ON MEAN FIELD SYSTEMS WITH MULTI-CLASSES

DUNG TIENTHUYEN*
Department of Applied Mathematics
Faculty of Applied Science
Ho Chi Minh City University of Technology, Vietnam National University Ho Chi Minh City
268 Ly Thuong Kiet, District 10, Ho Chi Minh City, Vietnam

SON LUV NGUYEN
Department of Mathematics
University of Puerto Rico, Rio Piedras campus
San Juan, PR 00925, Puerto Rico

NGUYEN HUU DU
Department of Mathematics, Mechanics and Informatics
Hanoi National University
334 Nguyen Trai, Thanh Xuan, Hanoi Vietnam

(Communicated by Jaime San Martin)

Dedicated to Professor Gang George Yin on the occasion of his 65th birthday.

Abstract. This work focuses on stochastic systems of weakly interacting particles containing different populations represented by multi-classes. The dynamics of each particle depends not only on the empirical measure of the whole population but also on those of different populations. The limits of such systems as the number of particles tends to infinity are investigated. We establish the existence, uniqueness, and basic properties of solutions to the limiting McKean-Vlasov equations of these systems and then obtain the rate of convergence of the sequences of empirical measures associated with the systems to their limits in terms of the p^{th} Monge-Wasserstein distance.

1. Introduction. Originated from statistical physics, mean-field models are concerned with stochastic systems containing a large number of particles having weak interactions. To overcome the complexity of interactions due to the large scale of system, all interactions with each particle are replaced by a single average interaction normally represented by empirical measure associated to system.

Studying the limits of mean-field models as the sizes of the systems tend to infinity has been a long-standing problem and presents many technical difficulties. There have been a vast amount of works dealing with the limit of the empirical measures of mean field systems such as propagation of chaos, law of large numbers, fluctuations, phase transitions, and large deviations (see [6, 11, 13, 23, 27, 28, 29] among others and references therein). Recently, the past decade has witnessed renewed interests in mean-field models in game theory since the seminal works

2010 Mathematics Subject Classification. Primary: 60F25, 93E03; Secondary: 60J60.
Key words and phrases. Mean-field model, stochastic differential equation, McKean-Vlasov equation, Monge-Wasserstein distance, multi-classes.

* Corresponding author: Dung Tien Nguyen.
The mean field interaction has been used to model the weak interaction between players in large population games and the limiting results are used to construct computable decentralized strategies.

Along with this renewed interest in the classical models, the studies for some other type of mean-field models were also carried out; see for example, models with a major particle which has an important impact to all other particles [17, 24, 26], models with space noises [12, 21], models with a common noise [8, 20], models with jumps [1], the regime-switching models [25, 31], and models with two-time scales [15]. Another type of mean-field models being investigated recently is the class of models with multi-classes. In these models, the particles come from finitely many different populations, types, or classes, which appear in social sciences [10], statistical mechanics [9], neurosciences [2], as well as finance [5]. In particular, in [10], the phase transition is studied for a multi-class mean field statistical mechanics model. Two different classes of particles are introduced to depict two interacting groups of spins. The model is then interpreted as a prototype of resident-immigrant cultural interaction. In [9], a two-population generalization of the classical mean field Ising model is considered. In [2], multi-class mean field model is used to describe the weak interaction of a large network of neurons of $P$ different populations. In [5], the authors use multi-class mean field model to reformulate a portfolio optimization problem with a very large total number of stocks as a sector-wise allocation problem in which each sector can be interpreted as a mean field class. In mean-field games, systems with multi-classes are frequently used to describe the heterogeneity of the population of agents (see [3, 17, 18, 19, 24]).

To the best of our knowledge, despite of many works on these models, each of them focuses on a specific system which linearly depends on either mean field terms from populations or mean field term from whole population. Models in a general setting, therefore, have not been considered. In this paper, we study a mean field model with multi-class in which the dynamic equation of each particle depends on mean field terms from both subclasses and the whole population. Unlike classical cases, our proposed problem is dealing with several empirical measures from different populations. As a result, we obtain the limiting equations as a system of McKean-Vlasov equations instead of a single one. We establish the existence, uniqueness, and some basic properties of the solutions to limiting McKean-Vlasov equations of the system and then obtain the rate of convergence of the sequence of empirical measures to their limits in terms of $p^{th}$ Monge-Wasserstein distance.

The paper is organized as follows. In Section 2, we begin with an introduction to the model of multi-class weakly interacting diffusions, the McKean-Vlasov equations and the system of associated limiting equations. The main results are also presented in this section. For the sake of the exposition, their proofs are aggregated in Section 3 together with some auxiliary results. Finally, proofs of these auxiliary results are placed in the Appendix.

2. Formulation and main results.

2.1. $N$-particle multi-class weakly interacting system. Let $d, K$ be positive integers and $\mathbb{K} = \{1, \ldots, K\}$. For each vector $x \in \mathbb{R}^d$ let $\delta_x$ denote the Dirac measure centered at $x$, i.e. $\delta_x(A) = 1$ if $x \in A$, and $\delta_x(A) = 0$ if $x \notin A$ for any Borel subset $A$ of $\mathbb{R}^d$. We consider the following mean field system with $K$ different classes

$$x_i^{(N)}(t) = b_{\theta_i} \left(t, x_i^{(N)}(t), \mu^{(N)}(t), \mu^{(N)}(t) \right) dt$$
distribution of particles in population

tion of particles in the whole population while, for each
th population if

\[ x^{(N)}_i(0) = x^{0}_i, \quad i = 1, \ldots, N, \quad 0 \leq t \leq T, \]

where \( B_1(\cdot), B_2(\cdot), \ldots \) are independent \( d \)-dimensional standard Brownian motions defined on a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\), the parameter \( \theta_i \in \mathbb{K}, \ i = 1, \ldots, N\), which indicates that the particle \( x^{(N)}_i \) belongs to the \( j \)th population if \( \theta_i = j \), and

\[
\begin{align*}
\mu^{(N)}(t) &= \frac{1}{N} \sum_{i=1}^{N} \delta_{x^{(N)}_i(t)}, \\
\mu^{(N)}_j(t) &= \frac{1}{\nu_j(N)} \sum_{1 \leq i \leq N, \theta_i = j} \delta_{x^{(N)}_i(t)}, \quad \text{for } j = 1, \ldots, K,
\end{align*}
\]

with \( \nu_j(N) = \sum_{i=1}^{N} \mathbb{1}_{\{\theta_i = j\}} \). Let \( \mathcal{K}_j, \ 1 \leq j \leq K \), denote the \( j \)th population of particles, i.e., \( x^{(N)}_i \in \mathcal{K}_j \) if \( \theta_i = j \). Note that \( \mu^{(N)}(t) \) is the empirical distribution of particles in the whole population while, for each \( j \), \( \mu^{(N)}_j(t) \) is the empirical distribution of particles in population \( j \). It is easy to verify that

\[
\mu^{(N)}(t) = \frac{1}{N} \sum_{j=1}^{K} \nu_j(N) \mu^{(N)}_j(t). \tag{2}
\]

Throughout the paper, we assume that for each \( j \in \mathbb{K} \) the initial conditions \( (x^{0}_{i,j}, i = 1, 2, \ldots) \) are independently identically distributed random vectors defined on \((\Omega, \mathcal{F}, \mathbb{P})\), the collection of random vectors \( \{x^{0}_{i,j}, B_i(\cdot) : i = 1, 2, \ldots; j \in \mathbb{K}\} \) is independent, and the parameters \( \theta_i, i \geq 1 \), satisfy the equations

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{\theta_i = j\}} = \lim_{N \to \infty} \frac{\nu_j(N)}{N} = \nu_j > 0, \quad \text{for } j \in \mathbb{K}. \tag{3}
\]

Now we will introduce a distance between two probability measures. For each metric space \((\mathcal{E}, d_{\mathcal{E}})\) let \( \mathcal{P}(\mathcal{E}) \) denote the set of all probability measures defined on the Borel \( \sigma \)-field \( \mathcal{B}(\mathcal{E}) \) of \( \mathcal{E} \). For an \( \mathbb{E} \)-valued random variable \( X \), we use the notation \( \mathcal{L}(X) \) to denote its probability distribution. That is, \( \mathcal{L}(X) \in \mathcal{P}(\mathcal{E}) \). For each function \( f : \mathcal{E} \to \mathbb{R} \) and \( \mu \in \mathcal{P}(\mathcal{E}) \) we denote \( (\mu, f) = \int_{\mathcal{E}} f(x) \mu(dx) \) if the integral exists. In what follows, for convenience, we denote \( \mathcal{P} = \mathcal{P}(\mathbb{R}^d) \). For \( x \in \mathbb{R}^d \) and \( A \in \mathbb{R}^{d \times m} \)

such that \( A = (a_{ij}) \), let \( |x| = \sqrt{x^T x} \) denote the usual Euclidian norm of \( x \), and \( |A| = \max_{i,j} |a_{ij}| \). For \( p \geq 1 \) and measures \( \mu \) and \( \eta \) in \( \mathcal{P} \), the Monge-Wasserstein distance \( \mathcal{W}_p \) is defined by

\[
\mathcal{W}_p(\mu, \eta) = \inf \left\{ \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p dm(x, y) \right)^{1/p} : m \in \Gamma(\mu, \eta) \right\},
\]

where \( \Gamma(\mu, \eta) \) denotes the set of all probability measures on \( \mathbb{R}^d \times \mathbb{R}^d \) with marginals \( \mu \) and \( \eta \). In addition, denote the \( p \)th moment with respect to a measure \( \mu \) by

\[
M_p(\mu) = \int_{\mathbb{R}^d} |x|^p d\mu(x). \tag{4}
\]

According to the Kantorovich-Rubinstein theorem, for \( p = 1 \), the \( \mathcal{W}_1 \) distance possesses the following dual representation
\[ W_1(\mu, \eta) = \sup \left\{ |(\mu, f) - (\eta, f)| : |f|_L \leq 1 \right\}, \]

where \( (\mu, f) = \int_{\mathbb{R}^d} f(x) \mu(dx) \) and
\[
|f|_L = \sup \left\{ \frac{|f(x) - f(y)|}{|x - y|} : x \neq y \text{ in } \mathbb{R}^d \right\}
\]
is the Lipschitz seminorm of real-valued function \( f \) on \( \mathbb{R}^d \).

To proceed, we make the following assumptions on the functions \( b_j(\cdot, \cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^d \times \mathcal{P} \times \mathcal{P}^K \to \mathbb{R}^d \) and \( \sigma_j(\cdot, \cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^d \times \mathcal{P} \times \mathcal{P}^K \to \mathbb{R}^d \times d \) for each \( j \in \mathbb{K} \).

**Assumption A.** For some \( p \geq 1 \) there exists a constant \( C > 0 \) such that for \( j \in \mathbb{K}, t > 0, x \in \mathbb{R}^d, y \in \mathbb{R}^d, \mu \in \mathcal{P}, \eta \in \mathcal{P}, (\mu_1, \ldots, \mu_K) \in \mathcal{P}^K, \) and \( \eta = (\eta_1, \ldots, \eta_K) \in \mathcal{P}^K, \) the following inequalities hold true.

1. \( |b_j(t, x, \mu, \eta)| + |\sigma_j(t, x, \mu, \eta)| \leq C \left( 1 + |x| + \langle \mu, \varphi \rangle + \sum_{j=1}^{K} (\mu_j, \varphi) \right), \) where \( \varphi(x) = |x| \).
2. \( |b_j(t, x, \mu, \mu) - b_j(t, y, \mu, \eta)| + |\sigma_j(t, x, \mu, \mu) - \sigma_j(t, y, \mu, \eta)| \leq C \left( |x - y| + \sum_{j=1}^{K} W_p(\mu, \eta) + \sum_{j=1}^{K} W_p(\mu_j, \eta_j) \right). \)

It is easily seen that the assumptions (A1) and (A2) respectively can be considered as usual linear growth and Lipschitz continuity conditions.

**Example 2.1.** (i) In [2, 7, 16], the authors consider a specific form of equation (1) where the coefficients are defined as follows
\[
b_\theta(t, x, \mu, \mu) = f_\theta(t, x) + \sum_{\gamma \in \mathbb{K}} \langle b_{\theta, \gamma}(x, \cdot), \mu_\gamma \rangle, \quad \text{and} \quad \sigma_\theta(t, x, \mu, \mu) \equiv I_d,
\]
for each \( \theta \in \mathbb{K}, \) where \( f_\theta(\cdot, \cdot) : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \) and \( b_{\theta, \gamma}(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \), \( I_d \) is the identity \( d \times d \) matrix, and \( \langle b_{\theta, \gamma}(x, \cdot), \mu \rangle = \int_{\mathbb{R}^d} b_{\theta, \gamma}(x, y) d\mu(y) \) for each \( \mu \in \mathcal{P}, \) \( \theta, \gamma \in \mathbb{K}. \) By virtue of (5), we can show that Assumption (A) is satisfied for \( p = 1 \) if the functions \( f_\theta(\cdot, \cdot) \) and \( b_{\theta, \gamma}(\cdot, \cdot) \) satisfy
\[
|b_{\theta, \gamma}(x, y)| \leq L,
\]
and
\[
|f_\theta(t, x) - f_\theta(t, x') + b_{\theta, \gamma}(x, y) - b_{\theta, \gamma}(x', y')| \leq L(|x - y| + |x' - y'|),
\]
for \( \theta, \gamma \in \mathbb{K}, t \in [0, T], \) and \( x, x', y, y' \in \mathbb{R}^d. \)

(ii) In [17, 18, 19], the following case was considered for models in mean field games
\[
b_\theta(t, x, \mu, \mu) = \langle f_\theta(x, \cdot), \mu \rangle, \quad \text{and} \quad \sigma_\theta(t, x, \mu, \mu) = \sigma I_d,
\]
for each \( \theta \in \mathbb{K}, \) where \( \sigma \) is a positive number and the function \( f_\theta(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) is bounded and Lipschitz continuous. Again, according to (5), Assumption (A) is satisfied for \( p = 1. \)

We have the following proposition. For convenience, its proof is given in Appendix A.1.

**Proposition 1.** Let \( p \geq 1. \) Under Assumption (A), if \( \sup_{i \geq 1, j \in \mathbb{K}} \mathbb{E}|x_{ij}|^p < \infty, \) then the following assertions hold.

(i) The equation (1) has a unique solution.
(ii) There exists a constant $C$ that depends only on $p$ such that
\[
\sup_{N \geq 1} \max_{1 \leq i \leq N} \mathbb{E} \left( \sup_{0 \leq t \leq T} \left| x_i^{(N)}(t) \right|^p \right) \\
\leq C \left( T^{p/2} + T^p + \sup_{i \geq 1, j \in \mathbb{K}} \mathbb{E} \left| x_{ij}^0 \right|^p \right) e^{CT^p} < \infty.
\] (6)

2.2. Limiting McKean-Vlasov processes. In this section, we will prove the existence, uniqueness, and some properties of of McKean-Vlasov stochastic differential equations associated to the equation (1). Similar to the classical McKean-Vlasov equation, since $\mu_i^{(N)}(t)$ is the empirical distribution of particles in population $K_j$, in view of (2), we would expect that, as $N \to \infty$, the limit $(\mu(t), \underline{\mu}(t))$ of $(\mu^{(N)}(t), \underline{\mu}^{(N)}(t))$ satisfies the following limiting equations
\[
y_{ij}(t) = b_j \left( t, y_j(t), \mu(t), \underline{\mu}(t) \right) dt + \sigma_j \left( t, y_j(t), \mu(t), \underline{\mu}(t) \right) d\bar{B}_j(t), \\
y_{ij}(0) = y_{ij}^0, \quad 1 \leq j \leq K, \\
\underline{\mu}(t) = (\mu_1(t), \ldots, \mu_K(t)) \text{ with } \mu_j(t) = \mathcal{L}(y_j(t)), \\
\mu(t) = \sum_{j=1}^K \nu_j \mu_j(t),
\] (7)
where $\bar{B}_1(t), \bar{B}_2(t), \ldots, \bar{B}_K(t)$ are $d$-dimensional standard Brownian motions, the initial value $y_{ij}^0$ has the same distribution as that of $x_{ij}^0$ for each $j \in \mathbb{K}$, and the set $\{y_{ij}^0, \bar{B}_j(\cdot) : j \in \mathbb{K}\}$ is independent and defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

Let $\mathcal{C}_t = C([0, t], (\mathbb{R}^d)^K)$ be the set of continuous functions $f : [0, t] \to (\mathbb{R}^d)^K$ equipped with the usual supremum norm and $\mathcal{P}_t = \mathcal{P}(\mathcal{C}_t)$ be the set of all probability measures on $\mathcal{C}_t$. Let us also introduce the $t^{th}$ Wasserstein distance on $\mathcal{P}_t$ as follows
\[
W_{p,t}(M_1, M_2) = \inf \left\{ \left[ \mathbb{E} \sup_{s \in [0, t]} |Y_1(s) - Y_2(s)|^p \right]^{1/p} : (Y_1, Y_2) \in \mathcal{C}_t \times \mathcal{C}_t, \\
M_1 = \mathcal{L}(Y_1), M_2 = \mathcal{L}(Y_2) \right\}.
\]
It is well-known that $(\mathcal{P}_t, W_{p,t})$ is a complete metric space (see [4, Theorem 2.5] for more details). For $1 \leq j \leq K$ let $e_j : (x_1, x_2, \ldots, x_K) \in (\mathbb{R}^d)^K \to x_j \in \mathbb{R}^d$ be the $j$th projection. For $s \leq t$ denote $p_x : f(\cdot) \in \mathcal{C}_t \to f(s) \in (\mathbb{R}^d)^K$.

The McKean-Vlasov equation (7) can be rewritten in the following equivalent form
\[
Y(t) = b(t, Y(t), M(t))dt + \sigma(t, Y(t), M(t))d\bar{B}(t), \quad 0 \leq t \leq T, \\
Y(0) = Y_0, \\
M(t) = \mathcal{L}(Y(t)), \quad 0 \leq t \leq T,
\] (8)
where $Y_0 = (y_{i1}^0, \ldots, y_{iK}^0)^\top$, $Y(t) = (y_1(t), \ldots, y_K(t))^\top$, and
\[
\bar{B}(t) = \left( \bar{B}_1(t), \ldots, \bar{B}_K(t) \right)^\top,
\]
\[
b(t, Y(t), M(t)) = (b_1(t, y_1(t), m(t), \underline{m}(t)), \ldots, b_K(t, y_K(t), m(t), \underline{m}(t)))^\top,
\]
\[
\sigma(t, Y(t), M(t)) = \text{diag} \left[ \sigma_1(t, y_1(t), m(t), \underline{m}(t)), \ldots, \sigma_K(t, y_K(t), m(t), \underline{m}(t)) \right],
\]
where \( m(t) = (m_1(t), \ldots, m_K(t)) \), \( m_j(t) = M(t) \circ e_j^{-1} \in \mathcal{P}(\mathbb{R}^d) \), \( j \in \mathbb{K} \), and \( m(t) = \sum_{j \in \mathbb{K}} \nu_j m_j(t) \).

For \( M \in \mathcal{P}_T \) denote \( M(t) = M \circ p_t^{-1} \). Define \( \Psi \) and \( \Phi \) the mappings which associate to each \( M \in \mathcal{P}_T \) the unique solution of the following equation and its law respectively

\[
Y(t) = b(t, Y(t), M(t))dt + \sigma(t, Y(t), M(t))d\tilde{B}(t), \quad 0 \leq t \leq T,
\]
\[
Y(0) = Y^0,
\]
i.e., \( \Psi(M) = Y \) is a solution of equation (9) and \( \Phi(M) = \mathcal{L}(\mathcal{Y}) \in \mathcal{P}_T \) is the law of \( Y \). Note that under Assumption (A), the functions \( b(\cdot, \cdot, M(t)) : [0, T] \times (\mathbb{R}^d)^K \to (\mathbb{R}^d)^K \) and \( \sigma(\cdot, M(t)) : [0, T] \times (\mathbb{R}^d)^K \to \mathbb{R}^{dK \times dK} \) also satisfy linear growth and Lipschitz conditions for any \( M(t) \in \mathcal{P}(\mathbb{R}^d)^K \). Therefore, equation (9) has a unique solution and \( \Psi \) and \( \Phi \) are well defined. Moreover, by a typical argument, we can show that \( \mathbb{E} \sup_{0 \leq s \leq T} |Y(s)|^p < \infty \) where \( Y = \Psi(M) \) for any \( M \in \mathcal{P}_T \). This implies that

\[
W_{p,T}(\Phi(M), \Phi^2(M)) < \infty, \quad \text{for all } M \in \mathcal{P}_T.
\]

Observe that if \( Y \) is a solution of equation (8) then its law is a fixed point of \( \Phi \), and conversely if \( \mu \) is such a fixed point of \( \Phi \), equation (9) defines a solution of equation (8). We have the following lemma which asserts that \( \Phi \) is a contraction mapping on \( \mathcal{P}_T \). Its proof is aggregated in Appendix A.2.

**Lemma 2.2.** For any \( 0 \leq t \leq T \) and measures \( M_1, M_2 \in \mathcal{P}_T \) we have

\[
|W_{p,t}(\Phi(M_1), \Phi(M_2))|^\varrho \leq C \int_0^t |W_{p,s}(M_1, M_2)|^\varrho \, ds,
\]

where \( \varrho = \max \{p, 2\} \) and \( C \) is a constant depending only on \( T \).

Fix a measure \( M_0 \in \mathcal{P}_T \). We consider the following recursive formula

\[
M_{k+1} = \Phi(M_k) = \Phi^{k+1}(M_0), \quad k \geq 0.
\]

Then (10) implies that \( W_{p,T}(M_1, M_2) < \infty \). Since \( W_{p,s}(M_1, M_2) \leq W_{p,t}(M_1, M_2) \) for \( 0 \leq s \leq t \), iterating inequality (11) yields

\[
W_{p,T}(M_{k+1}, M_{k+2}) \leq \frac{(CT)^{k+1}}{(k!)^\varrho} W_{p,T}(M_1, M_2),
\]

for all \( k \geq 0 \). It follows from inequality (12) that \( \{M_k\}_{k \geq 0} \) is a Cauchy sequence in complete metric space \( (\mathcal{P}_T, W_{p,T}) \). Therefore, this sequence converges to some measure \( M \) in \( \mathcal{P}_T \) which is a fixed point of \( \Phi \) on \( \mathcal{P}_T \). In addition, if \( t < T \) then the image of \( \mu(T) \) on \( C([0, t], (\mathbb{R}^d)^K) \) is also a fixed point of \( \Phi \). This property of consistency gives a fixed point \( \mu \) on \( C([0, \infty], (\mathbb{R}^d)^K) \), confirming the existence of solution of equation (7). The uniqueness of this solution also follows from Theorem 2.2. Thus, we have just established the following theorem.

**Theorem 2.3.** Assume that Assumption (A) holds. Then there exists a unique solution to the limiting equation (7).

The next proposition gives upper bounds for moments of solutions \((y_1, y_2, \ldots, y_K)\) of equation (7). Its proof is placed in Appendix A.3 to keep the presentation more transparent.
Proposition 2. Let $p \geq 1$. Under Assumption (A), if $\max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p < \infty$ then for any $T > 0$, there exists a constant $C$ such that

$$\max_{j \in \mathbb{K}} \left( \mathbb{E} \sup_{0 \leq t \leq T} |y_j(t)|^p \right) \leq C \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} < \infty. \quad (13)$$

Furthermore,

$$\max_{j \in \mathbb{K}} \left( \mathbb{E} \sup_{0 \leq r \leq s} |y_j(t + r) - y_j(t)|^p \right) \leq C(s^{p/2} + s^p) \left[ 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} \right], \quad (14)$$

where $T = t + s$ and the constant $C$ depends only on $p$.

The following theorem asserts the continuity of $\mu(t)$ in $t$ with respect to the Monge-Wasserstein distance and provides a bound for $W_p(\mu_j(0), \mu_j(t))$ for $j \in \mathbb{K}$ and $t > 0$. In order to achieve this, one needs an estimation for empirical measures of independent and identically distributed random vectors which requires the boundedness of higher-order moments of the initial values. More precisely, we shall assume that $\max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^q < \infty$ for some $q > p$ (see Theorem 3.1). The proof of the theorem below is postponed in Section 3.2.

Theorem 2.4. Let $p \geq 1$. Assume that $\max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^q < \infty$ for some $q > p$ and that Assumption (A) holds. Then there exists a constant $C$ such that for all $s > 0$ and $t > 0$,

$$|W_p(\mu(t + s), \mu(t))|^p \leq C(s^{p/2} + s^p) \left[ 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} \right],$$

and for each $j \in \mathbb{K}$,

$$|W_p(\mu_j(t + s), \mu_j(t))|^p \leq C(s^{p/2} + s^p) \left[ 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} \right],$$

where $T = t + s$ and the constant $C$ depends only on $p$.

2.3. Limiting system and approximation in Monge-Wasserstein distance. Let $y_j(t)$, $1 \leq j \leq K$, be the solution of equation (7), $\mu_j(t) = \mathcal{L}(y_j(t))$, $\mu(t) = \sum_{j=1}^K \nu_j \mu_j(t)$, and $\underline{\mu}(t) = (\mu_1(t), \mu_2(t), \ldots, \mu_K(t))$. In this section we will consider the system of infinite particles associated to (1) described by following equations

$$x_i(t) = b_{\theta_i} \left( t, x_i(t), \mu(t), \underline{\mu}(t) \right) dt + \sigma_{\theta_i} \left( t, x_i(t), \mu(t), \underline{\mu}(t) \right) dB_i(t), \quad 0 \leq t \leq T,$$

$$x_i(0) = x_{i,\theta_i}^0,$$  

(15)

for all $i = 1, 2, \ldots$, where the initial values $x_{i,\theta_i}^0$ and the Brownian motions $B_i(t)$ are as in equation (1). It can be shown that under Assumption (A), for each $i \geq 1$, equation (15) has a unique solution. Note that since $(B_i(t), B_j(t) : i = 1, 2, \ldots; j = 1, 2, \ldots, K)$ are independent and identically distributed, $(x_{i,j}^0, y_{j}^0 : i = 1, 2, \ldots; j = 1, 2, \ldots, K)$ are independent, and $(x_{i,j}^0, y_{j}^0 : i = 1, 2, \ldots)$ are identically distributed for each $j = 1, 2, \ldots, K$, we conclude that for all $i \in \mathcal{K}$, $x_i(t)$ and $y_j(t)$ have the same distribution $\mu_j(t)$, i.e. $\mathcal{L}(x_i(t)) = \mathcal{L}(y_j(t)) = \mu_j(t)$. Therefore, as a direct
consequence of Proposition 2, if \( \sup_{i \geq 1} \mathbb{E} |x_{i,\theta_i}^0|^p < \infty \) then

\[
\sup_{i \geq 1} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |x_i(t)|^p \right] = \max_{j \in \mathbb{K}} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |y_j(t)|^p \right] \leq C \left( T^{p/2} + \sup_{i \geq 1} \mathbb{E} |x_{i,\theta_i}^0|^p \right) e^{CT^p} < \infty. \tag{16}
\]

Next, we will estimate the rates of convergence of \( x_i^{(N)}(t) \) and \( \mu_j^{(N)}(t) \) to \( x_i(t) \) and \( \mu_j(t) \), respectively, in \( L_p \) and \( W_p \) senses. To do so, similar to the simpler case of approximating a distribution by empirical measures of independent and identically distributed random vectors (see Theorem 3.1), we need stronger conditions on the moments of the initial values. More precisely, we assume that \( \max_{i,j} \mathbb{E} |x_{ij}^0|^q < \infty \) for some \( q > \max \{p, 2\} \). Define

\[
g_1(p, q, N) = h(p, q, N) + h_\nu(N),
g_2(p, q, N) = [g_1(2, q, N)]^{p/2},
g(p, q, N) = \begin{cases} g_1(p, q, N), & \text{if } p \geq 2, 
g_2(p, q, N), & \text{if } 0 \leq p < 2, \end{cases} \tag{17}
\]

where

\[
h(p, q, N) = \begin{cases} N^{-1/2} + N^{-(q-p)/q}, & \text{if } p > d/2 \text{ and } q \neq 2p, 
N^{-1/2} \ln(1 + N) + N^{-(q-p)/q}, & \text{if } p = d/2 \text{ and } q \neq 2p, 
N^{-p/d} + N^{-(q-p)/q}, & \text{if } p \in (0, d/2) \text{ and } q \neq d/(d-p), \end{cases} \tag{18}
\]

and

\[
h_\nu(N) = \sup_{j \in \mathbb{K}} \left| \frac{\mu_j^{(N)}}{N} - \nu_j \right|. \tag{19}
\]

It is useful to notice that \( g_1(p, q, N) \leq C g_2(p, q, N) \) for \( 0 \leq p < 2 \), and thus \( g_1(p, q, N) \leq g(p, q, N) \) for all \( p, q, N \). The function \( h(p, q, N) \) comes from an approximation of a distribution by empirical measures (see [14, Theorem 1] or Theorem 3.1). For simplicity, \( h(p, q, N) \) is only defined for \( q \neq 2p \) if \( p \geq d/2 \) and \( q \neq d/(d-p) \) when some complicated terms involving logarithm functions will appear the two remaining cases as pointed out in [14].

Now we are in a position to state the following estimate the error in \( L_p \) sense of the approximation of finite systems by the limiting one. Its proof is provided in Section 3.3.

**Theorem 2.5.** Let \( p \geq 1 \). Under Assumption (A), if \( \max_{i,j} \mathbb{E} |x_{ij}^0|^q < \infty \) for some \( q > \max \{p, 2\} \), then there is a constant \( C \) independent of \( N \) such that

\[
\max_{1 \leq i \leq N} \mathbb{E} \left( \sup_{0 \leq t \leq T} |x_i^{(N)}(t) - x_i(t)|^p \right) \leq C g(p, q, N). \tag{20}
\]

Here \( g(p, q, N) \) is defined as in equation (17).

The approximation in Monge-Wasserstein distance of empirical measures in finite systems by the limiting measure is given in the following theorem for which the proof is aggregated in Section 3.4.
Theorem 2.6. Assume that Assumption (A) holds and \( \max_{i,j} \mathbb{E} |x_{ij}^0|^q < \infty \) for some \( q > \max \{p, 2\} \). Then
\[
\sup_{0 \leq t \leq T} \mathbb{E} \left| W_p \left( \mu^{(N)}(t), \mu(t) \right) \right|^p \leq C g(p, q, N).
\]
Moreover, for each \( j \in \mathcal{K} \),
\[
\sup_{0 \leq t \leq T} \mathbb{E} \left| W_p \left( \mu_j^{(N)}(t), \mu_j(t) \right) \right|^p \leq C g(p, q, N).
\]
As a consequence, \( \mu^{(N)}(t) \) converges weakly to \( \mu(t) \) for each \( t \in [0, T] \).

3. Proofs of main theorems.

3.1. Auxiliary results. We provide the following lemma without its proof. One may refer to [14, Theorem 1] for more details. For \( q > 0 \) and \( \eta \) be a probability measure in \( \mathcal{P} \), we define
\[
M_q(\eta) = \int |x| \eta(dx).
\]

Lemma 3.1. Let \( \eta \) be a probability measure in \( \mathcal{P} \) such that \( M_q(\eta) < \infty \) for some \( q > p > 0 \). Let \( \eta_N \) be the corresponding empirical measure of a sequence of \( N \) independent, \( \eta \)-distributed, and \( \mathbb{R}^d \)-valued random vectors \( X_1, X_2, \ldots, X_N \), i.e.,
\[
\eta_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}.
\]

Then there exists a constant \( C \) independent of \( N \) such that
\[
\mathbb{E} \left| W_p \left( \eta_N, \eta \right) \right|^p \leq C \left( M_q(\eta) \right)^{p/q} h(p, q, N),
\]
where \( h(p, q, N) \) is defined as in (18).

Remark 1. Similar to Proposition 2, it is easy to verify that if \( \sup_{j \in \mathcal{K}} \mathbb{E} |y_{ij}^0|^q < \infty \) for some \( q > p \) then
\[
\sup_{r \in [0, t]} M_q(\mu_j(r)) = \sup_{r \in [0, t]} \mathbb{E} |y_j(r)|^q < \infty.
\]
Thus, Theorem 3.1 implies that for all \( j = 1, \ldots, K \),
\[
\sup_{r \in [0, t]} \mathbb{E} \left| W_p \left( \overline{\mu}_j^{(N)}(r), \mu_j(r) \right) \right|^p \leq C h \left( p, q, \nu_j^{(N)}(r) \right),
\]
where
\[
\overline{\mu}_j^{(N)}(t) = \frac{1}{\nu_j^{(N)}} \sum_{1 \leq i \leq N, \theta_i = j} \delta_{x_i(t)}.
\]

Besides, it follows from (3) and (18) that \( h \left( p, q, \nu_j^{(N)}(r) \right) \) can be bounded by \( h(p, q, N) \) as follows
\[
h \left( p, q, \nu_j^{(N)}(r) \right) \leq C h \left( p, q, \nu_j^{(N)}(r) \right) \frac{1}{N} h(p, q, N) \leq C h(p, q, N) \quad \text{for all} \ j \in \mathcal{K}.
\]
Hence,
\[
\max_{j \in \mathcal{K}} \sup_{r \in [0, t]} \mathbb{E} \left| W_p \left( \overline{\mu}_j^{(N)}(r), \mu_j(r) \right) \right|^p \leq C h(p, q, N),
\]
where the constant \( C \) is independent of \( N \).

It is useful to mention the following property of Monge-Wasserstein distance. Its proof is accumulated in Appendix A.4.
Lemma 3.2. For each $1 \leq j \leq K$, let $\eta_j$ and $\varpi_j$ be probability measures and $\lambda_j$ and $\pi_j$ be nonnegative numbers such that $\sum_{j=1}^{K} \lambda_j = 1$ and $\sum_{j=1}^{K} \pi_j = 1$. Put $\eta = \sum_{j=1}^{K} \lambda_j \eta_j$ and $\varpi = \sum_{j=1}^{K} \pi_j \varpi_j$. Then for all $p > 1$,

$$\big|W_p(\eta, \varpi)\big|^p \leq C \sum_{j=1}^{K} \lambda_j \big|W_p(\eta_j, \varpi_j)\big|^p + C \sum_{j=1}^{K} |\lambda_j - \pi_j| M_p(\varpi_j),$$

where $C$ is a constant depending only on $p$ and $M_p(\varpi_j)$ is defined as in (4).

Remark 2. It can be derived from Theorem 3.2 and its proof that

(i) If $\lambda_j = \pi_j$ for $j \in \mathbb{I}$, that is, $\varpi = \sum_{j=1}^{K} \lambda_j \varpi_j$, then for all $p \geq 1$,

$$\big|W_p(\eta, \varpi)\big|^p \leq \sum_{j=1}^{K} \lambda_j \big|W_p(\eta_j, \varpi_j)\big|^p.$$

(ii) Moreover, if $u_i, v_i, 1 \leq i \leq n$, are $2n$ vectors in $\mathbb{R}^d$, $\eta_i = \varpi_i = \delta_{u_i}$, and $\lambda_i = \pi_i = \frac{1}{n}$ for all $1 \leq i \leq n$; that is, $\eta = \frac{1}{n} \sum_{i=1}^{n} \delta_{u_i}$ and $\varpi = \frac{1}{n} \sum_{i=1}^{n} \delta_{v_i}$, then for all $p \geq 1$,

$$\big|W_p(\eta, \varpi)\big|^p \leq \frac{1}{n} \sum_{i=1}^{n} |u_i - v_i|^p.$$

In order to prove Proposition 1 and Proposition 2, we need to use the following lemma which is a generalization of Gronwall’s inequality. See Appendix A.5 for its proof.

Lemma 3.3. Let $\mathcal{I}$ be an index set. For each $i \in \mathcal{I}$, let $(Z_i(t))_{t \geq 0}$ be a nonnegative stochastic process and $V_i(t) = \sup_{0 \leq s \leq t} Z_i(s)$. Assume that for some constants $p > 0$, $q > 0$, $\delta \geq 0$, $K_1 \geq 0$, $K_2 \geq 0$,

$$\sup_{i \in \mathcal{I}} \mathbb{E}[V_i(t)]^p \leq K_1 \sup_{i \in \mathcal{I}} \mathbb{E} \int_0^t \big[V_i(s)\big]^p ds + K_2 \sup_{i \in \mathcal{I}} \mathbb{E} \left[\int_0^t \big[Z_i(s)\big]^q ds\right]^{p/q} + \delta < \infty,$$

for all $t \in [0, T]$. Then there exists a constant $C$ that depends on $K_1$, $K_2$, $p$ and $q$ such that

$$\sup_{i \in \mathcal{I}} \mathbb{E}[V_i(T)]^p \leq 2\delta \exp \left(2K_1 + CK_2 T^{\max(p/q, 1)}\right).$$

In particular, if $q = 2$ and $p \geq 1$ then $C < 8$, and therefore,

$$\sup_{i \in \mathcal{I}} \mathbb{E}[V_i(T)]^p \leq 2\delta \exp \left(2K_1 T + 8K_2 T^{\max(p/2, 1)}\right).$$

3.2. Proof of Theorem 2.4. For each $i = 1, \ldots, N$, let $x_i(\cdot)$ be the solution of equation (15). Let $\mu(\cdot)$ and $\bar{\mu}(\cdot)$ be given as in (7) and $\bar{\mu}_j^{(N)}(\cdot)$ given as in (21). Then Theorem 3.2 yields an upper bound for the distance between two empirical measures

$$\max_{j \in \mathbb{K}} \mathbb{E} \left[ W_p(\bar{\mu}_j^{(N)}(t+s), \bar{\mu}_j^{(N)}(t)) \right]^p \leq \max_{j \in \mathbb{K}} \frac{1}{\nu_j^{(N)}} \mathbb{E} \left[ \sum_{1 \leq i \leq N, \delta_i = j} \left[ W_p(\delta_{x_i(t+s)}, \delta_{x_i(t)}) \right]^p \right].$$
we obtain

$$\leq \max_{j \in \mathbb{K}} \frac{1}{\nu_j \mathbb{N}} \mathbb{E} \sum_{1 \leq i \leq N, \theta_i = j} |x_i(t + s) - x_i(t)|^p$$

Applying Cauchy’s inequality, Hölder’s inequality, and Burkholder-Davis-Gundy inequality on the following equation

$$x_i(t + s) - x_i(t) = \int_t^{t+s} b_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \, dr + \int_t^{t+s} \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \, dB_i(r),$$

we obtain

$$\max_{i = 1, \ldots, N} \mathbb{E} (|x_i(t + s) - x_i(t)|^p) \leq C \max_{i = 1, \ldots, N} \mathbb{E} \left\{ \left( \int_t^{t+s} b_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \, dr \right)^p \right\}$$

$$+ C \max_{i = 1, \ldots, N} \mathbb{E} \left\{ \left( \int_t^{t+s} \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \, dB_i(r) \right)^p \right\} \leq C s^{p-1} \max_{i = 1, \ldots, N} \mathbb{E} \left( \int_t^{t+s} \left| b_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \right|^p \, dr \right)$$

$$+ C \max_{i = 1, \ldots, N} \mathbb{E} \left( \int_t^{t+s} \left| \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \right|^2 \, dr \right)^{p/2}. \leq C \left( 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} \right),$$

Thus, by Cauchy’s inequality, Assumption (A1), and inequality (16),

$$\max_{i = 1, \ldots, N} \mathbb{E} \left( \int_t^{t+s} \left| \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \right|^2 \, dr \right)^{p/2} \leq C \left( 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CT^p} \right)$$

where \( m_j = \min \{ l \in \mathbb{N} : 1 \leq l \leq N \text{ and } \theta_l = j \}. \) Similarly, we can also use inequality (16) to obtain an upper bound related to \( \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \) as follows

$$\leq \max_{i = 1, \ldots, N} \mathbb{E} \left( \left( \int_t^{t+s} \left| \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \right|^2 \, dr \right)^{p/2} \right) \leq C s^{p/2} \max_{i = 1, \ldots, N} \left( 1 + \sup_{r \in [t, t+s]} \left| \sigma_{\theta_i} (r, x_i(r), \mu(r), \underline{\mu}(r)) \right|^p \right) \leq C s^{p/2} \max_{i = 1, \ldots, N} \left( 1 + \sup_{r \in [t, t+s]} |x_i(r)|^p + \sum_{j = 1}^K \sup_{r \in [t, t+s]} |x_{m_j}(r)|^p \right) \leq C s^{p/2} \max_{i = 1, \ldots, N} \left( 1 + \sup_{r \in [t, t+s]} |x_i(r)|^p + \sum_{j = 1}^K \sup_{r \in [t, t+s]} |x_{m_j}(r)|^p \right).$$
On the other hand, Remark 1 shows that

\[ \max_{j \in \mathbb{K}} \mathbb{E} \left[ |W_p(E_j(N)(t+s), E_j(N)(t))|^p \right] \]

Hence, on one hand, it follows from inequalities (23) to (26) that

\[ \max_{j \in \mathbb{K}} \mathbb{E} \left[ |W_p(E_j(N)(t+s), E_j(N)(t))|^p \right] \leq C(s^{p/2} + s^p) \left[ 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CTp} \right]. \]

(27)

On the other hand, Remark 1 shows that

\[ \mathbb{E} \left[ |W_p(E_j(N)(t), \mu_j(t))|^p \right] \leq Ch(p, q, N), \]

and

\[ \mathbb{E} \left[ |W_p(E_j(N)(t+s), \mu_j(t+s))|^p \right] \leq Ch(p, q, N). \]

(29)

Therefore, by virtue of triangle inequality, we can deduce from inequalities (27) to (29) that

\[ \mathbb{E} |W_p(\mu_j(t+s), \mu_j(t))|^p \leq C(s^{p/2} + s^p) \left[ 1 + \left( T^{p/2} + T^p + \max_{j \in \mathbb{K}} \mathbb{E} |y_j(0)|^p \right) e^{CTp} \right] \]

\[ + Ch(p, q, N), \]

where \( C \) is a constant independent of \( N \). Letting \( N \to \infty \) in this inequality yields the desired estimate.

\[ \square \]

3.3. Proof of Theorem 2.5. Applying Cauchy’s inequality, Hölder’s inequality, and Burkholder-Davis-Gundy inequality on the following equation

\[ x_i^{(N)}(t) - x_i(t) \]

\[ = \int_0^t \left[ b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \underline{\mu}^{(N)}(r) \right) - b_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right] dr \]

\[ + \int_0^t \left[ \sigma_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \underline{\mu}^{(N)}(r) \right) - \sigma_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right] dB_i(r), \]

we obtain

\[ \sup_{1 \leq i \leq N} \mathbb{E} \left[ \left| \sup_{0 \leq s \leq t} \left[ x_i^{(N)}(s) - x_i(s) \right] \right|^p \right] \]

\[ \leq C \sup_{1 \leq i \leq N} \mathbb{E} \left\{ \sup_{0 \leq s \leq t} \left[ \int_0^s \left[ b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \underline{\mu}^{(N)}(r) \right) \right. \right. \right. \]

\[ - b_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right. \]

\[ \left. \left. - b_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right) dr \right|^p \}\]

\[ + C \sup_{1 \leq i \leq N} \mathbb{E} \left\{ \sup_{0 \leq s \leq t} \left[ \int_0^s \left[ \sigma_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \underline{\mu}^{(N)}(r) \right) \right. \right. \right. \]

\[ - \sigma_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right. \]

\[ \left. \left. \left. - \sigma_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right) dB_i(r) \right|^p \right\} \]

\[ \leq C T^{p-1} \sup_{1 \leq i \leq N} \mathbb{E} \int_0^t \left[ b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \underline{\mu}^{(N)}(r) \right) - b_{\theta_i} \left( r, x_i(r), \mu(r), \underline{\mu}(r) \right) \right]^p dr \]
\begin{align*}
&+ C \sup_{1 \leq i \leq N} \mathbb{E} \left\{ \left( \int_0^t \left| \sigma_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \bar{\mu}^{(N)}(r) \right) - \sigma_{\theta_i} (r, x_i(r), \mu(r), \bar{\mu}(r)) \right|^2 \, dr \right)^{p/2} \right\} \\
&=: A_1 + A_2.
\end{align*}

We will estimate two terms on the right hand side of inequality (30). First, by Assumption (A2) we have
\begin{align*}
A_1 := C t^{p-1} \sup_{1 \leq i \leq N} \mathbb{E} \int_0^t \left| b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \bar{\mu}^{(N)}(r) \right) - b_{\theta_i} (r, x_i(r), \mu(r), \bar{\mu}(r)) \right|^p \, dr
&= C t^{p-1} \sup_{1 \leq i \leq N} \mathbb{E} \int_0^t \sum_{j=1}^K \mathbb{I}_{\{\theta_i=j\}} \\
&\quad \times \left| b_j \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \bar{\mu}^{(N)}(r) \right) - b_j (r, x_i(r), \mu(r), \bar{\mu}(r)) \right|^p \, dr
\leq C t^{p-1} \int_0^t \left[ \sup_{1 \leq i \leq N} \mathbb{E} \left| x_i^{(N)}(r) - x_i(r) \right|^p \right] + \mathbb{E} \left| W_p (\mu^{(N)}(r), \mu(r)) \right|^p
&\quad + \sum_{j=1}^K \mathbb{E} \left| W_p (\bar{\mu}_j^{(N)}(r), \mu_j(r)) \right|^p \right] \, dr.
\end{align*}

Put \( \bar{\mu}^{(N)}(r) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i(r)} \). Let \( \bar{\mu}_j^{(N)}(r) \) be defined as in (21). In order to estimate \( W_p \left( \mu^{(N)}(r), \mu(r) \right) \), we will estimate \( W_p \left( \mu^{(N)}(r), \bar{\mu}^{(N)}(r) \right) \) and \( W_p \left( \bar{\mu}^{(N)}(r), \mu(r) \right) \).

It is easy to see that for each \( j \in \mathbb{K} \), \( \{x_i(r) : i \in K_j \} \) is an independent sequence of \( \mu_j(r) \)-distributed random vectors. By Remark 1, for each \( j \in \mathbb{K} \),
\begin{align*}
\mathbb{E} \left| W_p \left( \bar{\mu}_j^{(N)}(r), \mu_j(r) \right) \right|^p \leq C h(p, q, N).
\end{align*}

In addition, equation (19) implies \( \left| \frac{\mu_j^{(N)}}{N} - \nu_j \right| \leq h_\nu(N) \) for all \( j = 1, \ldots, K \). Thus, according to equation (3) and the fact \( \max_{j \in K} M_p(\mu_j(r)) < \infty \), we can derive from Theorem 3.2 that
\begin{align*}
\mathbb{E} \left| W_p \left( \bar{\mu}^{(N)}(r), \mu(r) \right) \right|^p
&\leq \mathbb{E} \left| W_p \left( \frac{1}{N} \sum_{j=1}^K \nu_j^{(N)} \bar{\mu}_j^{(N)}(r), \sum_{j=1}^K \nu_j \mu_j(r) \right) \right|^p
\leq C \sum_{j=1}^K \frac{\nu_j^{(N)}}{N} \mathbb{E} \left| W_p \left( \bar{\mu}_j^{(N)}(r), \mu_j(r) \right) \right|^p + C \sum_{j=1}^K \left| \frac{\nu_j^{(N)}}{N} - \nu_j \right| M_p(\mu_j(r))
\leq C [h(p, q, N) + h_\nu(N)] = C g_1(p, q, N).
\end{align*}

Besides, Theorem 3.2 (see (60)) also implies
\begin{align*}
\mathbb{E} \left| W_p \left( \mu^{(N)}(r), \bar{\mu}^{(N)}(r) \right) \right|^p
&\leq \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right)
\leq \sup_{1 \leq i \leq N} \mathbb{E} \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right).
\end{align*}
Therefore, a combination of inequalities (32) and (33) gives
\[
E \left| W_p \left( \mu^{(N)}(r), \mu(r) \right) \right|^p \\
\leq C \left( E \left| W_p \left( \mu^{(N)}(r), \mu^{(N)}(r) \right) \right|^p + E \left| W_p \left( \mu^{(N)}(r), \mu(r) \right) \right|^p \right) \\
\leq C \sum_{1 \leq i \leq N} E \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right) + C g_1(p, q, N).
\]
A similar argument also yields
\[
\sum_{j=1}^{K} E \left| W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^p \\
\leq \sum_{j=1}^{K} E \left| W_p \left( \mu_j^{(N)}(r), \mu_j^{(N)}(r) \right) \right|^p + \sum_{j=1}^{K} E \left| W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^p \\
\leq C \sum_{1 \leq i \leq N} E \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right) + C g_1(p, q, N).
\]
It follows from inequalities (31), (34) and (35) that
\[
A_1 \leq C \int_0^t \sup_{1 \leq i \leq N} E \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right) dr + C g_1(p, q, N).
\]
Next, in view of Assumption (A2) we get
\[
A_2 := C \sup_{1 \leq i \leq N} \left[ \left( \int_0^t \left| \sigma \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \mu(r) \right) \right|^2 dr \right)^{p/2} \right] \\
\leq C \sup_{1 \leq i \leq N} \left( \int_0^t \left| x_i^{(N)}(r) - x_i(r) \right|^2 dr \right)^{p/2} \\
+ C E \left( \int_0^t \left| W_p \left( \mu^{(N)}(r), \mu(r) \right) \right|^2 dr \right)^{p/2} \\
+ C \sup_{1 \leq i \leq N} \left( \int_0^t \left| \sum_{j=1}^{K} \left| W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^2 dr \right)^{p/2}.
\]
To proceed, we will estimate the last two terms in the right most hand side. We start with the last term. The second term then can be done in a similar way.

If \( p \geq 2 \) then Hölder’s inequality and (35) imply
\[
E \left( \int_0^t \left| \sum_{j=1}^{K} W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^2 dr \right)^{p/2} \\
\leq C E \left( \int_0^t \left| \sum_{j=1}^{K} W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^p dr \right) \\
\leq C \int_0^t \sup_{1 \leq i \leq N} E \left| x_i^{(N)}(r) - x_i(r) \right|^p dr + C g_1(p, q, N).
\]
If $0 \leq p < 2$ then Hölder’s inequality, Jensen’s inequality, and (35) yield

$$
\mathbb{E} \left( \int_0^t \sum_{j=1}^K \left| \mathcal{W}_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^2 dr \right)^{p/2} \\
\leq \left( \mathbb{E} \int_0^t \sum_{j=1}^K \left| \mathcal{W}_2 \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^2 dr \right)^{p/2} \\
\leq C \left( \int_0^t \sup_{1 \leq i \leq N} \mathbb{E} \left| x_i^{(N)}(r) - x_i(r) \right|^2 dr + g_1(2, q, N) \right)^{p/2} \\
\leq C K \sup_{1 \leq i \leq N} \left( \int_0^t \mathbb{E} \left| x_i^{(N)}(r) - x_i(r) \right|^2 dr \right)^{p/2} + C (g_1(2, q, N))^{p/2}.
$$

Note that in the last inequality we have used the fact that for all $i \in K_j$, the expectations $\mathbb{E} \left| x_i^{(N)}(r) - x_i(r) \right|^2$ remains the same for each fixed $j \in K$. By using (34) instead of (35), we can obtain a similar estimate for the second term in the right most hand side of (37). Therefore, for all $p \geq 1$,

$$
A_2 \leq C \int_0^t \sup_{1 \leq i \leq N} \mathbb{E} \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right) dr \\
+ C \sup_{1 \leq i \leq N} \mathbb{E} \left( \int_0^t \left| x_i^{(N)}(r) - x_i(r) \right|^2 dr \right)^{p/2} + C g(p, q, N).
$$

As a result, inequalities (30), (36) and (38) leads to

$$
\sup_{1 \leq i \leq N} \mathbb{E} \left( \sup_{0 \leq s \leq t} \left| x_i^{(N)}(s) - x_i(s) \right|^p \right) \\
\leq C \int_0^t \sup_{1 \leq i \leq N} \mathbb{E} \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right) dr \\
+ C \sup_{1 \leq i \leq N} \mathbb{E} \left( \int_0^t \left| x_i^{(N)}(r) - x_i(r) \right|^2 dr \right)^{p/2} + C g(p, q, N),
$$

which, in virtue of Theorem 3.3 with the index set $\mathcal{I} = \{1, \ldots, N\}$, $\delta = C g(p, q, N)$, and $q = 2$, implies the desired estimate (20).

### 3.4. Proof of Theorem 2.6

In view of Theorem 2.5 and the fact that $g_1(p, q, N) \leq g(p, q, N)$, inequality (34) yields

$$
\sup_{0 \leq t \leq T} \mathbb{E} \left| \mathcal{W}_p \left( \mu^{(N)}(t), \mu(t) \right) \right|^p \leq C g(p, q, N).
$$

This proves the first part of the theorem. To complete the proof, recall that

$$
\mu_j^{(N)}(t) = \frac{1}{\nu_j^{(N)}} \sum_{1 \leq i \leq N, \theta_i = j} \delta_{x_i(t)}^{(N)} \quad \text{and} \quad \bar{\mu}_j^{(N)}(t) = \frac{1}{\nu_j^{(N)}} \sum_{1 \leq i \leq N, \theta_i = j} \delta_{x_i(t)},
$$

where

$$
\nu_j^{(N)} = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\theta_i = j}. 
$$

Note that $\nu_j^{(N)} = \frac{1}{N}$ for all $j$. So the proof is complete.

\[\square\]
for each $j \in \mathbb{K}$. By the triangle inequality,

$$
E \left| W_p \left( \mu_j^{(N)}(t), \mu_j(t) \right) \right|^p
\leq C E \left| W_p \left( \mu_j^{(N)}(t), \mu_j^{(N)}(t) \right) \right|^p + C E \left| W_p \left( \mu_j^{(N)}(t), \mu_j(t) \right) \right|^p
=: A_3 + A_4.
$$

We can estimate $A_3$ by using Theorem 3.2 and Theorem 2.5 as follows

$$
A_3 := C E \left| W_p \left( \mu_j^{(N)}(t), \mu_j^{(N)}(t) \right) \right|^p
\leq \frac{1}{\mu_j^{(N)}} E \sum_{1 \leq i \leq N, \theta_i = j} \left| x_i^{(N)}(r) - x_i(r) \right|^p
\leq C \sup_{1 \leq i \leq N} E \left( \left| x_i^{(N)}(r) - x_i(r) \right|^p \right)
\leq C g(p, q, N).
$$

Next, according to Theorem 3.1, we have the following estimate for $A_4$

$$
A_4 := E \left| W_p \left( \mu_j^{(N)}(r), \mu_j(r) \right) \right|^p
\leq C h(p, q, N).
$$

Consequently, we can derive from inequalities (40) to (42) that

$$
E \left| W_p \left( \mu_j^{(N)}(t), \mu_j(t) \right) \right|^p
\leq C g(p, q, N),
$$

which implies the second claim of the theorem.

\textbf{Appendix A.}

\textbf{A.1. Proof of Proposition 1.}

(i) For each $X = (x_1, x_2, \ldots, x_N) \in (\mathbb{R}^d)^N$, denote

$$
\delta_X^{(N)} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i},
$$
$$
\delta_X^{(N), j} = \frac{1}{\mu_j^{(N)}} \sum_{1 \leq i \leq N, \theta_i = j} \delta_{x_i}, \quad j = 1, \ldots, K,
$$
$$
\delta_X^{(N)} = \left( \delta_X^{(N), 1}, \delta_X^{(N), 2}, \ldots, \delta_X^{(N), K} \right)^\top.
$$

Also, for $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_N) \in (\mathbb{R}^d)^N$, denote the functions $b^{(N)}(\cdot, \cdot) : [0, T] \times (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ and $\sigma^{(N)}(\cdot, \cdot) : [0, T] \times (\mathbb{R}^d)^N \to (\mathbb{R}^d)^{NK}$ as follows.

$$
b^{(N)}(t, X) = \left( b_{\theta_1}(t, x_1, \delta_X^{(N)}, \delta_X^{(N)}), \ldots, b_{\theta_N}(t, x_N, \delta_X^{(N)}, \delta_X^{(N)}) \right)^\top,
$$
$$
\sigma^{(N)}(t, X) = \text{diag} \left[ \sigma_{\theta_1}(t, x_1, \delta_X^{(N)}, \delta_X^{(N)}), \ldots, \sigma_{\theta_N}(t, x_N, \delta_X^{(N)}, \delta_X^{(N)}) \right].
$$

For $N \geq 1$, denote

$$
X^{(N)}(t) = \left( x_1^{(N)}(t), x_2^{(N)}(t), \ldots, x_N^{(N)}(t) \right)^\top,
$$

$$
X_0^{(N)} = \left( x_1^{0, \theta_1}, x_2^{0, \theta_2}, \ldots, x_N^{0, \theta_N} \right)^\top,
$$
$$
B^{(N)}(t) = (B_1(t), \ldots, B_N(t))^\top \in (\mathbb{R}^d)^N.
$$
Then the system (1) can be rewritten in the form
\[ dX^{(N)}(t) = b^{(N)} \left( t, X^{(N)}(t) \right) dt + \sigma^{(N)} \left( t, X^{(N)}(t) \right) dB^{(N)}(t), \]
\[ X^{(N)}(0) = X_0^{(N)}. \]

In virtue of Remark 2(ii), there exists a constant \( C = C(N) \) such that,
\[ \mathcal{W}_p \left( \delta^{(N)}_X, \delta^{(N)}_Y \right) + \mathcal{W}_p \left( \delta^{(N)}_X, \delta^{(N)}_Y, \right) \leq C \sum_{i=1}^N |x_i - y_i|, \quad j \in \mathbb{K}. \]

Therefore, Assumption (A) implies that \( b^{(N)}(\cdot, \cdot) \) and \( \sigma^{(N)}(\cdot, \cdot) \) satisfy the linear growth and Lipschitz continuity conditions. As a consequence, the stochastic differential equation (43) has a unique solution. This yields the existence and uniqueness of solution of (1).

(ii) Now we will verify the second assertion. For all \( t > 0 \) and each \( i = 1, \ldots, N, \)
put
\[ Z_i^{(N)}(t) = |x_i^{(N)}(t)| + \frac{1}{N} \sum_{k=1}^N |x_k^{(N)}(t)| + \sum_{j=1}^\nu_j^{(N)} \sum_{1 \leq k \leq N, \theta_k = j} |x_k^{(N)}(t)|. \]

By Hölder’s inequality,
\[ \left[ Z_i^{(N)}(t) \right]^p \leq C \left[ x_i^{(N)}(t) \right]^p + \frac{C}{N} \sum_{k=1}^N \left[ x_k^{(N)}(t) \right]^p + \sum_{j=1}^\nu_j^{(N)} \sum_{1 \leq k \leq N, \theta_k = j} \left[ x_k^{(N)}(t) \right]^p, \]
where \( C \) depends only on \( p. \) Recall that \( \varphi(x) = |x| \). The definitions of \( \mu^{(N)}(t) \) and \( \mu_j^{(N)}(t) \) yield that
\[ \langle \mu^{(N)}(t), \varphi \rangle = \langle \mu^{(N)}(t), \varphi \rangle + \sum_{j=1}^K \langle \mu_j^{(N)}(t), \varphi \rangle, \]
\[ \sum_{j=1}^K \langle \mu_j^{(N)}(t), \varphi \rangle = \sum_{j=1}^\nu_j^{(N)} \frac{1}{\nu_j^{(N)}} \sum_{1 \leq k \leq N, \theta_k = j} |x_k^{(N)}(t)|. \]
Hence, we can write
\[ Z_i^{(N)}(t) = |x_i^{(N)}(t)| + \langle \mu^{(N)}(t), \varphi \rangle + \sum_{j=1}^K \langle \mu_j^{(N)}(t), \varphi \rangle. \]

On the other hand, equation (7) gives
\[ \left[ x_i^{(N)}(t) \right]^p \leq C \left[ x_i^{(N)}(0) \right]^p + C \left| \int_0^t b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \mu_j^{(N)}(r) \right) dr \right|^p \]
\[ + C \left| \int_0^t \sigma_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \mu_j^{(N)}(r) \right) dB_i(r) \right|^p. \]
An application of Hölder’s inequality and Burkholder-Davis-Gundy inequality, in view of (45) and inequality (46), implies that for all \( i = 1, \ldots, N, \)
\[ \mathbb{E} \sup_{0 \leq s \leq t} \left| x_i^{(N)}(s) \right|^p \]
\[
\begin{align*}
&\leq C \mathbb{E} \left| x_i^{(N)}(0) \right|^p + C t^{p-1} \int_0^t \mathbb{E} \left| b_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \mathbf{\mu}^{(N)}(r) \right) \right|^p \, dr \\
&\quad + C \mathbb{E} \left[ \int_0^t \left| \sigma_{\theta_i} \left( r, x_i^{(N)}(r), \mu^{(N)}(r), \mathbf{\mu}^{(N)}(r) \right) \right|^2 \, dr \right]^{p/2} \\
&\leq C \mathbb{E} \left| x_i^{(N)}(0) \right|^p + C t^{p-1} \int_0^t \mathbb{E} \left[ 1 + \left| x_i^{(N)}(r) \right| + \left< \mu^{(N)}(r), \varphi \right> + \sum_{j=1}^K \left< \mu_j^{(N)}(r), \varphi \right> \right]^p \, dr \\
&\quad + C \mathbb{E} \left[ \int_0^t \left( 1 + \left| x_i^{(N)}(r) \right|^2 + \left< \mu^{(N)}(r), \varphi \right>^2 + \sum_{j=1}^K \left< \mu_j^{(N)}(r), \varphi \right>^2 \right) \, dr \right]^{p/2} \\
&\leq C \left( t^{p/2} + t^p + \mathbb{E} \left| x_i^{(N)}(0) \right|^p \right) + C t^{p-1} \int_0^t \mathbb{E} \left| Z_i^{(N)}(r) \right|^p \, dr \\
&\quad + C \mathbb{E} \left[ \int_0^t \left| Z_i^{(N)}(r) \right|^2 \, dr \right]^{p/2},
\end{align*}
\]
which, together with (44), leads to
\[
\mathbb{E} \sup_{0 \leq s \leq t} \left| Z_i^{(N)}(s) \right|^p \\
\leq C \mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| x_i^{(N)}(s) \right|^p + \frac{C}{N} \sum_{k=1}^N \mathbb{E} \sup_{0 \leq s \leq t} \left| x_k^{(N)}(s) \right|^p \\
&\quad + \sum_{j=1}^K \frac{C}{N^j} \sum_{1 \leq k \leq N, \theta_k = j} \mathbb{E} \sup_{0 \leq s \leq t} \left| x_k^{(N)}(s) \right|^p \right] \\
\leq C \left( t^{p/2} + t^p + \max_{k=1, \ldots, N} \mathbb{E} \left| x_k^{(N)}(0) \right|^p \right) + C \max_{k=1, \ldots, N} t^{p-1} \int_0^t \mathbb{E} \left| Z_k^{(N)}(r) \right|^p \, dr \\
&\quad + C \max_{k=1, \ldots, N} \mathbb{E} \left[ \int_0^t \left| Z_k^{(N)}(r) \right|^2 \, dr \right]^{p/2},
\]
Since the above inequality holds true for any \( i = 1, 2, \ldots, N \), by taking the maximum with that index we arrive at
\[
\max_{k=1, \ldots, N} \mathbb{E} \sup_{0 \leq s \leq t} \left[ Z_k^{(N)}(s) \right]^p \\
\leq C \left( t^{p/2} + t^p + \sup_{i \geq 1, j \in \mathcal{K}} \mathbb{E} \left| a_{ij}^{(0)} \right|^p \right) + C T^{p-1} \max_{k=1, \ldots, N} \int_0^t \mathbb{E} \left| Z_k^{(N)}(r) \right|^p \, dr \\
&\quad + C \max_{k=1, \ldots, N} \mathbb{E} \left[ \int_0^t \left| Z_k^{(N)}(r) \right|^2 \, dr \right]^{p/2}.
\]
In view of Theorem 3.3 with the index set \( I = \{1, \ldots, N\} \), \( K_1 = CT^{p-1} \), \( K_2 = C \), and \( q = 2 \) we obtain
\[
\max_{i=1, \ldots, N} \mathbb{E} \sup_{0 \leq s \leq t} \left[ Z_i^{(N)}(s) \right]^p \\
\leq C \left( t^{p/2} + t^p + \sup_{i \geq 1, j \in \mathcal{K}} \mathbb{E} \left| a_{ij}^{(0)} \right|^p \right) e^{C(T^{p} + T^{\max(p/2,1)})}.
\]
Hence,
\[
\sup_{N > 0} \max_{t_i = 1, \ldots, N} \mathbb{E} \sup_{0 \leq s \leq t} |x_i^{(N)}(s)|^p \\
\leq \sup_{N > 0} \max_{t_i = 1, \ldots, N} \mathbb{E} \sup_{0 \leq s \leq t} |Z_i^{(N)}(s)|^p \\
\leq C \left(T^{p/2} + T^p + \sup_{i \geq 1, j \in \mathbb{K}} \mathbb{E} |x_{ij}^0|^p \right) e^{C(T^p + T^{\max(p/2, 1)})} < \infty.
\]

This completes the proof. \(\square\)

A.2. **Proof of Theorem 2.2.** We first assume that \(p \geq 2\). Put \(Y_1 = \Psi(M_1)\) and \(Y_2 = \Psi(M_2)\). Then \(\Phi(M_1) = \mathcal{L}(Y_1)\) and \(\Phi(M_2) = \mathcal{L}(Y_2)\). Thus,
\[
|W_{p,t}(\Phi(M_1), \Phi(M_2))|^p \leq \mathbb{E} \sup_{r \in [0,t]} |Y_1(r) - Y_2(r)|^p. \quad (47)
\]

On the other hand,
\[
Y_1(t) = Y^0 + \int_0^t b(s, Y_1(s), M_1(s)) \, ds + \int_0^t \sigma(s, Y_1(s), M_1(s)) \, dB(s),
\]
\[
Y_2(t) = Y^0 + \int_0^t b(s, Y_2(s), M_2(s)) \, ds + \int_0^t \sigma(s, Y_2(s), M_2(s)) \, dB(s),
\]
where \(M_i(t) = M_i \circ \phi_t^{-1} \in \mathcal{P}(\mathbb{R}^d)^K\) for \(i = 1, 2\). Hence,
\[
Y_1(t) - Y_2(t) = \int_0^t [b(s, Y_1(s), M_1(s)) - b(s, Y_2(s), M_2(s))] \, ds \\
+ \int_0^t [\sigma(s, Y_1(s), M_1(s)) - \sigma(s, Y_2(s), M_2(s))] \, dB(s). \quad (49)
\]

For each \(i = 1, 2\) and \(j = 1, \ldots, K\) let \(\mu_{ij}(t) = M_i(t) \circ \phi_t^{-1} \in \mathcal{P}(\mathbb{R}^d)\). In addition, put \(\mu_i(t) = (\mu_{i1}(t), \ldots, \mu_{iK}(t))\) and denote \(m_i(t) = \sum_{j=1}^K \nu_j \mu_{ij}(t)\). Then for all \(s \in [0, t]\),
\[
W_p(\mu_{1j}(s), \mu_{2j}(s)) \leq W_{p,s}(M_1, M_2), \quad \text{for all } j \in \mathbb{K}, \quad (50)
\]
and thus, by **Theorem 3.2**,
\[
W_p(m_1(s), m_2(s)) \leq \left(\sum_{j=1}^K \nu_j [W_p(\mu_{1j}(s), \mu_{2j}(s))]^p\right)^{1/p} \leq CW_{p,s}(M_1, M_2). \quad (51)
\]

Hence, Assumption (A2), (50), and (51) imply that for all \(s \in [0, t]\),
\[
|b(t, Y_1(s), M_1(s)) - b(s, Y_2(s), M_2(s))| + |\sigma(s, Y_1(s), M_1(s)) - \sigma(s, Y_2(s), M_2(s))| \\
\leq C \sum_{j=1}^K |b_j(s, Y_1(s), m_1(s), \mu_{1j}(s)) - b_j(s, Y_2(s), m_2(s), \mu_{2j}(s))| \\
+ C \sum_{j=1}^K |\sigma_j(s, Y_1(s), m_1(s), \mu_{1j}(s)) - \sigma_j(s, Y_2(s), m_2(s), \mu_{2j}(s))| \\
\leq C |Y_1(s) - Y_2(s)| + CW_p(m_1(s), m_2(s)) + C \sum_{j=1}^K W_p(\mu_{1j}(s), \mu_{2j}(s)) \\
\leq C |Y_1(s) - Y_2(s)| + CW_{p,s}(M_1, M_2). \quad (52)
\]
Consequently, in view of Hölder’s inequality and Burkholder-Davis-Gundy inequality, a combination of equation (49) and inequality (52) yields

\[
E \sup_{r \in [0,t]} |Y_1(r) - Y_2(r)|^p \leq E \int_0^t \left| b(s, Y_1(s), M_1(s)) - b(s, Y_2(s), M_2(s)) \right|^p ds + E \int_0^t \left| \sigma(s, Y_1(s), M_1(s)) - \sigma(s, Y_1(s), M_1(s)) \right|^2 ds \]

\[
\leq C E \int_0^t |Y_1(r) - Y_2(r)|^p ds + C E \int_0^t |\mathcal{W}_{p,s}(M_1, M_2)|^p ds,
\]

which, in virtue of Gronwall’s inequality, leads to

\[
E \sup_{r \in [0,t]} |Y_1(r) - Y_2(r)|^p \leq C E \int_0^t |\mathcal{W}_{p,s}(M_1, M_2)|^p.
\]  

(53)

As a result, inequality (11) can be verified by using inequalities (47) and (53).

Next, if $1 \leq p < 2$ then it follows from inequality (47) and Jensen’s inequality that

\[
[\mathcal{W}_{p,t}(\Phi(M_1), \Phi(M_2))]^2 \leq \left( E \sup_{r \in [0,t]} |Y_1(r) - Y_2(r)|^p \right)^{2/p} \leq E \sup_{r \in [0,t]} |Y_1(r) - Y_2(r)|^2.
\]

A similar argument as in the previous step yields the desired estimate.

\[\square\]

### A.3. Proof of Proposition 2.

(i) We will first verify inequality (13). For all $t > 0$, put

\[
Z(t) = \sum_{k=1}^K |y_k(t)| + \sum_{k=1}^K E |y_k(t)|.
\]

Then we have

\[
\langle \mu(t), \varphi \rangle = \sum_{k=1}^K \nu_k \langle \mu_k(t), \varphi \rangle = \sum_{k=1}^K \nu_k \langle y_k(t), \varphi \rangle = \sum_{k=1}^K \nu_k E |y_k(t)| \leq CZ(t),
\]

(54)

and

\[
\sum_{k=1}^K \langle \mu_k(r), \varphi \rangle = \sum_{k=1}^K E |y_k(t)| \leq Z(t).
\]

(55)

On the other hand, equation (7) gives

\[
|y_j(t)|^p \leq C |y_j(0)|^p + C \left| \int_0^t b_j(r, y_j(r), \mu(r), \underline{\mu}(r)) dr \right|^p + C \left| \int_0^t \sigma_j(r, y_j(r), \mu(r), \underline{\mu}(r)) d\tilde{B}_j(r) \right|^p.
\]  

(56)

In view of inequalities (54) to (56), an application of Hölder’s inequality and Burkholder-Davis-Gundy inequality implies that for all $j \in \mathbb{K}$

\[
E \sup_{0 \leq s \leq t} |y_j(s)|^p \leq C E |y_j(0)|^p + Ct^{p-1} \int_0^t E |b_j(r, y_j(r), \mu(r), \underline{\mu}(r))|^p dr
\]

\[
+ E \left[ \int_0^t |\sigma_j(r, y_j(r), \mu(r), \underline{\mu}(r))|^2 dr \right]^{p/2}.
\]
Now we will prove inequality (14). Applying Cauchy's inequality, Hölder's inequality, and Burkholder-Davis-Gundy inequality on the following equation

\[ I \quad \text{Applying Theorem 3.3 with the index set } \mathcal{I} \text{ leads to} \]

\[
\mathbb{E} \sup_{0 \leq s \leq t} [Z(s)]^p \leq C \sum_{k=1}^{K} \mathbb{E} \sup_{0 \leq s \leq t} |y_k(s)|^p
\]

\[
\leq C \left( t^{p/2} + t^p + \max_{j \in \mathcal{K}} \mathbb{E} |y_j(0)|^p \right)
\]

\[
+ C t^{p-1} \int_0^t \mathbb{E} |Z(r)|^p \, dr + C \mathbb{E} \left[ \int_0^t |Z(r)|^2 \, dr \right]^{p/2},
\]

which leads to

\[
\mathbb{E} \sup_{0 \leq s \leq t} [Z(s)]^p \leq C \sum_{k=1}^{K} \mathbb{E} \sup_{0 \leq s \leq t} |y_k(s)|^p
\]

\[
\leq C \left( t^{p/2} + t^p + \max_{j \in \mathcal{K}} \mathbb{E} |y_j(0)|^p \right)
\]

\[
+ C t^{p-1} \int_0^t \mathbb{E} |Z(r)|^p \, dr + C \mathbb{E} \left[ \int_0^t |Z(r)|^2 \, dr \right]^{p/2}.
\]

Applying Theorem 3.3 with the index set \( \mathcal{I} \) being a singleton, we obtain

\[
\mathbb{E} \sup_{0 \leq t \leq T} [Z(t)]^p \leq C \left( T^{p/2} + T^p + \max_{j \in \mathcal{K}} \mathbb{E} |y_j(0)|^p \right) e^{C(T^p + T^{\max\{p/2,1\}})},
\]

from which inequality (13) follows.

(ii) Now we will prove inequality (14). Applying Cauchy's inequality, Hölder's inequality, and Burkholder-Davis-Gundy inequality on the following equation

\[
y_j(t + s) - y_j(t) = \int_t^{t+s} b_j (r, y_j(r), \mu(r), \underline{\mu}(r)) \, dr
\]

\[
+ \int_t^{t+s} \sigma_j (r, y_j(r), \mu(r), \underline{\mu}(r)) \, dB_j(r),
\]

we obtain

\[
\mathbb{E} \left[ \sup_{s_1 \in [0,s]} (|y_j(t + s_1) - y_j(t)|^p) \right]
\]

\[
\leq C \mathbb{E} \left\{ \sup_{s_1 \in [0,s]} \left| \int_t^{t+s_1} b_j (r, y_j(r), \mu(r), \underline{\mu}(r)) \, dr \right|^p \right\}
\]

\[
+ C \mathbb{E} \left\{ \sup_{s_1 \in [0,s]} \left| \int_t^{t+s_1} \sigma_j (r, y_j(r), \mu(r), \underline{\mu}(r)) \, dB_j(r) \right|^p \right\}
\]

\[
\leq Cs^{p-1} \int_t^{t+s} |b_j(r, y_j(r), \mu(r), \underline{\mu}(r))|^p \, dr
\]

\[
+ C \mathbb{E} \left[ \left( \int_t^{t+s} |\sigma_j (r, y_j(r), \mu(r), \underline{\mu}(r))|^2 \, dr \right)^{p/2} \right].
\]
By Cauchy’s inequality, Assumption (A1), and (13),
\[ E \left| b_j(r, y_j(r), \mu(r), \mu(r)) \right|^p \leq C E \left[ 1 + |y_j(r)| + \sum_{j=1}^{K} \nu_j E |y_j(r)| + \sum_{j=1}^{K} E |y_j(r)| \right]^p \]
\[ \leq C \left[ 1 + E \sum_{j=1}^{K} |y_j(r)|^p \right] \]
\[ \leq C \left[ 1 + E \sum_{j=1}^{K} \nu_j |y_j(r)| + K \sum_{j=1}^{K} E |y_j(r)| \right] \]
\[ \leq C \left[ 1 + \left( T^{p/2} + T^p + E |y_j(0)|^p \right) e^{CT^p} \right]. \]
(58)

Similarly, we can also use inequality (13) to obtain an upper bound related to \( \sigma_j(r, y_j(r), \mu(r), \mu(r)) \) as follows
\[ E \left[ \int_t^{t+s} \left| \sigma_j(r, y_j(r), \mu(r), \mu(r)) \right|^2 dr \right]^{p/2} \]
\[ \leq E \left[ \sup_{r \in [t,t+s]} \left| \sigma_j(r, y_j(r), \mu(r), \mu(r)) \right|^p \right] \]
\[ \leq C s^{p/2} E \left[ 1 + \sum_{j=1}^{K} \sup_{r \in [t,t+s]} |y_j(r)|^p \right] \]
\[ \leq C s^{p/2} \left[ 1 + \left( T^{p/2} + T^p + E |y_j(0)|^p \right) e^{CT^p} \right]. \]
(59)

It follows from inequalities (57) to (59) that inequality (14) holds.

\[ \square \]

A.4. Proof of Theorem 3.2. Put \( \vartheta = \sum_{j=1}^{K} \lambda_j \varpi_j \). It is clear that if \( \gamma_j \in \Gamma(\eta_j, \varpi_j) \) for each \( 1 \leq j \leq K \) and \( \gamma = \sum_{j=1}^{K} \lambda_j \gamma_j \), then \( \gamma \in \Gamma(\eta, \vartheta) \). Moreover,
\[ [W_p(\eta, \vartheta)]^p = \inf_{\gamma \in \Gamma(\eta, \vartheta)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\gamma(x, y) \]
\[ \leq \sum_{j=1}^{K} \lambda_j \inf_{\gamma_j \in \Gamma(\eta_j, \varpi_j)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\gamma_j(x, y) \]
\[ \leq \sum_{j=1}^{K} \lambda_j [W_p(\eta_j, \varpi_j)]^p. \]
(60)

On the other hand, it is known that Monge-Wasserstein distance is controlled by weighted total variation (see [30, Theorem 6.15]). Therefore,
\[ [W_p(\vartheta, \varpi)]^p \leq C \int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^p d\vartheta - \varpi(x) \]
\[ \leq C \sum_{j=1}^{K} \left| \lambda_j - \pi_j \right| \int_{\mathbb{R}^d} |x|^p d\varpi_j(x) \]
\[ \leq C \sum_{j=1}^{K} \left| \lambda_j - \pi_j \right| M_p(\varpi_j). \]
(61)
Inequalities (60) and (61), in view of H"older’s inequality, conclude the argument. □

A.5. **Proof of Theorem 3.3.** Let $Z(t), t \geq 0$, be a nonnegative stochastic process and $\rho > 0$. Denote $V(t) = \sup_{0 \leq s \leq t} Z(s)$. We claim that: For any $C_1 > 0$, there exists a constant $C_2 > 0$ such that

$$
\left( \int_0^t Z(s) ds \right)^\rho \leq C_1 [V(t)]^\rho + C_2 \int_0^t [Z(s)]^\rho ds,
$$

(62)

for all $t \in [0, T]$. In fact, we can choose

$$C_2 = \begin{cases} T^{\rho-1} & \text{if } \rho \geq 1, \\ \rho \left( \frac{1-\rho}{\rho - 1} \right)^{\frac{1}{\rho - 1}} & \text{if } 0 < \rho < 1. \end{cases}$$

Indeed, if $\rho \geq 1$ then H"older’s inequality implies that

$$
\left( \int_0^t Z(s) ds \right)^\rho \leq t^{\rho-1} \int_0^t [Z(s)]^\rho ds.
$$

If $0 < \rho < 1$, an application of Young’s inequality gives

$$
\left( \int_0^t Z(s) ds \right)^\rho \leq [V(t)]^{\rho(1-\rho)} \left( \int_0^t [Z(s)]^{\rho'} ds \right)^\rho

\leq C_1 [V(t)]^\rho + C_2 \int_0^t [Z(s)]^\rho ds,
$$

where $C_2 = \rho \left( \frac{1-\rho}{\rho - 1} \right)^{\frac{1}{\rho - 1}}$. Thus, inequality (62) holds for all $\rho > 0$ and $t \in [0, T]$.

Using inequality (62) with $Z_i^t$ in place of $Z$, $p/q$ in place of $\rho$, and $C_1 = \frac{1}{2K_2}$ gives

$$
\left( \int_0^t [Z_i(s)]^q ds \right)^{p/q} \leq \frac{1}{2K_2} [V_i(t)]^p + C_2 \int_0^t [Z_i(s)]^p ds

\leq \frac{1}{2K_2} [V_i(t)]^p + C_2 \int_0^t [V_i(s)]^p ds,
$$

(63)

with $C_2 = T^{p/q-1}$ when $p \geq q$ and $C_2 = \frac{p}{q} [2K_2(1-\frac{p}{q})]^{\frac{q-p}{q}}$ when $p < q$. As a result, inequalities (22) and (63) show that

$$
\sup_{i \in I} \mathbb{E}[V_i(t)]^p \leq K_1 \sup_{i \in I} \mathbb{E} \int_0^t [V_i(s)]^p ds + K_2 \sup_{i \in I} \mathbb{E} \left[ \int_0^t [Z_i(s)]^q ds \right]^{p/q} + \delta

\leq K_1 \sup_{i \in I} \mathbb{E} \int_0^t [V_i(s)]^p ds + \frac{1}{2} \sup_{i \in I} \mathbb{E}[V_i(t)]^p

+ K_2 C_2 \sup_{i \in I} \mathbb{E} \int_0^t [V_i(s)]^p ds + \delta.
$$

Here, $K_2 C_2 = K_2 T^{p/q-1}$ when $p \geq q$ and $K_2 C_2 = \frac{p}{q} [2(1-\frac{p}{q})]^{\frac{q-p}{q}} K_2^{p/q}$ when $p < q$. Thus, for any $0 \leq t \leq T$,

$$
\sup_{i \in I} \mathbb{E}[V_i(t)]^p \leq 2(K_1 + K_2 C_2) \int_0^t \mathbb{E}[V_i(s)]^p ds + 2\delta,
$$

which, in view of Gronwall’s inequality, leads to

$$
\sup_{i \in I} \mathbb{E}[V_i(T)]^p \leq 2\delta \exp \left( \int_0^T 2(K_1 + K_2 C_2) ds \right) \leq 2\delta \exp \left( 2K_1 T + CK_2 T^{\max(p/q,1)} \right),
$$
where constant $C$ depends only on $p$ and $q$. In particular, if $q = 2$ and $p \geq 1$ then $C < 8$, and therefore,

$$
\sup_{i \in I} E[V_i(t)]^p \leq 2\delta \exp \left(2K_1 T + 8K_2 T^{\max\{p/2,1\}}\right).
$$

This completes the proof.

**Acknowledgments.** The first author’s research was funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2019-20-01. The second author’s research was supported by a seed fund of Department of Mathematics at University of Puerto Rico, Rio Piedras campus. The third author’s research was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.03.2017.308.

**REFERENCES**

[1] L. Andreis, P. D. Pra and M. Fischer, McKean-Vlasov limit for interacting systems with simultaneous jumps, *Stochastic Analysis and Applications*, 36 (2018), 960–995.

[2] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul, Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons, *The Journal of Mathematical Neuroscience*, 2 (2012), Art. 10, 50 pp.

[3] A. Bensoussan, J. Frehse and P. Yam, *Mean Field Games and Mean Field Type Control Theory*, Springer Briefs in Mathematics, Springer New York, 2013.

[4] F. Bolley, Separability and completeness for the Wasserstein distance, *Séminaire de Probabilités XLI* (eds. C. Donati-Martin, M. Émery, A. Rouault, and C. Stricker), Berlin, Heidelberg: Springer Berlin Heidelberg, 1934 (2008), 371–377.

[5] V. S. Borkar and K. S. Kumar, McKean-Vlasov limit in portfolio optimization, *Stochastic Analysis and Applications*, 28 (2010), 884–906.

[6] A. Budhiraja, P. Dupuis and M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, *Ann. Probab.*, 40 (2012), 74–102.

[7] A. Budhiraja and R. Wu, Some fluctuation results for weakly interacting multi-type particle systems, *Stochastic Processes and their Applications*, 126 (2016), 2253–2296.

[8] R. Carmona, F. Delarue and D. Lacker, Mean field games with common noise, *Ann. Probab.*, 44 (2016), 3740–3803.

[9] F. Collet, Macroscopic limit of a bipartite Curie-Weiss model: A dynamical approach, *Journal of Statistical Physics*, 157 (2014), 1301–1319.

[10] P. Contucci, I. Gallo and G. Menconi, Phase transitions in social sciences: Two-populations mean field theory, *International Journal of Modern Physics B*, 22 (2008), 2199–2212.

[11] D. A. Dawson, Critical dynamics and fluctuations for a mean-field model of cooperative behavior, *Journal of Statistical Physics*, 31 (1983), 29–85.

[12] D. A. Dawson and J. Vaillancourt, Stochastic McKean-Vlasov equations, *Nonlinear Differential Equations and Applications NoDEA*, 2 (1995), 199–229.

[13] D. A. Dawson and J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, *Stochastics*, 20 (1987), 247–308.

[14] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, *Probability Theory and Related Fields*, 162 (2015), 707–738.

[15] T. A. Hoang, S. L. Nguyen and G. Yin, Near optimality and near equilibrium for controlled systems with wideband noise for hybrid systems, *Dynamics of Continuous Discrete and Impulsive Systems, Series A: Mathematical Analysis*, 23 (2016), 163–194.

[16] T. A. Hoang and G. Yin, Properties for a class of multi-type mean-field models, *Communications in Information and Systems*, 15 (2015), 489–519.

[17] M. Huang, Large-population LQG games involving a major player: The Nash certainty equivalence principle, *SIAM Journal on Control and Optimization*, 48 (2010), 3318–3353.

[18] M. Huang, P. E. Caines and R. P. Malhamé, Social optima in mean field LQG control: Centralized and decentralized strategies, *IEEE Trans. Automat. Control*, 57 (2012), 1736–1751.
[19] M. Huang, R. P. Malhamé and P. E. Caines, Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, *Communications in Information & Systems*, 6 (2006), 221–252.

[20] V. Kolokoltsov and M. Troeva, On the mean field games with common noise and the McKean-Vlasov SPDEs, *Stochastic Analysis and Applications*, 37 (2019), 522–549.

[21] T. G. Kurtz and J. Xiong, Particle representations for a class of nonlinear SPDEs, *Stochastic Processes and their Applications*, 83 (1999), 103–126.

[22] J. M. Lasry and P. L. Lions, Jeux à champ moyen. I - Le cas stationnaire, *Comptes Rendus Mathematique*, 343 (2006), 619–625.

[23] S. Méléard, Convergence of the fluctuations for interacting diffusions with jumps associated with boltzmann equations, *Stochastics and Stochastic Reports*, 63 (1998), 195–225.

[24] S. L. Nguyen and M. Huang, Linear-quadratic-Gaussian mixed games with continuum-parametrized minor players, *SIAM Journal on Control and Optimization*, 50 (2012), 2907–2937.

[25] S. L. Nguyen, G. Yin and T. A. Hoang, On laws of large numbers for systems with mean-field interactions and Markovian switching, *Stochastic Processes and their Applications*, In press (2019), Available from: https://doi.org/10.1016/j.spa.2019.02.014.

[26] M. Nourian and P. E. Caines, $\epsilon$-Nash mean field game theory for nonlinear stochastic dynamical systems with major and minor agents, *SIAM Journal on Control and Optimization*, 51 (2013), 3302–3331.

[27] K. Oelschlager, A martingale approach to the law of large numbers for weakly interacting stochastic processes, *The Annals of Probability*, 12 (1984), 458–479.

[28] T. Shiga and H. Tanaka, Central limit theorem for a system of Markovian particles with mean field interactions, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69 (1985), 439–459.

[29] A. S. Sznitman, *Topics in propagation of chaos*, *Ecole d’Eté de Probabilités de Saint-Flour XIX - 1989*, Ed. by P. L. Hennequin. Berlin, Heidelberg: Springer Berlin Heidelberg, 1464 (1991), 165–251.

[30] C. Villani, *Optimal transport: Old and new*, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Berlin, 2009.

[31] F. Xi and G. Yin, Asymptotic properties of a mean-field model with a continuous-state-dependent switching process, *Journal of Applied Probability*, 46 (2009), 221–243.

Received May 2018; revised August 2019.

E-mail address: dungnt@hcmut.edu.vn
E-mail address: sonluu.nguyen@upr.edu
E-mail address: dunh@vnu.edu.vn