Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 130 (2020) 2519–2551

stochastic processes and their applications

www.elsevier.com/locate/spa

Directed chain stochastic differential equations

Nils Detering, Jean-Pierre Fouque, Tomoyuki Ichiba*

Department of Statistics and Applied Probability, South Hall, University of California, Santa Barbara, CA 93106, USA

Received 14 December 2018; received in revised form 23 June 2019; accepted 12 July 2019
Available online 19 July 2019

Abstract

We propose a particle system of diffusion processes coupled through a chain-like network structure described by an infinite-dimensional, nonlinear stochastic differential equation of McKean–Vlasov type. It has both (i) a local chain interaction and (ii) a mean-field interaction. It can be approximated by a limit of finite particle systems, as the number of particles goes to infinity. Due to the local chain interaction, propagation of chaos does not necessarily hold. Furthermore, we exhibit a dichotomy of presence or absence of mean-field interaction, and we discuss the problem of detecting its presence from the observation of a single component process.

MSC: primary 60H10; secondary 60K35

Keywords: Interacting stochastic processes; Stochastic equation with constraints; Law of large numbers; Particle system approximation; Detecting mean-field

1. Introduction

Let us consider a directed graph (or oriented network) of vertices $\{1, \dots, n\}$ on a circle in the sense that each vertex i in the graph is the head of an arrow directed from its neighboring vertex $i+1$ for $i = 1, \dots, n-1$, and the boundary vertex n is the head of an arrow directed from the first vertex 1. On some probability space with independent Brownian motions $(W_{\cdot,i})$, $1 \le i \le n$, assigned to the vertices, we consider a process $X_{\cdot,i}$ defined by the following system of equations which incarnates this graph structure through drifts:

$$\begin{aligned} \mathrm{d}X_{t,i} &= h(X_{t,i}, X_{t,i+1})\mathrm{d}t + \mathrm{d}W_{t,i}\,; \quad t \ge 0, \quad i = 1, \dots, n-1, \\ \mathrm{d}X_{t,n} &= h(X_{t,n}, X_{t,1})\mathrm{d}t + \mathrm{d}W_{t,n}\,. \end{aligned} \tag{1.1}$$

* Corresponding author.
E-mail addresses: detering@pstat.ucsb.edu (N. Detering), fouque@pstat.ucsb.edu (J.-P. Fouque), ichiba@pstat.ucsb.edu (T. Ichiba).

https://doi.org/10.1016/j.spa.2019.07.009
0304-4149/

The initial values $X_{0,i}$ are independent and identically distributed random variables, independent of $(W_{\cdot,i})$, $1 \le i \le n$. Furthermore, $h : \mathbb{R}^2 \to \mathbb{R}$ is a Lipschitz function.

We view $(X_{\cdot,1}, \dots, X_{\cdot,n})$ as a particle system interacting through this particular directed graph. The system is invariant under a shift of the indexes of the particles. In particular, the law of $X_{\cdot,i}$ is the same as the law of $X_{\cdot,1}$ for every i and also the joint law of $(X_{\cdot,i}, X_{\cdot,i+1})$ is the same as the joint law of $(X_{\cdot,1}, X_{\cdot,2})$ for every i. Let us call such interaction in (1.1) a *directed chain* interaction. Note that if $h(x_1, x_2) = x_2 - x_1$, $(x_1, x_2) \in \mathbb{R}^2$, it is a simple Ornstein–Uhlenbeck type system (or a Gaussian cascade). Intuitively, because of the mean-reverting feature of Ornstein–Uhlenbeck type drifts, the particle $X_{\cdot,i}$ at vertex i in (1.1) tends to be close to the neighboring particle $X_{\cdot,i+1}$ *locally* under this particular choice of function h.

For comparison, on the same probability space, we also consider a typical mean-field interacting system where each particle is attracted towards the mean, defined by

$$\mathrm{d}X_{t,i} = \frac{1}{n}\sum_{\substack{j=1\\ j\neq i}}^{n} h\big(X_{t,i}, X_{t,j}\big)\mathrm{d}t + \mathrm{d}W_{t,i}\,; \quad t \ge 0, \quad i = 1, \dots, n. \tag{1.2}$$

This system (1.2) is invariant under permutations of indexes of particles, while the system (1.1) only possesses the shift invariance. Again, if $h(x_1, x_2) = x_2 - x_1$, $(x_1, x_2) \in \mathbb{R}^2$, the particle $X_{\cdot,i}$ at node i is *directly* attracted towards the mean $(X_{\cdot,1} + \cdots + X_{\cdot,n})/n$ of the system. This type of mean-field model has been considered in [6] as a Nash equilibrium of a stochastic game in the context of financial systemic risk. The drift in this system in contrast incarnates the structure of a complete graph.

Questions. What is the essential difference between the system (1.1) and (1.2) for large n? Can we detect the type of interaction from the single particle behavior at a vertex?

To answer these questions, let us fix $u \in [0, 1]$ and introduce a mixed system:

$$\begin{aligned}
\mathrm{d}X_{t,i} &= \Big(u \cdot h(X_{t,i}, X_{t,i+1}) + (1-u)\cdot \frac{1}{n}\sum_{j=1, j\neq i}^{n} h\big(X_{t,i}, X_{t,j}\big)\Big)\mathrm{d}t + \mathrm{d}W_{t,i}\,,\\
\mathrm{d}X_{t,n} &= \Big(u \cdot h(X_{t,n}, X_{t,1}) + (1-u)\cdot \frac{1}{n}\sum_{j=1, j\neq n}^{n} h\big(X_{t,n}, X_{t,j}\big)\Big)\mathrm{d}t + \mathrm{d}W_{t,n}
\end{aligned} \tag{1.3}$$

for $t \ge 0$, $i = 1, \dots, n-1$ with the initial random variables $X_{0,i}$, $1 \le i \le n$. If $u = 1$, (1.3) becomes (1.1), while if $u = 0$, (1.3) becomes (1.2).

The motivation of our study is to understand in a first instance the effect of the graph (network) structure on the stochastic system of interacting diffusions. Interacting diffusions have been studied in various contexts: nonlinear McKean–Vlasov equations, propagation of chaos results, large deviation results, stochastic control problems in large infinite particle systems, and their applications to Probability and Mathematical Physics, and more recently to Mathematical Economics and Finance in the context of the mean-field games. One of the advantages of introducing the mean-field dependence (1.2) and the corresponding limits, as $n \to \infty$, is to obtain a clear description of the complicated system, in terms of a representative particle, by the law of large numbers. As a result of the invariance under permutations of the indexes of particles, it often comes with the propagation of chaos, and then consequently the local dependence in the original system disappears in the limit. The single representative particle is characterized by a non-linear single equation, and the limiting distribution of many particles can be represented as a product measure. See Remark 3.2 in Section 3 for a short list of references and related research on propagation of chaos.

Here, in contrast, by breaking the invariance under permutation of particles, we consider the limit of the system (1.3) (or its slight generalization in the next section) as $n \to \infty$ and attempt to describe the presence of both, mean-field and local directed chain dependence in the interacting particles. In our directed chain dependence, conceptually there is a pair of representative particles in the limit: a particle (say $X_\cdot$) which corresponds to the *head* of an arrow and another particle (say $\widetilde{X}_\cdot$) which corresponds to the *tail* of the same arrow, i.e., the arrow directs from the particle $\widetilde{X}_\cdot$ to the particle $X_\cdot$. The marginal laws of $X_\cdot$ and $\widetilde{X}_\cdot$ are the same as a consequence of construction, and the dynamics of $X_\cdot$ is determined by its law, its position, the position of $\widetilde{X}_\cdot$ and a Brownian noise $B_\cdot$. As a result, our stochastic equation for the representative pair $(X_\cdot, \widetilde{X}_\cdot)$ is described in the limit by a weak solution to a single non-linear equation with constraints on the marginal law of particles (see (2.1)–(2.4)). The limiting distribution of a collection of particles is not necessarily a product measure, unless $u = 0$. When $u \in (0, 1]$, because of the local chain dependence, the single non-linear equation (2.1)–(2.2) with distributional constraints (2.3)–(2.4) has an infinite-dimensional nesting structure (see Remarks 2.4 and 3.1). Moreover, when $u \in (0, 1]$, essentially because of the violation of permutation invariance, the stochastic chaos does not propagate (see Remark 3.3). To our knowledge, our approach provides the first such instance in the context of particle approximation of the solution to a nonlinear stochastic equation of McKean–Vlasov type.

In Section 2 we discuss existence and uniqueness of the solution to a directed chain stochastic differential equation (2.1)–(2.2) for a representative pair $(X_\cdot, \widetilde{X}_\cdot)$ of interacting stochastic processes with distributional constraints (2.3)–(2.4). In Section 3 we propose a particle approximation of the solution to (2.1)–(2.2), we study the convergence of joint empirical measures (3.7) and an integral equation (3.11) with (3.12) for the limiting joint distribution in Propositions 3.1–3.2. Moreover, we provide a simple fluctuation estimate in Proposition 3.3. We will see that the joint law of adjacent two particles in the limit of interacting particle systems of type (1.3), as $n \to \infty$, can be described by the solution of the directed chain stochastic equation (2.1)–(2.2) under some assumptions. In Section 4, coming back to the above questions, we discuss the detection of the mean-field interaction as a filtering problem along with the systems of equations of Zakai and Kushner–Stratonovich type in Propositions 4.3–4.4. Then, we describe a connection to the infinite-dimensional Ornstein–Uhlenbeck process, and consequently, examine the corresponding Gaussian processes under presence or absence of the mean-field interaction in Section 4.2. The appendix includes some more technical proofs.

2. Directed chain stochastic equation with mean-field interaction

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, given a constant $u \in [0, 1]$ and a measurable functional $b : [0, \infty) \times \mathbb{R} \times \mathcal{M}(\mathbb{R}) \to \mathbb{R}$, let us consider a non-linear diffusion pair $(X_t^{(u)}, \widetilde{X}_t^{(u)})$, $t \ge 0$, described by the stochastic differential equation

$$\mathrm{d}X_t^{(u)} = b(t, X_t^{(u)}, F_t^{(u)})\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0\,, \tag{2.1}$$

driven by a Brownian motion $(B_t, t \ge 0)$, where $F_\cdot^{(u)}$ is the weighted probability measure

$$F_t^{(u)}(\cdot) := u \cdot \delta_{\widetilde{X}_t^{(u)}}(\cdot) + (1-u) \cdot \mathcal{L}_{X_t^{(u)}}(\cdot) \tag{2.2}$$

of the Dirac measure $\delta_{\widetilde{X}_t^{(u)}}(\cdot)$ of $\widetilde{X}_t^{(u)}$ and the law $\mathcal{L}_{X_t^{(u)}} = \mathrm{Law}(X_t^{(u)})$ of $X_t^{(u)}$ with corresponding weights $(u, 1-u)$ for $t \ge 0$. We shall assume that the law of $X_\cdot^{(u)}$ is identical to that of $\widetilde{X}^{(u)}_{\cdot}$, and $\widetilde{X}^{(u)}_{\cdot}$ is independent of the Brownian motion, i.e.,

$$\mathrm{Law}((X^{(u)}_t,\, t \ge 0)) \equiv \mathrm{Law}((\widetilde{X}^{(u)}_t,\, t \ge 0)) \quad \text{and} \quad \boldsymbol{\sigma}(\widetilde{X}^{(u)}_t,\, t \ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(B_t\,,\, t \ge 0)\,. \tag{2.3}$$

Let us also assume that the Brownian motion $B_{\cdot}$ is independent of the initial value $(X^{(u)}_0, \widetilde{X}^{(u)}_0)$. We assume the joint and marginal initial distributions of $(X^{(u)}_0, \widetilde{X}^{(u)}_0)$ are given and denoted by

$$\begin{aligned} \Theta &:= \mathrm{Law}(X^{(u)}_0, \widetilde{X}^{(u)}_0) = \mathrm{Law}(X^{(u)}_0) \otimes \mathrm{Law}(\widetilde{X}^{(u)}_0) = \theta^{\otimes 2}\,, \\ \theta &:= \mathrm{Law}(X^{(u)}_0) \equiv \mathrm{Law}(\widetilde{X}^{(u)}_0)\,. \end{aligned} \tag{2.4}$$

Here we assume $\widetilde{X}^{(u)}_{\cdot}$ is a copy of $X^{(u)}_{\cdot}$ which has the same law (2.3) as a random element in the space of continuous functions, however it is not necessarily independent of $X^{(u)}_{\cdot}$. They can be independent when $u = 0$, as in Remark 2.2. Rather, we are interested in the joint law of the pair $(X^{(u)}_{\cdot}, \widetilde{X}^{(u)}_{\cdot})$ which satisfies (2.1) and is generated from Brownian motion(s) in a non-linear way through their probability law for each $u \in [0, 1]$. The description (2.1) with the constraints (2.2)–(2.3) has an infinite-dimensional feature, because of non-trivial dependence between the unknown continuous processes $\widetilde{X}^{(u)}_{\cdot}$ and $X^{(u)}_{\cdot}$ in the space of continuous functions for every $u \in (0, 1]$. For a precise description of the infinite-dimensional nesting structure, see Remark 2.4 .

When $u \in (0, 1)$ we shall call (2.1) with (2.2)–(2.4) a *nonlinear, directed chain stochastic equation with mean-field interaction*. Let us denote by $\mathcal{M}(\mathbb{R})$ (and $\mathcal{M}(C([0, T], \mathbb{R}))$, respectively) the family of probability measures on $\mathbb{R}$ (and the space $C([0, T], \mathbb{R})$ of continuous functions equipped with the uniform topology on compact sets, respectively). Our following existence and uniqueness result relies on some standard assumptions to simplify the presentation.

Proposition 2.1. *Suppose that $b : [0, \infty) \times \mathbb{R} \times \mathcal{M}(\mathbb{R}) \to \mathbb{R}$ is Lipschitz, in the sense that there exists a measurable function $\widetilde{b} : [0, \infty) \times \mathbb{R} \times \mathbb{R}$ such that b is represented as*

$$b(t, x, \mu) = \int_{\mathbb{R}} \widetilde{b}(t, x, y)\mu(\mathrm{d}y)\,; \quad t \in [0, \infty)\,,\ x \in \mathbb{R}\,,\ \mu \in \mathcal{M}(\mathbb{R})\,, \tag{2.5}$$

and for every $T > 0$ there exists a constant $C_T > 0$ such that

$$|\widetilde{b}(t, x_1, y_1) - \widetilde{b}(t, x_2, y_2)| \le C_T(|x_1 - x_2| + |y_1 - y_2|)\,; \quad 0 \le t \le T\,. \tag{2.6}$$

With the same constant C_T, let us also assume that $\widetilde{b}$ is of linear growth, i.e.,

$$\sup_{0 \le s \le T} |\widetilde{b}(s, x, y)| \le C_T(1 + |x| + |y|)\,; \quad x, y \in \mathbb{R}\,. \tag{2.7}$$

Then, for each $u \in [0, 1]$ there exists a weak solution $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, $(X^{(u)}_{\cdot}, \widetilde{X}^{(u)}_{\cdot}, B_{\cdot})$ to the stochastic equation (2.1) with (2.2)–(2.4). This solution is unique in law.

Proof. First, observe that it is reduced to the well-known existence and uniqueness results of McKean–Vlasov equation, when $u = 0$. In particular, because of (2.3), in this case the joint distribution of $(X^{(u)}_{\cdot}, \widetilde{X}^{(u)}_{\cdot})$ is a product measure. Thus let us fix $u \in (0, 1]$ in the following, and also assume boundedness of the drift coefficients for the moment, i.e.,

$$|\widetilde{b}(t, x_1, y_1) - \widetilde{b}(t, x_1, y_2)| \le C_T((|x_1 - x_2| + |y_1 - y_2|) \wedge 1)\,; \quad t \ge 0\,, \tag{2.8}$$

in order to simplify our proof. We shall evaluate the Wasserstein distance $D_T(\mu_1, \mu_2)$ between two probability measures μ_1 and μ_2 on the space $C([0, T], \mathbb{R})$ of continuous functions, namely

$$D_t(\mu_1, \mu_2) := \inf\Big\{ \int (\sup_{0 \le s \le t} |X_s(\omega_1) - X_s(\omega_2)| \wedge 1) \mathrm{d}\mu(\omega_1, \omega_2) \Big\} \tag{2.9}$$

for $0 \le t \le T$, where the infimum is taken over all the joint distributions $\mu \in \mathcal{M}(C([0, T], \mathbb{R}) \times C([0, T], \mathbb{R}))$ such that their marginal distributions are μ_1 and μ_2, respectively, and the initial joint and marginal distributions are Θ and θ in (2.4), that is, $\mu|_{\{t=0\}} = \Theta$, $\mu_i|_{\{t=0\}} = \theta$,

$$\begin{aligned} &\mathrm{Law}(X_s(\omega_i), 0 \le s \le T) = \mu_i \quad \text{for } i = 1, 2 \\ &\quad \text{and } \mathrm{Law}(X_s(\omega_1), X_s(\omega_2), 0 \le s \le T) = \mu . \end{aligned}$$

Here $X_s(\omega) = \omega(s)$, $0 \le s \le T$ is the coordinate map of $\omega \in C([0, T], \mathbb{R})$. $D_T(\cdot, \cdot)$ defines a complete metric on $\mathcal{M}(C([0, T], \mathbb{R}))$, which gives the topology of weak convergence to it.

Given a probability measure $\mathrm{m} \in \mathcal{M}(C([0, T], \mathbb{R}))$ with initial law $\mathrm{m}_0 := \theta$ in (2.4) and the canonical process $\widetilde{X}^{\mathrm{m}}_\cdot$ of the law m with initial value $\widetilde{X}^{\mathrm{m}}_0 := \widetilde{X}^{(u)}_0$, and the initial variables $(X^{(u)}_0, \widetilde{X}^{(u)}_0)$ from (2.4), let us consider a map $\Phi : \mathcal{M}(C([0, T], \mathbb{R})) \mapsto \mathcal{M}(C([0, T], \mathbb{R}))$ such that

$$\Phi(\mathrm{m}) := \mathrm{Law}(X^{\mathrm{m}}_t, 0 \le t \le T) , \tag{2.10}$$

where on a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t \ge 0}$, given a fixed Brownian motion $B_\cdot$ on it, $X^{\mathrm{m}}_\cdot$ is defined from a solution $(X^{\mathrm{m}}_\cdot, \widetilde{X}^{\mathrm{m}}_\cdot)$ of the stochastic differential equation

$$\mathrm{d}X^{\mathrm{m}}_t = b(t, X^{\mathrm{m}}_t, u\, \delta_{\widetilde{X}^{\mathrm{m}}_t} + (1 - u)\mathrm{m}_t)\mathrm{d}t + \mathrm{d}B_t ; \quad 0 \le t \le T , \tag{2.11}$$

with the initial values $(X^{\mathrm{m}}_0, \widetilde{X}^{\mathrm{m}}_0) = (X^{(u)}_0, \widetilde{X}^{(u)}_0)$. That is, under the probability measure $\mathbb{P}$, $X^{\mathrm{m}}_\cdot$ is an $(\mathcal{F}_t)$-adapted process and the associated $(\mathcal{F}_t)$-adapted process $\widetilde{X}^{\mathrm{m}}_\cdot$ has the law

$$\mathrm{m} = \mathrm{Law}(\widetilde{X}^{\mathrm{m}}_t, 0 \le t \le T) \quad \text{with} \quad \mathrm{Law}(X^{\mathrm{m}}_0) = \mathrm{Law}(\widetilde{X}^{\mathrm{m}}_0) = \theta .$$

Here m_t in (2.11) is the marginal distribution of $\widetilde{X}^{\mathrm{m}}_t$ at time $t \ge 0$. Assume $B_\cdot$ is independent of the σ-field $\boldsymbol{\sigma}(\widetilde{X}^{\mathrm{m}}_t, 0 \le t \le T) \vee \boldsymbol{\sigma}(X^{\mathrm{m}}_0)$.

Thanks to the theory (e.g., [17]) of stochastic differential equation with Lipschitz condition (2.6) and the growth condition (2.7), a solution $X^{\mathrm{m}}_\cdot$ of (2.11) exists, given the probability measure $\mathrm{m} \in \mathcal{M}(C([0, T], \mathbb{R}))$, the initial values with the initial law (2.4) and the associated canonical process $\widetilde{X}^{\mathrm{m}}_\cdot$ of the law m. Hence, the map Φ is defined. Indeed, the solution $X^{\mathrm{m}}_\cdot$ in (2.11) can be given as a functional of m, $\widetilde{X}^{\mathrm{m}}_\cdot$ and $B_\cdot$, i.e., there exists a functional $\mathbf{\Phi} : [0, T] \times \mathcal{M}(C([0, T], \mathbb{R})) \times C([0, T], \mathbb{R}) \times C([0, T], \mathbb{R}) \times \mathbb{R} \to \mathbb{R}$ such that

$$X^{\mathrm{m}}_t = \mathbf{\Phi}(t, (\mathrm{m}_\cdot), (\widetilde{X}^{\mathrm{m}}_\cdot), (B_\cdot), X^{(u)}_0) ; \quad 0 \le t \le T , \tag{2.12}$$

where the value X^{m}_t at t is determined by the initial value $X^{\mathrm{m}}_0 = X^{(u)}_0$ with the law θ and the restrictions $(\mathrm{m}_s)_{0 \le s \le t}$, $(\widetilde{X}^{\mathrm{m}}_s)_{0 \le s \le t}$, $(B_s)_{0 \le s \le t}$ of elements on $[0, t]$ for $0 \le t \le T$. Note that here the filtration generated by $\widetilde{X}^{\mathrm{m}}_\cdot$ is not the Brownian filtration $(\mathcal{F}^B_t)_{t \ge 0}$ generated by the fixed Brownian motion $B_\cdot$ but we assume it is independent of $(\mathcal{F}^B_t)_{t \ge 0}$. Thus, we cannot expect the solution pair $(X^{\mathrm{m}}_\cdot, \widetilde{X}^{\mathrm{m}}_\cdot)$ to be a strong solution with respect to the filtration $(\mathcal{F}^B_t)_{t \ge 0}$, in general.

We shall find a fixed point m^* of this map Φ in (2.10), i.e., $\Phi(m^*) = m^*$ to show the uniqueness of solution to (2.1) with (2.2)–(2.3) in the sense of probability law.

For $\mathrm{m}_i \in \mathcal{M}(C([0,T],\mathbb{R}))$ with the initial law $\mathrm{m}_i|_{\{t=0\}} = \theta$, on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\ge 0}$, a fixed Brownian motion $B_{\cdot}$ on it, the initial values $(X_0^{(u)}, \widetilde{X}_0^{(u)})$ with the joint law Θ in (2.4), and the canonical process $\widetilde{X}_{\cdot}^{\mathrm{m}_i}$ with the initial value $\widetilde{X}_0^{\mathrm{m}_i} := \widetilde{X}_0^{(u)}$, let us consider $\Phi(\mathrm{m}_i) = \mathrm{Law}(X_t^{\mathrm{m}_i}, 0 \le t \le T)$ in (2.10), where $(X_{\cdot}^{\mathrm{m}_i}, \widetilde{X}_{\cdot}^{\mathrm{m}_i})$ satisfies $\mathrm{m}_i = \mathrm{Law}(\widetilde{X}_t^{\mathrm{m}_i}, 0 \le t \le T)$ and

$$X_t^{\mathrm{m}_i} = X_0^{(u)} + \int_0^t b(s, X_s^{\mathrm{m}_i}, u\,\delta_{\widetilde{X}_s^{\mathrm{m}_i}} + (1-u)\mathrm{m}_{i,s})\mathrm{d}t + \mathrm{d}B_t \,; \quad 0 \le t \le T\,, \quad i = 1,2\,.$$

Then, by the form (2.5) of b with the Lipschitz property (2.6) and the standard technique (see e.g., [27]) we obtain the estimates

$$\begin{aligned} |X_s^{\mathrm{m}_1} - X_s^{\mathrm{m}_2}| &\le \int_0^s |b(v, X_v^{\mathrm{m}_1}, u\delta_{\widetilde{X}_v^{\mathrm{m}_1}} + (1-u)\mathrm{m}_{1,v}) \\ &\qquad - b(v, X_v^{\mathrm{m}_2}, u\delta_{\widetilde{X}_v^{\mathrm{m}_2}} + (1-u)\mathrm{m}_{2,v})|\mathrm{d}v \\ &= \int_0^s \Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_1}, y)(u\delta_{\widetilde{X}_v^{\mathrm{m}_1}}(\mathrm{d}y) + (1-u)\mathrm{m}_{1,v}(\mathrm{d}y)) \\ &\qquad - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_2}, y)(u\delta_{\widetilde{X}_v^{\mathrm{m}_2}}(\mathrm{d}y) + (1-u)\mathrm{m}_{2,v}(\mathrm{d}y))\Big|\mathrm{d}v \\ &\le u \int_0^s \Big|\widetilde{b}(v, X_v^{\mathrm{m}_1}, \widetilde{X}_v^{\mathrm{m}_1}) - \widetilde{b}(v, X_v^{\mathrm{m}_2}, \widetilde{X}_v^{\mathrm{m}_2})\Big|\mathrm{d}v \\ &\qquad + (1-u)\int_0^s \Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_1}, y)\mathrm{m}_{1,v}(\mathrm{d}y) \\ &\qquad - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_2}, y)\mathrm{m}_{2,v}(\mathrm{d}y)\Big|\mathrm{d}v\,, \end{aligned} \tag{2.13}$$

where we evaluate the convex combination of the first term

$$\int_0^s \Big|\widetilde{b}(v, X_v^{\mathrm{m}_1}, \widetilde{X}_v^{\mathrm{m}_1}) - \widetilde{b}(v, X_v^{\mathrm{m}_2}, \widetilde{X}_v^{\mathrm{m}_2})\Big|\mathrm{d}v \le C_T\Big(\int_0^s ((|X_v^{\mathrm{m}_1} - X_v^{\mathrm{m}_2}| + |\widetilde{X}_v^{\mathrm{m}_1} - \widetilde{X}_v^{\mathrm{m}_2}|)\wedge 1)\mathrm{d}v\Big) \tag{2.14}$$

for every $0 \le s \le T$, and the second term with the integrand

$$\begin{aligned} &\Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_1}, y)\mathrm{m}_{1,v}(\mathrm{d}y) - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_2}, y)\mathrm{m}_{2,v}(\mathrm{d}y)\Big| \\ &\le \Big| \int_{\mathbb{R}} (\widetilde{b}(v, X_v^{\mathrm{m}_1}, y) - \widetilde{b}(v, X_v^{\mathrm{m}_2}, y))\mathrm{m}_{1,v}(\mathrm{d}y)\Big| \\ &\qquad + \Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_2}, z)\mathrm{m}_{1,v}(\mathrm{d}z) - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{\mathrm{m}_2}, y)\mathrm{m}_{2,v}(\mathrm{d}y)\Big| \\ &\le C_T(|X_v^{\mathrm{m}_1} - X_v^{\mathrm{m}_2}| \wedge 1) + C_T\, D_v(\mathrm{m}_1, \mathrm{m}_2)\,, \end{aligned} \tag{2.15}$$

where $D_v(\mathrm{m}_1, \mathrm{m}_2)$ is the Wasserstein distance in (2.9) between m_1 and m_2 in $[0, v]$ for $0 \le v \le T$. Here note that in the last equality of (2.15), we used (2.8) and an almost-sure inequality

$$\Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{m_2}, z) m_{1,v}(dz) - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{m_2}, y) m_{2,v}(dy) \Big|$$
$$= \Big| \mathbb{E}^{1,2}\big[\widetilde{b}(v, x, \omega_1) - \widetilde{b}(v, x, \omega_2)\big]\big|_{\{x = X_v^{m_2}\}} \Big| \le C_T \mathbb{E}^{1,2}\big[|\omega_{1,v} - \omega_{2,v}| \wedge 1\big],$$

where $\mathbb{E}^{1,2}$ is an expectation under a joint distribution of $(\omega_{1,v}, \omega_{2,v})$ (the value of $(\omega_1, \omega_2) \in \Omega_2 := C([0,T], \mathbb{R}^2)$ at time v) with fixed marginals $m_{1,v}$ and $m_{2,v}$ for every $0 \le v \le T$. Here, since the expectation on the left of $\le$ only depends on the marginals, taking the infimum on the right of $\le$ over all the joint distributions with fixed marginals $m_{1,v}$ and $m_{2,v}$, we obtained the last inequality in (2.15) from

$$\Big| \int_{\mathbb{R}} \widetilde{b}(v, X_v^{m_2}, z) m_{1,v}(dz) - \int_{\mathbb{R}} \widetilde{b}(v, X_v^{m_2}, y) m_{2,v}(dy) \Big| \le C_T D_v(m_1, m_2) ; \quad 0 \le v \le T .$$

Combining (2.13)–(2.15) and taking the supremum over $s \in [0,t]$, we obtain

$$\begin{aligned}
\sup_{0 \le s \le t} |X_s^{m_1} - X_s^{m_2}| \wedge 1 &\le C_T \int_0^t (|X_v^{m_1} - X_v^{m_2}| \wedge 1) dv \\
&\quad + C_T \int_0^t (u(|\widetilde{X}_v^{m_1} - \widetilde{X}_v^{m_2}| \wedge 1) + (1-u) D_v(m_1, m_2)) dv \\
&\le C_T \int_0^t (\sup_{0 \le s \le v} |X_s^{m_1} - X_s^{m_2}| \wedge 1) dv \\
&\quad + C_T \int_0^t (u(|\widetilde{X}_v^{m_1} - \widetilde{X}_v^{m_2}| \wedge 1) + (1-u) D_v(m_1, m_2)) dv
\end{aligned}$$

for every $0 \le t \le T$. Applying Gronwall's lemma, we obtain

$$\begin{aligned}
\sup_{0 \le s \le t} |X_s^{m_1} - X_s^{m_2}| \wedge 1 &\le C_T e^{C_T T} \int_0^t (u(|\widetilde{X}_v^{m_1} - \widetilde{X}_v^{m_2}| \wedge 1) + (1-u) D_v(m_1, m_2)) dv \\
&\le C_T e^{C_T T} \int_0^t (u(\sup_{0 \le s \le v} |\widetilde{X}_s^{m_1} - \widetilde{X}_s^{m_2}| \wedge 1) + (1-u) D_v(m_1, m_2)) dv
\end{aligned}$$

for every $0 \le t \le T$. Taking expectations of both sides and taking the infimum over all the joint measures with marginals (m_1, m_2) and initial law Θ in (2.4), we obtain

$$\begin{aligned}
D_t(\Phi(m_1), \Phi(m_2)) &\le C_T e^{C_T T} \int_0^t (u D_v(m_1, m_2) + (1-u) D_v(m_1, m_2)) dv \\
&= C_T e^{C_T T} \int_0^t D_v(m_1, m_2) dv
\end{aligned} \tag{2.16}$$

for every $0 \le t \le T$. Note that the upper bound in (2.16) is uniform over $u \in [0,1]$.

For every $m \in C([0,T], \mathbb{R})$ with initial marginal law $m|_{\{t=0\}} = \theta$, iterating (2.16) and the map Φ, k times, we observe the inequality

$$D_T(\Phi^{(k+1)}(m), \Phi^{(k)}(m)) \le \frac{(C_T T e^{C_T T})^k}{k!} \cdot D_T(\Phi(m), m) ; \quad k \in \mathbb{N}_0 , \tag{2.17}$$

and hence, we claim $\{\Phi^{(k)}(m), k \in \mathbb{N}_0\}$ forms a Cauchy sequence converging to a fixed point $m^* = \Phi(m^*)$ of Φ on $\mathcal{M}(C([0,T], \mathbb{R}))$. This fixed point $m^*(\cdot) = \mathbb{P}(X_\cdot \in \cdot)$ is a weak solution to (2.1) with (2.2)–(2.3). It is unique in the sense of probability distribution. To relax the condition (2.8) and to show the result under the weaker condition (2.6), we divide the time interval $[0, T]$ into time-intervals of short length and establish the uniqueness in the short time intervals, and then piece the unique solution together to get the global uniqueness by the standard method. □

Proposition 2.2. *In addition to the assumptions required in Proposition 2.1, let* $\mathbb{E}[|X_0|] < \infty$. *Then, the solution* $(X_\cdot, \widetilde{X}_\cdot)$, *given in Proposition 2.1, satisfies for every* $T > 0$

$$\mathbb{E}[\sup_{0\le t\le T} |X_t|] \le \big(\mathbb{E}[|X_0|] + \mathbb{E}[\sup_{0\le s\le T} |B_s|] + C_T T\big) e^{2C_T T} . \tag{2.18}$$

Proof. Suppose that $(X_\cdot, \widetilde{X}_\cdot)$ is the solution to (2.1) with (2.2)–(2.3) for a fixed $u \in [0, 1]$. Thanks to (2.7) and $\text{Law}(X_t) = \text{Law}(\widetilde{X}_t)$, $t \ge 0$, we have

$$\begin{aligned}|b(s, X_s, F_s^{(u)})| &= \Big| \int_{\mathbb{R}} \widetilde{b}(s, X_s, y) \mathrm{d}F_s^{(u)}(y) \Big| \\ &\le C_T \big(1 + u(|X_s| + |\widetilde{X}_s|) + (1-u)(|X_s| + \mathbb{E}[|X_s|])\big)\end{aligned}$$

for $0 \le s \le T$. Then we verify (2.18) by an application of Gronwall's lemma to

$$\begin{aligned}\mathbb{E}[\sup_{0\le s\le t} |X_s|] &\le \mathbb{E}[|X_0|] + \mathbb{E}[\sup_{0\le s\le t} |B_s|] \\ &\quad + C_T \mathbb{E}\Big[\int_0^t (1 + u(|X_s| + |\widetilde{X}_s|) + (1-u)(|X_s| + \mathbb{E}[|X_s|]))\mathrm{d}s\Big] \\ &\le \mathbb{E}[|X_0|] + \mathbb{E}[\sup_{0\le s\le t} |B_s|] + C_T T + \int_0^t 2C_T \mathbb{E}[\sup_{0\le u\le s} |X_u|]\mathrm{d}s ; \\ &\quad 0 \le t \le T. \quad \square\end{aligned}$$

Remark 2.1 (L^p *Estimates*)**.** We may extend Proposition 2.2 for the estimates of $\mathbb{E}[\sup_{0\le t\le T} |X_t|^p]$, assuming $\mathbb{E}[|X_0|^p] < +\infty$ for $p \ge 1$. □

Remark 2.2 (*Extreme Cases*)**.** In Proposition 2.1 the processes $(X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)})$, $u \in [0, 1]$ form a class of diffusions which contains two *extreme* cases $u = 0, 1$:

- When $u = 0$, we set $(X_\cdot^\bullet, \widetilde{X}_\cdot^\bullet) := (X_\cdot^{(0)}, \widetilde{X}_\cdot^{(0)})$ and distinguish it from other cases. $X_\cdot^\bullet$ satisfies a McKean–Vlasov diffusion equation

$$\mathrm{d}X_t^\bullet = b(t, X_t^\bullet, \mathcal{L}_{X_t^\bullet})\,\mathrm{d}t + \mathrm{d}B_t ; \quad t \ge 0 , \tag{2.19}$$

and the corresponding copy $\widetilde{X}_\cdot^\bullet$ does not appear, that is, we may take $\widetilde{X}_\cdot^\bullet$ independent of $X_\cdot^\bullet$ because of the solvability of (2.19) and the restriction (2.3).

- When $u = 1$, we set $(X_\cdot^\dagger, \widetilde{X}_\cdot^\dagger) := (X_\cdot^{(1)}, \widetilde{X}_\cdot^{(1)})$. The pair satisfies a stochastic equation

$$\mathrm{d}X_t^\dagger = b(t, X_t^\dagger, \delta_{\widetilde{X}_t^\dagger})\,\mathrm{d}t + \mathrm{d}B_t ; \quad t \ge 0 , \tag{2.20}$$

where $\widetilde{X}_\cdot^\dagger$ has the same law as $X_\cdot^\dagger$, independent of Brownian motion, i.e., $\text{Law}(X_\cdot^\dagger) = \text{Law}(\widetilde{X}_\cdot^\dagger)$ and $\boldsymbol{\sigma}(\widetilde{X}_t^\dagger, t \ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(B_t, t \ge 0)$. The corresponding non-linear contribution from the law $\text{Law}(X_\cdot^\dagger)$ of $X_\cdot^\dagger$ disappears from (2.20). □

Remark 2.3 (*Non-uniqueness*)**.** When $u \in (0, 1]$, it is simple to observe that the stochastic equation (2.1) with (2.2) but *without* the distributional constraints in (2.3) does not determine uniquely the joint law of $(X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)})$. For example, take a two-dimensional Brownian motion

$(B_\cdot, W_\cdot)$ and take $W_\cdot$ for $\widetilde{X}^{(u)}_\cdot$, i.e., $\widetilde{X}^{(u)}_\cdot \equiv W_\cdot$, then by the standard theory of stochastic differential equations we may construct a weak solution $(X^{(u)}_\cdot, \widetilde{X}^{(u)}_\cdot)$ for (2.1) with (2.2), in addition to the solution in Proposition 2.1. In this case, if $B_\cdot$ and $W_\cdot$ are independent, then the independence condition $\boldsymbol{\sigma}(\widetilde{X}^{(u)}_t \equiv W_t, t \ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(B_t, t \ge 0)$ holds but $\mathrm{Law}(X^{(u)}_\cdot) \neq \mathrm{Law}(\widetilde{X}^{(u)}_\cdot)$, in general. Thus the requirement (2.3) is crucial for the uniqueness of the solution. A recent work of Bayraktar et al. [4] introduces a pair of continuous stochastic processes coupled through the distribution of initial values without distributional constraints (2.3) for the study of randomized dynamic programming principle. □

Remark 2.4 (*Russian Nesting Doll Structure*)**.** When $u \in (0, 1]$, since $\widetilde{X}^{(u)}_\cdot$ has the same law as $X^{(u)}_\cdot$ in (2.3), the dynamics of $\widetilde{X}_\cdot$ is described by a similar equation as in (2.1), i.e.,

$$\mathrm{d}\widetilde{X}^{(u)}_t = b(t, \widetilde{X}^{(u)}_t, \widetilde{F}^{(u)}_t)\mathrm{d}t + \mathrm{d}\widetilde{B}_t\,; \quad t \ge 0\,, \tag{2.21}$$

where $\widetilde{B}$ is another Brownian motion but $\widetilde{F}^{(u)}_t$ is defined from another (unknown) copy $\widehat{X}_\cdot$ of $X_\cdot$,

$$\widetilde{F}^{(u)}_t = u \cdot \delta_{\widehat{X}^{(u)}_t} + (1-u) \cdot \mathcal{L}_{\widehat{X}^{(u)}_t}\,; \quad t \ge 0\,, \tag{2.22}$$

with $\mathrm{Law}(X^{(u)}_\cdot) = \mathrm{Law}(\widetilde{X}^{(u)}_\cdot) = \mathrm{Law}(\widehat{X}^{(u)}_\cdot)$, and $\boldsymbol{\sigma}(\widehat{X}^{(u)}_t, t \ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(\widetilde{\widetilde{B}}_t, t \ge 0)$. Thus it follows from Proposition 2.1 that the dynamics of $\widetilde{X}^{(u)}_\cdot$ depends on $\widehat{X}^{(u)}_\cdot$ and $\widetilde{\widetilde{B}}_\cdot$.

Repeating this argument, we see that the dynamics of $\widehat{X}^{(u)}_\cdot$ may depend on yet another copy and a Brownian motion, and then another copy and a Brownian motion, and so on. This dependence continues, and thus the dynamics of $X^{(u)}_\cdot$ may depend on the dynamics of infinitely many copies, as if we open infinitely many layers of Russian nesting doll “matryoshka”. Thus when $u \in (0, 1]$, the infinite-dimensional nesting structure naturally arises in the system (2.1)–(2.2). □

Remark 2.5 (*Generalization and Application*)**.** The set-up and conditions on the tamed drift function b in (2.1) can certainly be generalized and relaxed. Also, a Lipschitz continuous diffusion coefficient can be introduced in (2.1), instead of the unit diffusion coefficient. For such generalization in the McKean–Vlasov equation, see e.g., [12]. In a more realistic problem of large network objects (financial networks associated with blockchains, biological networks, neural networks, data networks etc.), it is of interest to analyze a more complicated infinite (random) *tree* structures rather than the simple local interaction of the infinite directed chains considered here. With these generalizations, it may also be natural to replace the current state space $\mathbb{R}$ of each particle by a locally compact, separable metric space E. Here, we take the simplest form (2.1)–(2.4) for the presentation of the essential idea of the infinite directed chain interaction. It can be seen as the sparse counterpart of a complete graph (as arising in the mean field setting) among the set of connected graphs. In the setup of unimodular Galton–Watson trees and related large sparse (but undirected) networks, we refer the reader to the interesting work of Lacker et al. [19] which came out after this work had been completed.

An interesting application of such generalized models in financial markets is modeling of stochastic volatility structures among financial asset price processes, that is, each $X_\cdot$ is a volatility process of a financial asset, so that the volatility processes of the financial assets have both a network structure and a mean-field interaction. The local network structure in this case could tie together companies from similar industries, with similar investments or operating under the same jurisdiction. Similar in the study of systemic risk, particle systems with coupled diffusions generated by sparse network structures are of particular importance.

Another interesting direction of research is to identify and explore the directed chain stochastic equations (2.1)–(2.4) or their variants, as Nash equilibria of stochastic games, where the representative pair of players interact optimally in the presence of both mean-field and network structure. This program was introduced for the mean-field games in [6] and substantial work has followed in the context of mean-field games and systemic risk analysis. The corresponding problem on chain-like networks is the object of a manuscript under preparation. □

3. Particle system approximation

We interpret the solution pair $(X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)})$ in Proposition 2.1 as a representative pair in the limits of the directed chain particle system (1.3) we introduced in Section 1, as $n \to \infty$. We view $X_\cdot^{(u)}$ as a particle which corresponds to the *head* of an arrow and $\widetilde{X}_\cdot^{(u)}$ as another particle which corresponds to the *tail* of the same arrow. Here u represents the strength of the directed chain dependence, comparative to the mean-field interaction. In this section we shall discuss this interpretation precisely by showing the limiting results in Propositions 3.1–3.2, as an extension from the stochastic chaos of Kac [15] (or propagation of chaos) towards a local dependence structure with mean-field interaction, and then discuss a fluctuation estimate in Proposition 3.3.

Let us consider a sequence of finite systems of particles $(X_{t,i}^{(u)},\ t \ge 0,\ i = 1, \ldots, n)$, $n \in \mathbb{N}$ defined by the system of stochastic differential equations

$$\mathrm{d}X_{t,i}^{(u)} = b\big(t, X_{t,i}^{(u)}, \widehat{F}_{t,i}^{(u)}\big)\mathrm{d}t + \mathrm{d}W_{t,i}\,; \quad t \ge 0, \quad i = 1, \ldots, n-1, \tag{3.1}$$

where $b : [0, \infty) \times \mathbb{R} \times \mathcal{M}(\mathbb{R}) \to \mathbb{R}$ is defined in (2.5) with the same assumptions (2.6)–(2.7) as in Proposition 2.1,

$$\widehat{F}_{t,i}^{(u)}(\cdot) := u \cdot \delta_{X_{t,i+1}^{(u)}}(\cdot) + (1-u) \cdot \frac{1}{n} \sum_{j=1}^{n} \delta_{X_{t,j}^{(u)}}(\cdot), \quad i = 1, \ldots, n-1$$

with the boundary particle

$$\mathrm{d}X_{t,n}^{(u)} = b\Big(t, X_{t,n}^{(u)}, u \cdot \delta_{X_{t,1}^{(u)}} + (1-u) \cdot \frac{1}{n} \sum_{j=1}^{n} \delta_{X_{t,j}^{(u)}}\Big)\mathrm{d}t + \mathrm{d}W_{t,n}\,. \tag{3.2}$$

Here $W_{\cdot,i}$, $i \in \mathbb{N}$ are standard independent Brownian motions on a filtered probability space, independent of the initial values $X_{0,i}^{(u)}$, $i = 1, \ldots, n$ and of $B_\cdot$ in (2.1). We assume the distribution of $X_{0,i}$ is *common* with $\mathbb{E}[|X_{0,1}|^2] < +\infty$ for $i = 1, \ldots, n$ and *independent* of each other.

Thanks to the assumption on b, the resulting particle system (3.1)–(3.2) is well-defined, and in particular, we have the law invariance $\mathrm{Law}(X_{\cdot,i}^{(u)}) = \mathrm{Law}(X_{\cdot,1}^{(u)})$, $i = 1, \ldots, n$,

$$\mathrm{Law}(X_{\cdot,i}^{(u)}, X_{\cdot,i+1}^{(u)}) = \mathrm{Law}(X_{\cdot,1}^{(u)}, X_{\cdot,2}^{(u)})\,; \quad i = 1 \ldots, n-1, \tag{3.3}$$

and more generally, the invariance under the shifts in one direction, i.e., for every fixed $k \le n-1$,

$$\mathrm{Law}(X_{\cdot,i}^{(u)}, X_{\cdot,i+1}^{(u)}, \ldots, X_{\cdot,i+k-1}^{(u)}) = \mathrm{Law}(X_{\cdot,1}^{(u)}, X_{\cdot,2}^{(u)}, \ldots, X_{\cdot,k}^{(u)})\,; \quad i = 1, \ldots, n-k+1\,. \tag{3.4}$$

Thus, it is natural to write $X_{\cdot,n+j}^{(u)} \equiv X_{\cdot,j}^{(u)}$, $j = 1, 2, \ldots$, so that (3.1) and (3.3)–(3.4) hold for $i = 1, \ldots, n$. The system (1.3) in Section 1 is a time-homogeneous special case of (3.1).

Under the setup of Proposition 2.2 we shall also consider a sequence of finite particle systems $\overline{X}_{t,i}$, $t \geq 0$, $i = 1, \ldots, n+1$, $n \geq 1$, defined recursively from the pair $(\overline{X}_{\cdot,n}, \overline{X}_{\cdot,n+1}) := (X^{(u)}_{\cdot}, \widetilde{X}^{(u)}_{\cdot})$ of the solution to (2.1) with (2.2)–(2.4), that is, the corresponding stochastic equation

$$\mathrm{d}\overline{X}_{t,n} = b(t, \overline{X}_{t,n}, u \cdot \delta_{\overline{X}_{t,n+1}} + (1-u) \cdot \mathcal{L}_{\overline{X}_{t,n}})\mathrm{d}t + \mathrm{d}W_{t,n}\,; \quad t \geq 0\,, \tag{3.5}$$

and then for $j = n-1, n-2, \ldots, 1$, given $\overline{X}_{\cdot,j+1}$, we solve

$$\mathrm{d}\overline{X}_{t,j} = b(t, \overline{X}_{t,j}, u \cdot \delta_{\overline{X}_{t,j+1}} + (1-u) \cdot \mathcal{L}_{\overline{X}_{t,j}})\mathrm{d}t + \mathrm{d}W_{t,j}\,; \quad t \geq 0 \tag{3.6}$$

with the restrictions for each pair $(\overline{X}_{\cdot,j}, \overline{X}_{\cdot,j+1})$, corresponding to (2.3). As a consequence of the proof of Proposition 2.2, we set the common law $\mathrm{m}^* = \mathrm{Law}(\overline{X}_{\cdot,i})$ for $i = 1, \ldots, n+1$, and we also assume the initial values are the same as $X^{(u)}_{0,i} = \overline{X}_{0,i}$, $i = 1, \ldots, n$ almost surely. Note that when $u = 0$, the particle system $\overline{X}_{t,i}$, $i = 1, \ldots, n+1$ induces a product measure; When $u \in (0, 1]$, the particle system forms a Russian doll nesting structure (see Remark 2.4 after the proof of Proposition 2.1)

For $n \geq 1$ with $X^{(u)}_{\cdot,n+1} \equiv X^{(u)}_{\cdot,1}$ let us assign the weight $1/n$ to the Dirac measure at $(X^{(u)}_{t,i}, X^{(u)}_{t,i+1})$ for $i = 1, \ldots, n$, and consider the law of the joint empirical measure process

$$\mathrm{M}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{(X^{(u)}_{t,i}, X^{(u)}_{t,i+1})}\,, \qquad \text{with the marginal} \qquad \mathrm{m}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{X^{(u)}_{t,i}}\,, \quad 0 \leq t \leq T\,, \tag{3.7}$$

in the space $\mathcal{M}(\Omega_2)$ of probability measures on the topological space $\Omega_2 := D([0, T], (\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1))$ of càdlàg functions on $[0, T]$ equipped with the Skorokhod topology, where $(\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1)$ is the space of probability measures on $\mathbb{R}^2$ equipped with the metric $\|\mu - \nu\|_1 := \sup_f \int_{\mathbb{R}^2} f(x)\mathrm{d}(\mu - \nu)(x)$. Here the supremum is taken over the bounded Lipschitz functions $f : \mathbb{R}^2 \to \mathbb{R}$ with $\sup_{x\in\mathbb{R}^2}|f(x)| \leq 1$ and $\sup_{x,y\in\mathbb{R}^2}|f(x) - f(y)|/\|x - y\| \leq 1$. By the construction the sequence of the law of the initial empirical measure converges to the Dirac measure concentrated in M_0 (say), i.e.,

$$\mathrm{Law}(\mathrm{M}_{0,n}) \xrightarrow[n\to\infty]{} \delta_{\mathrm{M}_0} \quad \text{weakly in} \quad \mathcal{M}((\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1))\,. \tag{3.8}$$

We denote by $\mathrm{m}_0(\mathrm{d}y) := \mathrm{M}_0(\mathbb{R} \times \mathrm{d}y) = \mathrm{M}_0(\mathrm{d}y \times \mathbb{R})$ the marginal of $\mathrm{M}_0 = \mathrm{m}_0 \otimes \mathrm{m}_0$.

Proposition 3.1. *Fix $u \in [0, 1]$. Under the same assumptions for the functional b as in Proposition 2.2, the law of joint empirical measure process $\mathrm{M}_{\cdot,n}$, defined in (3.7), of the finite particle system (3.1) with $X^{(u)}_{\cdot,n+1} \equiv X^{(u)}_{\cdot,1}$ converges in $\mathcal{M}(\Omega_2)$ to the Dirac measure concentrated in the deterministic measure-valued process $\mathrm{M}_t, 0 \leq t \leq T$, as $n \to \infty$, i.e.,*

$$\lim_{n\to\infty} Law\big(\mathrm{M}_{t,n}, 0 \leq t \leq T\big) = \delta_{(\mathrm{M}_t, 0 \leq t \leq T)} \quad in \quad \mathcal{M}(\Omega_2)\,. \tag{3.9}$$

The marginal laws of $\mathrm{M}_{\cdot}$ are the same, i.e.,

$$\mathrm{M}_t(\mathbb{R} \times \mathrm{d}y) = \mathrm{M}_t(\mathrm{d}y \times \mathbb{R}) =: \mathrm{m}_t(\mathrm{d}y)\,; \quad 0 \leq t \leq T\,, \tag{3.10}$$

and the joint $\mathrm{M}_{\cdot}$ *and its marginal* $\mathrm{m}_{\cdot}$ *satisfy the integral equation*

$$\int_{\mathbb{R}} g(x)\mathrm{m}_t(\mathrm{d}x) = \int_{\mathbb{R}} g(x)\mathrm{m}_0(\mathrm{d}x) + \int_0^t [\mathcal{A}_s(\mathrm{M})g]\,\mathrm{d}s\,; \quad 0 \le t \le T \tag{3.11}$$

for every test function $g \in C_c^2(\mathbb{R})$, *where*

$$\begin{aligned}\mathcal{A}_s(\mathrm{M})g := {} & u \int_{\mathbb{R}^2} \widetilde{b}(s, y_1, y_2)g'(y_1)\mathrm{M}_s(\mathrm{d}y_1\mathrm{d}y_2) \\ & + (1-u)\int_{\mathbb{R}^2} \widetilde{b}(s, y_1, y_2)g'(y_1)\mathrm{m}_s(\mathrm{d}y_1)\mathrm{m}_s(\mathrm{d}y_2) \\ & + \frac{1}{2}\int_{\mathbb{R}} g''(y_1)\mathrm{m}_s(\mathrm{d}y_1)\,; \quad 0 \le s \le T\,.\end{aligned} \tag{3.12}$$

Moreover, $\mathrm{M}_{\cdot}$ *is the joint distribution of the solution pair* $(X_{\cdot}^{(u)}, \widetilde{X}_{\cdot}^{(u)})$ *of* (2.1) *with* (2.2)–(2.4), *uniquely characterized by* (3.10)–(3.12) *in* Ω_1 *with the common marginal* $\mathrm{m}_{\cdot} = \mathit{Law}(X_{\cdot}) = \mathit{Law}(\widetilde{X}_{\cdot})$ *in Proposition* 2.1.

Remark 3.1. • When $u = 0$, the integral equation (3.11) for $\mathrm{M}_{\cdot}$ reduces to the McKean–Vlasov nonlinear integral equation only for the marginal $\mathrm{m}_{\cdot}$, i.e., for $0 \le t \le T$ and $g \in C_c^2(\mathbb{R})$

$$\begin{aligned}\int_{\mathbb{R}} g(x)\mathrm{m}_t(\mathrm{d}x) = {} & \int_{\mathbb{R}} g(x)\mathrm{m}_0(\mathrm{d}x) + \int_0^t \mathrm{d}s\Big[\int_{\mathbb{R}^2} \widetilde{b}(s, y_1, y_2)g'(y_1)\mathrm{m}_s(\mathrm{d}y_1)\mathrm{m}_s(\mathrm{d}y_2) \\ & + \frac{1}{2}\int_{\mathbb{R}} g''(y_1)\mathrm{m}_s(\mathrm{d}y_1)\Big].\end{aligned}$$

• When $u \in (0, 1]$, the integral equation (3.11) has an infinite-dimensional feature, because of marginal distributional constraints (3.10), as we discussed in Remark 2.4, i.e., the joint distribution $\mathrm{M}_{\cdot}$ appears in the infinitesimal generator (3.12) for the marginal distribution. □

Proof. The idea of the proof utilizes the assumptions on the coefficient b as in Proposition 2.2 and the law invariance (3.3) of the finite particle system (3.1). We take the martingale approach discussed in [25]. By the standard argument with Gronwall's lemma we claim

Lemma 3.1. (*a*) *For the joint empirical measure processes* $\mathrm{M}_{\cdot,n}$ *and its marginal* $\mathrm{m}_{\cdot,n}$ *there exist constants* $c_k(>0)$, $k = 1, \dots, 4$ *such that*

$$e^{-c_1 t}\int_{\mathbb{R}} |x|^2 \mathrm{dm}_{t,n}(x) - c_2 t\,, 0 \le t \le T$$

$$\Big(e^{c_3 t}\int_{\mathbb{R}} |x|^2 \mathrm{dm}_{t,n}(x) + c_4 t\,, 0 \le t \le T, \textit{respectively}\Big)$$

is a supermartingale (submartingale, respectively), and hence, so is

$$e^{-c_1 t}\int_{\mathbb{R}^2} \|y\|^2 \mathrm{dM}_{t,n}(y) - 2c_2 t\,, 0 \le t \le T$$

$$\Big(e^{c_3 t}\int_{\mathbb{R}^2} \|y\|^2 \mathrm{dM}_{t,n}(y) + 2c_4 t\,, 0 \le t \le T, \ \textit{respectively}\Big),$$

because $\sum_{i=1}^n |X_{\cdot,i}^{(u)}|^2 = \sum_{i=1}^n |X_{\cdot,i+1}^{(u)}|^2 = (1/2)\sum_{i=1}^n (|X_{\cdot,i}^{(u)}|^2 + |X_{\cdot,i+1}^{(u)}|^2)$.

(*b*) *Moreover, there exist constants* c_k , $k = 5, 6$, *such that for every* $t \le s \le T$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[2b(s, X^{(u)}_{s,i}, \widehat{F}^{(u)}_{t,i})(X_{s,i} - X_{t,i})|\mathcal{F}_t\big] \le c_5 \int_{\mathbb{R}} |x|^2 \mathrm{d}m_{t,n}(x) + c_6 \,.$$

Using this lemma and the Cauchy–Schwarz inequality, we claim that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|X^{(u)}_{t+\delta,i} - X^{(u)}_{t,i}||\mathcal{F}_t] &\le \Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[|X^{(u)}_{t+\delta,i} - X^{(u)}_{t,i}|^2|\mathcal{F}_t]\Big)^{1/2} \\ &= \Big(\int_t^{t+\delta}\frac{1}{n}\sum_{i=1}^{n}\Big(\mathbb{E}\Big[2(b, s, X_{s,i}, \widehat{F}^{(u)}_s)(X_{s,i} - X_{t,i})\Big|\mathcal{F}_t\Big] \\ &\quad + 1\Big)\mathrm{d}s\Big)^{1/2} \\ &\le \Big(c_5\int_{\mathbb{R}}|x|^2\mathrm{d}m_{t,n}(x) + c_6 + 1\Big)^{1/2}\sqrt{\delta}\end{aligned}$$

for $0 \le t \le T - \delta$. Thus, using these inequalities again with the supermartingale property, we claim that there exists an $\mathcal{F}_T$ -measurable random variable $\mathfrak{f}(\delta)$, such that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(|X^{(u)}_{t+\delta,i} - X^{(u)}_{t,i}|^2 + |X^{(u)}_{t+\delta,i+1} - X^{(u)}_{t,i+1}|^2)^{1/2}|\mathcal{F}_t] \le \mathbb{E}\big[\,\mathfrak{f}(\delta)\,|\,\mathcal{F}_t\big]; \quad 0 \le t \le T - \delta\,, \tag{3.13}$$

with $\lim_{\delta\to 0}\sup_{n\ge 1}\mathbb{E}[\mathfrak{f}(\delta)] = 0$. Here we set $X^{(u)}_{\cdot,n+1} \equiv X^{(u)}_{\cdot,1}$.

Moreover, by the super/submartingale properties in Lemma 3.1(a) we may evaluate the total variation $\|\mathrm{M}_{t,n}|_{B^c_\lambda}\|_{TV}$ of $\mathrm{M}_{\cdot,n}$ restricted outside the ball $B_\lambda := \{x \in \mathbb{R}^2 : \|x\| \le \lambda\}$ of radius $\lambda(> 0)$, i.e., for every $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}(\sup_{0\le t\le T}\|\mathrm{M}_{t,n}|_{B^c_\lambda}\|_{TV} > \varepsilon) &\le \mathbb{P}\Big(\sup_{0\le t\le T}\int_{\mathbb{R}^2}\|y\|^2\mathrm{d}\mathrm{M}_{t,n} > \lambda^2\varepsilon\Big) \\ &\le \mathbb{P}\Big(\sup_{0\le t\le T} e^{c_1 t}\Big(\int_{\mathbb{R}^2}\|y\|^2\mathrm{d}\mathrm{M}_{t,n} + 2c_2 t\Big) > \lambda^2\varepsilon\Big) \\ &\le \frac{1}{\lambda^2\varepsilon}\mathbb{E}\Big[e^{c_1 T}\Big(\int_{\mathbb{R}^2}\|y\|^2\mathrm{d}\mathrm{M}_{T,n} + 2c_2 T\Big)\Big] \\ &\le \frac{1}{\lambda^2\varepsilon}\mathbb{E}\Big[\Big(\int_{\mathbb{R}^2}\|y\|^2\mathrm{d}\mathrm{M}_{0,n} + 2c_4 T\Big)e^{(c_1+c_3)T} + 2c_2 T\Big].\end{aligned}$$

Taking sufficiently large λ, using Prohorov's theorem, we claim that $(\mathrm{M}_{t,n}, \; 0 \le t \le T)$, $n \ge 1$ of the empirical measures is tight in $(\mathcal{M}(\mathbb{R}^2), |\cdot|_1)$. Then combining this observation with (3.13), we claim by Theorem 8.6 (b) of Ethier and Kurtz [10] that the sequence $(\mathrm{M}_{t,n}$, $0 \le t \le T)$, $n \ge 1$ is relatively compact in the space $\mathcal{M}(\Omega_2)$, where $\mathcal{M}(\Omega_2)$ is equipped with the weak topology.

We shall characterize the limit points of $(\mathrm{M}_{t,n}, 0 \le t \le T)_{n\ge 1}$ as $n \to \infty$. Let us call a limit law $\mathrm{M}_t, 0 \le t \le T$. Thanks to the law invariance in the construction of (3.1), its marginals must be the same for every limit point, i.e., $\mathrm{M}_t(\mathbb{R}\times \mathrm{d}y) = \mathrm{M}_t(\mathrm{d}y \times \mathbb{R}) =: \mathrm{m}_t(\mathrm{d}y)$, $y \in \mathbb{R}$ with the initial marginal measure $\mathrm{m}_0(\mathrm{d}y)$. Applying Itô's formula to the system (3.1), we see

$$f(\langle \mathrm{m}_{t,n}, g\rangle) - f(\langle \mathrm{m}_{0,n}, g\rangle) - \int_0^t f'(\langle \mathrm{m}_{s,n}, g\rangle)$$
$$\times \Big(\frac{1}{n} \sum_{i=1}^{n} g'(X^{(u)}_{s,i}) b(s, X^{(u)}_{s,i}, \widehat{F}^{(u)}_{s,i}) + \frac{1}{2} \langle \mathrm{m}_{s,n}, g''\rangle \Big) \mathrm{d}s$$
$$- \frac{1}{2} \int_0^t f''(\langle \mathrm{m}_{s,n}, g\rangle) \frac{1}{n^2} \sum_{i=1}^{n} |g'(X^{(u)}_{s,i})|^2 \mathrm{d}s = \int_0^t f'(\langle \mathrm{m}_{s,n}, g\rangle) \frac{1}{n} \sum_{i=1}^{n} g'(X^{(u)}_{s,i}) \mathrm{d}W_{s,i} ,$$

is a martingale for every $f \in C^2_b(\mathbb{R})$, $g \in C^2_c(\mathbb{R})$, where we use the notation $\langle \mu, g\rangle := \int_{\mathbb{R}} g(x) \mathrm{d}\mu(x)$ for $\mu \in \mathcal{M}(\mathbb{R})$. Taking the limits with (3.8) and using the equivalence of certain martingales, we observe that $\exp(\sqrt{-1}\,\theta\, \eta_t)$, $0 \le t \le T$ is a martingale for every $\theta \in \mathbb{R}$, where we define

$$\eta_t := \int_{\mathbb{R}} g(x) \mathrm{d}\mathrm{m}_t(x) - \int_0^t [\mathcal{A}_s(\mathrm{M}_s) g] \mathrm{d}s$$

and $\mathcal{A}_t(\mathrm{M}_t) g$ as in (3.12) for $0 \le t \le T$. This implies that the characteristic function of η_t satisfies $\mathbb{E}[e^{\sqrt{-1}\theta\eta_t}] = \mathbb{E}[e^{\sqrt{-1}\theta\eta_0}] = e^{\sqrt{-1}\theta\langle \mathrm{m}_0, g\rangle}$ for $0 \le t \le T$, $\theta \in \mathbb{R}$, and hence, $\eta_t = \langle \mathrm{m}_0, g\rangle$ for every t in any countable subset of $[0, T]$ and for every g in any countable subset of $C^2_c(\mathbb{R})$. Because of the separability of $C^2_c(\mathbb{R})$ and right continuity of $t \mapsto \mathrm{M}_t$, we obtain

$$\int_{\mathbb{R}} g(x) \mathrm{d}\mathrm{m}_t(x) - \int_0^t [\mathcal{A}_s(\mathrm{M}_s) g] \mathrm{d}s = \eta_t = \langle \mathrm{m}_0, g\rangle = \int_{\mathbb{R}} g(x) \mathrm{d}\mathrm{m}_0(x)$$

for every $0 \le t \le T$ and $g \in C^2_c(\mathbb{R})$. Thus we claim $\mathrm{M}_\cdot$ satisfies the integral equation (3.11). With the uniqueness in Proposition 2.1 the last part of Proposition 3.1 can be shown as in Lemmas 8–10 of Oelschläger [25]. □

Proposition 3.1 describes the limiting system of (3.1)–(3.2), in terms of the joint distribution of two adjacent particles of the directed chain structure (2.1)–(2.2). Now let us fix $k (\ge 2)$ and define the empirical measure process $\mathrm{M}^{(j)}_{t,n}$, $t \ge 0$ of j consecutive particles from (3.1)–(3.2)

$$\mathrm{M}^{(j)}_{t,n} := \frac{1}{n} \sum_{i=1}^{n} \delta_{(X^{(u)}_{t,i}, \ldots, X^{(u)}_{t,i+j-1})} ; \quad j = 2, \ldots, k \tag{3.14}$$

with $\mathrm{M}^{(1)}_{t,n} := \mathrm{m}_{t,n}$ and $\mathrm{M}^{(2)}_{t,n} \equiv \mathrm{M}_{t,n}$ as in (3.7) in the space $\mathcal{M}(\Omega_j)$ of probability measures on the topological space $\Omega_j := D([0, T], (\mathcal{M}(\mathbb{R}^j), \|\cdot\|_1))$ of cádlág functions on $[0, T]$, equipped with the Skorokhod topology, where $(\mathcal{M}(\mathbb{R}^j), |\cdot|_1)$ is the space of probability measures on $\mathbb{R}^j$, a natural extension of $(\mathcal{M}(\mathbb{R}^2), |\cdot|_1)$ defined in the above for $j = 2, \ldots, k$. We shall consider their limits.

By the construction and the law of large numbers for the initial empirical measure, as in (3.8),

$$\mathrm{Law}(\mathrm{M}^{(k)}_{0,n}) \xrightarrow[n\to\infty]{} \delta_{\mathrm{M}^{(k)}_0} \quad \text{weakly in} \quad \mathcal{M}((\mathcal{M}(\mathbb{R}^k), \|\cdot\|_1)) , \tag{3.15}$$

where $\mathrm{M}^{(k)}_0 := \mathrm{m}_0^{\otimes k}$ is the k-tuple product measure of $\mathrm{m}_0 = \mathrm{Law}(X^{(u)}_{0,1})$. For $j = 1, \ldots, k+1$ let us denote by $\mathrm{M}^{(j)}_\cdot$ the joint probability measure induced by $(\overline{X}_{\cdot,1}, \ldots, \overline{X}_{\cdot,j})$ in (3.5)–(3.6), i.e.,

$$(\mathrm{M}^{(j)}_t, t \ge 0) := \mathrm{Law}((\overline{X}_{t,1}, \ldots, \overline{X}_{t,j}), t \ge 0) ; \quad j = 1, \ldots, k+1 . \tag{3.16}$$

Proposition 3.2. *Fix $u \in [0,1]$. Under the same assumptions as in Proposition 3.1, the law of the joint empirical measure process $\mathrm{M}^{(k)}_{\cdot,n}$ defined in (3.14) converges in $\mathcal{M}(\Omega_k)$ to the Dirac measure concentrated in the deterministic measure-valued process $\mathrm{M}^{(k)}_{\cdot}$ in (3.16), i.e.,*

$$\lim_{n\to\infty} Law(\mathrm{M}^{(k)}_{t,n}, 0 \le t \le T) = \delta_{(\mathrm{M}^{(k)}_t,\, 0\le t\le T)} \quad in \quad \mathcal{M}(\Omega_k)\,. \tag{3.17}$$

All the consecutive marginals of $\mathrm{M}^{(k)}_{\cdot}$ are the same, i.e.,

$$\begin{aligned} \mathrm{M}^{(j)}_t(\mathbb{R}\times \mathrm{d}y_1\times\cdots\times \mathrm{d}y_{j-1}) &= \mathrm{M}^{(j)}_t(\mathrm{d}y_1\times\cdots\times \mathrm{d}y_{j-1}\times\mathbb{R})\,,\\ \mathrm{M}^{(j)}_t(\mathbb{R}^2\times \mathrm{d}y_1\times\cdots\times \mathrm{d}y_{j-2}) &= \mathrm{M}^{(j)}_t(\mathbb{R}\times\mathrm{d}y_1\times\cdots\times \mathrm{d}y_{j-2}\times\mathbb{R})\\ &= \mathrm{M}^{(j)}_t(\mathrm{d}y_1\times\cdots\times \mathrm{d}y_{j-2}\times\mathbb{R}^2)\,,\\ \ldots,\quad \mathrm{M}^{(j)}_t(\mathbb{R}^{j-1}\times \mathrm{d}y_1) &= \mathrm{M}^{(j)}_t(\mathbb{R}^{j-2}\times\mathrm{d}y_1\times\mathbb{R}) = \ldots\\ &= \mathrm{M}^{(j)}_t(\mathrm{d}y_1\times\mathbb{R}^{j-1}) \end{aligned} \tag{3.18}$$

for $j = 2,\ldots,k$, $0\le t\le T$, and they also satisfy the system of integral equations

$$\begin{aligned} \int_{\mathbb{R}^j} g(x)\mathrm{M}^{(j)}_t(\mathrm{d}x) = \int_{\mathbb{R}^j} g(x)\mathrm{M}^{(j)}_0(\mathrm{d}x) + \int_0^t [\mathcal{A}^{(j)}_s(\mathrm{M}^{(j+1)})g]\mathrm{d}s\,;\\ 0\le t\le T\,,\ j = 1,\ldots,k \end{aligned} \tag{3.19}$$

for every test function $g\in C^2_c(\mathbb{R}^j)$, where

$$\begin{aligned} \mathcal{A}^{(j)}_s(\mathrm{M}^{(j+1)})g :=\ & u\int_{\mathbb{R}^{j+1}} \sum_{\ell=1}^{j} \widetilde{b}(s,y_\ell,y_{\ell+1})\frac{\partial g}{\partial x_\ell}(y_1,\ldots,y_j)\mathrm{M}^{(j+1)}_s(\mathrm{d}y_1\cdots\mathrm{d}y_{j+1})\\ & + (1-u)\int_{\mathbb{R}^{j+1}} \sum_{\ell=1}^{j} \widetilde{b}(s,y_\ell,y_{\ell+1})\frac{\partial g}{\partial x_\ell}(y_1,\ldots,y_j)\\ & \mathrm{M}^{(j)}_s(\mathrm{d}y_1\cdots\mathrm{d}y_j)\mathrm{m}_s(\mathrm{d}y_{j+1})\\ & + \frac{1}{2}\int_{\mathbb{R}^j} \sum_{\ell=1}^{j} \frac{\partial^2 g}{\partial x_\ell^2}(y_1,\ldots,y_j)\mathrm{M}^{(j)}_s(\mathrm{d}y_1\cdots\mathrm{d}y_j)\,;\\ & 0\le s\le T\,,\ j = 2,\ldots,k\,. \end{aligned} \tag{3.20}$$

Proof. We have shown (3.17) in the case $k = 2$ in Proposition 3.1. The relative compactness proof of $(\mathrm{M}^{(j)}_{t,n}, 0\le t\le T)$, $n\ge 1$ in $\mathcal{M}(\Omega_j)$ follows as in Proposition 3.1 *mutatis mutandis* for $j = 1,\ldots,k$. The limit points of $(M^{(k)}_{t,n}, 0\le t\le T)$ in (3.17) as $n\to\infty$ are characterized by (3.19), because for every test function $g\in C^2_c(\mathbb{R}^j)$, thanks to Itô's formula, it follows from

$$\begin{aligned} \mathrm{d}g(X^{(u)}_{t,i},\ldots,X^{(u)}_{t,i+j-1}) = & \sum_{\ell=1}^{j} \frac{\partial g}{\partial x_\ell}(X^{(u)}_{t,i},\ldots,X^{(u)}_{t,i+j-1})\mathrm{d}X^{(u)}_{t,\ell+i-1}\\ & + \frac{1}{2}\sum_{\ell=1}^{j}\sum_{m=1}^{j} \frac{\partial^2 g}{\partial x_\ell\partial x_m}(X^{(u)}_{t,i},\ldots,X^{(u)}_{t,i+j-1})\mathrm{d}\langle X^{(u)}_{\cdot,\ell+i-1}, X^{(u)}_{\cdot,m+i-1}\rangle_t \end{aligned}$$

that $\langle \mathrm{M}^{(j)}_{t,n}, g\rangle := \int_{\mathbb{R}^j} g(x)\mathrm{d}\mathrm{M}^{(j)}_{t,n}(\mathrm{d}x)$, $j = 1,\dots,k$ satisfy that for $0 \le t \le T$, $j = 1,\dots,k$,

$$f(\langle \mathrm{M}^{(j)}_{t,n}, g\rangle) - f(\langle \mathrm{M}^{(j)}_{0,n}, g\rangle) - \int_0^t f'(\langle \mathrm{M}^{(j)}_{s,n}, g\rangle)[\mathcal{A}^{(j)}\mathrm{M}^{(j+1)}_{\cdot,n} g]\mathrm{d}s$$
$$- \frac{1}{2n^2}\int_0^t f''(\langle \mathrm{M}^{(j)}_{s,n}, g\rangle)\sum_{i=1}^{n}\sum_{\ell=1}^{j}\Big(\frac{\partial g}{\partial x_\ell}\Big)^2 \mathrm{d}s = \int_0^t f'(\langle \mathrm{M}^{(j)}_{s,n}, g\rangle)\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{j}\Big(\frac{\partial g}{\partial x_\ell}\Big)\mathrm{d}W_{s,\ell},$$

for $f \in C^2(\mathbb{R})$, where $[\mathcal{A}^{(j)}_s \mathrm{M}^{(j+1)}_{\cdot,n} g]$ is defined as in (3.20). The condition (3.18) follows from the construction of (3.5)–(3.6). Applying the martingale argument again as in the proof of Proposition 3.1, we conclude the proof. □

Remark 3.2 (*Propagation of Chaos*). If $u = 0$ and if the initial law of $(X^{(u)}_0, \widetilde{X}^{(u)}_0)$ is a product measure, then $X^{(u)}_\cdot$ and $\widetilde{X}^{(u)}_\cdot$ in (2.1)–(2.2) are independent, as in Remark 2.2, and hence, the joint law of $(\overline{X}_{\cdot,1},\dots,\overline{X}_{\cdot,k})$ in (3.5)–(3.6) is the product measure. Thus in this case of $u = 0$, Proposition 3.2 corresponds to the classic propagation of chaos result (see Kac [15] for the original result for Boltzmann equation in Kinetic Theory, [13,14,21,22,26–28] for the advancement of theory for McKean–Vlasov and Boltzmann equations, [5,18,23,24] for recent developments of quantitative approach in propagation of chaos and references within them), where the limiting joint law takes the product form. This means the dependence between each particle $X^{(u)}_{\cdot,i}$ and another particle $X^{(u)}_{\cdot,j}$, $j \ne i$ diminishes in the limit, as $n \to \infty$. Chong and Klüppelberg [7] investigate linear partial mean field systems based on fairly general network structures in which both, propagation of chaos and local dependency arises jointly. □

Remark 3.3 (*Breaking Invariance Under Permutations*). When $u \in (0, 1]$, Proposition 3.2 implies that the local directed chain dependence among consecutive particles is preserved even in the limit as the number of particles go to infinity, in general. Thus if $u \in (0, 1]$, the limiting system (3.5)–(3.6) of (3.1)–(3.2) does not propagate the stochastic chaos, in contrast to the case $u = 0$.

This phenomenon can be seen as a consequence of breaking the invariance under permutations in the finite particle system (3.1)–(3.2), that is, the consecutive particles are invariant only under the shifts in one direction as in (3.3)–(3.4), and the finite particle system is not invariant under permutations, for example,

$$\mathrm{Law}(X^{(u)}_{\cdot,1}, X^{(u)}_{\cdot,2},\dots,X^{(u)}_{\cdot,n}) \ne \mathrm{Law}(X^{(u)}_{\cdot,n}, X^{(u)}_{\cdot,n-1},\dots,X^{(u)}_{\cdot,1})$$

unless $\widetilde{b}(\cdot,\cdot,\cdot) \equiv 0$. To our knowledge, our approach of breaking the invariance under permutations provides the first such instance of describing the dependence of the limiting system in the context of a particle system approximation to the solution of a nonlinear stochastic McKean–Vlasov equation. The simple case of a directed chain with its recursive structure sets itself apart from other network structures by allowing for a description by representative particles which solve a nonlinear McKean–Vlasov equation with distributional constraint. The analysis of this kind of "matryoshka" McKean–Vlasov equations might be of independent interest. □

Proposition 3.3. *In addition to the same assumptions for the functional* b *as in Proposition 2.2, we assume that the marginal distribution* $\mathrm{m}_t(\mathrm{d}y) = \mathrm{m}^*_t(\mathrm{d}y)$ *of* $(X^{(u)}_t, t \ge 0)$ *has the* *density* $m_t(\cdot)$ *(i.e.,* $\mathrm{m}_t(\mathrm{d}y) = m_t(y)\mathrm{d}y$, $y \in \mathbb{R}$*) with* $\int_{\mathbb{R}} |y|^2 \mathrm{m}_0(\mathrm{d}y) < \infty$ *and assume there exists a constant* C_T *such that*

$$\Big| \widetilde{b}(t, x_1, y_1) \cdot \frac{m_t(x_1)}{m_t(y_1)} - \widetilde{b}(t, x_2, y_2) \cdot \frac{m_t(x_2)}{m_t(y_2)} \Big| \le C_T(|x_1 - x_2| + |y_1 - y_2|) \tag{3.21}$$

for every $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, 2$, $0 \le t \le T$ *and*

$$\Big| \widetilde{b}(t, x, y) \cdot \frac{m_t(x)}{m_t(y)} \Big| \le C_T(1 + |x| + |y|) \tag{3.22}$$

for every $(x, y) \in \mathbb{R}^2$, $0 \le t \le T$. *Then for the difference between* (3.1)–(3.2) *and* (3.5)–(3.6) *we have the estimate*

$$\sup_{n \ge 1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}[\sup_{0 \le s \le t} |X_{s,i}^{(u)} - \overline{X}_{s,i}|] < \infty . \tag{3.23}$$

Proof. Substituting

$$\begin{aligned} \widetilde{b}(s, X_{s,i}^{(u)}, X_{s,j}^{(u)}) &= (\widetilde{b}(s, X_{s,i}^{(u)}, X_{s,j}^{(u)}) - \widetilde{b}(s, \overline{X}_{s,i}, X_{s,j}^{(u)})) \\ &\quad + (\widetilde{b}(s, \overline{X}_{s,i}, X_{s,j}^{(u)}) - \widetilde{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j})) + \widetilde{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j}) \end{aligned}$$

into the differences

$$\begin{aligned} X_{t,i}^{(u)} - \overline{X}_{t,i} &= u \int_0^t (\widetilde{b}(s, X_{s,i}^{(u)}, X_{s,i+1}^{(u)}) - \widetilde{b}(s, \overline{X}_{s,i}, \overline{X}_{s,i+1}))\mathrm{d}s \\ &\quad + (1-u) \int_0^t \Big(\frac{1}{n} \sum_{j=1}^{n} \widetilde{b}(s, X_{s,i}^{(u)}, X_{s,j}^{(u)}) - \int_{\mathbb{R}} \widetilde{b}(s, \overline{X}_{s,i}, y)\mathrm{m}^*(\mathrm{d}y) \Big)\mathrm{d}s \end{aligned} \tag{3.24}$$

for $i = 1, \ldots, n-1$, and the difference

$$\begin{aligned} X_{t,n}^{(u)} - \overline{X}_{t,n} &= u \int_0^t (\widetilde{b}(s, X_{s,n}^{(u)}, X_{s,1}^{(u)}) - \widetilde{b}(s, \overline{X}_{s,n}, \overline{X}_{s,n+1}))\mathrm{d}s \\ &\quad + (1-u) \int_0^t \Big(\frac{1}{n} \sum_{j=1}^{n} \widetilde{b}(s, X_{s,n}^{(u)}, X_{s,j}^{(u)}) - \int_{\mathbb{R}} \widetilde{b}(s, \overline{X}_{s,n}, y)\mathrm{m}^*(\mathrm{d}y) \Big)\mathrm{d}s \end{aligned} \tag{3.25}$$

at the boundary for $0 \le t \le T$, applying the triangle inequality and (2.6), and then taking the supremum, we obtain

$$\begin{aligned} &\sum_{i=1}^{n} \sup_{0 \le t \le T} |X_{t,i}^{(u)} - \overline{X}_{t,i}| \\ &\le 2C_T \int_0^T \sum_{i=1}^{n} \sup_{0 \le t \le s} |X_{t,i}^{(u)} - \overline{X}_{t,i}|\mathrm{d}s + 2C_T u \int_0^T (|X_{s,1}^{(u)} - \overline{X}_{s,n+1}| - |X_{s,1}^{(u)} - \overline{X}_{s,1}|)\mathrm{d}s \\ &\quad + (1-u) \int_0^T \frac{1}{n} \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} \overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j}) \Big| \mathrm{d}s \end{aligned} \tag{3.26}$$

$$\le 2C_T \int_0^T \sum_{i=1}^n \sup_{0\le t\le s} |X_{t,i}^{(u)} - \overline{X}_{t,i}| \mathrm{d}s + 2C_T u \int_0^T |\overline{X}_{s,n+1} - \overline{X}_{s,1}| \mathrm{d}s$$
$$+ (1-u)\int_0^T \frac{1}{n}\sum_{i=1}^n \Big| \sum_{j=1}^n \overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j})\Big| \mathrm{d}s \,,$$

where we set $\overline{b}(s,x,z) := \widetilde{b}(s,x,z) - \int_{\mathbb{R}} \widetilde{b}(s,x,y)\mathrm{m}^*(\mathrm{d}y)$ for $x, z \in \mathbb{R}$, $0 \le s \le T$. Here we used $|x| - |y| \le |x-y|$, $x, y \in \mathbb{R}$ in the last inequality, and this way we take care of the boundary particle. Note that $X_{\cdot,1}^{(u)} \equiv X_{\cdot,n+1}^{(u)}$ but $\overline{X}_{\cdot,1} \not\equiv \overline{X}_{\cdot,n+1}$.

After using Gronwall's lemma, taking expectation, we obtain

$$\sum_{i=1}^n \mathbb{E}[\sup_{0\le t\le T} |X_{t,i}^{(u)} - \overline{X}_{t,i}|] \le 2C_T e^{2C_T T}\mathbb{E}\Big[\int_0^T |\overline{X}_{s,n+1} - \overline{X}_{s,1}|\mathrm{d}s$$
$$+ \int_0^T \frac{1}{n}\sum_{i=1}^n \Big|\sum_{j=1}^n \overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,j})\Big|\mathrm{d}s\Big],$$

where there exists some constant $c > 0$ such that we evaluate the first term

$$\mathbb{E}\Big[\int_0^T |\overline{X}_{s,n+1} - \overline{X}_{s,1}|\mathrm{d}s\Big] \le \mathbb{E}\Big[\int_0^T (\sup_{0\le u\le T} |\overline{X}_{u,n+1}| + \sup_{0\le u\le T} |\overline{X}_{u,1}|)\mathrm{d}s\Big]$$
$$\le 2T(\mathbb{E}[|\overline{X}_{0,1}|] + c)e^{cT}\,, \tag{3.27}$$

by (2.18) in Proposition 2.2 and then with (3.21)–(3.22) we evaluate the second term

$$\sum_{i=1}^n \mathbb{E}\Big[\int_0^T \frac{1}{n}\Big|\sum_{j=1}^n \overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,j})\Big|\mathrm{d}s\Big] \le \sum_{i=1}^n \int_0^T \Big(\mathbb{E}\Big[\frac{1}{n^2}\Big|\sum_{j=1}^n \overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,j})\Big|^2\Big]\Big)^{1/2}\mathrm{d}s$$
$$\le c\sqrt{n} \tag{3.28}$$

by the Cauchy–Schwarz inequality and the (Markov) chain structure of the particle system $\overline{X}_{\cdot,i}$, $i = 1,\dots,n$, that is, by the map $\boldsymbol{\Phi}$ in (2.12), $\overline{X}_{\cdot,i} = \boldsymbol{\Phi}(\cdot, (\mathrm{m}_s^*)_{0\le s\le\cdot}, (\overline{X}_{s,i+1})_{0\le s\le\cdot}, (W_{s,i})_{0\le s\le\cdot})$ for $i = n-1,\dots,1$. Note that when $u \in (0,1]$, $\overline{X}_{\cdot,i}$ and $\overline{X}_{\cdot,j}$ are dependent for $i \ne j$, while $\overline{X}_{\cdot,i+1}$ and $W_{\cdot,i}$ are independent for $i = n-1,\dots,1$. An intuitive interpretation of the last inequality in (3.28) is that the dependence between $\overline{X}_{s,i}$ and $\overline{X}_{s,j}$ decays sufficiently fast, as $|i-j| \to \infty$. Its precise statement and some technical details are given in Appendix A.1.

Finally, combining these inequalities, we conclude the proof of (3.23) by

$$\sup_{n\ge1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[\sup_{0\le t\le T}|X_{t,i}^{(u)} - \overline{X}_{t,i}|] \le 2C_T e^{2C_T T}\sup_{n\ge1}\Big(\frac{2T}{\sqrt{n}}(\mathbb{E}[|\overline{X}_{0,1}|]+c)e^{cT}+c\Big) < \infty\,. \quad \square$$

Remark 3.4. The fluctuation results (central limit theorem and large deviations) suggested from Proposition 3.3 are ongoing research topics. We conjecture that Propositions 3.1–3.2 still hold if we replace (3.2) by another process, e.g., a standard Brownian motion, as long as the effect of the boundary process on the first two (or k) components in (3.1) diminishes sufficiently fast in the limit. The additional conditions (3.21)–(3.22) are used to evaluate the decay of asymptotic covariance between $\overline{X}_{\cdot,i}$ and $\overline{X}_{\cdot,i+n}$ as $n \to \infty$ (see Appendix A.1). In particular, the dependence between the first particle $\overline{X}_{\cdot,1}$ and the last particle $\overline{X}_{\cdot,n}$ of the directed chain diminishes in the limit. It is an ongoing project to see whether one may relax these conditions (3.21)–(3.22). $\square$

4. Detecting mean-field in the presence of directed chain interaction

In the weak solution $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, $(X_\cdot, \widetilde{X}_\cdot) := (X^{(u)}_\cdot, \widetilde{X}^{(u)}_\cdot)$, $B_\cdot$ from Proposition 2.1, the parametric value u in (2.1) indicates how much the particle $X_\cdot$ depends on the neighborhood particle $\widetilde{X}_\cdot$ in the directed chain, and $(1-u)$ indicates how much it depends on its law $\mathrm{Law}(X_t)$ for every $t \ge 0$. Let us consider the following detection problem of a single observer.

Detection Problem. Suppose that an observer only observes the single path X_t, $t \ge 0$ but does neither know the values $u \in [0, 1]$ nor $\widetilde{X}_t$, $t \ge 0$ in (2.1)–(2.4) under the same assumptions as Proposition 2.1. Only given the filtration $\mathcal{F}^X_t := \boldsymbol{\sigma}(X_s, 0 \le s \le t) \vee \mathcal{N}$, $t \ge 0$, augmented by the null sets $\mathcal{N}$, can the observer detect the value $u \in [0, 1]$ (and hence, the effect $1 - u$ of mean-field)?

In order to discuss this problem, it is natural to extend our consideration to the solution $(\overline{X}_{t,1}, \ldots, \overline{X}_{t,n+1})$, $t \ge 0$ of the system of the directed chain stochastic differential equations

$$\mathrm{d}\overline{X}_{t,i} = b(t, \overline{X}_{t,i}, F_{t,i})\,\mathrm{d}t + \mathrm{d}B_{t,i}\,; \quad i = 1, \ldots, n\,,\ t \ge 0\,, \tag{4.1}$$

as the solution to the system of the directed chain stochastic equations in (3.5)–(3.6) in Section 3, for arbitrary $n \in \mathbb{N}$, where $F_{s,i}$ is the random measure similar to (2.2), i.e.,

$$F_{s,i} := u \cdot \delta_{\overline{X}_{s,i+1}} + (1-u) \cdot \mathcal{L}_{\overline{X}_{s,i}}\,, \quad i = 1, \ldots, n\,,\ s \ge 0 \tag{4.2}$$

with the distributional constraints, such that the initial values $(\overline{X}_{0,1}, \ldots, \overline{X}_{0,n+1})$ are independently, identically distributed with finite second moments; The marginal law is identical

$$\mathrm{Law}(\{\overline{X}_{t,i}, t \ge 0\}) = \mathrm{Law}(\{\overline{X}_{t,1}, t \ge 0\})\,; \quad i = 1, 2, \ldots, n+1\,, \tag{4.3}$$

and the following independence relationships hold for independent standard Brownian motions $(B_{t,1}, \ldots, B_{t,n})$, $t \ge 0$

$$\boldsymbol{\sigma}(\{(\overline{X}_{t,n+1}, \ldots, \overline{X}_{t,i+1}), t \ge 0\}, \overline{X}_{0,i}) \perp\!\!\!\perp \boldsymbol{\sigma}(\{B_{t,i}, t \ge 0\})\,; \quad i = n, \ldots, 1\,. \tag{4.4}$$

The weak solution $(\overline{X}_{\cdot,1}, \ldots, \overline{X}_{\cdot,n+1})$ can be constructed as we considered in Section 3. Namely, we solve for $(\overline{X}_{\cdot,n}, \overline{X}_{\cdot,n+1})$ in (4.1) first as in Proposition 2.1 and then solve recursively for the directed chain system $(\overline{X}_{\cdot,k}, \overline{X}_{\cdot,k+1}, \ldots, \overline{X}_{\cdot,n+1})$ for $k = n, \ldots, 1$. We redefine

$$(X_\cdot, \widetilde{X}_\cdot) = (X^{(u)}_\cdot, \widetilde{X}^{(u)}_\cdot) := (\overline{X}_{\cdot,1}, \overline{X}_{\cdot,2}) \tag{4.5}$$

from the first two elements of $(\overline{X}_{\cdot,1}, \overline{X}_{\cdot,2}, \ldots, \overline{X}_{\cdot,n+1})$, and the observer only observes $X_\cdot = \overline{X}_{\cdot,1}$.

Remark 4.1 (*Variations of the Detection Problem*)**.** The setup of the detection problem would be different, if the observer observes the whole $(n+1)$ particles $(\overline{X}_{\cdot,1}, \ldots, \overline{X}_{\cdot,n+1})$ in (4.1) or if the observer observes the pre-limit system (1.3) of n particles. It would be interesting, yet out of scope of this current paper, to study the detection methods and to quantify the information gain/loss for different setups, and moreover to detect the presence of the propagation of chaos. These interesting open problems were suggested by the editors and the reviewers, while this paper was revised. □

Let us define the stochastic exponential

$$Z_t := \exp\Big(- \int_0^t b(s, X_s, F_s)\mathrm{d}B_s - \frac{1}{2} \int_0^t |b(s, X_s, F_s)|^2 \mathrm{d}s \Big)\,; \quad t \ge 0\,, \tag{4.6}$$

where $F_s := F_{s,1} = F_s^{(u)}$ from (2.2) and $B_\cdot = B_{\cdot,1}$, which satisfies $\mathrm{d}Z_t = -Z_t b(t, X_t, F_t)\,\mathrm{d}B_t$, $t \ge 0$. Here $\{Z_\cdot; \mathcal{F}_\cdot\}$ is a nonnegative, local martingale and hence, is a supermartingale with $\mathbb{E}[Z_t] \le 1$, $t \ge 0$.

If the Novikov condition (e.g., Corollary 3.5.13 of Karatzas and Shreve [17]) for $Z_\cdot$ holds, i.e.,

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_0^t |b(s, X_s, F_s)|^2 \mathrm{d}s\Big)\Big] < \infty\,; \quad t \ge 0\,, \tag{4.7}$$

then $Z_\cdot$ is a martingale. Since it is not always easy to verify the Novikov condition directly except for the Gaussian case (e.g., see Section 4.2) or for the bounded functional case (i.e., the functional $\widetilde{b}$ in (2.5) is bounded), we shall discuss the martingale property of $Z_\cdot$.

Let us assume the finite moment condition $\mathbb{E}[|X_0|^2] < \infty$ for the initial distribution θ in (2.4). Then under the linear growth condition (2.7) and this finite second moment condition, as in Proposition 2.2, we have $\mathbb{E}[\sup_{0\le t\le T} |X_t|^2] < +\infty$ (see Remark 2.1), and hence, combining with the inequalities $|b(s, X_s, F_s)| \le C_T(1 + |X_s| + u|\widetilde{X}_s| + (1-u)\mathbb{E}[|X_s|])$, $0 \le s \le T$ and $(a_1 + a_2 + a_3 + a_4)^2 \le 4(a_1^2 + a_2^2 + a_3^2 + a_4^2)$ for nonnegative reals $a_i \ge 0$, we obtain

$$\mathbb{E}\Big[\int_0^T |b(s, X_s, F_s)|^2 \mathrm{d}s\Big] \le 4C_T^2 T(1 + 3\mathbb{E}[\sup_{0\le s\le T} |X_s|^2]) < \infty\,. \tag{4.8}$$

Following the proof of Lemma 3.9 (and see also Exercise 3.11) of Bain and Crisan [2], in order to show $\mathbb{E}[Z_t] = 1$, $t \ge 0$, we consider for $\varepsilon > 0$,

$$\frac{Z_t}{1 + \varepsilon Z_t} = \frac{1}{1+\varepsilon} + \int_0^t \frac{Z_s b(s, X_s, F_s)}{(1+\varepsilon Z_s)^2}\mathrm{d}B_s - \int_0^t \frac{\varepsilon Z_s^2 |b(s, X_s, F_s)|^2}{(1+\varepsilon Z_s)^3}\mathrm{d}s\,; \quad t \ge 0\,, \tag{4.9}$$

and its expectation for $0 \le t \le T$

$$1 \ge \mathbb{E}[Z_t] \ge \mathbb{E}\Big[\frac{Z_t}{1+\varepsilon Z_t}\Big] = \frac{1}{1+\varepsilon} - \mathbb{E}\Big[\int_0^t \frac{\varepsilon Z_s^2 |b(s, X_s, F_s)|^2}{(1+\varepsilon Z_s)^3}\mathrm{d}s\Big] \tag{4.10}$$

where we used (4.8) to show that the stochastic integral in (4.9) is indeed a martingale and hence its expectation is zero. Thus, in order to verify $\mathbb{E}[Z_t] = 1$, $t \ge 0$, by letting $\varepsilon \downarrow 0$ in (4.10) and by the dominated convergence theorem, it suffices to check

$$\mathbb{E}\Big[\int_0^T Z_s |b(s, X_s, F_s)|^2 \mathrm{d}s\Big] < \infty\,; \quad T > 0\,. \tag{4.11}$$

Note that since $F_s = F_s^{(u)}$ in (2.2) depends on $\widetilde{X}_\cdot$, under the linear growth condition (2.7) on the functional b, the condition (4.11) is reduced to estimates for both

$$\mathbb{E}\Big[\int_0^T Z_s |X_s|^2 \mathrm{d}s\Big] < \infty \quad \text{and} \quad \mathbb{E}\Big[\int_0^T Z_s |\widetilde{X}_s|^2 \mathrm{d}s\Big] < \infty\,; \quad T > 0\,, \tag{4.12}$$

where the joint distribution of $(X_\cdot, Z_\cdot)$ is not the same as that of $(\widetilde{X}_\cdot, Z_\cdot)$.

Proposition 4.1. *In addition to the assumptions in Proposition 2.1, let us assume* $\mathbb{E}[|X_0|^2] < +\infty$. *Then the first inequality in (4.12) holds. Moreover, for* $i = 2, \dots, n$ *and for every* $T > 0$,

$$\mathbb{E}\Big[\int_0^T Z_s |X_{s,i+1}|^2 \mathrm{d}s\Big] < \infty \quad \textit{implies} \quad \mathbb{E}\Big[\int_0^T Z_s |X_{s,i}|^2 \mathrm{d}s\Big] < \infty\,,$$

where $(X_{\cdot,1}, \dots, X_{\cdot,n+1})$ *is defined from (4.1)–(4.4) with (4.5). In particular, if for every* $T > 0$, $\mathbb{E}[\int_0^T Z_s |X_{s,n+1}|^2 \mathrm{d}s] < \infty$, *then the second inequality in (4.12) holds.*

Proof. Under the assumptions in Proposition 2.1 with $\mathbb{E}[|X_0|^2] < \infty$, we consider for $\varepsilon > 0$

$$\mathrm{d}\Big(\frac{Z_t|X_t|^2}{1+\varepsilon Z_t|X_t|^2}\Big) = \frac{\mathrm{d}(Z_t|X_t|^2)}{(1+\varepsilon Z_t|X_t|^2)^2} - \frac{\varepsilon\,\mathrm{d}\langle Z.|X.|^2\rangle_t}{(1+\varepsilon Z_t|X_t|^2)^3}\,, \tag{4.13}$$

where $\mathrm{d}(Z_t|X_t|^2) = Z_t\mathrm{d}t + (2Z_tX_t - Z_t|X_t|^2)\mathrm{d}B_t$, $t \ge 0$. Taking the expectations, we claim

$$\mathbb{E}\Big[\frac{Z_t|X_t|^2}{1+\varepsilon Z_t|X_t|^2}\Big] \le \mathbb{E}\Big[\frac{|X_0|^2}{1+\varepsilon|X_0|^2}\Big] + \mathbb{E}\Big[\int_0^t \frac{Z_s}{(1+\varepsilon Z_s|X_s|^2)^2}\mathrm{d}s\Big]$$
$$\le \mathbb{E}[|X_0|^2] + \int_0^t \mathbb{E}[Z_s]\mathrm{d}s\,; \quad t \ge 0\,.$$

Here the stochastic integrals with respect to the Brownian motion in (4.13) are indeed martingales, as in Exercise 3.11 of Bain and Crisan [2]. Note also that b disappears in the evaluation. Since $\mathbb{E}[Z.] \le 1$, by letting $\varepsilon \downarrow 0$, we obtain the first inequality in (4.12) from $\mathbb{E}[Z_t|X_t|^2] \le \mathbb{E}[|X_0|^2] + t$.

For the second assertions, we replace $Z_t|X_t|^2/(1+\varepsilon|X_t|^2)$ in (4.13) by $Z_t|X_{t,i}|^2/(1+\varepsilon|X_{t,i}|^2)$, $t \ge 0$ for $i = 2, 3, \dots, n$. Thanks to (2.7) and $\mathbb{E}[Z.] \le 1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[\frac{Z_t|X_{t,i}|^2}{1+\varepsilon Z_t|X_{t,i}|^2}\Big] \le 1 + 4C_T\Big(1 + \mathbb{E}[\sup_{0\le s\le t}|X_{s,i}|^2] + \mathbb{E}\Big[\frac{Z_t|X_{t,i}|^2}{1+\varepsilon Z_t|X_{t,i}|^2}\Big]$$
$$+ \mathbb{E}\Big[\frac{Z_t|X_{t,i+1}|^2}{1+\varepsilon Z_t|X_{t,i+1}|^2}\Big]\Big)\,.$$

As in Remark 2.1, we may derive the estimate $\mathbb{E}[\sup_{0\le s\le T}|X_s|^2] < \infty$. Then applying the Gronwall inequality, we obtain the estimate that there exists a constant $c > 0$ such that

$$\mathbb{E}\Big[\frac{Z_t|X_{t,i}|^2}{1+\varepsilon Z_t|X_{t,i}|^2}\Big] \le c + 4C_T\mathbb{E}\Big[\frac{Z_t|X_{t,i+1}|^2}{1+\varepsilon Z_t|X_{t,i+1}|^2}\Big]$$
$$+ \int_0^t\Big(c + 4C_T\mathbb{E}\Big[\frac{Z_s|X_{s,i+1}|^2}{1+\varepsilon Z_s|X_{s,i+1}|^2}\Big]\Big)e^{4C_TT}\mathrm{d}s\,.$$

Integrating over $[0, T]$ with respect to t and letting $\varepsilon \downarrow 0$, we claim the conclusions. □

Let us assume that (4.7) or (4.12) holds. Then the stochastic exponential $(Z_t, \mathcal{F}_t)$, $t \ge 0$ in (4.6) is martingale. By Girsanov theorem, under a new probability measure $\mathbb{P}_0$ with expectation $\mathbb{E}_0$, defined by

$$(\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P})|_{\mathcal{F}_T} := Z_T \tag{4.14}$$

for every $T > 0$, we have the Kallianpur–Striebel formula: $\mathbb{P}_0$ ($\mathbb{P}$)-a.s.

$$\pi_t(\varphi) := \mathbb{E}[\varphi(\widetilde{X}_t)|\mathcal{F}_t^X] = \frac{\rho_t(\varphi)}{\rho_t(\mathbf{1})}\,,$$
$$\text{where} \quad \rho_t(\varphi) := \mathbb{E}_0[Z_t^{-1}\varphi(\widetilde{X}_t)|\mathcal{F}_T^X]\,; \quad 0 \le t \le T \tag{4.15}$$

for measurable function $\varphi : \mathbb{R} \to \mathbb{R}$ with $\mathbb{E}[|\varphi(\widetilde{X}_t)|] < \infty$. Its proof is a direct consequence of Proposition 3.16 and Exercise 5.1 of Bain and Crisan [2].

Given the observation $\mathcal{F}_T^X$, the conditional log-likelihood function $\mathbb{E}[\log(\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{P}_0)|_{\mathcal{F}_T}|\mathcal{F}_T^X]$ is

$$\mathbb{E}\big[-\log Z_T|\mathcal{F}_T^X\big] = \mathbb{E}\Big[\int_0^T b(s, X_s, F_s)\mathrm{d}X_s - \frac{1}{2}\int_0^T |b(s, X_s, F_s)|^2\mathrm{d}s\Big|\mathcal{F}_T^X\Big]\,.$$

Substituting the expression $b(s, X_s, F_s) = u\widetilde{b}(s, X_s, \widetilde{X}_s) + (1-u)\int_{\mathbb{R}} \widetilde{b}(s, X_s, y)\mathrm{m}_s(\mathrm{d}y)$, $s \ge 0$, we see it is a quadratic function of u. Thus the conditional log-likelihood is maximized at the conditional maximum likelihood estimator $\widehat{u}_T$ defined by

$$\widehat{u}_T := \Big[\mathbb{E}\Big[\int_0^T |\overline{b}(s, X_s, \widetilde{X}_s)|^2 \mathrm{d}s \Big| \mathcal{F}_T^X\Big]\Big]^{-1} \times \mathbb{E}\Big[\int_0^T \overline{b}(s, X_s, \widetilde{X}_s)\mathrm{d}_s\Big(X_s - \int_0^s \int_{\mathbb{R}} \widetilde{b}(u, X_u, y)\mathrm{m}_u(\mathrm{d}y)\mathrm{d}u\Big)\Big| \mathcal{F}_T^X\Big], \tag{4.16}$$

where $\overline{b}(s, x, z) := \widetilde{b}(s, x, z) - \int_{\mathbb{R}} \widetilde{b}(s, x, y)\mathrm{m}(\mathrm{d}y)$ is defined as in the proof of Proposition 3.3 for $x, z \in \mathbb{R}$, $0 \le s \le T$. The maximum likelihood estimator $\widehat{u}_T$ in (4.16) is well defined if the denominator is not zero, e.g., $\overline{b}(\cdot, \cdot, \cdot) \ne 0$.

The analysis of (4.16) is not straightforward due to the conditional expectation and the filtering feature. We shall discuss the filtering equations in Section 4.1 and then see the consistent estimators under the special linear case in Section 4.2. For the theory of parameter estimation in Stochastic Filtering, see e.g., chapter 17 of Liptser and Shiryayev [20].

4.1. Filtering equations

In the following let us assume under $\mathbb{P}_0$ defined in (4.14)

$$\mathbb{P}_0\Big(\int_0^t \big|\mathbb{E}_0\big[|b(s, X_s, F_s)| \,\big|\, \mathcal{F}_T^X\big]\big|^2 \mathrm{d}s < \infty\Big) = 1; \quad 0 \le t \le T. \tag{4.17}$$

Proposition 4.2. *Let us recall* (4.5) *and assume* (4.12) and (4.17). *For every* $\varphi \in C_0^2(\mathbb{R})$, *the conditional expectations* $\rho_\cdot(\varphi) = \mathbb{E}_0[Z_\cdot^{-1}\varphi(\widetilde{X}_\cdot)|\mathcal{F}_T^X]$ *in* (4.15) *satisfy*

$$\rho_t(\varphi) = \pi_0(\varphi) + \int_0^t \rho_{s,2}(\varphi b)\mathrm{d}X_s + \int_0^t \rho_{s,3}(\widetilde{\mathcal{A}}_s\varphi)\mathrm{d}s, \quad 0 \le t \le T, \tag{4.18}$$

where $\pi_0(\varphi) = \mathbb{E}[\varphi(\widetilde{X}_0)|\mathcal{F}_0^X] = \mathbb{E}[\varphi(\widetilde{X}_0)] = \mathbb{E}[\varphi(X_0)]$ *in* (4.15), $\rho_{s,2}(\varphi b)$ *and* $\rho_{s,3}(\widetilde{\mathcal{A}}_s\varphi)$ *are defined by*

$$\rho_{s,2}(\varphi b) := \mathbb{E}_0\big[Z_s^{-1}\varphi(\widetilde{X}_s)b(s, X_s, F_s)|\mathcal{F}_T^X\big], \tag{4.19}$$

$$\rho_{s,3}(\widetilde{\mathcal{A}}_s\varphi) := \mathbb{E}_0\Big[Z_s^{-1}\Big(\varphi'(\widetilde{X}_s)b(s, \widetilde{X}_s, F_{s,2}) + \frac{1}{2}\varphi''(\widetilde{X}_s)\Big)\Big|\mathcal{F}_T^X\Big], \quad 0 \le s \le T. \tag{4.20}$$

Here $F_\cdot = F_{\cdot,1}$ *and* $F_{\cdot,2}$ *are the random measures defined as in* (4.2) *from the law of* $X_\cdot = \overline{X}_{\cdot,1}$, $\widetilde{X}_\cdot = \overline{X}_{\cdot,2}$ *and* $\overline{X}_{\cdot,3}$, *in the solution* $(\overline{X}_{\cdot,1}, \overline{X}_{\cdot,2}, \overline{X}_{\cdot,3})$ *to the system* (4.1)–(4.2) *of the directed chain stochastic differential equation with the distributional constraints* (4.3)–(4.4).

Proof. The proof idea is a slight modification of Theorem 3.24 of Bain and Crisan [2]. For $\varphi \in C_0^2(\mathbb{R})$ let us take the semimartingale decomposition $\varphi(\widetilde{X}_\cdot) = \varphi(\widetilde{X}_0) + M_\cdot^\varphi + A_\cdot^\varphi$ of $\varphi(\widetilde{X}_\cdot)$, where $M_\cdot^\varphi$ and $A_\cdot^\varphi$ are the martingale and the finite variation terms, respectively,

$$\mathrm{d}M_t^\varphi := \varphi'(\widetilde{X}_t)\mathrm{d}B_{t,2}, \quad \mathrm{d}A_t^\varphi := \varphi'(\widetilde{X}_t)b(t, \widetilde{X}_t, F_{t,2})\mathrm{d}t + \frac{1}{2}\varphi''(\widetilde{X}_t)\mathrm{d}t,$$

and then consider $\widetilde{Z}_t^\varepsilon \cdot \varphi(\widetilde{X}_t)$, $t \ge 0$ for $\varepsilon > 0$, and its conditional expectation with respect to $\mathcal{F}_T^X$, where $\widetilde{Z}_t^\varepsilon := Z_t^{-1}/(1 + \varepsilon Z_t^{-1})$, $t \ge 0$. Since $\mathrm{d}Z_t^{-1} = Z_t^{-1}b(t, X_t, F_t)\mathrm{d}X_t$, $t \ge 0$, we have

$$\mathrm{d}\widetilde{Z}^{\varepsilon}_t = \frac{Z_t^{-1} b(t, X_t, F_t)}{(1+\varepsilon Z_t^{-1})^2}\mathrm{d}X_t - \frac{\varepsilon Z_t^{-2}|b(t, X_t, F_t)|^2}{(1+\varepsilon Z_t^{-1})^3}\mathrm{d}t\,; \quad t \ge 0\,.$$

Substituting these expressions for $M^{\varphi}_{\cdot}$, $A^{\varphi}_{\cdot}$ and $\widetilde{Z}^{\varepsilon}_{\cdot}$ into

$$\begin{aligned}\mathbb{E}_0[\widetilde{Z}^{\varepsilon}_t \varphi(\widetilde{X}_t)|\mathcal{F}^X_T] &= \mathbb{E}_0[\widetilde{Z}^{\varepsilon}_0 \varphi(\widetilde{X}_0)|\mathcal{F}^X_T] \\ &\quad + \mathbb{E}_0\Big[\int_0^t \widetilde{Z}^{\varepsilon}_s \mathrm{d}A^{\varphi}_s + \int_0^t \widetilde{Z}^{\varepsilon}_s \mathrm{d}M^{\varphi}_s + \int_0^t \varphi(\widetilde{X}_s)\mathrm{d}\widetilde{Z}^{\varepsilon}_s \Big| \mathcal{F}^X_T\Big],\end{aligned} \tag{4.21}$$

and taking the limits as $\varepsilon \downarrow 0$ under (4.12) and (4.17), we obtain (4.18) with (4.19)–(4.20). Indeed, we need (4.12) to show (4.11) and then with $\|\varphi\|_\infty := \sup_{x\in\mathbb{R}}|\varphi(x)|$ for $\varphi \in C_0^2(\mathbb{R})$,

$$\mathbb{E}_0\Big[\int_0^t \Big|\widetilde{Z}^{\varepsilon}_s \cdot \frac{\varphi(\widetilde{X}_s) b(s, X_s, F_s)}{(1+\varepsilon Z_s^{-1})}\Big|^2 \mathrm{d}s\Big] \le \|\varphi\|^2_\infty \mathbb{E}\Big[\int_0^t Z_s |b(s, X_s, F_s)|^2 \mathrm{d}s\Big] < \infty\,,$$

$$\mathbb{E}_0\Big[\int_0^t \mathbb{E}_0\Big[\Big|\widetilde{Z}^{\varepsilon}_s \cdot \frac{\varphi(\widetilde{X}_s) b(s, X_s, F_s)}{(1+\varepsilon Z_s^{-1})}\Big|^2 \Big| \mathcal{F}^X_T\Big]\mathrm{d}s\Big] \le \|\varphi\|^2_\infty \mathbb{E}\Big[\int_0^t Z_s |b(s, X_s, F_s)|^2 \mathrm{d}s\Big] < \infty\,,$$

and hence

$$\mathbb{E}_0\Big[\int_0^t \widetilde{Z}^{\varepsilon}_s \cdot \frac{\varphi(\widetilde{X}_s) b(s, X_s, F_s)}{(1+\varepsilon Z_s^{-1})}\mathrm{d}X_s \Big| \mathcal{F}^X_T\Big] = \int_0^t \mathbb{E}_0\Big[\widetilde{Z}^{\varepsilon}_s \cdot \frac{\varphi(\widetilde{X}_s) b(s, X_s, F_s)}{(1+\varepsilon Z_s^{-1})}\Big| \mathcal{F}^X_T\Big]\mathrm{d}X_s\,. \tag{4.22}$$

is a martingale under $\mathbb{P}_0$ for $0 \le t \le T$. We need (4.17) to verify that $\mathbb{P}_0$ a.s.

$$\begin{aligned}&\int_0^t \Big|\mathbb{E}_0\Big[\widetilde{Z}^{\varepsilon}_s \cdot \frac{\varphi(\widetilde{X}_s) b(s, X_s, F_s)}{(1+\varepsilon Z_s^{-1})}\Big|\mathcal{F}^X_T\Big] - \rho_{s,2}(\varphi b)\Big|^2 \mathrm{d}s \\ &\quad \le 4\|\varphi\|^2_\infty \int_0^t |\mathbb{E}_0[|b(s, X_s, F_s)|\big|\mathcal{F}^X_T]|^2 \mathrm{d}s < \infty\end{aligned}$$

for $0 \le t \le T$ and then by the dominated convergence theorem to show that as $\varepsilon \downarrow 0$, for a suitably chosen subsequence $\varepsilon_n \downarrow 0$, (4.22) converges $\mathbb{P}_0$-a.s. to the $\mathbb{P}_0$-local martingale $\int_0^{\cdot} \rho_{s,2}(\varphi b)\mathrm{d}X_s$. The convergence of the other terms in (4.21) along ε_n is relatively straightforward. □

More generally, for every $n \ge 2$ and every $k = 1, 2, \ldots, n$, given $\varphi(t, x) \in C^{1,2}([0, T] \times \mathbb{R}^k)$ with bounded support in $\mathbb{R}^k$ and bounded in time $0 \le t \le T$ let us recall (4.5), i.e., $\overline{X}_{\cdot,1} \equiv X_{\cdot}$, and define

$$\rho_{t,k}(\varphi) := \mathbb{E}_0[Z_t^{-1}\varphi(t, \overline{X}_{t,1}, \ldots, \overline{X}_{t,k})|\mathcal{F}^X_T]\,; \quad k = 2, \ldots, n\,,\ 0 \le t \le T\,, \tag{4.23}$$

and similarly, let us define the normalized version

$$\pi_{t,k}(\varphi) := \mathbb{E}[\varphi(t, \overline{X}_{t,1}, \ldots, \overline{X}_{t,k})\,|\,\mathcal{F}^X_t]\,; \quad k = 2, \ldots, n,\ 0 \le t \le T \tag{4.24}$$

for $\varphi \in C_0^2([0, T] \times \mathbb{R}^k)$, and for $(s, x) \in [0, T] \times \mathbb{R}^k$, $i = 1, 2, \ldots, n$

$$D_s\varphi(s, x) := \frac{\partial \varphi}{\partial s}(s, x)\,, \quad D_i\varphi(s, x) := \frac{\partial \varphi}{\partial x_i}(s, x)\,, \quad D_i^2\varphi(s, x) := \frac{\partial^2 \varphi}{\partial x_i^2}(s, x)\,.$$

Then with a similar reasoning as in the proof of Proposition 4.2, we obtain the following system (4.25) of Zakai equations for the (unnormalized) conditional expectations $\rho_{\cdot,k}$ of function of $(\overline{X}_{\cdot,1}, \ldots, \overline{X}_{\cdot,k})$, $k = 2, \ldots, n$ with respect to $\mathcal{F}^X_T$ and for arbitrary $n \ge 2$.

Proposition 4.3. *Under the same assumption as in Proposition 4.2, $\rho_{\cdot,k}(\varphi)$ in (4.23) satisfies*

$$\rho_{t,k}(\varphi) = \pi_{0,k}(\varphi) + \int_0^t \rho_{s,k}(\varphi b)\mathrm{d}X_s + \int_0^t \rho_{s,k+1}(\widetilde{A}_s\varphi)\mathrm{d}s\,, \tag{4.25}$$

where the integrands are defined by

$$\begin{aligned}\rho_{s,k+1}(\widetilde{\mathcal{A}}_s\varphi) := \mathbb{E}_0\Big[Z_s^{-1}\Big(&\sum_{i=1}^k D_i\varphi(s,\overline{X}_{s,1},\dots,\overline{X}_{s,k})\cdot b(s,\overline{X}_{s,i},F_{s,i})\\ &+\frac{1}{2}\sum_{i=1}^k D_i^2\varphi(s,\overline{X}_{s,1},\dots,\overline{X}_{s,k})\\ &+D_s\varphi(s,\overline{X}_{s,1},\dots,\overline{X}_{s,k}) + D_1\varphi(s,\overline{X}_{s,1},\dots,\overline{X}_{s,k})\\ &\cdot b(s,\overline{X}_{s,1},F_{s,1})\Big)\Big|\mathcal{F}_T^X\Big],\end{aligned}$$

and

$$\rho_{s,k}(\varphi b) := \mathbb{E}_0\big[Z_s^{-1}\varphi(s,\overline{X}_{s,1},\dots,\overline{X}_{s,k})b(s,X_{s,1},F_{s,1})\,|\,\mathcal{F}_T^X\big] \tag{4.26}$$

for $0\le s\le t\le T$, $k=2,\dots,n$ and for arbitrary $n\ge 2$.

Now under the assumption (4.12), we have the Kallianpur–Striebel formula: $\mathbb{P}_0$ ($\mathbb{P}$)-a.s.

$$\pi_{t,k}(\varphi) = \frac{\rho_{t,k}(\varphi)}{\rho_t(\mathbf{1})}\,;\quad k=2,\dots,n\,,\ 0\le t\le T\,.$$

Then it follows from Proposition 4.2 that

$$\rho_t(\mathbf{1}) = 1 + \int_0^t \rho_{s,2}(b)\mathrm{d}X_s = 1 + \int_0^t \rho_s(\mathbf{1})\pi_{s,2}(b)\mathrm{d}X_s\,;\quad 0\le t\le T\,. \tag{4.27}$$

For fixed $\varepsilon>0$, applying Ito's formula to $(1/2)\log(\varepsilon+|\rho_t(\mathbf{1})|^2)$ with (4.27), we obtain

$$\mathrm{d}\Big(\frac{1}{2}\log(\varepsilon+|\rho_t(\mathbf{1})|^2)\Big) = \frac{|\rho_t(\mathbf{1})|^2\pi_{t,2}(b)}{\varepsilon+|\rho_t(\mathbf{1})|^2}\mathrm{d}X_t + \frac{(\varepsilon-|\rho_t(\mathbf{1})|^2)|\rho_t(\mathbf{1})|^2|\pi_{t,2}(b)|^2}{2(\varepsilon+|\rho_t(\mathbf{1})|^2)^2}\mathrm{d}t$$

for $0\le t\le T$. Under the assumption of Proposition 4.3, letting $\varepsilon\downarrow 0$, by the dominated convergence theorem, we have $\log\rho_t(\mathbf{1}) = \int_0^t \pi_{s,2}(b)\mathrm{d}X_s - \frac{1}{2}\int_0^t|\pi_{s,2}(b)|^2\mathrm{d}s$, and hence

$$\mathbb{E}_0[Z_t^{-1}|\mathcal{F}_T^X] = \rho_t(\mathbf{1}) = \exp\Big(\int_0^t \pi_{s,2}(b)\mathrm{d}X_s - \frac{1}{2}\int_0^t|\pi_{s,2}(b)|^2\mathrm{d}s\Big)\,;\quad 0\le t\le T\,.$$

Proposition 4.4. *In addition to the assumption in Proposition 4.2, let us assume*

$$\mathbb{P}\Big(\int_0^t|\pi_{s,k}(b)|^2\mathrm{d}s<\infty\Big) = 1\,;\quad k=2,\dots,n\,,\ 0\le t\le T\,. \tag{4.28}$$

Then the conditional expectation $\pi_{\cdot,k}(\varphi)$ in (4.24) with respect to $\mathcal{F}_\cdot^X$ satisfies the following system of Kushner–Stratonovich equations

$$\pi_{t,k}(\varphi) = \pi_{0,k}(\varphi) + \int_0^t(\pi_{s,k}(\varphi b)-\pi_{s,k}(b)\pi_{s,k}(\varphi))\big(\mathrm{d}X_s-\pi_{s,k}(b)\mathrm{d}s\big) + \int_0^t \pi_{s,k+1}(\widetilde{\mathcal{A}}_s\varphi)\mathrm{d}s \tag{4.29}$$

where we define $\pi_{s,k}(\varphi b) := \rho_{s,k}(\varphi b)/\rho_s(\mathbf{1})$, $\pi_{s,k+1}(\widetilde{\mathcal{A}}_s\varphi) := \rho_{s,k+1}(\widetilde{\mathcal{A}}_s\varphi)/\rho_s(\mathbf{1})$ from (4.26) for $0\le t\le T$, $k=2,\dots,n$ and for arbitrary $n\ge 2$.

Proof. The proof is now straightforward by Itô's formula under the condition (4.28) thanks to the representations of $\rho.(\mathbf{1})$ and $\rho_{\cdot,k}(\varphi)$ in (4.25). $\square$

Remark 4.2. As in Remarks 2.4 and 3.1, when $u \in (0, 1]$, as a description of the conditional expectations $\pi_{\cdot,k}$, $\rho_{\cdot,k}$, $k = 2, \dots, n$ with respect to $\mathcal{F}_T^X$, the system (4.25) or the system (4.29) has an infinite-dimensional aspect. This is because $\rho_{\cdot,k}$ in (4.25) is represented by the integral of $\rho_{\cdot,k}$ with respect to $\mathrm{d}X_s$ and the integral of $\rho_{\cdot,k+1}$ with respect to $\mathrm{d}s$ and because $\rho_{\cdot,k+1}$ is the conditional expectation of function of $(\overline{X}_{\cdot,1}, \dots, \overline{X}_{\cdot,k+1})$ if $u \in (0, 1]$ for every $k = 2, \dots, n$. The system (4.29) for $\pi_{\cdot,k}$ has the same aspect, inherited from $\rho_{\cdot,k}$ in (4.25). Since n is arbitrary, the chain of such descriptions continues. Unique characterizations of conditional expectations as a solution to this infinite system of Zakai equations (4.25) and Kushner–Stratonovich equations (4.29) are out of scope of the current paper. Numerical methods, particle methods and their comparisons for solving such systems are also interesting ongoing projects. $\square$

4.2. Connection to the infinite-dimensional Ornstein–Uhlenbeck process

Let us take a time-homogeneous linear functional $b(t, x, \mu) := -\int_{\mathbb{R}}(x - y)\mu(\mathrm{d}y)$ for $t \ge 0$, $x \in \mathbb{R}$, $\mu \in \mathcal{M}(\mathbb{R})$ of mean-reverting type. Then, (2.1) is reduced to the stochastic differential equation

$$\mathrm{d}X_t^{(u)} = -\big(u\,(X_t^{(u)} - \widetilde{X}_t^{(u)}) + (1-u)(X_t^{(u)} - \mathbb{E}[X_t^{(u)}])\big)\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0 \tag{4.30}$$

for each $u \in [0, 1]$, where we recall that $\widetilde{X}^{(u)}$ has the same law as $X^{(u)}$ and is independent of the Brownian motion B.

In the case $u = 0$, we have $X_\cdot^{(0)} = X_\cdot^{\bullet}$, where $X_\cdot^{\bullet}$ is the solution of the pure McKean–Vlasov stochastic differential equation

$$\mathrm{d}X_t^{\bullet} = -(X_t^{\bullet} - \mathbb{E}[X_t^{\bullet}])\mathrm{d}t + \mathrm{d}B_t\,, \quad t \ge 0\,. \tag{4.31}$$

In the case $u = 1$, we have $X_\cdot^{(1)} = X_\cdot^{\dagger}$, where $X_\cdot^{\dagger}$ is given by

$$\mathrm{d}X_t^{\dagger} = -(X_t^{\dagger} - \widetilde{X}_t^{\dagger})\mathrm{d}t + \mathrm{d}B_t\,, \quad t \ge 0\,, \tag{4.32}$$

with $\widetilde{X}^{\dagger}$ having the same law as $X^{\dagger}$ and being independent of B.

Coming back to the general case with $u \in [0, 1]$ and setting a fixed initial value $X_0^{(u)} = 0$, we see that the expectations are constant in time

$$\mathbb{E}[X_t^{(u)}] = \mathbb{E}[\widetilde{X}_t^{(u)}] = \mathbb{E}[X_t^{\bullet}] = \mathbb{E}[X_t^{\dagger}] = 0\,, \quad t \ge 0\,,\ u \in [0, 1]\,, \tag{4.33}$$

with an explicitly solvable Gaussian pair $(X^{(u)}(t), \widetilde{X}^{(u)}(t))$ for $t \ge 0$, $u \in [0, 1]$

$$\begin{aligned} X_t^{(u)} &= \int_0^t e^{-(t-s)} u \widetilde{X}_s^{(u)} \mathrm{d}s + \int_0^t e^{-(t-s)} \mathrm{d}B_s\,, \\ \widetilde{X}_t^{(u)} &= \int_0^t \sum_{k=0}^{\infty} \mathfrak{p}_{0,k}(t-s; u)\,\mathrm{d}W_{s,k}\,, \quad \mathfrak{p}_{0,k}(t-s; u) := \frac{u^k (t-s)^k}{k!} e^{-(t-s)}\,, \end{aligned} \tag{4.34}$$

where $(W_\cdot^k, k \ge 0)$ is a sequence of independent, one-dimensional standard Brownian motions, independent of the Brownian motion $B(\cdot)$. Note that the integrand $\mathfrak{p}_{0,k}(t-s; u)$, $k \in \mathbb{N}_0$ in (4.34) is a (taboo) transition probability $\mathbb{P}(M(t-s) = k | M(0) = 0)$ of a continuous-time Markov chain $M(\cdot)$ in the state space $\mathbb{N}_0$ with generator matrix $\boldsymbol{Q} = (q_{i,j})_{i,j \in \mathbb{N}_0}$

with $q_{i,i+1} = u \in [0,1]$, $q_{i,i} = -1$ and $q_{i,j} = 0$ for the other entries $j \neq i, i+1$. When $u = 0$, $\boldsymbol{Q}$ is the generator of a Markov chain with jump rate 1 from state i and is killed immediately. When $u = 1$, $\boldsymbol{Q}$ is the generator of a Poisson process with rate 1. When $u \in (0,1)$, it jumps from i to $i+1$ with probability u and killed with probability $(1-u)$. Thus we interpret $\mathfrak{p}_{0,k}(t-s;u)$ as $(0,k)$-element of the $\mathbb{N}_0 \times \mathbb{N}_0$-dimensional matrix exponential $e^{(t-s)Q}$, i.e.,

$$(\mathfrak{p}_{i,j}(t-s;u) := \mathbb{P}(M(t-s) = j | M(0) = i), i, j \in \mathbb{N}_0) \equiv ((e^{(t-s)\boldsymbol{Q}})_{i,j}, i, j \in \mathbb{N}_0); \quad t \geq s \geq 0.$$

For the matrix exponential $e^{t\boldsymbol{Q}}$, $t \geq 0$ of such $\boldsymbol{Q}$, see for example, [11]. Then we have a Feynman–Kac representation formula

$$\widetilde{X}_t^{(u)} = \mathbb{E}^M \Big[\int_0^t \sum_{k=0}^{\infty} \mathbf{1}_{\{M(t-s)=k\}} \mathrm{d}W_{s,k} | M(0) = 0 \Big]; \quad t \geq 0, \tag{4.35}$$

where the expectation is taken with respect to the probability induced by the Markov chain $M(\cdot)$, independent of the Brownian motions $(W_{\cdot,k}, k \in \mathbb{N}_0)$.

Indeed, by Proposition 3.1, the solution (4.34) is obtained by an infinite particle approximation

$$\mathrm{d}X_{t,k}^{(u)} = -(X_{t,k}^{(u)} - uX_{t,k+1}^{(u)})\mathrm{d}t + \mathrm{d}W_{t,k}; \quad t \geq 0, \quad k \in \mathbb{N}_0 \tag{4.36}$$

of the simplified form of (4.30), that is,

$$\mathrm{d}X_t^{(u)} = -(X_t^{(u)} - u\widetilde{X}_t^{(u)})\mathrm{d}t + \mathrm{d}B_t; \quad t \geq 0.$$

Here we assume $\boldsymbol{\sigma}(X_{t,k+1}^{(u)}, t \geq 0)$ and $\boldsymbol{\sigma}(W_{t,k}, t \geq 0)$ are independent for every $k \in \mathbb{N}_0$. The infinite particle system (4.36) can be represented as an infinite-dimensional Ornstein–Uhlenbeck stochastic differential equation or more generally, stochastic evolution equation (see e.g., [1,3,8,9,16] for more general results in Hilbert spaces)

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{Q}\boldsymbol{X}_t\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t, \tag{4.37}$$

where $\boldsymbol{X}_\cdot := (X_{\cdot,k}^{(u)}, k \in \mathbb{N}_0)$ with $\boldsymbol{X}_0 = \boldsymbol{0}$, and $\boldsymbol{W}_\cdot := (W_{\cdot,k}, k \in \mathbb{N}_0)$. Note that the transition probabilities $\mathbb{P}(M(t) = k | M(0) = i) = (e^{t\boldsymbol{Q}})_{i,k}$, $i, k \in \mathbb{N}_0$ of the continuous-time Markov chain $M(\cdot)$ defined in the previous paragraph satisfy the backward Kolmogorov equation

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{t\boldsymbol{Q}} = \boldsymbol{Q}\, e^{t\boldsymbol{Q}}; \quad t \geq 0.$$

Thus, by Itô's formula we directly verify

$$\mathrm{d}\Big(\int_0^t e^{(t-s)\boldsymbol{Q}} \mathrm{d}\boldsymbol{W}_s \Big) = \Big(\boldsymbol{Q} \int_0^t e^{(t-s)\boldsymbol{Q}} \mathrm{d}\boldsymbol{W}_s \Big) \mathrm{d}t + \mathrm{d}\boldsymbol{W}_t; \quad t \geq 0,$$

and hence

$$\boldsymbol{X}_t = \int_0^t e^{(t-s)\boldsymbol{Q}}\, \mathrm{d}\boldsymbol{W}_s; \quad t \geq 0,$$

is a solution to (4.37). Therefore, (4.34) is the solution to (4.30). Although $\boldsymbol{Q}$ has the specific form here, it is easy to see that in general, the Feynman–Kac formula (4.35) still holds for the infinite-dimensional Ornstein–Uhlenbeck process with a class of generators $\boldsymbol{Q}$ which form a Banach algebra (e.g., the generator of the discrete-state, compound Poisson processes, see [11]).

Table 1
Different behaviors for different values of u in the linear, Gaussian case (4.30). Asymptotic variances are given in Remark 4.3. Dependence is described in (4.38).

u	Interaction type	Asymptotic variance	Asymptotic independence (Propagation of chaos)
$u = 0$	Purely mean-field (4.31)	Stabilized	Independent
$u \in (0, 1)$	Mixed interaction	Stabilized	Dependent
$u = 1$	Purely directed chain (4.32)	Explosive	Dependent

4.2.1. Asymptotic dichotomy

With $X_0^{(u)} = 0$ still, it follows from (4.34) that the auto covariance and cross covariance are

$$\mathbb{E}[X_s^{(u)} X_t^{(u)}] = \mathbb{E}[\widetilde{X}_s^{(u)} \widetilde{X}_t^{(u)}] = e^{-(t-s)} \int_0^s e^{-2v} I_0(2u\sqrt{(t-s+v)v})\mathrm{d}v \, ; \quad 0 \le s \le t \, ,$$
$$\mathbb{E}[X_s^{(u)} \widetilde{X}_t^{(u)}] = u \int_0^s e^{-(s-v)} \mathbb{E}[\widetilde{X}_v^{(u)} \widetilde{X}_t^{(u)}]\mathrm{d}v = u \int_0^s e^{-(s-v)} \mathbb{E}[X_v^{(u)} X_t^{(u)}]\mathrm{d}v \, ; \quad t, s \ge 0 \tag{4.38}$$

for $u \in [0, 1]$. Here $I_\nu(\cdot)$ is the modified Bessel function of the first kind with index ν, defined by

$$I_\nu(x) := \sum_{k=0}^{\infty} \frac{(x/2)^{2k+\nu}}{\Gamma(k+1)\cdot\Gamma(\nu+k+1)} \, ; \quad x > 0 \, , \nu \ge -1 \, .$$

Note that the Bessel functions $I_0(x)$ and $I_1(x)$ grow with the order of $O(e^x / \sqrt{2\pi x})$ as $x \to \infty$.

Remark 4.3 (*Asymptotic Dichotomy of* (4.30))**.** The asymptotic behaviors of their variances as $t \to \infty$ are dichotomous as in Table 1:

$$\mathrm{Var}(X_t^{(u)}) = \int_0^t e^{-2v} I_0(2uv)\mathrm{d}v = \begin{cases} O(1) \, , & u \in [0, 1) \, , \\ O(\sqrt{t}) \, , & u = 1 \, , \end{cases} \tag{4.39}$$

$$\text{with} \quad \mathrm{Var}(X_t^{(0)}) = \mathrm{Var}(X_t^{\bullet}) = \frac{1 - e^{-2t}}{2} \, ,$$
$$\mathrm{Var}(X_t^{(1)}) = \mathrm{Var}(X_t^{\dagger}) = t\, e^{-2t}(I_0(2t) + I_1(2t)) \, .$$

1. When $u \in [0, 1)$, the process $X_\cdot^{(u)}$ is positive recurrent and its stationary distribution is Gaussian with mean 0 and variance
$$\lim_{t\to\infty} \mathrm{Var}(X_t^{(u)}) = \int_0^\infty e^{-2v} I_0(2uv)\mathrm{d}v = \frac{1}{2\sqrt{1-u^2}} < \infty \, . \tag{4.40}$$
In particular, when $u = 0$, $X_\cdot^{(0)} = X_\cdot^{\bullet}$ is an Ornstein–Uhlenbeck process with a stationary Gaussian distribution of mean 0 and variance $1/2$, independent of $\widetilde{X}_\cdot^{(0)} = \widetilde{X}_\cdot^{\bullet}$.
2. When $u = 1$, the process $X_\cdot^{(1)} = X_\cdot^{\dagger}$ is a mean zero Gaussian process with growing variance of the order $O(\sqrt{t})$ with $\lim_{t\to\infty} \mathrm{Var}(X_t^{\dagger}) = \infty$, given by (4.39)

and covariances

$$\mathbb{E}[X^{\dagger}_s X^{\dagger}_t] = \mathbb{E}[\widetilde{X}^{\dagger}_s \widetilde{X}^{\dagger}_t] = e^{-(t-s)} \int_0^s e^{-2v} I_0(2\sqrt{(t-s+v)v}) \mathrm{d}v \, ; \quad 0 \le s \le t \, .$$

In particular, $\mathbb{E}[X^{\dagger}_s X^{\dagger}_t] = O(e^{-(t-2\sqrt{(t+s)s})} t^{-1/4})$ for large $t \to \infty$.

This asymptotic dichotomy is an answer to the first question posed in Section 1. Namely, the large system of type (1.1) diverges widely, while the large system of type (1.2) converges to the stationary distribution as $t \to \infty$ under the linear case of (4.30). □

Remark 4.4 (*Repulsive Case*)**.** Instead of mean-reverting, if the drift functional b is of *repulsive* type $b(t, x, \mu) := \int_{\mathbb{R}} (x-y)\mu(\mathrm{d}y)$, then the resulting paired process in (2.1) with $u = 1$ is described by

$$\mathrm{d}X^{\ddagger}_t = (X^{\ddagger}_t - \widetilde{X}^{\ddagger}_t)\mathrm{d}t + \mathrm{d}B_t \, ; \quad t \ge 0 \tag{4.41}$$

with the conditions (2.2)–(2.3). The solution with the initial values $X^{\ddagger}_0 = \widetilde{X}^{\ddagger}_0 = 0$ is given by

$$X^{\ddagger}_t = \int_0^t e^{t-s} \widetilde{X}^{\ddagger}_s \mathrm{d}s + \int_0^t e^{t-s} \mathrm{d}B_s \, , \quad \widetilde{X}^{\ddagger}_t = \int_0^t \sum_{k=0}^{\infty} e^{t-s} \cdot \frac{(-1)^k (t-s)^k}{k!} \mathrm{d}W_{s,k} \, ; \quad t \ge 0$$

for independent Brownian motions $W_{\cdot,k}$, $k \in \mathbb{N}_0$, independent of $B_{\cdot}$. In this case the variance grows *exponentially* fast, i.e., $\mathrm{Var}(X^{\ddagger}_t) = t e^{2t}(I_0(2t) - I_1(2t))$, $t \ge 0$. □

Remark 4.5 (*Discrete-time Time Series*)**.** The discrete-time version of (4.30) with distributional constraints can be defined by the difference equation for $(X_k, \widetilde{X}_k) = (X^{(u)}_k, \widetilde{X}^{(u)}_k)$, $k = 0, 1, \ldots$, for example, given constants $a \in (0, 1)$, $u \in [0, 1]$,

$$X_k = aX_{k-1} + (1-a)(u\widetilde{X}_{k-1} + (1-u)\mathbb{E}[X_{k-1}]) + \varepsilon_k \, ; \quad k = 1, 2, \ldots \tag{4.42}$$

where we assume $\mathrm{Law}(\{X_k, k = 0, 1, 2, \ldots\}) \equiv \mathrm{Law}(\{\widetilde{X}_k, k = 0, 1, 2, \ldots\})$ and the independently, identically distributed noise sequence ε_k, $k \ge 1$ is independent of $\widetilde{X}_{\cdot}$. We shall solve for the joint distribution of $(X_k, \widetilde{X}_k)$, $k \ge 0$. Again for simplicity, let us assume $X_0 = 0 = \widetilde{X}_0$. Then it reduces to $\mathbb{E}[X_k] = \mathbb{E}[\widetilde{X}_k]$, $k = 0, 1, 2, \ldots$, and hence to $X_k = aX_{k-1} + (1-a)u\widetilde{X}_{k-1} + \varepsilon_k$; $k = 1, 2, \ldots$ with distributional constraints.

By recursive substitutions, we have $X_1 = \varepsilon_1$, $X_2 = aX_1 + (1-a)u\widetilde{X}_1 + \varepsilon_2$, $X_3 = aX_2 + (1-a)u\widetilde{X}_2 + \varepsilon_3$, and $X_n = u\sum_{k=1}^{n-1} a^{k-1}(1-a)\widetilde{X}_{n-k} + \sum_{k=0}^{n-1} a^k \varepsilon_{n-k}$. Thanks to the distributional constraints, we represent the distribution of the solution to (4.42) as

$$X_n = \sum_{0 \le \ell \le k \le n-1} \binom{k}{\ell} u^{\ell}(1-a)^{\ell} a^{k-\ell} \varepsilon_{n-k,\ell}, \quad \widetilde{X}_n = \sum_{0 \le \ell \le k \le n-1} \binom{k}{\ell} u^{\ell}(1-a)^{\ell} a^{k-\ell} \varepsilon_{n-k,\ell+1},$$

where $\varepsilon_{n,m}$, $n, m \in \mathbb{N}$ are independently, identically distributed noise with $\varepsilon_{n,0} = \varepsilon_n$ for $n \in \mathbb{N}$. While the stochastic kernel in the stochastic integral in (4.34) for the solution to the continuous time equation (4.30) is a Poisson probability, the stochastic kernel for the solution to the discrete time equation (4.42) is a binomial probability. The variance and covariances can be calculated, e.g.,

$$\mathbb{E}[X_n^2] = \sum_{k=0}^{n-1} \sum_{\ell=0}^{k} \binom{k}{\ell}^2 u^{2\ell}(1-a)^{2\ell} a^{2(k-\ell)} = \sum_{k=0}^{n-1} u^k (1-a)^k \, {}_2F_1\Big(-k, -k, 1; \, \frac{a^2}{(1-a)^2}\Big),$$

where ${}_2F_1(\cdot)$ is the Gauss hypergeometric function. □

Remark 4.6. We may generalize these explicit examples in this section to time-inhomogeneous, linear equations, where the dependence on expectations and marginal laws remain to exist in the expressions. The resulting expressions would become more complicated. Here we demonstrate some simple examples with the connection to the infinite-dimensional Ornstein–Uhlenbeck processes. □

4.2.2. Consistent estimation

Let us denote by $\mathcal{F}_t, t \geq 0$ the filtration generated by the solution pair $(X_\cdot, \widetilde{X}_\cdot) := (X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)})$ in (4.30). Thanks to the Girsanov theorem, the log Radon–Nikodym derivative of the solution $\mathbb{P}^{(u)}$ with respect to the Wiener measure $\mathbb{P}_0$ is given by

$$\log \frac{\mathrm{d}\mathbb{P}^{(u)}}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_T} = \int_0^T (X_t - u\widetilde{X}_t)\mathrm{d}X_t + \frac{1}{2}\int_0^T (X_t - u\widetilde{X}_t)^2 \mathrm{d}t \,.$$

Thus given $\mathcal{F}_T^X$, the observer may maximize the conditional log likelihood function

$$\mathbb{E}\Big[-\log\Big(\frac{\mathrm{d}\mathbb{P}^{(u)}}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_T}\Big)\Big|\mathcal{F}_T^X\Big]$$

with respect to u, and formally obtain a unique maximizer, corresponding to (4.16),

$$\widehat{u} := \Big(\int_0^T \mathbb{E}[\widetilde{X}_t^2|\mathcal{F}_T^X]\mathrm{d}t\Big)^{-1} \cdot \mathbb{E}\Big[\int_0^T X_t\widetilde{X}_t\mathrm{d}t + \int_0^T \widetilde{X}_t\mathrm{d}X_t\,\Big|\,\mathcal{F}_T^X\Big] \tag{4.43}$$

as an estimator of u. Evaluation of these conditional expectations in (4.43) is a filtering problem.

The detailed study of $\widehat{u}$ in (4.43) still remains an open problem. If we replace $\widetilde{X}_\cdot$ by $X_\cdot$ in (4.43), then we obtain a modified estimator

$$\widehat{u}_m := \Big(\int_0^T X_t^2\mathrm{d}t\Big)^{-1} \cdot \Big(\int_0^T X_t^2\mathrm{d}t + \int_0^T X_t\mathrm{d}X_t\Big) = 1 - \Big(2\int_0^T X_t^2\mathrm{d}t\Big)^{-1}\big(T - X_T^2\big)\,. \tag{4.44}$$

It follows from (4.40) that $\lim_{T\to\infty}\widehat{u}_m = 1 - \sqrt{1-u^2} \leq u \in [0,1]$. Thus this modified estimator $\widehat{u}_m$ underestimates the value u asymptotically as $T \to \infty$.

Another typical method of estimation of u is known as the method of moments. We may obtain the method of moments estimator by matching the second moment in the limit, i.e.,

$$\widehat{u}_M := \Big[1 - \Big(\frac{2}{T}\int_0^T X_t^2\mathrm{d}t\Big)^{-2}\Big]^{1/2}. \tag{4.45}$$

It follows from (4.40) directly that $\lim_{T\to\infty}\widehat{u}_M = u \in [0,1]$. Thus this method of moments estimator $\widehat{u}_M$ is asymptotically consistent to the value $u \in [0,1]$, as $T \to \infty$.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We wish to thank Professors E. Bayraktar, I. Karatzas, K. Kardaras, H. Kawabi, K. Ramanan, M. Shkolnikov, H. Tanemura and Z. Zhang for several helpful discussions in the early stage of this paper. We are thankful to Drs. E. R. Fernholz and I. Karatzas for kindly sharing with us their unpublished manuscripts on the Gaussian cascades. We also appreciate the editors and the reviewers comments and suggestions. Research is supported in part by the National Science Foundation, USA under grants NSF-DMS-1409434 and NSF-DMS-1814091 for the second author and under grants NSF-DMS-13-13373 and NSF-DMS-16-15229 for the third author.

Appendix

A.1. Sketch of proof of (3.28)

We shall sketch the poof of (3.28) for Proposition 3.3 when $u > 0$. If $u = 0$, it reduces to the case of propagation of chaos results and it is given in [27]. First note that by the construction, $\overline{X}_{\cdot,i}$ in (3.5)–(3.6) is determined by the iteration $\overline{X}_{\cdot,i} = \mathbf{\Phi}(\cdot, (m_s)_{0\le s\le \cdot}, (\overline{X}_{s,i+1})_{0\le s\le \cdot}, (W_{s,i})_{0\le s\le \cdot})$ as in (2.12), where $\overline{X}_{\cdot,i+1}$ is independent of $W_{\cdot,i}$ for $i = n, n-1, \ldots, 1$, that is, with this random iterative map and a slight abuse of notation, we may write and view

$$\eta_{t,i} := \overline{X}_{t,i} = \mathbf{\Phi}\circ\mathbf{\Phi}\circ\cdots\circ\mathbf{\Phi}_t(\overline{X}_{\cdot,n+1}; W_{\cdot,i}, \ldots, W_{\cdot,n}) = \mathbf{\Phi}_t^{(n+1-i)}(\eta_{n+1}; W_{\cdot,i}, \ldots, W_{\cdot,n}) \tag{A.1}$$

for $0 \le t \le T$ as an element in the space $C([0,T],\mathbb{R}) = C([0,T])$ of continuous functions. Thus, η_i, $i = n+1, n, n-1, \ldots, 1$ possess a discrete-time Markov chain structure. In particular, for $j < k < i$, given η_k, the distribution of η_i and η_j are conditionally independent.

Let us write $\boldsymbol{W} := (W_{\cdot,1}, \ldots, W_{\cdot,n})$ for simplicity. For every Lipschitz function $\varphi(\cdot)$ with Lipschitz constant K, there exists a constant $c > 0$ such that the difference between the conditional expectation of $\varphi(\overline{X}_{t,1})$, given $\overline{X}_{s,n+1}$, $0 \le s \le T$ and the unconditional expectation of $\varphi(\overline{X}_{t,1})$ is bounded by

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T} |\mathbb{E}[\varphi(\overline{X}_{t,1})|\overline{X}_{s,n+1}, 0\le s\le T] - \mathbb{E}[\varphi(\overline{X}_{t,1})]|^2\Big] \\
&= \int_{C([0,T])} \sup_{0\le t\le T} \Big| \int_{C([0,T])} \Big(\mathbb{E}^{\boldsymbol{W}}[\varphi(\mathbf{\Phi}_t^{(n)}(\eta_{n+1}; \boldsymbol{W}_\cdot))] - \mathbb{E}^{\boldsymbol{W}}[\varphi(\mathbf{\Phi}_t^{(n)}(\widetilde{\eta}_{n+1}; \boldsymbol{W}_\cdot))]\Big) \\
&\quad \times \mathrm{m}(\mathrm{d}\widetilde{\eta}_{n+1})\Big|^2 \mathrm{m}(\mathrm{d}\eta_{n+1}) \\
&\le \int_{C([0,T])^2} \sup_{0\le t\le T} \mathbb{E}^{\boldsymbol{W}}[|\varphi(\mathbf{\Phi}_t^{(n)}(\eta_{n+1}; \boldsymbol{W})) - \varphi(\mathbf{\Phi}_t^{(n)}(\widetilde{\eta}_{n+1}; \boldsymbol{W}))|^2]\mathrm{m}(\mathrm{d}\widetilde{\eta}_{n+1})\mathrm{m}(\mathrm{d}\eta_{n+1}) \\
&\le K^2 \int_{C([0,T])^2} \mathbb{E}^{\boldsymbol{W}}[\sup_{0\le t\le T} |\mathbf{\Phi}_t^{(n)}(\eta_{n+1}) - \mathbf{\Phi}_t^{(n)}(\widetilde{\eta}_{n+1})|^2]\mathrm{m}(\mathrm{d}\widetilde{\eta}_{n+1})\mathrm{m}(\mathrm{d}\eta_{n+1}) \le \frac{c^n}{n!},
\end{aligned}$$

where $\mathbb{E}^{\boldsymbol{W}}$ is the expectation with respect to $\boldsymbol{W}$ and the last inequality is verified in a similar way as in the proof of Proposition 2.1, thanks to the Lipschitz continuity (2.6) of the functional $b(\cdot)$. Similarly, there exists a constant $c > 0$ such that we have the estimate

$$\mathbb{E}\Big[\sup_{0\le t\le T,\, x\in\mathbb{R}} \Big|\mathbb{E}[\varphi(x, \overline{X}_{t,j})|\overline{X}_{s,k}, 0\le s\le T] - \mathbb{E}[\varphi(x, \overline{X}_{t,j})]\Big|^2\Big] \le \frac{c^{k-j}}{(k-j)!}; \quad k > j \tag{A.2}$$

for a Lipschitz function $\varphi(x,y) : \mathbb{R}^2 \to \mathbb{R}$ with $|\varphi(x_1,y_1) - \varphi(x_2,y_2)| \le K(|x_1 - x_2| + |y_1 - y_2|)$.

Second, note that because of the definition of $\overline{b}(\cdot,\cdot,\cdot)$ appeared in (3.28), for every $j = 1,\dots,n$,

$$\mathbb{E}[\overline{b}(s,x,\overline{X}_{s,j})] = \int_{\mathbb{R}} \widetilde{b}(s,x,z)\mathrm{m}_s(\mathrm{d}z) - \int_{\mathbb{R}} \widetilde{b}(s,x,y)\mathrm{m}_s(\mathrm{d}y) = 0\,; \quad s \ge 0,\ x \in \mathbb{R}\,.$$

Combining this observation and the Markov chain structure with (A.2), for $j < k < i$ we evaluate

$$\begin{aligned}
&\mathbb{E}[\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,j})\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})] \\
&= \mathbb{E}[\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})\mathbb{E}[\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,j})|\overline{X}_{\cdot,i},\overline{X}_{\cdot,k}]] \\
&= \mathbb{E}[\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})\mathbb{E}[\overline{b}(s,x,\overline{X}_{s,j})|\overline{X}_{\cdot,i},\overline{X}_{\cdot,k}]\big|_{x=\overline{X}_{s,i}}] \\
&= \mathbb{E}[\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})\mathbb{E}[\overline{b}(s,x,\overline{X}_{s,j})|\overline{X}_{\cdot,k}]\big|_{x=\overline{X}_{s,i}}] \\
&\le (\mathbb{E}[|\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})|^2])^{1/2}\cdot(\mathbb{E}[|\mathbb{E}[\overline{b}(s,x,\overline{X}_{s,j})|\overline{X}_{\cdot,k}]|_{x=\overline{X}_{s,i}}|^2])^{1/2} \le C\cdot\Big[\frac{c^{k-j}}{(k-j)!}\Big]^{1/2},
\end{aligned} \tag{A.3}$$

where the constant c does not depend on (s,i,j,k) and we used the Lipschitz continuity of $b(\cdot)$ and a similar technique as in the proof of Proposition 2.2 to show $\sup_{0\le s\le T}(\mathbb{E}[|\overline{b}(s,\overline{X}_{s,i},\overline{X}_{s,k})|^2])^{1/2} \le C$ for some constant C which does not depend on (i,k). This is the case $1 \le j < k < i \le n$.

For the case $i < j < k$ or the case $j < i < k$ we need the estimates

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le t\le T,\,x\in\mathbb{R}}\Big|\mathbb{E}[\varphi(x,\overline{X}_{t,k})|\overline{X}_{s,j},0\le s\le T] - \mathbb{E}[\varphi(x,\overline{X}_{t,k})]\Big|^2\Big] \\
&\quad \le \frac{c^{k-j}}{(k-j)!}; \quad k > j\,.
\end{aligned} \tag{A.4}$$

This is similar to (A.2) but the condition in the conditional expectation is reverse in discrete-time. We shall construct time-reversal of the discrete-time Markov chain structure (A.1). To do so, as in Proposition 2.1, given the marginal law $\mathrm{m}(\cdot) = \mathrm{m}^*(\cdot)$ with the marginal density function $m_t : \mathbb{R} \to \mathbb{R}_+$ at time $t \ge 0$ in the assumptions (3.21)–(3.22) of Proposition 3.3, let us consider the following system of the directed chain stochastic equation with mean-field interaction for $(Y_\cdot, X_\cdot, \widetilde{X}_\cdot)$:

$$\begin{aligned}
\mathrm{d}X_t &= \Big[u\widetilde{b}(t,X_t,\widetilde{X}_t) + (1-u)\int_{\mathbb{R}}\widetilde{b}(t,X_t,z)\widehat{\mathrm{m}}_t(\mathrm{d}z)\Big]\mathrm{d}t + \mathrm{d}B_t\,, \\
\mathrm{d}Y_t &= \Big[u\widetilde{b}(t,Y_t,X_t)\cdot\frac{m_t(Y_t)}{m_t(X_t)} + (1-u)\int_{\mathbb{R}}\widetilde{b}(t,Y_t,z)\widehat{\mathrm{m}}_t(\mathrm{d}z)\Big]\mathrm{d}t + \mathrm{d}\widehat{B}_t
\end{aligned} \tag{A.5}$$

driven by independent Brownian motions $(B_\cdot, \widehat{B}_\cdot)$, where we assume the distributional constraints

$$\begin{aligned}
&\mathrm{Law}(\widetilde{X}_\cdot) = \mathrm{Law}(X_\cdot) = \mathrm{Law}(Y_\cdot)\,, \\
&\mathrm{Law}(X_0,\widetilde{X}_0,Y_0) = \mathrm{Law}(X_0)\otimes\mathrm{Law}(\widetilde{X}_0)\otimes\mathrm{Law}(Y_0)\,,
\end{aligned} \tag{A.6}$$

and $\widehat{\mathrm{m}}_t(\cdot)$ is the marginal law, i.e.,

$$\widehat{\mathrm{m}}_t = \mathrm{Law}(Y_t) = \mathrm{Law}(X_t) = \mathrm{Law}(\widetilde{X}_t)\,; \quad t \ge 0\,, \tag{A.7}$$

with the independence relations, similar to (2.3),

$$\boldsymbol{\sigma}(\widetilde{X}_t, t\ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(B_t, t\ge 0)\,, \quad \boldsymbol{\sigma}((\widetilde{X}_t, X_t), t\ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(\widehat{B}_t, t\ge 0)\,. \tag{A.8}$$

We claim that the conditional distribution of Y_t, given X_t, is the same as the conditional distribution of $\widetilde{X}_t$, given X_t, for every $t \geq 0$, i.e.,

$$\text{Conditional Law}(Y_t \mid X_t) = \text{Conditional Law}(\widetilde{X}_t \mid X_t); \quad t \geq 0. \tag{A.9}$$

with the condition

$$\mathrm{m}_t = \text{Law}(X_t) \equiv \text{Law}(\widetilde{X}_t) \equiv \text{Law}(Y_t) = \widehat{\mathrm{m}}_t; \quad t \geq 0. \tag{A.10}$$

Indeed, thanks to (3.21)–(3.22) and the fixed point argument, by some appropriate changes in the proof of Proposition 2.1, the weak solution $(Y_\cdot, X_\cdot, \widetilde{X}_\cdot)$ to (A.5) exists with the constraints (A.6)–(A.8), and its joint law and marginal laws are uniquely determined. Since the couple $(X_\cdot, \widetilde{X}_\cdot)$ also solves the first equation in (2.1), it follows from the construction of the system (A.5) and the uniqueness of (marginal) law in Proposition 2.1 that $\text{Law}(\widetilde{X}_\cdot) = \text{Law}(X_\cdot)$ with the marginal $\mathrm{m}(\cdot) = \mathrm{m}^*(\cdot)$ and its marginal density m_t, $t \geq 0$. Thus we obtain (A.10). Moreover, as in Proposition 3.1, its joint distribution $\mathrm{M}_\cdot$ of $(X_\cdot, \widetilde{X}_\cdot)$ satisfies the integral equation (3.11) with (3.12). Similarly, the joint distribution $\widehat{\mathrm{M}}_\cdot$ of $(Y_\cdot, X_\cdot)$ satisfies the integral equation

$$\int_{\mathbb{R}} g(x)\mathrm{m}_t(\mathrm{d}x) = \int_{\mathbb{R}} g(x)\mathrm{m}_0(\mathrm{d}x) + \int_0^t [\widehat{\mathcal{A}}_s(\widehat{\mathrm{M}})g]\,\mathrm{d}s; \quad 0 \leq t \leq T, \tag{A.11}$$

similar to (3.11), for every test function $g \in C_c^2(\mathbb{R})$, where

$$\begin{aligned} \widehat{\mathcal{A}}_s(\widehat{\mathrm{M}})g &:= u\int_{\mathbb{R}^2} \widetilde{b}(s, y_1, y_2)\cdot\frac{m_s(y_1)}{m_s(y_2)} g'(y_1)\widehat{\mathrm{M}}_s(\mathrm{d}y_1\mathrm{d}y_2) \\ &\quad + (1-u)\int_{\mathbb{R}^2} \widetilde{b}(s, y_1, y_2)g'(y_1)\mathrm{m}_s(\mathrm{d}y_1)\mathrm{m}_s(\mathrm{d}y_2) \\ &\quad + \frac{1}{2}\int_{\mathbb{R}} g''(y_1)\mathrm{m}_s(\mathrm{d}y_1); \quad 0 \leq s \leq T. \end{aligned} \tag{A.12}$$

The uniqueness of solution to this integral equation (A.11) may be shown as in Lemma 10 of Oelschläger [25]. Thus, comparing (3.11) with (A.11), we obtain (A.9) from (A.10) and the time-reversible relation

$$m_s(y_1)\,\widehat{\mathrm{M}}_s(\mathrm{d}y_1\mathrm{d}y_2) = m_s(y_2)\,\mathrm{M}_s(\mathrm{d}y_1\mathrm{d}y_2); \quad 0 \leq s \leq T, \ (y_1, y_2) \in \mathbb{R}^2.$$

Thus, thanks again to the Lipschitz continuity (3.21) and linear growth condition (3.22), repeating the derivation of (A.2) but now with this reversed discrete-time Markov chain relationship (A.9), we obtain (A.4). Hence both for the cases $j < i < k$ and $i < j < k$ there exist constants c and C such that

$$\mathbb{E}[\overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j})\overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,k})] \leq C\cdot\Big[\frac{c^{k-j}}{(k-j)!}\Big]^{1/2}.$$

Therefore, we conclude (3.28), because there exist constants $c, C > 0$ such that

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbb{E}[\overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,j})\overline{b}(s, \overline{X}_{s,i}, \overline{X}_{s,k})] \\ &\quad \leq \frac{2C}{n}\sum_{j=1}^{n}\sum_{k=j}^{n}\Big[\frac{c^{k-j}}{(k-j)!}\Big]^{1/2} \leq 2C\sum_{k=0}^{\infty}\Big[\frac{c^k}{k!}\Big]^{1/2} < +\infty. \end{aligned}$$

References

[1] S.R. Athreya, R.F. Bass, E.A. Perkins, Hölder Norm estimates for elliptic operators on finite and infinite-dimensional spaces, Trans. Amer. Math. Soc 357 (2005) 5001–5029.
[2] A. Bain, D. Crisan, Fundamentals of stochastic filtering, in: Stochastic Modeling and Applied Probability, Vol. 60, Springer, 2009.
[3] A.G. Batt, G. Kallianpur, R.L. Karandikar, J. Xiong, On interacting systems of Hilbert-space-valued differential equations, Appl. Math. Optim. 37 (1998) 151–188.
[4] E. Bayraktar, A. Cosso, H. Pham, Randomized dynamic programming principle and Feynman–Kac representation for optimal control of McKean-Vlasov dynamics, Trans. Amer. Math. Soc. 370 (2018) 2115–2160.
[5] F. Bolley, A. Guillin, F. Malrieu, Trend to equilibrium and particle approximation for a weakly self consistent Vlasov-Fokker–Planck equation, ESAIM: M2AN 44 (2010) 867–884.
[6] R. Carmona, J.-P. Fouque, L.-H. Sun, Mean field games and systemic risk, Comm. in Math. Sci. 13 (4) (2015) 911–933.
[7] C. Chong, C. Klüppelberg, Partial mean field limits in heterogeneous networks. Preprint arXiv:1507.01905, 2017.
[8] G. Da Prato, J. Zabczyk, Stochastic Equations in Infinite Dimensions, Cambridge University Press, Cambridge, UK, 1992.
[9] D.A. Dawson, Stochastic evolution equations, Math. Biosci. 15 (1972) 287–316.
[10] S.N. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1985.
[11] D. Friedman, Markov Chains, Holden-Day, San Francisco, 1971.
[12] T. Funaki, A certain class of diffusion processes associated with nonlinear parabolic equations, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 67 (1984) 331–348.
[13] C. Graham, Mckean-Vlasov Ito-Skorohod equations, and nonlinear diffusions with discrete jump sets, Stoch. Process. Appl. 40 (1992) 69–82.
[14] C. Graham, S. Méléard, Stochastic particle approximations for generalized Boltzmann models and convergence estimates, Ann. Probab. 25 (1997) 115–132.
[15] M. Kac, Foundation of kinetic theory, in: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, Vol. III, University of California Press, Berkeley and Los Angeles, 1956.
[16] G. Kallianpur, J. Xiong, Stochastic Differential Equations in Infinite Dimensions, in: Lecture Notes-Monograph Series, vol. 26, Institute of Mathematical Statistics, 1995.
[17] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer-Verlag, NY, 1991.
[18] V.N. Kolokoltsov, Nonlinear Markov Processes and Kinetic Equations, in: Cambridge Tracts in Mathematics, vol. 182, Cambridge University Press, Cambridge, 2010.
[19] D. Lacker, K. Ramanan, R. Wu, Large sparse networks of interacting diffusions. Preprint arXiv:1904.02585, 2019.
[20] R.S. Liptser, A.N. Shiryayev, Statistics of Random Processes. I & II, in: Applications of Mathematics Stochastic Modeling and Applied Probability, vol. 5–6, Springer, New York, 2001.
[21] H.P. McKean J.R., An exponential formula for solving Boltzmann's equation for a Maxwellian gas, J. Combin. Theory 2 (1967) 258–372.
[22] S. Méléard, Asymptotic behavior of some interacting particle systems; McKean-Vlasov and Boltzmann models, in: Probabilistic Models for Nonlinear Partial Differential Equations, Montecatini Terme, in: Lecture Notes in Mathematics, vol. 1627, 1995, pp. 42–95.
[23] S. Mischler, C. Mouhot, Kac's program in kinetic theory, Invent. Math. 193 (2013) 1–147.
[24] S. Mischler, C. Mouhot, B. Wennberg, A new approach to quantitative propagation of chaos for drift diffusion and jump processes, Probab. Theory Related Fields 161 (2015) 1–59.
[25] K. Oelschläger, A Martingale approach to the law of large numbers for weakly interacting stochastic processes, Ann. Probab. 2 (1984) 458–479.
[26] A.S. Sznitman, Nonlinear reflecting diffusion processes and the propagation of chaos and fluctuations associated, J. Funct. Anal. 56 (1984) 311–336.
[27] A.S. Sznitman, Topics in propagation of chaos, in: École d'Été de Probabilités de Saint-Flour XIX – 1989, in: Lecture Notes in Math., vol. 1464, Springer Berlin, 1991, pp. 165–251.
[28] H. Tanaka, Probabilistic treatment of the Boltzmann equation of Maxwellian molecules, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 46 (1978) 67–105.