

SWITCHING DIFFUSIONS WITH MEAN-FIELD INTERACTIONS: LIMIT RESULTS, MAXIMUM PRINCIPLE, AND NON-MARKOV SYSTEMS

GEORGE YIN

*Department of Mathematics, University of Connecticut
Storrs, CT 06269
ORCID: 0000-0002-2951-0704 E-mail: gyin@uconn.edu*

SON LUU NGUYEN

*University of Puerto Rico, Rio Piedras campus
San Juan, PR 00936, USA
ORCID: 0000-0002-6082-6686 E-mail: sonluu.nguyen@upr.edu*

Abstract. This paper is devoted to switching diffusions with mean-field interactions. We first review some of the recent results. Then we examine a case of systems not driven by Brownian motion but stationary mixing processes. We obtain the limit of the systems by weak convergence analysis together with our limit results on a law of large numbers for switching diffusion processes with mean-field terms.

1. Introduction. In this work, we study switching diffusions with mean-field interactions. The motivation stems from two lines of work. One of them is to treat hybrid diffusions that include both continuous dynamics and discrete events, in which the discrete events are modeled as a continuous-time Markov chain taking values in a finite state space $\mathcal{M} = \{1, \dots, m_0\}$. The other line is the consideration of systems with mean-field interactions. Both lines are of considerable current interests. What we are looking here is at the intersection of the two lines.

Before proceeding further, let us briefly recall some of the works in each of these

2010 *Mathematics Subject Classification*: 60J25; 60J27; 60J60; 93E20.

Key words and phrases: mean-field model, Markovian switching diffusion, law of large number, McKean–Vlasov equation, maximum principle, non-Markov model, weak convergence.

The paper is in final form and no version of it will be published elsewhere.

areas. The study of switching diffusion stems from the need of treating hybrid systems. Typically, in such systems continuous dynamics and discrete events are intertwined. The system is running in continuous time. The continuous dynamics are as given by solutions of differential equations or stochastic differential equations. The discrete events, however, take values in a finite set. In this paper, we are mainly concerned ourselves with systems that may be called Markovian switching diffusions. A representative of such systems is the book of Mao and Yuan [MY]. It is called a Markovian switching diffusion because the Markov chain modulating the system is independent of the Brownian motion in the continuous part of the dynamic system. A related work is [YZ2], in which the switching process itself is not necessarily Markov but depending on the continuous states. One of the reasons that the hybrid diffusions have drawn much attention is the pressing needs in such emerging applications as social networks, cyber-physical systems, and networked control systems; see for example, [HY, TY, YI, YZ1, ZY1] among others.

Next consider systems with mean-field interactions, which has a long history. In statistical physics, it is known that many body problems are difficult to deal with. It is suggested that one studies the behavior of high-dimensional random (stochastic) models by a simpler model that approximates the original systems using averaging over the many bodies. The rationale is to replace all interactions to any one body with an average or effective interactions. The intuition has been around for a long time, but it was until 1983, Dawson settled the matter in [D] by providing a mathematically rigorous proof. He showed that a law of large numbers holds for an empirical measure associated with a diffusion process together with a phase transition property. This work promoted much subsequent study; see for example, [DG, DV, DZ, G] and references therein. This line of work has been much extended, for example, in [BF, CM, O, S]. Consideration of mean-field models associated with control, optimization, and game problems was started in the 2000s by Huang, Caines, and Malhamé [HM, HC] and independently by Lasry and Lions [LL, Lb]. The book by Bensoussan, Frehse, and Yam [BY] provided an illuminating presentation and discussion of certain aspects of mean-field games and mean-field type controls, describing their similarities and differences together with a unified approach for treating them. A more analytic approach can be found in the book by Kolokoltsov [K], where the author named the processes under consideration nonlinear Markov processes.

Considering the intersections of the switching diffusions and mean-field interactions, in this paper, our objectives include two parts. In the first part, we provide a survey of some recent limit results as the number of players getting large; we also deal with controlled switching diffusions with mean-field terms. After reviewing these properties, in the second part, we devote our attention to a case of stochastic systems involving mean-field interactions. It is similar to the first part, but the systems are non-Markov. The rationale is that very often the Markovian systems are only a idealization. In applications, very often we need to deal with non-Markov systems. We consider the case that an ordinary differential equation involving mean-field terms subject to wide-band noise perturbations. The wide-band noise is scaled by a small parameter ε . As $\varepsilon \rightarrow 0$, the wide-band noise tends to a limit that is a Brownian motion. For this case, we carry out the analysis and obtain the limit process. Then we show that there is also a law of large numbers in this case. The desired McKean–Vlasov equation is obtained.

The rest of the paper is arranged as follows. Section 2 presents the formulation and some of our recent results including a law of large numbers and a maximum principle for the controlled systems under consideration. Section 3 is devoted to obtaining the weak convergence and the limit McKean–Vlasov equation for non-Markovian systems. Finally, Section 4 concludes the paper with some further remarks.

2. Formulation and some recent results. Consider the following system of switching diffusions:

$$dx_i(t) = b\left(x_i(t), \frac{1}{N} \sum_{j=1}^N \delta_{x_j(t)}, \alpha(t_-)\right) dt + \sigma\left(x_i(t), \frac{1}{N} \sum_{j=1}^N \delta_{x_j(t)}, \alpha(t_-)\right) dw_i(t), \quad (1)$$

for $i = 1, 2, \dots, N$, where $\delta_x(\cdot)$ denotes the Dirac measure centered at x with $x \in \mathbb{R}^d$, $w_1(\cdot), w_2(\cdot), \dots, w_N(\cdot)$ are N independent d -dimensional standard Brownian motions, and $\alpha(\cdot)$ is a Markov chain taking values in a finite state space $\mathcal{M} = \{1, 2, \dots, m_0\}$ with a generator $Q = (q_{i_0 j_0})_{i_0, j_0 \in \mathcal{M}}$ satisfying the following properties: $q_{i_0 j_0} \geq 0$ for $i_0 \neq j_0 \in \mathcal{M}$ and $\sum_{j_0 \in \mathcal{M}} q_{i_0 j_0} = 0$ for each $i_0 \in \mathcal{M}$. We shall use the following notation.

- $C_b(\mathbb{R}^d)$: space of bounded continuous functions on \mathbb{R}^d
- $C_b^k(\mathbb{R}^d)$: C^k functions with bounded partial derivatives
- C_c^k : C^k functions with compact support
- E : a metric space
- $\mathcal{B}(E)$: Borel σ -field on E
- $\mathcal{P}(E)$: space of probability measures on $(E, \mathcal{B}(E))$ with weak topology
- $C([0, T], E)$: space of continuous functions with sup metric
- $D([0, T], E)$: space of all càdlàg functions with Skorohod topology
- $\mathcal{M} := \{1, 2, \dots, m_0\}$: state space of the Markov chain
- $d_{\mathcal{M}}$: metric on \mathcal{M} , $d_{\mathcal{M}}(i_0, j_0) = 1 - \delta_{i_0, j_0}$ for $i_0, j_0 \in \mathcal{M}$.
- $D_f([0, T], \mathcal{M})$: subspace of $D([0, T], \mathcal{M})$ with finite jumps
- \mathcal{M}_1 : space of probability measures on \mathbb{R}^d
- $\mu_N(t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i(t)}$
- $\langle \mu, f \rangle = \int_{\mathbb{R}^d} f(x) \mu(dx)$: for $\mu \in \mathcal{M}_1$
- $M_{i_0 j_0}(t) = [M_{i_0 j_0}](t) - \langle M_{i_0 j_0} \rangle(t)$: martingale associate with the Markov chain, where

$$[M_{i_0 j_0}](t) = \sum_{0 \leq s \leq t} \mathbb{1}(\alpha(s_-) = i_0) \mathbb{1}(\alpha(s) = j_0),$$

$$\langle M_{i_0 j_0} \rangle(t) = \int_0^t q_{i_0 j_0} \mathbb{1}(\alpha(s_-) = i_0) ds$$

if $i_0 \neq j_0 \in \mathcal{M}$ and $M_{i_0 i_0}(t) \equiv 0$

- For $\mu \in \mathcal{M}_1$, $f(\cdot, \cdot, i_0) \in C_b(\mathbb{R} \times \mathbb{R}^d)$, $g(\cdot, i_0) \in C_b(\mathbb{R}^d)$,

$$\langle \mu, f(t, \cdot, i_0) \rangle = \int_{\mathbb{R}^d} f(t, x, i_0) \mu(dx), \quad \langle \mu, g(\cdot, i_0) \rangle = \int_{\mathbb{R}^d} g(x, i_0) \mu(dx)$$

- $\|\cdot\|_{TV}$: total variation metric on \mathcal{M}_1
- $\|\mu - \eta\|_{BL} = \sup \left\{ |\langle \mu, f \rangle - \langle \eta, f \rangle| : \|f\| \leq 1, \sup_{x \neq y \in \mathbb{R}^d} \frac{|f(x) - f(y)|}{|x - y|} \leq 1 \right\}$

- $(\mathcal{M}_1, \|\cdot\|_{BL})$ is a complete and separable metric space
- $d((\mu, i_0), (\eta, j_0)) = \|\mu - \eta\|_{BL} + d_{\mathcal{M}}(i_0, j_0), \forall \mu, \eta \in \mathcal{M}_1, i_0, j_0 \in \mathcal{M}$
- $\varphi_1(x) = |x|$ and $\varphi_2(x) = |x|^2$
- $\mathcal{F}_{t-}^\alpha = \sigma\{\alpha(s) : 0 \leq s < t\}$
- $\mathcal{F}_t^{N,\alpha} = \sigma\{w_i(s), \alpha(s) : 0 \leq s \leq t, 1 \leq i \leq N\}$
- For a random variable ς on $(\Omega, \mathcal{F}, \mathbb{P})$,
 - $\mathcal{L}(\varsigma)$: distribution (law of ς)
 - $\eta_t = \mathcal{L}(\varsigma | \mathcal{F}_{t-}^\alpha)$: conditional law given \mathcal{F}_{t-}^α
 $(\mathbb{E}(f(\varsigma) | \mathcal{F}_{t-}^\alpha) = \int_{\mathbb{R}^d} f(x) \eta_t(dx), \forall f \in C_b(\mathbb{R}^d))$
- \mathcal{P}_N : the induced probability measure of $(\mu_N(\cdot), \alpha(\cdot))$ on $D([0, T], \mathcal{M}_1 \times \mathcal{M})$
 [Note \mathcal{P}_N concentrates on the set $C([0, T], \mathcal{M}_1) \times D_f([0, T], \mathcal{M})$, a closed subspace of $D([0, T], \mathcal{M}_1 \times \mathcal{M})$.]

We will use the following assumptions.

(H1) For each $i_0 \in \mathcal{M}$, $b(\cdot, \cdot, i_0) : \mathbb{R}^d \times \mathcal{M}_1 \rightarrow \mathbb{R}^d$ and $\sigma(\cdot, \cdot, i_0) : \mathbb{R}^d \times \mathcal{M}_1 \rightarrow \mathbb{R}^{d \times d}$ are Lipschitz continuous in that, there is a constant L such that

$$|b(x, \mu, i_0) - b(y, \eta, i_0)| + |\sigma(x, \mu, i_0) - \sigma(y, \eta, i_0)| \leq L(|x - y| + \|\mu - \eta\|_{BL}),$$

for all $x, y \in \mathbb{R}^d$ and $\mu, \eta \in \mathcal{M}_1$.

(H2) The \mathbb{R}^d -valued function $b(\cdot, \cdot, \cdot)$ satisfies

$$|b(x, \mu, i_0)| \leq C(1 + |x| + \langle \mu, \varphi_1 \rangle), \quad (x, \mu, i_0) \in \mathbb{R}^d \times \mathcal{M}_1 \times \mathcal{M},$$

for some constant C , and the matrix-valued function $\sigma(\cdot, \cdot, \cdot)$ is bounded.

For $f(\cdot, i_0) \in C_c^2(\mathbb{R}^d)$ and $(x, \mu, i_0) \in \mathbb{R}^d \times \mathcal{M}_1 \times \mathcal{M}$ define the operator

$$\begin{aligned} \mathcal{L}(\mu)f(x, i_0) &= b'(x, \mu, i_0)\nabla_x f(x, i_0) + \frac{1}{2}(a(x, \mu, i_0)\nabla_x)' \nabla_x f(x, i_0) \\ &\quad + \sum_{j_0 \in \mathcal{M}} q_{i_0 j_0}(f(x, j_0) - f(x, i_0)), \end{aligned} \quad (2)$$

where

$$a(x, \mu, i_0) = \sigma(x, \mu, i_0)\sigma'(x, \mu, i_0) \in \mathbb{R}^{d \times d}.$$

The following theorem provides a law of large numbers for the weakly interacting systems under consideration.

THEOREM 2.1. *Assume (H1), (H2), and*

$$\sup_{N \in \mathbb{N}} \mathbb{E} \langle \mu_N(0), \varphi_2 \rangle < \infty, \quad \mathcal{L}(\mu_N(0)) \Rightarrow \delta_{\mu_0} \text{ in } \mathcal{P}(\mathcal{M}_1, \|\cdot\|_{BL}).$$

Then $(\mu_N(\cdot), \alpha(\cdot))$ converges weakly to a process $(\mu_\alpha(\cdot), \alpha(\cdot))$, where

$$(\mu_\alpha(t), \alpha(t)) = (\mathcal{L}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t)), \quad 0 \leq t \leq T,$$

and $y(t), 0 \leq t \leq T$, is the unique solution of the following stochastic differential equation

$$\begin{cases} dy(t) = b(y(t), \mathcal{L}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dt + \sigma(y(t), \mathcal{L}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dw(t), \\ \mathcal{L}(y(0)) = \mu_0, \end{cases}$$

where $w(\cdot)$ is a standard Brownian motion independent of $\alpha(\cdot)$.

The limit of $\mu_N(\cdot)$ can also be characterized as a solution of a stochastic McKean-Vlasov equation with Markovian switching as follows.

THEOREM 2.2. *Assume (H1) and (H2). Then the system of integral equations*

$$\begin{aligned} \langle \mu(t), f(\cdot, \alpha(t)) \rangle &= \langle \mu_0, f(\cdot, \alpha(0)) \rangle + \int_0^t \langle \mu(s), \mathcal{L}(\mu(s))f(\cdot, \alpha(s_-)) \rangle ds \\ &\quad + \sum_{i_0, j_0 \in \mathbb{S}} \int_0^t \langle \mu(s), f(\cdot, j_0) - f(\cdot, i_0) \rangle dM_{i_0 j_0}(s), \end{aligned} \quad (3)$$

where $0 \leq t \leq T$ and $f(\cdot, i_0) \in C_c^2(\mathbb{R}^d)$ for each $i_0 \in \mathcal{M}$, has a unique solution in $D([0, T], \mathcal{M}_1)$. Moreover, this solution equals $\mathcal{L}(y(t) \mid \mathcal{F}_{t-}^\alpha)$ for all $0 \leq t \leq T$, where $y(t)$ is the unique solution of

$$\begin{cases} dy(t) = b(y(t), \mu_\alpha(t), \alpha(t_-)) dt + \sigma(y(t), \mu_\alpha(t), \alpha(t_-)) dw(t), & \mathcal{L}(y(0)) = \mu_0, \\ \mu_\alpha(t) = \mathcal{L}(y(t) \mid \mathcal{F}_{t-}^\alpha), \end{cases}$$

where $w(\cdot)$ is a standard Brownian motion independent of $\alpha(\cdot)$.

The above theorems were obtained in [NH]. We merely cite the results here and omit the verbatim proofs. To proceed, we consider an associated controlled dynamic system. For simplicity, in what follows we will take $d = 1$.

Let U , the set where the controls take values in, be a nonempty and convex subset of \mathbb{R} , and \mathcal{U} be the class of measurable, \mathcal{F}_t -adapted, and square integrable processes $u(\cdot, \cdot) : [0, T] \times \Omega \rightarrow U$. We call \mathcal{U} the set of admissible controls. For each $u(\cdot) \in \mathcal{U}$, we consider the following controlled stochastic differential equation

$$\begin{aligned} dx(t) &= b(t, x(t), \mathbb{E}(\psi(x(t)) \mid \mathcal{F}_{t-}^\alpha), u(t), \alpha(t_-)) dt \\ &\quad + \sigma(t, x(t), \mathbb{E}(\varphi(x(t)) \mid \mathcal{F}_{t-}^\alpha), u(t), \alpha(t_-)) dw(t), \\ x(0) &= x_0, \end{aligned} \quad (4)$$

where x_0 is a real number, and $b(\cdot, \cdot, \cdot, \cdot, \cdot), \sigma(\cdot, \cdot, \cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^2 \times U \times \mathcal{M} \rightarrow \mathbb{R}$ and $\psi(\cdot), \varphi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ are real functions.

For simplicity of notation, we keep using the letters b and σ for the drift and diffusion coefficients of the controlled problem in the remaining part of this section. Note that these functions are just a special case of drift and diffusion coefficients in (1). For each $(t, x, y, u, i_0) \in [0, T] \times \mathbb{R}^2 \times U \times \mathcal{M}$, let $b_x(t, x, y, u, i_0)$, $b_y(t, x, y, u, i_0)$, and $b_u(t, x, y, u, i_0)$ respectively denote the partial derivative of b with respect to the variable x, y and u at the point (t, x, y, u, i_0) . The partial derivatives of $\sigma(\cdot, \cdot, \cdot, \cdot, \cdot)$ are defined by a similar way. With a little abuse of notation, we will use the subscript t to denote the value of a function at t instead of its derivative. For example, $x_t = x(t)$ and $u_t = u(t)$.

The mean-field SDE (4) is obtained as the mean-square limit as $N \rightarrow \infty$ of a system

of interacting particles of the form

$$\begin{aligned}
 dx^{i,N}(t) = & b\left(t, x^{i,N}(t), \frac{1}{N} \sum_{i=1}^N \psi(x^{i,N}(t)), u(t), \alpha(t-)\right) dt \\
 & + \sigma\left(t, x^{i,N}(t), \frac{1}{n} \sum_{i=1}^N \varphi(x^{i,N}(t)), u(t), \alpha(t-)\right) dw^i(t),
 \end{aligned}$$

where $(w^i(t), i \geq 1)$ is a collection of independent standard Brownian motions. Note that for more generality, we consider the mean-field term as nonlinear functions of the state with the use of $\psi(\cdot)$ and $\varphi(\cdot)$, respectively. Moreover, in (4), the conditional expectations $\mathbb{E}(\psi(x(t)) | \mathcal{F}_{t-}^\alpha)$ and $\mathbb{E}(\varphi(x(t)) | \mathcal{F}_{t-}^\alpha)$ appear instead of the expectations $\mathbb{E}(\psi(x(t)))$ and $\mathbb{E}(\varphi(x(t)))$ because of the effect of the common switching process $\alpha(t)$. Because all the particles depend on the history of this process, their average (mean-field term) must depend on the history of α_t . Note also that in (4), the conditional expectations $\mathbb{E}(\psi(x(t)) | \mathcal{F}_{t-}^\alpha)$ are used for more generality than the usual mean. The idea is that we may consider the conditional moments beyond the usual consideration of mean (e.g., second moment, variance, as well as higher moments). The motivation stems from the applications of cyber-physical systems and social network modeling. The boundedness essentially ensure the finite moment, which can be done by using the usual truncated moments. This is analogous to the treatment of diffusion processes, we truncate the function involved.

The control problem we are interested in is to minimize the cost functional given by

$$\begin{aligned}
 J(u) = \mathbb{E} \left[\int_0^T h(t, x(t), \mathbb{E}(\phi(x(t)) | \mathcal{F}_{t-}^\alpha), u(t), \alpha(t-)) dt \right. \\
 \left. + g(x(T), \mathbb{E}(\eta(x(T)) | \mathcal{F}_T^\alpha), \alpha(T)) \right], \quad (5)
 \end{aligned}$$

where $h(\cdot, \cdot, \cdot, \cdot) : [0, T] \times \mathbb{R}^2 \times U \times \mathcal{M} \rightarrow \mathbb{R}$, $g(\cdot, \cdot, \cdot) : \mathbb{R}^2 \times \mathcal{M} \rightarrow \mathbb{R}$, and $\phi(\cdot), \eta(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ are given functions. For $(t, x, y, u, i_0) \in [0, T] \times \mathbb{R}^2 \times U \times \mathcal{M}$, $h_x(t, x, y, u, i_0)$, $h_y(t, x, y, u, i_0)$, $g_x(x, y, i_0)$, and $g_y(x, y, i_0)$ denote the partial derivatives of h and g in a usual way.

Define the Hamiltonian

$$H(t, \hat{x}, u, p, q, i_0) = h(t, x_1, x_2, u, i_0) + b(t, x_1, x_3, u, i_0)p + \sigma(t, x_1, x_4, u, i_0)q,$$

where $\hat{x} = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$. We make the following assumptions.

- (A1) The functions $\psi(\cdot)$, $\phi(\cdot)$, $\varphi(\cdot)$, and $\eta(\cdot)$ are continuously differentiable; $g(\cdot, \cdot, i_0)$ is continuously differentiable with respect to (x, y) ; $b(\cdot, \cdot, \cdot, i_0)$, $\sigma(\cdot, \cdot, \cdot, i_0)$, and $h(\cdot, \cdot, \cdot, i_0)$ are continuous in t and continuously differentiable with respect to (x, y, u) .
- (A2) In (A1), for each t and i_0 , all derivatives of $\psi(\cdot)$, $\phi(\cdot)$, $\varphi(\cdot)$, $g(\cdot, \cdot, i_0)$, $b(t, \cdot, \cdot, i_0)$, $\sigma(t, \cdot, \cdot, i_0)$, and $h(t, \cdot, \cdot, i_0)$ with respect to x, y , and u are Lipschitz continuous and bounded.
- (A3) The functions $\psi(\cdot)$, $\phi(\cdot)$, $\varphi(\cdot)$, and $\eta(\cdot)$ are convex, the function $g(\cdot, \cdot, \cdot)$ is convex in (x, y) , and the Hamiltonian $H(\cdot, \cdot, \cdot, \cdot, \cdot)$ is convex in (\hat{x}, u) .
- (A4) The functions $b_y(\cdot, \cdot, \cdot, \cdot)$, $\sigma_y(\cdot, \cdot, \cdot, \cdot)$, $h_y(\cdot, \cdot, \cdot, \cdot)$, and $g_y(\cdot, \cdot, \cdot)$ are nonnegative.

For simplicity, for a random variable x , $H(t, x, u, p, q, i_0)$ will be used instead of $H(t, x, \mathbb{E}(\phi(x)|\mathcal{F}_{t-}^\alpha), \mathbb{E}(\psi(x)|\mathcal{F}_{t-}^\alpha), \mathbb{E}(\varphi(x)|\mathcal{F}_{t-}^\alpha), u, p, q, i_0)$ with a little abuse of notation. That is,

$$H(t, x, u, p, q, i_0) = h(t, x, \mathbb{E}(\phi(x)|\mathcal{F}_{t-}^\alpha), u, i_0) + b(t, x, \mathbb{E}(\psi(x)|\mathcal{F}_{t-}^\alpha), u, i_0)p + \sigma(t, x, \mathbb{E}(\varphi(x)|\mathcal{F}_{t-}^\alpha), u, i_0)q. \tag{6}$$

Let

$$\lambda_t * dM_t = \sum_{i_0, j_0 \in \mathcal{M}} \lambda_{i_0 j_0}(t) dM_{i_0 j_0}(t).$$

Then

$$\int_0^t \lambda_s * dM_s = \sum_{i_0, j_0 \in \mathcal{M}} \int_0^t \lambda_{i_0 j_0}(s) dM_{i_0 j_0}(s).$$

The proofs of the following theorems are also omitted. They can be found in [NY].

THEOREM 2.3. *Under assumptions (A1) and (A2), if $\bar{u}(\cdot)$ is an optimal control with state trajectory $\bar{x}(\cdot)$, then there exists a triple $(\bar{p}, \bar{q}, \bar{\lambda})$ of adapted processes that satisfies the BSDE*

$$\begin{aligned} d\bar{p}_t &= -[\bar{b}_x(t)\bar{p}_t + \bar{\sigma}_x(t)\bar{q}_t + \bar{h}_x(t)] dt \\ &\quad - \left[\mathbb{E}(\bar{b}_y(t)\bar{p}_t | \mathcal{F}_{t-}^\alpha) \bar{\psi}_x(t) + \mathbb{E}(\bar{\sigma}_y(t)\bar{q}_t | \mathcal{F}_{t-}^\alpha) \bar{\varphi}_x(t) + \mathbb{E}(\bar{h}_y(t) | \mathcal{F}_{t-}^\alpha) \bar{\phi}_x(t) \right] dt \\ &\quad + \bar{q}_t dw(t) + \bar{\lambda}_t * dM_t, \end{aligned} \tag{7}$$

$$\bar{p}_T = \bar{g}_x(T) + \mathbb{E}(\bar{g}_y(T) | \mathcal{F}_{T-}^\alpha) \bar{\eta}_x(T). \tag{8}$$

such that

$$\frac{d}{du} H(t, \bar{x}_t, \bar{u}_t, \bar{p}_t, \bar{q}_t, \alpha_{t-})(v - \bar{u}_t) \geq 0, \quad dt \text{ d}\mathbb{P}\text{-a.s. on } [0, T] \times \Omega \text{ for any } v \in U,$$

where

$$\begin{aligned} \bar{b}(t) &= b(t, \bar{x}_t, \mathbb{E}(\psi(\bar{x}_t) | \mathcal{F}_{t-}^\alpha), \bar{u}_t, \alpha_{t-}), \\ \bar{\sigma}(t) &= \sigma(t, \bar{x}_t, \mathbb{E}(\varphi(\bar{x}_t) | \mathcal{F}_{t-}^\alpha), \bar{u}_t, \alpha_{t-}), \\ \bar{h}(t) &= h(t, \bar{x}_t, \mathbb{E}(\phi(\bar{x}_t) | \mathcal{F}_{t-}^\alpha), \bar{u}_t, \alpha_{t-}), \\ \bar{g}(t) &= g(\bar{x}_t, \mathbb{E}(\eta(\bar{x}_t) | \mathcal{F}_{t-}^\alpha), \alpha_t), \\ \bar{\phi}(t) &= \phi(\bar{x}_t), \quad \bar{\varphi}(t) = \varphi(\bar{x}_t), \quad \bar{\psi}(t) = \psi(\bar{x}_t), \quad \bar{\eta}(t) = \eta(\bar{x}_t). \end{aligned}$$

Recall that to reduce notation, we have used $\bar{p}_t, \bar{q}_t, \bar{\lambda}_t, \bar{x}_t, \bar{u}_t$, and α_{t-} to denote $\bar{p}(t), \bar{q}(t), \bar{\lambda}(t), \bar{x}(t), \bar{u}(t)$, and $\alpha(t-)$, respectively. We also obtain the following sufficient condition.

THEOREM 2.4. *Assume that assumptions (A1)–(A4) hold. Let $\bar{u}(\cdot)$ be a control in \mathcal{U} with the corresponding state trajectory $\bar{x}(\cdot)$. Let $(\bar{p}, \bar{q}, \bar{\lambda})$ be the solution to the adjoint equation. If*

$$H(t, \bar{x}_t, \bar{u}_t, \bar{p}_t, \bar{q}_t) = \inf_{v \in U} H(t, \bar{x}_t, v, \bar{p}_t, \bar{q}_t), \quad dt \text{ d}\mathbb{P}\text{-a.s. on } [0, T] \times \Omega, \tag{9}$$

then $\bar{u}(\cdot)$ is an optimal control.

3. Weak convergence and limit McKean–Vlasov equations for non-Markovian systems. Our main concern here is to be able to treat certain non-Markov driving noise. Almost all the stochastic models considered thus far have been centered around the “Markovian” formulation. That is, although allowing the treatment of the noise effect, the formulation has been confined to the cases that the driving noise is given by a Brownian motion or a jump process. Because of the Markovian formulation, we have good technical machineries to handle the systems. A question naturally arises. What can we do about a non-Markovian system? We show in this section, with a class of non-Markovian systems, we can still carry out a lot of work. For simplicity of notation, we consider the state, the functions involved and the Brownian motions are all one-dimensional. Consider the following system with mean-field interactions

$$\dot{x}_i^\varepsilon(t) = \tilde{b}\left(x_i^\varepsilon(t), \frac{1}{N} \sum_{j=1}^N x_j^\varepsilon(t), \alpha(t_-)\right) + \frac{1}{\varepsilon} \tilde{\sigma}\left(x_i^\varepsilon(t), \frac{1}{N} \sum_{j=1}^N x_j^\varepsilon(t), \alpha(t_-)\right) \xi_i^\varepsilon(t),$$

$$1 \leq i \leq N, \quad (10)$$

where $\tilde{b}(\cdot)$ and $\tilde{\sigma}(\cdot)$ are real-valued but otherwise similar to $b(\cdot)$ and $\sigma(\cdot)$ defined before and satisfy the similar conditions (to be specified shortly), and $\xi_i^\varepsilon(t) = \xi_i(t/\varepsilon^2)$ such that $\xi_i(\cdot)$ is a bounded stationary mixing process. Note that in (10), $x_i^\varepsilon(t) = x_{i,N}^\varepsilon(t)$. For simplicity, we omit the subscript N and assume $x_i^\varepsilon(\cdot)$ is scalar. Vector-valued $x_i^\varepsilon(\cdot)$ can be treated with no essential difficulty and only notation is more complex. In addition, we assume that $x_i^\varepsilon(0) = x_0$ for each $\varepsilon > 0$ and $i = 1, 2, \dots, N$. For each $\varepsilon > 0$ and positive integer N set

$$\mathcal{F}_{t_-}^\alpha = \sigma\{\alpha(s) : 0 \leq s < t\}, \quad \mathcal{F}_t^\varepsilon = \mathcal{F}_t^{\varepsilon, N} = \mathcal{F}_{t_-}^\alpha \vee \sigma\{\xi_i(s) : 0 \leq s \leq \frac{t}{\varepsilon^2}, i = 1, 2, \dots, N\},$$

and

$$\mathbf{x}^\varepsilon(\cdot) = (x_1^\varepsilon(\cdot), x_2^\varepsilon(\cdot), \dots, x_N^\varepsilon(\cdot))'. \quad (11)$$

Equation (10) can be written in a vector form. To do so, for $\mathbf{x} = (x_1, \dots, x_N)' \in \mathbb{R}^N$ and $\iota \in \mathcal{M}$ define

$$\begin{aligned} \mathbf{b}(\mathbf{x}, \iota) &= (\tilde{b}(x_1, \bar{\mathbf{x}}, \iota), \dots, \tilde{b}(x_N, \bar{\mathbf{x}}, \iota))' \in \mathbb{R}^N, \\ \boldsymbol{\sigma}(\mathbf{x}, \iota) &= \text{diag}(\tilde{\sigma}(x_1, \bar{\mathbf{x}}, \iota), \dots, \tilde{\sigma}(x_N, \bar{\mathbf{x}}, \iota)) \in \mathbb{R}^{N \times N}, \end{aligned} \quad (12)$$

where $\bar{\mathbf{x}} = (1/N) \sum_{j=1}^N x_j \in \mathbb{R}$ and $\text{diag}(c_1, \dots, c_N)$ is the diagonal matrix with entries c_1, \dots, c_N where $c_i \in \mathbb{R}$. Then (10) can be rewritten as

$$\dot{\mathbf{x}}^\varepsilon(t) = \mathbf{b}(\mathbf{x}^\varepsilon(t), \bar{\mathbf{x}}^\varepsilon(t), \alpha(t_-)) + \frac{1}{\varepsilon} \boldsymbol{\sigma}(\mathbf{x}^\varepsilon(t), \bar{\mathbf{x}}^\varepsilon(t), \alpha(t_-)) \boldsymbol{\xi}^\varepsilon(t), \quad (13)$$

where $\mathbf{x}^\varepsilon(t)$ is as in (11), $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)' \in \mathbb{R}^N$, and $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$.

As illustrated in [Kh, p. 34], under such driven processes $\{\xi_i(\cdot) : i = 1, \dots, N\}$, $\mathbf{x}^\varepsilon(\cdot)$ is not Markovian. In [Kh, Chapters 8–10] a collection of examples in applications is provided. In [vK], van Kampen noted that “non-Markov is the rule, Markov is the exception.” Thus there is a real need to consider non-Markovian systems. To proceed, we first state the conditions needed to get the weak convergence and law of large number result.

(H3) For each $i = 1, \dots, N$, the process $\xi_i^\varepsilon(t) = \xi_i(t/\varepsilon^2)$ such that $\xi_i(\cdot)$ is a stationary ϕ -mixing process that is bounded, that satisfies $\mathbb{E}\xi_i(t) = 0$, and that has right continuous sample paths and a mixing measure $\tilde{\phi}(t)$ satisfying

$$\int_0^\infty \tilde{\phi}^{1/2}(t) dt < \infty.$$

Moreover, $\xi_1(t), \xi_2(t), \dots, \xi_N(t)$ are independent for each t and

$$\int_0^\infty \mathbb{E}\xi_i(u)\xi_i(0) du = \Sigma, \quad i = 1, \dots, N.$$

(H4) The following conditions are satisfied.

- Conditions (H1) and (H2) hold.
- For each $i_0 \in \mathcal{M}$, $\tilde{b}(\hat{x}, \bar{x}, i_0)$ and $\tilde{\sigma}(\hat{x}, \bar{x}, i_0)$ are differentiable with respect to \hat{x} and \bar{x} . Denote the partial derivatives with respect to \hat{x} and \bar{x} by $\tilde{\sigma}_{\hat{x}}$ and $\tilde{\sigma}_{\bar{x}}$, respectively; assume that $\tilde{\sigma}_{\hat{x}}$ and $\tilde{\sigma}_{\bar{x}}$ are bounded.

REMARK 3.1. In (H3), for simplicity, we have assumed that for each i , $\xi_i(\cdot)$ has the same mixing measure.

We proceed to obtain a weak convergence results. A weak limit is obtained by first letting $\varepsilon \rightarrow 0$ which is a mean field switched diffusion system with N processes. Then we take $N \rightarrow \infty$ to obtain the second one in the spirit of Theorem 2.2 which is again a law of large number.

3.1. Weak convergence. We first state the main convergence theorem. In this process, we keep N fixed and take limit only as $\varepsilon \rightarrow 0$. For $(\hat{x}, \bar{x}, i_0) \in \mathbb{R}^2 \times \mathcal{M}$ let

$$\begin{aligned} \bar{b}(\hat{x}, \bar{x}, i_0) &= \tilde{b}(\hat{x}, \bar{x}, i_0) + \tilde{\sigma}_{\hat{x}}(\hat{x}, \bar{x}, i_0)\tilde{\sigma}(\hat{x}, \bar{x}, i_0)\Sigma, \\ \bar{b}_N(\hat{x}, \bar{x}, i_0) &= \bar{b}(\hat{x}, \bar{x}, i_0) + \frac{1}{N} \tilde{\sigma}_{\bar{x}}(\hat{x}, \bar{x}, i_0)\tilde{\sigma}(\hat{x}, \bar{x}, i_0)\Sigma, \\ \bar{\sigma}(\hat{x}, \bar{x}, i_0) &= \sqrt{2\Sigma}\tilde{\sigma}(\hat{x}, \bar{x}, i_0). \end{aligned} \tag{14}$$

We have the following theorem.

THEOREM 3.2. Assume (H3) and (H4) and the stochastic differential equation (15) has a unique solution (unique in the sense of distribution). Then for each $i = 1, \dots, N$, $\mathbf{x}^\varepsilon(\cdot)$ converges weakly to $\mathbf{x}(\cdot) = (x_1(\cdot), x_2(\cdot), \dots, x_N(\cdot))$ where $x_i(\cdot)$ is the unique solution of the stochastic differential equation

$$dx_i(t) = \bar{b}_N\left(x_i(t), \frac{1}{N} \sum_{j=1}^N x_j(t), \alpha(t_-)\right) dt + \bar{\sigma}\left(x_i(t), \frac{1}{N} \sum_{j=1}^N x_j(t), \alpha(t_-)\right) dw_i(t), \tag{15}$$

where $w_i(\cdot)$, $i = 1, \dots, N$, are independent Brownian motions and $\bar{b}_N(\cdot, \cdot, \cdot)$ and $\bar{\sigma}(\cdot, \cdot, \cdot)$ are defined as in (14).

REMARK 3.3. Note that both $\mathbf{x}^\varepsilon(\cdot)$ and $\mathbf{x}(\cdot)$ depend on N as well. So strictly speaking, these should have been written as $\mathbf{x}^{\varepsilon, N}(\cdot)$ and $\mathbf{x}^N(\cdot)$, respectively. However, for notational simplicity, in the above and the following computation, we suppress the dependence on N although keeping in mind this dependence is there.

To proceed, note that equation (15) can be written in a vector form. Define

$$\begin{aligned} \bar{\mathbf{b}}_N(\mathbf{x}, \iota) &= (\bar{b}_N(x_1, \bar{\mathbf{x}}, \iota), \dots, \bar{b}_N(x_N, \bar{\mathbf{x}}, \iota))' \in \mathbb{R}^N, \\ \bar{\boldsymbol{\sigma}}(\mathbf{x}, \iota) &= \text{diag}(\bar{\sigma}(x_1, \bar{\mathbf{x}}, \iota), \dots, \bar{\sigma}(x_N, \bar{\mathbf{x}}, \iota)) \in \mathbb{R}^{N \times N}, \end{aligned} \tag{16}$$

for $\mathbf{x} = (x_1, \dots, x_N)'$ and $\iota \in \mathcal{M}$. Then we can rewrite (15) as

$$d\mathbf{x}(t) = \bar{\mathbf{b}}_N(\mathbf{x}(t), \alpha(t_-)) dt + \bar{\boldsymbol{\sigma}}(\mathbf{x}(t), \alpha(t_-)) dw(t), \tag{17}$$

where $w(t) = (w_1(t), \dots, w_N(t))' \in \mathbb{R}^N$. The limit Markov process $(\mathbf{x}(\cdot), \alpha(\cdot))$ is a solution of the martingale problem with the operator $\mathcal{L} = \mathcal{L}^{(N)}$ which is defined as for each (real-valued function) $f(\cdot, \cdot) : \mathbb{R}^N \times \mathcal{M} \mapsto \mathbb{R}$, and each $\iota \in \mathcal{M}$, $f(\cdot, \iota) \in C_c^2(\mathbb{R}^N)$ such that

$$\mathcal{L}f(\mathbf{x}, \iota) = \nabla f'(\mathbf{x}, \iota) \bar{\mathbf{b}}_N(\mathbf{x}, \iota) + \frac{1}{2} \text{tr}[\nabla^2 f(\mathbf{x}, \iota) \bar{\mathbf{A}}(\mathbf{x}, \iota)] + Qf(\mathbf{x}, \iota), \tag{18}$$

where $\bar{\mathbf{A}}(\mathbf{x}, \iota) = \bar{\boldsymbol{\sigma}}(\mathbf{x}, \iota) \bar{\boldsymbol{\sigma}}'(\mathbf{x}, \iota)$ and

$$Qf(\mathbf{x}, \iota) = \sum_{j \in \mathcal{M}} q_{\iota j} f(\mathbf{x}, j).$$

Under (H3) and (H4), the martingale problem with the operator \mathcal{L} has a unique solution [XZ]. In order to construct perturbed functions to prove the weak convergence, it is more convenient to consider each component of $\mathbf{x}(t)$. We therefore separate the operator \mathcal{L} into several components such that each of which corresponds to one component of $\mathbf{x}(t)$ as follows: For each $\mathbf{x} = (x_1, x_2, \dots, x_N)'$ in \mathbb{R}^N , $\iota \in \mathcal{M}$, $f(\cdot, \iota) \in C_c^2(\mathbb{R}^N)$ (C^2 function with compact support), set

$$\begin{aligned} \mathcal{L}_i f(\mathbf{x}, \iota) &= f_{x_i}(\mathbf{x}, \iota) \bar{b}_N\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) + \frac{1}{2} f_{x_i x_i}(\mathbf{x}, \iota) \left[\bar{\sigma}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \right]^2 \\ &= f_{x_i}(\mathbf{x}, \iota) \left[\bar{b}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) + \tilde{\sigma}_x\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \tilde{\sigma}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \Sigma \right. \\ &\quad \left. + \frac{1}{N} \tilde{\sigma}_{\bar{x}}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \tilde{\sigma}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \Sigma \right] \\ &\quad + f_{x_i x_i}(\mathbf{x}, \iota) \left[\tilde{\sigma}\left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota\right) \right]^2 \Sigma. \end{aligned} \tag{19}$$

Now we have

$$\mathcal{L}f(\mathbf{x}, \iota) = \sum_{i=1}^N \mathcal{L}_i f(\mathbf{x}, \iota) + Qf(\mathbf{x}, \iota).$$

In fact, $f(\mathbf{x}(t), \alpha(t)) - \int_0^t \mathcal{L}f(\mathbf{x}(s), \alpha(s)) ds$ is a martingale; the martingale is driven by two martingales, one of them is driven by the Brownian motion and the other is driven by a compensated jump process (Poisson process) due to the Markov chain. We refer the reader to [YZ2, Chapter 2] for further details.

Using weak convergence methods, we shall show that the non-Markovian systems with mean-field interactions has a limit. The limit can be characterized as the solution of the martingale problem with operator \mathcal{L} . The corresponding stochastic differential equation

is given by (15). Roughly, the non-Markovian systems with driving noise $\xi_i^\varepsilon(\cdot)$ yields a limit in which the $\xi_i^\varepsilon(\cdot)$ are “replaced” by independent Brownian motions $w_i(\cdot)$.

Because the solution of (10) is not *a priori* bounded, we use a truncation device as in [Kh, p. 83]. Let ℓ be a fixed but otherwise arbitrary integer. Let $B_\ell = \{x : |x| \leq \ell\}$, the ball of radius ℓ centered at the origin. Consider a truncated process $\mathbf{x}^{\varepsilon,\ell}(\cdot) = (x_1^{\varepsilon,\ell}(\cdot), x_2^{\varepsilon,\ell}(\cdot), \dots, x_N^{\varepsilon,\ell}(\cdot))'$ defined by $x_i^{\varepsilon,\ell}(0) = x_0$, and $x_i^{\varepsilon,\ell}(t) = x_i^\varepsilon(t)$ on the set $\{t \leq \tau\}$, and

$$\lim_{m \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \mathbb{P}(\sup_{t \leq T} |x_i^{\varepsilon,\ell}(t)| \geq m) = 0, \quad 1 \leq i \leq N, \tag{20}$$

where $\tau = \inf\{s : \max_{1 \leq j \leq N} |x_j^\varepsilon(s)| > \ell\}$, the first exit time of $x_j^\varepsilon(\cdot)$'s from the ℓ -ball B_ℓ . Such a truncated process can be obtained by using a truncation function ρ^ℓ defined as follows

$$\rho^\ell(x) = \begin{cases} 1, & \text{if } x \in B_\ell, \\ 0, & \text{if } x \in \mathbb{R}^d - B_{\ell+1}, \\ \text{smooth,} & \text{otherwise.} \end{cases}$$

For $i = 1, \dots, N$, define the i th component $x_i^{\varepsilon,\ell}(\cdot)$ of this desired truncated process as the solution of

$$\dot{x}_i^{\varepsilon,\ell}(t) = \tilde{b}^\ell(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)) + \frac{1}{\varepsilon} \tilde{\sigma}^\ell(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)) \xi_i^\varepsilon(t), \tag{21}$$

where

$$\begin{aligned} \tilde{b}^\ell(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \nu) &= \tilde{b}(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \nu) \rho^\ell(x_i^{\varepsilon,\ell}(t)), \\ \tilde{\sigma}^\ell(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \nu) &= \tilde{\sigma}(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \nu) \rho^\ell(x_i^{\varepsilon,\ell}(t)). \end{aligned} \tag{22}$$

From the definition of $x_i^{\varepsilon,\ell}(\cdot)$, it is readily seen that

$$\mathbb{P}(\sup_{t \leq T} |x_i^{\varepsilon,\ell}(t)| \geq m) \leq \frac{1}{m} \mathbb{E} \sup_{t \leq T} |x_i^{\varepsilon,\ell}(t)| \rightarrow 0 \text{ as } m \rightarrow \infty.$$

It thus confirms that $\mathbf{x}^{\varepsilon,\ell}(\cdot)$ is an ℓ -truncation process [Kh]. We will first prove that for a fixed ℓ , $\{\mathbf{x}^{\varepsilon,\ell}(\cdot) : \varepsilon > 0\}$ is tight. Then we establish its weak convergence. Letting $\ell \rightarrow \infty$, we finally show that the untruncated process $\{\mathbf{x}^\varepsilon(\cdot) : \varepsilon > 0\}$ also converges. To proceed, we state a lemma first.

LEMMA 3.4. *Under (H3)–(H4), $\{\mathbf{x}^{\varepsilon,\ell}(\cdot) : \varepsilon > 0\}$ is tight in $D([0, T], \mathbb{R}^N)$, where $D([0, T], \mathbb{R}^N)$ is the space of \mathbb{R}^N -valued functions defined on $[0, T]$ that are right continuous and that have left limit, endowed with the Skorohod topology.*

REMARK 3.5. Although the solution of (10) is not Markovian, we can still define an operator \mathcal{L}^ε by using the idea of p -limit [Kh, p. 38]. The p -limit requires the sequence under consideration to be integrable and the convergence to the limit taking place in the

L^1 sense. Let h and g be in $C_c^2(\mathbb{R})$. With $f(\cdot) \in D(\mathcal{L}^\varepsilon)$, the domain of \mathcal{L}^ε and $\mathcal{L}^\varepsilon h = g$, if

$$p\text{-}\lim_{\delta \rightarrow 0} \left[\frac{\mathbb{E}_t^\varepsilon h(t + \delta) - h(t)}{\delta} - g(t) \right] = 0,$$

where \mathbb{E}_t^ε denotes the conditional expectation with respect to $\mathcal{F}_t^\varepsilon$. Then \mathcal{L}^ε is still a type of infinitesimal operator. Moreover, for $h(\cdot) \in D(\mathcal{L}^\varepsilon)$,

$$\begin{aligned} h(t) - \int_0^t \mathcal{L}^\varepsilon h(u) \, du & \text{ is a martingale, and} \\ \mathbb{E}_t^\varepsilon h(t + s) - h(t) & = \int_t^{t+s} \mathbb{E}_t^\varepsilon \mathcal{L}^\varepsilon h(u) \, du \quad \text{w.p.1;} \end{aligned} \tag{23}$$

see [Kh, Theorem 1, p. 39] and [EK]. The first line above confirms the martingale property and the second line is a Dynkin like formula with \mathbb{E} replaced by the conditional expectation \mathbb{E}_t^ε .

Proof of Lemma 3.4. We shall use the idea of p -limit, the truncation device, and the perturbed test function methods (see [Kh, Chapter 4] and [KY, p. 172]). Mainly, we define a perturbation that is small in magnitude and that results in the desired cancellation. Note that for each $\iota \in \mathcal{M}$ and each $f(\cdot, \iota) \in C_c^2(\mathbb{R}^N)$, it is easy to see that

$$\mathcal{L}^{\varepsilon, \ell} f(\mathbf{x}, \iota) = \sum_{i=1}^N \mathcal{L}_i^{\varepsilon, \ell} f(\mathbf{x}, \iota) + Qf(\mathbf{x}, \iota), \quad \mathbf{x} = (x_1, x_2, \dots, x_N)' \in \mathbb{R}^N,$$

where for each $i = 1, \dots, N$,

$$\mathcal{L}_i^{\varepsilon, \ell} f(\mathbf{x}, \iota) = f_{x_i}(\mathbf{x}, \iota) \tilde{b}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) + \frac{1}{\varepsilon} f_{x_i}(\mathbf{x}, \iota) \tilde{\sigma}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \xi_i^\varepsilon(t). \tag{24}$$

We have

$$\begin{aligned} \mathcal{L}_i^{\varepsilon, \ell} f(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) & = f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{b}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \\ & \quad + \frac{1}{\varepsilon} f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_i^\varepsilon(t). \end{aligned} \tag{25}$$

The notation reflects the dependence on the truncation level ℓ . For each i , the term

$$\frac{1}{\varepsilon} f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_i^\varepsilon(t)$$

is undesirable due to the order $O(\varepsilon^{-1})$. We wish to “average” it out. To do so, we proceed to define a perturbation. To emphasize the dependence on $x_i^\varepsilon(t)$ and $\alpha(t)$, we define a perturbation by

$$\begin{aligned} f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-), t) \\ = \frac{1}{\varepsilon} \int_t^T \mathbb{E}_t^\varepsilon f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_i^\varepsilon(u) \, du. \end{aligned} \tag{26}$$

Because $\mathbb{E}\xi_i(t) = 0$ for each $t > 0$,

$$\begin{aligned}
 & f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &= \frac{1}{\varepsilon} \int_t^T f_{x_i}(x_i^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell\left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)\right) [\mathbb{E}_t^\varepsilon \xi_i^\varepsilon(u) - \mathbb{E} \xi_i^\varepsilon(u)] du. \quad (27)
 \end{aligned}$$

Using $\xi_i^\varepsilon(t) = \xi_i(t/\varepsilon^2)$ and a change of variable $u \mapsto u/\varepsilon^2$, the above can be rewritten as

$$\begin{aligned}
 & \sup_{t \in [0, T]} |f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t)| \\
 &= \varepsilon \sup_{t \in [0, T]} \left| \int_{t/\varepsilon^2}^{T/\varepsilon^2} f_{x_i}(x_i^{\varepsilon,\ell}(t), \alpha(t_-), t) \tilde{\sigma}^\ell\left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)\right) \right. \\
 & \quad \left. \times [\mathbb{E}_t^\varepsilon \xi_i^\varepsilon(u) - \mathbb{E} \xi_i^\varepsilon(u)] du \right| \\
 &\leq K\varepsilon \sup_{t \in [0, T]} \int_{t/\varepsilon^2}^\infty \tilde{\phi}(u - (t/\varepsilon^2)) du = O(\varepsilon).
 \end{aligned}$$

Thus the perturbation is small in magnitude. We proceed to demonstrate that it results in desired cancellations. Define

$$f^\varepsilon(t) = f(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) + \sum_{i=1}^N f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t).$$

Because of the cancellation of the last term in (25),

$$\begin{aligned}
 \mathcal{L}^{\varepsilon,\ell} f^\varepsilon(t) &= \frac{\partial}{\partial t} f^\varepsilon(\alpha(t_-), t) + \sum_{i=1}^N \mathcal{L}_i^{\varepsilon,\ell} f^\varepsilon(\alpha(t_-), t) + Qf^\varepsilon(\alpha(t_-), t) \\
 &= \sum_{i=1}^N f_{x_i}(x_i^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{b}^\ell\left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)\right) \\
 & \quad + \sum_{k=1}^N \sum_{i=1}^N \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i}(x_i^{\varepsilon,\ell}(t), \alpha(t_-), t) \tilde{\sigma}^\ell\left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)\right) \right]_{x_k} \xi_i(u) du \\
 & \quad \times \tilde{\sigma}^\ell\left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-)\right) \xi_k^\varepsilon(t) + Qf(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) + O(\varepsilon), \quad (28)
 \end{aligned}$$

where $O(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. (Detailed calculation will be carried out in (31), (33), and (34)). In the next to the last line above, the term $[\dots]_{x_i}$ denotes the partial derivative of the term $[\dots]$ with respect to x_i at $x = x_i^{\varepsilon,\ell}(t)$. Because $\{\xi_i(t)\}$ is bounded and stationary mixing, it can be verified that $\mathcal{L}^{\varepsilon,\ell} f^\varepsilon(t)$ is uniformly integrable. It then follows from [Kh, Theorem 3.4, p. 48], $\{\mathbf{x}^{\varepsilon,\ell}(\cdot) : \varepsilon > 0\}$ is tight as desired. The lemma is proved. ■

LEMMA 3.6. *Under the conditions of Lemma 3.4, $\{\mathbf{x}^{\varepsilon,\ell}(\cdot) : \varepsilon > 0\}$ converges weakly to $\mathbf{x}^\ell(\cdot)$ as $\varepsilon \rightarrow 0$, where $\mathbf{x}^\ell(\cdot)$ is the solution of the martingale problem with operator \mathcal{L}^ℓ given by*

$$\mathcal{L}^\ell f(\mathbf{x}, \iota) = \sum_{i=1}^N \mathcal{L}_i^\ell f(\mathbf{x}, \iota) + Qf(\mathbf{x}, \iota),$$

with

$$\begin{aligned} \mathcal{L}_i^\ell f(\mathbf{x}, \iota) &= f_{x_i}(\mathbf{x}, \iota) \left[\tilde{b}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) + \tilde{\sigma}_x^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \tilde{\sigma}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \Sigma \right. \\ &\quad \left. + \frac{1}{N} \tilde{\sigma}_x^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \tilde{\sigma}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \Sigma \right] \\ &\quad + f_{x_i x_i}(\mathbf{x}, \iota) \left[\tilde{\sigma}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) \right]^2 \Sigma, \quad (29) \end{aligned}$$

for any $f(\mathbf{x}, \iota) \in C_c^2$ with $\tilde{b}^\ell(\cdot, \cdot, \cdot)$ and $\tilde{\sigma}^\ell(\cdot, \cdot, \cdot)$ defined as in (22).

Proof. By Lemma 3.4, $\{\mathbf{x}^{\varepsilon, \ell}(\cdot)\}$ is tight in $D([0, T], \mathbb{R}^d)$. By Prohorov’s theorem (see [Kh, p. 28] or [EK, p. 104]), we may extract a weakly convergent subsequence. Select such a subsequence, still denote the sequence by $\{\mathbf{x}^{\varepsilon, \ell}(\cdot)\}$ for simplicity, and denote the limit by $\mathbf{x}^\ell(\cdot)$. We proceed to characterize the limit process. By the Skorohod representation (see [Kh, pp. 29-30] or [EK, p. 102]), we assume with a slight abuse of notation that $\mathbf{x}^{\varepsilon, \ell}(\cdot)$ converges to $\mathbf{x}^\ell(\cdot)$ in the sense of with probability one. Also recall our notation of $\tilde{\sigma}_x^\ell$ and $\tilde{\sigma}_x^\ell$ for the partial derivatives of $\tilde{\sigma}$ with respect to the first and second variable x_i and $\sum_{j=1}^N x_j/N$, respectively.

To characterize the limit, we use the perturbed test function methods [Kh]. For each $\iota \in \mathcal{M}$ and any $f(\cdot, \iota) \in C_c^2(\mathbb{R}^N)$, define $f_1^\varepsilon(\cdot, \cdot)$ as in the tightness proof. Using the notation just mentioned above, we have

$$\begin{aligned} f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-), t) &= \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \mathbb{E}_t^\varepsilon \xi_i(u) du. \end{aligned}$$

To proceed, we first observe that because of the chain rule of differentiations,

$$\frac{\partial}{\partial x_i} \tilde{\sigma}^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) = \tilde{\sigma}_x^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right) + \frac{1}{N} \tilde{\sigma}_x^\ell \left(x_i, \frac{1}{N} \sum_{j=1}^N x_j, \iota \right). \quad (30)$$

Detailed calculation yields that

$$\begin{aligned} \mathcal{L}_k^{\varepsilon, \ell} f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-), t) &= \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \\ &\quad \times \tilde{b}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(u) du \\ &+ \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \\ &\quad \times \tilde{b}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(u) du \end{aligned}$$

$$\begin{aligned}
& + \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \\
& \quad \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(t) du \\
& + \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \\
& \quad \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(t) du, \tag{31}
\end{aligned}$$

and

$$\begin{aligned}
& \frac{\partial}{\partial t} f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
& = -\frac{1}{\varepsilon} f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i^\varepsilon(t). \tag{32}
\end{aligned}$$

Because of the ϕ -mixing property, for some generic positive constant $K > 0$,

$$\begin{aligned}
& \left| \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \right. \\
& \quad \left. \times \tilde{b}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i^\varepsilon(t) du \right| \leq K \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \tilde{\phi}(u - (t/\varepsilon^2)) du \leq K \varepsilon. \tag{33}
\end{aligned}$$

Likewise,

$$\begin{aligned}
& \left| \varepsilon \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] \right. \\
& \quad \left. \times \tilde{b}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(t) du \right| \leq K \varepsilon. \tag{34}
\end{aligned}$$

Thus we need only work on the remaining two terms in (31). Define

$$\begin{aligned}
& f_{2,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
& = \int_t^T \int_{s/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left\{ \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \right. \\
& \quad \left. \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] [\xi_i(u) \xi_k^\varepsilon(s) - \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(s))] \right\} du ds,
\end{aligned}$$

$$\begin{aligned}
 & f_{3,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &= \int_t^T \int_{s/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left\{ \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \right. \\
 & \quad \left. \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] [\xi_i(u) \xi_k^\varepsilon(s) - \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(s))] \right\} du ds. \tag{35}
 \end{aligned}$$

Note that

$$\begin{aligned}
 & f_{2,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &= \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \\
 & \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] \int_t^T \int_{s/\varepsilon^2}^{T/\varepsilon^2} [\mathbb{E}_t^\varepsilon(\xi_i(u) \xi_k^\varepsilon(s)) - \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(s))] du ds \tag{36}
 \end{aligned}$$

and that

$$\begin{aligned}
 & \left| \int_t^T \int_{s/\varepsilon^2}^{T/\varepsilon^2} [\mathbb{E}_t^\varepsilon(\xi_i(u) \xi_k^\varepsilon(s)) - \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(s))] du ds \right| \\
 & \leq \int_t^\infty \int_{s/\varepsilon^2}^\infty \left| \mathbb{E}_t^\varepsilon(\xi_i(u) \xi_k(s/\varepsilon^2)) - \mathbb{E}(\xi_i(u) \xi_k(s/\varepsilon^2)) \right| du ds \\
 & = \varepsilon^2 \int_{t/\varepsilon^2}^\infty \int_v^\infty \left| \mathbb{E}_t^\varepsilon(\xi_i(u) \xi_k(v)) - \mathbb{E}(\xi_i(u) \xi_k(v)) \right| du dv \\
 & \leq K \varepsilon^2 \int_{t/\varepsilon^2}^\infty \int_v^\infty \tilde{\phi}^{1/2}(u-v) \tilde{\phi}^{1/2}(v-t/\varepsilon^2) du dv = O(\varepsilon^2). \tag{37}
 \end{aligned}$$

Similar to the case of $f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t)$, it can be verified that

$$\sup_{0 \leq t \leq T} |f_{p,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t)| = O(\varepsilon^2) \text{ for } p = 2, 3 \text{ and } 1 \leq i, k \leq N.$$

Thus, the perturbations are all small. Moreover, for each $l = 1, 2, \dots, N$, we obtain

$$\mathcal{L}_l^{\varepsilon,\ell} [f_{2,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) + f_{3,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t)] = e_l^\varepsilon(t), \tag{38}$$

where $e_l^\varepsilon(t) = O(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. In addition,

$$\begin{aligned}
 & \frac{\partial}{\partial t} f_{2,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &= - \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \\
 & \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} [\mathbb{E}_t^\varepsilon(\xi_i(u) \xi_k^\varepsilon(t)) - \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t))] du
 \end{aligned}$$

$$\begin{aligned}
 &= - \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] du \\
 &\quad \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(t) \\
 &\quad + \left[f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \\
 &\quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du. \tag{39}
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 &\frac{\partial}{\partial t} f_{3,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &= - \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}_t^\varepsilon \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_i(u) \right] du \\
 &\quad \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \xi_k^\varepsilon(t) \\
 &\quad + \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right. \\
 &\quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon,\ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du. \tag{40}
 \end{aligned}$$

Now, we define the perturbed function as follows

$$\begin{aligned}
 \tilde{f}^\varepsilon(t) &= f(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) + \sum_{i=1}^N f_{1,i}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) \\
 &\quad + \sum_{i=1}^N \sum_{k=1}^N (f_{2,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t) + f_{3,i,k}^\varepsilon(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-), t)). \tag{41}
 \end{aligned}$$

Using (31), (39), and (40), applying $\mathcal{L}^{\varepsilon,\ell}$ to $\tilde{f}^\varepsilon(t)$, and recalling that $\xi_i^\varepsilon(t) = \xi_i(t/\varepsilon^2)$, upon cancellations, direct calculation yields that

$$\begin{aligned}
 \mathcal{L}^{\varepsilon,\ell} \tilde{f}^\varepsilon(t) &= \frac{\partial}{\partial t} \tilde{f}^\varepsilon(t) + \sum_{i=k}^N \mathcal{L}_k^{\varepsilon,\ell} f^\varepsilon(t) + Q f^\varepsilon(\alpha(t_-), t) \\
 &= \sum_{i=1}^N f_{x_i}(\mathbf{x}^{\varepsilon,\ell}(t), \alpha(t_-)) \tilde{b}^\ell \left(x_i^{\varepsilon,\ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon,\ell}(t), \alpha(t_-) \right)
 \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^N \sum_{k=1}^N \left[f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du \\
& + \sum_{i=1}^N \sum_{k=1}^N \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du \\
& + Qf(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) + O(\varepsilon), \tag{42}
\end{aligned}$$

Observe that $\mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) = 0$ if $i \neq k$. Under the conditions of Lemma 3.6, as $\varepsilon \rightarrow 0$, for each $i = 1, \dots, N$,

$$\int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_i(t/\varepsilon^2)) du \rightarrow \int_0^\infty \mathbb{E}(\xi_i(u) \xi_i(0)) du = \Sigma. \tag{43}$$

Hence, using (30), we can demonstrate that

$$\begin{aligned}
& \sum_{i=1}^N \sum_{k=1}^N \left[f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \frac{\partial}{\partial x_k} \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du \\
& \rightarrow \sum_{i=1}^N \left[f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}_x^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \Sigma \right] \\
& + \frac{1}{N} \sum_{i=1}^N \left[f_{x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}_x^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \Sigma \right] \tag{44}
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{i=1}^N \sum_{k=1}^N \left[f_{x_i x_k}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right. \\
& \quad \left. \times \tilde{\sigma}^\ell \left(x_k^{\varepsilon, \ell}, \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right] \int_{t/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}(\xi_i(u) \xi_k^\varepsilon(t)) du \\
& \rightarrow \sum_{i=1}^N f_{x_i x_i}(\mathbf{x}^{\varepsilon, \ell}(t), \alpha(t_-)) \left[\tilde{\sigma}^\ell \left(x_i^{\varepsilon, \ell}(t), \frac{1}{N} \sum_{j=1}^N x_j^{\varepsilon, \ell}(t), \alpha(t_-) \right) \right]^2 \Sigma. \tag{45}
\end{aligned}$$

In view of the estimates obtained so far, putting the pieces together, and using [Kh, Theorem 2, p. 44], we have proved that $\mathbf{x}^{\varepsilon, \ell}(\cdot)$ converges weakly to $\mathbf{x}^\ell(\cdot)$ such that $\mathbf{x}^\ell(\cdot)$ is the solution of martingale problem with operator \mathcal{L}^ℓ . ■

LEMMA 3.7. *Under the conditions of Lemma 3.4, $\mathbf{x}^\varepsilon(\cdot)$ converges weakly to $\mathbf{x}(\cdot)$.*

Idea of proof. Here, we can use the idea similar to the last part of [Kh, p. 46] to prove that the untruncated process $\{\mathbf{x}^\varepsilon(\cdot)\}$ also converges. Denote by \mathbb{P} and \mathbb{P}^ℓ the measures induced by $\mathbf{x}(\cdot)$ and $\mathbf{x}^\ell(\cdot)$, respectively. Note that the martingale problem has a unique solution on $[0, T]$. Thus the measure \mathbb{P} is unique. Then for each $T < \infty$, \mathbb{P} is the same with \mathbb{P}^ℓ on all Borel subsets of the set of paths in $D([0, \infty) : \mathbb{R}^N)$ with values in the product of ℓ -balls B_ℓ^N for all $t \leq T$. In addition, the existence of the unique solution of the martingale problem implies $\lim_{\ell \rightarrow \infty} \mathbb{P}(\sup_{t \leq T} |\mathbf{x}(t)| \leq \ell) = 1$. Thus as in the reference mentioned above, the desired convergence follows. Thus the proof of Theorem 3.2 is complete. ■

3.2. Mean-field limit. Using the results obtained so far, we can carry out analysis similar to that of Theorem 2.2 to prove the law of large number for the limit interacting diffusion system (15). More precisely, for each positive integer N , consider

$$\mu_N(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{x_i(\cdot)},$$

where $\mathbf{x}(\cdot) = (x_1(\cdot), x_2(\cdot), \dots, x_N(\cdot))'$ with $x_i(\cdot)$ for $i = 1, \dots, N$, being the solutions to the limit system (15). In view of Remark 3.3, both $\mathbf{x}^\varepsilon(\cdot)$ and $\mathbf{x}(\cdot)$ should have really been written as $\mathbf{x}^{\varepsilon, N}(\cdot)$ and $\mathbf{x}^N(\cdot)$, respectively. That is, they are N dependent. However, we have used the simplified notation to suppress the N -dependence. The next result is concerned with $N \rightarrow \infty$;

THEOREM 3.8. *Under assumptions of Theorem 3.2, $(\mu_N(\cdot), \alpha(\cdot))$ converges weakly to a process $(\mu(\cdot), \alpha(\cdot))$ where*

$$(\mu(t), \alpha(t)) = (\mathcal{L}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t)), \quad 0 \leq t \leq T,$$

and $y(t)$, $0 \leq t \leq T$, is the unique solution of the following McKean–Vlasov stochastic differential equation

$$dy(t) = \bar{b}(y(t), \mathbb{E}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dt + \bar{\sigma}(y(t), \mathbb{E}(y(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dw(t) \tag{46}$$

where $\tilde{w}(\cdot)$ is a standard Brownian motion independent of $\alpha(\cdot)$.

The proof is similar to that of Theorem 2.2 (see [NY]). We only note that

$$\sup_{\hat{x}, \bar{x}, i_0} |\bar{b}_N(\hat{x}, \bar{x}, i_0) - \bar{b}(\hat{x}, \bar{x}, i_0)| \leq \frac{C}{N}$$

for some constant C and that the term with order $O(N^{-1})$ disappears in the limit.

Next, for each $i = 1, 2, \dots$ the following McKean–Vlasov stochastic differential equation has a unique solution

$$dy_i(t) = \bar{b}(y_i(t), \mathbb{E}(y_i(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dt + \bar{\sigma}(y_i(t), \mathbb{E}(y_i(t) | \mathcal{F}_{t-}^\alpha), \alpha(t_-)) dw_i(t).$$

We observe that

$$\mathbb{E}(y_i(t) | \mathcal{F}_T^\alpha) = \mathbb{E}(y_i(t) | \mathcal{F}_{t-}^\alpha), \quad \mathbb{E}(x_i(t) | \mathcal{F}_T^\alpha) = \mathbb{E}(x_i(t) | \mathcal{F}_{t-}^\alpha), \quad 0 \leq t \leq T.$$

In addition, $x_1(\cdot), x_2(\cdot), \dots, x_N(\cdot)$ are conditionally independent given \mathcal{F}_T^α and identically distributed. Thus by using the Gronwall's inequality and the conditional expectation $\mathbb{E}(\cdot | \mathcal{F}_T^\alpha)$ instead of the expectation $\mathbb{E}(\cdot)$, the standard approach (see [S, NN]) yields the following estimate.

LEMMA 3.9. *There exists a constant C independent of i, N such that for all $1 \leq i \leq N$,*

$$\sup_{0 \leq t \leq T} \mathbb{E}(|x_i(t) - y_i(t)|^2 | \mathcal{F}_{t-}^\alpha) \leq \frac{C}{N}. \quad (47)$$

Using the results obtained so far, we see that $x_i^\varepsilon(t)$ converges weakly to $y(t)$ as $\varepsilon \rightarrow 0$ and $N \rightarrow \infty$ for each $i = 1, \dots, N$ and $0 \leq t \leq T$, that is

$$\lim_{N \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \mathbb{E}[f(x_i^\varepsilon(t))] = \mathbb{E}[f(y(t))]$$

for each bounded continuous function $f(\cdot)$ where $y(\cdot)$ is the solution of the McKean–Vlasov equation (46).

4. Further remarks. This paper reviews some recent results on switching diffusions with mean-field interactions together with the associated control problems. It also examines a non-Markovian case. The non-Markovian system is treated by using repeated limits. It is possible that we investigate the limit problem in the last part with $\varepsilon \rightarrow 0$ and $N = N_\varepsilon \rightarrow \infty$.

Acknowledgments. The research of G. Yin was supported in part by the Air Force Office of Scientific Research under grant FA9550-18-1-0268. The research of S. L. Nguyen was supported in part by the University of Puerto Rico under grant FIPI-5310020.00.

References

- [BF] J. Baladron, D. Fasoli, O. Faugeras, J. Touboul, *Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons*, J. Math. Neuroscience 2 (2012), art. 10.
- [BY] A. Bensoussan, J. Frehse, P. Yam, *Mean Field Games and Mean Field Type Control Theory*, SpringerBriefs Math., Springer, New York, 2013.
- [CM] P. Contucci, I. Gallo, G. Menconi, *Phase transitions in social sciences: two-populations mean field theory*, Internat. J. Modern Phys. B 22 (2008), 2199–2212.
- [CD] O. Costa, F. Dufour, *Continuous Average Control of Piecewise Deterministic Markov Processes*, SpringerBriefs Math., Springer, New York, 2013.
- [D] D. A. Dawson, *Critical dynamics and fluctuations for a mean-field model of cooperative behavior*, J. Statist. Phys. 31 (1983), 29–85.
- [DG] D. A. Dawson, J. Gärtner, *Large deviations from the McKean–Vlasov limit for weakly interacting diffusions*, Stochastics 20 (1987), 247–308.
- [DV] D. A. Dawson, J. Vaillancourt, *Stochastic McKean–Vlasov equations*, NoDEA Nonlinear Differential Equations Appl. 2 (1995), 199–229.
- [DZ] D. A. Dawson, X. Zheng, *Law of large numbers and central limit theorem for unbounded jump mean-field models*, Adv. in Appl. Math. 12 (1991), 293–326.
- [EK] S. N. Ethier, T. G. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley Ser. Probab. Math. Statist. Wiley, New York, 1986.

- [G] J. Gärtner, *On the McKean–Vlasov limit for interacting diffusions*, Math. Nachr. 137 (1988), 197–248.
- [HY] M. Hamdi, V. Krishnamurthy, G. Yin, *Tracking a Markov-modulated stationary degree distribution of a dynamic random graph*, IEEE Trans. Inform. Theory 60 (2014), 6609–6625.
- [HM] M. Huang, P. E. Caines, R. P. Malhamé, *Individual and mass behavior in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions*, in: 42nd IEEE Conference on Decision Control, 2003, 98–103.
- [HC] M. Huang, R. P. Malhamé, P. E. Caines, *Large population stochastic dynamic games: Closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle*, Commun. Inf. Syst. 6 (2006), 221–252.
- [vK] N. G. van Kampen, *Remarks on non-Markov processes*, Brazilian J. Physics 28 (1998), 90–96.
- [K] V. N. Kolokoltsov, *Nonlinear Markov Processes and Kinetic Equations*, Cambridge Tracts in Math. 182, Cambridge Univ. Press, Cambridge, 2010.
- [Kh] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.
- [KY] H. J. Kushner, G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, second ed., Appl. Math. (N.Y.) 35, Springer, New York, 2003.
- [LL] J. M. Lasry, P.-L. Lions, *Jeux à champ moyen. I. Le cas stationnaire*, C. R. Math. Acad. Sci. Paris 343 (2006), 619–625.
- [Lb] J. M. Lasry, P.-L. Lions, *Jeux à champ moyen. II. Horizon fini et contrôle optimal*, C. R. Math. Acad. Sci. Paris 343 (2006), 679–684.
- [MY] X. Mao, C. Yuan, *Stochastic Differential Equations with Markovian Switching*, Imperial College Press, London, 2006.
- [M] H. P. McKean, *A class of Markov processes associated with nonlinear parabolic equations*, Proc. Nat. Acad. Sci. U.S.A. 56 (1966), 1907–1911.
- [NN] D. T. Nguyen, S. L. Nguyen, N. H. Du, *On mean field systems with multi-classes*, Discrete Contin. Dyn. Syst. 40 (2020), 683–707.
- [NH] S. L. Nguyen, G. Yin, T. A. Hoang, *On laws of large numbers for systems with mean-field interactions and Markovian switching*, Stochastic Process. Appl. 130 (2020), 262–296.
- [NY] S. L. Nguyen, D. T. Nguyen, G. Yin, *A stochastic maximum principle for switching diffusions using conditional mean fields with applications to control problems*, ESAIM Control Optim. Calc. Var. 26 (2020), Paper no. 69.
- [O] K. Oelschläger, *A martingale approach to the law of large numbers for weakly interacting stochastic processes*, Ann. Probab. 12 (1984), 458–479.
- [SZ] S. P. Sethi, Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Systems Control Found. Appl., Birkhäuser, Boston, 1994.
- [S] A. S. Sznitman, *Topics in propagation of chaos*, in: École d’Été de Probabilités de Saint-Flour XIX – 1989, Lecture Notes in Math. 1464, Springer, Berlin, 1991, 165–251.
- [TY] K. Tran, G. Yin, L. Y. Wang, *A generalized Goodwin business cycle model in random environment*, J. Math. Anal. Appl. 438 (2016), 311–327.
- [XZ] F. Xi, C. Zhu, *On the martingale problem and Feller and strong Feller properties for weakly coupled Lévy type operators*, Stochastic Process. Appl. 128 (2018), 4277–4308.
- [YI] G. Yin, V. Krishnamurthy, C. Ion, *Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization*, SIAM J. Optim. 14 (2004), 1187–1215.

- [YZ1] G. Yin, H. Q. Zhang, Q. Zhang, *Applications of Two-Time-Scale Markovian Systems*, Science Press, Beijing, China, 2013.
- [YZ] G. Yin, Q. Zhang, *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*, second ed., Stoch. Model. Appl. Probab. 37, Springer, New York, 2013.
- [YZ2] G. Yin, C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*, Springer, New York, 2010.
- [ZY] X. Y. Zhou, G. Yin, *Markowitz's mean-variance portfolio selection with regime switching: A continuous-time model*, SIAM J. Control Optim. 42 (2003), 1466–1482.
- [ZY1] C. Zhu, G. Yin, *On competitive Lotka–Volterra model in random environments*, J. Math. Anal. Appl. 357 (2009), 154–170.