Smoothness of Directed Chain Stochastic Differential Equations and its Applications (Probability Symposium)

ICHIBA, TOMOYUKI

Smoothness of Directed Chain Stochastic Differential Equations and its Applications

TOMOYUKI ICHIBA
Department of Statistics & Applied Probability
University of California Santa Barbara

1 Introduction

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we shall consider the following system of stochastic differential equations for a pair $(X^{\theta}_{\cdot}, \widetilde{X}^{\theta}_{\cdot})$ of N-dimensional stochastic processes:

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

for $t \geq 0$ with the distributional constraint

$$[X^{\theta}_t, t\geq 0] := \mathrm{Law}(X^{\theta}_t, t\geq 0) = \mathrm{Law}(\widetilde{X}_t, t\geq 0) =: [\widetilde{X}_t, t \geq 0], \tag{2}$$

where $V_i : [0,T]\times\mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N)\times\mathbb{R}^N \to \mathbb{R}^N$, $i = 0,1,\ldots,d$ are some smooth coefficients, $B := (B^1, \cdots, B^d)$ is the standard d-dimensional Brownian motion. We assume the initial value $\theta \in \mathcal{P}_2(\mathbb{R}^N)$ is independent of B and $\widetilde{X}_0$, and $\widetilde{X}_0$ is independent of B. Here, $\mathcal{P}_2(\mathbb{R}^N)$ is the set of probability 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 $W_2(\mu,\nu) := \inf_{\Pi\in\mathcal{P}_{\mu,\nu}}(\int_{M\times M} d^2(x,y)\Pi(\mathrm{d}x,\mathrm{d}y))^{1/2}$, where $\mathcal{P}_{\mu,\nu}$ denotes the class of probability measures on $M\times M$ with marginals μ and ν. Note that the law $[X^{\theta}_{\cdot}]$ of $X^{\theta}_{\cdot}$ depends on the law $[\widetilde{X}_{\cdot}]$ of $\widetilde{X}_{\cdot}$ and they are the same marginal law. Setting $B^0_t \equiv t$, $t\geq 0$, the above equation is rewritten as

$$\begin{aligned} X^{\theta}_t &= \theta + \sum_{i=0}^{d}\int_0^t V_i(s, X^{\theta}_s, [X^{\theta}_s], \widetilde{X}_s)\mathrm{d}B^i_s; \quad t\geq 0, \\ [\widetilde{X}_t, \geq 0] &= \mathrm{Law}\,(\widetilde{X}_t, t\geq 0) = \mathrm{Law}\,(X^{\theta}_t, t\geq 0) = [X^{\theta}_t, t\geq 0]. \end{aligned} \tag{3}$$

We call the system (1) with the constraint (2) the system of *directed chain stochastic differential equation*.

For example, with $N = 1$, $u \in [0,1]$, and some smooth functions $b_{0,i} : \mathbb{R}_+ \times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, for $i = 0,1,\ldots,d$, we define the coefficients

$$V_i(t,x,\mu,y) := u\, b_{0,i}(t,x,y) + (1-u)\int_{\mathbb{R}} b_{0,i}(t,x,z)\mathrm{d}\mu(z)$$

as a linear combination of two terms. When $u = 0$, the equation becomes a McKean-Vlasov equation; When $u = 1$, there is no contribution from the distribution $[X^{\theta}_{\cdot}]$.

Proposition 1 (Uniqueness of weak solution)**.** *Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^N)$ be a fixed reference measure. Suppose that V_i, $i = 0,1,\ldots,d$ are Lipschitz continuous and grow at most linearly in the sense that for every $T > 0$, there exists a constant c_T such that for every $0\leq t\leq T$, $x_1, y_1, x, y \in \mathbb{R}^N$, $\mu_1,\mu_2 \in \mathcal{P}_2(\mathbb{R}^N)$,*

$$\sup_i |V_i(t,x_1,\mu_1,y_1) - V_i(t,x_2,\mu_2,y_2)| \leq c_T(|x_1-x_2| + |y_1-y_2| + W_2(\mu_1,\mu_2), \tag{4}$$

$$\sup_i \sup_{0\leq t\leq T} |V_i(t,x,\mu,y)| \leq c_T(1+|x|+|y|+W_2(\mu,\mu_0)). \tag{5}$$

Then there exists a unique weak solution $(X^{\theta}_{\cdot}, \widetilde{X}_{\cdot}, B)$ $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ to the system (1) *of stochastic differential equations with the distributional constraint* (2)*.*

The analysis of the special case with $N = d = 1$, $V_1 \equiv 1$ is considered in [DFI]. The name, directed chain, is coined from the fact that the joint distribution of $(X^{\theta}_{\cdot}, \widetilde{X}_{\cdot})$ in (1) can be approximated by the limit of the joint distribution of $(X^1_{\cdot}, X^2_{\cdot})$ from a finite particle system on the vertexes $i = 1, \ldots, n$, where the process $X^i_{\cdot}$ at vertex i depends on $X^{i+1}_{\cdot}$ at vertex $i+1$ via the equation $\mathrm{d}X^i_t = V_0(t, X^i_t, \overline{\mu}_t, X^{i+1}_t)\mathrm{d}t + \mathrm{d}B^i(t)$ with the empirical measure $\overline{\mu}_t := n^{-1}\sum_{i=1}^{n} \boldsymbol{\delta}_{X^i_t}$ of the particle system for $i = 1, \ldots, n-1$ and $\mathrm{d}X^n_t = V_0(t, X^n_t, \overline{\mu}_t, X^1_t)\mathrm{d}t + \mathrm{d}B^n(t)$, $t \ge 0$. Here, $\boldsymbol{\delta}_x$ is the Dirac measure at the point x. Under some reasonable assumptions, the joint distribution of $(X^1_{\cdot}, X^2_{\cdot})$ converges weakly to that of $(X^{\theta}_{\cdot}, \widetilde{X}_{\cdot})$ in (1), as $n \to \infty$.

The motivation of studying (1) comes from the interacting particles of sparse network [2], [10], [16] as well as the mean field games [5], [7], [11], [13], [18], particularly on the infinite random graph. In this short note, we discuss the smoothness of the joint distribution. Smoothness of solution to McKean-Vlasov equation has been studied by [1], [8], [9].

2 Smoothness

2.1 Lion's derivatives in the Wasserstein space $\mathcal{P}_2$

Let us recall the Wasserstein distance between two measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ is written as

$$W_2(\mu, \nu) := \inf\{\|X - Y\|_2 : [X] = \mu, [Y] = \nu\}.$$

For a function $u : \mathcal{P}_2 \to \mathbb{R}$, we denote by U “extension” (or lift) to $L^2(\Omega', \mathcal{F}', \mathbb{P}')$ defined by

$$U(X) = u(\mathrm{Law}(X)), \quad \mathrm{Law}(X) = [X] = \mu.$$

Here, $(\Omega', \mathcal{F}', \mathbb{P}')$ is an atomless Polish space. Following [6], we say u is differentiable at $[X] \in \mathcal{P}$, if there exists X' such that $[X'] = [X]$ and the lift U is Fréchet differentiable at X'.

For example, when $u : \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{R}$ is given by

$$u(\mu) := \prod_{i=1}^{n} \int_{\mathbb{R}^N} \varphi_i(x)\mathrm{d}\mu(x)$$

for some smooth functions $\varphi_i \in C_c^{\infty}(\mathbb{R}^N)$, then $U(X)$ and its gradient $\mathcal{D}U(X)$ are given by

$$U(X) := \prod_{i=1}^{n} \mathbb{E}[\varphi_i(X)]; \quad [X] = \mu, \quad \mathcal{D}U(X) = \sum_{i=1}^{n} \big(\prod_{j \ne i} \mathbb{E}[\varphi_j(X)]\big) D\varphi_i(X),$$

and hence, for every $v \in \mathbb{R}^N$, $\mu \in \mathcal{P}_2(\mathbb{R}^N)$,

$$\mathcal{D}_{\mu}u(\mu)(v) = \sum_{i=1}^{n} \Big(\prod_{j \ne i} \int_{\mathbb{R}^N} \varphi_j(z)\mathrm{d}\mu(z)\Big) D\varphi_i(v),$$

which does not depend on the random vector X.

2.2 Smoothness of coefficients

We say $V : \mathbb{R}_+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}^N$ belongs to $\mathcal{C}^{1,1,1}_{b,\mathrm{Lip}}$, if each component V^i of $V = (V^1, \ldots, V^N)$ has bounded, Lipschitz continuous derivatives $\partial_x V^i$, $\tilde{\partial} V^i$ in the second and fourth variables, respectively, in the sense of P.L. Lions [6] with at most linear growth property, i.e., there exists a constant $c > 0$ such that

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

$$|\partial_\mu V^i(t,x,\mu,y,v) - \partial_\mu V^i(t,x',\mu',y',v)| \le c(|x-x'| + |y-y'| + |v-v'| + W_2(\mu,\mu'))$$

for $(t,x,\mu,y,v),(t,x',\mu',y',v') \in [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N$. Moreover, we say V belongs to $\mathcal{C}^{k,k,k}_{b,\mathrm{Lip}}$, if it has bounded, Lipschitz, k times derivatives $\partial_x^\gamma \widetilde{\partial}^{\widetilde{\gamma}} \partial_v^\beta \partial_\mu^\alpha V^i$ in multi-indexes $(\alpha,\beta,\gamma,\widetilde{\gamma})$, $|\alpha| + |\beta| + |\gamma| + |\widetilde{\gamma}| \le k$ with at most linear growth property.

Now we consider the pathwise-unique, strong solution to auxiliary stochastic equation

$$X_t^{x,[\theta]} = x + \sum_{i=0}^{d} \int_0^t V_i(s, X_s^{x,[\theta]}, [X_s^\theta], \widetilde{X}_s) \mathrm{d}B_s^i \,, \tag{6}$$

given the solution pair $(X_\cdot^\theta, \widetilde{X}_\cdot)$ in (1). More specifically, we set $\widetilde{X}_0 =: \widetilde{\theta}$ and

$$X^{x,[\theta],\widetilde{\theta}} = x + \sum_{i=0}^{d} \int_0^t V_i(s, X_s^{x,[\theta],\widetilde{\theta}}, [X_s^\theta], \widetilde{X}_s) \mathrm{d}B_s^i \,. \tag{7}$$

Then by the pathwise uniqueness, we have $X_s^{x,[\theta],\widetilde{\theta}}|_{x=\theta} = X_s^\theta$; $0 \le s \le T$.

2.3 Flow property

For different initial points $x, x' \in \mathbb{R}^N$, the corresponding solutions $X^{x,[\theta],\widetilde{\theta}}$ and $X^{x',[\theta],\widetilde{\theta}}$ in (7) satisfy that for every $T > 0$, there exists a constant $c_T > 0$ such that

$$\mathbb{E}[\sup_{t \le s \le T} |X^{x,[\theta],\widetilde{\theta}} - X^{x',[\theta],\widetilde{\theta}}|^2] \le c_T |x - x'|^2$$

by Lipschitz continuity and Burkholder-Davis-Gundy inequality. With a slightly abuse of notation, we write $X_\cdot^{t,x,[\theta],\widetilde{\theta}}$ for the process $X_\cdot^{\cdot,[\theta],\widetilde{\theta}}$ with $X_t^{t,[\theta],\widetilde{\theta}} = x$, and $(X_\cdot^{t,\theta}, \widetilde{X}_\cdot^{t,\widetilde{\theta}})$ for the process $(X_\cdot^{\cdot,\theta}, \widetilde{X}_\cdot^{\cdot,\widetilde{\theta}})$ with $(X_t^{t,\theta}, \widetilde{X}_t^{t,\widetilde{\theta}}) = (\theta, \widetilde{\theta})$, we have the flow property

$$(X_r^{s, X_s^{t,x,[\theta],\widetilde{\theta}}, [X_s^{t,\theta}], \widetilde{X}_s^{t,\widetilde{\theta}}}, X_r^{s, X_s^{t,\theta}}, \widetilde{X}_r^{s, \widetilde{X}_s^{t,\widetilde{\theta}}}) = (X_r^{t,x,[\theta],\theta}, X_r^{t,\theta}, \widetilde{X}_r^{t,\widetilde{\theta}}); \quad 0 \le t \le s \le r \le T \,.$$

2.4 Partial Malliavin Calculus

Let us consider the Malliavin derivative operator $\boldsymbol{D}$ and its adjoint operator $\boldsymbol{\delta}$. Let σ be the $N \times d$ matrix with columns $V_1, \dots, V_d$. If there is **no** interaction with the neighborhood process $\widetilde{X}_\cdot$, the McKean-Vlasov equation in (6) has the derivative

$$\partial_x X_t^{x,[\theta]} = \boldsymbol{D}_r X_t^{x,[\theta]} \sigma^\top (\sigma\sigma^\top)^{-1} (r, X_r^{x,[\theta]}, [X_r^\theta] \quad) \partial_x X_r^{x,[\theta]}; \quad r \le t \,,$$

however, because of the interaction with $\widetilde{X}_\cdot$, in general,

$$\partial_x X_t^{x,[\theta]} \ne \boldsymbol{D}_r X_t^{x,[\theta]} \sigma^\top (\sigma\sigma^\top)^{-1} (r, X_r^{x,[\theta]}, [X_r^\theta], \widetilde{X}_r) \partial_x X_r^{x,[\theta]}; \quad r \le t \,,$$

To overcome this difficulty, we shall apply the following partial Malliavin derivatives from [15], [19].

Let us take the rational numbers $\mathbb{Q}_T := \mathbb{Q} \cap [0,T]$ in $[0,T]$ and define the σ-field $\mathcal{G} := \sigma(\{\widetilde{X}_t, t \in \mathbb{Q}_T\})$ (countably generated) and the family of subspaces defined by the orthogonal complement

$$K(\omega) := \langle \boldsymbol{D}\widetilde{X}_t(\omega), t \in \mathbb{Q}_T \rangle^\perp$$

to the subspace generated by $\{\boldsymbol{D}\widetilde{X}_t(\omega), t \in \mathbb{Q}_T\}$. Then the family $\mathcal{H} := \{K(\omega), \omega \in \Omega\}$ has a measurable projection. We define the partial derivative operator $\boldsymbol{D}^{\mathcal{H}} : \mathbb{D}^{1,2} \to L^2(\Omega, \mathcal{H})$, namely, for $F \in \mathbb{D}^{1,2}$, $\boldsymbol{D}^{\mathcal{H}} F = \text{Proj}_{\mathcal{H}}(\boldsymbol{D}F) = \text{Proj}_{K(\omega)}(\boldsymbol{D}F)(\omega)$ with associated norm

$$\|F\|_{\mathbb{D}^{k,p}_{\mathcal{H}}} := (\mathbb{E}[|F|^p] + \sum_{j=1}^{k} \mathbb{E}[\|\boldsymbol{D}^{\mathcal{H},(j)} F\|^p_{\mathcal{H}}])^{1/p},$$

where $\boldsymbol{D}^{(j)}$ is the j-th order derivative and $\boldsymbol{D}^{\mathcal{H},(j)} F := \text{Proj}_{\mathcal{H}}(\boldsymbol{D}^{(j)} F) = \text{Proj}_{K(\omega)}(\boldsymbol{D}^{(j)} F)(\omega)$.

Similar to the Malliavin calculus, there is an adjoint operator $\boldsymbol{\delta}_{\mathcal{H}}(u) := \boldsymbol{\delta}(\text{Proj}_{\mathcal{H}}(u))$ of $\boldsymbol{D}^{\mathcal{H}}$ if $\text{Proj}_{\mathcal{H}} u \in \text{Dom}(\boldsymbol{\delta})$, as well as the integration by parts formula $\mathbb{E}[\langle u, \boldsymbol{D}^{\mathcal{H}} F\rangle] = \mathbb{E}[\langle \text{Proj}_{\mathcal{H}} u, \boldsymbol{D}F\rangle] = \mathbb{E}[F\, \boldsymbol{\delta}_{\mathcal{H}} u]$ for any $u \in \text{Dom}(\boldsymbol{\delta}_{\mathcal{H}})$, $F \in \mathbb{D}^{1,2}$.

Let E be a separable Hilbert space. For $r \in \mathbb{R}, q, M \in \mathbb{N}$ let us define the family $\mathbb{K}^q_r(E, M)$ of processes $\Psi : [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \to \mathbb{D}^{M,\infty}(E)$ satisfying the following:

$$(t, x, [\theta]) \mapsto \partial^\gamma_x \partial^\beta_v \partial^\alpha_\mu \Psi(t, x, [\theta], v) \in L^p(\Omega)$$

exists and continuous for all $p \geq 1$ and multi-indexes (α, β, γ) with $|\alpha| + |\beta| + |\gamma| \leq M$, and

$$\sup_{v \in (\mathbb{R}^N)^{\sharp\beta}} \sup_{t \in [0,T]} \frac{1}{t^{r/2}} \|\partial^\gamma_x \partial^\beta_v \partial^\alpha_\mu \Psi(t, x, [\theta], v)\|_{\mathbb{D}^{m,p}_{\mathcal{H}}(E)} \leq C(1 + |x| + \|\theta\|_2)^q$$

for every $p \geq 1$, $m \in \mathbb{N}$ and multi-indexes (α, β, γ) with $|\alpha| + |\beta| + |\gamma| + m \leq M$. This is a modification of $\mathbb{K}^q_r$ in [9] for the smoothness of the density function of $X^{x,[\theta]}$.

Proposition 2. *Assume $V_i \in C^{1,1,1}_{b,Lip}(\mathbb{R}_+ \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$. There exists a modification of $X^{x,[\theta]}$ such that the map $x \mapsto X^{x,[\theta]}_t$ is almost surely differentiable, and for $t \geq 0$,*

$$\partial_x X^{x,\theta}_t = Id_N + \sum_{i=0}^{d} \int_0^t \partial V_i(s, X^{x,[\theta]}_s, [X^\theta_s], \widetilde{X}_s) \partial_x X^{x,[\theta]}_s \mathrm{d}B^i_s.$$

The maps $\theta \mapsto X^\theta_t$, $\theta \mapsto X^{x,[\theta]}_t$ are Fréchet differentiable in $L^2(\Omega)$ with gradients $\mathcal{D}X^{x,[\theta]}_t$ and $\mathcal{D}X^{x,[\theta]}_t$ satisfying

$$\mathcal{D}X^{x,[\theta]}_t(\gamma) = \sum_{i=0}^{d} \int_0^t [\partial V_i \mathcal{D}X^{x,[\theta]}_s + \widetilde{\partial} V_i \mathcal{D}\widetilde{X}_s(\gamma) + \mathcal{D}V'_i(\mathcal{D}X^\theta_s(\gamma))] \mathrm{d}B^i_s,$$

$$\mathcal{D}X^\theta_t(\gamma) = \gamma + \sum_{i=0}^{d} \int_0^t [\partial V_i \mathcal{D}X^\theta_s(\gamma) + \widetilde{\partial} V_i \mathcal{D}\widetilde{X}_s(\gamma) + \mathcal{D}V'_i(\mathcal{D}X^\theta_s(\gamma))] \mathrm{d}B^i_s,$$

for $\gamma \in L^2(\Omega)$, $t \geq 0$.

Moreover, the map $[\theta] \mapsto X^{x,[\theta]}_t$ is differentiable with the derivative $\partial_\mu X^{x,[\theta]}_t$ satisfying

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

where $(X_s^{\theta'})'$ is a copy of X_s^{θ}, $\partial_x(X_s^{v,[\theta]})'$ is a copy of $\partial_x X_s^{v,[\theta]}$ and $\partial_\mu(X_s^{\theta',[\theta]})' = \partial_\mu(X_s^{x,[\theta]})'_{x=\theta'}$ on a probability space with $\mathcal{D}X_t^{x,[\theta]}(\gamma) = \mathbb{E}'[\partial_\mu X_t^{x,[\theta]}(\theta')\gamma']$. Furthermore, $X_t^{x,[\theta]}, X_t^{\theta} \in \mathbb{D}^{1,\infty}$, and $\boldsymbol{D}_r^{\mathcal{H}} X^{x,[\theta]} = (\boldsymbol{D}_r^{\mathcal{H},j}(X^{x,[\theta]})^i)_{1\le j\le N, 1\le i\le d}$ satisfies, for $0 \le r \le t$

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

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

2.5 Characterization of the auxiliary process

Assume $V_i \in C_{b,\mathrm{Lip}}^{k,k,k}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$ for $i = 1, \dots, d$. Then the map satisfies

$$(t, x, [\theta]) \mapsto X_t^{x,[\theta]} \in \mathbb{K}_0^1(\mathbb{R}^N, k)\,.$$

If, in addition, V_i are uniformly bounded, then $(t, x, [\theta]) \mapsto X_t^{x,[\theta]} \in \mathbb{K}_0^0(\mathbb{R}^N, k)$. Proof is based on the first order derivatives (cf. [9]).

Now we define operators $I_{(i)}^j$, $j = 1, 2, 3$, $\mathcal{I}_{(i)}^1$, $\mathcal{I}_{(i)}^3$ on $\Psi \in \mathbb{K}_r^q(\mathbb{R}, n)$ with $\alpha = (i)$, $(t, x, [\theta]) \in [0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N)$,

$$\begin{aligned}
I_{(i)}^1(\Psi)(t,x,[\theta]) &:= \frac{1}{\sqrt{t}}\boldsymbol{\delta}_{\mathcal{H}}\big(r \mapsto \Psi(t,x,[\theta])(\sigma^\top(\sigma\sigma^\top)^{-1}(r, X_r^{x,\theta}, [X_r^{\theta}], \widetilde{X}_r)\partial_x X_r^{x,\mu})_i\big)\,,\\
I_{(i)}^2(\Psi)(t,x,[\theta]) &:= \sum_{j=1}^{N} I_{(j)}^1((\partial_x X_t^{x,\mu})_{j,i}^{-1}\Psi(t,x,[\theta]))\,,\\
I_{(i)}^3(\Psi)(t,x,[\theta]) &:= I_{(i)}^1(\Psi)(t,x,[\theta]) + \sqrt{t}\partial^i\Psi(t,x,[\theta])\\
\mathcal{I}_{(i)}^1(\Psi)(t,x,[\theta],v_1) &:= \frac{1}{\sqrt{t}}\delta_{\mathcal{H}}\Big(r \mapsto \big(\sigma^\top\big(\sigma\sigma^\top\big)^{-1}(r, X_r^{x,\mu}, [X_r^{\theta}], \widetilde{X}_r)\,,\\
&\qquad \partial_x X_r^{x,\mu}(\partial_x X_t^{x,\mu})^{-1}\partial_\mu X_t^{x,[\theta]}(v_1)\big)_i\Psi(t,x,[\theta])\Big)\,,\\
\mathcal{I}_{(i)}^3(\Psi)(t,x,[\theta],v_1) &:= \mathcal{I}_{(i)}^1(\Psi)(t,x,[\theta],v_1) + \sqrt{t}(\partial_\mu\Psi)_i(t,x,[\theta],v_1)\,.
\end{aligned} \tag{8}$$

2.6 Integration-by-parts formulae

Assume $V_i \in C_{b,\mathrm{Lip}}^{k,k,k}([0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N) \times \mathbb{R}^N; \mathbb{R}^N)$ and also the uniform ellipticity of the diffusion coefficients. For $f \in C_b^\infty(\mathbb{R}^N, \mathbb{R})$, $\Psi \in \mathbb{K}_r^q(\mathbb{R}, n)$, we have

- If $|\alpha| \le n \wedge k$, then

$$\mathbb{E}\big[\partial_x^\alpha\big(f\big(X_t^{x,[\theta]}\big)\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)I_\alpha^1(\Psi)(t,x,[\theta])\big]\,;$$

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

$$\mathbb{E}\big[(\partial^\alpha f)\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)I_\alpha^2(\Psi)(t,x,[\theta])\big]\,;$$

- If $|\alpha| \le n \wedge k$, then

$$\partial_x^\alpha\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big] = t^{-|\alpha|/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)I_\alpha^3(\Psi)(t,x,[\theta])\big]\,;$$

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

$$\partial_x^\alpha\mathbb{E}\big[(\partial^\beta f)\big(X_t^{x,[\theta]}\big)\Psi(t,x,[\theta])\big] = t^{-(|\alpha|+|\beta|)/2}\mathbb{E}\big[f\big(X_t^{x,[\theta]}\big)I_\alpha^3\big((I_\beta^2\Psi)\big)(t,x,[\theta])\big]\,.$$

For $f \in C_b^\infty(\mathbb{R}^N, \mathbb{R})$ and $\Psi \in \mathbb{K}_r^q(\mathbb{R}, n)$, we have

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

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

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

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

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

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

For every $f \in C_k^\infty(\mathbb{R}^N;\mathbb{R})$, multi-index α on $\{1,\dots,N\}$ with $|\alpha| \le k-2$,

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

where δ_x is a Dirac point mass at $x \in \mathbb{R}^N$, and

$$J_{(i)}(\Phi)(t,x) \;:=\; I_{(i)}^3(\Phi)(t,x,\delta_x) + \mathcal{I}_{(i)}^3(t,x,\delta_x)\,;\quad t \ge 0$$

with $J_\alpha(\Phi) \;:=\; J_{\alpha_n}\circ J_{\alpha_{n-1}}\circ\cdots\circ J_{\alpha_1}(\Phi)$. Particularly, there exists a constant $c > 0$ such that

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

for $0 \le t \le T$, $x \in \mathbb{R}^N$. Moreover, with $|\alpha| + |\beta| \le k-2$,

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

and $I_\beta^2(J_\alpha(1)) \in \mathbb{K}_0^{4|\alpha|+3|\beta|}(\mathbb{R}, k-2-|\alpha|-|\beta|)$. Thus, $X_t^{x,\delta_x} \;=\; X_t^\theta|_{\theta=x}$ has a probability density function $p(t,x,z)$ such that $(x,z) \mapsto \partial_x^\alpha\partial_z^\beta p(t,x,z)$ exists and is continuous.

2.7 Smoothness of the joint density

Proposition 3. *Let α, β be multi-indices on $\{1,\dots,N\}$ and $k \ge |\alpha|+|\beta|+N+2$. Under these assumptions of the uniform ellipticity of σ and the smoothness of coefficients $V_i \in C_{b,Lip}^{k,k,k}$, the solution X_t^θ to the directed chain SDE* (1) *with $\theta \equiv x \in \mathbb{R}^N$ at time $t \ge 0$ has a density $p(t,x,\cdot)$ such that $(x,z) \mapsto \partial_x^\alpha\partial_z^\beta p(t,x,z)$ exist and is continuous. Moreover, there exists a constant C which depends on T, N and bounds on the coefficients, such that*

$$|\partial_x^\alpha\partial_z^\beta p(t,x,z)| \;\le\; C(1+|x|)^{4|\alpha|+3|\beta|+3N}\,t^{-(N+|\alpha|+|\beta|)\,/\,2} \tag{9}$$

for $t \in (0,T]$, $x, z \in \mathbb{R}^N$. Furthermore, if V_i, $i = 0,\dots,d$ are bounded, then

$$|\partial_x^\alpha\partial_z^\beta p(t,x,z)| \;\le\; Ct^{-(N+|\alpha|+|\beta|)\,/\,2}\exp\Big(-\frac{C|x-z|^2}{t}\Big) \tag{10}$$

for $t \in (0,T]$, $x, z \in \mathbb{R}^N$.

The above existence and smoothness results on the marginal density $p(t,x,z)$ of a single particle can be extended to the joint distribution of adjacent particles. That is, We extend the pair $(X^{\theta}_{\cdot}, \widetilde{X}_{\cdot})$ to consider the system $(\widetilde{X}^0, \widetilde{X}^1, \dots, \widetilde{X}^m)$, such that the joint distribution of adjacent pair is determined by the directed chain stochastic differential equation 1, namely, $[\widetilde{X}^{k-1}_{\cdot}, \widetilde{X}^k_{\cdot}] = [X^{\theta}_{\cdot}, \widetilde{X}_{\cdot}]$ for $k = 1, \dots, m$.

Corollary. *Under the same assumptions on the coefficients, the joint density of* $(\widetilde{X}^0_t, \widetilde{X}^1_t, \dots, \widetilde{X}^m_t)$ *exists and is continuous for* $t \ge 0$. *Particularly, the joint density of* $(X^{\theta}_{\cdot}, \widetilde{X}_{\cdot})$ *exists and is continuous.*

The applications of the smoothness of the joint distribution are the recursive factorization of the first order Markov random field [16], some connection to a class of non-linear partial differential equations, smoothness of the filtering equation and the analysis of master equation of the mean-field game and the mean-field control problems on the directed chain graph.

2.8 Relation to PDE

Let us consider time-homogeneous coefficients. For the function $U(t,x,[\theta]) := \mathbb{E}[g(X^{x,[\theta]}_t, [X^{\theta}_t])]$, $t \in [0,T]$, $x \in \mathbb{R}^N$, by the flow property, we have

$$U(t+h, x, [\theta]) = \mathbb{E}[g(X^{x,[\theta]}_{t+h}, [X^{\theta}_{t+h}])] = \mathbb{E}[U(t, X^{x,[\theta]}_h, [X^{\theta}_h])]$$

for $t \ge 0$, $0 \le t \le T - h$. Then we come up with a PDE of the form

$$(\partial_t - \mathcal{L})U(t,x,[\theta]) = 0, \quad (t,x,[\theta]) \in (0,T] \times \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N),$$

$$U(0,x,[\theta]) = g(x,[\theta]), \quad (x,[\theta]) \in \mathbb{R}^N \times \mathcal{P}_2(\mathbb{R}^N),$$

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

$$\begin{aligned}\mathcal{L}F(x,[\theta]) :=& \mathbb{E}\Big[\sum_{i=1}^N V_0^i(x,[\theta],\tilde{\theta})\partial_{x_i}F(x,[\theta]) + \frac{1}{2}\sum_{i,j=1}^N [\sigma\sigma^{\top}(x,[\theta],\tilde{\theta})]_{i,j}\partial_{x_i}\partial_{x_j}F(x,[\theta])\Big] \\ &+ \mathbb{E}\Big[\sum_{i=1}^N V_0^i(\theta,[\theta],\tilde{\theta})\partial_{\mu}F(x,[\theta],\theta)_i + \frac{1}{2}\sum_{i,j=1}^N [\sigma\sigma^{\top}(\theta,[\theta],\tilde{\theta})]_{i,j}\partial_{v_j}\partial_{\mu}F(x,[\theta],\theta)_i\Big]\end{aligned} \tag{11}$$

cf. [4], [9] for McKean-Vlasov SDE.

2.9 Relation to Mimicking problem

The mimicking problem is to obtain the marginal distribution of some non-Markovian process by a unique strong solution to the stochastic differential equation

$$\mathrm{d}Y_t = b_0(Y_t)\mathrm{d}t + b_1(Y_t)\mathrm{d}B^y(t); \quad t \ge 0, \quad Y_0 := \xi \tag{12}$$

for Y with some smooth functions $b_0 : \mathbb{R}^N \to \mathbb{R}^N, b_1 : \mathbb{R}^N \to \mathbb{R}^{N \times N}$. B^y is the n-dimensional standard Brownian motion. cf. [3], [12], [17].

Conversely, it follows from the smoothness of the solution in Proposition 3 that there exist $(X_{\cdot}, \widetilde{X}_{\cdot})$ and functions V_i, $i = 0,1$, such that $(X_0, \widetilde{X}_0)$ are independent and

$$[Y_{\cdot}] = [X_{\cdot}] = [\widetilde{X}_{\cdot}],$$

where the pair $(X_\cdot, \widetilde{X}_\cdot)$ satisfies the directed chain equation

$$\mathrm{d}X_t = V_0(X_t, \widetilde{X}_t)\mathrm{d}t + V_1(X_t, \widetilde{X}_t)\mathrm{d}B_t\,; \quad t \ge 0\,, \tag{13}$$

driven by another standard Brownian motion B independent of $\widetilde{X}$.

Research supported in part by the National Science Foundation under grant DMS-20-08427. Part of research is joint work [14] with M. MIN.

Bibliography

[1] BAÑOS, D. (2018). The Bismut-Elworthy-Li Formula for Mean-Field Stochastic Differential Equations. *Annales de l'Insitut Henri Poincaré* **54** 220-233.

[2] BAYRAKTAR, E. & WU, R. (2021). Graphon particle system: Uniform-in-time concentration bounds. *arXiv: 2105.11040.*

[3] BRUNICK, G. & SHREVE, S. (2010). Mimicking an Itô process by a solution of a stochastic differential equation. *Annals of Applied Probability* **23** 1584–1628.

[4] BUCKDAHN, R., LI, J., PENG, S., & RAINER, C. (2017). Mean-field stochastic differential equations and associated PDEs. *Annals of Probability* **45** 824–878.

[5] CAINES, P.E. & HUANG, M. (2019). Graphon Mean Field Games and the GMFG Equations: ε-Nash Equilibria. In *2019 IEEE 58th Conference on Decision and Control (CDC)* 286–292.

[6] CARDALIAGUET, P. (2010). Notes on Mean Field Games. From P.-L. Lions lecture at College de France.

[7] CARMONA & DELARUE (2015). Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics. *Annals of Probability* **43** 2647–2700.

[8] CHAUDRU DE RAYNAL, P.-E. (2020). Strong Well-Posedness of McKean-Vlasov Stochastic Differential Equation with Hölder Drift. *Stochastic Process and their Applications* **130** 79–107.

[9] CRISAN, D. & MCMURRAY, E. (2018). Smoothing properties of McKean-Vlasov SDEs. *Probability Theory and Related Fields* **171** 97–148.

[10] DETERING, N, FOUQU, J.-P. & ICHIBA, T. (2020). Directed chain stochastic differential equations. *Stochastic Processes and their Applications* **130** 2519–2551.

[11] FENG, Y., FOUQUE J.-P., & ICHIBA, T. (2021). Linear-quadratic stochastic differential games on directed chain networks. *Journal of Mathematics and Statistical Science* **7** 25–67.

[12] GYÖNGY, I. (1986) Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Theory Related Fields* **71** 501–516.

[13] HUANG, M., MALHAMÉ, R.P., & CAINES, P.E. (2006). Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6** 221–251.

[14] ICHIBA, T. & MING, M. (2022). Smoothness of Directed Chain Stochastic Differential Equations. *arXiv: 2202.09354*.

[15] KUSUOKA, S. & STROOCK, D. W. (1984). The partial Malliavin calculus and its application to non-linear filtering. *Stochastics* **12** 83–142.

[16] LACKER, D., RAMANAN, K. & WU, R. (2021). Locally interacting diffusions as Markov random fields on path space. *Stochastic Processes and their Applications* **140** 81–114.

[17] LACKER, D., SHKOLNIKOV, M. & ZHANG, J. (2022). Superposition and mimicking theorems for conditional McKean-Vlasov equations. to appear in *Journal of European Mathematical Society*.

[18] LASRY, J. & LIONS, P. (2007). Mean field games. *Jpn. J. Math.* **2**, 229–260.

[19] NUALART, D. & ZAKAI, M. (1989). The partial Malliavin calculus. *Séminaire de Probabilités XXIII* 362–381.

Department of Statistics and Applied Probability,
South Hall
University of California,
Santa Barbara, CA 93106
E-mail address: *ichiba@pstat.ucsb.edu*

カリフォルニア大学サンタバーバラ　一場　知之