Electron. J. Probab. **29** (2024), article no. 127, 1–28.
ISSN: 1083-6489 https://doi.org/10.1214/24-EJP1192

Smoothness of directed chain stochastic differential equations*

Tomoyuki Ichiba† Ming Min‡

Abstract

We study the smoothness of the solution of the directed chain stochastic differential equations, where each process is affected by its neighborhood process in an infinite directed chain graph, introduced by Detering et al. (2020). Because of the auxiliary process in the chain-like structure, classic methods of Malliavin derivatives are not directly applicable. Namely, we cannot make a connection between the Malliavin derivative and the first order derivative of the state process. It turns out that the partial Malliavin derivatives can be used here to fix this problem.

Keywords: directed chain SDEs; smoothness; partial Malliavin calculus; diffusions on graph.
MSC2020 subject classifications: 60H07; 60H30; 60K35.
Submitted to EJP on December 30, 2022, final version accepted on August 23, 2024.

1 Introduction

The main objective of this paper is to study the existence and regularity of the densities of the directed chain stochastic differential equations: given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, the directed chain McKean-Vlasov stochastic differential equation (or directed chain SDE for short) for a pair $(X^\theta_\cdot, \widetilde{X}_\cdot)$ of N-dimensional stochastic processes considered here is of the form

$$X^\theta_t = \theta + \int_0^t V_0(s, X^\theta_s, \mathrm{Law}(X^\theta_s), \widetilde{X}_s)\,\mathrm{d}s + \sum_{i=1}^d \int_0^t V_i(s, X^\theta_s, \mathrm{Law}(X^\theta_s), \widetilde{X}_s)\,\mathrm{d}B^i_s, \tag{1.1}$$

for $t \geq 0$ with the distributional constraint

$$[X^\theta_t, t \geq 0] := \mathrm{Law}(X^\theta_t, t \geq 0) = \mathrm{Law}(\widetilde{X}_t, t \geq 0) =: [\widetilde{X}_t, t \geq 0],$$

*A part of the research was supported by the National Science Foundation grant NSF DMS-2008427.
†Department of Statistics and Applied Probability, South Hall, University of California, Santa Barbara, CA 93106, USA. E-mail: ichiba@pstat.ucsb.edu
‡Department of Statistics and Applied Probability, South Hall, University of California, Santa Barbara, CA 93106, USA. E-mail: m_min@pstat.ucsb.edu

where V_i, $i = 0, 1, \ldots, d$ are some smooth coefficients, $B_\cdot := (B_\cdot^1, \ldots, B_\cdot^d)$ is a standard d-dimensional Brownian motion independent of the initial state $X_0^\theta = \theta$ and of $\widetilde{X}_\cdot$, and X_0^θ is independent of $\widetilde{X}_0$. Throughout the paper, $[\xi]$ denotes the law of a generic random element ξ. Here, each coefficient V_i in (1.1) depends on time s, the value X_s^θ, its law $\mathrm{Law}(X_s^\theta) =: [X_s^\theta]$ and the other $\widetilde{X}_s$ of the pair for $s \geq 0$. The law $[X_\cdot^\theta]$ depends on the law $[\widetilde{X}_\cdot]$ through (1.1) and they are the same marginal law. We show that the above directed chain SDE has a unique weak solution in section 2.

This kind of directed chain structure was first proposed by [13] in a simpler form. Schematically, this structure can be written as an infinite chain of stochastic equations for $(X_{1,\cdot}, X_{2,\cdot}, \ldots)$,

$$\begin{aligned} \mathrm{d}X_{1,t} &= b(t, X_{1,t}, F_{1,t})\,\mathrm{d}t + \mathrm{d}B_{1,t}, \\ \mathrm{d}X_{2,t} &= b(t, X_{2,t}, F_{2,t})\,\mathrm{d}t + \mathrm{d}B_{2,t} \\ &\vdots \\ \mathrm{d}X_{i,t} &= b(t, X_{i,t}, F_{i,t})\,\mathrm{d}t + \mathrm{d}B_{i,t}, \\ &\vdots \end{aligned} \tag{1.2}$$

where $F_{i,t} := u\delta_{X_{i+1,t}} + (1-u)\mu_{i,t}$ is the mixture distribution term of the measure-dependent drift coefficient b with the marginal law $\mu_{i,t} := \mathrm{Law}(X_{i,t})$ of $X_{i,t}$ for $t \geq 0$, $\delta_{X_{i+1,t}}$ is the Dirac measure at $X_{i+1,t}$, a fixed constant $u \in [0,1]$ measures the common amount of dependency of $X_{i,\cdot}$ on its neighborhood value $X_{i+1,\cdot}$, and $B_{1,\cdot}, B_{2,\cdot,\ldots}$ are independent standard Brownian motions. We also assume that the initial value $X_{i,0}$ is independent of $B_{i,\cdot}$, and $X_{i+1,\cdot}$ and $B_{i,\cdot}$ are independent for $i = 1, 2, \ldots$. In particular, the drift b in [13] has the following form $b(t,x,\mu) := \int_{\mathbb{R}} \widetilde{b}(t,x,y)\mu(\mathrm{d}y)$ with some Lipschitz continuous function $\widetilde{b}$. See also Figure 1 in section 4.

The stochastic processes on infinite graphs including the directed chain structure have drawn much attention recently (e.g., see [28] for their application to the generative adversarial network). Stochastic Differential Games on the directed chain have been studied in [14] and on the extended version of random directed networks in [15] as well as on the general random graph (e.g., [25]). [24] discuss the Markov random field property over both finite and countably infinite graphs with local interactions through the drift coefficients. Another related topic is the Graphon particle system. There are a sequence of works on Graphon particle system and Graphon Mean Field Games, [1, 4, 5] just to name a few. [2] introduced the uniform-in-time exponential concentration bounds related to the graphon particle system and its finite particle approximations. Here, we are interested in the existence and smoothness of the density of the directed chain SDE (2.1)-(2.2). It should be emphasized here that in this problem, we need notions of derivatives in the space of measures, which is used frequently in the theory of Mean Field Games.

In most cases, Malliavin calculus is a foundation for analyzing the smoothness of the density of stochastic differential equations. It has been widely used in investigating the density of diffusions [20], [22], [23] and then applied into many different scenarios. The authors in [8] use Malliavin calculus to derive smoothing properties of solution to stochastic differential equations with jumps. The smoothing properties of McKean-Vlasov SDEs have been studied in [10], which is closely related to our purpose. However, because of the appearance of the auxiliary process $\widetilde{X}$, the crucial step making connections between the Malliavin derivative and the first-order derivative of the state process fails, please see Question 3.3 for the details. To our best knowledge, we did not find any work studying the smoothness property of such weak solutions of stochastic differential equations.

For the purpose of resolving this problem and using the Malliavin derivatives, we should freeze the auxiliary process $\widetilde{X}$. This inspired us to consider another closely related and well-developed tool, the partial Malliavin calculus. Partial Malliavin calculus is first introduced by [18] for the constant case, where the projections are taken on a fixed Hilbert subspace, and applied to prove some regularity results in non-linear filtering theory. Another work developing the partial Malliavin calculus is [16], by which the authors were able to complete the proof of some results in [27] on the long-time asymptotic of stochastic oscillatory integrals. We mainly adopt the framework from a later work by Nualart and Zakai [30], where the projection is taken on a family of subspaces which are defined as the orthogonal complement to the subspaces generated by $\widetilde{X}$ in (1.1). We remark that our method is potentially applied to analyze the smoothness of the (conditional) law of weak solutions of stochastic differential equations in a similar setting. For example, the study of nonlinear filtering problem and the corresponding stochastic partial differential equations.

This paper is structured as follows. In section 2, we first introduce the differentiation in the space of measures and multi-index notation in section 2.1, and then prove the existence, uniqueness and some regularity results on the solutions of generalized directed chain SDEs in Propositions 2.2-2.3. In section 3, we prepare the notions of the partial Malliavin calculus and give the Kusuoka-Stroock process for the proof of our smoothness of densities, which will be stated in section 4. Our proofs follow the idea of [10], where we first derive integration by parts formulae for the directed chain SDEs via the partial Malliavin derivatives, instead of the Malliavin derivatives, as in [10]. The main result is stated in Theorem 4.11 with some applications in section 4.

2 Preliminaries and directed chain SDEs

In this section, we first prepare some notations and the notion of differentiation in $\mathcal{P}_2(\mathbb{R}^N)$, where $\mathcal{P}_2(\mathbb{R}^N)$ is the space of all measures on $\mathbb{R}^N$ with finite second moments, and then establish the weak solutions of the directed chain SDEs.

2.1 Notations and basic setup

Rather than the directed chain SDE of the type given in [13], we consider the SDE in a more general setup, allowing the diffusion coefficients non-constant. Given a probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, the directed chain McKean-Vlasov SDE (or directed chain SDE for short) is of the form

$$X_t^\theta = \theta + \int_0^t V_0(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\,\mathrm{d}s + \sum_{i=1}^d \int_0^t V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\,\mathrm{d}B_s^i, \tag{2.1}$$

$$\text{with the constraint} \quad [X_t^\theta, t \geq 0] = [\widetilde{X}_t, t \geq 0], \tag{2.2}$$

where $B_s := (B_s^1, \ldots, B_s^d)$ is a standard d dimensional Brownian motion and $\widetilde{X}_s \in L^2(\Omega \times [0,T], \mathbb{R}^N)$ is an adapted random process independent of all the Brownian motions $B^i, i = 1, \ldots, d$ and initial state θ.

Moreover, we assume that $V_0, V_i : [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^N$, where $\mathcal{P}_2(\mathbb{R}^N)$ is the set of measures on $\mathbb{R}^N$ with finite second moments. We equip $\mathcal{P}_2(\mathbb{R}^N)$ with the 2-Wasserstein metric, W_2. For a general metric space (M, d), we define the 2-Wasserstein metric on $\mathcal{P}_2(M)$ by

$$W_2(\mu,\nu) = \inf_{\Pi \in \mathcal{P}_{\mu,\nu}} \left(\int_{M\times M} d(x,y)^2 \Pi(\,\mathrm{d}x, \,\mathrm{d}y) \right)^{1/2},$$

where $\mathcal{P}_{\mu,\nu}$ denotes the class of measures on $M \times M$ with marginals μ and ν.

We denote L^p norm on $(\Omega, \mathcal{F}, \mathbb{P})$ by $\|\cdot\|_p$, $p \geq 1$ and for every $t \geq 0$, we also introduce the space $\mathcal{S}_t^p$ of continuous $\mathbb{F}$ adapted process φ on $[0,t]$, satisfying

$$\|\varphi\|_{\mathcal{S}_t^p} = \big(\mathbb{E} \sup_{s\in[0,t]} |\varphi_s|^p\big)^{1/p} < \infty.$$

Let us introduce more notations in accordance with [10]. We will write $[\theta] = \delta_x$ if the initial state of this SDE is a fixed real vector $x \in \mathbb{R}^N$. We use $\mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}(\mathbb{R}^+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$ for the class of functions, with a precise definition given in Definition 2.1, where the notion of derivatives with respect to measure is adopted from P.-L. Lions' lecture notes at the *Collège de France*, recorded in a set of notes [6], very well exposed in [7] and also adopted by [10].

Differentiability in $\mathcal{P}_2(\mathbb{R}^N)$ Lion's notion of differentiability with respect to measure of functions $U : \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}$ is to define a lifted function U' on the Hilbert space $L^2(\Omega'; \mathbb{R}^N)$ over probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, where Ω' is a Polish space and $\mathbb{P}'$ is an atomless measure, such that $U'(X') = U([X'])$ for $X' \in L^2(\Omega'; \mathbb{R}^N)$ and $[X'] = [X]$. Thus, we are able to express the derivative of U w.r.t. measure $\mu = [X]$ term as the Fréchet derivative of U' w.r.t. X' whenever it exists, which can be written as an element of $L^2(\Omega'; \mathbb{R}^N)$ by identifying $L^2(\Omega'; \mathbb{R}^N)$ and its dual. This gradient in a direction $\gamma' \in L^2(\Omega'; \mathbb{R}^N)$ is given by

$$\mathcal{D}U'(X')(\gamma') = \langle \mathcal{D}U'(X'), \gamma' \rangle = \mathbb{E}'\big[\mathcal{D}U'(X') \cdot \gamma'\big],$$

where $\mathbb{E}'$ is the expectation under $\mathbb{P}'$. By [6, Theorem 6.2], the distribution of this gradient depends only on the measure μ, exists uniquely and can be written as

$$\partial_\mu U(\mu, X') := \mathcal{D}U'(X') = \xi(X') \in L^2(\Omega'; \mathbb{R}^N).$$

This definition of the derivative with respect to measure can be extended to higher orders by thinking of $\partial_\mu U(\mu, \cdot) : \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^N$ as a function, and the derivative is well defined for each of its components as in the following. For each $\mu \in \mathcal{P}_2(\mathbb{R}^N)$, there exists a unique version of such function $\partial_\mu U(\mu, \cdot)$ which is assumed to be a priori continuous (see the discussion in [10]).

Multi-index To get a more general result, we extend the derivatives to higher order. For a function $f : \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}^N$, we can apply the above discussion straightforwardly to each component $f = (f^1, \dots, f^N)$. Then the derivatives $\partial_\mu f^i, 1 \leq i \leq N$ takes values in $\mathbb{R}^N$, and we denote $(\partial_\mu f^i)_j : \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ for $j = 1, \dots, N$. For a fixed $v \in \mathbb{R}^N$, we are able to differentiate $\mathcal{P}_2 \ni \mu \mapsto (\partial_\mu f^i)_j(\mu, v) \in \mathbb{R}$ again to get the second order derivative. If the derivative of this mapping exists and there is a continuous version of

$$\mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \times \mathbb{R}^N \ni (\mu, v_1, v_2) \mapsto \partial_\mu (\partial_\mu f^i)_j(\mu, v_1, v_2) \in \mathbb{R}^N,$$

then it is unique. It is natural to have a multi-index notation $\partial_\mu^{(j,k)} f^i := (\partial_\mu (\partial_\mu f^i)_j)_k$ to ease the notation. Similarly, for higher derivatives, if for each $(i_0, \dots, i_n) \in \{1, \dots, N\}^{n+1}$,

$$\underbrace{\partial_\mu (\partial_\mu \dots (\partial_\mu}_{n \text{ times}} f^{i_0})_{i_1} \dots)_{i_n}$$

exists, we denote this $\partial_\mu^\alpha f^{i_0}$ with $\alpha = (i_1, \dots, i_n)$ and $|\alpha| = n$. Each derivative in μ is a function of an extra variable with $\partial_\mu^\alpha f^{i_0} : \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^n \to \mathbb{R}$. We always denote these variables, by $v_1, \dots, v_n$, i.e.,

$$\mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^n \ni (\mu, v_1, \dots, v_n) \mapsto \partial_\mu^\alpha f^{i_0}(\mu, v_1, \dots, v_n) \in \mathbb{R}.$$

When there is no confusion, we will abbreviate $(v_1, \dots, v_n)$ to $\boldsymbol{v} \in (\mathbb{R}^N)^n$, so that

$$\partial_\mu^\alpha f^{i_0}(\mu, \boldsymbol{v}) = \partial_\mu^\alpha f^{i_0}(\mu, v_1, \dots, v_n),$$

and use notation

$$|\boldsymbol{v}| := |v_1| + \cdots + |v_n|,$$

with $|\cdot|$ the Euclidean norm on $\mathbb{R}^N$. Then it makes sense to discuss the derivatives of the function $\partial_\mu^\alpha f^{i_0}$ with respect to the variables $v_1, \dots, v_n$.

If, for some $j \in \{1, \dots, N\}$ and all $(\mu, v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n) \in \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^{n-1}$,

$$\mathbb{R}^N \ni v_j \mapsto \partial_\mu^\alpha f^{i_0}(\mu, v_1, \dots, v_n) \in \mathbb{R}$$

is l-times continuously differentiable, we denote the derivatives $\partial_{v_j}^{\beta_j} \partial_\mu^\alpha f^{i_0}$, for β_j a multi-index on $\{1, \dots, N\}$ with $|\beta_j| \le l$. Similar to the above, we will denote by $\boldsymbol{\beta}$ the n-tuple of multi-indices $(\beta_1, \dots, \beta_n)$. We also associate a length to $\boldsymbol{\beta}$ by

$$|\boldsymbol{\beta}| := |\beta_1| + \cdots + |\beta_n|,$$

and denote $\#\boldsymbol{\beta} := n$. Then we denote by $\mathcal{B}_n$ the collection of all such $\boldsymbol{\beta}$ with $\#\boldsymbol{\beta} = n$, and $\mathcal{B} := \cup_{n \ge 1} \mathcal{B}_n$. Again, to lighten the notation, we use

$$\partial_{\boldsymbol{v}}^{\boldsymbol{\beta}} \partial_\mu^\alpha f^i(\mu, \boldsymbol{v}) := \partial_{v_n}^{\beta_n} \cdots \partial_{v_1}^{\beta_1} \partial_\mu^\alpha f^i(\mu, v_1, \dots, v_n).$$

The coefficients $V_0, \dots, V_d : [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^N$ depend on a time variable, two Euclidean variables as well as the measure variable. So whether the order of taking derivatives matters is a question. Fortunately, a result from [3, Lemma 4.1] tells us that derivatives commute when the mixed derivatives are Lipschitz continuous. However, it should be emphasized that we could not interchange the order of ∂_μ and ∂_v, since the coefficients would not depend on the extra variable $\boldsymbol{v}$ before taking derivatives with respect to measure.

Definition 2.1 ($\mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}$)**.** *We have the following definitions:*

(a) We use $\partial_x, \tilde{\partial}$ to denote the derivative with respect to the second and fourth Euclidean variables in V_0, V_i's, respectively.

(b) Let $V : \mathbb{R}^+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^N$ with components $V^1, \dots, V^N : \mathbb{R}^+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$. We say $V \in \mathcal{C}_{b,\mathrm{Lip}}^{1,1,1}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$ if the following is true: for each $i = 1, \dots, N$, $\partial_\mu V^i$, $\partial_x V^i$ and $\tilde{\partial} V^i$ exist. Moreover, assume the boundedness of the derivatives for all $(t, x, \mu, y, v) \in [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \times \mathbb{R}^N$,

$$|\partial_x V^i(t, x, \mu, y)| + |\tilde{\partial} V^i(t, x, \mu, y)| + |\partial_\mu V^i(t, x, \mu, y, v)| \le C.$$

In addition, suppose that $\partial_\mu V^i$, $\partial_x V^i$ and $\tilde{\partial} V^i$ are all Lipschitz in the sense that for all $(t, x, \mu, y, v), (t, x', \mu', y', v') \in [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \times \mathbb{R}^N$,

$$\begin{aligned}
\left|\partial_\mu V^i(t, x, \mu, y, v) - \partial_\mu V^i(t, x', \mu', y', v')\right| &\le C(|x - x'| + |y - y'| + |v - v'| \\
&\quad + W_2(\mu, \mu')), \\
\left|\partial_x V^i(t, x, \mu, y) - \partial_x V^i(t, x', \mu', y')\right| &\le C(|x - x'| + |y - y'| + W_2(\mu, \mu')), \\
\left|\tilde{\partial} V^i(t, x, \mu, y) - \tilde{\partial} V^i(t, x', \mu', y')\right| &\le C(|x - x'| + |y - y'| + W_2(\mu, \mu')).
\end{aligned}$$

(c) We write $V \in \mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$, if the following holds true: for each $i = 1, \dots, N$, and all multi-indices α, $\tilde{\gamma}$ and γ on $\{1, \dots, N\}$ and all $\boldsymbol{\beta} \in \mathcal{B}$ satisfying $|\alpha| + |\boldsymbol{\beta}| + |\gamma| + |\tilde{\gamma}| \le k$, the derivative

$$\partial_x^\gamma \tilde{\partial}^{\tilde{\gamma}} \partial_{\boldsymbol{v}}^{\boldsymbol{\beta}} \partial_\mu^\alpha V^i(t, x, \mu, y, \boldsymbol{v})$$

exists and is bounded, Lipschitz continuous with respect to x, μ, y and ν uniformly in $t \in [0,T]$.

(d) We write $h \in \mathcal{C}^{k,k}_{b,\mathrm{Lip}}([0,T]\times\mathbb{R}^N\times\mathbb{R}^N;\mathbb{R}^N)$, if the mapping h does not depend on a measure variable and all the other conditions are satisfied in (c).

2.2 Solutions of directed chain SDEs

The existence and uniqueness of weak solutions of directed chain SDEs are given in Proposition 2.2. The constraint (2.2) plays an essential role here in governing the uniqueness.

Proposition 2.2. *Suppose that $V_i, i = 0, 1, \dots, d$ are Lipschitz in the sense that for every $T > 0$, there exists a constant C_T such that*

$$\sup_i |V_i(t,x_1,\mu_1,y_1) - V_i(t,x_2,\mu_2,y_2)| \le C_T(|x_1-x_2| + |y_1-y_2| + W_2(\mu_1,\mu_2)), \quad 0 \le t \le T. \tag{2.3}$$

With the same constant C_T, let us also assume that V_i's have at most linear growth, i.e.

$$\sup_{0\le t\le T} |V_i(t,x,\mu,y)| \le C_T(1 + |x| + |y| + W_2(\mu,\mu_0)) \tag{2.4}$$

where $\mu_0 \in \mathcal{P}_2(\mathbb{R}^N)$ is fixed. Then there exists a unique weak solution to the stochastic differential equation of the directed chain (2.1)-(2.2).

The proof is similar to the proof for [13, Proposition 2.1] with a small generalization. Due to the appearance of the neighborhood process, we cannot expect a strong solution of the directed chain SDEs (2.1) (*cf.* Proposition 2.1 of [13]).

Proof. We shall evaluate the Wasserstein distance between two probability measures μ_1, μ_2 on the space $C([0,T],\mathbb{R}^N)$ of continuous functions, namely

$$D_t(\mu_1,\mu_2) := \inf\left\{ \int \Big(\sup_{0\le s\le t} |X_s(\omega_1) - X_s(\omega_2)|^2 \wedge 1 \Big)\, \mathrm{d}\mu(\omega_1,\omega_2) \right\}^{1/2} \tag{2.5}$$

for $0 \le t \le T$, where the infimum is taken over all the joint measure μ on $C([0,T],\mathbb{R}^N) \times C([0,T],\mathbb{R}^N)$ such that their marginals are μ_1, μ_2, and the initial joint distribution is the product measure $\theta \otimes \theta$, the initial marginals are θ. Here, $X_s(\omega) = \omega(s), 0 \le s \le T$ is the coordinate map of $\omega \in C([0,T],\mathbb{R}^N)$. $D_T(\cdot,\cdot)$ defines a complete metric on $\mathcal{M}(C([0,T],\mathbb{R}^N))$, which gives the weak topology to it.

Given the distribution $m = \mathrm{Law}(\widetilde{X}) \in \mathcal{M}(C([0,T],\mathbb{R}^N))$ of $\widetilde{X}$ that is independent of B and X_0, it is well known that the following stochastic differential equation

$$\mathrm{d}X^m_t = V_0(t, X^m_t, m_t, \widetilde{X}_t)\,\mathrm{d}t + \sum_{i=1}^d V_i(t, X^m_t, m_t, \widetilde{X}_t)\,\mathrm{d}B^i_t \tag{2.6}$$

has a unique solution, based on the Lipschitz and linear growth condition on coefficients, thanks to the iteration scheme for the stochastic equation with random coefficients $V_i(t,\cdot,m_t,\widetilde{X}_t)$, $i = 0,\dots,d$, $0 \le t \le T$. Note that since $\widetilde{X}$ is independent of Brownian motion B, we can only expect the solution to exist in the weak sense.

Define a map $\Phi : \mathcal{M}(C([0,T],\mathbb{R}^N)) \to \mathcal{M}(C([0,T],\mathbb{R}^N))$ by $\Phi(m) := \mathrm{Law}(X^m_\cdot)$. We shall find a fixed point m^* for the map Φ such that $\Phi(m^*) = m^*$ to show the uniqueness of the solution in the weak sense.

Assume $m_1 = \mathrm{Law}(\widetilde{X}^1)$ and $m_2 = \mathrm{Law}(\widetilde{X}^2)$, then by rewriting (2.6) we have

$$X^{m_i}_t = \theta + \int_0^t V_0(t, X^{m_i}_t, m_{i,t}, \widetilde{X}^i_t)\,\mathrm{d}s + \sum_{i=1}^d \int_0^t V_i(t, X^{m_i}_t, m_{i,t}, \widetilde{X}^i_t)\,\mathrm{d}B^i_s, \quad i = 1,2.$$

Note that here we fix the initial state to be the same θ for both X^{m_1} and X^{m_2}. Let m be a joint distribution of m_1, m_2 and $\mathbb{E}^m$ be the expectation under m. Then

$$\begin{aligned}
&\mathbb{E}^m\big[\sup_{0\le s\le t}(X_s^{m_1}-X_s^{m_2})^2\big]\\
&\le 2\mathbb{E}^m\bigg[\sup_{0\le s\le t}\int_0^s\big(V_0(v,X_v^{m_1},m_{1,v},\widetilde{X}_v^1)-V_0(v,X_v^{m_2},m_{2,v},\widetilde{X}_v^2)\big)^2\,\mathrm{d}v\bigg]\\
&\quad+2^d\sum_{i=1}^d\mathbb{E}^m\bigg[\sup_{0\le s\le t}\int_0^s\big(V_i(v,X_v^{m_1},m_{1,v},\widetilde{X}_v^1)-V_i(v,X_v^{m_2},m_{2,v},\widetilde{X}_v^2)\big)^2\,\mathrm{d}v\bigg]\\
&\le 2^{d+3}(d+1)C_T^2\mathbb{E}^m\bigg[\sup_{0\le s\le t}\int_0^s\big((X_v^{m_1}-X_v^{m_2})^2+W_2(m_{1,v},m_{2,v})^2+(\widetilde{X}_v^1-\widetilde{X}_v^2)^2\big)\wedge 1\,\mathrm{d}v\bigg]\\
&\le C\cdot\mathbb{E}^m\bigg[\int_0^t\sup_{0\le v\le s}(X_v^{m_1}-X_v^{m_2})^2\wedge 1\,\mathrm{d}s\bigg]+C\int_0^t W_2(m_{1,s},m_{2,s})^2\wedge 1\,\mathrm{d}s\\
&\quad+C\cdot\mathbb{E}^m\bigg[\int_0^t\sup_{0\le v\le s}(\widetilde{X}_v^1-\widetilde{X}_v^2)^2\wedge 1\,\mathrm{d}s\bigg]\\
&=C\int_0^t\mathbb{E}^m\big[\sup_{0\le v\le s}(X_v^{m_1}-X_v^{m_2})^2\wedge 1\big]\,\mathrm{d}s+C\int_0^t W_2(m_{1,s},m_{2,s})^2\wedge 1\,\mathrm{d}s\\
&\quad+C\int_0^t\mathbb{E}^m\big[\sup_{0\le v\le s}(\widetilde{X}_v^1-\widetilde{X}_v^2)^2\wedge 1\big]\,\mathrm{d}s \qquad (2.7)
\end{aligned}$$

where we replace $2^{d+3}(d+1)C_T$ by C. Note that by construction,

$$W_2(m_{1,s},m_{2,s})^2\wedge 1\le D_s(m_1,m_2)^2.$$

By taking infimum over all m such that its marginals are m_1, m_2, the third term in (2.7) is bounded by

$$C\int_0^t D_s(m_1,m_2)^2\,\mathrm{d}s.$$

Hence we get

$$D_t(\Phi(m_1),\Phi(m_2))^2\le C\int_0^t D_s(\Phi(m_1),\Phi(m_2))^2\,\mathrm{d}s+2C\int_0^t D_s(m_1,m_2)^2\,\mathrm{d}s.$$

Then by applying Gronwall's lemma, we get

$$D_t(\Phi(m_1),\Phi(m_2))^2\le 2Ce^{CT}\int_0^t D_s(m_1,m_2)^2\,\mathrm{d}s. \qquad (2.8)$$

For every $m\in\mathcal{M}(C([0,T],\mathbb{R}^N))$, let $m_1=m$, $m_2=\Phi(m)$, we get by iterating (2.8),

$$D_T(\Phi^{(k+1)}(m),\Phi^{(k)}(m))\le\sqrt{\frac{(2CTe^{CT})^k}{k!}}D_T(\Phi(m),m),\quad\forall k\in\mathbb{N}. \qquad (2.9)$$

This implies that $\{\Phi^{(k)}(m),k\in\mathbb{N}\}$ forms a Cauchy sequence converging to a fixed point m^*. This m^* is the weak solution to directed chain SDE (2.1)-(2.2). □

Proposition 2.3 (Regularity)**.** *If $\theta\in L^2(\Omega)$, the solution of directed chain SDE* (2.1)-(2.2) *satisfies*

$$\|X^\theta\|_{\mathcal{S}_T^2}\le C(1+\|\theta\|_2),$$

where $C=C(T)$, under the assumption of Proposition 2.2.

Proof. The proof follows from a similar procedures as [13, Proposition 2.2]. □

2.3 Flow property

In the last part of this section, we discuss the flow property of directed SDEs informally. After establishing the weak solution to the exact directed chain SDE (2.1), we also consider the auxiliary process $X_\cdot^{x,[\theta]}$ that satisfies

$$X_\cdot^{x,[\theta]} = x + \int_0^\cdot V_0(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s)\,\mathrm{d}s + \sum_{i=1}^d \int_0^\cdot V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s)\,\mathrm{d}B_s^i, \tag{2.10}$$

where $x \in \mathbb{R}^N$ is a fixed initial point and $\widetilde{X}_\cdot$ is the neighborhood process satisfying the constraints (2.2), i.e., $\mathrm{Law}(\widetilde{X}_\cdot) = \mathrm{Law}(X_\cdot^\theta) = [X_\cdot^\theta]$. Note that $X_\cdot^{x,[\theta]}$ in (2.10), driven by the same Brownian motion B, is solvable with pathwise uniqueness, given the (fixed) unique, weak solution $(X_\cdot^\theta, \widetilde{X}_\cdot, B_\cdot)$ in Proposition 2.2, for each $x \in \mathbb{R}^N$. Thus we consider the weak solution $(X_\cdot^{x,[\theta]}, X_\cdot^\theta, \widetilde{X}_\cdot, B_\cdot)$ for the system (2.1) and (2.10).

Proposition 2.4 (Regularity)**.** *Under the assumption in Proposition 2.2, for every $\theta \in L^2(\Omega)$, $T > 0$ and $p \geq 2$, there exists a constant $C = C(T,p)$ such that the solution of* (2.10) *satisfies*

$$\|X^{x,[\theta]}\|_{\mathcal{S}_T^p} \leq C(1 + \|\theta\|_2 + |x|).$$

Proof. The proof follows from the Burkholder-Davis-Gundy inequality and Proposition 2.3, which is also satisfied by $\widetilde{X}$. □

For the explanation purpose, we will add a superscript $\tilde{\theta}$ such that $X_t^{x,[\theta],\tilde{\theta}} := X_t^{x,[\theta]}$ and $\widetilde{X}_t^{\tilde{\theta}} := \widetilde{X}_t$ to emphasize the neighborhood process start at $\tilde{\theta}$, independent of θ. This notation is only used in this subsection. Thus, with the notation $B_t^0 \equiv t$, $t \geq 0$, (2.10) is read as

$$X_t^{x,[\theta],\tilde{\theta}} = x + \sum_{i=0}^d \int_0^t V_i(s, X_s^{x,[\theta],\tilde{\theta}}, [X_s^\theta], \widetilde{X}_s^{\tilde{\theta}})\,\mathrm{d}B_s^i, \quad t \geq 0. \tag{2.11}$$

For different initial points x, x' and the corresponding solutions $X_\cdot^{x,[\theta],\tilde{\theta}}$ and $X_\cdot^{x',[\theta],\tilde{\theta}}$, we have the following estimate: there exists a constant $C > 0$ such that

$$\mathbb{E}\big[\sup_{t \leq s \leq T} \big|X_s^{x,[\theta],\tilde{\theta}} - X_s^{x',[\theta],\tilde{\theta}}\big|^2\big] \leq C|x - x'|^2$$

again by the Lipschitz continuity and the Burkholder-Davis-Gundy inequality. By the pathwise uniqueness of $X_\cdot^{x,[\theta],\theta}$, given the pair $(X_\cdot^\theta, \widetilde{X}_\cdot^{\tilde{\theta}})$, it follows

$$X_s^{x,[\theta],\tilde{\theta}}\Big|_{x=\theta} = X_s^\theta, \quad 0 \leq s \leq T. \tag{2.12}$$

Now, with some abuse of notations, we denote by $X_\cdot^{t,x,[\theta],\tilde{\theta}}$ the solution to (2.11) with $X_t^{t,x,[\theta],\tilde{\theta}} = x$, denote by $(X_\cdot^{t,\theta}, \widetilde{X}_\cdot^{t,\tilde{\theta}})$ the solution to (2.1) with $(X_t^{t,\theta}, \widetilde{X}_t^{t,\tilde{\theta}}) = (\theta, \tilde{\theta})$. It follows from (2.12) that by the strong Markov property, for $0 \leq t \leq s \leq r \leq T$, we have the flow property

$$(X_r^{s,X_s^{t,x,[\theta],\tilde{\theta}},[X_s^{t,\theta}],\widetilde{X}_s^{t,\tilde{\theta}}}, X_r^{s,X_s^{t,\theta}}, \widetilde{X}_r^{s,\widetilde{X}_s^{t,\tilde{\theta}}}) = (X_r^{t,x,[\theta],\tilde{\theta}}, X_r^{t,\theta}, \widetilde{X}_r^{t,\tilde{\theta}}). \tag{2.13}$$

We close section 2 at this point. After the introduction of the partial Malliavin derivatives, we will revisit the directed chain SDE and study the regularities of its derivatives.

3 Partial Malliavin calculus

In this section, we will briefly review the Malliavin calculus, following [29], and introduce the partial Malliavin derivatives for our problem.

Malliavin Calculus Let $H := L^2([0,T],\mathbb{R}^d)$ be the Hilbert space, where we define Gaussian process, and $\mathcal{S}$ be the set of smooth functionals of the form

$$F(\omega) = f\bigg(\int_0^T h_1(t)\cdot \mathrm{d}B_t(\omega),\dots,\int_0^T h_n(t)\cdot \mathrm{d}B_t(\omega)\bigg),$$

where $f\in\mathcal{C}_p^\infty(\mathbb{R}^n;\mathbb{R})$, $h_1,\dots,h_n\in H$ and $\int_0^T h_i(t)\cdot \mathrm{d}B_t = \sum_{j=1}^d \int_0^T h_i^j(t)\,\mathrm{d}B_t^j$.

Then the Malliavin derivative of F, denoted by $\boldsymbol{D}F\in L^2(\Omega;H)$ is given by:

$$\boldsymbol{D}F = \sum_{i=1}^n \partial^i f\bigg(\int_0^T h_1(t)\cdot \mathrm{d}B_t(\omega),\dots,\int_0^T h_n(t)\cdot \mathrm{d}B_t(\omega)\bigg)h_i. \tag{3.1}$$

As stated in [29], because of the isometry $L^2(\Omega\times[0,T];\mathbb{R}^d)\simeq L^2(\Omega;H)$, we are able to identify $\boldsymbol{D}F$ with a process $(\boldsymbol{D}_rF)_{r\in[0,T]}$ taking values in $\mathbb{R}^d$. Moreover, the set of smooth functionals, denoted by $\mathcal{S}$, is dense in $L^p(\Omega)$ for any $p\geq 1$ and $\boldsymbol{D}$ is closable as an operator from $L^p(\Omega)$ to $L^p(\Omega;H)$. We define $\mathbb{D}^{1,p}$ as the closure of the set $\mathcal{S}$ within $L^p(\Omega;\mathbb{R}^d)$ with respect to the norm

$$\|F\|_{\mathbb{D}^{1,p}} = \big(\mathbb{E}|F|^p + \mathbb{E}\|\boldsymbol{D}F\|_H^p\big)^{\frac{1}{p}}.$$

The higher order Malliavin derivatives are defined similarly, denoted by $\boldsymbol{D}^{(k)}F$, which is a random variable with values in $H^{\otimes k}$ defined as

$$\boldsymbol{D}^{(k)}F := \sum_{i_1,\dots,i_k=1}^n \partial^{(i_1,\dots,i_k)} f\bigg(\int_0^T h_1(t)\cdot \mathrm{d}B_t(\omega),\dots,\int_0^T h_n(t)\cdot \mathrm{d}B_t(\omega)\bigg)h_{i_1}\otimes\cdots\otimes h_{i_k}.$$

We define $\mathbb{D}^{k,p}$ to be the closure of the set of smooth functions $\mathcal{S}$ with respect to the norm:

$$\|F\|_{\mathbb{D}^{k,p}} = \big(\mathbb{E}|F|^p + \sum_{j=1}^k \mathbb{E}\|\boldsymbol{D}^{(j)}F\|_H^p\big)^{\frac{1}{p}}.$$

The Malliavin derivative is also well defined for the general E-valued random variables, where E is some separable Hilbert space, and we write $\mathbb{D}^{1,p}(E)$ as the closure of $\mathcal{S}$ under some appropriate metric with respect to E. We will use the notation $\mathbb{D}^{1,\infty}=\cap_{p\geq 1}\mathbb{D}^{1,p}$. The adjoint operator of $\boldsymbol{D}$ is introduced as follows.

Definition 3.1 (Definition 1.3.1, [29])**.** *We denote by δ the adjoint of the operator $\boldsymbol{D}$. That is, δ is an unbounded operator on $L^2(\Omega;H)$ with values in $L^2(\Omega)$ such that*

1. *The domain of δ, denoted by $\mathrm{Dom}\,\delta$, is the set of H-valued square integrable random variables $u\in L^2(\Omega;H)$ such that*

$$\big|\mathbb{E}[\langle \boldsymbol{D}F,u\rangle_H]\big|\leq c\|F\|_2$$

   for all $F\in\mathbb{D}^{1,2}$, where c is a constant depending on u.
2. *If u belongs to $\mathrm{Dom}\,\delta$, then $\delta(u)$ is the element of $L^2(\Omega)$ characterized by*

$$\mathbb{E}[F\delta(u)] = \mathbb{E}[\langle \boldsymbol{D}F,u\rangle_H]$$

   for any $F\in\mathbb{D}^{1,2}$.

3.1 Partial Malliavin calculus

The following remark motivates us to use the partial Malliavin calculus.

Remark 3.2. Because of the appearance of a neighborhood process $\widetilde{X}_\cdot$, we propose the following problem. We note that almost everything satisfied by the McKean-Vlasov SDE in [10] is also satisfied by our directed chain SDE. However, we cannot directly apply their approach to argue the existence, continuity and differentiability of the density function of $X_t^{x,[\theta]}$. The reason is that a key step connecting the Malliavin derivative and $\partial_x X_t^{x,[\theta]}$, which is defined in (2.10), may not hold in our case, that is, in general, the identity:

$$\partial_x X_t^{x,[\theta]} = \boldsymbol{D}_r X_t^{x,[\theta]} \sigma^\top \left(\sigma\sigma^\top\right)^{-1}(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r)\partial_x X_r^{x,[\theta]} \tag{3.2}$$

does not hold for any $r \leq t$. Thus, we cannot directly make use of the integration by parts formulae in [10], and hence, we cannot argue the smoothness of $X_t^{x,[\theta]}$.

Question 3.3. How can we make connections between the first order derivative $\partial_x X_t^{x,[\theta]}$ and the Malliavin derivatives similar to (3.2), which would render us to apply integration by parts formula?

To address Question 3.3, we consider the partial Malliavin derivative in [30]. Let $\mathcal{G} := \sigma(\{\widetilde{X}_{t_i}, \forall t_i \in \mathbb{Q}_T\})$ be the sigma algebra generated by the neighborhood process at all rational times, where $\mathbb{Q}_T = \mathbb{Q} \cap [0,T]$ denotes the collection of all rational numbers in $[0,T]$. Due to the continuity of $\widetilde{X}$, considering all rational time stamps is equivalent to considering the whole time interval $[0,T]$, that is, $\mathcal{G} = \sigma(\widetilde{X}_s, 0 \leq s \leq T)$. We associate with $\mathcal{G}$ the family of subspaces defined by the orthogonal complement to the subspace generated by $\{\boldsymbol{D}\widetilde{X}_{t_i}(\omega), t_i \in \mathbb{Q}_T\}$, i.e.,

$$K(\omega) = \langle \boldsymbol{D}\widetilde{X}_{t_i}(\omega), t_i \in \mathbb{Q}_T \rangle^\perp.$$

Since $\mathcal{G}$ is generated by countably many random variables, we say it is countably smoothly generated. Then the family $\mathcal{H} := \{K(\omega), \omega \in \Omega\}$ has a measurable projection by this countably smoothness of $\mathcal{G}$. We define the partial Malliavin derivative operator as $\boldsymbol{D}^{\mathcal{H}}$.

Definition 3.4 (Definition 2.1, [30])**.** *We define the partial derivative operator $\boldsymbol{D}^{\mathcal{H}} : \mathbb{D}^{1,2} \to L^2(\Omega, \mathcal{H})$ as the projection of $\boldsymbol{D}$ on $\mathcal{H}$, that is, for any $F \in \mathbb{D}^{1,2}$,*

$$\boldsymbol{D}^{\mathcal{H}} F = \mathrm{Proj}_{\mathcal{H}}(\boldsymbol{D}F) = \mathrm{Proj}_{K(\omega)}(\boldsymbol{D}F)(\omega).$$

This operator, similar to $\boldsymbol{D}$, admits an identification with a process $(\boldsymbol{D}_r^{\mathcal{H}})_{r\in[0,T]}$. Moreover, we define the norm associated with $\boldsymbol{D}^{\mathcal{H}}$ by

$$\|F\|_{\mathbb{D}^{k,p}_{\mathcal{H}}} = \left(\mathbb{E}|F|^p + \sum_{j=1}^{k} \mathbb{E}\|\boldsymbol{D}^{\mathcal{H},(j)} F\|_H^p\right)^{\frac{1}{p}},$$

where $\boldsymbol{D}^{\mathcal{H},(j)}$ is defined as

$$\boldsymbol{D}^{\mathcal{H},(j)} F = \mathrm{Proj}_{\mathcal{H}}(\boldsymbol{D}^{(j)}F) = \mathrm{Proj}_{K(\omega)}(\boldsymbol{D}^{(j)}F)(\omega).$$

Now we have the important fact that $\boldsymbol{D}^{\mathcal{H}}\widetilde{X}_t = 0$. This is because $\widetilde{X}_t$ is $\mathcal{G}$ measurable and hence equivalently

$$\boldsymbol{D}\widetilde{X}_t \in \langle \boldsymbol{D}\widetilde{X}_{t_i}, t_i \in \mathbb{Q}_T \cup \{t\} \rangle; \quad t \in [0,T]. \tag{3.3}$$

Then the projection of $\boldsymbol{D}\widetilde{X}_t$ onto the orthogonal of $\langle \boldsymbol{D}\widetilde{X}_{t_i}, t_i \in \mathbb{Q}_T \cup \{t\} \rangle$ must be zero. Similar to the common Malliavin calculus, we have an adjoint operator of $\boldsymbol{D}^{\mathcal{H}}$, which is denoted by $\delta_{\mathcal{H}}$, as well as the integration by parts formula for the partial Malliavin calculus.

Definition 3.5 (Definition 2.3, [30])**.** *Set* $\mathrm{Dom}\,\delta_{\mathcal{H}} = \{u \in L^2(\Omega; H) : \mathrm{Proj}_{\mathcal{H}} u \in \mathrm{Dom}\,\delta\}$. *For any* $u \in \mathrm{Dom}\,\delta_{\mathcal{H}}$*, set* $\delta_{\mathcal{H}}(u) = \delta(\mathrm{Proj}_{\mathcal{H}} u)$.

Following Definition 3.4 and 3.5, we have the integration by parts formula for $\boldsymbol{D}^{\mathcal{H}}$ and $\delta_{\mathcal{H}}$

$$\mathbb{E}[\langle h, \boldsymbol{D}^{\mathcal{H}} F\rangle] = \mathbb{E}[\langle \mathrm{Proj}_{\mathcal{H}} h, \boldsymbol{D}F\rangle] = \mathbb{E}[F\delta_{\mathcal{H}}(h)]. \tag{3.4}$$

Kusuoka-Stroock Processes In order to derive the differentiability of the density function, we mimic the procedure in [10] and need to develop the integration-by-parts formulae introduced in the works of [21] and [19].

Definition 3.6 (Definition 2.8 in [10])**.** *Let* E *be a separable Hilbert space and let* $r \in \mathbb{R}$, $q, M \in \mathbb{N}$. *We denote by* $\mathbb{K}_r^q(E, M)$ *the set of processes* $\Psi : [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{D}_{\mathcal{H}}^{M,\infty}(E)$ *satisfying the following:*

1. *For any multi-indices* $\alpha, \boldsymbol{\beta}, \gamma$ *satisfying* $|\alpha| + |\boldsymbol{\beta}| + |\gamma| \le M$*, the function*

$$[0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \ni (t, x, [\theta]) \mapsto \partial_x^\gamma \partial_{\boldsymbol{v}}^{\boldsymbol{\beta}} \partial_\mu^\alpha \Psi(t, x, [\theta], v) \in L^p(\Omega)$$

   exists and is continuous for all $p \ge 1$.

2. *For any* $p \ge 1$ *and* $m \in \mathbb{N}$ *with* $|\alpha| + |\boldsymbol{\beta}| + |\gamma| + m \le M$*, we have*

$$\sup_{v \in (\mathbb{R}^N)^{\#\boldsymbol{\beta}}} \sup_{t \in (0,T]} t^{-r/2} \left\| \partial_x^\gamma \partial_{\boldsymbol{v}}^{\boldsymbol{\beta}} \partial_\mu^\alpha \Psi(t, x, [\theta], v) \right\|_{\mathbb{D}_{\mathcal{H}}^{m,p}(E)} \le C\,(1 + |x| + \|\theta\|_2)^q.$$

In our discussion, we do not consider the differentiability of the process X with respect to the initial state of its neighborhood $\widetilde{X}$. This above definition of $\mathbb{K}_r^q(E)$ is almost the same as the definition in [10, Definition 2.8], except for the norm. The reason is that we only care about the existence and smoothing properties of the density function of $X^{x,[\theta]}$ and have to use the partial Malliavin calculus. We note that although the norms are different, all the regularity results under the norm $\|\cdot\|_{\mathbb{D}^{k,p}}$ also hold under our norm $\|\cdot\|_{\mathbb{D}_{\mathcal{H}}^{k,p}}$ because of the Hölder's inequality. We also define $\mathbb{D}_{\mathcal{H}}^{k,\infty} := \cap_{p \ge 1} \mathbb{D}_{\mathcal{H}}^{k,p}$. To obtain the smoothness of density functions of a process starting from a fixed initial point, we use $\mathcal{K}_r^q(\mathbb{R}, M)$ as the class of Kusuoka-Stroock processes which do not depend on a measure term. By [10, Lemma 2.11], if $\Psi \in \mathbb{K}_r^q(E, M)$, then $\Phi(t, x, y) := \Psi(t, x, \delta_x, y) \in \mathcal{K}_r^q(E, M)$.

4 Smoothness of densities

4.1 Regularities of solutions of directed chain SDEs

For the purpose of establishing the integration by parts formulae for the directed chain SDEs and applying the results in [10, Theorem 6.1], we only need to check all the regularities conditions with respect to the parameters (θ, x) contained in [10, Section 3].

Proposition 4.1 (First-order derivatives)**.** *Suppose that* $V_0, \dots, V_d \in \mathcal{C}_{b,Lip}^{1,1,1}(\mathbb{R}^+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$. *Then the following statements hold:*

1. *There exists a modification of* $X^{x,[\theta]}$ *such that for all* $t \in [0,T]$*, the map* $x \mapsto X_t^{x,[\theta]}$ *is* $\mathbb{P}$*-a.s. differentiable. We denote the derivative by* $\partial_x X^{x,[\theta]}$ *and note that it solves the following SDE*

$$\partial_x X_t^{x,[\theta]} = Id_N + \sum_{i=0}^{d} \int_0^t \left\{ \partial V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s) \partial_x X_s^{x,[\theta]} \right\} \mathrm{d}B_s^i \tag{4.1}$$

   for every $t \in [0,T]$.

2. *For all $t \in [0,T]$, the maps $\theta \mapsto X_t^\theta$ and $\theta \mapsto X_t^{x,[\theta]}$ are Fréchet differentiable in $L^2(\Omega)$, i.e. there exists a linear continuous map $\mathcal{D}X_t^\theta : L^2(\Omega) \to L^2(\Omega)$ such that for all $\gamma \in L^2(\Omega)$,*

$$\|X_t^{\theta+\gamma} - X_t^\theta - \mathcal{D}X_t^\theta(\gamma)\|_2 = o(\|\gamma\|_2) \quad \text{as} \quad \|\gamma\|_2 \to 0,$$

and similarly for $X_t^{x,[\theta]}$. These processes satisfy the following stochastic differential equations

$$\begin{aligned}\mathcal{D}X_t^{x,[\theta]}(\gamma) = \sum_{i=0}^d \int_0^t \Big[&\partial V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s)\mathcal{D}X_s^{x,[\theta]}(\gamma) \\ &+ \tilde{\partial} V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s)\mathcal{D}\widetilde{X}_s(\gamma) + \mathcal{D}V_i'(s, X_s^{x,[\theta]}, X_s^\theta, \widetilde{X}_s)(\mathcal{D}X_s^\theta(\gamma))\Big]\,\mathrm{d}B_s^i,\end{aligned} \tag{4.2}$$

$$\begin{aligned}\mathcal{D}X_t^\theta(\gamma) = \gamma + \sum_{i=0}^d \int_0^t \Big[&\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}X_s^\theta(\gamma) + \tilde{\partial} V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}\widetilde{X}_s(\gamma) \\ &+ \mathcal{D}V_i'(s, X_s^\theta, X_s^\theta, \widetilde{X}_s)(\mathcal{D}X_s^\theta(\gamma))\Big]\,\mathrm{d}B_s^i\end{aligned} \tag{4.3}$$

where V_i' is the lifting of V_i. Moreover, for each $x \in \mathbb{R}^N$, $t \in [0,T]$, the map $\mathcal{P}_2 \ni [\theta] \mapsto X_t^{x,[\theta]} \in L^2(\Omega)$ is differentiable. So, $\partial_\mu X_t^{x,[\theta]}(v)$ exists and it satisfies the following equation

$$\begin{aligned}\partial_\mu X_t^{x,[\theta]}(v) = \sum_{i=0}^d \int_0^t \Big\{&\partial V_i\big(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s\big)\partial_\mu X_s^{x,[\theta]}(v) \\ &+ \tilde{\partial} V_i\big(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s\big)\partial_\mu \widetilde{X}_s(v) \\ &+ \mathbb{E}'\Big[\partial_\mu V_i\big(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s, (X_s^{v,[\theta]})'\big)\partial_x (X_s^{v,[\theta]})'\Big] \\ &+ \mathbb{E}'\Big[\partial_\mu V_i\big(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})'\big)\partial_\mu (X_s^{\theta',[\theta]})'(v)\Big]\Big\}\,\mathrm{d}B_s^i,\end{aligned} \tag{4.4}$$

where $(X_s^{\theta'})'$ is a copy of X_s^θ on the probability space $(\Omega', \mathcal{F}', \mathbb{P}')$. Similarly, $\partial_x (X_s^{v,[\theta]})'$ is a copy of $\partial_x X_s^{v,[\theta]}$ and $\partial_\mu (X_s^{\theta',[\theta]})' = \partial_\mu (X_s^{x,[\theta]})'\big|_{x=\theta'}$. Finally, the following representation holds for all $\gamma \in L^2(\Omega)$:

$$\mathcal{D}X_t^{x,[\theta]}(\gamma) = \mathbb{E}'[\partial_\mu X_t^{x,[\theta]}(\theta')\gamma']. \tag{4.5}$$

3. *For all $t \in [0,T]$, $X_t^{x,[\theta]}, X_t^\theta \in \mathbb{D}^{1,\infty}$. Moreover, $\boldsymbol{D}_r^{\mathcal{H}} X^{x,[\theta]} = \Big(\boldsymbol{D}_r^{\mathcal{H},j}(X^{x,[\theta]})^i\Big)_{\substack{1\le j\le N\\ 1\le i\le d}}$ satisfies, for $0 \le r \le t$*

$$\boldsymbol{D}_r^{\mathcal{H}} X_t^{x,[\theta]} = \sigma\big(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r\big) + \sum_{i=0}^d \int_r^t \Big(\partial V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s)\boldsymbol{D}_r^{\mathcal{H}} X_s^{x,[\theta]}\Big)\,\mathrm{d}B_s^i, \tag{4.6}$$

where $\sigma\big(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r\big)$ is the $N \times d$ matrix with columns $V_1, \dots, V_d$.

Proof. 1. The SDE of $X^{x,[\theta]}$ satisfies a classical SDE with adapted coefficients, by [17, Theorem 7.6.5] there exists a modification of $X_t^{x,[\theta]}$ which is continuously differentiable in x, and the first derivative satisfies (4.1).

2. The maps $\theta \mapsto X_t^\theta$ and $\theta \mapsto X_t^{x,[\theta]}$ are Fréchet differentiable in a similar fashion of [10, Proposition 3.1], where we can construct the proof through an iterative scheme as in [9, Lemma 4.17]. Then (4.2) and (4.3) follow from direct computation.

Let us first rewrite the equation for $\mathcal{D}X_t^\theta(\gamma)$ in terms of the lifting V',

$$\mathcal{D}X_t^\theta(\gamma) = \gamma + \sum_{i=0}^{d} \int_0^t \Big[\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}X_s^\theta(\gamma) + \tilde{\partial} V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}\widetilde{X}_s(\gamma) + \mathbb{E}'\big[\partial_\mu V_i'(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})')(\mathcal{D}(X_s^{\theta'})'(\gamma'))\big]\Big] \mathrm{d}B_s^i. \tag{4.7}$$

We then consider the equation that we are going to prove for $\partial_\mu X_s^{\theta',[\theta]}(v)$, evaluated at $v = \theta''$ and multiplied by γ'' with both random variables defined on a probability space $(\Omega'', \mathcal{F}'', \mathbb{P}'')$. Then taking expectation with respect to $\mathbb{P}''$, we get

$$\begin{aligned}\mathbb{E}''\big[\partial_\mu X_t^{\theta',[\theta]}(\theta'')\gamma''\big] = \sum_{i=0}^{d} \int_0^t \Big\{ &\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathbb{E}''[\partial_\mu X_s^{\theta',[\theta]}(\theta'')\gamma''] \\ &+ \tilde{\partial} V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathbb{E}''[\partial_\mu \widetilde{X}_s\gamma''] \\ &+ \mathbb{E}''\mathbb{E}'\Big[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'',[\theta]})')\partial_x(X_s^{\theta'',[\theta]})'\gamma''\Big] \\ &+ \mathbb{E}'\big[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})')\mathbb{E}''[\partial_x(X_s^{\theta',[\theta]})'(\theta'')\gamma'']\big]\Big\} \mathrm{d}B_s^i.\end{aligned} \tag{4.8}$$

Note that since (γ'', θ'') are defined on a separate probability space, we have $\mathbb{E}''[\partial_\mu \widetilde{X}_s\gamma''] = \mathcal{D}\widetilde{X}_s(\gamma)$ and

$$\mathbb{E}''\mathbb{E}'\big[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'',[\theta]})')\partial_x(X_s^{\theta'',[\theta]})'\gamma''\big] = \mathbb{E}'[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})')\partial_x(X_s^{\theta',[\theta]})'\gamma'].$$

Then the dynamic of $\mathbb{E}''[\partial_\mu X_t^{\theta',[\theta]}(\theta'')\gamma'']$ reduces to

$$\begin{aligned}\mathbb{E}''\big[\partial_\mu X_t^{\theta',[\theta]}(\theta'')\gamma''\big] = \sum_{i=0}^{d} \int_0^t \Big\{ &\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathbb{E}''[\partial_\mu X_s^{\theta',[\theta]}(\theta'')\gamma''] \\ &+ \tilde{\partial} V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}\widetilde{X}_s(\gamma) \\ &+ \mathbb{E}'\big[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})')\big[\partial_x(X_s^{\theta',[\theta]})'\gamma' + \mathbb{E}''[\partial_x(X_s^{\theta'',[\theta]})'(\theta'')\gamma'']\big]\Big\} \mathrm{d}B_s^i.\end{aligned} \tag{4.9}$$

By (4.1), we can evaluate the equation at $x = \theta$, multiply by x, and derive a dynamic of $\partial_x X_t^{\theta,[\theta]}\gamma$. It can be seen that $\partial_x X_t^{\theta,[\theta]}\gamma + \mathbb{E}''[\partial_\mu X_t^{\theta',[\theta]}(\theta'')\gamma'']$ is equal to

$$\begin{aligned}\gamma + \sum_{i=0}^{d} \int_0^t \Big\{ &\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathbb{E}''[\partial_\mu X_s^{\theta',[\theta]}(\theta'')\gamma''] + \tilde{\partial} V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s)\mathcal{D}\widetilde{X}_s(\gamma) \\ &+ \mathbb{E}'\big[\partial_\mu V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s, (X_s^{\theta'})')\big[\partial_x(X_s^{\theta',[\theta]})'\gamma' + \mathbb{E}''[\partial_x(X_s^{\theta'',[\theta]})'(\theta'')\gamma'']\big]\Big\} \mathrm{d}B_s^i.\end{aligned} \tag{4.10}$$

We observe that this dynamic is identical to the dynamic for $\mathcal{D}X_t^\theta(\gamma)$ in (4.7) and hence they are identical by uniqueness. Similarly, by using this result for $\mathcal{D}X_t^\theta(\gamma)$ and the same procedures, we are able to derive that $\mathbb{E}''[\partial_\mu X_t^{x,[\theta]}(\theta'')\gamma'']$ is equal to $\mathcal{D}X_t^{x,[\theta]}(\gamma)$. So (4.5) is proved. Moreover, $\partial_\mu X_t^{x,[\theta]}(v)$ exists and satisfies equation (4.4) by its definition.

3. We first deduce the Malliavin derivative for X^θ. Consider the Picard iteration given by

$$\begin{aligned} X_t^{\theta,0} &= \theta, \\ X_t^{\theta,k+1} &= \theta + \sum_{i=0}^{d} \int_0^t V_i(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)\,\mathrm{d}B_s^i, \end{aligned}$$

where $\widetilde{X}^k$ is a copy of $X^{\theta,k}$ independent of the Brownian motion and θ. We have shown that such an iteration induces a Cauchy sequence $\{\Phi^{(k)}(\mathrm{Law}(X^{\theta,0})), k \in \mathbb{N}\}$ in Proposition 2.2 and a weak solution of the directed chain SDE. Since V_0, V_i are bounded and continuously differentiable, we have

$$\boldsymbol{D}_r^{\mathcal{H},l}[V_i^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)] = \partial V_i^j \boldsymbol{D}_r^{\mathcal{H},l} X_s^{\theta,k},$$

where we omit the arguments in V_i's for notation simplicity. Note that $|\partial V_i^j| \le K$ for some constant $K > 0$. We can then deduce $V_i^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k) \in \mathbb{D}^{1,\infty}$ by [29, Proposition 1.5.5]. Moreover, the Ito integral

$$\int_0^t V_i^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)\,\mathrm{d}B_s^i, \quad i = 1, \dots, d$$

belongs to $\mathbb{D}^{1,2}$ and for $r \le t$, we have

$$\begin{aligned} \boldsymbol{D}_r^{\mathcal{H},l}\Big[\int_0^t V_i^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)\,\mathrm{d}B_s^i\Big] &= V_l^j(r, X_r^{\theta,k}, [\widetilde{X}_r^k], \widetilde{X}_r^k) \\ &\quad + \int_r^t \boldsymbol{D}_r^{\mathcal{H},l}[V_i^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)]\,\mathrm{d}B_s^i. \end{aligned}$$

On the other hand, the Lebesgue integral $\int_0^t V_0^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)\,\mathrm{d}s$ is also in the space $\mathbb{D}^{1,2}$ and have the dynamics

$$\boldsymbol{D}_r^{\mathcal{H},l}\Big[\int_0^t V_0^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)\,\mathrm{d}s\Big] = \int_0^t \boldsymbol{D}_r^{\mathcal{H},l}[V_0^j(s, X_s^{\theta,k}, [\widetilde{X}_s^k], \widetilde{X}_s^k)]\,\mathrm{d}s.$$

Therefore, the dynamic of $\boldsymbol{D}_r^{\mathcal{H},l}[X_t^{\theta,k+1}]$ has exactly the form of (4.6) by the chain rule of the Malliavin derivative. Due to the reason that $\widetilde{X}^k$ and $X^{\theta,k}$ has the same distribution, by Doob's maximal inequality and Burkholder's inequality,

$$\mathbb{E}[\sup_{0\le s\le t} |\boldsymbol{D}_r^{\mathcal{H},l} X_s^{\theta,k}|^p] \le c_1,$$

where c_1 is a constant that depends only on K, d, p for $p \ge 2$. Moreover, we define a metric similar to (2.5) but raise the power to general $p \ge 1$,

$$D_{t,p}(\mu_1, \mu_2) := \inf\left\{\int (\sup_{0\le s\le t} |X_s(\omega_1) - X_s(\omega_2)|^p \wedge 1)\,\mathrm{d}\mu(\omega_1, \omega_2)\right\}^{1/p}.$$

We then have the following,

$$D_{t,p}(m^{k+1}, m^k)^2 \le c_1 \int_0^t D_{s,p}(\mathrm{Law}(X^{\theta,k}), \mathrm{Law}(X^{\theta,k-1}))\,\mathrm{d}s + c_2 \int_0^t D_s(m^k, m^{k-1})^2\,\mathrm{d}s,$$

by a similar approach as in the proof of Proposition 2.2, where c_1, c_2 are positive constants depending on K, d, p and $m^k = \mathrm{Law}(\boldsymbol{D}_r^{\mathcal{H},l} X^{\theta,k})$. By iteration, we find that $\{m^k, k \in \mathbb{N}\}$ forms a Cauchy sequence in space $\mathcal{M}(C([0,T], \mathbb{R}^{N\times d}))$ and has a limit. We have now proved that

$$\boldsymbol{D}_r^{\mathcal{H}} X_t^\theta = \sigma\big(r, X_r^\theta, [X_r^\theta], \widetilde{X}_r\big) + \sum_{i=0}^{d} \int_r^t \Big(\partial V_i(s, X_s^\theta, [X_s^\theta], \widetilde{X}_s) \boldsymbol{D}_r^{\mathcal{H}} X_s^\theta\Big)\, \mathrm{d}B_s^i, \tag{4.11}$$

and the solution of $\boldsymbol{D}_r^{\mathcal{H}} X_t^\theta$ exists uniquely in the weak sense. In the iteration, it can easily be proved by induction that $X^{\theta,k} \in \mathbb{D}^{1,\infty}$ and the sequence $\boldsymbol{D}_r^{\mathcal{H}} X_t^{\theta,k}$ is uniformly bounded in $L^p(\Omega; H)$ for $p \geq 2$. Therefore, we have $X_t^\theta \in \mathbb{D}_{\mathcal{H}}^{1,\infty}$. The proof for $X_t^{x,[\theta]}$ is similar, we can set $X_t^{x,[\theta],0} = \theta$ add another equation for $X_t^{x,[\theta],k}$ into the above Picard iteration

$$X_t^{x,[\theta],k+1} = x + \sum_{i=0}^{d} \int_0^t V_i(s, X_s^{x,[\theta],k}, [\widetilde{X}_s^k], \widetilde{X}_s^{x,k})\, \mathrm{d}B_s^i.$$

Then the procedures are the same as the deduction for $\boldsymbol{D}_r^{\mathcal{H}} X^\theta$.

□

For the purpose of more general applications, we want to make sure that the density for directed chain SDE is at least second order differentiable in $(x, [\theta])$, hence we need to extend the above first order regularities to higher orders. Following [10], we provide a result for the general case, which characterizes $X_t^{x,[\theta]}$ as a Kusuoka-Stroock process.

Theorem 4.2. *Suppose $V_0, \ldots, V_d \in \mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}([0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N)\times\mathbb{R}^N; \mathbb{R}^N)$, then it follows $(t,x,[\theta]) \mapsto X_t^{x,[\theta]} \in \mathbb{K}_0^1(\mathbb{R}^N, k)$. If, in addition, $V_0, \ldots, V_d$ are uniformly bounded, then $(t,x,[\theta]) \mapsto X_t^{x,[\theta]} \in \mathbb{K}_0^0(\mathbb{R}^N, k)$.*

Note that [10, Proposition 6.7 and 6.8] can be extended to our directed chain case, since the coefficients $V_i : [0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N)\times\mathbb{R}^N \to \mathbb{R}^N$ in the directed chain SDEs can be written as a map of the form $\Omega\times[0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N) \ni (\omega,t,x,\mu) \mapsto a(\omega,t,x,\mu) \in \mathbb{R}^N$. This is because the auxiliary dependence on the neighborhood in the coefficients can be thought of as the dependence on an initial state x, initial distribution μ and independent Brownian motions, which are implied in the term ω. Moreover, we are able to take care of the extra term with $\mathcal{D}\widetilde{X}_s$ due to the differentiability and regularity of V_i.

Similarly to Proposition 4.1, each type of derivative (w.r.t. x, μ or v) of $X_t^{x,[\theta]}$ satisfies a linear equation. We will introduce a general linear equation, derive some a priori L^p estimates on the solution, and then show that this linear equation is again differentiable under some conditions in the next lemma. Whenever we say a_k, $k = 1,2,3$, we also mean $\tilde{a}_1$.

Lemma 4.3. *Let v_r be one element of the tuple $\boldsymbol{v} = (v_1, \ldots, v_{\#\boldsymbol{v}})$ and $Y^{x,[\theta]}(\boldsymbol{v})$ solve the following SDE*

$$\begin{aligned} Y_t^{x,[\theta]}(\boldsymbol{v}) = a_0 + \sum_{i=0}^{d} \int_0^t \Big\{ & a_1^i(s,x,[\theta]) Y_s^{x,[\theta]}(\boldsymbol{v}) + \tilde{a}_1^i(s,x,[\theta]) \widetilde{Y}_s(\boldsymbol{v}) + a_2^i(s,x,[\theta],\boldsymbol{v}) \\ & + \mathbb{E}'\big[a_3^i(s,x,[\theta],\theta')(Y_s^{\theta',[\theta]})'(\boldsymbol{v}) + \sum_{r=1}^{\#\boldsymbol{v}} a_3^i(s,x,[\theta],\theta')(Y_s^{v_r,[\theta]})'(\boldsymbol{v})\big]\Big\}\, \mathrm{d}B_s^i, \end{aligned} \tag{4.12}$$

where, for all $i = 1, \dots, d$, the coefficients $(t, x, [\theta], \boldsymbol{v}) \mapsto a_k(t, x, [\theta], \boldsymbol{v})$ are continuously in $L^p(\Omega)$ $\forall p \geq 1$, $k = 1, 2, 3$ and

$$\begin{aligned}
&a_0 \in \mathbb{R}^N, \\
&a_1, \tilde{a}_1 : \Omega \times [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}^{N \times N} \\
&a_2 : \Omega \times [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^{\#\boldsymbol{v}} \to \mathbb{R}^N \\
&a_3^i : \Omega' \times \Omega \times [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^{N \times N}.
\end{aligned}$$

In (4.12), $(Y^{\theta', [\theta]})'$ is a copy of Y^θ on the probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ where the initial state is θ'. Similarly, $(Y^{v_r, [\theta]})'$ is a copy of $Y^{x, [\theta]}$ on the probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ where the initial state is v_r. $\widetilde{Y}$ is the neighborhood process, which has the same law as Y^θ and is independent of Brownian motion B and θ. We make the following boundedness assumptions

1. $\sup_{x \in \mathbb{R}^N, [\theta] \in \mathcal{P}_2(\mathbb{R}^N), \boldsymbol{v} \in (\mathbb{R}^N)^{\#\boldsymbol{v}}} \|a_2(\cdot, x, [\theta], \boldsymbol{v})\|_{\mathcal{S}_T^p} < \infty$,
2. $a_1, \tilde{a}_1$ *and* a_3 *are uniformly bounded,*
3. $\sup_{x \in \mathbb{R}^N, [\theta] \in \mathcal{P}_2(\mathbb{R}^N), \boldsymbol{v} \in (\mathbb{R}^N)^{\#\boldsymbol{v}}} \|a_2(\cdot, x, [\theta], \boldsymbol{v})\|_{\mathcal{S}_T^2} < \infty$

and then we have the following estimate for $C = C(p, T, a_1, a_3)$

$$\|Y^{x, [\theta]}(\boldsymbol{v})\|_{\mathcal{S}_T^p} \leq C(|a_0| + \|a_2(\cdot, x, [\theta], \boldsymbol{v})\|_{\mathcal{S}_T^p} + \|a_2(\cdot, x, [\theta], \boldsymbol{v})\|_{\mathcal{S}_T^2}).$$

Moreover, we also get that the mapping

$$[0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^{\#\boldsymbol{v}} \ni (t, x, [\theta], \boldsymbol{v}) \mapsto Y_t^{x, [\theta]}(\boldsymbol{v}) \in L^p(\Omega)$$

is continuous.

Proof. Note that $\|\widetilde{Y}(\boldsymbol{v})\|_{\mathcal{S}_T^p} = \|(Y_s^{\theta', [\theta]})'(\boldsymbol{v})\|_{\mathcal{S}_T^p}$ since they have the same distribution. The rest proof is identical to [10, Lemma 6.7] by using Gronwall's lemma and the Burkholder-Davis-Gundy inequality a couple times. □

We now consider the differentiability of the generic process satisfying the linear equation in Lemma 4.3. To ease the burden on notation, we omit the $(t, x, [\theta])$ in a_k, and write $a_k\big|_{v=\theta'}$ to denote $a_k(s, x, [\theta], \theta')$, for example.

Proposition 4.4. *Suppose that the process $Y^{x, [\theta]}(\boldsymbol{v})$ is as in Lemma 4.3. In addition to the assumptions of Lemma 4.3, we introduce the following differentiability assumptions:*

(a) For $k = 1, 2, 3$, all $(s, [\theta], \boldsymbol{v}) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^{\#\boldsymbol{v}}$ and each $p \geq 1$, $\mathbb{R}^N \ni x \mapsto a_k(s, x, [\theta], \boldsymbol{v}) \in L^p(\Omega)$ is differentiable.

(b) For $k = 1, 2, 3$, all $(s, [\theta], x) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N$ and each $p \geq 1$, $(\mathbb{R}^N)^{\#\boldsymbol{v}} \ni \boldsymbol{v} \mapsto a_k(s, x, [\theta], \boldsymbol{v}) \in L^p(\Omega)$ are differentiable.

(c) For all $(s, x, \boldsymbol{v}) \in [0,T] \times \mathbb{R}^N \times (\mathbb{R}^N)^{\#\boldsymbol{v}}$ the mapping $L^2(\Omega) \ni \theta \mapsto a_2(s, x, [\theta], \boldsymbol{v}) \in L^2(\Omega)$ is Fréchet differentiable.

(d) $a_k(s, x, [\theta], \boldsymbol{v}) \in \mathbb{D}_{\mathcal{H}}^{1,\infty}$ for $k = 1, 2, 3$ and all $(s, x, [\theta], \boldsymbol{v}) \in [0,T] \times \mathcal{P}_2(\mathbb{R}^N) \times (\mathbb{R}^N)^{\#\boldsymbol{v}}$. Moreover, we assume the following estimates on the Malliavin derivatives hold.

$$\sup_{r \in [0,T]} \mathbb{E}\left[\sup_{s \in [0,T]} |\boldsymbol{D}_r^{\mathcal{H}} a_k(s, x, [\theta], \boldsymbol{v})|^p \right] < \infty, \quad k = 0, 1, 2, 3.$$

Then, for all $t \in [0,T]$ the following hold:

1. *Under assumption (a), $x \mapsto Y_t^{x,[\theta]}(\boldsymbol{v})$ is differentiable in $L^p(\Omega)$ for all $p \geq 1$ and*

$$\partial_x Y_t^{x,[\theta]}(\boldsymbol{v}) :\overset{L^p}{=} \lim_{h\to 0} \frac{1}{|h|}\bigg(Y_t^{x+h,[\theta]}(\boldsymbol{v}) - Y_t^{x,[\theta]}(\boldsymbol{v})\bigg),$$

where the limit is taken in L^p sense, satisfies

$$\begin{aligned}\partial_x Y_t^{x,[\theta]}(\boldsymbol{v}) = \sum_{i=0}^{d}\int_0^t \Big\{ &\partial_x a_1^i Y_s^{x,[\theta]}(\boldsymbol{v}) + a_1^i \partial_x Y_s^{x,[\theta]}(\boldsymbol{v}) + \partial_x a_2^i \\ &+ \mathbb{E}'\bigg[\partial_x a_3^i\big|_{v=\theta'}(Y_s^{\theta',[\theta]})'(\boldsymbol{v}) + \sum_{r=1}^{\#\boldsymbol{v}} \partial_x a_3^i\big|_{v=v_r}(Y_s^{\theta',[\theta]})'(\boldsymbol{v})\bigg]\Big\}\,\mathrm{d}B_s^i\end{aligned}$$

2. *Under assumption (b), $\boldsymbol{v} \mapsto Y_t^{x,[\theta]}(\boldsymbol{v})$ is differentiable in $L^p(\Omega)$ for all $p \geq 1$ and*

$$\partial_{\boldsymbol{v}} Y_t^{x,[\theta]}(\boldsymbol{v}) :\overset{L^p}{=} \lim_{h\to 0} \frac{1}{|h|}\bigg(Y_t^{x,[\theta]}(\boldsymbol{v}+h) - Y_t^{x,[\theta]}(\boldsymbol{v})\bigg)$$

satisfies

$$\begin{aligned}\partial_{v_j} Y_t^{x,[\theta]}(\boldsymbol{v}) = \sum_{i=0}^{d}\int_0^t \Big\{ &a_1^i \partial_{v_j} Y_s^{x,[\theta]}(\boldsymbol{v}) + \tilde{a}_1^i \partial_{v_j}\widetilde{Y}_s(\boldsymbol{v}) + \partial_{v_j} a_2^i \\ &+ \mathbb{E}'\bigg[\partial_v a_3^i\big|_{v=v_j}(Y_s^{v_j,[\theta]})'(\boldsymbol{v})\bigg] + \mathbb{E}'\bigg[a_3^i\big|_{v=v_j}\partial_x(Y_s^{v_j,[\theta]})'(\boldsymbol{v}) \\ &+ a_3^i\big|_{v=\theta'}\partial_{v_j}(Y_s^{\theta',[\theta]})'(\boldsymbol{v}) + \sum_{r=1}^{\#\boldsymbol{v}} a_3^i\big|_{v=v_r}\partial_{v_j}(Y_s^{v_r,[\theta]})'(\boldsymbol{v})\bigg]\Big\}\,\mathrm{d}B_s^i.\end{aligned}$$

3. *Under assumption (a), (b) and (c), the maps $\theta \mapsto Y_t^{\theta,[\theta]}(\boldsymbol{v})$ and $\theta \mapsto Y_t^{x,[\theta]}(\boldsymbol{v})$ are Fréchet differentiable for all $(x, \boldsymbol{v}) \in \mathbb{R}^N \times (\mathbb{R}^N)^{\#\boldsymbol{v}}$, so $\partial_\mu Y_t^{x,[\theta]}(\boldsymbol{v})$ exists and it satisfies*

$$\begin{aligned}\partial_\mu Y_t^{x,[\theta]}(\boldsymbol{v},\hat{v}) = \sum_{i=0}^{d}\int_0^t \Big\{ &\partial_\mu a_1^i Y_s^{x,[\theta]}(\boldsymbol{v}) + a_1^i \partial_\mu Y_s^{x,[\theta]}(\boldsymbol{v},\hat{v}) + \partial_\mu \tilde{a}_1^i \widetilde{Y}_s(\boldsymbol{v}) + a_1^i \partial_\mu \widetilde{Y}_s(\boldsymbol{v},\hat{v}) \\ &+ \partial_\mu a_2^i + \mathbb{E}'\bigg[\partial_\mu a_3^i(Y_s^{\theta',[\theta]})'(\boldsymbol{v}) + \partial_v a_3^i(Y_s^{\hat{v},[\theta]})'(\boldsymbol{v}) + a_3^i\big|_{v=\theta'}\partial_\mu(Y_s^{\theta',[\theta]})'(\boldsymbol{v},\hat{v})\bigg] \\ &+ \mathbb{E}'\bigg[a_3^i\big|_{v=\hat{v}}\partial_x(Y_s^{\hat{v},[\theta]})'(\boldsymbol{v}) + \sum_{r=1}^{\#\boldsymbol{v}} a_3^i\big|_{v=v_r}\partial_\mu(Y_s^{v_r,[\theta]})'(\boldsymbol{v},\hat{v})\bigg]\Big\}\,\mathrm{d}B_s^i.\end{aligned}$$

Moreover, we have the representation, for all $\gamma \in L^2(\Omega)$,

$$\mathcal{D}\bigg(Y_t^{\theta,[\theta]}(\boldsymbol{v})\bigg)(\gamma) = \bigg(\partial_x Y_t^{x,[\theta]}(\boldsymbol{v})\gamma + \mathbb{E}''\big[\partial_\mu Y_t^{x,[\theta]}(\boldsymbol{v},\theta'')\gamma''\big]\bigg)\bigg|_{x=\theta}.$$

4. *Under assumption (d), $Y_t^{x,[\theta]} \in \mathbb{D}^{1,\infty}$ and $\boldsymbol{D}_r^{\mathcal{H}} Y_t^{x,[\theta]}$ satisfies*

$$\begin{aligned}\boldsymbol{D}_r^{\mathcal{H}} Y_t^{x,[\theta]}(\boldsymbol{v}) = &\Big(a_1^j Y_r^{x,[\theta]} + \tilde{a}_1^j \widetilde{Y}_r + a_2^j + \mathbb{E}'\big[a_3^j(Y_s^{x,[\theta]})'(\boldsymbol{v})\big]\Big)_{j=1,\dots,d} \\ &+ \sum_{i=0}^{d}\int_0^t \Big\{ \boldsymbol{D}_r^{\mathcal{H}} a_1^i Y_s^{x,[\theta]}(\boldsymbol{v}) + \boldsymbol{D}_r^{\mathcal{H}} \tilde{a}_1^i \widetilde{Y}_s + a_1^i \boldsymbol{D}_r^{\mathcal{H}} Y_s^{x,[\theta]}(\boldsymbol{v}) + \tilde{a}_1^i \boldsymbol{D}_r^{\mathcal{H}} \widetilde{Y}_s \\ &+ \boldsymbol{D}_r^{\mathcal{H}} a_2^i + \mathbb{E}'\big[\boldsymbol{D}_r^{\mathcal{H}} a_3^i\big|_{v=\theta'}(Y_s^{x,[\theta]})'(\boldsymbol{v})\big]\Big\}\,\mathrm{d}B_s^i.\end{aligned}$$

Moreover, the following bound holds:

$$\sup_{r\le t}\mathbb{E}\left[|\boldsymbol{D}_r^{\mathcal{H}}Y_t^{x,[\theta]}(\boldsymbol{v})|^p\right]\le C\sup_{r\le t}\mathbb{E}\left[\sup_{r\le t\le T}\left(|\boldsymbol{D}_r^{\mathcal{H}}a_1|^p+|\boldsymbol{D}_r^{\mathcal{H}}\tilde{a}_1|^p\right)\right] \tag{4.13}$$

The limits in the above are taken in the L^p sense. When we say $k=1,2,3$ for the assumptions, we also mean $\tilde{a}_1$.

Proof. See Propositions 4.1 and [10, Proposition 6.8] for the proof. □

We are now ready to prove Theorem 4.2.

Proof of Theorem 4.2. The proof follows identically the proof of [10, Theorem 3.2], where we apply Lemma 4.3 and Proposition 4.4. □

4.2 Integration by parts formulae

Now we introduce some operators acting on the Kusuoka-Stroock processes. These operators will be used later in the integration by parts formulae. We first make the following common assumption on uniform ellipticity.

Assumption 4.5 (Uniform Ellipticity)**.** *Let $\sigma:[0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N)\times\mathbb{R}^N\to\mathbb{R}^{N\times d}$ be given by*

$$\sigma(t,z,\mu,\tilde{z}):=[V_1(t,z,\mu,\tilde{z}),\dots,V_d(t,z,\mu,\tilde{z})].$$

We assume that there exists $\epsilon>0$ such that, for all $\xi\in\mathbb{R}^N$, $z\in\mathbb{R}^N$ and $\mu\in\mathcal{P}_2(\mathbb{R}^N)$,

$$\xi^\top\sigma(t,z,\mu,\tilde{z})\sigma(t,z,\mu,\tilde{z})^\top\xi\ge\epsilon|\xi|^2.$$

For a function $\Psi:[0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N)\to\mathbb{D}^{n,\infty}$, the following operators acting on Kusuoka-Stroock processes in $\mathbb{K}_r^q(\mathbb{R},n)$ with multi-index $\alpha=(i)$ and $(t,x,[\theta])\in[0,T]\times\mathbb{R}^N\times\mathcal{P}_2(\mathbb{R}^N)$ are given by

$$\begin{aligned}
I_{(i)}^1(\Psi)(t,x,[\theta])&:=\frac{1}{\sqrt{t}}\delta_{\mathcal{H}}\left(r\mapsto\Psi(t,x,[\theta])\left(\sigma^\top\left(\sigma\sigma^\top\right)^{-1}(r,X_r^{x,[\theta]},[X_r^\theta],\widetilde{X}_r)\partial_xX_r^{x,[\theta]}\right)_i\right)\\
I_{(i)}^2(\Psi)(t,x,[\theta])&:=\sum_{j=1}^N I_{(j)}^1\left(\left(\partial_xX_t^{x,[\theta]}\right)_{j,i}^{-1}\Psi(t,x,[\theta])\right),\\
I_{(i)}^3(\Psi)(t,x,[\theta])&:=I_{(i)}^1(\Psi)(t,x,[\theta])+\sqrt{t}\partial^i\Psi(t,x,[\theta]),\\
\mathcal{I}_{(i)}^1(\Psi)(t,x,[\theta],v_1)&:=\frac{1}{\sqrt{t}}\delta_{\mathcal{H}}\Bigg(r\mapsto\left(\sigma^\top\left(\sigma\sigma^\top\right)^{-1}(r,X_r^{x,[\theta]},[X_r^\theta],\widetilde{X}_r)\right.\\
&\qquad\left.\partial_xX_r^{x,[\theta]}(\partial_xX_t^{x,[\theta]})^{-1}\partial_\mu X_t^{x,[\theta]}(v_1)\right)_i\Psi(t,x,[\theta])\Bigg),\\
\mathcal{I}_{(i)}^3(\Psi)(t,x,[\theta],v_1)&:=\mathcal{I}_{(i)}^1(\Psi)(t,x,[\theta],v_1)+\sqrt{t}(\partial_\mu\Psi)_i(t,x,[\theta],v_1).
\end{aligned}$$

For a general multi-index $\alpha=(\alpha_1,\dots,\alpha_n)$, we inductively define

$$I_\alpha^1:=I_{\alpha_n}^1\circ I_{\alpha_{n-1}}^1\circ\cdots\circ I_{\alpha_1}^1,$$

the definition of the other operators are analogue to I_α^1. The following proposition follows directly from our previous discussion and the definition of the Kusuoka-Stroock process.

Proposition 4.6. *If $V_0, \dots, V_d \in \mathcal{C}^{k,k,k}_{b,\mathrm{Lip}}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$, Assumption 4.5 holds and $\Psi \in \mathbb{K}^q_r(\mathbb{R}, n)$, then $I^1_\alpha(\Psi)$ and $I^3_\alpha(\Psi)$, are all well-defined for $|\alpha| \le (k \wedge n)$. $I^2_\alpha(\Psi), \mathcal{I}^1_\alpha(\Psi)$ and $\mathcal{I}^3_\alpha(\Psi)$ are well defined for $|\alpha| \le n \wedge (k-2)$. Moreover,*

$$\begin{aligned} I^1_\alpha(\Psi), I^3_\alpha(\Psi) &\in \mathbb{K}^{q+2|\alpha|}_r(\mathbb{R}, (k \wedge n) - |\alpha|), \\ I^2_\alpha(\Psi) &\in \mathbb{K}^{q+3|\alpha|}_r(\mathbb{R}, [n \wedge (k-2)] - |\alpha|), \\ \mathcal{I}^1_\alpha(\Psi), \mathcal{I}^3_\alpha(\Psi) &\in \mathbb{K}^{q+4|\alpha|}_r(\mathbb{R}, [n \wedge (k-2)] - |\alpha|). \end{aligned}$$

If $\Psi \in \mathbb{K}^0_r(\mathbb{R}, n)$ and $V_0, \dots, V_d$ are uniformly bounded, then

$$\begin{aligned} I^1_\alpha(\Psi), I^3_\alpha(\Psi) &\in \mathbb{K}^0_r(\mathbb{R}, (k \wedge n) - |\alpha|), \\ I^2_\alpha(\Psi) &\in \mathbb{K}^0_r(\mathbb{R}, [n \wedge (k-2)] - |\alpha|), \\ \mathcal{I}^1_\alpha(\Psi), \mathcal{I}^3_\alpha(\Psi) &\in \mathbb{K}^0_r(\mathbb{R}, [n \wedge (k-2)] - |\alpha|). \end{aligned}$$

From now on, the Integration by Parts Formulae (IBPF) follow in the same way as [10, Sec 4.] by replacing $\boldsymbol{D}, \delta$ by $\boldsymbol{D}^{\mathcal{H}}, \delta_{\mathcal{H}}$ and using integral by parts for this partial Malliavin derivative.

Integration by parts formulae in the space variable are established in the following Proposition.

Proposition 4.7 (Proposition 4.1, [10])**.** *Let $f \in \mathcal{C}^\infty_b(\mathbb{R}^N, \mathbb{R})$ and $\Psi \in \mathbb{K}^q_r(\mathbb{R}, n)$, then*

1. *If $|\alpha| \le n \wedge k$, then*

$$\mathbb{E}\big[\partial^\alpha_x \big(f\big(X^{x,[\theta]}_t\big)\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X^{x,[\theta]}_t\big) I^1_\alpha(\Psi)(t,x,[\theta])\big].$$

2. *If $|\alpha| \le n \wedge (k-2)$, then*

$$\mathbb{E}\big[(\partial^\alpha f)\big(X^{x,[\theta]}_t\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X^{x,[\theta]}_t\big) I^2_\alpha(\Psi)(t,x,[\theta])\big].$$

3. *If $|\alpha| \le n \wedge k$, then*

$$\partial^\alpha_x \mathbb{E}\big[f\big(X^{x,[\theta]}_t\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X^{x,[\theta]}_t\big) I^3_\alpha(\Psi)(t,x,[\theta])\big].$$

4. *If $|\alpha| + |\beta| \le n \wedge (k-2)$, then*

$$\partial^\alpha_x \mathbb{E}\big[(\partial^\beta f)\big(X^{x,[\theta]}_t\big)\Psi(t,x,[\theta])\big] = t^{-(|\alpha|+|\beta|)/2}\mathbb{E}\big[f\big(X^{x,[\theta]}_t\big) I^3_\alpha\big(I^2_{(}\beta\Psi)\big)(t,x,[\theta])\big].$$

Proof. 1. First, we note that Equation (4.1) satisfied by $\partial_x X^{x,[\theta]}_t$ and Equation (4.6) satisfied by $\boldsymbol{D}^{\mathcal{H}}_r X^{x,[\theta]}_t$ are the same except their initial condition. It therefore follows from our discussion of partial Malliavin derivative that

$$\partial_x X^{x,[\theta]}_t = \boldsymbol{D}^{\mathcal{H}}_r X^{x,[\theta]}_t \sigma^\top \big(\sigma\sigma^\top\big)^{-1}(r, X^{x,[\theta]}_r, [X^\theta_r], \widetilde{X}_r)\partial_x X^{x,[\theta]}_r. \tag{4.14}$$

We are then allowed to compute the followings for $f \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R})$,

$$\begin{aligned}
\mathbb{E}\big[\partial_x\big(f\big(X_t^{x,[\theta]}\big)\big)\Psi(t,x,[\theta])\big] &= \mathbb{E}\big[\partial f\big(X_t^{x,[\theta]}\big)\partial_x X_t^{x,[\theta]}\Psi(t,x,[\theta])\big] \\
&= \frac{1}{t}\mathbb{E}\bigg[\int_0^t \partial f\big(X_t^{x,[\theta]}\big)\partial_x X_t^{x,[\theta]}\Psi(t,x,[\theta])\,\mathrm{d}r\bigg] \\
&= \frac{1}{t}\mathbb{E}\bigg[\int_0^t \partial f\big(X_t^{x,[\theta]}\big)\boldsymbol{D}_r^{\mathcal{H}} X_t^{x,[\theta]}\sigma^\top\big(\sigma\sigma^\top\big)^{-1}(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r) \\
&\qquad\qquad \times \partial_x X_r^{x,[\theta]}\Psi(t,x,[\theta])\,\mathrm{d}r\bigg] \\
&= \frac{1}{t}\mathbb{E}\bigg[\int_0^t \boldsymbol{D}_r^{\mathcal{H}} f\big(X_t^{x,[\theta]}\big)\sigma^\top\big(\sigma\sigma^\top\big)^{-1}(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r) \\
&\qquad\qquad \times \partial_x X_r^{x,[\theta]}\Psi(t,x,[\theta])\,\mathrm{d}r\bigg] \\
&= \frac{1}{t}\mathbb{E}\bigg[f\big(X_t^{x,[\theta]}\big)\delta_{\mathcal{H}}\bigg(r \mapsto \Psi(t,x,[\theta]) \\
&\qquad\qquad \times \big(\sigma^\top\big(\sigma\sigma^\top\big)^{-1}(r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r)\partial_x X_r^{x,[\theta]}\big)\bigg)\bigg],
\end{aligned}$$

where we have applied partial Malliavin calculus integration by parts from Equation (3.4) in the last equality. This proves the result for $|\alpha| = 1$. By Proposition 4.6, $I_\alpha^1(\Psi) \in \mathbb{K}_r^{q+2}(\mathbb{R}, (k \wedge n) - 1)$ when $|\alpha| = 1$. We can then repeat the above procedures iteratively to get to desired result.

2. By the chain rule,

$$\begin{aligned}
\mathbb{E}\big[(\partial^i f)\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big] &= \sum_{j=1}^N \mathbb{E}\bigg[\partial_{x_i}\big(f\big(X_t^{x,[\theta]}\big)\big)\bigg(\big(X_t^{x,[\theta]}\big)^{-1}\bigg)^{j,i}\Psi(t,x,[\theta])\bigg] \\
&= t^{-1/2}\sum_{j=1}^N \mathbb{E}\bigg[f\big(X_t^{x,[\theta]}\big)I_{(j)}^1\bigg(\bigg(\big(X_t^{x,[\theta]}\big)^{-1}\bigg)^{j,i}\Psi(t,x,[\theta])\bigg)\bigg] \\
&= t^{1/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)I_{(i)}^2(\Psi)(t,x,[\theta])\big],
\end{aligned}$$

where we apply the result in part 1 to the second equality. From Proposition 4.6, $I_{(i)}^2(\Psi) \in \mathbb{K}_r^{q+3}(\mathbb{R}, (n \wedge (k-2)) - 1)$, so since $|\alpha| \leq (n \wedge (k-2))$, the proof follows from applying the same arguments for another $|\alpha| - 1$ times.

3. By part 1 and direct computation,

$$\begin{aligned}
\partial_x^i\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big] &= \mathbb{E}\big[\partial_x^i f\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta]) + f\big(X_t^{x,[\theta]}\big)\partial_x^i\Psi(t,x,[\theta])\big] \\
&= t^{-1/2}\mathbb{E}\bigg[f\big(X_t^{x,[\theta]}\big)\big\{I_i^1(\Psi)(t,x,[\theta]) + \sqrt{t}\partial_x^i\Psi(t,x,[\theta])\big\}\bigg],
\end{aligned}$$

which proves the result for $|\alpha| = 1$. Again, we have $I_\alpha^3(\Psi) \in \mathbb{K}_r^{q+2}(\mathbb{R}, (k \wedge n) - 1)$ when $|\alpha| = 1$. Then the proof follows from iterative implementation of the above procedure.

4. This part follows from parts 2 and 3 directly.

□

Similarly to integration by parts in the space variable, we can also derive integration by parts in the measure variable as follows.

Proposition 4.8 (Proposition 4.2, [10])**.** *Let $f \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R})$ and $\Psi \in \mathbb{K}_r^q(\mathbb{R}, n)$, then*

1. *If* $|\beta| \le n \wedge (k-2)$, *then*

$$\mathbb{E}\big[\partial_\mu^\beta\big(f\big(X_t^{x,[\theta]}\big)\big)(\boldsymbol{v})\Psi(t,x,[\theta])\big] = t^{-|\beta|/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\mathcal{I}_\beta^1(\Psi)(t,x,[\theta],\boldsymbol{v})\big].$$

2. *If* $|\beta| \le n \wedge (k-2)$, *then*

$$\partial_\mu^\beta\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big](\boldsymbol{v}) = t^{-|\beta|/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\mathcal{I}_\beta^3(\Psi)(t,x,[\theta],\boldsymbol{v})\big].$$

3. *If* $|\alpha|+|\beta| \le n \wedge (k-2)$, *then*

$$\partial_\mu^\beta\mathbb{E}\big[(\partial^\alpha f)\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big](\boldsymbol{v}) = t^{-(|\alpha|+|\beta|)/2}\mathbb{E}\Big[f\big(X_t^{x,[\theta]}\big)\mathcal{I}_\beta^3\big(I_\alpha^2(\Psi)\big)(t,x,[\theta],\boldsymbol{v})\Big].$$

Proof. The proofs use the same idea as Proposition 4.7 and Equation (4.14). □

We now consider the integration by parts formulae for the derivatives of the mapping:

$$x \mapsto \mathbb{E}[f(X_t^{x,\delta_x})].$$

Let us introduce the following operator acting on $\mathbb{K}_r^q(\mathbb{R},M)$, the set of the Kusuoka-Stroock processes that do not depend on measure μ. For $\alpha=(i)$,

$$J_{(i)}(\Phi)(t,x) := I_{(i)}^3(\Phi)(t,x,\delta_x) + \mathcal{I}_{(i)}^3(\Phi)(t,x,\delta_x)$$

and for $\alpha=(\alpha_1,\dots,\alpha_n)$, $J_\alpha(\Phi) := J_{\alpha_n}\circ\cdots\circ J_{\alpha_1}(\Phi)$.

Theorem 4.9. *Let* $f \in \mathcal{C}_b^\infty(\mathbb{R}^N;\mathbb{R})$. *For all multi-indices* α *on* $\{1,\dots,N\}$ *with* $|\alpha| \le k-2$,

$$\partial_x^\alpha\mathbb{E}\big[f\big(X_t^{x,\delta_x}\big)\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X_t^{x,\delta_x}\big)J_\alpha(1)(t,x)\big].$$

In particular, we get the following bound,

$$\big|\partial_x^\alpha\mathbb{E}\big[f\big(X_t^{x,\delta_x}\big)\big]\big| \le C\|f\|_\infty t^{-|\alpha|/2}(1+|x|)^{4|\alpha|}$$

Proof. Since δ_x depends on x, we have

$$\partial_x^i\mathbb{E}\big[f\big(X_t^{x,\delta_x}\big)\big] = \partial_z^i\mathbb{E}\big[f\big(X_t^{z,\delta_x}\big)\big]\big|_{z=x} + \partial_\mu^i\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\big](v)\big|_{[\theta]=\delta_x,v=x},$$

then for $|\alpha|=1$ the result yields by Proposition 4.7 and 4.8. The proof is completed by repeating this procedure for another $|\alpha|-1$ times. □

The following corollary is useful for the smoothness of densities of directed-chain SDEs.

Corollary 4.10. *Let* $f \in \mathcal{C}_b^\infty(\mathbb{R}^N;\mathbb{R})$, α *and* β *be multiindices on* $\{1,\dots,N\}$ *with* $|\alpha|+|\beta| \le k-2$. *Then,*

$$\partial_x^\alpha\mathbb{E}\big[(\partial^\beta f)\big(X_t^{x,\delta_x}\big)\big] = t^{-\frac{|\alpha|+|\beta|}{2}}\mathbb{E}\big[f\big(X_t^{x,\delta_x}\big)I_\beta^2(J_\alpha(1))(t,x)\big]$$

and $I_\beta^2(J_\alpha(1)) \in \mathbb{K}_0^{4|\alpha|+3|\beta|}(\mathbb{R},k-2-|\alpha|-|\beta|)$.

Proof. The proof is derived from Theorem 4.9 and Proposition 4.7. □

4.3 Smooth densities

We are now ready to prove the main theorem of this section.

Theorem 4.11. *We assume that Assumption 4.5 holds and* $V_0, \dots, V_d \in \mathcal{C}^{k,k,k}_{b,\mathrm{Lip}}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$. *Let* α, β *be multiindices in* $\{1, \dots, N\}$ *and let* $k \geq |\alpha| + |\beta| + N + 2$.

1. *Assume also the initial state for directed chain SDE is* $\theta \equiv x$, *i.e.* $[\theta] = \delta_x$. *Then the directed chain SDE* (2.1) *coincides with the alternative SDE* (2.10). *For all* $t \in [0,T]$, X_t^{x,δ_x} *has a density* $p(t,x,\cdot)$ *such that* $(x,z) \mapsto \partial_x^\alpha \partial_z^\beta p(t,x,z)$ *exists and is continuous. Moreover, there exists a constant* C *which depends on* T, N *and bounds on the coefficients, such that for all* $t \in (0,T]$

$$|\partial_x^\alpha \partial_z^\beta p(t,x,z)| \leq C(1+|x|)^{4|\alpha|+3|\beta|+3N} t^{-\frac{1}{2}(N+|\alpha|+|\beta|)}, \quad x \in \mathbb{R}^N,\ z \in \mathbb{R}^N. \tag{4.15}$$

If $V_0, \dots, V_d$ *are bounded, then the following estimate holds, for all* $t \in (0,T]$

$$|\partial_x^\alpha \partial_z^\beta p(t,x,z)| \leq C t^{-\frac{1}{2}(N+|\alpha|+|\beta|)} \exp\big(-C\frac{|z-x|^2}{t}\big), \quad x \in \mathbb{R}^N,\ z \in \mathbb{R}^N.$$

2. *The alternative SDE* (2.10) *has a density* $p_\theta(t,x,\cdot)$ *such that* $(x,z) \mapsto \partial_x^\alpha \partial_z^\beta p_\theta(t,x,z)$ *exists and is continuous. Moreover, there exists a constant* C *which depends on* T, N *and bounds on the coefficients, such that for all* $t \in (0,T]$

$$|\partial_x^\alpha \partial_z^\beta p_\theta(t,x,z)| \leq C(1+|x|)^{2|\alpha|+3|\beta|+3N} t^{-\frac{1}{2}(N+|\alpha|+|\beta|)}, \quad x \in \mathbb{R}^N,\ z \in \mathbb{R}^N. \tag{4.16}$$

If $V_0, \dots, V_d$ *are bounded, then the following estimate holds, for all* $t \in (0,T]$

$$|\partial_x^\alpha \partial_z^\beta p_\theta(t,x,z)| \leq C t^{-\frac{1}{2}(N+|\alpha|+|\beta|)} \exp\big(-C\frac{|z-x|^2}{t}\big), \quad x \in \mathbb{R}^N,\ z \in \mathbb{R}^N.$$

Proof. 1. The proof is verbatim to Theorem 6.1 of [10] by applying our integration by parts formulae established in Corollary 4.10 and Lemma 3.1 in [31].

2. The inequality (4.16) is similar to the inequality (4.15) but with a different exponent. The procedures to derive this exponent is briefly discussed and the rest procedures are the same as Theorem 6.1 of [10].

Let $\eta = (1,2,3,\dots,N)$ and $\mathbb{1}_{\{z<z_0\}} := \prod_{i=1}^N \mathbb{1}_{\{z^i<z_0^i\}}$. For any $g \in \mathcal{C}_0^\infty(\mathbb{R}^N, \mathbb{R})$ define the function f as

$$f(z_0) := \int_{\mathbb{R}^N} g(z) \mathbb{1}_{\{z<z_0\}}\, \mathrm{d}z$$

is in $\mathcal{C}_p^\infty(\mathbb{R}^N, \mathbb{R})$ and we have $\partial^\eta f = g$. Therefore,

$$\begin{aligned}
\partial_x^\alpha \mathbb{E}\big[(\partial^\beta g)\big(X_t^{x,[\theta]}\big)\big] &= \partial_x^\alpha \mathbb{E}\big[(\partial^{\beta*\eta} f)\big(X_t^{x,[\theta]}\big)\big] \\
&= t^{-\frac{|\alpha|+|\beta|+N}{2}} \mathbb{E}\big[f\big(X_t^{x,[\theta]}\big) I_\alpha^3\big(I^2_{\beta*\eta}(1)\big)(t,x,[\theta])\big] \\
&= t^{-\frac{|\alpha|+|\beta|+N}{2}} \mathbb{E}\bigg[\bigg(\int_{\mathbb{R}^N} g(z) \mathbb{1}_{\{z<X_t^{x,[\theta]}\}}\, \mathrm{d}z\bigg) I_\alpha^3\big(I^2_{\beta*\eta}(1)\big)(t,x,[\theta])\bigg] \\
&= t^{-\frac{|\alpha|+|\beta|+N}{2}} \int_{\mathbb{R}^N} g(z) \mathbb{E}\big[\mathbb{1}_{\{z<X_t^{x,[\theta]}\}} I_\alpha^3\big(I^2_{\beta*\eta}(1)\big)(t,x,[\theta])\big]\, \mathrm{d}z,
\end{aligned} \tag{4.17}$$

where we have used $\partial^\eta f = g$, Proposition 4.7, and Fubini's theorem. The exponent $2|\alpha|+3|\beta|+3N$ comes from Proposition 4.6 and $I_\alpha^3\big(I^2_{\beta*\eta}(1)\big)(t,x,[\theta])$ being a Kusuoka-Stroock process.

□

Theorem 4.11 presents the smoothness result for X_t^{x,δ_x} and it can be generalized to X_t^θ with a general initial distribution $[\theta]$.

Corollary 4.12. *Suppose Assumption 4.5 holds and $V_0, \dots, V_d \in \mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$. Let θ be a random variable in $\mathbb{R}^N$ with finite moments of all orders. For any multi-index β on $\{1, \dots, N\}$ such that $k \geq |\beta| + N + 2$, we see that for all $t \in [0,T]$, X_t^θ has a density $p_\theta(t, \cdot)$ such that $z \mapsto \partial_z^\beta p_\theta(t,z)$ exists and is continuous.*

Proof. If θ is degenerate and $\theta \equiv x$, then the results follow directly from Theorem 4.11 part 1. If θ is non-degenerate, we use the relation

$$p_\theta(t,z) = \int p_\theta(t,x,z)[\theta](\,\mathrm{d}x),$$

and hence, the existence and regularities of $p_\theta(t,x,z)$ is given in Theorem 4.11 part 2. The proof is done by taking the expectation on both sides of the inequality (4.16) with respect to the initial distribution θ and applying the dominated convergence theorem, where we use the assumption that θ has finite moments. □

The above existence and smoothness results on the marginal density of a single object can be extended to the joint distribution for any number of adjacent particles. Namely, for a fixed integer $m \geq 1$, we may construct the system of stochastic processes $(\widetilde{X}_\cdot^0, \widetilde{X}_\cdot^1, \widetilde{X}_\cdot^2, \dots, \widetilde{X}_\cdot^m)$ such that $(\widetilde{X}_\cdot^{m-1}, \widetilde{X}_\cdot^m) \equiv (X_\cdot^\theta, \widetilde{X}_\cdot)$ in (1.1), and $\widetilde{X}_\cdot^i$ depends on the adjacent process $\widetilde{X}_\cdot^{i+1}$ and Brownian motion $\widetilde{B}_\cdot^i$, independent of $\widetilde{X}^{i+1}$, in the same fashion as $(X_\cdot^\theta, \widetilde{X}_\cdot)$ in (1.1) for $i = 0, \dots, m-1$.

Corollary 4.13. *Suppose that Assumption 4.5 holds and $V_0, \dots, V_d \in \mathcal{C}_{b,\mathrm{Lip}}^{k,k,k}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$ and θ has finite moments. Then the joint density of the process $(X_\cdot^\theta, \widetilde{X}_\cdot^1, \widetilde{X}_\cdot^2, \dots, \widetilde{X}_\cdot^m)$ exists and is continuous at any $t \in [0,T]$, where $\widetilde{X}_\cdot^1 \equiv \widetilde{X}_\cdot$ and $\widetilde{X}_\cdot^i$ depend on $\widetilde{X}_\cdot^{i+1}$ in the same way as $(X_\cdot^\theta, \widetilde{X}_\cdot)$ in* (1.1).

Proof. We consider the process evolving in space $\mathbb{R}^{(m+1)N}$ defined by

$$Y_\cdot := (\widetilde{X}_\cdot^0, \widetilde{X}_\cdot^1, \widetilde{X}_\cdot^2, \dots, \widetilde{X}_\cdot^m)$$

and the neighborhood process $\widetilde{Y}_\cdot := (\widetilde{X}_\cdot^{m+1}, \widetilde{X}_\cdot^{m+2}, \dots, \widetilde{X}_\cdot^{2m+1})$. Now $(Y_\cdot, \widetilde{Y}_\cdot)$ satisfies the directed chain structure and it can be proved that this new directed chain SDE structure $Y_\cdot$ also satisfies Assumption 4.5. Hence, the existence and continuity follow from Theorem 4.11 and Corollary 4.12. In particular, if $m = 1$, the coupled process $Y_\cdot$ is defined by

$$Y_t = Y_0 + \sum_{i=1}^{2d} \int_0^T V_i^y(s, Y_s, \mathrm{Law}(Y_s), \widetilde{Y}_s)\,\mathrm{d}B_s^{y,i},$$

where the diffusion coefficients V_i^y, $i = 1, \dots, 2d$ are given by

$$V_i^y := \begin{cases} \big(V_i(s, X_s, \mathrm{Law}(X_s), \widetilde{X}_s^1), \mathbf{0}\big)^T \in \mathbb{R}^{2N}, & i = 1, \dots d, \\ \big(\mathbf{0}, V_{i-d}(s, \widetilde{X}_s^1, \mathrm{Law}(\widetilde{X}_s^1), \widetilde{X}_s^2)\big)^T \in \mathbb{R}^{2N}, & i = d+1, \dots 2d, \end{cases}$$

B^y is independent standard Brownian motions in $\mathbb{R}^{2d}$ and $\mathbf{0} \in \mathbb{R}^N$ is a zero vector. □

4.4 Markov random fields

The existence of density in Theorem 4.11 is closely related to the local Markov property (or Markov random fields) of the directed chain structure. Here, we shall briefly elaborate the relation. A similar topic has been studied by [24] on the undirected graph with locally interactions only on the drift terms. Their approach is to apply a change of

measure under which the diffusion coefficients at one vertex of the undirected graph do not depend on the diffusions at the other vertexes of the graph, to get the factorization of the probability measure. Usually, this Markov property is only discussed for the undirected graph or directed acyclic graph. The finite particle system that approximates the directed chain structure discussed in [13] admits a loop structure in the finite graph. More precisely, the finite system of n particles $(X^{(n)}_{1,\cdot}, \dots, X^{(n)}_{n,\cdot})$ is constructed in a loop of size n so that $X^{(n)}_{1,\cdot}$ depends on $X^{(n)}_{2,\cdot}$, $X^{(n)}_{2,\cdot}$ depends on $X^{(n)}_{3,\cdot}$, $\dots$, $X^{(n)}_{n-1,\cdot}$ depends on $X^{(n)}_{n,\cdot}$ and $X^{(n)}_{n,\cdot}$ depends on $X^{(n)}_{1,\cdot}$. However, when the size n of this loop is forced to be infinity, that is, $n \to \infty$, we can then treat the dependence of the system on any finite subgraph as the system on an acyclic graph [13, Section 3], as (1.2) in our paper. An illustration is given in Figure 1.

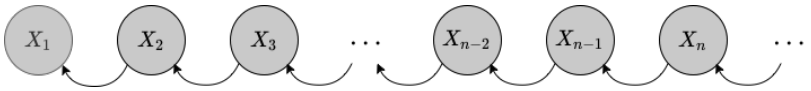

Figure 1: This figure shows a finite cut of the infinite directed chain, i.e., X_k is affected by X_{k+1}.

Proposition 4.14. *The directed chain SDEs described in* (1.2) *form first-order Markov random fields, or we say it has the local Markov property.*

We follow the notation and terminology in [26]. Given a directed graph $G = (V, E)$ with vertices V and edges E, for a vertex $\nu \in V$, let $\mathcal{X}_\nu$ denote the generic space of vertex ν and $\mathrm{pa}(\nu) \in V$ denote all its parents. In the infinite directed chain case, $\mathrm{pa}(X_{k,\cdot}) = X_{k+1,\cdot}$.

Definition 4.15 (Recursive Factorization)**.** *Given a directed graph $G = (V, E)$, we say the probability distribution P^G admits a recursive factorization according to G, if there exists non-negative functions, henceforth referred to as kernels, $k^\nu(\cdot,\cdot), \nu \in V$ defined on $\mathcal{X}_\nu \times \mathcal{X}_{\mathrm{pa}(\nu)}$, such that*

$$\int k^\nu(y_\nu, x_{\mathrm{pa}(\nu)})\mu_\nu(\,\mathrm{d}y_\nu) = 1$$

and P^G has density f^G with respect to a product measure μ, which is defined on the product space $\prod_{\nu\in V}\mathcal{X}_\nu$ by μ_ν a measure defined on each $\mathcal{X}_\nu$, where

$$f^G(x) = \prod_{\nu\in V} k^\nu(x_\nu, x_{\mathrm{pa}(\nu)}).$$

Proof of Proposition 4.14. Thanks to the special structure of the chain, it can be shown that the distribution of the chain satisfies the *recursive factorization* property, where the existence and continuity of the kernel functions are given by Theorem 4.11 and Corollary 4.12. For it, on a filtered probability space, let us consider a system of directed chain diffusion $X_{i,t}$, $i \in \mathbb{N}$, $t \geq 0$ on the infinitely graph with vertices $\mathbb{N} = \{1, 2, \dots\}$. Firstly, coupled diffusion $(X_{1,\cdot}, X_{2,\cdot}) \equiv (X^\theta_\cdot, \widetilde{X}_\cdot)$ satisfies the directed chain stochastic equation and has a continuous density according to Corollary 4.13 and we denote this joint density by $g(\cdot,\cdot) : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. We then construct the chain recursively according to the following rule: Given $X_{k,\cdot}$, initial state $X_{k+1,0}$ and Brownian motion $B_{k+1,\cdot}$ independent of $(X_{1,\cdot}, \dots, X_{k,\cdot}, X_{k+1,0})$, we construct $X_{k+1,\cdot}$ according to the distribution of $(X^\theta_\cdot, \widetilde{X}_\cdot)$.

Defining the kernel functions in the following way

$$k^\nu(x_\nu, x_{\mathrm{pa}(\nu)}) := \begin{cases} g(x_\nu, x_{\mathrm{pa}(\nu)}), & \text{if } \nu = X_1, \\ (g(x_\nu))^{-1} g(x_\nu, x_{\mathrm{pa}(\nu)}), & \text{if } \nu = X_k,\ k \geq 2, \end{cases}$$

i.e., the conditional density of $X_{k+1,t}$ given $X_{k,t}$ for $k \geq 2$, proves the *recursive factorization* property of the chain on any finite cut $(X_{1,t}, X_{2,t}, \ldots, X_{m,t})$, $\forall m \in \mathbb{N}$ of the infinite chain for any $t \in [0,T]$, as well as the local Markov property following from [26, Theorem 3.27], which is also called the first order Markov random field in the context of [24]. This result can also be verified by a filtering problem built upon this directed chains structure, which we omit due to the page limitation. □

4.5 Relation to PDE

We have constructed the integration by parts formulae to argue that the density of directed chain SDEs is smooth in section 4.2, which is also the tool for constructing solutions to a related PDE problem. To simplify notation, we will omit the time dependency in coefficients of SDEs through this section, i.e. we will write $V(X_t^{x,[\theta]}, [X_t^\theta], \widetilde{X}_t) := V(t, X_t^{x,[\theta]}, [X_t^\theta], \widetilde{X}_t)$. In particular, we are interested in the function

$$U(t,x,[\theta]) = \mathbb{E}[g(X_t^{x,[\theta]}, [X_t^\theta])],$$

$t \in [0,T]$, $x \in \mathbb{R}^N$ for some sufficiently smooth function g. Here $X_\cdot^\theta$ is the solution of (2.1)-(2.2) with random initial θ and $X_t^{x,[\theta]}$ is the solution to (2.10) with deterministic initial x. They depend on a neighborhood process $\widetilde{X}_\cdot$ with an initial independent random vector $\tilde{\theta}$. Recall the flow property (2.13) in section 2.3. It follows that for every $0 \leq t \leq t+h \leq T$, $x \in \mathbb{R}^d$,

$$U(t+h,x,[\theta]) = \mathbb{E}[g(X_{t+h}^{x,[\theta]}, [X_{t+h}^\theta])] = \mathbb{E}\big[U(t, X_h^{x,[\theta]}, [X_h^\theta])\big].$$

Hence

$$\begin{aligned} &U(t+h,x,[\theta]) - U(t,x,[\theta]) \\ &= U(t,x,[X_h^\theta]) - U(t,x,[\theta]) + \mathbb{E}\big[U(t, X_h^{x,[\theta]}, [X_h^\theta]) - U(t,x,[X_h^\theta])\big] \\ &= I - \mathbb{E}[J], \end{aligned} \tag{4.18}$$

where we define $I = U(t,x,[X_h^\theta]) - U(t,x,[\theta])$ and $J = U(t, X_h^{x,[\theta]}, [X_h^\theta]) - U(t,x,[X_h^\theta])$. Applying the chain rule introduced in [9] to I and Ito's formula to J, we have

$$\begin{aligned} I = \int_0^h \mathbb{E}\Big[&\sum_{i=1}^N V_0^i(X_r^\theta, [X_r^\theta], \widetilde{X}_r) \partial_\mu U(t,x,[X_r^\theta], X_r^\theta)_i \\ &+ \frac{1}{2} \sum_{i,j=1}^N [\sigma\sigma^\top(X_r^\theta, [X_r^\theta], \widetilde{X}_r)]_{i,j} \partial_{v_j} \partial_\mu U(t,x,[X_r^\theta], X_r^\theta)_i \Big] \, \mathrm{d}r, \\ J = \int_0^h &\sum_{i=1}^N V_0^i(X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r) \partial_{x_i} U(t, X_r^{x,[\theta]}, [X_h^\theta]) \, \mathrm{d}r \\ &+ \frac{1}{2} \int_0^h \sum_{i,j=1}^N [\sigma\sigma^\top(X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r)]_{i,j} \partial_{x_i} \partial_{x_j} U(t, X_r^{x,[\theta]}, [X_h^\theta]) \, \mathrm{d}r \\ &+ \int_0^h \sum_{j=1}^d \sum_{i=1}^N V_j^i(X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r) \partial_{x_i} U(t, X_r^{x,[\theta]}, [X_h^\theta]) \, \mathrm{d}B_r^j. \end{aligned}$$

For the meaning of the differential operator with respect to the measure ∂_μ appearing in I, we refer to Section 2.1. Then let us plug I, J into (4.18) and take expectation, divide by h on both sides, and send h to 0, we will end up with a PDE of the form given below

$$\begin{aligned} (\partial_t - \mathcal{L}) U(t,x,[\theta]) &= 0 \quad \text{for } (t,x,[\theta]) \in (0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N), \\ U(0,x,[\theta]) &= g(x,[\theta]) \quad \text{for } (x,[\theta]) \in \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N), \end{aligned} \tag{4.19}$$

where $g : \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}$ and the operator $\mathcal{L}$ acts on smooth enough functions $F : \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}^N$ defined by

$$\begin{aligned}\mathcal{L}F(x,[\theta]) =&\mathbb{E}\bigg[\sum_{i=1}^N V_0^i(x,[\theta],\tilde{\theta})\partial_{x_i}F(x,[\theta]) + \frac{1}{2}\sum_{i,j=1}^N [\sigma\sigma^\top(x,[\theta],\tilde{\theta})]_{i,j}\partial_{x_i}\partial_{x_j}F(x,[\theta])\bigg]\\ &+\mathbb{E}\bigg[\sum_{i=1}^N V_0^i(\theta,[\theta],\tilde{\theta})\partial_\mu F(x,[\theta],\theta)_i + \frac{1}{2}\sum_{i,j=1}^N [\sigma\sigma^\top(\theta,[\theta],\tilde{\theta})]_{i,j}\partial_{v_j}\partial_\mu F(x,[\theta],\theta)_i\bigg]. \quad (4.20)\end{aligned}$$

The expectation in the first line of (4.20) is taken with respect to the random variable $\tilde{\theta}$ due to the appearance of the neighborhood process in the difference J, while the expectation in the second line is taken with respect to the joint distribution of $\theta, \tilde{\theta}$, as an application of the chain rule introduced in [9] to the difference I.

Evidently, a proper condition for the initial g is needed to establish the existence of the solution to PDE (4.19). Such a directed chain type SDE has not been considered before, the closest works are related to the existence of solutions to Kolmogorov PDE on the Wasserstein space; see [32, 11, 12]. In [3], g is assumed to have bounded second-order derivatives. The smoothness on g is relaxed in [10]. In particular, they assume g belongs to a class of functions that can be approximated by a sequence of functions with polynomial growth, and also satisfy certain growth conditions on its derivatives. Hence, they claim that g is not necessarily differentiable. We emphasize that a detailed discussion on the choice of assumptions in g is beyond the scope of this paper, but we conjecture that some similar results should also hold for our case and will include this in our future research.

References

[1] Erhan Bayraktar, Suman Chakraborty, and Ruoyu Wu, *Graphon mean field systems*, arXiv preprint arXiv:2003.13180 (2020). MR4663492

[2] Erhan Bayraktar and Ruoyu Wu, *Graphon particle system: Uniform-in-time concentration bounds*, arXiv preprint arXiv:2105.11040 (2021). MR4514837

[3] Rainer Buckdahn, Juan Li, Shige Peng, and Catherine Rainer, *Mean-field stochastic differential equations and associated PDEs*, The Annals of Probability **45** (2017), no. 2, 824 – 878. MR3630288

[4] Peter E. Caines and Minyi Huang, *Graphon mean field games and the GMFG equations*, 2018 IEEE Conference on Decision and Control (CDC), 2018, pp. 4129–4134. MR4139292

[5] Peter E. Caines and Minyi Huang, *Graphon mean field games and the GMFG equations: ϵ-nash equilibria*, 2019 IEEE 58th Conference on Decision and Control (CDC), 2019, pp. 286–292. MR4340663

[6] Pierre Cardaliaguet, *Notes on mean field games. from p.-l.lions lectures at college de france (2010).*, Online Notes Available at, 2010.

[7] René Carmona and François Delarue, *Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics*, The Annals of Probability **43** (2015), no. 5, 2647 – 2700. MR3395471

[8] Thomas Cass, *Smooth densities for solutions to stochastic differential equations with jumps*, Stochastic Processes and their Applications **119** (2009), no. 5, 1416–1435. MR2513114

[9] Jean-François Chassagneux, Dan Crisan, and François Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, arXiv preprint arXiv:1411.3009 (2014). MR4493576

[10] Dan Crisan and Eamon McMurray, *Smoothing properties of McKean–Vlasov sdes*, Probability Theory and Related Fields **171** (2018), no. 1, 97–148. MR3800831

[11] Paul-Eric Chaudru de Raynal and Noufel Frikha, *From the backward kolmogorov pde on the wasserstein space to propagation of chaos for mckean-vlasov sdes*, Journal de Mathématiques Pures et Appliquées **156** (2021), 1–124. MR4338452

[12] Paul-Eric Chaudru de Raynal and Noufel Frikha, *Well-posedness for some non-linear sdes and related pde on the wasserstein space*, Journal de Mathématiques Pures et Appliquées **159** (2022), 1–167. MR4377993

[13] Nils Detering, Jean-Pierre Fouque, and Tomoyuki Ichiba, *Directed chain stochastic differential equations*, Stochastic Processes and their Applications **130** (2020), no. 4, 2519–2551.

[14] Yichen Feng, Jean-Pierre Fouque, and Tomoyuki Ichiba, *Linear-quadratic stochastic differential games on directed chain networks*, Journal of mathematics and statistical science **7** (2021), no. 2.

[15] Yichen Feng, Jean-Pierre Fouque, and Tomoyuki Ichiba, *Linear-quadratic stochastic differential games on random directed networks*, Journal of mathematics and statistical science **7** (2021), no. 3.

[16] Nobuyuki Ikeda, Ichiro Shigekawa, and Setsuo Taniguchi, *The Malliavin calculus and long time asymptotics of certain Wiener integrals*, Miniconference on Linear Analysis and Functional Spaces, Australian National University, Mathematical Sciences Institute, 1985, pp. 46–113. MR0825515

[17] Hiroshi Kunita, *Stochastic differential equations and stochastic flows of diffeomorphisms*, Ecole d'été de probabilités de Saint-Flour XII-1982, Springer, 1984, pp. 143–303. MR0876080

[18] S. Kusuoka and D. Stroock, *The partial Malliavin calculus and its application to non-linear filtering*, Stochastics **12** (1984), no. 2, 83–142. MR0747781

[19] S Kusuoka and D Stroock, *Applications of the Malliavin calculus, part iii*, J. Fac. Sci. Univ. Tokyo Sect IA Math **34** (1987), 391–442. MR0914028

[20] Shigeo Kusuoka, *Applications of the Malliavin calculus, part iii*, Journal of the Faculty of Science. University of Tokyo. Section IA. Mathematics **32** (1984), 271–306. MR0780762

[21] Shigeo Kusuoka, *Malliavin calculus revisited*, Journal of Mathematical Sciences-University of Tokyo **10** (2003), no. 2, 261–278. MR1987133

[22] Shigeo Kusuoka and Daniel Stroock, *Applications of the Malliavin calculus, part i*, Stochastic Analysis (Kiyosi Itô, ed.), North-Holland Mathematical Library, vol. 32, Elsevier, 1984, pp. 271–306. MR0780762

[23] Shigeo Kusuoka and Daniel W. Stroock, *Applications of the Malliavin calculus. ii*, Journal of the Faculty of Science, the University of Tokyo. Sect. 1 A, Mathematics **32** (1985), 1–76. MR0783181

[24] Daniel Lacker, Kavita Ramanan, and Ruoyu Wu, *Locally interacting diffusions as markov random fields on path space*, Stochastic Processes and their Applications **140** (2021), 81–114. MR4276494

[25] Daniel Lacker and Agathe Soret, *A case study on stochastic games on large graphs in mean field and sparse regimes*, arXiv preprint arXiv:2005.14102 (2020). MR4435025

[26] S.L. Lauritzen, *Graphical models*, Oxford Statistical Science Series, Clarendon Press, 1996. MR1419991

[27] P Malliavin et al., *Sur certaines intégrales stochastiques oscillantes*, C.R. SEANCES ACAD. SCI., 295, 295-300 (1982). MR0681601

[28] Ming Min, Ruimeng Hu, and Tomoyuki Ichiba. Directed chain generative adversarial networks, *Proceedings of the 40th International Conference on Machine Learning*, ICML'23 JMLR.org, 2023, 1033, 19 pages.

[29] David Nualart, *Malliavin calculus and its applications*, CBMS Regional Conference Series in Mathematics, no. 110, American Mathematical Soc., 2009. MR2498953

[30] David Nualart and Moshe Zakai, *The partial Malliavin calculus*, Séminaire de Probabilités XXIII, Springer, 1989, pp. 362–381. MR1022924

[31] Setsuo Taniguchi, *Applications of Malliavin's calculus to time-dependent systems of heat equations*, Osaka Journal of Mathematics **22** (1985), no. 2, 307 – 320. MR0800974

[32] Alvin Tse, *Higher order regularity of nonlinear fokker-planck pdes with respect to the measure component*, Journal de Mathématiques Pures et Appliquées **150** (2021), 134–180. MR4248465

Acknowledgments. We are grateful to both Reviewers and Associate Editor for correcting our mistakes and providing insightful feedback.