Malliavin differentiability of McKean-Vlasov SDEs with locally Lipschitz coefficients

Gonçalo dos Reis
University of Edinburgh, School of Mathematics, Edinburgh, UK
and
Center for Mathematics and Applications (NOVA Math),
NOVA School of Science and Technology (NOVA FCT), PT
e-mail: `G.dosReis@ed.ac.uk`

Zac Wilde
University of Edinburgh, School of Mathematics, Edinburgh, UK
e-mail: `Z.J.Wilde@sms.ed.ac.uk`

Resumo: Nesta pequena nota estabelecemos a diferenciabilidade no sentido do cálculo de Malliavin das soluções das Equações Diferenciais Estocásticas de McKean-Vlasov (MV-SDEs), onde se assume que o coeficiente de tendência ('drift') é localmente Lipschitz e satisfaz uma condição Lipschitz 'lateral' enquanto o coeficiente de difusão satisfaz a condição Lipschitz usual.

Como contribuição secundária, é investigado como é que a diferenciabilidade no sentido do cálculo de Malliavin se transfere através do limite de sistemas de partículas que converge para a solução da MV-SDE original. Estabelecer este resultado requer que os coeficientes da equação sejam diferenciáveis na componente espaço e na componente da medida de probabilidade, e logo o resultado apresentado não é tão geral quanto o resultado da primeira parte do artigo (onde não é necessária a diferenciabilidade em medida, apenas a propriedade Lipschitz na métrica de Wasserstein). No entanto, a metodologia é de interesse independente por ser esta a primeira vez a aparecer na literatura (ao melhor do nosso conhecimento). A apresentação da secção é de carácter didático e a secção conclui com uma discussão alargada sobre técnicas de molificação para a derivada de Lions (i.e., derivadas na variável da medida de probabilidade).

Abstract: In this short note, we establish Malliavin differentiability of McKean-Vlasov Stochastic Differential Equations (MV-SDEs) with drifts satisfying both a locally Lipschitz and a one-sided Lipschitz assumption, and where the diffusion coefficient is assumed to be uniformly Lipschitz in its variables.

As a secondary contribution, we investigate how Malliavin differentiability transfers across the interacting particle system associated with the McKean-Vlasov equation to its limiting equation. This final result

requires both spatial and measure differentiability of the coefficients and doubles as a standalone result of independent interest since the study of Malliavin derivatives of weakly interacting particle systems seems novel to the literature. The presentation is didactic and finishes with a discussion on mollification techniques for the Lions derivative.

Keywords: McKean-Vlasov SDEs, Malliavin differentiability, superlinear growth, interacting particle systems

Contents

1 Introduction

The main object of our study are McKean-Vlasov Stochastic Differential Equations (MV-SDE), also known as mean-field equations or distribution-dependent SDEs. They differ from standard SDEs by means of the presence of the law of the solution process in the coefficients. Namely

$$dZ_t = b(t, Z_t, \mu_t)dt + \sigma(t, Z_t, \mu_t)dW_t, \quad X_0 = \xi,$$

for some measurable coefficients, where $\mu_t = \text{Law}(Z_t)$ denotes the law of process Z at time t (with W a Brownian motion; ξ a random initial condition). Similar to standard SDEs, MV-SDEs are shown to be well-posed under a variety of frameworks, for instance, under locally Lipschitz and super-linear growth conditions alongside random coefficients, see e.g. [AdRR$^+$22] or [dRST19]. In this setting there are also many studies on their numerical approximation e.g. [dRES22, CdR22, CdR24], ergodicity [CdS25] and large deviations [dRST19, AdRR$^+$22].

Many mean-field models exhibit drift dynamics that include superlinear growth and non-global Lipschitz growth, for example, mean-field models for neuronal activity (e.g. stochastic mean-field FitzHugh-Nagumo models or the network of Hodgkin-Huxley neurons) [BFFT12], [BCC11], [BFT15] appearing in biology or the physics of modelling batteries [DGG$^+$11],[DFG$^+$16]. Quoting [GPV20], systems of weakly-interacting particles and their limiting processes, so-called McKean-Vlasov or mean-field equations appear in a wide variety of applications, ranging from plasma physics and galactic dynamics to mathematical biology, the social sciences, active media, dynamical density functional theory (DDFT) and machine learning. They can also be used in models for co-operative behavior, opinion formation, risk management, as well as in algorithms for global optimization.

Our 1st contribution: Malliavin differentiability of MV-SDEs under locally Lipschitz conditions. We extend Malliavin variational results to McKean-Vlasov SDEs with locally Lipschitz drifts satisfying a so-called one-sided Lipschitz condition. The result is new to the best of our knowledge. Malliavin differentiability of MV-SDEs has been addressed in [CM18, Proposition 3.1] and [RW19], and in both cases, their assumptions revolve around the differentiable Lipschitz case. Our proof methodology is inspired by that of [CM18] – both there and here, the result is established by appealing to the celebrated [Nua06, Lemma 1.2.3].

Our 2nd contribution: transfer of Malliavin differentiability across the particle system limit. Another large aspect of McKean-Vlasov SDE theory, is the study of the large weakly-interacting particle systems and their particle limit that recovers the MV-SDE in the limit. This latter limit result is called Propagation of Chaos [Szn91] (also [BCC11],[GPV20], [CCD22], [AdRR$^+$22]). In the second part of this note, we study the Malliavin differentiability of the interacting particle system and how the Malliavin regularity transfers across the particle limit to the limiting equation. To

the best of our knowledge, this particular proof methodology is new to the literature.

From a methodological viewpoint, our point of attack is the *projections over empirical measures* approach [CCD22, CD18a, dRP23]. This approach allows us to use the best available Malliavin differentiability results for standard (multidimensional) SDEs [Nua06, IdRS19], and then carry them to the MV-SDE setting via the particle limit using Propagation of Chaos and [Nua06, Lemma 1.2.3]. Our variational results are limited only by the SDE results we cite. If better results are found, one only needs to replace the reference in the appropriate place. Lastly, in relation to our 1st contribution, this 2nd contribution is established under a full global Lipschitz and differentiability (space and measure) assumption on the coefficients.

Organization of the paper. In section 2, we set notation and review a few concepts necessary for the main constructions. In section 3, we prove Malliavin differentiability of MV-SDEs under superlinear drift growth assumptions and in section 4 we prove Malliavin differentiability of MV-SDEs under the weaker global Lipschitz assumptions via the convergence of interacting particle systems, providing a lengthy remark about mollification in Wasserstein spaces.

2   Notation and preliminary results

2.1   Notation and Spaces

For collections of vectors, let the upper indices denote the distinct vectors, whereas the lower index is a vector component, i.e. x_j^l denote the j-th component of l-th vector. Let $\mathbf{x} = (x^1, \cdots, x^N)$ denote a vector in $\mathbb{R}^{dN}$ where $x^i := (x_1^i, \cdots, x_d^i)$ for $i = 1, \cdots, N$. For matrices M and N of agreeing dimensions, define the inner product $M : N = \text{Trace}(M^T N)$ and the norm induced by this inner product (the Hilbert-Schmidt norm) as $|M| = \sqrt{\text{Trace}(M^T M)}$.

For $g : \mathbb{R}^m \to \mathbb{R}^n$, writing $g(x) = (g^1(x), \cdots, g^n(x))$ and define $\nabla_x g$ as the Jacobian matrix $(g_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ where $g_{ij} = \partial_{x^i} g^j$.

Take $T \in [0, \infty)$ and let $(\Omega, \mathbb{F}, \mathcal{F}, \mathbb{P})$ be a filtered probability space carrying a m-dimensional Brownian Motion on the interval $[0, T]$ and $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$. The filtration is the one generated by the Brownian motion and augmented by the $\mathbb{P}$-null sets, and with an additionally sufficiently rich

sub σ-algebra $\mathcal{F}_0$ independent of W. We denote by $\mathbb{E}[\cdot] = \mathbb{E}^{\mathbb{P}}[\cdot]$ the usual expectation operator with respect to $\mathbb{P}$.

The space of probability measures on $\mathbb{R}^d$ with finite second moment, $\mathcal{P}_2(\mathbb{R}^d)$ is Polish under the 2-Wasserstein distance

$$W_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \Big(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \Big)^{\frac{1}{2}}, \quad \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d),$$

where $\Pi(\mu, \nu)$ is the set of couplings for μ and ν such that $\pi \in \Pi(\mu, \nu)$ is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\cdot \times \mathbb{R}^d) = \mu$ and $\pi(\mathbb{R}^d \times \cdot) = \nu$. Let $\text{Supp}(\mu)$ denote the support of $\mu \in \mathcal{P}(\mathbb{R}^d)$.

Let $p \in [2, \infty)$. We introduce the following spaces.

- Let $\mathcal{X}$ be a metric space. We denote by $C(\mathcal{X})$ as the space of continuous functions $f : \mathcal{X} \to \mathbb{R}$ endowed with the uniform norm and $C_b(\mathcal{X})$ its subspace of bounded functions endowed with the sup norm $\|f\|_\infty = \sup_{x \in \mathcal{X}} |f(x)| < \infty$; For $k \in \mathbb{N}$ denote $\mathcal{C}^k(\mathbb{R}^d)$ the space of k-times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^d$, equipped with a collection of seminorms $\{\|g\|_{\mathcal{C}^p(K)} := \sup_{x \in K}(|g(x)| + \sum_{j=1}^k |\partial_x^j g(x)|), g \in \mathcal{C}^k(\mathbb{R}^d)\}$, indexed by the compact subsets $K \subset \mathbb{R}^d$.

- $L^p(\Omega) := L^p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^d)$, $t \in [0, T]$ is the space of $\mathbb{R}^d$-valued $\mathcal{F}_t$-measurable random variables $X : \Omega \to \mathbb{R}^d$ with norm $\|X\|_{L^p(\Omega)} = \mathbb{E}[|X|^p]^{1/p} < \infty$.

- $\mathcal{S}^p([0, T]) := \mathcal{S}^p([0, T], \mathbb{P}; \mathbb{R}^d)$ is the space of $\mathbb{R}^d$-valued measurable $\mathbb{F}$-adapted processes $(Y_t)_{t \in [0, T]}$ satisfying $\|Y\|_{\mathcal{S}^p([0, T])} = \mathbb{E}[\sup_{t \in [0, T]} |Y(t)|^p]^{1/p} < \infty$.

2.2 Malliavin Calculus

Let $\mathcal{H}$ be a Hilbert space and $W : \mathcal{H} \to L^2(\Omega)$ a Gaussian random variable. The space $W(\mathcal{H})$ endowed with an inner product $\langle W(h_1), W(h_2) \rangle = \mathbb{E}[W(h_1) W(h_2)]$ is a Gaussian Hilbert space. Let $C_p^\infty(\mathbb{R}^n; \mathbb{R})$ be the space of infinitely differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ which have partial derivatives of all orders, each with polynomial growth. Let $\mathbb{S}$ be the collection of random variables $F : \Omega \to \mathbb{R}$ such that for $n \in \mathbb{N}$, $f \in C_p^\infty(\mathbb{R}^n; \mathbb{R})$ and $h_i \in \mathcal{H}$ can be written as $F = f(W(h_1), \ldots, W(h_n))$. Then we define the derivative of F to be the $\mathcal{H}$-valued random variable

$$DF = \sum_{i=1}^n \partial_{x_i} f(W(h_1), \ldots, W(h_n)) h_i.$$

The Malliavin derivative from $L^p(\Omega)$ into $L^p(\Omega, \mathcal{H})$ is closable and the domain of the operator is defined to be $\mathbb{D}^{1,p}$, defined to be the closure of the of the set $\mathbb{S}$ with respect to the norm

$$\|F\|_{1,p} = \left[\mathbb{E}[|F|^p] + \mathbb{E}[\|DF\|_{\mathcal{H}}^p]\right]^{\frac{1}{p}}.$$

We also define the directional Malliavin derivative $D^h F = \langle DF, h\rangle_{\mathcal{H}}$ for any choice of $h \in \mathcal{H}$. For more details, see [Nua06].

3   Malliavin differentiability under local Lipschitz assumptions

3.1   McKean-Vlasov Equations with locally Lipschitz coefficients

In this manuscript, we work with so-called McKean-Vlasov SDEs described by the following dynamics for $0 \leq t \leq T < \infty$,

$$dZ_t = b(t, Z_t, \mu_t)dt + \sum_{l=1}^m \sigma^l(t, Z_t, \mu_t)dW_t^l, \quad Z_0 = \xi, \tag{1}$$

where μ_t denotes the law of the process Z at time t, i.e. $\mu_t = \mathbb{P} \circ Z_t^{-1}$ and W^l, $l = 1, \ldots, m$ are 1-dimensional independent Brownian motions. We write $W = (W^1, \ldots, W^m)$ as the corresponding $m-$dimensional Brownian motion. In this chapter, we work under the locally Lipschitz case as the below assumption describes.

3.1.1   Assumptions

Assumption 3.1. *Let $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and for $l = 1, \ldots, m$, $\sigma^l : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$. Then there $\exists L > 0$ such that:*

1. *For some $p \geq 2$, ξ is $\mathcal{F}_0$-measurable and $\xi \in L^p(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$.*

2. *σ^l is continuous in time and Lipschitz in space-measure $\forall t \in [0, T]$, $\forall x, x' \in \mathbb{R}^d$ and $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ we have*

$$|\sigma^l(t, x, \mu) - \sigma^l(t, x', \mu')| \leq L\Big(|x - x'| + W_2(\mu, \mu')\Big).$$

3. b is continuous in time and satisfies the one-sided Lipschitz condition *in space and is Lipschitz in measure:* $\forall t \in [0, T]$, $\forall x, x' \in \mathbb{R}^d$ *and* $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ *we have that*

$$\langle x - x', b(t, x, \mu) - b(t, \omega, x', \mu) \rangle_{\mathbb{R}^d} \le L|x - x'|^2,$$
$$|b(t, x, \mu) - b(t, x, \mu')| \le L W_2(\mu, \mu').$$

4. b is Locally Lipschitz: $\forall t \in [0, T]$, $\forall \mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\forall x, x' \in \mathbb{R}^d$ *such that* $|x|, |x'| < N$ *we have that* $\exists L_N > 0$ *such that*

$$|b(t, x, \mu) - b(t, x', \mu)| \le L_N |x - x'|.$$

Throughout, denote by σ, the $d \times m$ matrix with columns $(\sigma^1, \cdots, \sigma^m)$.

Observe that time continuity is a sufficient condition for integrability of b and σ^l since we are working on a compact time interval. The above assumption ensures existence, uniqueness and related stability while the following will be used to ensure the differentiability results.

Assumption 3.2. *Let Assumption 3.1 hold. For any $t \ge 0$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ the maps $x \mapsto b(t, x, \mu)$ and $x \mapsto \sigma^l(t, x, \mu)$ are $C^1(\mathbb{R}^d)$. The derivative maps are jointly continuous in their variables.*

Remark 3.3. *One recognises that the above assumptions can be weakened in a few ways. For instance, the Lipschitz constant can be a non-negative function of time L_t, under an L^1-integrability condition: $\int_0^T L_s ds < \infty$. Further, the time continuity $t \mapsto b(t, 0, \delta_0)$ can be exchanged for an integrability condition: $\int_0^T |b(s, 0, \delta_0)| ds < \infty$, while the time-continuity of $t \mapsto \sigma(t, 0, \delta_0)$ can be exchanged for a square-integrability condition: $\int_0^T |\sigma(s, 0, \delta_0)|^2 ds < \infty$. We leave these points open for the interested reader.*

3.1.2 Well-posedness and moment estimates

The first result establishes well-posedness, moment estimates and continuity in time for the solution of (1).

Theorem 3.4. *Let Assumption 3.1 hold with some $p \ge 2$. Then, MV-SDE (1) is well-posed and has a unique solution $Z \in \mathcal{S}^p([0, T])$. Moreover, it*

satisfies

$$\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t|^p\right] \leq Ce^{CT}\left(\mathbb{E}[\,|\xi|^p] + \mathbb{E}\left[\left(\int_0^T |b(s,0,\delta_0)|ds\right)^p\right]\right.$$
$$\left. + \mathbb{E}\left[\left(\int_0^T |\sigma(s,0,\delta_0)|^2 ds\right)^{\frac{p}{2}}\right]\right),$$

for some positive constant C. Lastly, Z has $\mathbb{P}$-almost surely continuous paths and its law $[0,T] \ni t \mapsto \mu_t$ is continuous under the W_2-distance.

This result also yields estimates for standard SDEs (which have no measure dependency).

Proof. Well-posedness and the moment estimate follow from [AdRR$^+$22, Theorem 3.2] as their assumption (with $\mathcal{D} = \mathbb{R}^d$, x_0 and deterministic continuous maps b, σ) subsumes our Assumption 3.1. The continuity of the sample paths of Z and its law in W_2 is trivial. $\qquad\square$

3.2 Malliavin differentiability with locally Lipschitz coefficients

We state the first main result of this work, the Malliavin differentiability of the solution of (1).

Theorem 3.5. *Let $p \geq 2$. Let Assumption 3.2 hold. Denote by Z the unique solution (1) in $\mathcal{S}^p([0,T])$. Then Z is Malliavin differentiable, i.e. $Z \in \mathbb{D}^{1,2}(\mathcal{S}^2) \cap \mathbb{D}^{1,p}(\mathcal{S}^p)$, and the Malliavin derivative against W satisfies for $0 \leq s \leq t \leq T$,*

$$D_s Z_t = \sigma(s, Z_s, \mu_s) + \int_s^t (\nabla_x b)(r, Z_r, \mu_r) D_s Z_r dr$$
$$+ \sum_{l=1}^m \int_s^t (\nabla_x \sigma^l)(r, Z_r, \mu_r) D_s Z_r dW_r^l. \qquad (2)$$

If $s > t$ then $D_s Z_t = 0$ $\mathbb{P}$-a.s.
  Moreover, we have

$$\sup_{0 \leq s \leq T} \|D_s Z\|_{\mathcal{S}^p([0,T])}^p \leq \sup_{0 \leq s \leq T} \mathbb{E}\left[\sup_{0 \leq t \leq T} |D_s Z_t|^p\right] \leq C(1 + \|Z\|_{\mathcal{S}^p([0,T])}^p)$$
$$\leq C(1 + \|\xi\|_{L^p(\Omega)}^p) < \infty,$$

and, reflecting that DZ_t is a Hilbert space valued random variable we have

$$\mathbb{E}\Big[\sup_{0 \leq t \leq T} \big(\int_0^T |D_s Z_t|^2 ds \big)^{\frac{p}{2}} \Big] \leq C\big(1 + \|\xi\|_{L^p(\Omega)}^p \big) < \infty. \tag{3}$$

This proof is inspired by that appearing in [CM18] but with critical differences to allow for the superlinear growth of the drift and the general Malliavin differentiability of [IdRS19].

We comment that using the proof methodology we present below, the result above can be extended in several ways – we leave these as open questions. The first is to allow for random drift and diffusion coefficients: [AdRR$^+$22, Theorem 3.2] provides well-posedness and moment estimates and concluding Malliavin differentiability via [IdRS19, Theorem 3.2 or Theorem 3.7].

Another open setting, with continuous deterministic coefficients, is to establish our Theorem 3.5 with a drift map b having super-linear growth in the measure component: for instance, by allowing convolution type measure dependencies. Lastly, the differentiability requirements for b and σ can be weakened via mollification; see [IdRS19, Remark 3.4].

Example 3.6 (Linear interaction kernels)**.** *Take MV-SDE* (1) *with solution* $(Z_t, \mu_t)_{t \geq 0}$ *under the assumptions of Theorem 3.5 and let the drift function* b *take a specific convolutional form. Concretely, let* $b(t, x, \mu) = (\tilde{b} * \mu)(x) = \int_{\mathbb{R}^d} \tilde{b}(x - y)\mu(dy)$ *for some* $\tilde{b} \in C^1(\mathbb{R}^d)$ *and let* $\sigma = \sigma_0 I_d$ *for* $\sigma_0 \neq 0$ *a constant.*

Then, we have for any $0 \leq s \leq t \leq T$

$$D_s Z_t = \sigma_0 \exp\Big(\int_s^t (\nabla_x \tilde{b} * \mu_r)(Z_r) dr \Big). \tag{4}$$

We can in fact reduce the C^1 in space differentiability assumption to a Lipschitz one, exploiting [Nua06, Proposition 1.2.4]. *That is, we can take a sequence of mollifiers $\tilde{b}_n := \tilde{b} * \rho_n$ for a smoothing kernel $(\rho_n)_n$ such that* $C^\infty(\mathbb{R}^d) \ni \rho_n(x) \to x$ *uniformly, and the expression* (4) *still holds.*

We can generalise the above form of the drift to linear interaction kernels of the form $b(t, x, \mu) = \int_{\mathbb{R}^d} \widehat{b}(x, y)\mu(dy)$, where $\widehat{b} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $x \mapsto \widehat{b}(x, \cdot)$ is Lipschitz uniformly to obtain for any $0 \leq s \leq t \leq T$

$$D_s Z_t = \sigma_0 \exp\Big(\int_s^t \Big\{ \int_{\mathbb{R}^d} (\nabla_x \widehat{b})(Z_r, y)\mu_r(dy) \Big\} dr \Big).$$

Proof of Theorem 3.5. The Malliavin differentiability of (1) is shown by appealing to [Nua06, Lemma 1.2.3]. One builds a convenient sequence of Picard iterations which converge to the McKean-Vlasov Equation and use [Nua06, Lemma 1.2.3] to ensure that the limit is also Malliavin differentiable in $\mathbb{D}^{1,2}(\mathcal{S}^2)$. Lastly, by showing that the Malliavin derivative of Z is $\mathcal{S}^p$-integrable, then [Nua06, Proposition 1.5.5] yields that $Z \in \mathbb{D}^{1,p}(\mathcal{S}^p)$.

Step 1.0. The Picard sequence. We start by defining a Picard sequence approximation for (1), namely set $Z^0 = \xi$ and $\mu_t^0 = \mathbb{P} \circ \xi^{-1} = \text{Law}(\xi)$ for any $t \geq 0$; we have $t \mapsto \mu_t^0$ is a W_2-continuous map. For any $n \geq 1$ define

$$dZ_t^{n+1} = b(t, Z_t^{n+1}, \mu_t^n)dt + \sum_{l=1}^m \sigma^l(t, Z_t^{n+1}, \mu_t^n)dW_t^l, \quad Z_0^{n+1} = \xi. \quad (5)$$

(5) is a standard SDE with added time dependence induced by $t \mapsto \mu_t^n$ with drift b satisfying a one-sided Lipschitz condition (in space) and σ uniformly Lipschitz (in space).

Step 1.1. Existence and uniqueness of Z^n. Take μ^{n-1} such that $t \mapsto b(t, x, \mu_t^{n-1})$ and $t \mapsto \sigma^l(t, x, \mu_t^{n-1})$ are continuous, then (a slight variation of) Theorem 3.4, given Assumption 3.1, yields the existence of a unique solution $Z^{n+1} \in \mathcal{S}^2([0, T])$. Moreover, an easy variation of [dRST19, Proposition 3.4] yields that for $n \geq 0$, $t \mapsto \mu_t^{n+1}$ is continuous (in 2-Wasserstein distance) if $t \mapsto \mu_t^n$ is. We can conclude that $\{Z^n\}_{n \geq 0}$ exists and is well defined.

Using that $W_2(\delta_0, \mu_\cdot^n)^2 \leq \mathbb{E}[|Z_\cdot^n|^2]$ and Theorem 3.4, we have

$$\|Z^{n+1}\|_{\mathcal{S}^2([0,T])}^2$$
$$\leq C\Big(1 + \|\xi\|_{L^2(\Omega)}^2 + \int_0^T \mathbb{E}[|Z_r^n|^2]dr\Big)$$
$$\leq C\Big(1 + \|\xi\|_{L^2(\Omega)}^2 + \int_0^T \|Z^n\|_{\mathcal{S}^2([0,r])}^2 dr\Big)$$
$$\leq C\Big(1 + \|\xi\|_{L^2(\Omega)}^2 + \int_0^T \Big\{C(1 + \|\xi\|_{L^2(\Omega)}^2 + \int_0^r \|Z^{n-1}\|_{\mathcal{S}^2([0,s])}^2 ds\Big\}dr\Big)$$
$$\leq \cdots \leq C(1 + \|\xi\|_{L^2(\Omega)}^2)\Big(\sum_{j=0}^n \frac{(CT)^j}{j!}\Big)\|Z^0\|_{\mathcal{S}^2([0,T])}^2$$
$$\leq C(1 + \|\xi\|_{L^2(\Omega)}^2)e^{CT}\|Z^0\|_{\mathcal{S}^2([0,T])}^2,$$

where we iterated the initial estimate on $[0, T]$ over small subintervals $[0, r]$ leading to a known *simplex* estimate. We conclude that

$$\sup_{n \geq 0}\big\{\|Z^n\|_{\mathcal{S}^2([0,T])} + \sup_{0 \leq t \leq T} W_2(\delta_0, \mu_t^n)\big\} < \infty. \quad (6)$$

Step 1.2. Convergence of Z^n. Recall that (5) is a standard SDE, thus standard SDE stability estimation arguments apply. We sketch such argument only and invite the reader to inspect the proof of [dRST19, Proposition 3.3] for the full details. Take the SDE for the difference of $Z^{n+1} - Z^n$, i.e.

$$Z_t^{n+1} - Z_t^n = \int_0^T \left[b(s, Z_s^{n+1}, \mu_s^n) - b(s, Z_t^n, \mu_s^{n-1})\right]ds$$
$$+ \sum_{l=1}^m \int_0^T \left[\sigma^l(s, Z_s^{n+1}, \mu_s^n) - \sigma^l(s, Z_t^n, \mu_s^{n-1})\right]dW_s^l.$$

Applying Itô's formula to $|Z_t^{n+1} - Z_t^n|^2$, using the growth assumptions on b, σ and taking the supremum over time and expectations, we use the BDG inequality and then the Grönwall inequality to obtain

$$\|Z^{n+1} - Z^n\|_{\mathcal{S}^2([0,T])}^2$$
$$\leq C_T \left(\int_0^T W_2(\mu_r^n, \mu_r^{n-1})^2 dr\right) \leq C_T \int_0^T \mathbb{E}[|Z_r^n - Z_r^{n-1}|^2]]dr$$
$$\leq C_T \int_0^T \|Z^n - Z^{n-1}\|_{\mathcal{S}^2([0,r])}^2 dr \leq \cdots \leq \frac{(C_T T)^n}{n!} \|Z^1 - Z^0\|_{\mathcal{S}^2([0,T])}^2,$$

where we used the same *simplex* trick as in *Step 1.1* above. We conclude that Z^n converges to the solution of (1) in $\mathcal{S}^2([0,T])$ as $\|Z^1 - Z^0\|_{\mathcal{S}^2([0,T])}^2$ is bounded and independent of n.

Step 2. Malliavin differentiability for Z^n. Now, under our Assumption 3.2, [IdRS19, Corollary 3.5] holds applied to yield the Malliavin differentiability of Z^n for each fixed n; critically, due to the equation's coefficients being deterministic, that corollary does not require $p > 2$ and it holds for any $p \geq 2$ (in particular, our case here for the time being $p = 2$). We have that the Malliavin Derivative DZ^{n+1} satisfies $D_s Z_t^{n+1} = 0$ for $0 \leq t < s \leq T$, while for $0 \leq s \leq t \leq T$ it is given by the SDE dynamics

$$D_s Z_t^{n+1} = \sigma(s, Z_s^{n+1}, \mu_s^n) + \int_s^t (\nabla_x b)(r, Z_r^{n+1}, \mu_r^n) D_s Z_r^{n+1} dr$$
$$+ \sum_{l=1}^m \int_s^t (\nabla_x \sigma^l)(r, Z_r^{n+1}, \mu_r^n) D_s Z_r^{n+1} dW_r^l.$$

Step 3. Uniform bound on DZ^n. We remark that the SDE for DZ^n is linear and satisfies [IdRS19, Assumption 2.4] with their $b(s, \omega), \sigma(s, \omega)$ set

to zero, hence [IdRS19, Theorem 2.5] applies to yield

$$
\begin{aligned}
\|D.Z^n\|_{\mathcal{S}^2([0,T])}^2 = \mathbb{E}\big[\sup_{0\leq t\leq T}|D.Z_t^n|^2\big] &\leq C\mathbb{E}\big[\,|\sigma(\cdot,Z_{\cdot}^n,\mu_{\cdot}^{n-1})|^2\,\big] \\
&\leq C\Big(1+\|Z^n\|_{\mathcal{S}^2([0,T])}^2+W_2^2(\mu_{\cdot}^{n-1},\delta_0)\Big) \\
&\leq C\Big(1+\|Z^n\|_{\mathcal{S}^2([0,T])}^2+\|Z^{n-1}\|_{\mathcal{S}^p([0,T])}^2\Big) \\
&\leq C(1+\|\xi\|_{L^2(\Omega)}^2)<\infty, \tag{7}
\end{aligned}
$$

where we used the linear growth of σ and its time continuity property. The constant C depends heavily on the constants appearing in Assumption 3.1 and the upper bound on $t\mapsto\sigma(t,0,\delta_0)$ stemming from its continuity over the compact $[0,T]$. Using that (6) provides an estimate of $\|Z^n\|_{\mathcal{S}^2([0,T])}$ uniform over n and uniform over the Malliavin derivative parameter, we conclude taking supremum that

$$
\sup_{n\in\mathbb{N}}\sup_{s\in[0,T]}\mathbb{E}\big[\sup_{t\in[0,T]}|D_sZ_t^n|^2\big]\leq C(1+\|\xi\|_{L^2(\Omega)}^2)<\infty. \tag{8}
$$

We also establish the inequality for DZ_t^n that yields its interpretation as Hilbert space valued random variable. Simple manipulations and using (8) above easily yields estimate (3) when $p=2$ (when we pass to the limit in $n\to\infty$). Concretely, we have

$$
\begin{aligned}
\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^T|D_sZ_t^n|^2ds\Big] &\leq C\mathbb{E}\Big[\int_0^T\big\{\sup_{t\in[0,T]}|D_sZ_t^n|^2\big\}ds\Big] \\
&\leq C\int_0^T\mathbb{E}\Big[\sup_{t\in[0,T]}|D_sZ_t^n|^2\Big]ds \\
&\leq C\sup_{s\in[0,T]}\mathbb{E}\Big[\sup_{t\in[0,T]}|D_sZ_t^n|^2\Big] \\
&\leq C(1+\|\xi\|_{L^2(\Omega)}^p)<\infty, \tag{9}
\end{aligned}
$$

where we firstly used Jensen's inequality and moved the supremum inside the integral, then used Fubini and dominated the integral by the supremum over the integration variable in a way that (7) can be used. Taking supremum over n on both sides and using (8) yields (3).

By applying [Nua06, Lemma 1.2.3], we conclude that the limit of Z is Malliavin differentiable and the limit of DZ^n gives its Malliavin derivative (identified by Equation (2)). The moment estimates for DZ^n, holding uniformly over n, yield the moment estimate for DZ.

Step 4. Higher-order moments on DZ and conclusion for $\mathbb{D}^{1,p}(\mathcal{S}^p)$.

Recall that $p \geq 2$. From Theorem 3.4 we have that $Z \in \mathcal{S}^p([0,T])$ and the estimates

$$\|Z\|^p_{\mathcal{S}^p([0,T])} \leq C(1 + \|\xi\|^p_{L^p(\Omega)}) \quad \text{and} \quad \sup_{0 \leq t \leq T} W_2^2(\mu_t, \delta_0) \leq C(1 + \|\xi\|^2_{L^2(\Omega)}).$$

Noticing now that (2) is a standard linear SDE (in DZ) with random (time-continuous) coefficients satisfying a one-sided Lipschitz condition, i.e. the SDE for DZ is linear and satisfies [IdRS19, Assumption 2.4] with their $b(s,\omega), \sigma(s,\omega)$ set to zero, hence the moment estimate of [IdRS19, Theorem 2.5] applies with general $p \geq 2$ and yields

$$\|D.Z\|^p_{\mathcal{S}^p([0,T])} = \mathbb{E}\big[\sup_{0 \leq t \leq T} |D.Z^n_t|^p \big] \leq C\mathbb{E}\big[\, |\sigma(\cdot, Z_., \mu_.)|^p \,\big]$$

$$\leq C\Big(1 + \|Z^n\|^p_{\mathcal{S}^p([0,T])} + \sup_{0 \leq t \leq T} W_2^p(\mu_t, \delta_0)\Big)$$

$$\leq C(1 + \|\xi\|^p_{L^p(\Omega)}) < \infty.$$

Thus, akin to Estimate (8), we obtain from the above inequality

$$\sup_{s \in [0,T]} \mathbb{E}\big[\sup_{t \in [0,T]} |D_s Z_t|^p \big] = \sup_{s \in [0,T]} \|D_s Z\|^p_{\mathcal{S}^p([0,T])} \leq C(1 + \|\xi\|^p_{L^p(\Omega)}). \quad (10)$$

Further, following the footsteps of (9) but this time in p-moment norms we have

$$\mathbb{E}\Big[\sup_{t \in [0,T]} \Big(\int_0^T |D_s Z_t|^2 ds \Big)^{\frac{p}{2}} \Big] \leq C\mathbb{E}\Big[\int_0^T \big\{ \sup_{t \in [0,T]} |D_s Z_t|^p \big\} ds \Big]$$

$$\leq C \int_0^T \mathbb{E}\Big[\sup_{t \in [0,T]} |D_s Z_t|^p \Big] ds$$

$$\leq C \sup_{s \in [0,T]} \mathbb{E}\Big[\sup_{t \in [0,T]} |D_s Z_t|^p \Big]$$

$$\leq C(1 + \|\xi\|^p_{L^p(\Omega)}) < \infty,$$

where we firstly used Jensen's inequality and moved the supremum inside the integral, then used Fubini and dominated the integral by the supremum over the integration variable in a way that (10) can be used. This shows (3).

Having shown that $Z \in \mathcal{S}^p([0,T])$ and that DZ has higher-order p-moments (in the sense of (10) and (3)) we conclude via [Nua06, Proposition 1.5.5] that $Z \in \mathbb{D}^{1,p}(\mathcal{S}^p)$ (and space interpolation that $Z \in \mathbb{D}^{1,q}(\mathcal{S}^q)$ for any $q \in [2, p]$).

$\square$

4 Malliavin differentiability via the interacting particle system

The main goal of this section is to explore, in a didactic fashion, how much of Theorem 3.5 can be recovered under the interacting particle system approach. A somewhat close approach has been taken, for instance, in [HAHT21] and [CdRSW24].

To simplify arguments we will work under the further restriction of full Lipschitz conditions on the MV-SDE's coefficients and thus abdicate the more general super-linear growth and one-sided Lipschitz assumption.

4.1 The interacting and non-interacting particle system

We introduce the interacting particle system (IPS) associated to McKean-Vlasov SDE (1). Consider the system of SDEs for $i = 1, \cdots, N$:

$$dX_t^i = b(t, X_t^i, \bar{\mu}_t^N)dt + \sum_{l=1}^m \sigma^l(t, X_t^i, \bar{\mu}_t^N)dW_t^{l,i}, \quad X_0^i = \xi^i, \qquad (11)$$

where $\bar{\mu}_t^N(dy) = \frac{1}{N}\sum_{k=1}^N \delta_{X_t^i}(dy)$ and for $l = 1, \ldots, m$, $\{W^{l,i}\}_{i=1,\cdots,N}$, are independent 1-dimensional Brownian motions and $\{\xi^i\}_{i=1,\cdots,N}$ are i.i.d. copies of ξ; the $(W^{l,i}, \xi^i)_i$ in (11) are independent of W^l, ξ in (1) (and in fact, live in different probability spaces). We write $W^i = (W^{1,i}, \ldots, W^{m,i})$ to be the corresponding $m-$dimensional Brownian motions for $i = 1, \ldots, N$. The dependence on the empirical distributions in the coefficients introduces non-linearity into the system in the form of self-interaction; hence we refer to the above set of equations as an *interacting particle system* (IPS).

Since (1) and (11) live in different probability spaces, we construct an auxiliary *non-interacting particle system* (non-IPS) as living in the same probability space as (11). For $i = 1, \cdots, N$,

$$dZ_t^i = b(t, Z_t^i, \mu_t^i)dt + \sum_{l=1}^m \sigma^l(t, Z_t^i, \mu_t^i)dW_t^{l,i}, \qquad Z_0^i = \xi^i, \qquad (12)$$

where μ^i is defined as the law of Z^i. In this case, the $\{Z^i\}_{i=1,\cdots,N}$ are independent of each other, since the $(W^i, \xi^i)_i$ are all i.i.d. and $\mu^i = \mu^j = \mu$ $\forall i, j = 1, \cdots, N$ where μ denotes the law of the McKean-Vlasov SDE (1). In essence, (12) is a decoupled system of N copies of (1). From direct inspection of (12), we have the following lemma regarding the cross-Malliavin derivatives $D^j Z^i$ for $i \neq j$.

Lemma 4.1. *Assume* (12) *is well-posed. Then, the cross-Malliavin derivatives of the solution* $\{Z^i\}_{i=1,\cdots,N}$ *to* (12) *are all zero. That is,*

$$D_s^j Z_t^i = 0 \quad \text{for any} \quad j \neq i, 1 \leq i, j \leq N, \quad s, t \in [0, T].$$

Proof. One sees that Z^i acts independently of any other Brownian motion W^j for $j \neq i$. The result follows immediately from the definition of the Malliavin derivative. □

4.1.1 Preliminaries

In this section, we work under stronger assumptions requiring b and σ to be globally space-measure Lipschitz. Formally we state the framework as follows:

Assumption 4.2. *Let* $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ *and for* $l = 1, \ldots, m$, $\sigma^l : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ *be progressively measurable deterministic maps and* $\exists L > 0$ *such that:*

1. *For some* $p \geq 2$, $\xi^i \in L^p(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^d)$ *for* $i = 1, \cdots, N$,

2. σ^l *is continuous in time and Lipschitz in space-measure* $\forall t \in [0, T]$, $\forall x, x' \in \mathbb{R}^d$ *and* $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ *we have*

$$|\sigma^l(t, x, \mu) - \sigma^l(t, x', \mu')| \leq L\Big(|x - x'| + W_2(\mu, \mu')\Big).$$

3. b *is continuous in time and Lipschitz in space-measure* $\forall t \in [0, T]$, $\forall x, x' \in \mathbb{R}^d$ *and* $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$ *we have*

$$|b(t, x, \mu) - b(t, x', \mu')| \leq L\Big(|x - x'| + W_2(\mu, \mu')\Big).$$

A quick inspection of (11) highlights that this system of equations can be seen as a system in $(\mathbb{R}^d)^N$ as opposed to N equations valued in $\mathbb{R}^d$. The former has a few advantages and to formalize it we introduce the notion of an *empirical projection* introduced in [CD18a, Definition 5.34].

Definition 4.3 (Empirical projection of a map). *Given* $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ *and* $N \in \mathbb{N}$, *define the empirical projection* u^N *of* u *via* $u^N : (\mathbb{R}^d)^N \to \mathbb{R}^d$, *such that*

$$u^N(x^1, \ldots, x^N) := u(\bar{\mu}^N), \quad \text{with} \quad \bar{\mu}^N(dy) := \frac{1}{N} \sum_{l=1}^N \delta_{x^l}(dy),$$

for $x^l \in \mathbb{R}^d, l = 1, \ldots, N$.

We can use a similar notion of mapping points onto empirical projections to express the interacting particle system as a high-dimensional SDE. Hence we can interpret system (11) as a system in $(\mathbb{R}^d)^N$ with $B^N : [0,T] \times (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ and $\Sigma^N : [0,T] \times (\mathbb{R}^d)^N \to \mathbb{R}^{mN \times dN}$,

$$d\mathbf{X}_t = B^N(t, \mathbf{X}_t)dt + \Sigma^N(t, \mathbf{X}_t)d\mathbf{W}_t, \quad \mathbf{X}_0 = \boldsymbol{\xi}, \tag{13}$$

for $\mathbf{X} = (X^1, \cdots, X^N)$, $\mathbf{W} = (W^1, \cdots, W^N)$ and $\boldsymbol{\xi} = (\xi^1, \cdots, \xi^N)$ where for $t \in [0,T]$ and $\mathbf{x} = (x^1, \cdots, x^N) \in \mathbb{R}^{dN}$, $x^i \in \mathbb{R}^d$, $i = 1, \cdots, N$ we have

$$\begin{aligned}
B^N(t, \mathbf{x}) &= (b_1^N(t, \mathbf{x}), \cdots, b_N^N(t, \mathbf{x})) \\
&:= (b(t, x^1, \bar{\mu}^N), \cdots, b(t, x^N, \bar{\mu}^N)) \\
\Sigma^N(t, \mathbf{x}) &= \mathrm{diag}(\sigma_1^N(t, \mathbf{x}), \cdots, \sigma_N^N(t, \mathbf{x})) \\
&:= \mathrm{diag}(\sigma(t, x^1, \bar{\mu}^N), \cdots, \sigma(t, x^N, \bar{\mu}^N)),
\end{aligned}$$

with the relation between $\mathbf{x}$ and $\bar{\mu}^N$ being as highlighted in Definition 4.3. The next result shows that from the Lipschitz properties of b, σ in space and measure, one can show that $\mathbf{x} \mapsto B^N(\cdot, \mathbf{x}), \Sigma^N(\cdot, \mathbf{x})$ are uniformly Lipschitz (uniformly in time).

Lemma 4.4. *Under Assumption 3.1, the maps B^N and Σ^N in (13) are globally Lipschitz in their spatial variables.*

Proof. Let $\mathbf{x} = (x^1, \cdots, x^N), \mathbf{y} = (y^1, \cdots, y^N) \in \mathbb{R}^{dN}$, for $x^i, y^i \in \mathbb{R}^d$, $i = 1, \cdots, N$ then

$$\begin{aligned}
|B^N(t, \mathbf{x}) - B^N(t, \mathbf{y})|^2 &= \sum_{k=1}^N \left| b(t, x^k, \bar{\mu}^N(\mathbf{x}) - b(t, y^k, \bar{\mu}^N(\mathbf{y}) \right|^2 \\
&\leq L^2 \sum_{k=1}^N \left(|x^k - y^k| + W_2(\bar{\mu}^N(\mathbf{x}), \bar{\mu}^N(\mathbf{y})) \right)^2 \\
&\leq 2L^2 \sum_{k=1}^N |x^k - y^k|^2 + W_2^2(\bar{\mu}^N(\mathbf{x}), \bar{\mu}^N(\mathbf{y})) \leq 4L^2 |\mathbf{x} - \mathbf{y}|^2,
\end{aligned}$$

with the final inequality arising from the fact

$$W_2 \left(\frac{1}{N} \sum_{i=1}^N \delta_{x^i}, \frac{1}{N} \sum_{i=1}^N \delta_{y^i} \right) \leq \left(\frac{1}{N} \sum_{i=1}^N |x^i - y^i|^2 \right)^{\frac{1}{2}} = \frac{1}{\sqrt{N}} |\mathbf{x} - \mathbf{y}|. \tag{14}$$

The proof is similar for $\mathbb{R}^{dN} \ni \mathbf{x} \mapsto \Sigma^N(\mathbf{x}) \in \mathbb{R}^{mN \times dN}$. $\qquad \square$

4.1.2 Classical results

We briefly recall classical results involving the relationship between systems described above. Well-posedness follows from classic literature, while the second result is classically known as Propagation of Chaos (PoC), which ascertains convergence of X^i to Z^i as $N \to \infty$ (as respective laws) [Szn91].

Proposition 4.5 (Well-posedness and Propagation of Chaos)**.** *Let Assumption 4.2 hold. Then, the solutions to the systems* (11) *and* (12)*, given by* $\{X_t^i\}_{i=1,\cdots,N}$ *and* $\{Z_t^i\}_{i=1,\cdots,N}$ *respectively are well-posed, unique and square integrable. It holds that*

$$\sup_{N \in \mathbb{N}} \max_{1 \leq i \leq N} \left\{ \mathbb{E}\big[\sup_{0 \leq t \leq T} |Z_t^i|^2 \big] + \mathbb{E}\big[\sup_{0 \leq t \leq T} |X_t^i|^2 \big] \right\} \leq C(1 + \mathbb{E}[|\xi^{\cdot}|^2])e^{CT} < \infty, \tag{15}$$

where the involved constant C depends on d, m, L and the quantity

$$\int_0^T |b(t, 0, \delta_0)| dt + \int_0^T |\sigma(t, 0, \delta_0)|^2 dt,$$

but independent of N. Moreover, we have for any $i = 1, \cdots, N$,

$$\lim_{N \to \infty} \sup_{0 \leq t \leq T} \mathbb{E}[W_2^2(\mu_t^i, \bar{\mu}_t^N)] = 0 \ and \lim_{N \to \infty} \max_{1 \leq i \leq N} \mathbb{E}\Big[\sup_{0 \leq t \leq T} \Big| X_t^i - Z_t^i \Big|^2 \Big] = 0. \tag{16}$$

Proof. Under Lipschitz conditions this result is classical. Well-posedness of (12) follows directly from Theorem 3.4 as it is a non-interacting particle system (thus well-posedness of the initial McKean-Vlasov equation suffices).

As for system (11), Lemma 4.4 ensures the coefficients are uniformly Lipschitz (as maps in $(\mathbb{R}^d)^N$) and thus well-posedness (for fixed N) follows from general SDE theory [GK80, Theorem 1]. One can conclude the uniform in N estimates of (15) for $\{X^i\}_i$ by mimicking the arguments used in the proof of Theorem 3.5.

The convergence results of (16) follows from [Car16, Lemma 1.9 and Theorem 1.10]. □

4.1.3 A primer on Lions derivatives

To consider the calculus for the mean-field setting, one requires to build a suitable differentiation operator on 2-Wasserstein space. Among the

numerous notions of differentiability of a function u defined over the $\mathcal{P}_2(\mathbb{R}^d)$, we try to follow the approach introduced by Lions in his lectures at Collège de France. A comprehensive collection of recent results was done in the joint monographs of Carmona and Delarue [CD18a], [CD18b]. In line with the construction we assume our probability space to be an atomless Polish space [CD18a, Chapter 5].

We consider a canonical lifting of the function $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ to $\tilde{u} : L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) \ni X \to \tilde{u}(X) = u(Law(X)) \in \mathbb{R}^d$. We can say that u is L-differentiable at μ, if $\tilde{u}$ is Fréchet differentiable at some X, such that $\mu = \mathbb{P} \circ X^{(-1)}$. Denoting the gradient by $D\tilde{u}$ and using a Hilbert structure of the L^2 space, we can identify $D\tilde{u}$ as an element of L^2. It has been shown that $D\tilde{u}$ is a $\sigma(X)$-measurable random variable and given by the function $Du(\mu)(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, depending on the law of X and satisfying $Du(\mu)(\cdot) \in L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu; \mathbb{R}^d)$. Hereinafter the L-derivative of u at μ is the map $\partial_\mu u(\mu)(\cdot) : \mathbb{R}^d \ni v \to \partial_\mu u(\mu)(v) \in \mathbb{R}^d$, satisfying $D\tilde{u}(X) = \partial_\mu u(\mu)(X)$. We always denote $\partial_\mu u$ as the version of the L-derivative that is continuous in product topology of all components of u.

Definition 4.3 relates the spatial derivatives of u^N with the Lions derivative of the measure function u. Such is stated next; see also [CD18a, Proposition 5.35 (p.399)].

Proposition 4.6. *Let $u : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ be a continuously L-differentiable map, then, for any $N > 1$, the empirical projection u^N is differentiable in $(\mathbb{R}^d)^N$ and for all $x^1, \cdots, x^N \in \mathbb{R}^d$ we have the following relation:*

$$\partial_{x^j} u^N(x^1, \ldots, x^N) = \frac{1}{N} \partial_\mu u\Big(\frac{1}{N} \sum_{l=1}^N \delta_{x^l}\Big)(x^j).$$

4.2 Exploring Malliavin differentiability via interacting particle system limits

The novelty in this section lies not within the results, as these are implied directly by those in our Section 3.2 or [CM18], but in the proof methodology via limits of interacting particle systems.

Assumption 4.7. *Let Assumption 4.2 hold.*

1. *The functions b, σ^l, $l = 1, \ldots, m$ are continuously differentiable in their spatial variables and their spatial derivative maps are continuous in time. Further, the maps $(\nabla_x b)$, $(\nabla_x \sigma^l)(t, x, \mu)$ are uniformly bounded for $l = 1, \ldots, m$ (over all variables $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$).*

2. *For any $t \in [0,T]$ the maps $\mu \mapsto b(t,x,\mu)$ and $\mu \mapsto \sigma^l(t,x,\mu)$, $l = 1,\ldots,m$, are $\mathbb{P}$-a.s. continuous in topology, induced by the Wasserstein metric and L-differentiable $\mathbb{P}$-a.s. at every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Moreover, $\partial_\mu b(t,x,\mu)(v)$ and $\partial_\mu \sigma^l(t,x,\mu)(v)$ have μ-versions such that $\partial_\mu b(t,x,\mu)(v)$ and $\partial_\mu \sigma^l(t,x,\mu)(v)$ are $\mathbb{P}$-a.s. joint-continuous at every quadruple (t,x,μ,v) with $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, $v \in Supp(\mu)$ and uniformly bounded (in all variables).*

Focusing on the technique of limits of particle systems, we establish the following result that covers the Malliavin differentiability of the particle system, the Propagation of Chaos result and how to transfer the Malliavin regularity to the limiting McKean-Vlasov SDE.

Proposition 4.8. *Let Assumption 4.7 hold. Then the solution $\{Z^i\}_{i=1,\cdots,N}$ to (12) and the solution $\{X^i\}_{i=1,\cdots,N}$ to (11) are Malliavin differentiable with Malliavin derivatives $\{D^j Z^i\}_{i,j=1,\cdots,N}$ and $\{D^j X^i\}_{i,j=1,\cdots,N}$ respectively.*

For $\{D^j Z^i\}_{i,j=1,\cdots,N}$ we have that:

- *For any $j \neq i$, $1 \leq i,j \leq N$ $s,t \in [0,T]$ that $D_s^j Z_t^i = 0$ (Lemma 4.1).*

- *When $j = i$ then $D^i Z^i$ satisfies for $0 \leq s \leq t \leq T$*

$$D_s^i Z_t^i = \sigma(s, Z_s^i, \mu_s^i) + \int_s^t (\nabla_x b)(r, Z_r^i, \mu_r^i) D_s^i Z_r^i dr$$
$$+ \sum_{l=1}^m \int_s^t (\nabla_x \sigma^l)(r, Z_r^i, \mu_r^i) D_s^i Z_r^i dW_r^{l,i}. \qquad (17)$$

If $s > t$ then $D_s^i Z_t^i = 0$ $\mathbb{P}$-almost surely.

- *Moreover, for some $C > 0$ dependent on T and L but not on N, (c.f. Theorem 3.5)*

$$\sup_{0 \leq s \leq T} \|D_s^i Z^i\|_{\mathcal{S}^2([0,T])}^2 \leq \sup_{0 \leq s \leq T} \mathbb{E}\Big[\sup_{0 \leq t \leq T} |D_s^i Z_t^i|^2 \Big]$$
$$\leq C(1 + \|Z^i\|_{\mathcal{S}^2([0,T])}^2) \leq C(1 + \|\xi^{\cdot}\|_{L^2(\Omega)}^2) < \infty.$$

For $\{D^j X^i\}_{i,j=1,\cdots,N}$, the Malliavin derivative of (11), we have that:

- *If $s > t$ then $D_s^j X_t^i = 0$ $\mathbb{P}$-almost surely for any $i,j = 1,\cdots,N$.*

- *If $0 \le s \le t$ then $D_s^j X_t^i$ satisfies*

$$
\begin{aligned}
D_s^j X_t^i = {} & \sigma(s, X_s^i, \bar{\mu}_s^N) \mathbb{1}_{i=j} \\
& + \int_s^t \Big\{ (\nabla_x b)(r, X_r^i, \bar{\mu}_r^N) D_s^j X_r^i && (18) \\
& \qquad + \frac{1}{N} \sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k \Big\} \, dr \\
& + \sum_{l=1}^m \int_s^t \Big\{ (\nabla_x \sigma^l)(r, X_r^i, \bar{\mu}_r^N) D_s^j X_r^i \\
& \qquad + \frac{1}{N} \sum_{k=1}^N (\partial_\mu \sigma^l)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k \Big\} \, dW_r^i. && (19)
\end{aligned}
$$

- *Moreover, there exists a constant $C > 0$ depending on T and L but not on N such that*

$$
\sup_{0 \le s \le T} \mathbb{E}\left[\sup_{0 \le t \le T} |D_s^j X_t^i|^2 \right] \le C\Big(\mathbb{1}_{i=j} + \frac{1}{N} \Big).
$$

Finally, for any $s, t \in [0, T]$,

- $D_s^j X_t^i \to 0$ *as $N \to \infty$ for $j \ne i$ in $L^2(\Omega)$ and almost surely, and*

- $D_s^i X_t^i \to D_s^i Z_t^i$ *as $N \to \infty$ in $L^2(\Omega)$.*

Proof. Rewrite the IPS (11) in integral form: for $i = 1, \ldots, N$,

$$
\begin{aligned}
X_t^i & = \xi^i + \int_0^t b(r, X_r^i, \bar{\mu}_r^N) dr + \sum_{l=1}^m \int_0^t \sigma^l(r, X_r^i, \bar{\mu}_r^N) dW_r^{l,i}, && (20) \\
& = \xi^i + \int_0^t b_i^N(r, X_r^1, \cdots, X_r^N) dr \\
& \qquad + \sum_{l=1}^m \int_0^t \sigma_i^{l,N}(r, X_r^1, \cdots, X_r^N) dW_r^{l,i}, && (21)
\end{aligned}
$$

where (21) uses the empirical projection representation (Definition 4.3) of the coefficients in (20); that is $b_i^N(t, X^1, \cdots, X^N) := b(t, X^i, \bar{\mu}^N)$ and similarly for $\sigma_i^{l,N}$. In view of Assumption 4.7, it is easy to conclude that the coefficients $b_i^N, \sigma_i^{l,N}$ of (20) are Lipschitz continuous (via Lemma 4.4) and also differentiable in their variables (via Proposition 4.6). From the results

in [Nua06] or [IdRS19] for classical SDEs we conclude immediately that the Malliavin derivatives of X^i exist, are unique and are square-integrable. Furthermore, if $s > t$ then $D_s^j X_t^i = 0$ $\mathbb{P}$-a.s. for any $i, j = 1, \cdots, N$.

From application of the chain rule, the Malliavin derivative is written for $0 \leq s \leq t \leq T < \infty$ as

$$D_s^j X_t^i = \sigma_i^N(s, X_s^1, \cdots, X_s^N)\mathbb{1}_{i=j} + \int_s^t \sum_{k=1}^N (\partial_{x^k} b_i^N)(r, X_r^1, \cdots, X_r^N)D_s^j X_r^k dr$$

$$+ \sum_{l=1}^m \int_s^t \sum_{k=1}^N (\partial_{x^k}\sigma_i^{l,N})(r, X_r^1, \cdots, X_r^N)D_s^j X_r^k dW_r^{l,i}.$$

Exploiting Proposition 4.6 and reverting the empirical projection maps to their original form, we rewrite this as

$$D_s^j X_t^i = \sigma(s, X_s^i, \bar{\mu}_s^N)\mathbb{1}_{i=j}$$

$$+ \int_s^t \left\{ (\nabla_x b)(r, X_r^i, \bar{\mu}_r^N)D_s^j X_r^i \right.$$

$$\left. + \frac{1}{N}\sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k)D_s^j X_r^k \right\} dr$$

$$+ \sum_{l=1}^m \int_s^t \left\{ (\nabla_x \sigma^l)(r, X_r^i, \bar{\mu}_r^N)D_s^j X_r^i \right.$$

$$\left. + \frac{1}{N}\sum_{k=1}^N (\partial_\mu \sigma^l)(r, X_r^i, \bar{\mu}_r^N)(X_r^k)D_s^j X_r^k \right\} dW_r^i.$$

noting that the derivatives $\partial_{x^i} b_i^N$ and $\partial_{x^i}\sigma_i^{l,N}$ produce two components as opposed to the one produced by the cross-terms $\partial_{x^k} b_i^N$ and $\partial_{x^k}\sigma_i^{l,N}$, $k \neq i$.

Step 1. Preliminary manipulations. Applying Itô's formula to (18),

$$|D_s^j X_t^i|^2 = |\sigma(s, X_s^i, \bar{\mu}_s^N)|^2 \mathbb{1}_{i=j} \tag{22}$$

$$+ \int_s^t 2D_s^j X_r^i : \left\{ (\nabla_x b)(r, X_r^i, \bar{\mu}_r^N)D_s^j X_r^i \right. \tag{23}$$

$$\left. + \frac{1}{N}\sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k)D_s^j X_r^k \right\}$$

$$+ \sum_{l=1}^m \left| (\nabla_x \sigma^l)(r, X_r^i, \bar{\mu}_r^N)D_s^j X_r^i + \frac{1}{N}\sum_{k=1}^N (\partial_\mu \sigma^l)(r, X_r^i, \bar{\mu}_r^N)(X_r^k)D_s^j X_r^k \right|^2 dr$$

$$+ M_N(s, t), \tag{24}$$

where M_N is a Hilbert space-valued local martingale term. In fact, by standard SDE well-posedness theory, e.g. [GK80, Theorem 1], for any $s \in [0, T]$, a L^2-integrable solution to Equation (22) exists, allowing us to conclude M_N is a proper martingale for finite N.[1]

Let L be the Lipschitz constant which bounds the spatial and measure derivatives of b and σ^l, $l = 1, \ldots, m$. Temporarily fix $s \in [0, T]$. We have

$$
\begin{aligned}
&\mathbb{E}\big[\sup_{0 \le t \le T} |D_s^j X_t^i|^2 \big] \\
&\quad \le \mathbb{E}\big[|\sigma(s, X_s^i, \bar{\mu}_s^N) \mathbb{1}_{i=j}|^2 \big] + \mathbb{E}\big[\sup_{0 \le t \le T} |M_N(t, s)| \big] \\
&\quad + \mathbb{E}\bigg[\sup_{0 \le t \le T} \int_s^t 2L |D_s^j X_r^i|^2 + 2mL^2 |D_s^j X_r^i|^2 \\
&\qquad\qquad + \frac{2L |D_s^j X_r^i|}{N} \sum_{k=1}^N |D_s^j X_r^k| + 2mL^2 \Big| \frac{1}{N} \sum_{k=1}^N |D_s^j X_r^k| \Big|^2 \, dr \bigg] \\
&\quad \le \mathbb{E}\big[|\sigma(s, X_s^i, \bar{\mu}_s^N) \mathbb{1}_{i=j}|^2 \big] + \kappa \mathbb{E}\bigg[\sup_{0 \le t \le T} \int_s^t |D_s^j X_r^i|^2 + \frac{1}{N} \sum_{k=1}^N |D_s^j X_r^k|^2 \, dr \bigg],
\end{aligned}
\tag{25}
$$

for some $\kappa > 0$, by applying the Young and Jensen inequalities, where we used the BDG inequality to control the martingale term. Further, one can bound $\sup_{0 \le s \le T} \mathbb{E}[|\sigma(s, X_s^i, \bar{\mu}_s^N)|^2]$ uniformly in N, justified by well-posedness and the uniform in N moment bounds of the SDE-IPS (and also using the continuity assumption on $t \mapsto \sigma(t, 0, \delta_0)$). That is, by linear growth of σ, properties of the Wasserstein metric and Proposition 4.5,

$$
\begin{aligned}
&\sup_{0 \le s \le T} \mathbb{E}\big[|\sigma(s, X_s^i, \bar{\mu}_s^N)|^2 \big] \\
&\qquad \le C\Big(1 + \sup_{0 \le s \le T} \mathbb{E}\big[|X_s^i|^2 \big] + \frac{1}{N} \sum_{k=1}^N \sup_{0 \le s \le T} \mathbb{E}\big[|X_s^k|^2 \big] \Big) \\
&\qquad \le C(1 + \mathbb{E}[|\xi^\cdot|^2]) e^{CT} =: \alpha,
\end{aligned}
\tag{26}
$$

for some $C > 0$ independent of N, observing that the $\{\xi^j\}_j$ are i.i.d.

---

[1] We note that [GK80, Theorem 1] provides wellposedness and moment estimates for the SDE, but critically, such estimates are *dependent on N* and explode as $N \nearrow +\infty$. Nonetheless, for any fixed N the estimates suffice to ensure that $\mathbb{E}[M_N(s, t)] = 0$ for any s, t. Obtaining moment estimates uniformly in N is done in a subsequent step of the proof. This is the exact same as in the proof of Proposition 4.5.

Step 2. Controlling the empirical mean of the Malliavin derivative. We now aim to gain control over the quantity $\frac{1}{N}\sum_{i=1}^{N} D_s^j X_r^i$. Averaging (25) over the index i and once more using that the $\{\xi^j\}_j$ are i.i.d., we have

$$\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^{N}\sup_{0\leq t\leq T}|D_s^j X_t^i|^2\Big]$$

$$\leq \frac{\alpha}{N} + \mathbb{E}\bigg[\frac{\kappa}{N}\sum_{i=1}^{N}\int_s^T\Big\{|D_s^j X_r^i|^2 + \frac{1}{N}\sum_{m=1}^{N}|D_s^j X_r^m|^2\Big\}dr\bigg]$$

$$\leq \frac{\alpha}{N} + 2\kappa\mathbb{E}\bigg[\frac{1}{N}\int_s^T\sum_{i=1}^{N}|D_s^j X_r^i|^2 dr\bigg]. \tag{27}$$

Taking the supremum inside the expectation means we are not able to apply Grönwall's inequality directly. However, we have that

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|D_s^j X_r^i|^2\right] \leq \frac{\alpha}{N} + 2\kappa\mathbb{E}\bigg[\frac{1}{N}\int_s^r\sum_{i=1}^{N}|D_s^j X_\rho^i|^2 d\rho\bigg].$$

Applying Grönwall's inequality yields

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}|D_s^j X_r^i|^2\right] \leq \frac{\alpha}{N}\sup_{0\leq r,s\leq T}\Big(1 + e^{2\kappa(r-s)}(e^{-2\kappa s} - e^{-2\kappa r})\Big) \leq \frac{\alpha}{N}e^{2\kappa T}. \tag{28}$$

Hence, substituting (28) back into (27), we get

$$\sup_{0\leq s\leq T}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\sup_{0\leq t\leq T}|D_s^j X_t^i|^2\right] \leq \frac{2\alpha}{N}(1 + \kappa T e^{2\kappa T}) := \Psi_N.$$

It is immediate to see that Ψ_N is uniformly bounded over N, hence

$$\lim_{N\to\infty}\sup_{0\leq s\leq T}\mathbb{E}\left[\sup_{0\leq t\leq T}\frac{1}{N}\sum_{i=1}^{N}|D_s^j X_t^i|^2\right] = 0 \qquad \text{for any} \quad j = 1,\cdots,N.$$

Using Jensen's inequality and (28), observe that

$$\sup_{0\leq s\leq T}\mathbb{E}\left[\sup_{0\leq t\leq T}\frac{1}{N}\sum_{i=1}^{N}|D_s^j X_t^i|^2\right] \leq \Psi_N \to 0 \quad \text{as } N \to +\infty \tag{29}$$

$$\Rightarrow \quad \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}|D_s^j X_t^k| = 0 \text{ } \mathbb{P}\text{-a.s.}$$

Step 3. Convergence of the individual terms. Injecting the bound Ψ_N into (25), we obtain

$$\sup_{0 \leq s \leq T} \mathbb{E}\left[\sup_{0 \leq t \leq T} |D_s^j X_t^i|^2\right] \leq \alpha \mathbb{1}_{i=j} + \kappa T \Psi_N + \kappa \int_0^T \mathbb{E}\left[|D_s^j X_t^i|^2\right] dr.$$

Applying Grönwall's inequality again, we bound

$$\mathbb{E}\left[|D_s^j X_t^i|^2\right] \leq \left(\alpha \mathbb{1}_{i=j} + \kappa T \Psi_N\right) + \kappa \int_s^t e^{\kappa(t-s-r)} \left(\alpha \mathbb{1}_{i=j} + \kappa T \Psi_N\right) dr$$

$$\Rightarrow \mathbb{E}\left[|D_s^j X_t^i|^2\right] \leq e^{\kappa T} \left(\alpha \mathbb{1}_{i=j} + \kappa T \Psi_N\right).$$

Hence,

$$\sup_{0 \leq s \leq T} \mathbb{E}\left[\sup_{0 \leq t \leq T} |D_s^j X_t^i|^2\right] \leq \left(\alpha \mathbb{1}_{i=j} + \kappa T \Psi_N\right)\left(1 + \kappa T e^{\kappa T}\right). \qquad (30)$$

Hence, applying an identical argument to Equation (9) we have:

$$\sup_{N \in \mathbb{N}} \max_{1 \leq i,j \leq N} \mathbb{E}\left[\sup_{0 \leq t \leq T} \int_0^T |D_s^j X_t^i|^2 ds\right] < \infty.$$

In particular, we have that $X_t^i \in \mathbb{D}^{1,2}(\mathcal{S}^2)$ uniformly in N.

Aside, we obtain that $D_s^j X_t^i \to 0$ in $\mathbb{D}^{1,2}(\mathcal{S}^2)$ for $i \neq j$ as the size of interacting particle system $N \to \infty$. This is in line with Lemma 4.1, since referring back to our non-IPS analogy, particles essentially become more conditionally independent as the particle system size gets larger.

We can now apply [Nua06, Lemma 1.2.3]. Let $\{Z^i\}_{i=1,\cdots,N}$ denote the solution to the non-IPS (12). By Proposition 4.5, we have

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} \mathbb{E}\left[\sup_{0 \leq t \leq T} \left|X_t^i - Z_t^i\right|^2\right] = 0.$$

The established uniform in N upper bound on $\mathbb{E}[|D_s^j X_r^i|^2]$, given by (30), allows us to conclude that $Z^i \in \mathbb{D}^{1,2}(\mathcal{S}^2)$ and $D_s^j Z_t^i = \lim_{N \to \infty} D_s^j X_t^i$ as prescribed by [Nua06, Lemma 1.2.3].

Step 4. Recovering the limiting equation and conclusion. It remains only to identify and confirm the stochastic differential equation the limiting object of $D_s^j X_t^i$ as $N \to \infty$ satisfies.

We now consider the L^2-limit of the right hand side of (18). First, note that the MV-SDE (17) is an affine SDE with random coefficients

that are space-measure Lipschitz, hence existence and uniqueness follows from [Mao08, Theorem 2.1]. Define $\delta\sigma^l_s := \sigma^l(s, X^i_s, \bar{\mu}^N_s) - \sigma^l(s, Z^i_s, \mu^i_s)$, $l = 1, \ldots, m$ and $\delta\sigma_s$ to be the $d \times m$ matrix valued process with columns $\delta\sigma^l_s$. By our Lipschitz assumption on σ^l, for all $s \in [0, T]$:

$$|\delta\sigma_s| \le L\Big(|X^i_s - Z^i_s| + W_2(\bar{\mu}^N_s, \mu^i_s)\Big) \to 0, \tag{31}$$

as $N \to \infty$ in $L^2(\Omega)$ by Proposition 4.5. Computing the difference between the SDEs (17) and (18),

$$D^j_s X^i_t - D^j_s Z^i_t = \delta\sigma_s + \int_s^t \theta_r dr + \int_s^t \Big\{ \frac{1}{N} \sum_{k=1}^N (\partial_\mu b)(r, X^i_r, \bar{\mu}^N_r)(X^k_r) D^j_s X^k_r \Big\} dr$$

$$+ \sum_{l=1}^m \int_s^t \eta^l_r dW^{l,i}_r + \sum_{l=1}^m \int_s^t \Big\{ \frac{1}{N} \sum_{k=1}^N (\partial_\mu \sigma^l)(r, X^i_r, \bar{\mu}^N_r)(X^k_r) D^j_s X^k_r \Big\} dW^{l,i}_r,$$

where

$$\theta_r := (\nabla_x b)(r, X^i_r, \bar{\mu}^N_r) D^j_s X^i_r - (\nabla_x b)(r, Z^i_r, \mu^i_r) D^j_s Z^i_r,$$
$$\eta^l_r := (\nabla_x \sigma^l)(r, X^i_r, \bar{\mu}^N_r) D^j_s X^i_r - (\nabla_x \sigma^l)(r, Z^i_r, \mu^i_r) D^j_s Z^i_r.$$

Squaring both sides of the equation, expanding the squares and taking expectations, we have

$$\|D^j_s X^i_t - D^j_s Z^i_t\|^2_{L^2(\Omega)} - \Phi_N = \Big\| \int_s^t \theta_r dr \Big\|^2_{L^2(\Omega)} + \Big\| \sum_{l=1}^m \int_s^t \eta^l_r dW^{l,i}_r \Big\|^2_{L^2(\Omega)}, \tag{32}$$

where we define

$$
\Phi_N := \mathbb{E}\Big[|\delta\sigma_s|^2\Big] + 2\mathbb{E}\Big[\delta\sigma_s : \int_s^t \theta_r dr\Big]
$$

$$
+ 2\mathbb{E}\Big[\delta\sigma_s : \int_s^t \frac{1}{N}\sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k dr\Big]
$$

$$
+ 2\mathbb{E}\Big[\int_s^t \theta_r dr : \int_s^t \frac{1}{N}\sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k dr\Big]
$$

$$
+ 2\mathbb{E}\Big[\sum_{l=1}^m \int_s^t \eta_r^l dr : \int_s^t \frac{1}{N}\sum_{k=1}^N (\partial_\mu \sigma^l)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k dr\Big]
$$

$$
+ \mathbb{E}\Big[\Big|\int_s^t \frac{1}{N}\sum_{k=1}^N (\partial_\mu b)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k dr\Big|^2\Big]
$$

$$
+ \mathbb{E}\Big[\sum_{l=1}^m \int_s^t \Big|\frac{1}{N}\sum_{k=1}^N (\partial_\mu \sigma^l)(r, X_r^i, \bar{\mu}_r^N)(X_r^k) D_s^j X_r^k\Big|^2 dr\Big].
$$

The term Φ_N is a sequence which converges to zero as $N \to \infty$. This is justified by repeated use of the Cauchy-Schwarz and Jensen inequalities, over the bounds (29) and (31); the $L^2(\Omega)$ boundedness of θ_r and η_r is justified by the Lipschitz property of σ^l and b (i.e. $\nabla_x \sigma^l$ and $\nabla_x b$ are uniformly bounded maps). The remaining term on the left hand side of (32) converges to zero by the conclusion of [Nua06, Lemma 1.2.3] (shown in Step 3 of this proof). Hence so must the right hand side of (32) and we obtain that

$$
\lim_{N\to\infty} \int_s^t (\nabla_x b)(r, X_r^i, \bar{\mu}_r^N) D_s^j X_r^i \, dr = \int_s^t (\nabla_x b)(r, Z_r^i, \mu_r^i) D_s^j Z_r^i \, dr,
$$

and

$$
\lim_{N\to\infty} \int_s^t (\nabla_x \sigma^l)(r, X_r^i, \bar{\mu}_r^N) D_s^j X_r^i \, dW_r^i = \int_s^t (\nabla_x \sigma^l)(r, Z_r^i, \mu_r^i) D_s^j Z_r^i dW_r^{l,i}.
$$

in the L^2-sense and the it follows that MV-SDE (17) is identified as the limit of (18).

$\square$

Remark 4.9 (Mollification: Lifting Assumption 4.7(2.); the measure differentiability requirement)**.** *We do not carry out the analysis here, but by drawing on techniques of mollification in Wasserstein spaces it seems possible*

to remove the measure differentiability assumption of Proposition 4.8. There are several ways to carry out mollification over the space of measures, with the main difficulty being that the Wasserstein space of measures is an infinite dimensional one.

[Inf-Sup convolution]: Motivated by the study of Hamilton-Jacobi equations in infinite dimensional spaces, Lasry-Lions [LL86] propose inf-sup convolution (in Hilbert spaces) to show that bounded uniformly continuous scalar functions defined on a Hilbert space $\mathcal{H}$ can be uniformly approximated by functions belonging to the class of differentiable maps with Lipschitz derivatives. Critically, their methodology preserves, in the infinite dimensional space, certain good properties that other known mollifications at the time were not known to. As an example, in [DDJ23] and working on the torus, the the inf-sup convolution techniques of [LL86] are used to establish optimal rates for the convergence problem in mean field control. Notably, the mollification procedures of [DDJ23] rely on the properties of the Hilbert Sobolev space H^{-s} for $s > d/2 + 1$, i.e., the dual of the Hilbert space of functions with s generalized derivatives in L^2.

[Smoothing via truncating Fourier expansions]: In [CD22], working over the probability measure $\mathcal{P}(\mathbb{T}^d)$ on the d-dimensional torus $\mathbb{T}^d$, a mollification procedure of real-valued functions on $\mathcal{P}(\mathbb{T}^d)$ based on the Fourier coefficients of the measure is introduced. Very roughly, the mollification is carried out by truncating higher-order coefficients of the Fourier expansion at a conveniently chosen threshold and combining with a Fejér kernel (see their Definition 3.13). It is currently open how to extend the analysis from $\mathbb{T}^d$ to $\mathbb{R}^d$.

[Smooth approximation over L^∞ under W_1]: In [CZ21, Section 3], the authors construct a novel mollification technique for measure functionals in $C(\mathcal{P}_1(\mathbb{R}^d))$ under the 1-Wasserstein distance and the approximation is carried in L^∞ (the subset of bounded RVs with norm $\|X\|_{L^\infty(\Omega)} = \text{ess sup}_{\omega \in \Omega} |X(\omega)|$). Critically, their mollified map satisfies uniformly the same Lipschitz property of the original functional. Such property does not hold under W_2 (unless the functional is also W_1-Lipschitz) but a convergence result is provided (see their Theorem 3.1).

[Smoothing the approximating particle projection]: The interacting particle system point of view allows us to benefit from a finite dimensional framework and draw from standard mollification arguments (in $(\mathbb{R}^d)^N$). The most suitable regularisation method for our IPS approach has been shown in [CCD22] (also [CD18a, CD18b] and the complementary [dRP23] adding all the missing details of the proofs of [CD18a, CD18b]). There, regularization

is applied directly to the empirical projection map u^N (of Definition 4.3 and not u) via convolution with a smooth kernel. The method allows to control the derivatives of the mollified map by the Lipschitz constant.

In [CM23], the authors close the two open questions left by [CZ21]. They show that when u is W_2-continuous there exists a sequence $\{u_k\}_k \in C^\infty(\mathscr{P}_2(\mathbb{R}^d))$ converging to u uniformly on compact subsets of $\mathscr{P}_2(\mathbb{R}^d)$. If u is additionally uniformly (resp. Lipschitz) continuous then each u_k is also uniformly (resp. Lipschitz) continuous, with the same modulus of continuity (resp. Lipschitz constant) as u – this closes [CZ21, Remark 3.2(i)]. Moreover, for $u \in C^1(\mathscr{P}_2(\mathbb{R}^d))$ (resp. $u \in C^2(\mathscr{P}_2(\mathbb{R}^d))$) we show that the convergence also holds for the first-order derivative (resp. second-order derivatives) – this closes [CZ21, Remark 3.2(ii)]. The smooth approximating sequence $\{u_k\}_k \in C^\infty(\mathscr{P}_2(\mathbb{R}^d))$ of [CM23] is constructed relying on the empirical distribution, similarly to what is done in [CD18a, Theorem 5.92] in the proof of Itô's formula along a flow of measures, although there it is not really a smoothing as both u and u_k are of class $C^2(\mathscr{P}_2(\mathbb{R}^d))$, but rather a way to approximate a function u on $\mathscr{P}_2(\mathbb{R}^d)$ by functions defined on finite-dimensional spaces.

Acknowledgment

References

[AdRR+22] D. Adams, G. dos Reis, R. Ravaille, W. Salkeld, and J. Tugaut, *Large deviations and exit-times for reflected McKean-Vlasov equations with self-stabilising terms and superlinear drifts*, Stochastic Process. Appl. **146** (2022), 264–310. MR4374937

[BCC11] F. Bolley, J. A. Cañizo, and J. A. Carrillo, *Stochastic mean-field limit: non-Lipschitz forces and swarming*, Math Models Methods Appl Sci **21** (2011).

[BFFT12] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul, *Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons*, The Journal of Mathematical Neuroscience **2** (2012May), no. 1, 10.

[BFT15] M. Bossy, O. Faugeras, and D. Talay, *Clarification and complement to "mean-field description and propagation of chaos in networks of Hodgkin–Huxley and Fitzhugh–Nagumo neurons"*, The Journal of Mathematical Neuroscience (JMN) **5** (2015), no. 1, 19.

[Car16] R. Carmona, *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*, Financial Mathematics, vol. 1, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016. MR3629171

[CCD22] J.-F. Chassagneux, D. Crisan, and F. Delarue, *A probabilistic approach to classical solutions of the master equation for large population equilibria*, Mem. Amer. Math. Soc. **280** (2022), no. 1379, v+123. MR4493576

[CD18a] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications I*, Probability Theory and Stochastic Modelling, vol. 83, Springer-Verlag, 2018.

[CD18b] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications II*, Probability Theory and Stochastic Modelling, vol. 84, Springer-Verlag, 2018.

[CD22] A. Cecchin and F. Delarue, *Weak solutions to the master equation of potential mean field games*, arXiv preprint arXiv:2204.04315 (2022).

[CdR22] X. Chen and G. dos Reis, *A flexible split-step scheme for solving McKean-Vlasov stochastic differential equations*, Appl. Math. Comput. **427** (2022), Paper No. 127180, 23. MR4413221

[CdR24] X. Chen and G. dos Reis, *Euler simulation of interacting particle systems and McKean-Vlasov SDEs with fully super-linear growth drifts in space and interaction*, IMA J. Numer. Anal. **44** (2024), no. 2, 751–796. MR4727110

[CdS25] X. Chen, G. dos Reis, and W. Stockinger, *Wellposedness, exponential ergodicity and numerical approximation of fully super-linear McKean–Vlasov SDEs and associated particle systems*, Electronic Journal of Probability **30** (2025), 1–50.

[CdRSW24] X. Chen, G. dos Reis, W. Stockinger, and Z. Wilde, *Improved weak convergence for the long time simulation of mean-field Langevin equations*, arXiv preprint arXiv:2405.01346 (2024). To appear in Electr. J. of Prob.

[CM18] D. Crisan and E. McMurray, *Smoothing properties of McKean-Vlasov SDEs*, Probab. Theory Related Fields **171** (2018), no. 1-2, 97–148. MR3800831

[CM23] A. Cosso and M. Martini, *On smooth approximations in the Wasserstein space*, Electron. Commun. Probab. **28** (2023), Paper No. 30, 11. MR4627412

[CZ21] M. Chenchen and J. Zhang, *Wellposedness of second order master equations for mean field games with nonsmooth data*, Memoirs of the American Mathematical Society (2021).

[DDJ23] S. Daudin, F. Delarue, and J. Jackson, *On the optimal rate for the convergence problem in mean field control*, arXiv preprint arXiv:2305.08423 (2023).

[DFG⁺16] W. Dreyer, P. K. Friz, P. Gajewski, C. Guhlke, and M. Maurelli, *Stochastic model for LFP-electrodes*, WIAS preprint no. 2329, 2016.

[DGG⁺11] W. Dreyer, M. Gaberšček, C. Guhlke, R. Huth, and J. Jamnik, *Phase transition in a rechargeable lithium battery*, European Journal of Applied Mathematics **22** (2011), no. 3, 267–290.

[dRES22] G. dos Reis, S. Engelhardt, and G. Smith, *Simulation of McKean-Vlasov SDEs with super-linear growth*, IMA J. Numer. Anal. **42** (2022), no. 1, 874–922. MR4367675

[dRP23] G. dos Reis and V. Platonov, *Itô-Wentzell-Lions Formula for Measure Dependent Random Fields under Full and Conditional Measure Flows*, Potential Anal. **59** (2023), no. 3, 1313–1344. MR4647952

[dRST19] G. dos Reis, W. Salkeld, and J. Tugaut, *Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law*, Ann. Appl. Probab. **29** (2019), no. 3, 1487–1540. MR3914550

[GK80] I. Gyöngy and N. V. Krylov, *On stochastic equations with respect to semimartingales. I*, Stochastics **4** (1980/81), no. 1, 1–21. MR587426

[GPV20] S. N. Gomes, G. A. Pavliotis, and U. Vaes, *Mean field limits for interacting diffusions with colored noise: phase transitions and spectral numerical methods*, Multiscale Modeling & Simulation **18** (2020), no. 3, 1343–1370.

[HAHT21] A.-L. Haji-Ali, H. Hoel, and R. Tempone, *A simple approach to proving the existence, uniqueness, and strong and weak convergence rates for a broad class of mckean–vlasov equations*, arXiv preprint arXiv:2101.00886 (2021).

[IdRS19] P. Imkeller, G. dos Reis, and W. Salkeld, *Differentiability of SDEs with drifts of super-linear growth*, Electron. J. Probab. **24** (2019), Paper No. 3, 43. MR3916323

[LL86] J.-M. Lasry and P.-L. Lions, *A remark on regularization in Hilbert spaces*, Israel J. Math. **55** (1986), no. 3, 257–266. MR876394

[Mao08] X. Mao, *Stochastic differential equations and applications*, Horwood, 2008.

[Nua06] D. Nualart, *The Malliavin calculus and related topics*, Vol. 1995, Springer, 2006.

[RW19] P. Ren and F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications*, J. Differential Equations **267** (2019), no. 8, 4745–4777. MR3983053

[Szn91] A.-S. Sznitman, *Topics in Propagation of Chaos*, École d'été de probabilités de saint-flour xix - 1989, 1991.