^n\right]$$

$$+ \frac{8}{\alpha n^2} E\left[\int_t^T \left|\sum_{j=1}^n \left[f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right]\right|^2 du \mid \mathcal{F}_t^n\right].$$

Choose $\alpha = 8L_f + L_F + 1$. Then, we have

$$\alpha|Y_t^{1,n} - Y_t^1|^2 + E\left[\int_t^T |Z_s^{1,n} - Z_s^1|^2 ds \mid \mathcal{F}_t^n\right]$$

$$\leq \alpha E\left[\int_t^T \left\{(L_F(2\alpha+2) + 1)|Y_u^{1,n} - Y_u^1|^2\right\} du \mid \mathcal{F}_t^n\right]$$

$$+ 4L_f E\left[\frac{1}{n}\sum_{j=1}^n \int_t^T |Y_u^{j,n} - Y_u^j|^2 du \mid \mathcal{F}_t^n\right]$$

$$+ \frac{8}{n^2} E\left[\int_t^T \left|\sum_{j=1}^n \left[f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right]\right|^2 du \mid \mathcal{F}_t^n\right]. \quad (3.19)$$

Since $E[|Y_t^{1,n} - Y_t^1|^2] = E[|Y_t^{j,n} - Y_t^j|^2]$ for all j we can apply Gronwall's inequality to get

$$|Y_t^{1,n} - Y_t^1|^2 \leq \frac{C_{T,F,f}}{n^2} E\left[\int_0^T \left|\sum_{j=1}^n \left[f_u(Y_u^1, Y_u^j, Z_u^1)\right.\right.\right.$$

$$\left.\left.\left. - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right]\right|^2 du \mid \mathcal{F}_t^n\right]$$

for a constant $C_{T,F,f}$ depending only on T, L_f and L_F. Now, denoting by f^ℓ the ℓth component of f and using that Y_u^i and Y_u^j have the same law we deduce that

$$E\left[|Y_t^{1,n} - Y_t^1|^2\right] \leq \frac{C_{T,F,f}}{n^2} E\left[\int_0^T \left|\sum_{j=1}^n \left[f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right]\right|^2 du\right]$$

$$\leq \frac{C_{T,F,f}}{n^2} \int_0^T E\left[\sum_{\ell=1}^m \sum_{j,k=1}^n \left(f_u^\ell(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u^\ell(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right)\right.$$

$$\left. \times \left(f_u^\ell(Y_u^1, Y_u^k, Z_u^1) - \int_{\mathbb{R}^m} f_u^\ell(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right)\right] du$$

$$\leq \frac{C_{T,F,f}}{n^2} \int_0^T \sum_{j=1}^n E\left[\left|f_u(Y_u^1, Y_u^j, Z_u^1) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right|^2\right]$$

$$= \frac{C_{T,F,f}}{n^2} \int_0^T \sum_{j=1}^n E\left[\mathrm{Var}(f_u(Y_u^1, Y_u^j, Z_u^1))\right]\, du.$$

The equality before the last one follows from the fact that Y_t^i, $j = 1, \ldots, n$ are i.i.e. for all t. Since the law $\mathcal{L}(Y_u^1)$ of Y_u^1 satisfies Talagrand's T_2 inequality with a constant $C_{F,f,G,T}$ which depends only on the Lipschitz constants of F, f and G and of T (and which does not depend on the dimensions) (see [5, Theorem 1.3]) it follows e.g. by [54, Theorem 22.17] (see also [49, Section 7]) that $\mathcal{L}(Y_u^1)$ satisfies the Poincaré inequality with the same constant. That is, it holds that

$$\mathrm{Var}(f_u(x, Y_u^j, z)) \le C_{F,f,G,T} \int_{\mathbb{R}^m} |\partial_y f_u(x, y, z)|^2\, d\mathcal{L}(Y_u^j)(y) \quad \text{for all } x, z \text{ fixed.}$$

Since f is Lipschitz continuous, L_f^2 is an upper bound for the integral in the right hand side above (uniformly in x, z). Therefore, we have

$$E\left[|Y_t^{1,n} - Y_t^1|^2\right] \le \sum_{j=1}^n \frac{C}{n^2} \int_0^T E\left[\int_{\mathbb{R}^m} |\partial_y f_u(Y_u^j, y, Z_u^1)|^2\, d\mathcal{L}(Y_u^j)(y)\right]\, du$$
$$\le \frac{C}{n}$$

for some constant C depending on $T, L_F\ L_f$ and the Lipschitz constant of G.

Now, showing that $E[\int_0^T |Z_u^{1,n} - Z_u^1|^2\, du] \le C/n$ follows by (3.19). In fact, that inequality implies

$$E\left[\int_0^T |Z_s^{1,n} - Z_s^1|^2 ds\right] \le \int_0^T CE\left[|Y_s^{1,n} - Y_s^1|^2\right]\, ds$$
$$+ \frac{8}{n^2} E\left[\int_0^T \left|\frac{1}{n} \sum_{j=1}^n f_u(Y_u^{1,n}, Y_u^{j,n}, Z^{1,n}) - \int_{\mathbb{R}^m} f_u(Y_u^1, y, Z_u^1)\, d\mathcal{L}(Y_u^1)(y)\right|^2 ds\right].$$

We have just seen that, up to the factor CT (for some constant $C > 0$), the first term is smaller than the second one, which in turn is less than C/n for some constant $C > 0$. This completes the proof. □

3.4 Proof of Theorem 2.15

The main argument for the proof is the probabilistic representation of parabolic PDEs on the Wasserstein space given in [20]. This representation links the PDE (2.17) to the (decoupled) McKean Vlasov FBSDE

$$\begin{cases} X_s^{t,\xi} = \xi + \int_t^s B(X_u^{t,\xi}, \mathcal{L}(X_u^{t,\xi}))\, du + \int_t^s \sigma(X_u^{t,\xi}, \mathcal{L}(X_u^{t,\xi}))\, dW_u^1 \\ Y_s^{t,\xi} = G(X_T^{t,\xi}, \mathcal{L}(X_T^{t,\xi})) + \int_s^T F(X_u^{t,\xi}, Y_u^{t,\xi}, Z_u^{t,\xi}, \mathcal{L}(X_u^{t,\xi}), \mathcal{L}(Y_u^{t,\xi}))\, du - \int_s^T Z_u^{t,\xi}\, dW_u. \end{cases}$$
(3.20)

In preparation for the proof of Theorem 2.15, let us first show the following Lipschitz-type property for the solution of (3.20).

Lemma 3.4. *If the functions B, F and G are L_F Lipschitz continuous, then there is a constant C depending only on L_F and T such that*

$$E\left[|Y_t^{t,\xi} - Y_t^{t,\xi'}|^2\right] \le CE\left[|\xi - \xi'|^2\right]$$

for every square integrable and $\mathcal{F}_t$-measurable random variables ξ, ξ'.

Proof. This follows by standard arguments, thus we give only the main steps of the proof. Firstly, apply Itô's formula to $|Y_s^{t,\xi} - Y_s^{t,\xi'}|^2$, $s \in [t, T]$ and then using Lipschitz continuity of G and F and Young's inequality to obtain

$$E[|Y_t^{t,\xi} - Y_t^{t,\xi'}|^2] \le L_F E[|X_T^{t,\xi} - X_T^{t,\xi'}|^2 + \mathcal{W}_2^2(\mathcal{L}(X_T^{t,\xi}), \mathcal{L}(X_T^{t,\xi'}))]$$
$$+ (5L_F + 1)E\Big[\int_t^T |Y_u^{t,\xi} - Y_u^{t,\xi'}|^2$$
$$+ |X_u^{t,\xi} - X_u^{t,\xi'}|^2 + \mathcal{W}_2^2(\mathcal{L}(X_u^{t,\xi}), \mathcal{L}(X_u^{t,\xi'})) + \mathcal{W}_2^2(\mathcal{L}(Y_u^{t,\xi}), \mathcal{L}(Y_u^{t,\xi'}))\Big].$$

Since we have $\mathcal{W}_2^2(\mathcal{L}(X_u^{t,\xi}), \mathcal{L}(X_u^{t,\xi'})) \le E[|X_u^{t,\xi} - X_u^{t,\xi'}|^2]$ and a similar inequality for $Y^{t,\xi}$ and $Y^{t,\xi'}$, applying Gronwall's inequality yields (with $\tilde{L}_F := 5L_F + 1$)

$$E[|Y_t^{t,\xi} - Y_t^{t,\xi'}|^2] \le e^{\tilde{L}_F T} \tilde{L}_F (1 + T) E\Big[\sup_{u \in [0,T]} |X_u^{t,\xi} - X_u^{t,\xi'}|^2 \Big]. \tag{3.21}$$

Next, we can estimate $|X_u^{t,\xi} - X_u^{t,\xi'}|^2$ using Lipschitz continuity of b and σ and Burkholder-Davis-Gundy inequality to obtain

$$E\Big[\sup_{u \in [0,T]} |X_u^{t,\xi} - X_u^{t,\xi'}|^2 \Big] \le E\Big[|\xi - \xi'|^2 + C \int_0^T |X_u^{t,\xi} - X_u^{t,\xi'}|^2 + \mathcal{W}_2^2(\mathcal{L}(X_u^{t,\xi}), \mathcal{L}(X_u^{t,\xi'})) \, du \Big]$$

where C is a constant depending only on L_F and the universal constant in Burkholder-Davis-Gundy inequality. Thus, applying Gronwall's inequality yields

$$E\Big[\sup_{u \in [0,T]} |X_u^{t,\xi} - X_u^{t,\xi'}|^2 \Big] \le e^{CT} E\Big[|\xi - \xi'|^2 \Big].$$

Combine this with (3.21) to conclude. $\qquad\square$

Proof of Theorem 2.15. First note that the PDE (2.18) admits a (classical) solution. Assumption (PDE1) says that B is differentiable on $\mathcal{P}_2(\mathbb{R}^d)$. Thus, it follows by [16, Proposition 5.35] that, for every $x \in \mathbb{R}^d$, the projection of B on $\mathbb{R}^d \times (\mathbb{R}^d)^n$ given by $(x, \boldsymbol{x}) \mapsto \mathbf{B}(x, \boldsymbol{x}) := B(x, L^n(\boldsymbol{x}))$ is again differentiable, with

$$\partial_{x_i} \mathbf{B}(\cdot, \boldsymbol{x}) = \frac{1}{n} \partial_\mu B(\cdot, L^n(\boldsymbol{x}))(x_i), \quad i = 1, \ldots, n. \tag{3.22}$$

One similarly shows that the respective projections $\boldsymbol{\sigma}$, $\mathbf{F}$ and $\mathbf{G}$ of σ, F and G on finite dimensional spaces (with appropriate dimensions) are three times differentiable, and by the identity (3.22) the derivatives of first and second order are bounded by L_F. In particular, the bound does not depend on n. Therefore, it follows from [51, Theorem 3.2] that the PDE (2.18) admits a solution $v^{i,n} : [0, T] \times (\mathbb{R}^d)^n \to \mathbb{R}^m$ with the probabilistic representation

$$v^{i,n}\big(s, X_s^{1,n,t,x_1}, \ldots, X_s^{n,n,t,x_n}\big) = Y_s^{i,n,t,\boldsymbol{x}} \quad \text{for all} \quad s \ge t, \ (t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^n,$$

where $(X^{i,n,t,x_i}, Y^{i,n,t,\boldsymbol{x}}, Z^{i,n,t,\boldsymbol{x}})_{i=1,\ldots,n,t \ge 0}$ solves the decoupled FBSDE

$$\begin{cases} X_s^{i,n,t,x_i} = x_i + \int_t^s B(X_u^{i,n,t,x_i}, L^n(\mathbf{X}_u^{t,\boldsymbol{x}})) \, du + \int_t^s \sigma(X_u^{i,n,t,x_i}, L^n(\mathbf{X}_u^{t,\boldsymbol{x}})) \, dW_u \\ Y_s^{i,n,t,\boldsymbol{x}} = G(X_T^{i,n,t,x_i}, L^n(\mathbf{X}_T^{t,\boldsymbol{x}})) + \int_s^T F(X_u^{i,n,t,x_i}, Y_u^{i,n,t,\boldsymbol{x}}, Z_u^{i,n,t,\boldsymbol{x}}, L^n(\mathbf{X}_u^{t,\boldsymbol{x}}), L^n(\mathbf{Y}_u^{t,\boldsymbol{x}})) \, du \\ \quad - \int_s^T Z_u^{i,n,t,\boldsymbol{x}} \, dW_u \\ i = 1, \ldots, n \end{cases}$$
$$\tag{3.23}$$

with $\mathbf{X}^{t,\boldsymbol{x}} := (X_u^{1,n,t,x_1}, \ldots, X_u^{n,n,t,x_n})$ and $\mathbf{Y}^{t,\boldsymbol{x}} := (Y_u^{1,n,t,\boldsymbol{x}}, \ldots, Y_u^{n,n,t,\boldsymbol{x}})$. In particular $v^{i,n}(t, \boldsymbol{x}) = Y_t^{i,n,t,\boldsymbol{x}}$. Naturally, it follows by standard SDE and BSDE theories (see e.g. [50]) that the system (3.23) is well-posed, since by (PDE1) and the identity (3.22), the functions $\mathbf{B}, \boldsymbol{\sigma}$ and $\mathbf{F}$ are Lipschitz continuous (recall they are defined on finite dimensional spaces).

On the other hand, the PDE (2.17) is connected to the decoupled McKean-Vlasov FBSDE (3.20) whose solution is the triple $(X^{t,\xi}, Y^{t,\xi}, Z^{t,\xi})_t$, with $\xi \in L^2(\Omega, \mathcal{F}_t, P; \mathbb{R}^d)$ fixed. Since B is differentiable on $\mathcal{P}_2(\mathbb{R}^d)$, it follows by definition of $\partial_\mu B$ that for every $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, one has

$$B(x,\mu) - B(x,\mu') = \int_0^1 \int_{\mathbb{R}^d} \partial_\mu B(x, \lambda\mu + (1-\lambda)\mu')(y)(\mu - \mu')(dy)d\lambda.$$

Since $y \mapsto \partial_\mu B(x,\mu)(y)$ is L_F-Lipschitz continuous, it follows by Kantorovich-Rubinstein formula that

$$|B(x,\mu) - B(x,\mu')| \le \int_0^1 L_F \mathcal{W}_1(\mu, \mu') \, d\lambda = L_F \mathcal{W}_2(\mu, \mu').$$

That is, B is Lipschitz continuous. One similarly shows that the functions σ, F and G are Lipschitz continuous on their respective domain. Thus, the equation (3.20) is well-posed, see e.g. [53] and [16].

By the assumption (PDE) and Lemma (3.4) we can apply [20, Proposition 5.2], to obtain that the PDE (2.17) is well-posed and that its solution V satisfies

$$V(s, X_s^{t,\xi}, \mathcal{L}(X_s^{t,\xi})) = Y_s^{t,\xi}$$

for all $0 \le t \le s \le T$ and $\xi \in L^2(\Omega, \mathcal{F}_t, P; \mathbb{R}^d)$. In particular, $V(t, \xi, \mu) = Y_t^{t,\xi}$ for all $(t, \xi, \mu) \in [0,T] \times L^2(\Omega, \mathcal{F}_t, P; \mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ with $\mathcal{L}(\xi) = \mu$.

Let ξ_i, $i = 1, \ldots, n$, be i.i.d. random variables in $L^2(\Omega, \mathcal{F}_t, P; \mathbb{R}^d)$ with common law μ and denote $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_n)$. Then we have

$$E\Big[\sup_{t\in[0,T]} |v^{1,n}(t, \xi_1, \ldots, \xi_n) - V(t, \xi_1, \mathcal{L}(\xi^1))|^2\Big] = E\Big[\sup_{t\in[0,T]} |Y_t^{1,n,t,\boldsymbol{\xi}} - Y_t^{t,\xi_1}|^2\Big]. \quad (3.24)$$

To conclude, it remains to estimate the convergence rate of the right hand side of (3.24). This can be done as in the proof of Theorem 2.9. The only difference here being that the generator and terminal condition of the n-particle system "depends on (i, n)" through the processes X^{i,n,t,ξ_i} and their empirical measure. But thanks to the Lipschitz continuity property of F and G (proved above) this does not cause much problems. One can follow the estimations in the proofs of Lemma 3.3 and Theorem 2.9, adding the terms

$$\delta_s^n := |X_s^{1,n,t,\xi_1} - X_s^{t,\xi_1}|^2 \quad \text{and} \quad \eta_s^n := \mathcal{W}_2^2(L^n(\mathbf{X}_s^{t,\boldsymbol{\xi}_1}), \mathcal{L}(X_s^{t,\xi_1})).$$

In fact, using the arguments leading to Equation (3.12), Equation (3.15) and Equation (3.17) respectively, we obtain

$$E\Big[\sup_{t\in[0,T]} |Y_t^{1,n,t,\boldsymbol{\xi}} - Y_t^{t,\xi_1}|^2\Big] \le 16 L_F^2 e^{2L_F T} E\Big[\int_0^T \big(\delta_u^n + \eta_u^n + \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\big) \, du$$
$$+ \int_0^T |Z_u^{1,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi}|^2 \, du\Big] + 2L_G^2 E\Big[\delta_T^n + \eta_T^n\Big],$$

$$E\left[|Y_s^{1,n,t,\boldsymbol{\xi}} - Y_s^{t,\xi}|^2\right] \le e^{(4+L_F)L_F T}\Bigg(L_G E\left[\delta_T^n + \eta_T^n\right]$$

$$+ L_F E\Big[\int_0^T \left(\delta_u^n + \eta_u^n + \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\right) du\Big]\Bigg)$$

$$\le e^{(4+L_F)L_F T}(TL_F \vee L_G \vee L_F)\Bigg(E\Big[\sup_{s\in[0,T]} \delta_s^n + \sup_{s\in[0,T]} \eta_s^n\Big]$$

$$+ E\Big[\int_0^T \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\, du\Big]\Bigg) \tag{3.25}$$

and

$$E\Big[\int_0^T |Z_u^{1,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi}|^2\, du\Big] \le (L_F+1)\Bigg(L_G E\Big[\sup_{s\in[0,T]} \delta_s^n + \sup_{s\in[0,T]} \eta_s^n\Big]$$

$$+ \int_0^T (5+L_F)L_F E\Big[|Y_u^{1,n,t,\boldsymbol{\xi}} - Y_u^{t,\xi}|^2\Big]\, du$$

$$+ L_F E\Big[\int_0^T \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\, du\Big]\Bigg).$$

Therefore, there is a constant $C_{L_F,T}$ depending only on L_F and T such that

$$E\Big[\sup_{t\in[0,T]} |Y_t^{1,n,t,\boldsymbol{\xi}} - Y_t^{t,\xi}|^2\Big]$$

$$\le C_{L_F,T}\Big(E\Big[\sup_{s\in[0,T]} \delta_s^n + \sup_{s\in[0,T]} \eta_s^n\Big] + \int_0^T E\Big[\mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\Big]\, du\Big). \tag{3.26}$$

Moreover, by the theory of (forward) propagation of chaos, see e.g. [17, Theorem 2.12], it holds that

$$E\Big[\sup_{s\in[0,T]} \delta_s^n + \sup_{s\in[0,T]} \eta_s^n\Big] \le C\varepsilon_n. \tag{3.27}$$

It remains to estimate $E\big[\mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\big]$. This is done as follows. We apply Itô's formula to $|Y_s^{i,n,t,\boldsymbol{\xi}} - Y_s^{t,\xi}|^2$. This yields, thanks to the Lipschitz continuity of F and G,

$$|Y_s^{i,n,t,\boldsymbol{\xi}} - Y_s^{t,\xi}|^2 \le L_F\left(\delta_T^n + \eta_T^n\right) + L_F \int_s^T \Big[\Big(\delta_u^n + \eta_u^n + |Z_u^{i,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi}| + |Y_u^{i,n,t,\boldsymbol{\xi}} - Y_u^{t,\xi}|$$

$$+ \mathcal{W}_2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\Big)|Y_u^{i,n,t,\boldsymbol{\xi}} - Y_u^{t,\xi}| - \frac{1}{2}|Z_u^{i,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi}|^2\Big]\, du$$

$$- \int_s^T (Y_u^{i,n,t,\boldsymbol{\xi}} - Y_u^{t,\xi})(Z_u^{i,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi})\, dW_u.$$

Thus, using successively Young's inequality with a constant $\alpha > 0$ and Gronwall's inequality we have

$$|Y_s^{i,n,t,\boldsymbol{\xi}} - Y_s^{t,\xi}|^2 \le C_{L_F,\alpha,T} E\Big[\left(\delta_T^n + \eta_T^n\right) + \int_s^T \left(\delta_u^n + \eta_u^n + \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi}))\right)$$

$$+ \Big(\frac{1}{2\alpha} - \frac{1}{2}\Big)|Z_u^{i,n,t,\boldsymbol{\xi}} - Z_u^{t,\xi}|^2\, du \mid \mathcal{F}_s\Big]$$

$$\le C_{L_F,T} E\Big[\sup_{s\in[0,T]} \delta_s^n + \sup_{s\in[0,T]} \eta_s^n + \int_0^T \mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_u^{t,\xi})) \mid \mathcal{F}_s\Big],$$

where the second inequality follows by choosing $\alpha > 1$ and for some constant $C_{L_F,T} > 0$. Since

$$\mathcal{W}_2^2(L^n(\mathbf{Y}_u^{t,\xi}), \mathcal{L}(Y_u^{t,\xi})) \leq \frac{1}{n}\sum_{i=1}^{n}|Y_u^{i,n,t,\boldsymbol{\xi}} - Y_u^{t,\xi}|^2,$$

we obtain by Gronwall's inequality that

$$E\Big[\mathcal{W}_2^2(L^n(\mathbf{Y}_s^{t,\boldsymbol{\xi}}), \mathcal{L}(Y_s^{t,\xi}))\Big] \leq C_{L_F,T}E\Big[\sup_{s\in[0,T]}\delta_s^n + \sup_{s\in[0,T]}\eta_s^n\Big]$$

and it thus follows from (3.26) and (3.27) that

$$E\Big[\sup_{t\in[0,T]}|Y_t^{1,n,t,\boldsymbol{\xi}} - Y_t^{t,\xi}|^2\Big] \leq C_{L_F,T}\varepsilon_n$$

for some constant $C_{L_F,T} > 0$. Combining this with (3.24) leads to (2.19).

To prove (2.21), let $(\xi_1,\ldots,\xi_n)$ be n i.i.d. $\mathcal{F}_t$-measurable, square integrable random variables. It follows by triangle inequality that

$$
\begin{aligned}
&|v^{i,n}(t,\xi_1,\ldots,\xi_n) - V(t,\xi_i,L^n(\boldsymbol{\xi}))|\\
&\leq |v^{i,n}(t,\xi_1,\ldots,\xi_n) - V(t,\xi_i,\mathcal{L}(\xi_1))| + |V(t,\xi_i,\mathcal{L}(\xi_1)) - V(t,\xi_i,L^n(\boldsymbol{\xi}))|\\
&\leq |v^{i,n}(t,\xi_1,\ldots,\xi_n) - V(t,\xi_i,\mathcal{L}(\xi_1)| + C\mathcal{W}_2(\mathcal{L}(\xi_1),L^n(\boldsymbol{\xi})),
\end{aligned}
$$

where the second inequality follows by Lipschitz continuity of V given in [20, Proposition 5.2]. Therefore, we obtain by (2.19) and [25, Theorem 1] that

$$E\left[|v^{i,n}(t,\xi_1,\ldots,\xi_n) - V(t,\xi_i,L^n(\boldsymbol{\xi}))|^2\right] \leq C_{L_F,T}\left(\varepsilon_n + r_{n,d,k,2}\right).$$

This concludes the proof. $\qquad\square$

References

[1] Beatrice Acciaio, Julio Backhoff-Veraguas, and Rene Carmona, *Extended mean field control problems: stochastic maximum principle and transport perspective*, SIAM J. Cont. Optim., to appear (2018). MR4029806

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[3] Alexander Aurell, *Mean-field type games between two players driven by backward stochastic differential equations*, Games **9** (2018), no. 4, Paper No. 88, 26. MR3893930

[4] Alexander Aurell and Boualem Djehiche, *Modeling tagged pedestrian motion: A mean-field type game approach*, Transportation research part B: methodological **121** (2019), 168–183.

[5] Daniel Bartl and Ludovic Tangpi, *Functional inequalities for forward and backward diffusions*, Preprint (2019). MR4136474

[6] Alain Bensoussan, K. C. J. Sung, Sheung Chi Phillip Yam, and Siu-Pang Yung, *Linear-quadratic mean field games*, J. Optim. Theory Appl. **169** (2016), no. 2, 496–529. MR3489817

[7] M. Bossy and Denis Talay, *Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation*, Ann. Appl. Probab. **6** (1996), no. 3, 818–861. MR1410117

[8] Philippe Briand, Pierre Cardaliaguet, Paul-Eric Chaudru De Raynal, and Ying Hu, *Forward and backward stochastic differential equations with normal constraints in law*, Stochastic Processes and their Applications **130** (2020), no. 12, 7021–7097. MR4167200

[9] Philippe Briand and Romuald Elie, *A simple constructive approach to quadratic bsdes with or without delay*, Stochastic processes and their applications **123** (2013), no. 8, 2921–2939. MR3062430

[10] Philippe Briand and Hélène Hibon, *Particles systems for mean reflected bsdes*, Stochastic Processes and their Applications **131** (2021), 253–275. MR4157714

[11] Rainer Buckdahn, Boualem Djehiche, Juan Li, and Shige Peng, *Mean-field backward stochastic differential equations: A limit approach*, Ann. Probab. **37** (2009), no. 4, 1524–1565. MR2546754

[12] Pierre Cardaliaguet, *The convergence problem in mean field games with local coupling*, Appl. Math. Optim. **76** (2017), no. 1, 177–215. MR3679342

[13] Pierre Cardaliaguet, *A short course on mean field games*, Unpublished lecture notes, 2018.

[14] Pierre Cardaliaguet, François Delarue, J.-M. Lasry, and P.-L. Lions, *The master equation and the convergence problem in mean-field game*, Princeton University Press, 2019. MR3967062

[15] Pierre Cardaliaguet and Charles-Albert Lehalle, *Mean field game of controls and an application to trade crowding*, Math. Financ. Econ. **12** (2018), no. 3, 335–363. MR3805247

[16] René Carmona and François Delarue, *Probabilistic theory of mean field games with applications. I*, Probability Theory and Stochastic Modelling, vol. 83, Springer, Cham, 2018, Mean field FBSDEs, control, and games. MR3752669

[17] René Carmona and François Delarue, *Probabilistic theory of mean field games with applications. II*, Probability Theory and Stochastic Modelling, vol. 84, Springer, Cham, 2018, Mean field games with common noise and master equations. MR3753660

[18] René Carmona, Jean-Pierre Fouque, and Li-Hsien Sun, *Mean field games and systemic risk*, Commun. Math. Sci. **13** (2015), no. 4, 911–933. MR3325083

[19] Patrick Chan and Ronnie Sircar, *Bertrand and Cournot mean field games*, Appl. Math. Optim. **71** (2015), no. 3, 533–569. MR3359708

[20] Jean-François Chassagneux, Dan Crisan, and François Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, Forthcoming in Memoirs of the AMS (arXiv:1411.3009), 2014. MR3332710

[21] Patrck Cheridito and Kihun Nam, *Multidimensional quadratic and subquadratic BSDEs with special structure*, Stochastics **87** (2015), no. 5, 871–884. MR3390237

[22] Patrick Cheridito and Kihun Nam, *BSDEs with terminal conditions that have bounded Malliavin derivative*, J. Funct. Anal. **266** (2014), no. 3, 1257–1285. MR3146818

[23] François Delarue, Daniel Lacker, and Kavita Ramanan, *From the master equation to mean field game limit theory: Large deviations and concentration of measure*, The Annals of Probability **48** (2020), no. 1, 211–263. MR4079435

[24] Nicole El Karoui, Shige Peng, and Marie Claire Quenez, *Backward stochastic differential equations in finance*, Math. Finance **1** (1997), no. 1, 1–71. MR1434407

[25] Nicolas Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Relat. Fields **162** (2015), 707–738. MR3383341

[26] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions*, Math. Nach. **137** (1988), no. 1, 197–248. MR0968996

[27] Diogo A. Gomes, Stefania Patrizi, and Vardan Voskanyan, *On the existence of classical solutions for stationary extended mean field games*, Nonlinear Anal. **99** (2014), 49–79. MR3160525

[28] Nathael Gozlan, *A characterization of dimension free concentration in terms of transporation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498. MR2573565

[29] Joseph Horowitz and Rajeeva Karandikar, *Mean rate of convergence of empirical measures in the Wasserstein metric*, Journal of Computational and Applied Mathematics **55** (1994), 261–273. MR1329874

[30] Kaitong Hu, Zhenjie Ren, and Junjian Yang, *Principal-agent problem with multiple principals*, Preprint (2019).

[31] Ying Hu and Shanjian Tang, *Multi-dimensional backward stochastic differential equations of diagonally quadratic generators*, Stoch. Proc. Appl. **126** (2016), no. 4, 1066–1086. MR3461191

[32] Minyi Huang, Peter E. Caines, and Roland P. Malhamé, *Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ-Nash equilibria*, IEEE Trans. Automat. Control **52** (2007), no. 9, 1560–1571. MR2352434

[33] Minyi Huang, Roland P. Malhamé, and Peter E. Caines, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst. **6** (2006), no. 3, 221–251. MR2346927

[34] Peter Imkeller and Gonzalos Dos Reis, *Path regularity and explicit convergence rate for BSDE with truncated quadratic growth*, Stoch. Proc. Appl. **120** (2010), 348–379. MR2584898

[35] Pierre-Emmanuel Jabin and Zhenfu Wang, *Mean field limit and propagation of chaos for Vlasov systems with bounded forces*, J. Funct. Anal. **271** (2016), no. 12, 3588–3627. MR3558251

[36] Pierre-Emmanuel Jabin and Zhenfu Wang, *Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels*, Invent. Math. **214** (2018), no. 1, 523–591. MR3858403

[37] Jean-François Jabir, *Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation.*, Preprint (2019).

[38] Asgar Jamneshan, Michael Kupper, and Peng Luo, *Multidimensional quadratic BSDEs with separated generators*, Electron. Commun. Probab. **22** (2017), no. 58, 10. MR3718708

[39] M. Kac, *Foundations of kinetic theory*, Proc. Third Berkeley Symp. on Math. Stat. and Prob. **3** (1956), 171–197. MR0084985

[40] Michael Kupper, Peng Luo, and Ludovic Tangpi, *Multidimensional Markovian FBSDEs with superquadratic growth*, Stoch. Proc. Appl. **129** (2019), no. 3, 902–923. MR3913273

[41] Daniel Lacker, *On a strong form of propagation of chaos for Mckean-Vlasov equations*, Electron. Commun. Probab. **23** (2018), no. 45, 1–11. MR3841406

[42] J.-M. Lasry and P.-L. Lions, *Jeux à champs moyen. I le cas stationnaire*, C. R. Acad. Sci. Paris **343** (2006), no. 9, 619–625. MR2269875

[43] Jean-Michel Lasry and Pierre-Louis Lions, *Mean field games*, Jpn. J. Math. **2** (2007), no. 1, 229–260. MR2295621

[44] Mathieu Laurière and Ludovic Tangpi, *Convergence of large population games to mean field games with interaction through the controls*, To appear in SIAM Journal on Mathematical Analysis (preprint arXiv:2004.08351) (2020).

[45] P-L. Lions, *Cours du Collège de France*, https://www.college-de-france.fr/site/pierre-louis-lions/_course.htm, 2007-2012.

[46] Peng Luo and Ludovic Tangpi, *Solvability of coupled FBSDEs with diagonally quadratic generators*, Stoch. Dyn. **17** (2017), no. 6, 1750043. MR3685636

[47] François Malrieu, *Convergence to equilibrium for granular media equations and their numerical schemes*, Ann. Appl. Probab. **13** (2003), no. 2, 540–560. MR1970276

[48] Henry P. McKean, *Propagation of chaos for a class of non-linear parabolic equations*, Lecture Series in Differential Equations, Catholic Univ. Washington D.C., 1967, pp. 41–57. MR0233437

[49] Felix Otto and Cedric Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), 361–400. MR1760620

[50] E. Pardoux and S. Peng, *Adapted solution of a backward stochastic differential equation*, Systems Control Letter **14** (1990), 55–61. MR1037747

[51] Etienne Pardoux and Shige Peng, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Stochastic Partial Differential Equations and Their Applications (Boris L. Rozovskii and Richard B Sowers, eds.), Springer Berlin Heidelberg, 1992, pp. 200–217. MR1176785

[52] Mi. Shkolnikov, *Large systems of diffusions interacting through their ranks*, Stoch. Proc. Appl. **122** (2012), no. 4, 1730–1747. MR2914770

[53] Alain-Sol Sznitman, *Topics in propagation of chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, Lecture Notes in Math., vol. 1464, Springer, Berlin, 1991, pp. 165–251. MR1108185

[54] C. Villani, *Optimal transport: Old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer-Verlag, Berlin, Heidelberg, 2009. MR2459454

[55] Hao Xing and Gordan Žitković, *A class of globally solvable Markovian quadratic BSDE systems and applications*, Forthcoming in Ann. Probab. (2016). MR3758736