Malliavin differentiability of McKean-Vlasov SDEs with common noise*

Jianhai Bao†   Gonçalo dos Reis‡ §   Zac Wilde¶

Abstract

We establish the Malliavin differentiability of McKean-Vlasov stochastic differential equations (MV-SDEs) with common noise under the global Lipschitz assumption in the space variable and the measure variable. Our result gives also meaning to the Malliavin derivative of the conditional law with respect to the common noise. As an application, we derive an integration by parts formula on the Wiener space for the class of common noise MV-SDEs under consideration.

Keywords: McKean-Vlasov stochastic differential equation; common noise; conditional measure flow; Lions derivative; Malliavin calculus; integration by parts formula.
MSC2020 subject classifications: Primary 60H07, Secondary 60H10; 60H30.
Submitted to ECP on January 11, 2025, final version accepted on October 1, 2025.

1 Introduction

McKean-Vlasov stochastic differential equations (SDEs for short) are a class of SDEs whose coefficients have dependence not only on the state process but also on the law of the solution process [4]. In the present work, we turn to focus on McKean-Vlasov SDEs with common noise, which, in literature, are also termed as the conditional McKean-Vlasov SDEs [10]. As far as such SDEs are concerned, there are two sources of noise input into the system, in which one is an idiosyncratic source of randomness and the other is a so-called common noise (i.e., the path $(\omega_t^0)_{t\geq 0}$ driven by the Brownian motion which we condition upon pathwise). More precisely, in this paper we work on the following dynamics on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma^0(t, X_t, \mu_t)\mathrm{d}W_t^0 + \sigma^1(t, X_t, \mu_t)\mathrm{d}W_t^1, \quad t > 0; \quad X_0 = \xi, \qquad (1.1)$$

where $b : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma^0 : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\otimes m_0}$, and $\sigma^1 : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\otimes m}$, the Brownian motion $(W_t^1)_{t\geq 0}$, supported on $(\Omega^1, \mathcal{F}^1, (\mathcal{F}_t^1)_{t\geq 0}, \mathbb{P}^1)$,

†Center for Applied Mathematics, Tianjin University, 300072 Tianjin, P.R. China. E-mail: jianhaibao@tju.edu.cn
‡University of Edinburgh, School of Mathematics, Edinburgh, EH9 3FD, UK. E-mail: G.dosReis@ed.ac.uk
§Centro de Matemática e Aplicações (Nova MATH), FCT, UNL, 2829-516 Caparica, Portugal
¶University of Edinburgh, School of Mathematics, Edinburgh, EH9 3FD, UK. E-mail: zwilde@ed.ac.uk

is referred to as an m-dimensional idiosyncratic noise, the Brownian motion $(W_t^0)_{t\geq 0}$, carried on $(\Omega^0, \mathcal{F}^0, (\mathcal{F}_t^0)_{t\geq 0}, \mathbb{P}^0)$, as an m_0-dimensional common noise, and $\mu_t := \mu_t(\omega^0) = \text{Law}(X_t(\omega^0, \cdot))$ for $\mathbb{P}^0$-almost ω^0 (i.e., $\mu_t = \mathbb{P}^1 \circ (\mathbb{E}[X_t|\mathcal{F}_t^0])^{-1})$. Hereinbefore, $\mathcal{P}(\mathbb{R}^d)$ is the space of probability measures over $\mathbb{R}^d$, and μ_t can be understood as a random variable mapping $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ to $\mathcal{P}(\mathbb{R}^d)$; see [5, Section 4.3] or [7] for related details.

As mentioned previously, the drift and diffusion coefficients in (1.1) depend on the path $(\omega_t^0)_{t\geq 0}$. Thus, one can interpret (1.1) in the light of the theory on rough differential equations (RDEs for short). One of our contributions lies in extending [3], which addressed the Malliavin differentiability of standard RDEs, to the mean-field setting, where the rough path under consideration is a Brownian motion path. According to [5, p.110-112], the random distribution flow $(\mu_t)_{t>0}$ corresponding to (1.1) solves the following SPDE:

$$\mathrm{d}\mu_t = \Big(-\operatorname{div}\big(b(\cdot, \mu_t)\mu_t\big) + \frac{1}{2}\operatorname{trace}\big(\nabla^2((\sigma^1(\sigma^1)^\top)(\cdot, \mu_t)\mu_t)\big) \Big)\,\mathrm{d}t - \operatorname{div}\big((\sigma^0(\cdot, \mu_t)\mathrm{d}W_t^0)\mu_t\big),$$

where $(\sigma^1)^\top$ denotes the transpose of σ^1, ∇^2 means the second-order gradient operator, and div represents the divergence operator. The aforementioned SPDE, which is also named as the nonlinear stochastic Fokker-Planck equation, is understood in the weak sense. It is worthy to stress that the flow (μ_t) involved in (1.1) is random whereas the same counterpart associated with the standard McKean-Vlasov SDEs (without common noise) is deterministic. This discrepancy introduces involved difficulties into the analysis of the Malliavin differentiability for $(X_t)_{t\geq 0}$ governed by (1.1).

In the past few years, the Malliavin differentiability of classical SDEs advanced greatly. In particular, via a Picard approximation approach, [11, Theorem 2.2.1] handled the case that the drift term and the diffusion term involved are globally Lipschitz. Nevertheless, the Malliavin differentiability for SDEs with drifts of super-linear growth is left open for a long time, which is extremely unexpected. In [9], the issue mentioned above on the Malliavin differentiability is addressed by the aid of stochastic Gâteaux differentiability and ray absolute continuity. Recently, the research on the Malliavin differentiability for McKean-Vlasov SDEs has also made some progress; see, for instance, [14, Proposition 3.5] for the globally Lipschitz setting. Since the distribution variable involved in the coefficients associated with McKean-Vlasov SDEs is deterministic, there is no additional term in the linear SDE solved by the Malliavin derivative; see, e.g., [14, (3.26)] or [8, (2)]. When the coefficients of McKean-Vlasov SDEs under consideration is Lipschitz continuous in the measure variable under the L^2-Wasserstein distance, the Malliavin differentiability was established in [8, Theorem 3.5] and [8, Proposition 4.8] via the Picard iteration trick and the interacting particle system (IPS for short) limits when the underlying drifts satisfy the one-sided Lipschitz condition and fulfil the globally Lipschitz condition in the spatial variable, respectively.

Our contribution: Malliavin differentiability of McKean-Vlasov SDEs with common noise under the global Lipschitz assumption. Specially, we extend the classical Malliavin variational results to McKean-Vlasov SDEs with common noise, where the deterministic coefficients satisfy a global space-measure Lipschitz condition. To derive the Malliavin differentiability with respect to the common noise, we adopt a similar approach to that of [8] by appealing to the celebrated [11, Lemma 1.2.3]. In detail, we employ the following strategy: we first study the Malliavin differentiability of the underlying IPS followed by passing to the limit via Propagation of Chaos (PoC for short), and then studying how the Malliavin regularity transfers across the particle limit to the limiting equation. Essentially from (1.1), our construction allows one to give meaning to the Malliavin derivative of the conditional distribution flow μ_t in relation to W^0; see Theorem 3.1 below.

With regarding to the McKean-Vlasov SDEs, the Bismut-Elworthy-Li formulae were provided in [6] for the associated decoupled SDEs with deterministic initial value. Later on, the Bismut formula for Lions derivatives was initially established in [14] for distribution dependent SDEs, which the initial value is a random variable. Subsequently, [14] was extended to distribution-path dependent SDEs, where the associated drift term satisfies a monotone condition in the spatial variable. In the present paper, as an application of our main theory derived, we seek to establish a conditional integration by parts formula (instead of the corresponding Bismut-Elworthy-Li formula, where the order between the Markov operator and the L-derivative is different).

2 Notation and preliminary results

2.1 Notation and spaces

For $M, N \in \mathbb{R}^{m \otimes n}$, define the inner product $\langle M, N \rangle_{\mathrm{HS}} = \mathrm{Trace}(M^\top N)$, which induces the associated Hilbert-Schmidt norm $\| \cdot \|_{\mathrm{HS}}$. For $g : \mathbb{R}^m \to \mathbb{R}^n$, write $g(x) = (g^1(x), \ldots, g^n(x))$, and define $\nabla_x g$ as the Jacobian matrix $(\partial_i g^j)_{i,j}$ with $\partial_i := \frac{\partial}{\partial x^i}$ for $1 \le i \le m$, $1 \le j \le n$. For $t > 0$ and a vector-valued or matrix-valued function f defined on $\mathbb{R}^d$, we set the uniform $\|f\|_{0,t} := \sup_{0 \le s \le t} \|f(s)\|_{\mathrm{HS}}$. For a given terminal $T > 0$, write $\mathcal{C}_T := C([0, T]; \mathbb{R}^d)$, which is endowed with the uniform norm $\| \cdot \|_{\mathcal{C}_T}$, as the path space of continuous functions $\omega : [0, T] \to \mathbb{R}^d$. Throughout the paper, we shall work on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \ge 0}, \mathbb{P})$, where $\Omega = \Omega^0 \times \Omega^1$, $(\mathcal{F}, \mathbb{P})$ is the completion of $(\mathcal{F}^0 \times \mathcal{F}^1, \mathbb{P}^0 \times \mathbb{P}^1)$, and $(\mathcal{F}_t)_{t \ge 0}$ is the complete and right continuous augmentation of $(\mathcal{F}_t^0 \times \mathcal{F}_t^1)_{t \ge 0}$. Let $\mathbb{E}^0$, $\mathbb{E}^1$ and $\mathbb{E}$ denote the expectation operators under the probability measures $\mathbb{P}^0$, $\mathbb{P}^1$, and $\mathbb{P}$, respectively. The space of probability measures over $\mathbb{R}^d$ with finite rth moment for $r \ge 1$, written as $\mathcal{P}_r(\mathbb{R}^d)$, is a Polish space under the Wasserstein distance:

$$\mathcal{W}_r(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^r \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{r}}, \quad \mu, \nu \in \mathcal{P}_r(\mathbb{R}^d),$$

where $\Pi(\mu, \nu)$ (i.e., the family of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$) is the set of couplings for μ and ν (that is, $\pi \in \Pi(\mu, \nu)$ satisfying $\pi(\cdot \times \mathbb{R}^d) = \mu(\cdot)$ and $\pi(\mathbb{R}^d \times \cdot) = \nu(\cdot)$). Let $\mathrm{Supp}(\mu)$ denote the support of the probability measure μ.

2.2 Basics on Malliavin calculus

In this subsection, we aim to recall some basics on Malliavin calculus to make the content self-contained; see e.g. the monographs [11] and [16, pages 8-9]. For $T > 0$ fixed, denote by $\mathcal{H}$ the Cameron-Martin space defined as below:

$$\mathcal{H} = \left\{ h \in \mathcal{C}_T \Big| h_0 = \mathbf{0}, h_t' \text{ exists a.e. in } t, \|h\|_{\mathcal{H}}^2 := \int_0^T |h_t'|^2 \mathrm{d}t < \infty \right\},$$

which is a Hilbert space with the inner product $\langle f, g \rangle_{\mathcal{H}} := \int_0^T \langle f_t', g_t' \rangle \mathrm{d}t$. Let μ_T be the distribution of $W_{[0,T]} = (W_t)_{t \in [0,T]}$ on the path space $\mathcal{C}_T$ of paths starting at the origin, where $(W_t)_{t \ge 0}$ is a d-dimensional Brownian motion. In fact, μ_T is the so-called Wiener measure so that the coordinate process $W_t(\omega) = \omega_t$ is a d-dimensional Brownian motion. For $F \in L^2(\mathcal{C}_T, \mu_T)$, $F(W_{[0,T]})$ is called Malliavin differentiable along direction $h \in \mathcal{H}$ if the directional derivative $D_h F(W_{[0,T]}) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(F(W_{[0,T]} + \varepsilon h) - F(W_{[0,T]}))$ exists in $L^2(\mathcal{C}_T, \mu_T)$. If $\mathcal{H} \ni h \mapsto D_h F \in L^2(\mathcal{C}_T, \mu_T)$ is bounded, via the Riesz representation theorem, then there exists a unique $DF(W_{[0,T]}) \in L^2(\mathcal{C}_T \to \mathcal{H}, \mu_T)$ such that $\langle DF(W_{[0,T]}), h \rangle = D_h F(W_{[0,T]})$ holds true μ_T-a.s. In this case, we write $F(W_{[0,T]}) \in \mathcal{D}(D)$ (the domain of D) and call $DF(W_{[0,T]})$ the Malliavin gradient (derivative) of $F(W_{[0,T]})$.

The adjoint operator $(\delta, \mathcal{D}(\delta))$ of $(D, \mathcal{D}(D))$ is called the Malliavin divergence. Via the integration by parts formula, we have

$$\mathbb{E}(D_h F) = \mathbb{E}(F\delta(f)), \quad F \in \mathcal{D}(D), \ h \in \mathcal{D}(\delta).$$

In particular, when $h \in L^2(\mathcal{C}_T \to \mathcal{H}, \mu_T)$ is an adapted stochastic process, then $h \in \mathcal{D}(\delta)$ and $\delta(h) = \int_0^T \langle h'_t, \mathrm{d}W_t \rangle$. Otherwise, $\delta(h)$ is the Skorokhod stochastic integral.

2.3 Lions differentiability and the empirical projection map

To begin, we recall the notion of the Lions derivative; see e.g. [14, Definition 1.1], which indeed coincides with the one given in [4] via a lift-up approach. The functional $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is called L-differentiable at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ if the functional $L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu) \ni \phi \mapsto f(\mu \circ (I + \phi)^{-1})$ is Fréchet differentiable at $\mathbf{0} \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$, where I means the identity map and $\mu \circ (I + \phi)^{-1}$ is the push-forward measure of μ by $I + \phi$. Namely, there is (hence, unique) $\gamma \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that

$$\lim_{\mu(|\phi|^2) \to 0} \frac{f(\mu \circ (I + \phi)^{-1}) - f(\mu) - \mu(\langle \gamma, \phi \rangle)}{\sqrt{\mu(|\phi|^2)}} = 0.$$

In this case, we write $\partial_\mu f(\mu) = \gamma$ and call it the L-derivative of f at μ. For a vector-valued or matrix-valued function f on $\mathbb{R}^d$, the L-derivative (if it exists) is defined in the component-wise way. If the L-derivative $\partial_\mu f(\mu)$ exists for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, then f is called L-differentiable. If, moreover, for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a μ-version $\partial_\mu f(\mu)$ such that $\partial_\mu f(\mu)(x)$ is jointly continuous in $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, we write $f \in C^{(1,0)}(P_2(\mathbb{R}^d))$. g is called differentiable on $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ if, for any $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, $g(\cdot, \mu)$ is differentiable at x and $g(x, \cdot)$ is L-differentiable at μ. If moreover, $\nabla g(\cdot, \mu)(x)$ and $\partial_\mu g(x, \cdot)(\mu)(y)$ are jointly continuous in $(x, y, \mu) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, where ∇ is the gradient operator on $\mathbb{R}^d$, we write $g \in C^{1,(1,0)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$. In addition, for $g \in C^{1,(1,0)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, we write $g \in C_b^{1,(1,0)}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ when the gradient and the L-derivative are uniformly bounded in all variables.

Below, we retrospect the definition on the empirical project of a map; see, for instance, [4, Definition 5.34].

Definition 2.1. *Given* $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ *and* $N \in \mathbb{N}$, *the empirical projection map* u^N *of* u *is defined via the following relation:*

$$u^N : (\mathbb{R}^d)^N \ni \mathbf{x}^N := (x^1, \ldots, x^N) \mapsto u(\bar{\mu}^N) \quad \text{with} \quad \bar{\mu}^N(\mathrm{d}y) := \frac{1}{N} \sum_{i=1}^N \delta_{x^i}(\mathrm{d}y).$$

Let $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ be a continuously L-differentiable map and u^N be the associated empirical projection map. Then, according to [4, Proposition 5.35], u^N is differentiable in $(\mathbb{R}^d)^N$ and satisfies the identity:

$$(\nabla_j u^N)(\mathbf{x}^N) = N^{-1} \partial_\mu u(\bar{\mu}^N)(x^j), \quad \mathbf{x}^N \in (\mathbb{R}^d)^N, \tag{2.1}$$

where $\nabla_j u$ denotes the gradient in the j-th component x^j.

2.4 McKean-Vlasov SDEs with common noise and associated IPS

Throughout the paper, we assume the following Assumptions.

($\mathbf{H}_1$) there is a constant $L > 0$ such that for $t \in [0, T]$, $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$|b(t, x, \mu) - b(t, y, \nu)| \vee \|\sigma^0(t, x, \mu) - \sigma^0(t, y, \nu)\|_{\mathrm{HS}} \vee \|\sigma^1(t, x, \mu) - \sigma^1(t, y, \nu)\|_{\mathrm{HS}}$$
$$\leq L(|x - y| + W_2(\mu, \nu)).$$

$(\mathbf{H}_2)$ $b(t, \cdot, \cdot), \sigma^0(t, \cdot, \cdot), \sigma^1(t, \cdot, \cdot) \in C_b^{1,(1,0)}(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ for all $t \in [0, T]$.

It is standard that (1.1) is strongly well-posed and has a uniform second-order moment bound in a finite-time interval; see, for example, [5, Proposition 2.8] and [2, Proposition 2.2] for more details.

Theorem 2.2. *Under* $(\mathbf{H}_1)$, *there exists a unique solution to* (1.1) *in* $L^2(\mathcal{C}_T; \mathbb{P})$ *for a finite horizon* $T > 0$, *and there is a constant* $K = K(L, T) > 0$ *such that* $\mathbb{E}\|X\|_{\mathcal{C}_T}^2 \leq K(1 + \mathbb{E}|\xi|^2)$ *as long as the initial value* $X_0 = \xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$.

In the sequel, we copy the dynamics of $(X_t)_{t \geq 0}$ solving (1.1) to form a *non-interacting particle system (non-IPS for short)* written as below: for any $i = 1, \ldots, N$,

$$\mathrm{d}X_t^i = b(t, X_t^i, \mu_t^i)\mathrm{d}t + \sigma^0(t, X_t^i, \mu_t^i)\mathrm{d}W_t^0 + \sigma^1(t, X_t^i, \mu_t^i)\mathrm{d}W_t^{1,i}, \quad t > 0; \quad X_0^i = \xi^i, \quad (2.2)$$

where $(\xi^i)_{1 \leq i \leq N}$ are $L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ i.i.d. random variables with the same distribution as ξ (all are independent), $((W_t^{1,i})_{t \geq 0})_{1 \leq i \leq N}$ are mutually independent Brownian motion on $(\Omega^1, \mathcal{F}^1, (\mathcal{F}^1)_{t \geq 0}, \mathbb{P}^1)$. In terms of [5, Proposition 2.11], one has $\mu_t^i := \mathrm{Law}(X_t^i(\omega^0, \cdot)) = \mu_t^i$ for $\mathbb{P}^0$-almost any ω^0, and meanwhile $\mu_t^i = \mu_t$ for any $t \geq 0$, where $(\mu_t)_{t \geq 0}$ is determined by (1.1). The dynamics of a corresponding IPS to (2.2) is formulated as follows:

$$\mathrm{d}X_t^{i,N} = b(t, X_t^{i,N}, \bar{\mu}_t^N)\mathrm{d}t + \sigma^0(t, X_t^{i,N}, \bar{\mu}_t^N)\mathrm{d}W_t^0 + \sigma^1(t, X_t^{i,N}, \bar{\mu}_t^N)\mathrm{d}W_t^{1,i}, \quad (2.3)$$

for any $i = 1, \ldots, N$, where $X_0^{i,N} := \xi^i$, and the empirical measure $\bar{\mu}_t^N := \frac{1}{N} \sum_{j=1}^N \delta_{X_t^{j,N}}$.

The following Lemma plays a vital role in analysing the Malliavin differentiability for McKean-Vlasov SDEs with common noise.

Lemma 2.3. *Under* $(\mathbf{H}_1)$, *the IPS* (2.3) *has a unique solution* $((X_t^{i,N})_{0 \leq t \leq T})_{1 \leq i \leq N}$ *and there exists a constant* $K = K(T, L) > 0$ *such that* $\max_{1 \leq l \leq N} \mathbb{E}\|X^l\|_{\mathcal{C}_T}^2 \leq K(1 + \mathbb{E}|\xi|^2)$ *in case of* $\xi \in L^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$. *Moreover, for* $((X_t^{i,N})_{0 \leq t \leq N})_{1 \leq i \leq N}$ *the solution to* (2.2),

$$\lim_{N \to \infty} \left(\max_{1 \leq i \leq N} \mathbb{E}\|X_\cdot^{i,N} - X_\cdot^i\|_{\mathcal{C}_T}^2 + \sup_{0 \leq t \leq T} \mathbb{E}\mathcal{W}_2(\bar{\mu}_t^N, \mu_t^1)^2 \right) = 0. \quad (2.4)$$

Proof. Note that (2.3) can be rewritten in the state space $(\mathbb{R}^d)^N$. Correspondingly, under $(\mathbf{H}_1)$, the lift-up drift term and the diffusion term satisfy the global Lipschitz condition; see e.g. [8, Lemma 4.4]. Therefore, it is more or less standard that (2.3) is strongly well-posed and the uniform second-order moment estimate is available. The assertion (2.4) has been established in [5, Theorem 2.12] so we herein omit the details. $\square$

Remark 2.4. Concerning the validity of [2, Proposition 2.5], the higher-order (i.e., $p > 4$) moment of the initial distribution is necessitated. Nevertheless, for our purpose, we merely require a qualitative (rather than quantitative) estimate (2.4) so the requirement on the higher-order moment on the initial distribution is necessary.

3 Malliavin differentiability

In this section, we aim at showing that the Malliavin differentiability of $(X_t)_{t \in [0,T]}$ with respect to the common noise and the idiosyncratic noise, one by one.

Theorem 3.1. *Assume that* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$ *hold. Then, the Malliavin derivative of* $(X_t)_{t \geq 0}$ *solving* (1.1) *with respect to the common Brownian motion* W^0, *written as* $D^0 X$, *satisfies*

the SDE: for all $0 \leq s \leq t \leq T$

$$
\begin{aligned}
D_s^0 X_t = \sigma^0(s, X_s, \mu_s) \\
+ \int_s^t \left(\nabla b(r, X_r, \mu_r) D_s^0 X_r \mathbb{E}^1 \left[\partial_\mu b(r, z, \mu_r)(X_r) D_s^0 X_r \right] \big|_{z=X_r} \right) \mathrm{d}r \\
+ \int_s^t \left(\nabla \sigma^0(r, X_r, \mu_r) D_s^0 X_r \mathbb{E}^1 \left[\partial_\mu \sigma^0(r, z, \mu_r)(X_r) D_s^0 X_r \right] \big|_{z=X_r} \right) \mathrm{d}W_r^0 \\
+ \int_s^t \left(\nabla \sigma^1(r, X_r, \mu_r) D_s^0 X_r + \mathbb{E}^1 \left[\partial_\mu \sigma^1(r, z, \mu_r)(X_r) D_s^0 X_r \right] \big|_{z=X_r} \right) \mathrm{d}W_r^1.
\end{aligned}
\tag{3.1}
$$

where $D_s^0 X_t = \mathbf{0}_{d \times m_0}$ *for* $s > t$. *Moreover, there exists a unique solution to* (3.1) *and a positive constant* $K = K(T, L)$ *such that for any*

$$
\sup_{0 \leq s \leq T} \mathbb{E} \left[\left\| D_s^0 X_{\cdot} \right\|_{\mathcal{C}_T}^2 \right] \leq K(1 + \mathbb{E}[|\xi|^2]).
\tag{3.2}
$$

To proceed, we make two comments on Theorem 3.1.

Remark 3.2. Obviously, Theorem 3.1, SDE (3.1) and (3.2) can be written just the same for the non-IPS (X^i) of (2.2) leading to $(D^0 X^i)_i$ and (3.3): namely, for $i = 1, \ldots, N$,

$$
\begin{aligned}
D_s^0 X_t^i = \sigma^0(s, X_s^i, \mu_s^i) \\
+ \int_s^t \left(\nabla b(r, X_r^i, \mu_r^i) D_s^0 X_r^i + \mathbb{E}^1 \left[\partial_\mu b(r, z, \mu_r^i)(X_r^i) D_s^0 X_r^i \right] \big|_{z=X_r^i} \right) \mathrm{d}r \\
+ \int_s^t \left(\nabla \sigma^0(r, X_r^i, \mu_r^i) D_s^0 X_r^i + \mathbb{E}^1 \left[\partial_\mu \sigma^0(r, z, \mu_r^i)(X_r^i) D_s^0 X_r^i \right] \big|_{z=X_r^i} \right) \mathrm{d}W_r^0 \\
+ \int_s^t \left(\nabla \sigma^1(r, X_r^i, \mu_r^i) D_s^0 X_r^i + \mathbb{E}^1 \left[\partial_\mu \sigma^1(r, z, \mu_r^i)(X_r^i) D_s^0 X_r^i \right] \big|_{z=X_r^i} \right) \mathrm{d}W_r^{1,i}.
\end{aligned}
\tag{3.3}
$$

For each i the process $D^0 X^i$ of (3.3) shares the same distribution as $D^0 X$ of (3.1).

As a second observation, in (3.3) the first stochastic integral is equal to

$$
\sum_{j=1}^{m_0} \int_s^t \left(\nabla \sigma_j^0(r, X_r^i, \mu_r^i) D_s^0 X_r^i + \mathbb{E}^1 \left[\partial_\mu \sigma_j^0(r, z, \mu_r^i)(X_r^i) D_s^0 X_r^i \right] \big|_{z=X_r^i} \right) \mathrm{d}W_r^{0,j},
\tag{3.4}
$$

where $W^{0,j}$ denotes the j-th component of W^0, and σ_i^0 stands for the i-th column vector of σ^0. The second stochastic integral in (3.3) can be understood in the same manner.

Remark 3.3. A close inspection of the proof for Theorem 3.1 reveals that (i) the qualitative conditional PoC and (ii) the uniform (in particle number N) second-order moment boundedness of $D^0 X^{i,N}$ is essential. Provided that the drift b fulfils a monotone condition in the spatial variable and is Lipschitz continuous under the L^2-Wasserstein distance in the measure variable, along with the precondition that σ^0, σ^1 are Lipschitz continuous in the spatial variable and the measure variable, the essentials (i) and (ii) mentioned previously can also be established. For the monotone setting, there is much more sophisticated work to be done in order to implement Step 2 and Step 3 in the proof of Theorem 3.1. Nonetheless, in the present work, we aim to reveal the spirit on the exploration of Malliavin differentiability for McKean-Vlasov SDEs with common noise. So, the extension of Theorem 3.1 to the setting of the locally Lipschitz coefficients is left for the future work.

Below, we move on to complete the proof of Theorem 3.1, which is separated into four steps.

Proof. The proof's main argument is that the Malliavin regularity of the IPS (2.3) carries through to the non-IPS (2.2) in the particle limit N. Thus, we present our proof from the point of view of the Malliavin differentiability of (2.2) and (3.3) and not (1.1) and (3.1); the result is, of course, the same.

Step 0. Basic property and well-posedness of the SDE (3.3) *and* (3.1). Due to $(\mathbf{H}_2)$, there exists $C > 0$ such that for all $A \in \mathbb{R}^{d \otimes m_0}$, $t \in [0, T]$ and $(x, y, \mu) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$,

$$\|g(t, x, \mu)A\|_{\mathrm{HS}} \leq C\|A\|_{\mathrm{HS}} \quad \text{and} \quad \|h(t, x, \mu, y)A\|_{\mathrm{HS}} \leq C\|A\|_{\mathrm{HS}},$$

where $g = \nabla b, \nabla \sigma_i^0, \nabla \sigma_i^1$ and $h(t, x, \mu, y) = \partial_\mu b(t, x, \mu)(y), \partial_\mu \sigma_i^0(t, x, \mu)(y), \partial_\mu \sigma_i^1(t, x, \mu)(y)$. With the previous estimates at hand, (3.2) can be obtained from the Burkholder-Davis-Gundy inequality and the Gronwall inequality in case (3.3) is strongly well-posed.

Below, we merely show that (3.3) is well-posed for the case $s = 0$ since the general case is analogous. To show the strong well-posedness of (3.3) with $s = 0$, we appeal to the Banach fixed point approach (similarly to [13, Theorem 1.1]). Define the set

$$\mathbb{H} = \left\{ f : [0, T] \times \Omega \to \mathbb{R}^{d \otimes m_0} \text{ is progressively measurable s.t. } \int_0^T \mathbb{E}\|f(t)\|_{\mathrm{HS}}^2 \, dt < \infty \right\},$$

which is a complete metric space under the metric: $\rho_\lambda(f, g) := \left(\int_0^T e^{-\lambda t} \mathbb{E}\|f_t - g_t\|_{\mathrm{HS}}^2 dt \right)^{\frac{1}{2}}$, $f, g \in \mathbb{H}$, for each $\lambda > 0$. Now, we define a map $\eta \ni \mathbb{H} \mapsto \Phi(\eta)$ to be the right hand side of (3.3) with $D^0 X^i$ therein being replaced by η. Under $(\mathbf{H}_1)$ and $(\mathbf{H}_2)$, we can infer that $\Phi : \mathbb{H} \to \mathbb{H}$ and Φ is a contractive map when $\lambda > 0$ is chosen large enough. Subsequently, the Banach fixed point theorem enables us to conclude well-posedness.

Step 1. L^2-boundedness of $D_s^0 X_t^{i,N}$. Starting from the IPS (2.3), as a classical SDE, take its D^0 Malliavin derivative: $D_s^0 X_t^{i,N}$. Clearly, $D_s^0 X_s^{i,N} = 0$ for $s > t$, while we obtain from (2.1) (see also [7]) that for $0 \leq s \leq t \leq T < \infty$,

$$dD_s^0 X_t^{i,N} = \left(\nabla b(t, X_t^{i,N}, \bar{\mu}_t^N) D_s^0 X_t^{i,N} + \frac{1}{N} \sum_{k=1}^N \partial_\mu b(t, X_t^{i,N}, \bar{\mu}_t^N)(X_t^{k,N}) D_s^0 X_t^{k,N} \right) dt$$

$$+ \sum_{j=1}^{m_0} \left(\nabla \sigma_j^0(t, X_t^{i,N}, \bar{\mu}_t^N) D_s^0 X_t^{i,N} \right.$$

$$\left. + \frac{1}{N} \sum_{k=1}^N \partial_\mu \sigma_j^0(t, X_t^{i,N}, \bar{\mu}_t^N)(X_t^{k,N}) D_s^0 X_t^{k,N} \right) dW_t^{0,j}$$

$$+ \sum_{j=1}^{m} \left(\nabla \sigma_j^1(t, X_t^{i,N}, \bar{\mu}_t^N) D_s^0 X_t^{i,N} \right.$$

$$\left. + \frac{1}{N} \sum_{k=1}^N \partial_\mu \sigma_j^1(t, X_t^{i,N}, \bar{\mu}_t^N)(X_t^{k,N}) D_s^0 X_t^{k,N} \right) dW_t^{1,i,j}$$

$$=: A_{s,t}^i dt + \sum_{j=1}^{m_0} B_{s,t}^{i,j} dW_t^{0,j} + \sum_{j=1}^{m} C_{s,t}^{i,j} dW_t^{1,i,j},$$

(3.5)

where $D_s^0 X_s^{i,N} = \sigma^0(s, X_s^{i,N}, \bar{\mu}_s^N)$. For any fixed N, $(D^0 X^{i,N})_{s \leq t}$ is well-defined and well posed by the results in [9] and with exploding norms as $N \to \infty$. Next, we show that such norms do not explode. Indeed, applying Itô's formula yields that

$$d\|D_s^0 X_t^{i,N}\|_{\mathrm{HS}}^2 = 2\langle D_s^0 X_t^{i,N}, A_{s,t}^i \rangle_{\mathrm{HS}} dt + \sum_{j=1}^{m_0} \left\|B_{s,t}^{i,j}\right\|_{\mathrm{HS}}^2 dt + \sum_{i=1}^{m} \left\|C_{s,t}^{i,j}\right\|_{\mathrm{HS}}^2 dt + dM_{s,t}^i,$$

where $M_{s,t}^i$ a martingale term.

Initially fixing s, applying the supremum over t and taking said expectation, one implements the Young and Jensen inequalities and makes use of $(\mathbf{H_2})$ to obtain the existence of some $C_1 > 0$ (independent of N, but dependent on d, m and m_0) such that for all $0 \leq s \leq t \leq T$, (with a slight abuse of notation as we write $\|\cdot\|_{\mathcal{C}_t}$ and not $\|\cdot\|_{\mathcal{C}_{s,t}}$)

$$
\begin{aligned}
\mathbb{E}\big[\big\|D_s^0 X_{\cdot}^{i,N}\big\|_{\mathcal{C}_t}^2\big] &\leq \mathbb{E}\big[\|\sigma^0(s, X_s^{i,N}, \bar{\mu}_s^N)\|_{\mathrm{HS}}^2\big] \\
&\quad + C_1 \mathbb{E}\Big[\int_s^t \Big(\|D_s^0 X_t^{i,N}\|_{\mathrm{HS}}^2 + \frac{1}{N}\sum_{k=1}^N \|D_s^0 X_t^{k,N}\|_{\mathrm{HS}}^2\Big)\mathrm{d}r\Big],
\end{aligned} \tag{3.6}
$$

where the $\mathbb{E}\big[\sup_{0\leq t\leq T}|M_N(s,t)|\big]$ was treated using the Burkholder-Davis-Gundy inequality. By invoking $(\mathbf{H_1})$, along with Lemma 2.3, it follows that for some constants $C_2, C_3 > 0$ (independent of N), we have

$$
\begin{aligned}
\mathbb{E}\big[\|\sigma^0(s, X_s^{i,N}, \bar{\mu}_s^N)\|_{\mathrm{HS}}^2\big] &\leq C_2\Big(1 + \mathbb{E}[|X_s^{i,N}|^2] + \frac{1}{N}\sum_{k=1}^N \mathbb{E}[|X_s^{k,N}|^2]\Big) \\
&\leq C_3(1 + \mathbb{E}[|\xi|^2]).
\end{aligned} \tag{3.7}
$$

Averaging over the particle index $i = 1, \ldots, N$, we have

$$
\frac{1}{N}\sum_{i=1}^N \mathbb{E}\big[\big\|D_s^0 X_{\cdot}^{i,N}\big\|_{\mathcal{C}_t}^2\big] \leq C_3(1 + \mathbb{E}[|\xi|^2]) + 2C_1 \int_s^t \Big[\frac{1}{N}\sum_{i=1}^N \mathbb{E}\|D_s^0 X_r^{i,N}\|_{\mathrm{HS}}^2\Big]\mathrm{d}r.
$$

Subsequently, Gronwall's inequality enables us to derive that

$$
\frac{1}{N}\sum_{i=1}^N \mathbb{E}\big[\big\|D_s^0 X_{\cdot}^{i,N}\big\|_{\mathcal{C}_T}^2\big] \leq C_3(1 + \mathbb{E}[|\xi|^2])\mathrm{e}^{2C_1 T}.
$$

Injecting the above estimate back into (3.6) and exploiting (3.7), we obtain

$$
\begin{aligned}
\mathbb{E}\big[\big\|D_s^0 X_{s,\cdot}^{i,N}\big\|_{\mathcal{C}_t}^2\big] &\leq C_3(1 + \mathbb{E}[|\xi|^2]) \\
&\quad + C_1\Big[\int_s^t \Big(\mathbb{E}\|D_s^0 X_r^{i,N}\|_{\mathrm{HS}}^2 + C_3(1 + \mathbb{E}[|\xi|^2])\mathrm{e}^{2C_1 T}\Big)\mathrm{d}r\Big].
\end{aligned}
$$

Once more, applying Gronwall's inequality yields that

$$
\sup_{N\in\mathbb{N}} \max_{1\leq i\leq N} \sup_{0\leq s\leq T} \mathbb{E}\big[\big\|D_s^0 X_{\cdot}^{i,N}\big\|_{\mathcal{C}_T}^2\big] < \infty.
$$

This, besides Lemma 2.3, implies that $X^i \in \mathbb{D}^{1,2}$ and $D_s^0 X_t^i = \lim_{N\to\infty} D_s^0 X_t^{i,N}$ in $L^2(\mathcal{H})$ weakly by applying [11, Lemma 1.2.3 and Proposition 1.5.5]. Two points need to be remarked. Firstly, we still need to identify the SDE that this limiting object satisfies. Secondly, for any $i = 1, \ldots, N$ $\mathbb{P}$-a.s. the processes $X_t^i, X_t^{i,N}$ and $D_s^0 X_t^i, D_s^0 X_t^{i,N}$ are continuous in time t (given s).

Step 2. Convergence of the measure derivative terms. Fix i. We now prove convergence of the measure derivative terms in (3.5) to the corresponding terms in (3.3) (and (3.4)). Recall that μ^i is a Ω^0-random variable only. Define the auxiliary quantities

$$
\Lambda_{s,t}^{b,i} := \frac{1}{N}\sum_{k=1}^N \partial_\mu b(t, X_t^{i,N}, \bar{\mu}_t^N)(X_t^{k,N})D_s^0 X_t^{k,N} - \mathbb{E}^1\big[\partial_\mu b(t, z, \mu_t^i)(X_t^i)D_s^0 X_t^i\big]\big|_{z=X_t^i}
$$

and similarly, comparing (3.4) to (3.5) and accordingly, define $\Lambda_{s,t}^{\sigma_j^0,i}$, $\Lambda_{s,t}^{\sigma_j^1,i}$. Note the quantities in the difference are well defined from the results established in the previous

step. Taking the conditional square expectation,

$$
\sup_{0\le s\le T}\mathbb{E}^1\bigg[\sup_{0\le r\le T}\bigg\|\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^{i,N},\bar\mu_r^N)(X_r^{k,N})D_s^0X_r^{k,N}
$$
$$
-\,\mathbb{E}^1\big[\partial_\mu b(r,z,\mu_r^i)(X_r^i)D_s^0X_r^i\big]\big|_{z=X_r^i}\bigg\|_{\mathrm{HS}}^2\bigg]
$$
$$
\le 4\sup_{0\le s\le T}\mathbb{E}^1\bigg[\sup_{0\le r\le T}\bigg\|\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^{i,N},\bar\mu_r^N)(X_r^{k,N})D_s^0X_r^{k,N}
$$
$$
-\,\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^{i,N},\mu_r^i)(X_r^{k,N})D_s^0X_r^{k,N}\bigg\|_{\mathrm{HS}}^2\bigg]
$$
$$
+\,4\sup_{0\le s\le T}\mathbb{E}^1\bigg[\sup_{0\le r\le T}\bigg\|\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^{i,N},\mu_r^i)(X_r^{k,N})D_s^0X_r^{k,N}
$$
$$
-\,\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^i,\mu_r^i)(X_r^k)D_s^0X_r^{k,N}\bigg\|_{\mathrm{HS}}^2\bigg]
$$
$$
+\,4\sup_{0\le s\le T}\mathbb{E}^1\bigg[\sup_{0\le r\le T}\bigg\|\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^i,\mu_r^i)(X_r^k)D_s^0X_r^{k,N}
$$
$$
-\,\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^i,\mu_r^i)(X_r^k)D_s^0X_r^k\bigg\|_{\mathrm{HS}}^2\bigg]
$$
$$
+\,4\sup_{0\le s\le T}\mathbb{E}^1\bigg[\sup_{0\le r\le T}\bigg\|\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,X_r^i,\mu_r^i)(X_r^k)D_s^0X_r^k
$$
$$
-\,\mathbb{E}^1\big[\partial_\mu b(r,z,\mu_r^i)(X_r^i)D_s^0X_r^i\big]\big|_{z=X_r^i}\bigg\|_{\mathrm{HS}}^2\bigg].
$$

For the first term. Using Lemma 2.3, the joint continuity of $\partial_\mu b$ and Jensen's Inequality we bound $\mathbb{P}^0$-a.s. this term by $4\sup_{0\le s\le T}\mathbb{E}^1\big[\sup_{0\le r\le T}|\frac{1}{N}\sum_{k=1}^{N}D_s^0X_r^k|^2\big]\cdot\varepsilon_N$ for some bounded sequence $(\varepsilon_N)_{N\ge1}\to 0$ as $N\to\infty$. The $\mathbb{P}^0$-a.s. boundedness of $\sup_{0\le s\le T}\mathbb{E}^1\big[\sup_{0\le r\le T}|\frac{1}{N}\sum_{k=1}^{N}D_s^0X_r^{k,N}|^2\big]$, which is justified as in *Step 1*, implies this converges $\mathbb{P}^0$-a.s. to zero in the limit $N\to\infty$.

For the second term: the (conditional) Propagation of Chaos ascertains that $X_r^{k,N}$ converges $\mathbb{P}^0$-a.s. to X_r^k in $L^2(\Omega^1)$ and since, by $(\mathbf{H}_2)$, $\partial_\mu b$ is jointly continuous and uniformly bounded then the convergence results follows immediately from the $L^2(\Omega^1)$-boundedness of $D_s^0X_r^{k,N}$ (uniformly in N) from *Step 1* and conditional dominated convergence.

For the third term. We also have convergence to zero, $\mathbb{P}^0$-a.s. in the limit $N\to\infty$. This is justified by boundedness and uniform continuity of $\partial_\mu b$ and the already established fact (in the proof's *Step 1*) that $D_s^0X_t^i=\lim_{N\to\infty}D_s^0X_t^{i,N}$ in $L^2(\mathcal{C}_T)$-weakly, before once again applying conditional dominated convergence.

For the fourth term. For the particle system $\{(X^k,D^0X^k)\}_k$, one observes by strong symmetry that the particles are exchangeable (and pairwise conditionally independent with uniformly (in N) finite second moments). That is,

$$
\{(X_r^{\pi(1)},D_s^0X_r^{\pi(1)}),\ldots,(X_r^{\pi(N)},D_s^0X_r^{\pi(N)})\}\overset{\mathrm{Law}}{=}\{(X_r^1,D_s^0X_r^1),\ldots,(X_r^N,D_s^0X_r^N)\}
$$

for any permutation π of $\{1,\ldots,N\}$. By the strong Law of Large Numbers, $\mathbb{P}^0\otimes\mathbb{P}^1$-a.s.

$$
\lim_{N\to\infty}\bigg|\Big\{\frac{1}{N}\sum_{k=1}^{N}\partial_\mu b(r,z,\mu_r^i)(X_r^k)D_s^0X_r^k-\mathbb{E}^1\big[\partial_\mu b(r,z,\mu_r^i)(X_r^i)D_s^0X_r^i\big]\Big\}\bigg|_{z=X_r^i}\bigg|^2=0,
$$

where the case $k = i$ is trivial using the boundedness of $\partial_\mu b$, the L^2-bounds of $D^0 X^\cdot$ established earlier and that the weight $1/N$ is vanishing; the continuity in time of the processes extends the result uniformly over $[s, T]$; the result holds with $z = X_r^i$ omitted and for any $z \in \mathbb{R}^d$. Uniform boundedness of $\partial_\mu b$, and the $L^2(\Omega^1)$-uniform bounds on $D_s^0 X_t^i$ and $D_s^0 X_t^{i,N}$, justified in *Step 1*, allow us to apply conditional dominated convergence and conclude this final term converges to zero $\mathbb{P}^0 \otimes \mathbb{P}^1$-a.s. Thus for any fixed $i = 1, \ldots, N$ we have $\mathbb{P}^0$-a.s.

$$\limsup_{N \to \infty} \mathbb{E}^1 \Big[\sup_{0 \le s \le r \le T} |\Lambda_{s,r}^{b,i}|^2 \Big] = 0.$$

Gathering the arguments, an application of the Cauchy-Schwarz inequality then yields the required convergence

$$\int_s^t \frac{1}{N} \sum_{k=1}^N \partial_\mu b(r, X_r^{i,N}, \bar{\mu}_r^N)(X_r^{k,N}) D_s^0 X_r^{k,N} \mathrm{d}r$$

$$\longrightarrow \int_s^t \mathbb{E}^1 \Big[\partial_\mu b(r, z, \mu_r^i)(X_r^i) D_s^0 X_r^i \Big] \Big|_{z = X_r^i} \mathrm{d}r \quad \text{as } N \to \infty \tag{3.8}$$

in $L^2(\Omega^1)$ for any $t \in [s, T]$ conditional on a path ω^0 $\mathbb{P}^0$-a.s. To deal with the corresponding stochastic integral terms involving the measure derivatives of σ_j^0, $j = 1, \ldots, m_0$ and σ_j^1, $j = 1, \ldots, m$, we apply the above method, replacing the Cauchy-Schwarz inequality with an additional Itô Isometry argument, localising and arguing over quadratic variation.

Step 3. Identifying the limiting equation of (3.5). With the results of *Step 2* at hand, the remaining arguments are classical and thus presented in a streamlined fashion. Fix i. Set for $s \in [0, T]$

$$\delta \sigma_s^{0,i} := \sigma^0\big(s, X_s^{i,N}, \bar{\mu}_s^N\big) - \sigma^0\big(s, X_s^i, \mu_s^i\big),$$

and compare the McKean-Vlasov SDEs (3.5) and (3.3). We prove next that (3.3) is the L^2-limit equation of (3.5) (for any fixed i as $N \to \infty$):

$$D_s^0 X_t^{i,N} - D_s^0 X_t^i = \delta \sigma_s^{0,i} + \int_s^t \widehat{A}_{s,r}^i \, \mathrm{d}r + \int_s^t \Lambda_{s,r}^{b,i} \, \mathrm{d}r$$
$$+ \sum_{j=1}^{m_0} \int_s^t \big(\widehat{B}_{s,r}^{i,j} + \Lambda_{s,r}^{\sigma_j^0,i} \big) \, \mathrm{d}W_r^{0,j} + \sum_{j=1}^m \int_s^t \big(\widehat{C}_{s,r}^{i,j} + \Lambda_{s,r}^{\sigma_j^1,i} \big) \, \mathrm{d}W_r^{1,i,j},$$

where the Λ processes were defined in Step 2 and we define additionally

$$\widehat{A}_{s,r}^i := \nabla b(r, X_r^{i,N}, \bar{\mu}_r^N) D_s^0 X_r^{i,N} - \nabla b(r, X_r^i, \mu_r^i) D_s^0 X_r^i,$$
$$\widehat{B}_{s,r}^{i,j} := \nabla \sigma_j^0(r, X_r^{i,N}, \bar{\mu}_r^N) D_s^0 X_r^{i,N} - \nabla \sigma_j^0(r, X_r^i, \mu_r^i) D_s^0 X_r^i, \quad j = 1, \ldots, m_0$$
$$\widehat{C}_{s,r}^{i,j} := \nabla \sigma_j^1(r, X_r^{i,N}, \bar{\mu}_r^N) D_s^0 X_r^{i,N} - \nabla \sigma_j^1(r, X_r^i, \mu_r^i) D_s^0 X_r^i, \quad j = 1, \ldots, m.$$

Taking the square expectation, and using the Itô isometry, we obtain that

$$\mathbb{E}\Big[\|D_s^0 X_t^{i,N} - D_s^0 X_t^i\|_{\mathrm{HS}}^2 \Big] - \Psi_N = \mathbb{E}\left[\left\| \int_s^t \widehat{A}_{s,r}^i \mathrm{d}r \right\|_{\mathrm{HS}}^2 \right] + \mathbb{E}\left[\sum_{j=1}^{m_0} \int_s^t \|\widehat{B}_{s,r}^{i,j}\|_{\mathrm{HS}}^2 \mathrm{d}r \right]$$
$$+ \mathbb{E}\left[\sum_{j=1}^m \int_s^t \|\widehat{C}_{s,r}^{i,j}\|_{\mathrm{HS}}^2 \mathrm{d}r \right], \tag{3.9}$$

where

$$\Psi_N := \mathbb{E}\left[\|\delta\sigma_s^{0,i}\|_{\mathrm{HS}}^2\right] + 2\mathbb{E}\left[\int_s^t \langle\delta\sigma_s^{0,i}, \widehat{A}_{s,r}^i\rangle_{\mathrm{HS}}\mathrm{d}r\right] + 2\mathbb{E}\left[\int_s^t \langle\delta\sigma_s^{0,i}, \Lambda_{s,r}^{b,i}\rangle_{\mathrm{HS}}\mathrm{d}r\right]$$

$$+ 2\mathbb{E}\left[\left\langle\int_s^t \widehat{A}_{s,r}^i\mathrm{d}r, \int_s^t \Lambda_{s,r}^{b,i}\mathrm{d}r\right\rangle_{\mathrm{HS}}\right] + 2\mathbb{E}\left[\sum_{j=1}^{m_0}\int_s^t \langle\widehat{B}_{s,r}^{i,j}, \Lambda_{s,r}^{\sigma_j^0,i}\rangle_{\mathrm{HS}}\mathrm{d}r\right]$$

$$+ 2\mathbb{E}\left[\sum_{j=1}^{m}\int_s^t \langle\widehat{C}_{s,r}^{i,j}, \Lambda_{s,r}^{\sigma_j^1,i}\rangle_{\mathrm{HS}}\mathrm{d}r\right] + \mathbb{E}\left[\left\|\int_s^t \Lambda_{s,r}^{b,i}\mathrm{d}r\right\|_{\mathrm{HS}}^2\right]$$

$$+ \mathbb{E}\left[\sum_{j=1}^{m_0}\int_s^t \|\Lambda_{s,r}^{\sigma_j^0,i}\|_{\mathrm{HS}}^2\mathrm{d}r\right] + \mathbb{E}\left[\sum_{j=1}^{m}\int_s^t \|\Lambda_{s,r}^{\sigma_j^1,i}\|_{\mathrm{HS}}^2\mathrm{d}r\right],$$

noting that since W^0 and $W^{1,\cdot}$ are independent no cross variation terms appear.

The joint continuity property of σ^0 and the (conditional) Propagation of Chaos implies $L^2(\mathcal{C}_T)$ and $\mathbb{P}^0$-a.s. convergence of $\delta\sigma_\cdot^{0,i}$ and $\mathbb{E}^1[\|\delta\sigma_\cdot^{0,i}\|_{\mathrm{HS}}^2] \to 0$ respectively as $N \to \infty$. The last three terms of Ψ_N vanish using the results of *Step 2* (possibly along a subsequence if needed). The remaining five terms also vanish in the N limit using classic arguments: one mimics as needed the arguments used in *Step 2*, using the tower property $\mathbb{E}[\cdot] = \mathbb{E}[\mathbb{E}^1[\cdot]]$ to draw on the conditional Propagation of chaos (2.4), using the joint continuity and uniform boundedness of the derivatives $\nabla b, \nabla\sigma^0$ and $\nabla\sigma^1$ plus the L^2-convergence (possibly along a subsequence) and L^2-bounds of $D^0 X^{i,N}, D^0 X^i$ coupled with repeated use of the Cauchy-Schwarz and Jensen inequalities and the convergence of the Λ terms established in *Step 2*. In conclusion, the LHS of (3.9) converges to zero (possibly along a subsequence if needed), thus so must its RHS (using additionally the continuity in time of the processes involved) and it follows that (3.3) is precisely the L^2-limiting equation of the IPS (3.5). □

The Malliavin differentiability of (1.1) with respect to the idiosyncratic W^1 is classical.

Theorem 3.4. *Assume* $(\mathbf{H}_1)$ *and let* b, σ^0, σ^1 *be continuously differentiable in space with uniformly bounded spatial derivatives. Then, the Malliavin derivative of (1.1) with respect to the idiosyncratic Brownian motion* $(W_t^1)_{t\geq 0}$ *satisfies the linear SDE: for any* $t \geq s$,

$$D_s^1 X_t = \sigma^1(s, X_s, \mu_s) + \int_s^t \nabla b(r, X_r, \mu_r)D_s^1 X_r\mathrm{d}r + \int_s^t \nabla\sigma^0(r, X_r, \mu_r)D_s^1 X_r\mathrm{d}W_r^0$$

$$+ \int_s^t \nabla\sigma^1(r, X_r, \mu_r)D_s^1 X_r\mathrm{d}W_r^1,$$

and $D_s^1 X_t = \mathbf{0}_{d\times m}$ *for any* $s > t$. *Moreover, there is a constant* $K = K(T, L) > 0$ *such that*

$$\sup_{0\leq s\leq T}\mathbb{E}\left[\|D_s^1 X_\cdot\|_{\mathcal{C}_T}^2\right] \leq K(1 + \mathbb{E}|\xi|^2).$$

We point that using further approximation arguments, this result holds only under $(\mathbf{H}_1)$ (understanding $\nabla b, \nabla\sigma^0, \nabla\sigma^1$ in a generalised sense, see [11, Prop. 1.2.4]).

Proof. The Malliavin derivative with respect to the idiosyncratic noise $(W_t^1)_{t\geq 0}$ of $(X_t)_{t\geq 0}$ (1.1) will ignore the conditional dependence on $\omega^0 \in \Omega^0$, in the sense that the Cameron-Martin perturbations are solely against the Brownian motion that is not being conditioned upon. Thus, the result follows promptly from the same arguments as in [8, Theorem

3.4] and [6, Proposition 3.1]. Methodologically: via a Picard iteration argument build a convenient sequence of *standard SDEs* that converges to the McKean-Vlasov SDE, apply a suitable Malliavin differentiability result for regular SDEs [9] to the sequence and then pass to limit using the closedness of the Malliavin operator [11, Lemma 1.2.3]. □

4 Conditional integration by parts formula

For a random variable η, we write $[\eta]$ as its law. For convenience, denote the solution to (1.1) as $(X_t^{[\xi]})_{t \geq 0}$ to highlight the distribution of ξ, and, for each $t \geq 0$, write $\mu_t^\xi := \mathrm{Law}(X_t^{[\xi]}(\omega^0, \cdot))$. Consider the decoupled SDE associated with (1.1):

$$\mathrm{d}X_t^{x,[\xi]} = b\big(t, X_t^{x,[\xi]}, \mu_t^\xi\big)\mathrm{d}t + \sigma^0\big(t, X_t^{x,[\xi]}, \mu_t^\xi\big)\mathrm{d}W_t^0 + \sigma^1\big(t, X_t^{x,[\xi]}, \mu_t^\xi\big)\mathrm{d}W_t^1, \quad t > 0, \quad (4.1)$$

with the initial value $X_0^{x,[\xi]} = x \in \mathbb{R}^d$. The SDE above is a standard one with random coefficients and is related but decoupled from (1.1) since the measure term is exogenous (and also independent of x). Its well-posedness follows from [12, Theorem 3.1.1], observing the integrability of the maps $t \mapsto b(t, 0, \mu_t^\xi)$ and $t \mapsto \sigma^l(t, 0, \mu_t^\xi)$, $l = 0, 1$. Results regarding the first variation process with respect to the initial condition follow the almost identical formulation as the unconditional McKean-Vlasov SDE, as presented in [6]. We state these for completeness.

Lemma 4.1. *Let* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$ *hold. Then, the map* $\mathbb{R}^d \ni x \mapsto X_\cdot^{x,[\xi]} \in L^2(\mathcal{C}_T, \mathbb{P})$ *is* $\mathbb{P}$-*a.s. continuously differentiable and* $w_t^x := (\nabla X_t^{\cdot,[\xi]})(x)$ *satisfies the affine SDE: for any* $0 \leq s \leq t \leq T$,

$$w_t^x = I_{d \times d} + \int_s^t \nabla b(r, \cdot, \mu_r^\xi)(X_r^{x,[\xi]})w_r^x \mathrm{d}r + \int_s^t \nabla \sigma^0(r, \cdot, \mu_r^\xi)(X_r^{x,[\xi]})w_r^x \mathrm{d}W_r^0 \qquad (4.2)$$
$$+ \int_s^t \nabla \sigma^1(r, \cdot, \mu_r^\xi)(X_r^{x,[\xi]})w_r^x \mathrm{d}W_r^1.$$

Proof. Given $X^{x,[\xi]}$ and μ^ξ, and knowing $\nabla b, \nabla \sigma^0, \nabla \sigma^1$ are uniformly bounded, then the linear SDE (4.2) is well-posed; see e.g. [12, 9]. For the differentiability statement, recall that the measure component in (4.1) is independent of the spatial variable x and is exogenous given with good integrability properties (Theorem 2.2). The differentiability of $x \mapsto X^{x,[\xi]}$ follows as a particular case of the results from [9]. Refer to [5, Theorem 5.29], which discusses differentiability in a decoupled McKean-Vlasov FBSDE case. □

Comparing the SDE in Theorem 3.4 with the SDE (4.2), we note that these are the same type affine SDEs but with different dimensions and initial conditions. Working under the (uniform ellipticity) assumption that there exists some $\delta > 0$ such that for all $\zeta \in \mathbb{R}^d$ and $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, $\zeta^\top((\sigma^1(\sigma^1)^\top)(t, x, \mu))\zeta \geq \delta|\zeta|^2$, one establishes, exactly as in [6, Proposition 4.1], the following integration by parts formula. Let $f \in C_b^\infty(\mathbb{R}^d)$ and $\Phi : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{D}^{1,2}$ be a map whose measure derivative and spatial derivative maps are L^1-integrable and of linear growth. Then it holds that

$$\mathbb{E}^1\Big[\nabla f(X_t^{\cdot,[\xi]})(x)\Phi(t, x, [\xi])\Big] = \frac{1}{t}\mathbb{E}^1\Big[f(X_t^{x,[\xi]})\delta^1\big(r \mapsto g(r)^\top \Phi(t, x, [\xi])\big)\Big],$$

where $g(r) := ((\sigma^1)^\top(\sigma^1(\sigma^1)^\top)^{-1})(r, X_r^{x,[\xi]}, \mu_r^\xi)(\nabla X_r^{\cdot,[\xi]})(x)$ and δ^1 is the Skorokhod operator with respect to the idiosyncratic Brownian motion.

More of interest is an *integration by parts formula for Equation* (1.1) *with respect to the measure variable*. We first need the following lemma.

Lemma 4.2. *Under* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$*, there exists a modification of* $X_t^{[\xi]}$ *such that the map* $L^2(\Omega) \ni \xi \mapsto X^\xi \in L^2(\mathcal{C}_T, \mathbb{P})$ *is* $\mathbb{P}$*-a.s. Fréchet differentiable in* $L^2(\Omega)$*. Hence* $\partial_\mu X_t^{[\xi]}$ *exists and for* $v \in \mathbb{R}^d$*,* $\Gamma_t^\xi(v) := \partial_\mu X_t^{[\xi]}(v)$ *satisfies the dynamics:*

$$\Gamma_t^\xi(v) = \boldsymbol{\Psi}_t(X) + \int_0^t \left(\nabla b(r, \cdot, \mu_r^\xi)(X_r^{[\xi]})\Gamma_r^\xi(v) + \mathbb{E}^1 \left[\partial_\mu b(r, \eta)(X_r^{[\xi]})\Gamma_r^\xi(v) \right] \Big|_{\eta = (X_r^{[\xi]}, \mu_r^\xi)} \right) dr$$

$$+ \int_0^t \left(\nabla \sigma^0(r, \cdot, \mu_r^\xi)(X_r^{[\xi]})\Gamma_r^\xi(v) + \mathbb{E}^1 \left[\partial_\mu \sigma^0(r, \eta)(X_r^{[\xi]})\Gamma_r^\xi(v) \right] \Big|_{\eta = (X_r^{[\xi]}, \mu_r^\xi)} \right) dW_r^0$$

$$+ \int_0^t \left(\nabla \sigma^1(r, \cdot, \mu_r^\xi)(X_r^{[\xi]})\Gamma_r^\xi(v) + \mathbb{E}^1 \left[\partial_\mu \sigma^1(r, \eta)(X_r^{[\xi]})\Gamma_r^\xi(v) \right] \Big|_{\eta = (X_r^{[\xi]}, \mu_r^\xi)} \right) dW_r^1, \qquad (4.3)$$

where setting $\Xi_r := (r, X_r^{[\xi]}, \mu_r^\xi)$ *for* $r \geq 0$ *we have*

$$\boldsymbol{\Psi}_t(X) := \int_0^t \partial_\mu b(\Xi_r)(X_r^{v,[\xi]})\nabla_x X_r^{v,[\xi]} dr + \int_0^t \partial_\mu \sigma^0(\Xi_r)(X_r^{v,[\xi]})\nabla_x X_r^{v,[\xi]} dW_r^0$$

$$+ \int_0^t \partial_\mu \sigma^1(\Xi_r)(X_r^{v,[\xi]})\nabla_x X_r^{v,[\xi]} dW_r^1.$$

Proof. See [15, Lemma 3.5] and replace their Λ_t^μ by $\tilde{\Lambda}_t^\xi := [\xi] \circ (\mathbb{E}[X_t^{x,[\xi]} | \mathcal{F}_t^0])^{-1}$, yielding their *"image-dependent SDE"*. See also [5, Lemma 5.27]. $\qquad \square$

We now state the below integration by parts formula that is not covered by [15].

Lemma 4.3. *Assume that* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$ *hold. Suppose further that there exists some* $\delta > 0$ *such that* $\zeta^\top ((\sigma^0(\sigma^0)^\top)(t, x, \mu))\zeta \geq \delta |\zeta|^2$ *for all* $\zeta \in \mathbb{R}^d$ *and* $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$*. Then, there exists an* $h \in \mathcal{H}$ *(given explicitly in the proof) such that the following integration by parts formula holds: for* $v \in \mathbb{R}^d$ *and* $f \in C_b^2(\mathbb{R}^d)$*,*

$$\mathbb{E}^1 \left[\partial_\mu (f(X_t^{[\xi]}))(v) \right] = \frac{1}{t} \mathbb{E}^1 \left[f(X_t^{[\xi]})\delta^0 (r \mapsto h(r, v)^\top) \right].$$

where δ^0 *denotes the Skorokhod operator with respect to the common Brownian motion; the* L^2*-adjoint to the Malliavin divergence.*

Proof. Compare (3.3) and (4.3): these equations are driven by the same dynamics but with different dimensions and initial conditions. Arguing as in [6], we have, with $\boldsymbol{\Psi}_t(X)$ as above, the $\mathbb{P}$-a.s. representation[1]

$$\partial_\mu X_t^{[\xi]}(v) = D_r^0 X_t^{[\xi]} h(r, v),$$

where

$$h(r, v) := \left((\sigma^0)^\top (\sigma^0(\sigma^0)^\top)^{-1} \right)(r, X_r^{[\xi]}, \mu_r^\xi) \left\{ \boldsymbol{\Psi}_t(X) - \boldsymbol{\Psi}_r(X) + \partial_\mu X_r^{[\xi]}(v) \right\}.$$

---

[1] From (4.3), it is easy to see that $\Gamma_\cdot^\xi(v)$ can be written with the help of $\Psi_\cdot(X)$

$$\Gamma_t^\xi(v) = \Gamma_r^\xi(v) + \Psi_t(X) - \Psi_r(X) + \int_r^t \cdots \text{``extra terms''} \cdots.$$

For the relationship between $\partial_\mu X_t^\xi(v)$ and $\Gamma_t^\xi(v)$, we formally multiply $((\sigma^0(\sigma^0)^\top)^{-1})(r)(\Gamma_r^\xi(v) + \Psi_t(X) - \Psi_r(X))$ on both sides of (3.3) but as the term is not adapted it cannot be moved inside the stochastic integrals. Thus, methodologically, the relations between $\partial_\mu X_t^\xi(v)$ and $\Gamma_t^\xi(v)$ does not follow via strong uniqueness.

Uniform boundedness of $\partial_\mu\big(f(X_t^{[\xi]})\big)$ follows from [5, Theo. 5.29]. Thus, Malliavin integration by parts implies that for $v \in \mathbb{R}^d$

$$\mathbb{E}^1\Big[\partial_\mu\big(f(X_t^{[\xi]}))(v)\Big] = \frac{1}{t}\mathbb{E}^1\Big[\int_0^t D_r^0 X_t^{[\xi]} h(r)(\nabla f)(X_t^{[\xi]})\mathrm{d}r\Big]$$

$$= \frac{1}{t}\mathbb{E}^1\Big[\int_0^t h(r)^\top D_r^0 f(X_t^{[\xi]})\mathrm{d}r\Big] = \frac{1}{t}\mathbb{E}^1\Big[f(X_t^{[\xi]})\delta^0\big(r \mapsto h(r,v)^\top\big)\Big]. \ \square$$

References

[1] J. Bao, P. Ren and F.-Y. Wang. Bismut formula for Lions derivative of distribution-path dependent SDEs. J. Differential Equations, 282:285–329, 2021 MR4219322

[2] S. Biswas, C. Kumar, Neelima, G. dos Reis, and C. Reisinger. An explicit Milstein-type scheme for interacting particle systems and McKean-Vlasov SDEs with common noise and non-differentiable drift coefficients. *Ann. Appl. Probab.*, 34(2):2326–2363, 2024. MR4728171

[3] F. Bugini, M. Coghi, and T. Nilssen. Malliavin calculus for rough stochastic differential equations. *arXiv preprint arXiv:2402.12056*, 2024.

[4] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer-Verlag, 2018. MR3752669

[5] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II*, volume 84 of *Probability Theory and Stochastic Modelling*. Springer-Verlag, 2018. MR3753660

[6] D. Crisan and E. McMurray. Smoothing properties of McKean-Vlasov SDEs. *Probab. Theory Related Fields*, 171(1-2):97–148, 2018. MR3800831

[7] G. dos Reis and V. Platonov. Itô-Wentzell-Lions Formula for Measure Dependent Random Fields under Full and Conditional Measure Flows. *Potential Anal.*, 59(3):1313–1344, 2023. MR4647952

[8] G. dos Reis and Z. Wilde. Malliavin differentiability of McKean-Vlasov SDEs with locally Lipschitz coefficients. *arXiv preprint arXiv:2310.13400*, 2023.

[9] P. Imkeller, G. dos Reis, and W. Salkeld. Differentiability of SDEs with drifts of super-linear growth. *Electron. J. Probab.*, 24:Paper No. 3, 43, 2019. MR3916323

[10] D. Lacker, M Shkolnikov and J. Zhang. Superposition and mimicking theorems for conditional McKean-Vlasov equations, *J. Eur. Math. Soc. (JEMS)*, 25 (8):3229–3288, 2023. MR4612111

[11] D. Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006. MR1344217

[12] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. MR2329435

[13] P. Ren. Extrinsic derivative formula for distribution dependent SDEs. *Bernoulli*, 31 (3), 2508–2524, 2025. MR4890839

[14] P. Ren and F.-Y. Wang. Bismut formula for Lions derivative of distribution dependent SDEs and applications. *J. Differential Equations*, 267(8), 4745–4777, 2019 MR3983053

[15] F.-Y. Wang. Image-dependent conditional McKean-Vlasov SDEs for measure-valued diffusion processes. *J. Evol. Equ.*, 21(2):2009–2045, 2021. MR4278420

[16] F.-Y. Wang. Harnack inequalities for stochastic partial differential equations. Springer, New York, 2013. MR3099948