Bismut Formula for Lions Derivative of Distribution-Path Dependent SDEs *

Jianhai Bao[a)], **Panpan Ren**[b)], **Feng-Yu Wang**[a),b)]

[a)] Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

[b)] Department of Mathematics, Swansea University, Fabian Way, Skewen, SA1 8EN, UK

jianhaibao13@gmail.com, 673788@swansea.ac.uk, wangfy@tju.edu.cn

November 24, 2021

Abstract

To characterize the regularity of distribution-path dependent SDEs in initial distributions variable as probability measures on the path space, we introduce the intrinsic and Lions derivatives in the space of probability measures on Banach spaces, and prove the chain rule for the Lions derivative in the distribution of Banach-valued random variables. By using Malliavin calculus, we establish the Bismut type formula for the Lions derivatives of functional solutions to SDEs with distribution-path dependent drifts. When the noise term is also path dependent so that the Bismut formula is invalid, we establish the asymptotic Bismut formula. Both non-degenerate and degenerate noises are considered. The main results of this paper generalize and improve the corresponding ones derived recently in the literature for the classical SDEs with memory and McKean-Vlasov SDEs without memory.

AMS subject Classification: 60J60, 58J65.
Keywords: McKean-Vlasov SDEs with memory, Bismut formula, asymptotic Bismut formula, Malliavin calculus, Lions derivative

1 Introduction

To characterize stochastic systems with evolutions affected by both history and micro environment, the distribution-path dependent SDEs have been considered in [20, 28], where the Harnack type inequalities, ergodicity and long time large deviation principles are investigated. This type SDEs generalize the McKean-Vlasov (distribution dependent or mean-field) SDEs and path dependent (functional) SDEs (or SDEs with memory). Both have been studied intensively in the literature; see, for instance, the monograhs [6, 9] and references within.

On the other hand, as a powerful tool in the study of regularity for diffusion processes, a derivative formula on diffusion semigroups was established first by Bismut in [7] using Malliavin calculus, and then by Elworthy-Li in [12] using a martingale argument. Hence, this type derivative

*Supported in part by NNSFC (11771326, 11831014, 11921001).

formula is named as Bsimut formula or Bismut-Elworthy-Li formula in the literature. Moreover, a new coupling method (called coupling by change of measures) was introduced to establish derivative formulas and Harnack inequalities for SDEs and SPDEs; see, for example, [33] and references therein. Due to its wide applications, the Bismut type formulas have been investigated for different models; see, for instance, [10, 25, 30, 31, 38, 40] for SDEs/SPDEs driven by jump processes, [16, 17, 24, 35, 34, 37, 39] for hypoelliptic diffusion semigroups, and [2, 14, 15] for SDEs with fractional noises.

Recently, the Bismut type formulas have been established in [4] for the Gâteaux derivative of functional solutions to path dependent SDEs, in [26] for the Lions derivative of solutions to McKean-Vlasov SDEs. See also [3, 11] for the study of derivative in the initial points for McKean-Vlasov SDEs, and Lions derivative for solutions to the de-coupled SDEs (which do not depend on the distribution of its own solution) associated with McKean-Vlasov SDEs. In these references, the noise term is distribution-path independent. However, when the noise term is path dependent, the distribution of the solution is no longer differentiable in the initial value, so that the Bismut type formula is invalid. In this case, a weaker derivative formula, called asymptotic Bismut formula, has been established in [22].

The aim of this paper is to establish (asymptotic) derivative formulas for the Lions derivative w.r.t. the initial distribution of distribution-path dependent SDEs, such that results derived in [4, 22, 26] are generalized and improved. To this end, we will identify the Lions derivative of the solution in initial distributions by the Malliavin derivative along an adapted direction, so that the integration by parts formula for the Malliavin derivative gives rise to a formula of the Lions derivative.

Since the functional solution of a path-distribution dependent SDE takes values in the path space $C([-r_0,0];\mathbb{R}^d)$, where $r_0>0$ is the length of memory, to investigate the regularities of the solution in initial distributions, we need to introduce and study derivatives in probability measures on the path space, which is new in the literature.

For a fixed number $r_0>0$, the path space $\mathscr{C}:=C([-r_0,0];\mathbb{R}^d)$ is a separable Banach space under the uniform norm

$$\|\xi\|_{\mathscr{C}}:=\sup_{-r_0\le\theta\le0}|\xi(\theta)|,\quad \xi\in\mathscr{C}.$$

For $t\ge0$ and $f\in C([-r_0,\infty);\mathbb{R}^d)$, the $\mathscr{C}$-valued function $(f_t)_{t\ge0}$ defined by

$$f_t(\theta)=f(t+\theta),\quad \theta\in[-r_0,0]$$

is called the segment (or window) process of $(f(t))_{t\ge-r_0}$. Let $\mathscr{L}_\xi$ stand for the distribution of a random variable ξ. When different probability measures are concerned, we also denote $\mathscr{L}_\xi$ by $\mathscr{L}_\xi|\mathbb{P}$ to emphasize the reference probability measure $\mathbb{P}$. Let $\mathscr{P}(\mathscr{C})$ be the collection of all probability measures on $\mathscr{C}$ and, for $p>0$, $\mathscr{P}_p(\mathscr{C})$ the set of probability measures on $\mathscr{C}$ with finite p-th moment, i.e.,

$$\mathscr{P}_p(\mathscr{C})=\big\{\mu\in\mathscr{P}(\mathscr{C}):\|\mu\|_p:=\{\mu(\|\cdot\|_{\mathscr{C}}^p)\}^{\frac1p}<\infty\big\},$$

where $\mu(f):=\int f\mathrm{d}\mu$ for a measurable function f. Then $\mathscr{P}_p(\mathscr{C})$ is a Polish space under the $\mathbb{W}_p$-Wasserstein distance defined by

$$\mathbb{W}_p(\mu,\nu)=\inf_{\pi\in\mathcal{C}(\mu,\nu)}\left(\int_{\mathscr{C}\times\mathscr{C}}\|\xi-\eta\|_{\mathscr{C}}^p\pi(\mathrm{d}\xi,\mathrm{d}\eta)\right)^{\frac{1}{p\vee1}},\quad \mu,\nu\in\mathscr{P}_p(\mathscr{C}),\quad p>0,$$

where $\mathcal{C}(\mu,\nu)$ is the set of all couplings of μ and ν.

Consider the following McKean-Vlasov SDE with memory (also called distribution-path dependent SDE):

E1
$$\mathrm{d}X(t) = b(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\mathrm{d}W(t), \quad t \geq 0, \tag{1.1}$$

where $(W(t))_{t\geq 0}$ is an m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, and

$$b : [0,\infty) \times \mathscr{C} \times \mathscr{P}(\mathscr{C}) \to \mathbb{R}^d, \quad \sigma : [0,\infty) \times \mathscr{C} \times \mathscr{P}(\mathscr{C}) \to \mathbb{R}^d \otimes \mathbb{R}^m$$

are measurable satisfying the following assumption.

(A) Let $p \in [1,\infty)$.

(A_1) b and σ are bounded on bounded subsets of $[0,\infty) \times \mathscr{C} \times \mathscr{P}_p(\mathscr{C})$.

(A_2) For any $T > 0$, there is a constant $K \geq 0$ such that

$$\begin{aligned}&2\langle \xi(0) - \eta(0), b(t,\xi,\mu) - b(t,\eta,\nu)\rangle^+ + \|\sigma(t,\xi,\mu) - \sigma(t,\eta,\nu)\|_{\mathrm{HS}}^2 \\ &\leq K\big\{\|\xi - \eta\|_{\mathscr{C}}^2 + W_p(\mu,\nu)^2\big\}, \quad \xi,\eta \in \mathscr{C}, \mu,\nu \in \mathscr{P}_p(\mathscr{C}), t \in [0,T].\end{aligned}$$

(A_3) When $p \in [1,2]$, $\sigma(t,\xi,\mu) = \sigma(t,\xi(0))$ depends only on t and $\xi(0)$.

For any $\mathscr{F}_0$-measurable random variable $X_0 \in \mathscr{C}$, an adapted continuous process $(X(t))_{t\geq 0}$ is called a solution with the initial value X_0, if $\mathbb{P}$-a.s.

$$X(t) = X(0) + \int_0^t b(s, X_s, \mathscr{L}_{X_s})\mathrm{d}s + \int_0^t \sigma(s, X_s, \mathscr{L}_{X_s})\mathrm{d}W(s), \quad t \geq 0,$$

where the segment process $(X_t)_{t\geq 0}$ associated with the solution process

$$X(t) := X(t)1_{(0,\infty)}(t) + X_0(t)1_{[-r_0,0]}(t), \quad t \geq -r_0$$

is called a functional solution to (1.1).

According to Lemma 3.1 below, under the assumption **(A)**, for any $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$, (1.1) has a unique functional solution $(X_t)_{t\geq 0}$ satisfying

$$\mathbb{E}\Big(\sup_{0\leq s\leq t} \|X_t\|_{\mathscr{C}}^p\Big) < \infty, \quad t > 0.$$

To emphasize the initial distribution, we denote the functional solution by X_t^μ if $\mathscr{L}_{X_0} = \mu$. In this paper, we aim to investigate the Lions derivative of the functional $\mu \mapsto (P_t f)(\mu)$, where

SM
$$(P_t f)(\mu) := \mathbb{E} f(X_t^\mu), \quad t > 0, f \in \mathscr{B}_b(\mathscr{C}). \tag{1.2}$$

This refers to the regularity of the law $\mathscr{L}_{X_t^\mu}$ w.r.t. the initial distribution μ. Due to the weak uniqueness ensured by Lemma 3.1 below, $(P_t f)(\mu)$ is a function of μ; i.e., if $\widetilde{X}_t^\mu$ also solves (1.1) with $\mathscr{L}_{\widetilde{X}_0^\mu} = \mu$, then $\mathscr{L}_{X_t^\mu} = \mathscr{L}_{\widetilde{X}_t^\mu}$ so that the definition of $(P_t f)(\mu)$ does not depend on the choice of solutions.

The remainder of this paper is organized as follows. Since $\mathscr{C}$ is a Banach space, in Section 2 we extend the notion of Lions derivative introduced in [8] for functions on the Wasserstein space $\mathscr{P}_2(\mathbb{R}^d)$ to probability measures on Banach spaces, and establish a derivative formula in the distribution of Banach-valued random variables, where $\mathscr{P}_2(\mathbb{R}^d)$ is the set of all probability measures on $\mathbb{R}^d$ with finite second-order moment. In Section 3, under **(A)** we prove the well-posedness of solutions to (1.1), which generalizes the corresponding results derived in [20] for $p = 2$ and in [28] for Lipschitz continuous $b(t, \cdot)$. In Sections 4 and 5 we calculate the Malliavin derivative of X_t^μ w.r.t the Brownian motion $W(t)$, and the Lions derivative of X_t^μ in the initial distribution μ, respectively. Finally, in Sections 6 and 7 we establish respectively the Bismut type formula for the Lions derivative of $(P_tf)(\mu)$ in μ when $\sigma(t, \xi, \mu) = \sigma(t, \xi(0))$ depends only on t and $\xi(0)$, and the asymptotic Bismut formula for the Lions derivative of $(P_tf)(\mu)$ in μ in case of $\sigma(t, \xi, \mu) = \sigma(t, \xi)$ (i.e., the diffusion term is path-dependent but independent of measure argument μ).

We would like to emphasize that even come back to Mckean-Vlasov SDEs without memory studied in [26] and the classical SDEs with memory considered in [22], our main conditions are weaker since the drift term can be non-Lipschitz continuous.

2 Derivatives in probability measures on a separable Banach space

In this part, we introduce the intrinsic and Lions derivatives for probability measures on a separable Banach space, and establish the chain rule for the distribution of Banach-valued random variables. These will be used to establish the (asymptotic) Bismut type formulas for the intrinsic and Lions derivatives of $(P_tf)(\mu)$.

The intrinsic derivative was first introduced in [1] on the configuration space over Riemannian manifolds, while the Lions derivative (denoted by L-derivative in the literature) was developed on the Wasserstein space $\mathscr{P}_2(\mathbb{R}^d)$ from Lions' lectures [8] on mean-field games. The relation between them has been clarified in the recent paper [27], where the latter is a stronger notion than the former and they coincide if both exist.

Let $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ be a separable Banach space, and let $(\mathbb{B}^*, \|\cdot\|_{\mathbb{B}^*})$ be its dual space. For any $p \in [1, \infty)$, denote $p^* = \frac{p}{p-1}$ when $p > 1$ and $p^* = \infty$ as $p = 1$. Let $\mathscr{P}(\mathbb{B})$ be the class of all probability measures on $\mathbb{B}$ equipped with the weak topology. Then

$$\mathscr{P}_p := \big\{\mu \in \mathscr{P}(\mathbb{B}) : \|\mu\|_p := \{\mu(\|\cdot\|_{\mathbb{B}}^p)\}^{\frac{1}{p}} < \infty\big\}$$

is a Polish space under the L^p-Wasserstein distance

$$\mathbb{W}_p(\mu_1, \mu_2) := \inf_{\pi \in \mathcal{C}(\mu_1,\mu_2)} \left(\int_{\mathbb{B}\times\mathbb{B}} \|x - y\|_{\mathbb{B}}^p \pi(\mathrm{d}x, \mathrm{d}y)\right)^{\frac{1}{p}},$$

where $\mathcal{C}(\mu_1, \mu_2)$ is the set of all couplings of μ_1 and μ_2.

For any $\mu \in \mathscr{P}_p(\mathbb{B})$, the tangent space at μ is given by

$$T_{\mu,p} = L^p(\mathbb{B} \to \mathbb{B}; \mu) := \big\{\phi : \mathbb{B} \to \mathbb{B} \text{ is measurable with } \mu(\|\phi\|_{\mathbb{B}}^p) < \infty\big\},$$

which is a Banach space under the norm $\|\phi\|_{T_{\mu,p}} := \{\mu(\|\phi\|_{\mathbb{B}}^p)\}^{\frac{1}{p}}$, and its dual space is

$$T_{\mu,p}^* = L^{p^*}(\mathbb{B} \to \mathbb{B}^*; \mu) := \big\{\psi : \mathbb{B} \to \mathbb{B}^* \text{ is measurable with } \|\psi\|_{T_{\mu,p}^*} := \big\|\|\psi\|_{\mathbb{B}^*}\big\|_{L^{p^*}(\mu)} < \infty\big\}.$$

Definition 2.1. Let $f : \mathscr{P}_p(\mathbb{B}) \to \mathbb{R}$ be a continuous function for some $p \in [1, \infty)$, and let Id be the identity map on $\mathbb{B}$.

(1) f is called intrinsically differentiable at a point $\mu \in \mathscr{P}_p(\mathbb{B})$, if

$$T_{\mu,p} \ni \phi \mapsto D^L_\phi f(\mu) := \lim_{\varepsilon \downarrow 0} \frac{f(\mu \circ (\mathrm{Id} + \varepsilon\phi)^{-1}) - f(\mu)}{\varepsilon} \in \mathbb{R}$$

is a well-defined bounded linear functional. In this case, the unique element $D^L f(\mu) \in T^*_{\mu,p}$ satisfying

$${}_{T^*_{\mu,p}}\langle D^L f(\mu), \phi \rangle_{T_{\mu,p}} := \int_{\mathbb{B}} {}_{\mathbb{B}^*}\langle D^L f(\mu)(x), \phi(x) \rangle_{\mathbb{B}} \mu(\mathrm{d}x) = D^L_\phi f(\mu), \quad \phi \in T_{\mu,p}$$

is called the intrinsic derivative of f at μ.

If moreover

$$\lim_{\|\phi\|_{T_{\mu,p}} \downarrow 0} \frac{|f(\mu \circ (\mathrm{Id} + \phi)^{-1}) - f(\mu) - D^L_\phi f(\mu)|}{\|\phi\|_{T_{\mu,p}}} = 0,$$

f is called L-differentiable at μ with the L-derivative (i.e., Lions derivative) $D^L f(\mu)$.

(2) We write $f \in C^1(\mathscr{P}_p(\mathbb{B}))$ if f is L-differentiable at any point $\mu \in \mathscr{P}_p(\mathbb{B})$, and the L-derivative has a version $D^L f(\mu)(x)$ jointly continuous in $(x, \mu) \in \mathbb{B} \times \mathscr{P}_p(\mathbb{B})$. If moreover $D^L f(\mu)(x)$ is bounded, we denote $f \in C^1_b(\mathscr{P}_p(\mathbb{B}))$.

TO **Theorem 2.1.** *Let $f : \mathscr{P}_p(\mathbb{B}) \to \mathbb{R}$ be continuous for some $p \in [1, \infty)$, and let $(\xi_\varepsilon)_{\varepsilon \in [0,1]}$ be a family of $\mathbb{B}$-valued random variables on a Polish complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\dot{\xi}_0 := \lim_{\varepsilon \downarrow 0} \frac{\xi_\varepsilon - \xi_0}{\varepsilon}$ exists in $L^p(\Omega)$.*

(1) *Let $\mu_0 = \mathscr{L}_{\xi_0}$ be atomless. If f is L-differentiable such that $D^L f(\mu_0)$ has a continuous version satisfying*

GRW
$$\|D^L f(\mu_0)(x)\|_{\mathbb{B}^*} \le C(1 + \|x\|_{\mathbb{B}}^{p/p^*} 1_{\{p>1\}}), \quad x \in \mathbb{B} \tag{2.1}$$

for some constant $C > 0$, then

CR
$$\lim_{\varepsilon \downarrow 0} \frac{f(\mathscr{L}_{\xi_\varepsilon}) - f(\mathscr{L}_{\xi_0})}{\varepsilon} = \mathbb{E}[{}_{\mathbb{B}^*}\langle D^L f(\mu_0)(\xi_0), \dot{\xi}_0 \rangle_{\mathbb{B}}]. \tag{2.2}$$

(2) *If f is L-differentiable in a neighbourhood O of μ_0 such that $D^L f$ has a version jointly continuous in $(x, \mu) \in \mathbb{B} \times O$ satisfying*

GRW'
$$\|D^L f(\mu)(x)\|_{\mathbb{B}^*} \le C(1 + \|x\|_{\mathbb{B}}^{p/p^*} 1_{\{p>1\}}), \quad (x, \mu) \in \mathbb{B} \times O \tag{2.3}$$

for some constant $C > 0$, then (2.2) *holds.*

To prove this result, we need the following lemma similar to [18, Lemma A.2] for the special case that $\mathscr{P}_p(\mathscr{B}) = \mathscr{P}_2(\mathbb{R}^d)$ (i.e., $p = 2$ and $\mathbb{B} = \mathbb{R}^d$).

LN1 **Lemma 2.2.** *Let $\{(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)\}_{i=1,2}$ be two atomless, Polish complete probability spaces, and let X_i be $\mathbb{B}$-valued random variables on these two probability spaces respectively such that $\mathscr{L}_{X_1}|\mathbb{P}_1 = \mathscr{L}_{X_2}|\mathbb{P}_2$. Then for any $\varepsilon > 0$, there exist measurable maps*

$$\tau : \Omega_1 \to \Omega_2, \quad \tau^{-1} : \Omega_2 \to \Omega_1$$

such that

$$\begin{aligned}
&\mathbb{P}_1(\tau^{-1} \circ \tau = \mathrm{Id}_{\Omega_1}) = \mathbb{P}_2(\tau \circ \tau^{-1} = \mathrm{Id}_{\Omega_2}) = 1,\\
&\mathbb{P}_1 = \mathbb{P}_2 \circ \tau, \quad \mathbb{P}_2 = \mathbb{P}_1 \circ \tau^{-1},\\
&\|X_1 - X_2 \circ \tau\|_{L^\infty(\mathbb{P}_1)} + \|X_2 - X_1 \circ \tau^{-1}\|_{L^\infty(\mathbb{P}_2)} \le \varepsilon,
\end{aligned}$$

where Id_{Ω_i} stands for the identity map on $\Omega_i, i = 1, 2$.

Proof. Since $\mathbb{B}$ is separable, there is a measurable partition $(A_n)_{n\ge1}$ of $\mathbb{B}$ such that $\mathrm{diam}(A_n) < \varepsilon$, $n \ge 1$. Let $A_n^i = \{X_i \in A_n\}, n \ge 1, i = 1, 2$. Then $(A_n^i)_{n\ge1}$ forms a measurable partition of Ω_i so that $\sum_{n\ge1} A_n^i = \Omega_i, i = 1, 2$, and, due to $\mathscr{L}_{X_1}|\mathbb{P}_1 = \mathscr{L}_{X_2}|\mathbb{P}_2$,

$$\mathbb{P}_1(A_n^1) = \mathbb{P}_2(A_n^2), \quad n \ge 1.$$

Since the probabilities $(\mathbb{P}_i)_{i=1,2}$ are atomless, according to [19, Theorem C in Section 41], for any $n \ge 1$ there exist measurable sets $\widetilde{A}_n^i \subset A_n^i$ with $\mathbb{P}_i(A_n^i \setminus \widetilde{A}_n^i) = 0, i = 1, 2$, and a measurable bijective map

$$\tau_n : \widetilde{A}_n^1 \to \widetilde{A}_n^2$$

such that

$$\mathbb{P}_1|_{\widetilde{A}_n^1} = \mathbb{P}_2 \circ \tau_n|_{\widetilde{A}_n^1}, \quad \mathbb{P}_2|_{\widetilde{A}_n^2} = \mathbb{P}_1 \circ \tau_n^{-1}|_{\widetilde{A}_n^2}.$$

By $\mathrm{diam}(A_n) < \varepsilon$ and $\mathbb{P}_i(A_n^i \setminus \widetilde{A}_n^i) = 0$, we have

$$\|(X_1 - X_2 \circ \tau_n)1_{\widetilde{A}_n^1}\|_{L^\infty(\mathbb{P}_1)} \vee \|(X_2 - X_1 \circ \tau_n^{-1})1_{\widetilde{A}_n^2}\|_{L^\infty(\mathbb{P}_2)} \le \varepsilon.$$

Then the proof is finished by taking, for fixed points $\hat{\omega}_i \in \Omega_i, i = 1, 2$,

$$\tau(\omega_1) := \begin{cases} \tau_n(\omega_1), & \text{if } \omega_1 \in \widetilde{A}_n^1 \text{ for some } n \ge 1,\\ \hat{\omega}_2, & \text{otherwise,} \end{cases}$$

$$\tau^{-1}(\omega_2) := \begin{cases} \tau_n^{-1}(\omega_2), & \text{if } \omega_2 \in \widetilde{A}_n^2 \text{ for some } n \ge 1,\\ \hat{\omega}_1, & \text{otherwise.} \end{cases}$$

□

Proof of Theorem 2.1. Without loss of generality, we may and do assume that $\mathbb{P}$ is atomless. Otherwise, by taking

$$(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}}) := (\Omega \times [0,1], \mathscr{F} \times \mathscr{B}([0,1]), \mathbb{P} \times \mathrm{d}s), \quad (\widetilde{\xi}_\varepsilon)(\omega, s) := \xi_\varepsilon(\omega) \text{ for } (\omega, s) \in \widetilde{\Omega},$$

where $\mathscr{B}([0,1])$ is the completion of the Borel σ-algebra on $[0,1]$ w.r.t. the Lebesgue measure $\mathrm{d}s$, we have

$$\mathscr{L}_{\widetilde{\xi}_\varepsilon}|\widetilde{\mathbb{P}} = \mathscr{L}_{\xi_\varepsilon}|\mathbb{P}, \quad \mathbb{E}[_{\mathbb{B}^*}\langle D^L f(\mu_0)(\xi_0), \dot{\xi}_0\rangle_{\mathbb{B}}] = \widetilde{\mathbb{E}}[_{\mathbb{B}^*}\langle D^L f(\mu_0)(\widetilde{\xi}_0), \dot{\widetilde{\xi}}_0\rangle_{\mathbb{B}}].$$

In this way we go back to the atomless situation.

(1) Let $\mathscr{L}_{\xi_0} = \mu_0 \in \mathscr{P}_p(\mathbb{B})$ be atomless. In this case, $(\mathbb{B}, \mathscr{B}(\mathbb{B}), \mu_0)$ is an atomless Polish complete probability space, where $\mathscr{B}(\mathbb{B})$ is the μ_0-complete Borel σ-algebra of $\mathbb{B}$. By Lemma 2.2, for any $n \geq 1$ we find measurable maps

$$\tau_n : \Omega \to \mathbb{B}, \quad \tau_n^{-1} : \mathbb{B} \to \Omega$$

such that

AB3
$$\begin{aligned} &\mathbb{P}(\tau_n^{-1} \circ \tau_n = \mathrm{Id}_\Omega) = \mu_0(\tau_n \circ \tau_n^{-1} = \mathrm{Id}) = 1, \\ &\mathbb{P} = \mu_0 \circ \tau_n, \quad \mu_0 = \mathbb{P} \circ \tau_n^{-1}, \\ &\|\xi_0 - \tau_n\|_{L^\infty(\mathbb{P})} + \|\mathrm{Id} - \xi_0 \circ \tau_n^{-1}\|_{L^\infty(\mu_0)} \leq \frac{1}{n}, \end{aligned} \tag{2.4}$$

where $\mathrm{Id} = \mathrm{Id}_{\mathbb{B}}$ is the identity map on $\mathbb{B}$.

Since f is L-differentiable at μ_0, there exists a decreasing function $h : [0,1] \to [0, \infty)$ with $h(r) \downarrow 0$ as $r \downarrow 0$ such that

AB4
$$\sup_{\|\phi\|_{L^p(\mu_0)} \leq r} \big| f(\mu_0 \circ (\mathrm{Id} + \phi)^{-1}) - f(\mu_0) - D_\phi^L f(\mu_0) \big| \leq r h(r), \quad r \in [0,1]. \tag{2.5}$$

Since $\mathscr{L}_{\xi_\varepsilon - \xi_0} \in \mathscr{P}_p(\mathbb{B})$, by (2.4) we have

AB5
$$\phi_{n,\varepsilon} := (\xi_\varepsilon - \xi_0) \circ \tau_n^{-1} \in T_{\mu,p}, \quad \|\phi_{n,\varepsilon}\|_{T_{\mu,p}} = \|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})}. \tag{2.6}$$

Next, (2.4) implies

ABB0
$$\mathscr{L}_{\tau_n + \xi_\varepsilon - \xi_0} = \mathbb{P} \circ (\tau_n + \xi_\varepsilon - \xi_0)^{-1} = (\mu_0 \circ \tau_n) \circ (\tau_n + \xi_\varepsilon - \xi_0)^{-1} = \mu_0 \circ (\mathrm{Id} + \phi_{n,\varepsilon})^{-1}. \tag{2.7}$$

Moreover, by $\frac{\xi_\varepsilon - \xi_0}{\varepsilon} \to \dot{\xi}_0$ in $L^p(\mathbb{P})$ as $\varepsilon \downarrow 0$, we find a constant $c \geq 1$ such that

ABB
$$\|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})} \leq c\varepsilon, \quad \varepsilon \in [0,1]. \tag{2.8}$$

Combining (2.4)-(2.8) leads to

AB6
$$\begin{aligned} &\big| f(\mathscr{L}_{\tau_n + \xi_\varepsilon - \xi_0}) - f(\mathscr{L}_{\xi_0}) - \mathbb{E}[{}_{\mathbb{B}^*}\langle (D^L f)(\mu_0)(\tau_n), (\xi_\varepsilon - \xi_0) \rangle_{\mathbb{B}}] \big| \\ &= \big| f(\mu_0 \circ (\mathrm{Id} + \phi_{n,\varepsilon})^{-1}) - f(\mu_0) - D_{\phi_{n,\varepsilon}}^L f(\mu_0) \big| \\ &\leq \|\phi_{n,\varepsilon}\|_{T_{\mu,p}} h(\|\phi_{n,\varepsilon}\|_{T_{\mu,p}}) = \|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})} h(\|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})}), \quad \varepsilon \in [0, c^{-1}]. \end{aligned} \tag{2.9}$$

Since $f(\mu)$ is continuous in μ and $D^L f(\mu_0)(x)$ is continuous in x, by (2.1) and (2.4), we may apply the dominated convergence theorem to deduce from (2.9) with $n \to \infty$ that

$$\big| f(\mathscr{L}_{\xi_\varepsilon}) - f(\mathscr{L}_{\xi_0}) - \mathbb{E}[{}_{\mathbb{B}^*}\langle (D^L f)(\mu_0)(\xi_0), (\xi_\varepsilon - \xi_0) \rangle_{\mathbb{B}}] \big| \leq \|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})} h(\|\xi_\varepsilon - \xi_0\|_{L^p(\mathbb{P})}), \quad \varepsilon \in [0, c^{-1}].$$

Combining this with (2.8) and $h(r) \to 0$ as $r \to 0$, we prove (2.2).

(2) When μ_0 has an atom, we take a $\mathbb{B}$-valued bounded random variable X which is independent of $(\xi_\varepsilon)_{\varepsilon \in [0,1]}$ and $\mathscr{L}_X$ does not have an atom. Then $\mathscr{L}_{\xi_0 + sX + r(\xi_\varepsilon - \xi_0)} \in \mathscr{P}_p(\mathscr{B})$ does not have atom for any $s > 0, \varepsilon \in [0,1]$. By conditions in Theorem 2.1(2), there exists a small constant $s_0 \in (0,1)$

such that for any $s, \varepsilon \in (0, s_0]$, we may apply (2.2) to the family $\xi_0 + sX + (r+\delta)(\xi_\varepsilon - \xi_0)$ for small $\delta > 0$ to conclude

$$
\begin{aligned}
f(\mathscr{L}_{\xi_\varepsilon + sX}) - f(\mathscr{L}_{\xi_0 + sX}) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\delta} f(\mathscr{L}_{\xi_0 + sX + (r+\delta)(\xi_\varepsilon - \xi_0)})\big|_{\delta=0}\, \mathrm{d}r \\
&= \int_0^1 \mathbb{E}[_{\mathbb{B}^*}\langle D^L f(\mathscr{L}_{\xi_0 + sX + r(\xi_\varepsilon - \xi_0)})(\xi_0 + sX + r(\xi_\varepsilon - \xi_0)), \xi_\varepsilon - \xi_0 \rangle_{\mathbb{B}}]\, \mathrm{d}r.
\end{aligned}
$$

By conditions in Theorem 2.1(2), we may let $s \downarrow 0$ to derive

$$
f(\mathscr{L}_{\xi_\varepsilon}) - f(\mathscr{L}_{\xi_0}) = \int_0^1 \mathbb{E}[_{\mathbb{B}^*}\langle D^L f(\mathscr{L}_{\xi_0 + r(\xi_\varepsilon - \xi_0)})(\xi_0 + r(\xi_\varepsilon - \xi_0)), \xi_\varepsilon - \xi_0 \rangle_{\mathbb{B}}]\, \mathrm{d}r, \quad \varepsilon \in (0, s_0).
$$

Multiplying both sides by ε^{-1} and letting $\varepsilon \downarrow 0$, we finish the proof. $\square$

3 Well-posedness of (1.1)

When $p = 2$, the existence and uniqueness of strong solutions to (1.1) follows from [20, Theorem 3.1]; see also [28, Theorem 3.1] for $p \geq 2$, where $b(t, \xi, \mu)$ is Lipschitz continuous in $(\xi, \mu) \in \mathscr{C} \times \mathscr{P}_p(\mathscr{C})$. In the following result, the drift $b(t, \xi, \mu)$ may be non-Lipschitz continuous w.r.t. ξ.

Lem01 **Lemma 3.1.** *Assume* **(A)** *for some $p \in [1, \infty)$ and let $T \geq 0$. There exists a constant $c > 0$ such that for any $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$,* (1.1) *has a functional solution $X_{[0,T]} := (X_t)_{t \in [0,T]}$ satisfying*

ESTY (3.1)
$$
\mathbb{E}\Big(\sup_{0 \leq t \leq T} \|X_t\|_{\mathscr{C}}^p\Big) \leq c\Big(1 + \mathbb{E}\|X_0\|_{\mathscr{C}}^p\Big),
$$

and any two functional solutions $X_{[0,T]}$ and $Y_{[0,T]}$ satisfy

ESTY' (3.2)
$$
\mathbb{E}\Big(\sup_{0 \leq t \leq T} \|X_t - Y_t\|_{\mathscr{C}}^p\Big) \leq c\, \mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^p.
$$

Consequently, the SDE (1.1) *is strongly and weakly well-posed.*

Proof. By Itô's formula and BDG's inequality, it is easy to derive estimates (3.1) and (3.2) from assumption **(A)**. In particular, the strong uniqueness holds. Next, according to [29, Theorem 2.3], the assumption **(A)** implies the well-posedness of the SDE with memory

$$
\mathrm{d}X(t) = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W(t)
$$

for any $\mu_\cdot \in C([0,T]; \mathscr{P}_p(\mathscr{C}))$. As shown in the proof of [21, Lemma 2.1], the weak well-posedness of (1.1) follows from the strong one. So, it remains to prove the strong existence, for which we only need to find a constant $t_0 \in (0, T)$ independent of the initial value X_0 such that the SDE (1.1) is well-posed up to time t_0. Indeed, once this is proved, we may solve the SDE (1.1) from time t_0 up to $2t_0$ so that (1.1) is strongly well-posed up to time $2t_0$. Repeating the previous procedure for finite many times we obtain the strong well-posedness up to time T. Below we construct the solution by using an iterating argument as in [36].

Let $\mu = \mathscr{L}_{X_0} \in \mathscr{P}_p(\mathscr{C})$ and consider the SDE

$$
\mathrm{d}X^{(0)}(t) = b(t, X_t^{(0)}, \mu)\mathrm{d}t + \sigma(t, X_t^{(0)}, \mu)\mathrm{d}W(t), \quad t \geq 0, \quad X_0^{(0)} = X_0.
$$

According to [29, Theorem 2.3], **(A)** implies that this SDE has a unique solution. Using **(A)** and applying Itô's formula to $\xi^{(0)}(t) := (1+|X^{(0)}(t)|^2)^{\frac{p}{2}}$, we find a constant $c_1 > 0$ such that

$$\mathrm{d}\xi^{(0)}(t) \le c_1(1+\|\xi^{(0)}_t\|_{\mathscr{C}})\mathrm{d}t + p(1+|X^{(0)}(t)|^2)^{\frac{p-2}{2}}\big\langle X^{(0)}(t), \sigma(t, X^{(0)}_t, \mu)\mathrm{d}W(t)\big\rangle, \quad t\in[0,T].$$

For any $n\ge 1$, let $\tau_n = \inf\big\{t\ge 0 : \|X^{(0)}_t\|_{\mathscr{C}} \ge n\big\}$. By BDG's inequality and **(A)**, we find constants $c_2, c_3 > 0$ such that

$$\begin{aligned}\mathbb{E}\|\xi^{(0)}_{t\wedge\tau_n}\|_{\mathscr{C}} - \mathbb{E}\|\xi^{(0)}_0\|_{\mathscr{C}} &\le \mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|\xi^{(0)}(s)|\Big)\\ &\le c_1\int_0^t \mathbb{E}\|\xi^{(0)}_{s\wedge\tau_n}\|_{\mathscr{C}}\mathrm{d}s + c_2\mathbb{E}\Big(\int_0^{t\wedge\tau_n}\|\xi^{(0)}_s\|^2_{\mathscr{C}}\mathrm{d}s\Big)^{\frac{1}{2}}\\ &\le \frac{1}{2}\mathbb{E}\|\xi^{(0)}_{t\wedge\tau_n}\|_{\mathscr{C}} + c_3\int_0^t\mathbb{E}\|\xi^{(0)}_{s\wedge\tau_n}\|_{\mathscr{C}}\mathrm{d}s, \quad t\in[0,T].\end{aligned}$$

By Gronwall's lemma and $\mathbb{E}\|\xi^{(0)}_{t\wedge\tau_n}\|_{\mathscr{C}} < \infty$, it follows that

$$\mathbb{E}\|\xi^{(0)}_{t\wedge\tau_n}\|_{\mathscr{C}} \le 2\,\mathrm{e}^{2c_3 t}\mathbb{E}\|\xi^{(0)}_0\|_{\mathscr{C}}, \quad n\ge 1, t\in[0,T].$$

By Fatou's lemma for $n\to\infty$ and using the definition of $\xi^{(0)}_t$, we find a constant $c>0$ such that

P00

$$\mathbb{E}\|X^{(0)}_t\|^p_{\mathscr{C}} \le c(1+\mathbb{E}\|X_0\|^p_{\mathscr{C}})\mathrm{e}^{ct}, \quad t\in[0,T]. \tag{3.3}$$

If, for some $n\ge 1$, we have constructed a continuous adapted process $X^{(n-1)}_t \in \mathscr{C}$ with $X^{(n-1)}_0 = X_0$ such that

P0*0

$$\mathbb{E}\Big(\sup_{0\le s\le t}\|X^{(n-1)}_s\|^p_{\mathscr{C}}\Big) < \infty, \quad t\ge 0, \tag{3.4}$$

consider the SDE with memory

P01

$$\mathrm{d}X^{(n)}(t) = b\big(t, X^{(n)}_t, \mathscr{L}_{X^{(n-1)}_t}\big)\mathrm{d}t + \sigma\big(t, X^{(n)}_t, \mathscr{L}_{X^{(n-1)}_t}\big)\mathrm{d}W(t), \quad t\in[0,T], X^{(n)}_0 = X_0. \tag{3.5}$$

By [29, Theorem 2.3], **(A)** implies that (3.4) has a unique solution, and the argument leading to (3.3) yields

P0*1

$$\mathbb{E}\Bigg(\sup_{0\le s\le t}\|X^{(n)}_s\|^p_{\mathscr{C}}\Bigg) < \infty, \quad t\in[0,T]. \tag{3.6}$$

In this way we have constructed a sequence of continuous adapted process $\{(X^{(n)}_t)_{t\in[0,T]}\}_{n\ge 1}$ on $\mathscr{C}$. It suffices to find a constant $t_0 > 0$ independent of X_0 such that

CS

$$\lim_{n,m\to\infty}\mathbb{E}\Big(\sup_{0\le s\le t_0}\|X^{(n)}_s - X^{(m)}_s\|^p_{\mathscr{C}}\Big) = 0. \tag{3.7}$$

Indeed, this implies the existence of an adapted continuous process $(X_t)_{t\in[0,t_0]}$ such that

$$\lim_{n\to\infty}\mathbb{E}\Big(\sup_{0\le s\le t_0}\|X^{(n)}_s - X_s\|^p_{\mathscr{C}}\Big) = 0,$$

which, together with **(A)** and

$$X^{(n)}(t) = X_0(0) + \int_0^t b\big(X_s^{(n)}, \mathscr{L}_{X_s^{(n-1)}}\big)\mathrm{d}s + \int_0^t \sigma\big(X_s^{(n)}, \mathscr{L}_{X_s^{(n-1)}}\big)\mathrm{d}W(s), \quad t \in [0, t_0]$$

due to (3.5), implies that $(X_t)_{t\in[0,t_0]}$ solves (1.1) up to time t_0 with

$$\mathbb{E}\bigg(\sup_{0\le s\le t_0} \|X_s\|_{\mathscr{C}}^p\bigg) < \infty.$$

It remains to verify (3.7). To this end, for $n \ge 2$ we denote $\mu_t^{(n)} = \mathscr{L}_{X_t^{(n)}}$ and set

$$\Psi^{(n)}(t) := X^{(n)}(t) - X^{(n-1)}(t), \quad t \in [0, T].$$

By (3.5), we have

$$\begin{aligned} \mathrm{d}\Psi^{(n)}(t) &= \big\{b(t, X_t^{(n)}, \mu_t^{(n-1)}) - b(t, X_t^{(n-1)}, \mu_t^{(n-2)})\big\}\mathrm{d}t \\ &\quad + \big\{\sigma(t, X_t^{(n)}, \mu_t^{(n-1)}) - \sigma(t, X_t^{(n-1)}, \mu_t^{(n-2)})\big\}\mathrm{d}W(t), \quad \Psi_0^{(n)} = \mathbf{0}. \end{aligned}$$

By Itô's formula and **(A)**, we find a constant $c_1 > 0$ such that

$$\begin{aligned} \mathrm{d}|\Psi^{(n)}(t)|^p &\le c_1\big\{\|\Psi_t^{(n)}\|_{\mathscr{C}}^p + \mathbb{W}_p(\mu_t^{(n-1)}, \mu_t^{(n-2)})^p\big\}\mathrm{d}t \\ &\quad + p|\Psi^{(n)}(t)|^{p-2}\big\langle\Psi^{(n)}(t), \{\sigma(t, X_t^{(n)}, \mu_t^{(n-1)}) - \sigma(t, X_t^{(n-1)}, \mu_t^{(n-2)})\}\mathrm{d}W(t)\big\rangle, \ \Psi_0^{(n)} = \mathbf{0}. \end{aligned}$$

By BDG's inequality and noting that, for $p \in [1, 2]$, the coefficient $\sigma(t, \xi, \mu)$ depends only on (t, ξ) so that **(A)** implies

$$\|\sigma(t, X_t^{(n)}, \mu_t^{(n-1)}) - \sigma(t, X_t^{(n-1)}, \mu_t^{(n-2)})\|_{\mathrm{HS}}^2 \le K\|\Psi_t^{(n)}\|_{\mathscr{C}}^2,$$

there exist constants $c_2, c_3 > 0$ such that

$$\begin{aligned} \mathbb{E}\Big(\sup_{0\le s\le t} \|\Psi_s^{(n)}\|_{\mathscr{C}}^p\Big) &= \mathbb{E}\Big(\sup_{0\le s\le t} |\Psi^{(n)}(s)|^p\Big) \\ &\le c_1 \int_0^t \big\{\mathbb{E}\|\Psi_s^{(n)}\|_{\mathscr{C}}^p + \mathbb{W}_p(\mu_s^{(n-1)}, \mu_s^{(n-2)})^p\big\}\mathrm{d}s \\ &\quad + c_2\mathbb{E}\bigg(\int_0^t \|\Psi_s^{(n)}\|_{\mathscr{C}}^{2(p-1)}\big\{\|\Psi_s^{(n)}\|_{\mathscr{C}}^2 + 1_{\{p\ge2\}}\mathbb{W}_p(\mu_s^{(n-1)}, \mu_s^{(n-2)})^2\big\}\mathrm{d}s\bigg)^{\frac{1}{2}} \\ &\le \frac{1}{2}\mathbb{E}\Big(\sup_{0\le s\le t} \|\Psi_s^{(n)}\|_{\mathscr{C}}^p\Big) + c_2 \int_0^t \big\{\mathbb{E}\|\Psi_s^{(n)}\|_{\mathscr{C}}^p + \mathbb{W}_p(\mu_s^{(n-1)}, \mu_s^{(n-2)})^p\big\}\mathrm{d}s. \end{aligned}$$

By Gronwall's lemma, (3.4), (3.6) and noting that

$$\mathbb{W}_p(\mu_s^{(n-1)}, \mu_s^{(n-2)})^p \le \mathbb{E}\|\Psi_s^{(n-1)}\|_{\mathscr{C}}^p,$$

we obtain

$$\mathbb{E}\Big(\sup_{0\le s\le t} \|\Psi_s^{(n)}\|_{\mathscr{C}}^p\Big) \le 2c_2 t\mathrm{e}^{2c_2 t}\mathbb{E}\Big(\sup_{0\le s\le t} \|\Psi_s^{(n-1)}\|_{\mathscr{C}}^p\Big), \quad t \in [0, T].$$

Taking $t_0 \in (0, T]$ such that $2c_2 t_0 \mathrm{e}^{2c_2 t_0} \le \frac{1}{2}$, we arrive at

$$\mathbb{E}\bigg(\sup_{0\le s\le t_0} \|\Psi_s^{(n)}\|_{\mathscr{C}}^p\bigg) \le \frac{1}{2}\mathbb{E}\bigg(\sup_{0\le s\le t_0} \|\Psi_s^{(n-1)}\|_{\mathscr{C}}^p\bigg), \quad n \ge 2.$$

This and $\mathbb{E}\big(\sup_{0\le s\le t_0} \|\Psi_s^{(1)}\|_{\mathscr{C}}^p\big) < \infty$ imply (3.7). $\square$

4 The Malliavin derivative of X_t^μ

Consider the separable Banach space $\mathscr{C}$ with the uniform norm $\|\xi\|_{\mathscr{C}} := \sup_{t\in[-r_0,0]}|\xi(t)|$. For a Gâteaux differentiable matrix-valued function f on $\mathscr{C}$, let

$$\|\nabla f(\xi)\| = \sup_{\eta\in\mathscr{C},\|\eta\|_{\mathscr{C}}\le 1}\|(\nabla_\eta f)(\xi)\|_{\mathrm{HS}},\quad \xi\in\mathscr{C},$$

where

$$(\nabla_\eta f)(\xi) := \lim_{\varepsilon\downarrow 0}\frac{f(\xi+\varepsilon\eta)-f(\xi)}{\varepsilon}.$$

Besides **(A)**, we will need the following assumption. A function f on $\mathscr{C}$ is called C^1-smooth, denoted by $f\in C^1(\mathscr{C})$, if it is Gâteaux differentiable with derivative $\nabla f(\xi)$ continuous in ξ. Moreover, if the derivative is bounded, we write $f\in C_b^1(\mathscr{C})$. It is well known that a function $f\in C^1(\mathscr{C})$ is Fréchet differentiable.

(B) Let $p\in[1,\infty)$. $\sigma(t,\xi,\mu)$ and $b(t,\xi,\mu)$ are C^1-smooth in $\xi\in\mathscr{C}$ and L-differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$, and satisfy the following conditions.

(B_1) $\sigma(t,\xi,\mu)$ is bounded and $(\nabla_\eta\sigma)(t,\cdot,\mu)(\xi)$ is continuous in $(\xi,\eta)\in\mathscr{C}\times\mathscr{C}$, and there exist increasing functions $K_1,K_2:[0,\infty)\to[0,\infty)$ such that

$$\|(\nabla b)(t,\cdot,\mu)(\xi)\|\le K_1(t)\big\{1+\|\xi\|_{\mathscr{C}}^{\frac{(p-2)^+}{2}}+K_2(\|\mu\|_p)\big\},\quad (t,\xi,\mu)\in[0,\infty)\times\mathscr{C}\times\mathscr{P}_p(\mathscr{C}).$$

(B_2) $b(t,\xi,\cdot),\sigma(t,\xi,\cdot)\in C^1(\mathscr{P}_p(\mathscr{C}))$ with

$$\sup_{(t,\xi,\mu)\in[0,T]\times\mathscr{C}\times\mathscr{P}_p(\mathscr{C})}\{\mu(\|D^Lb(t,\xi,\cdot)(\mu)(\cdot)\|_{\mathscr{C}^*}^2)+\mu(\|D^L\sigma(t,\xi,\cdot)(\mu)(\cdot)\|_{\mathscr{C}^*}^2)\}<\infty,\ T>0.$$

(B_3) If $p\in[1,2)$, then there exists an increasing function $K:[0,\infty)\to[0,\infty)$ such that

$$\|\sigma(t,\xi,\mu)\|\le K(t)\big(1+\|\xi\|^{\frac{p}{2}}\big),\quad \xi\in\mathscr{C}.$$

(B_4) If $p\in[1,2)$, then $\sigma(t,\xi,\mu)=\sigma(t,\xi(0))$ only depends on t, and $\xi(0)$.

Since **(B)** ensures that $b(t,\xi,\mu)$ and $\sigma(t,\xi,\mu)$ are Gâteaux differentiable in $\xi\in\mathscr{C}$, (**A**) implies

M2

$$\begin{aligned}&2\langle\xi(0),(\nabla_\xi b)(t,\cdot,\mu)(\eta)\rangle^+ +\|(\nabla_\xi\sigma)(t,\cdot,\mu)(\eta)\|_{\mathrm{HS}}^2\\&\le K\|\xi\|_{\mathscr{C}}^2,\quad \xi,\eta\in\mathscr{C},\mu\in\mathscr{P}_p(\mathscr{C}),t\in[0,T].\end{aligned}\tag{4.1}$$

For any $T>0$, set $\mathscr{C}_T := C([0,T];\mathbb{R}^d)$ and consider the Cameron-Martin space

$$\mathcal{H}=\left\{h\in\mathscr{C}_T\,\middle|\,h(0)=\mathbf{0},\dot h(t)\text{ exists a.e. }t,\|h\|_{\mathcal{H}}:=\left(\int_0^T|\dot h(t)|^2\mathrm{d}t\right)^{\frac12}<\infty\right\}.$$

By the pathwise uniqueness of (1.1), we may regard X_t^μ as a $\mathscr{C}$-valued function of X_0^μ and W, and investigate its Malliavin derivative w.r.t. the Brownian motion W. For any $h\in L^\infty(\Omega\to\mathcal{H},\mathbb{P})$ and $\varepsilon\ge 0$, consider the SDE

EE43

$$\begin{aligned}&\mathrm{d}X^{h,\varepsilon,\mu}(t)=\{b(t,X_t^{h,\varepsilon,\mu},\mu_t)+\varepsilon\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)\dot h(t)\}\mathrm{d}t+\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)\mathrm{d}W(t),\\&t\in[0,T],X_0^{h,\varepsilon,\mu}=X_0^\mu,\mu_t:=\mathscr{L}_{X_t^\mu}.\end{aligned}\tag{4.2}$$

When h is adapted, according to the proof of Lemma 3.1, assumption **(A)** implies the existence and uniqueness of this SDE.

The directional Malliavin derivative of $X^\mu(t)$ along h is given by

$$D_h X^\mu(t) := \lim_{\varepsilon \to 0} \frac{X^{h,\varepsilon,\mu}(t) - X^\mu(t)}{\varepsilon}$$

provided the limit exists in $L^2(\Omega \to C([0,T];\mathbb{R}^d), \mathbb{P})$. To prove the existence of this limit, we first present the following lemma.

Lem1 **Lemma 4.1.** *Assume* **(A)** *and let* (B_3) *hold if* $p \in [1,2)$*. Let* $h \in L^\infty(\Omega \to \mathcal{H}, \mathbb{P})$ *which is adapted if* $\sigma(t,\xi,\mu)$ *depends on* ξ*, and let* $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$*. Then there exists a constant* $c > 0$ *such that*

EE5
$$\mathbb{E}\Big(\sup_{0 \le s \le T} \|X_s^{h,\varepsilon,\mu} - X_s^\mu\|_{\mathscr{C}}^{2\vee p}\Big) \le c\,\varepsilon^{2\vee p}, \quad \varepsilon \in [0,1]. \tag{4.3}$$

Proof. Below, we only consider the case that h is adapted and $\sigma(t,\xi,\mu)$ depends on ξ, since the proof for the setup that $\sigma(t,\xi,\mu)$ is independent of ξ is even simpler.

Let $Z^{h,\varepsilon}(t) = \frac{X^{h,\varepsilon,\mu}(t) - X^\mu(t)}{\varepsilon}$ and

$$\tau_n = \inf\{t \ge 0 : \|X_t^\mu\|_{\mathscr{C}} + \|X_t^{h,\varepsilon,\mu}\|_{\mathscr{C}} \ge n\}, \quad n \ge 1.$$

By (1.1) and (4.2), we have

EE4
$$\begin{aligned} \mathrm{d}Z^{h,\varepsilon}(t) = &\Big\{ \frac{b(t, X_t^{h,\varepsilon,\mu}, \mu_t) - b(t, X_t^\mu, \mu_t)}{\varepsilon} + \sigma(t, X_t^{h,\varepsilon,\mu}, \mu_t)\dot{h}(t) \Big\}\mathrm{d}t \\ &+ \frac{\sigma(t, X_t^{h,\varepsilon,\mu}, \mu_t) - \sigma(t, X_t^\mu, \mu_t)}{\varepsilon}\mathrm{d}W(t), \quad Z_0^{h,\varepsilon} = \mathbf{0}. \end{aligned} \tag{4.4}$$

Applying Itô's formula and taking **(A)** and $Z_0^\varepsilon = \mathbf{0}$ into account yields, for $q := 2 \vee p$,

*PB
$$\begin{aligned} |Z^{h,\varepsilon}(t \wedge \tau_n)|^q \le &\frac{q}{2} \int_0^{t\wedge\tau_n} \Big\{ \frac{2}{\varepsilon} \langle Z^{h,\varepsilon}(s), b(s, X_s^{h,\varepsilon,\mu}, \mu_s) - b(s, X_s^\mu, \mu_s)\rangle \\ &+ \frac{q-1}{\varepsilon^2} \|\sigma(s, X_s^{h,\varepsilon,\mu}, \mu_s) - \sigma(s, X_s^\mu, \mu_s)\|_{\mathrm{HS}}^2 \Big\}\mathrm{d}s + N^\varepsilon(t) + M^\varepsilon(t) \\ \le &\, c \int_0^{t\wedge\tau_n} \|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q \mathrm{d}s + N^\varepsilon(t) + M^\varepsilon(t), \end{aligned} \tag{4.5}$$

for some constant $c > 0$, where, by setting $r^0 = 1$ for $r \in [0,\infty)$ in case of $p = 1$,

$$\begin{aligned} N^\varepsilon(t) &:= q \int_0^{t\wedge\tau_n} |Z^{h,\varepsilon}(s)|^{q-1} \sigma(s, X_s^{h,\varepsilon,\mu}, \mu_s)\dot{h}(s)|\mathrm{d}s, \\ M^\varepsilon(t) &:= \frac{q}{\varepsilon} \int_0^{t\wedge\tau_n} |Z^{h,\varepsilon}(s)|^{q-2} \langle Z^{h,\varepsilon}(s), (\sigma(s, X_s^{h,\varepsilon,\mu}, \mu_s) - \sigma(s, X_s^\mu, \mu_s))\mathrm{d}W(s)\rangle. \end{aligned}$$

Let $\psi > 0$ be a constant such that $\|h\|_{\mathcal{H}} \le \psi$ due to $h \in L^\infty(\Omega \to \mathcal{H}, \mathbb{P})$. By Hölder's and Young's inequalities, Lemma 3.1, **(A)** and (B_3) when $p \in [1,2)$, we find constants $c_0, c_1 > 0$ such that

EE1
$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|N^\varepsilon(s)|\Big) &\le q\psi\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^{2(q-1)}\int_0^{t\wedge\tau_n}\|\sigma(s,X_s^{h,\varepsilon,\mu},\mu_s)\|^2\mathrm{d}s\Big)^{1/2}\\ &\le \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^q\Big) + c_0\mathbb{E}\bigg(\int_0^t(1+\|X_s^{h,\varepsilon,\mu}\|_{\mathscr{C}}^{2\wedge p})\mathrm{d}s\bigg)^{\frac{2\vee p}{2}}\\ &\le \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^2\Big) + c_1,\quad t\in[0,T].\end{aligned}\tag{4.6}$$

By **(A)** and the BDG inequality, there exist constants $c_2, c_3 > 0$ such that

EE2
$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|M^\varepsilon(s)|\Big) &\le c_2\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}\|Z_s^\varepsilon\|_{\mathscr{C}}^q\int_0^{t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\Big)^{1/2}\\ &\le \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\Big) + c_3\int_0^t\mathbb{E}\|Z_{s\wedge\tau_n}^{h,\varepsilon}\|_{\mathscr{C}}^q\mathrm{d}s.\end{aligned}\tag{4.7}$$

Combining (4.5)-(4.7) we find a constant $c > 0$ such that

$$\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\Big) \le c + c\int_0^t\mathbb{E}\|Z_{s\wedge\tau_n}^{h,\varepsilon}\|_{\mathscr{C}}^q\mathrm{d}s < \infty,\quad t\in[0,T],\ \varepsilon\in[0,1].$$

By first applying Gronwall's inequality then letting $n\to\infty$, we derive (4.3). □

Lem5 **Lemma 4.2.** *Assume* **(A)** *and* **(B)** . *For any $X_0^\mu \in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$ and $h \in L^\infty(\Omega\to\mathcal{H},\mathbb{P})$ which is adapted if $\sigma(t,\xi,\mu)$ depends on ξ, the limit*

ELL
$$D_hX_t^\mu := \lim_{\varepsilon\downarrow 0}\frac{X_t^{h,\varepsilon,\mu}-X_t^\mu}{\varepsilon},\quad t\in[0,T]\tag{4.8}$$

exists in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$, and it is the unique solution of the following SDE with memory

EE3
$$\begin{aligned}\mathrm{d}w^h(t) &= \big\{(\nabla_{w_t^h}b)(t,\cdot,\mu_t)(X_t^\mu) + \sigma(t,X_t^\mu,\mu_t)\dot{h}(t)\big\}\mathrm{d}t\\ &\quad + (\nabla_{w_t^h}\sigma)(t,\cdot,\mu_t)(X_t^\mu)\mathrm{d}W(t),\ t\in[0,T],\ w_0^h = \mathbf{0}, \mu_t := \mathscr{L}_{X_t^\mu}.\end{aligned}\tag{4.9}$$

Proof. By (4.1) and the boundedness of σ due to (B_1), for any adapted $h \in L^2(\Omega\to\mathcal{H},\mathbb{P})$, the SDE (4.9) has a unique solution in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$ and for some constant $C > 0$,

FF
$$\mathbb{E}\Big(\sup_{0\le t\le T}\|w_t^h\|_{\mathscr{C}}^2\Big) \le C\mathbb{E}\|h\|_{\mathcal{H}}^2 < \infty.\tag{4.10}$$

So, it remains to prove that the limit in (4.8) exists in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$, and it solves (4.9). Let $\Lambda^{h,\varepsilon}(t) = Z^{h,\varepsilon}(t) - w^h(t)$, where $Z^{h,\varepsilon}(t) := \frac{X^{h,\varepsilon,\mu}(t)-X^\mu(t)}{\varepsilon}$ as before. Then, it suffices to verify

W1
$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\sup_{0\le s\le T}|\Lambda^{h,\varepsilon}(s)|^2\Big) = 0.\tag{4.11}$$

Observe that (4.3) and (4.10) imply

JH
$$\mathbb{E}\Big(\sup_{0\le s\le T}|\Lambda^{h,\varepsilon}(s)|^2\Big)<\infty. \tag{4.12}$$

By (4.4) and (4.9), we have

ELL2
$$\mathrm{d}\Lambda^{h,\varepsilon}(t)=\big\{(\nabla_{\Lambda_t^\varepsilon}b)(t,\cdot,\mu_t)(X_t^\mu)+\Gamma_1^\varepsilon(t)\big\}\mathrm{d}t+\big\{(\nabla_{\Lambda_t^\varepsilon}\sigma)(t,\cdot,\mu_t)(X_t^\mu)+\Gamma_2^\varepsilon(t)\big\}\mathrm{d}W(t), \tag{4.13}$$

where

*WQ
$$\begin{aligned}\Gamma_1^\varepsilon(t):&=(\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)-\sigma(t,X_t^\mu,\mu_t))\dot h(t)\\ &\quad+\int_0^1\big\{(\nabla_{Z_t^{h,\varepsilon}}b)(t,\cdot,\mu_t)(X_t^\mu+\theta(X_t^{h,\varepsilon,\mu}-X_t^\mu))-(\nabla_{Z_t^{h,\varepsilon}}b)(t,\cdot,\mu_t)(X_t^\mu)\big\}\mathrm{d}\theta\\ \Gamma_2^\varepsilon(t):&=\int_0^1\big\{(\nabla_{Z_t^{h,\varepsilon}}\sigma)(t,\cdot,\mu_t)(X_t^\mu+\theta(X_t^{h,\varepsilon,\mu}-X_t^\mu))-(\nabla_{Z_t^{h,\varepsilon}}\sigma)(t,\cdot,\mu_t)(X_t^\mu)\big\}\mathrm{d}\theta.\end{aligned} \tag{4.14}$$

Obviously, when $\sigma(t,\xi,\mu)=\sigma(t,\mu)$ does not depend on ξ, the noise term in (4.13) disappears so that the SDE reduces to an ODE for which we can allow h to be non-adapted. Applying Itô's formula yields

$$\begin{aligned}&|\Lambda^{h,\varepsilon}(t)|^2\le\int_0^t\big\{2\langle\Lambda^{h,\varepsilon}(s),(\nabla_{\Lambda_s^{h,\varepsilon}}b)(s,\cdot,\mu_s)(X_s^\mu)\rangle+2\|(\nabla_{\Lambda_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^\mu)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s\\ &+2\int_0^t\big\{\langle\Lambda^{h,\varepsilon}(s),\Gamma_1^\varepsilon(s)\rangle+\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s+2\int_0^t\big\langle\Lambda^{h,\varepsilon}(s),\{(\nabla_{\Lambda_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^\mu)+\Gamma_2^\varepsilon(s)\}\mathrm{d}W(s)\big\rangle\\ &=:\Upsilon_1^\varepsilon(t)+\Upsilon_2^\varepsilon(t)+\Upsilon_3^\varepsilon(t).\end{aligned}$$

Obviously, (4.1) implies

F1
$$\mathbb{E}\Big(\sup_{0\le s\le t}\Upsilon_1^\varepsilon(s)\Big)\le 3K\int_0^t\mathbb{E}\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s, \tag{4.15}$$

while Cauchy-Schwarz's inequality gives

F2
$$\mathbb{E}\Big(\sup_{0\le s\le t}|\Upsilon_2^\varepsilon(s)|\Big)\le\int_0^t\{2\mathbb{E}|\Lambda^{h,\varepsilon}(s)|^2+\mathbb{E}|\Gamma_1^\varepsilon(s)|^2+2\,\mathbb{E}\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\}\mathrm{d}s. \tag{4.16}$$

Next, by (4.1) and BDG's inequality, we find constants $c_1,c_2>0$ such that

F3
$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}\Upsilon_3^\varepsilon(s)\Big)&\le c_1\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\int_0^t\|(\nabla_{\Lambda_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^\mu)+\Gamma_2^\varepsilon(s)\|^2\mathrm{d}s\Big)^{1/2}\\ &\le\frac12\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big)+c_2\int_0^t\{\mathbb{E}\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2+\mathbb{E}\|\Gamma_2^\varepsilon(s)\|^2\}\mathrm{d}s.\end{aligned} \tag{4.17}$$

Combining (4.15), (4.16) with (4.17), there exists a constant $c_3>0$ such that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big)\le c_3\int_0^t\mathbb{E}\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s+c_3\int_0^t\{\mathbb{E}|\Gamma_1^\varepsilon(s)|^2+\mathbb{E}\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\}\mathrm{d}s.$$

By Gronwall's inequality and (4.12), this implies

W2 (4.18) $$\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big)\le c_3\mathrm{e}^{c_3t}\mathbb{E}\int_0^t\big\{|\Gamma_1^\varepsilon(s)|^2+\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s.$$

Moreover, by (4.14), we have

KB1 (4.19) $$|\Gamma_1^\varepsilon(s)|^2+\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\le I_\varepsilon(t)|\dot h(t)|^2+J_\varepsilon(t)\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2,$$

where according to (B_1) and (4.1) we find a constant $c(T)>0$ increasing in T such that

$$\begin{aligned}
I_\varepsilon(t)&:=2\|\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)-\sigma(t,X_t^\mu,\mu_t)\|^2,\\
J_\varepsilon(t)&:=2\int_0^1\{\|(\nabla b)(t,\cdot,\mu_t)(X_t^\mu+\theta(X_t^{h,\varepsilon,\mu}-X_t^\mu))-(\nabla b)(t,\cdot,\mu_t)(X_t^\mu)\|^2\\
&\qquad\qquad+\|(\nabla\sigma)(t,\cdot,\mu_t)(X_t^\mu+\theta(X_t^{h,\varepsilon,\mu}-X_t^\mu))-(\nabla\sigma)(t,\cdot,\mu_t)(X_t^\mu)\|^2\}\mathrm{d}\theta\\
&\quad\le c(T)\big(1+\|X_t^\mu\|_{\mathscr{C}}^{p-2}+\|X_t^{h,\varepsilon,\mu}-X_t^\mu\|_{\mathscr{C}}^{p-2}+K_2(\|\mu_t\|_p^2)\big),\quad t\in[0,T].
\end{aligned}$$

By **(B)** and (4.3), $I_\varepsilon(t)$ is bounded and $I_\varepsilon(t)\to0$ in probability as $\varepsilon\to0$. So, the dominated convergence theorem yields

KB2 (4.20) $$\lim_{\varepsilon\to0}\mathbb{E}\int_0^T I_\varepsilon(t)|\dot h(t)|^2\mathrm{d}t=0.$$

Below we complete the proof of (4.11) by considering two different cases.

(1) When $p>2$, (3.1) and (4.3) imply that $\{\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2(1+\|X_t^\mu\|_{\mathscr{C}}^{p-2})\}_{\varepsilon\in[0,1]}$ is uniformly integrable in $L^1(\mathbb{P})$ and

$$\mathbb{E}[\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2\|X_t^{h,\varepsilon,\mu}-X_t^\mu\|_{\mathscr{C}}^{p-2}]=\varepsilon^{p-2}\mathbb{E}\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^p\le c\,\varepsilon^{p-2}\to0\qquad\text{as }\varepsilon\to0.$$

Then, by the dominated convergence theorem, (4.3) and $J_\varepsilon(t)\to0$ in probability, we arrive at

$$\lim_{\varepsilon\to0}\mathbb{E}\int_0^T J_\varepsilon(t)\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}t=0.$$

This, together with (4.19) and (4.20), implies

XTD (4.21) $$\lim_{\varepsilon\to0}\mathbb{E}\int_0^T\big\{|\Gamma_1^\varepsilon(t)|^2+\|\Gamma_2^\varepsilon(t)\|_{\mathrm{HS}}^2\big\}\mathrm{d}t=0$$

so that (4.11) follows from (4.18).

(2) When $p\in[1,2]$, (B_1) and (3.1) imply $J_\varepsilon(t)\le K$ for some constant K depending on T. Then,

bb (4.22) $$\mathbb{E}\int_0^t\big\{|\Gamma_1^\varepsilon(s)|^2+\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s\le\varepsilon_T+2K\int_0^t\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s,\quad t\in[0,T],$$

where, by the dominated convergence theorem,

$$\varepsilon_T:=\int_0^T\mathbb{E}\big[I_\varepsilon(t)|\dot h(t)|^2+J_\varepsilon(t)\|w_t^h\|_{\mathscr{C}}^2\big]\mathrm{d}t\to0\quad\text{ as }\varepsilon\to0.$$

Substituting (4.22) into (4.18) and using Gronwall's lemma, we derive

$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big)\le \lim_{\varepsilon\to 0}\varepsilon_T \mathrm{e}^{(c_3+2K)T}=0.$$

Therefore, (4.11) holds. □

Let $(D,\mathscr{D}(D))$ be the Malliavin gradient with adjoint (i.e., Malliavin divergence) $(D^*,\mathscr{D}(D^*))$. Then,

ITP
$$\mathbb{E}[D_hF]=\mathbb{E}[FD^*(h)],\quad F\in\mathscr{D}(D), h\in\mathscr{D}(D^*). \tag{4.23}$$

In particular, if $h\in L^2(\Omega\to\mathcal{H},\mathbb{P})$ is adapted, then $h\in\mathscr{D}(D^*)$ and

ITP'
$$D^*(h)=\int_0^T\langle\dot h(t),\mathrm{d}W(t)\rangle, \tag{4.24}$$

see, for example, [23].

P4.3 **Proposition 4.3.** *Assume* **(A)** *and* **(B)**. *For any $h\in\mathscr{D}(D^*)$ which is adapted if $\sigma(t,\xi,\mu)$ depends on ξ,* (4.9) *has a unique functional solution satisfying* (4.10) *for some constant $C>0$, and for any $f\in C_b^1(\mathscr{C})$,*

*WPB
$$\mathbb{E}\big[(\nabla_{w_T^h}f)(X_T^\mu)\big]=\mathbb{E}\big[f(X_T^\mu)D^*(h)\big]. \tag{4.25}$$

Proof. As explained in the proof of Lemma 4.2, the first assertion follows from assumptions **(A)** and **(B)**. So it suffices to prove (4.25).

We first consider $h\in L^\infty(\Omega\to\mathcal{H},\mathbb{P})\cap\mathscr{D}(D^*)$. By Lemma 4.2, the chain rule and (4.23), we obtain

bb1
$$\mathbb{E}\big[(\nabla_{w_T^h}f)(X_T^\mu)\big]=\mathbb{E}\big[D_h\{f(X_T^\mu)\}\big]=\mathbb{E}\big[f(X_T^\mu)D^*(h)\big]. \tag{4.26}$$

In general, for adapted $h\in\mathscr{D}(D^*)$, we choose $(h_n)_{n\ge0}\subset L^\infty(\Omega\to\mathcal{H},\mathbb{P})\cap\mathscr{D}(D^*)$ such that

ITP2
$$\lim_{n\to\infty}\mathbb{E}\big[\|h_n-h\|_{\mathcal{H}}^2+|D^*(h_n)-D^*(h)|^2\big]=0. \tag{4.27}$$

In terms of (4.26), we have

LDB
$$\mathbb{E}\big[(\nabla_{w_T^{h_n}}f)(X_T^\mu)\big]=\mathbb{E}\big[f(X_T^\mu)D^*(h_n)\big],\quad n\ge1. \tag{4.28}$$

By **(A)** and **(B)**, we find a constant $C>0$ such that

$$\mathbb{E}\|w_T^{h_n}-w_T^h\|_{\mathscr{C}}^2\le C\mathbb{E}\|h-h_n\|_{\mathcal{H}}^2.$$

This, together with $f\in C_b^1(\mathscr{C})$ and (4.27), the desired formula (4.25) by taking $n\to\infty$ in (4.28). □

5 The Gâteaux and intrinsic derivatives

For fixed $p \in [2,\infty)$ and $X_0^\mu \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ with the distribution μ, let $(X_t^\mu)_{t\geq 0}$ be the unique solution to (1.1) starting from X_0^μ. To calculate the intrinsic derivative of X_t^μ w.r.t. μ, we consider the tangent space $T_{\mu,p} := L^p(\mathscr{C} \to \mathscr{C}, \mu)$, where $\mathscr{C} := C([-r_0,0];\mathbb{R}^d)$ endowed with the uniform norm $\|\xi\|_{\mathscr{C}} := \sup_{t\in[-r_0,0]} |\xi(t)|$ is a separable Banach space with the dual space $\mathscr{C}^*$ consisting of all bounded linear functionals $\alpha : \mathscr{C} \to \mathbb{R}$. We denote the dualization between $\mathscr{C}^*$ and $\mathscr{C}$ by ${}_{\mathscr{C}^*}\langle \alpha, \xi\rangle_{\mathscr{C}} = \alpha(\xi)$ for $\alpha \in \mathscr{C}^*, \xi \in \mathscr{C}$. For any $\mu \in \mathscr{P}_p(\mathscr{C})$ and $\phi \in T_{\mu,p}$, let

$$\mu^\phi = \mu \circ (\mathrm{Id} + \phi)^{-1} = \mathscr{L}_{(\mathrm{Id}+\phi)(X_0^\mu)}.$$

Let $(X_t^{\mu^\phi})_{t\geq 0}$ be the functional solution to (1.1) with $X_0^{\mu^\phi} := (\mathrm{Id} + \phi)(X_0^\mu)$, and denote

$$\mu_t^\phi = \mathscr{L}_{X_t^{\mu^\phi}}, \quad t \geq 0.$$

Then the directional intrinsic derivative of X_t^μ along ϕ is given by

D12 $$D_\phi^L X_t^\mu := \lim_{\varepsilon\to 0} \frac{X_t^{\mu^{\varepsilon\phi}} - X_t^\mu}{\varepsilon} \tag{5.1}$$

provided the limit above exists.

More generally, for $\xi \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ and $\varepsilon \in [0,1]$, we let $X_t^{\varepsilon\xi,\mu}$ be the functional solution to (1.1) with $X_0^{\varepsilon\xi,\mu} := \varepsilon\xi + X_0^\mu$, and denote $\mu_t^{\xi,\varepsilon} = \mathscr{L}_{X_t^{\varepsilon\xi,\mu}}$. Then the Gâteaux derivative of X_t^μ along ξ is

D12' $$\nabla_\xi X_t^\mu := \lim_{\varepsilon\to 0} \frac{X_t^{\varepsilon\xi,\mu} - X_t^\mu}{\varepsilon} \tag{5.2}$$

provided the limit above exists. Obviously,

D12'' $$\nabla_\xi X_t^\mu = D_\phi^L X_t^\mu \qquad \text{if } \xi = \phi(X_0^\mu). \tag{5.3}$$

To prove the existence of $\nabla_\xi X_t^\mu$, we need the following lemma.

L1 **Lemma 5.1.** *Assume* **(A)** *and that* (B_4) *holds when* $p \in [1,2)$. *For any* $T > 0$ *and* $q \geq p$, *there exists a constant* $c > 0$ *such that*

D3 $$\mathbb{E}\Big(\sup_{0\leq s\leq t} \|X_s^{\varepsilon\xi,\mu} - X_s^\mu\|_{\mathscr{C}}^q\Big) \leq \varepsilon^p\, \mathrm{e}^{ct}\mathbb{E}\|\xi\|_{\mathscr{C}}^q, \quad t \in [0,T],\ \varepsilon \in [0,1],\ \xi \in L^q(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P}). \tag{5.4}$$

Proof. Set $\Phi^{\xi,\varepsilon}(t) := \frac{X^{\varepsilon\xi,\mu}(t) - X^\mu(t)}{\varepsilon}, t \geq -r_0, \varepsilon > 0$. Since $X_t^{\varepsilon\xi,\mu}$ and X_t^μ solve (1.1) with the initial values $X_0^{\varepsilon\xi,\mu}$ and X_0^μ, respectively, one has

D11 $$\begin{aligned} \mathrm{d}\Phi^{\xi,\varepsilon}(t) &= \frac{1}{\varepsilon}\{b(t, X_t^{\varepsilon\xi,\mu}, \mu_t^{\xi,\varepsilon}) - b(t, X_t^\mu, \mu_t)\}\mathrm{d}t \\ &\quad + \frac{1}{\varepsilon}\{\sigma(t, X_t^{\varepsilon\xi,\mu}, \mu_t^{\xi,\varepsilon}) - \sigma(t, X_t^\mu, \mu_t)\}\mathrm{d}W(t), \quad t \geq 0, \Phi_0^{\xi,\varepsilon} = \xi. \end{aligned} \tag{5.5}$$

By **(A)** and (B_4) when $p \in [1,2)$, and applying Itô's formula and the fact that

$$\mathbb{W}_p(\mu_s^{\xi,\varepsilon}, \mu_s)^p \leq \mathbb{E}\|X_s^{\varepsilon\xi,\mu} - X_s^\mu\|_{\mathscr{C}}^p = \varepsilon^p\mathbb{E}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^p \leq \varepsilon^p\{\mathbb{E}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\}^{\frac{p}{q}},$$

we find a constant $c_1 > 0$ such that

D03

$$
\begin{aligned}
|\Phi^{\xi,\varepsilon}(t)|^q &\le \frac{q}{2}\int_0^t |\Phi^{\xi,\varepsilon}(s)|^{q-2}\Big\{\frac{2}{\varepsilon}\langle \Phi^{\xi,\varepsilon}(s), b(s, X_s^{\varepsilon\xi,\mu}, \mu_s^{\xi,\varepsilon}) - b(s, X_s^{\mu}, \mu_s)\rangle \\
&\quad + \frac{q-1}{\varepsilon^2}\|\sigma(s, X_s^{\varepsilon\xi,\mu}, \mu_s^{\xi,\varepsilon}) - \sigma(s, X_s^{\mu}, \mu_s)\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s + M^{\varepsilon}(t) \\
&\le c_1\int_0^t \big(\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q + \mathbb{E}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\big)\mathrm{d}s + M^{\varepsilon}(t), \quad t \ge 0,
\end{aligned} \tag{5.6}
$$

where

$$
M^{\varepsilon}(t) := \frac{q}{\varepsilon}\int_0^t |\Phi^{\xi,\varepsilon}(s)|^{q-2}\big\langle \Phi^{\xi,\varepsilon}(s), (\sigma(s, X_s^{\varepsilon\xi,\mu}, \mu_s^{\xi,\varepsilon}) - \sigma(s, X_s^{\mu}, \mu_s))\mathrm{d}W(s)\big\rangle.
$$

Next, by BDG's inequality and **(A)**, there exist some constants $c_2, c_3 > 0$ such that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t} M^{\varepsilon}(s)\Big) &\le \frac{c_2}{\varepsilon}\mathbb{E}\bigg(\sup_{0\le s\le t}|\Phi^{\xi,\varepsilon}(s)|^q\int_0^t |\Phi^{\xi,\varepsilon}(s)|^{q-2}\|\sigma(s, X_s^{\varepsilon\xi,\mu}, \mu_s^{\xi,\varepsilon}) - \sigma(s, X_s^{\mu}, \mu_s)\|^2\mathrm{d}s\bigg)^{\frac{1}{2}} \\
&\le \frac{1}{2}\mathbb{E}\Big(\sup_{0\le s\le t}|\Phi^{\xi,\varepsilon}(s)|^q\Big) + c_3\mathbb{E}\int_0^t \|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\mathrm{d}s.
\end{aligned}
$$

Combining this with (5.6), we derive

$$
\mathbb{E}\Big(\sup_{0\le s\le t}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\Big) \le 2\mathbb{E}\|\Phi_0^{\xi,\varepsilon}\|_{\mathscr{C}}^q + c_4\int_0^t \mathbb{E}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\mathrm{d}s, \quad t \ge 0
$$

for some constant $c_4 > 0$. By stopping at an exit time as in the proof of Lemma 4.1, we may assume $\mathbb{E}\Big(\sup_{0\le s\le t}\|\Phi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^q\Big) < \infty$, such that (5.4) follows from Gronwall's inequality. $\square$

Consider the following SDE with memory

D007

$$
\begin{aligned}
\mathrm{d}v^{\xi}(t) &= \Big\{(\nabla_{v_t^{\xi}} b)(t, \cdot, \mu_t)(X_t^{\mu}) + (\mathbb{E}_{\mathscr{C}^*}\langle D^L b(t, \eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}})\Big|_{\eta = X_t^{\mu}}\Big\}\mathrm{d}t \\
&\quad + \Big\{(\nabla_{v_t^{\xi}}\sigma)(t, \cdot, \mu_t)(X_t^{\mu}) + (\mathbb{E}_{\mathscr{C}^*}\langle D^L\sigma(t, \eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}})\Big|_{\eta = X_t^{\mu}}\Big\}\mathrm{d}W(t), \quad v_0^{\xi} = \xi,
\end{aligned} \tag{5.7}
$$

where $\mu_t := \mathscr{L}_{X_t^{\mu}}$ and

$$
\begin{aligned}
{}_{\mathscr{C}^*}\langle D^L b(\eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}} &:= ({}_{\mathscr{C}^*}\langle D^L b_i(\eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}})_{1\le i\le d} \in \mathbb{R}^d \\
{}_{\mathscr{C}^*}\langle D^L\sigma(\eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}} &:= ({}_{\mathscr{C}^*}\langle D^L\sigma_{ij}(\eta, \cdot)(\mu_t)(X_t^{\mu}), v_t^{\xi}\rangle_{\mathscr{C}})_{1\le i\le d, 1\le j\le m} \in \mathbb{R}^d \otimes \mathbb{R}^m, t \ge 0.
\end{aligned}
$$

Let $p \ge 2$. By **(A)** and **(B)**, this linear SDE has a unique solution. Moreover, by Itô's formula and BDG's inequality, we find a constant $c > 0$ such that

F05

$$
\mathbb{E}[\|v_t^{\xi}\|_{\mathscr{C}}^p] \le c\,\mathbb{E}[\|\xi\|_{\mathscr{C}}^p], \quad t \in [0, T], \quad \xi \in L^q(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P}). \tag{5.8}
$$

Lem **Lemma 5.2.** *Assume* **(A)** *and* **(B)** *for some $p \ge 2$. Then for any $\xi \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$, the limit in* (5.2) *exists in $L^2(\Omega \to C([0,T]; \mathscr{C}), \mathbb{P})$ and it gives rise to the unique functional solution of* (5.7).

Proof. Let $\Xi_t^{\xi,\varepsilon}=\Phi_t^{\xi,\varepsilon}-v_t^{\xi}$, where $(\Phi_t^{\xi,\varepsilon})_{t\ge0}$ solves (5.5). To end the proof, it suffices to prove

D10 (5.9)
$$\lim_{\varepsilon\to0}\mathbb{E}\Big(\sup_{0\le t\le T}\|\Xi_t^{\xi,\varepsilon}\|_{\mathscr{C}}^2\Big)=0,\quad T>0.$$

Set
$$X^{\varepsilon,\theta}(t):=X^{\mu}(t)+\theta(X^{\varepsilon\xi,\mu}(t)-X^{\mu}(t)),\ t\ge -r_0,\ \theta\in[0,1].$$

By (5.5), (5.7) and Theorem 2.1, we obtain

$$\begin{aligned}\mathrm{d}\Xi^{\xi,\varepsilon}(t)=&\Big\{(\nabla_{\Xi_t^{h,\varepsilon}}b)(t,\cdot,\mu_t)(X_t^{\mu})+(\mathbb{E}_{\mathscr{C}^*}\langle(D^Lb(t,\eta,\cdot))(\mu_t)(X_t^{\mu}),\Xi_t^{\xi,\varepsilon}\rangle_{\mathscr{C}})\Big|_{\eta=X_t^{\mu}}+\Upsilon_1^{\varepsilon}(t)\Big\}\mathrm{d}t\\&+\Big\{(\nabla_{\Xi_t^{h,\varepsilon}}\sigma)(t,\cdot,\mu_t)(X_t^{\mu})+(\mathbb{E}_{\mathscr{C}^*}\langle(D^L\sigma(t,\eta,\cdot))(\mu_t)(X_t^{\mu}),\Xi_t^{\xi,\varepsilon}\rangle_{\mathscr{C}})\Big|_{\eta=X_t^{\mu}}+\Upsilon_2^{\varepsilon}(t)\Big\}\mathrm{d}W(t),\end{aligned}$$

where

$$\begin{aligned}\Upsilon_1^{\varepsilon}(t):=&\int_0^1\{(\nabla_{\Phi_t^{\xi,\varepsilon}}b)(t,\cdot,\mu_t^{\xi,\varepsilon})(X_t^{\varepsilon,\theta})-(\nabla_{\Phi_t^{\xi,\varepsilon}}b)(t,\cdot,\mu_t)(X_t^{\mu})\}\mathrm{d}\theta\\&+\int_0^1\Big\{(\mathbb{E}_{\mathscr{C}^*}\langle(D^Lb(t,\eta,\cdot))(\mathscr{L}_{X_t^{\varepsilon,\theta}})(X_t^{\varepsilon,\theta})-(D^Lb(t,\eta,\cdot))(\mu_t)(X_t^{\mu}),\Phi_t^{\xi,\varepsilon}\rangle_{\mathscr{C}})\Big\}\Big|_{\eta=X_t^{\mu}}\mathrm{d}\theta,\\\Upsilon_2^{\varepsilon}(t):=&\int_0^1\{(\nabla_{\Phi_t^{\xi,\varepsilon}}\sigma)(t,\cdot,\mu_t^{\xi,\varepsilon}))(X_t^{\varepsilon,\theta})-(\nabla_{\Phi_t^{\xi,\varepsilon}}\sigma)(t,\cdot,\mu_t)(X_t^{\mu})\}\mathrm{d}\theta\\&+\int_0^1\Big\{(\mathbb{E}_{\mathscr{C}^*}\langle(D^L\sigma(t,\eta,\cdot))(\mathscr{L}_{X_t^{\varepsilon,\theta}})(X_t^{\varepsilon,\theta})-(D^L\sigma(t,\eta,\cdot))(\mu_t)(X_t^{\mu}),\Phi_t^{\xi,\varepsilon}\rangle_{\mathscr{C}})\Big\}\Big|_{\eta=X_t^{\mu}}\mathrm{d}\theta.\end{aligned}$$

By Itô's formula, we obtain

YPP (5.10)
$$|\Xi^{\xi,\varepsilon}(t)|^2\le\Theta_1^{\varepsilon}(t)+\Theta_2^{\varepsilon}(t)+\Theta_3^{\varepsilon}(t)+\Theta_4^{\varepsilon}(t),\quad t\ge0,$$

where

$$\begin{aligned}\Theta_1^{\varepsilon}(t):=&\int_0^t\Big\{2\langle\Xi^{\xi,\varepsilon}(s),(\nabla_{\Xi_s^{\xi,\varepsilon}}b)(s,\cdot,\mu_s)(X_s^{\mu})\rangle+3\,\|(\nabla_{\Xi_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^{\mu})\|_{\mathrm{HS}}^2\\&+2\big\langle\Xi^{\xi,\varepsilon}(s),\{\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(s,\eta,\cdot)(\mu_s)(X_s^{\mu}),\Xi_s^{\xi,\varepsilon}\rangle_{\mathscr{C}}\}\big\rangle\Big|_{\eta=X_s^{\mu}}\\&+3\|(\mathbb{E}_{\mathscr{C}^*}\langle(D^L\sigma(s,\eta,\cdot))(\mu_s)(X_s^{\mu}),\Xi_s^{\xi,\varepsilon}\rangle_{\mathscr{C}})\|_{\mathrm{HS}}^2\Big|_{\eta=X_s^{\mu}}\Big\}\mathrm{d}s,\\\Theta_2^{\varepsilon}(t):=&\int_0^t\{3\|\Upsilon_2^{\varepsilon}(s)\|_{\mathrm{HS}}^2+2\langle\Xi^{\xi,\varepsilon}(s),\Upsilon_1^{\varepsilon}(s)\rangle\}\mathrm{d}s,\\\Theta_3^{\varepsilon}(t):=&2\int_0^t\big\langle\Xi^{\xi,\varepsilon}(s),\{(\nabla_{\Xi_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^{\mu})\\&+(\mathbb{E}_{\mathscr{C}^*}\langle(D^L\sigma(s,\eta,\cdot))(\mu_s)(X_s^{\mu}),\Xi_s^{\xi,\varepsilon}\rangle_{\mathscr{C}})+\Upsilon_2^{\varepsilon}(s)\}\Big|_{\eta=X_s^{\mu}}\mathrm{d}W(s)\big\rangle.\end{aligned}$$

By (4.1) and **(B)**, we find a constant $c_1>0$ such that for any $t\in[0,T]$,

D01 (5.11)
$$\mathbb{E}\Big(\sup_{0\le s\le t}\Theta_1^{\varepsilon}(s)\Big)\le c_1\int_0^t\Big\{\mathbb{E}\|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2+\mathbb{E}|\Xi^{\xi,\varepsilon}(s)|\sqrt{\mathbb{E}\|\Xi^{\xi,\varepsilon}(s)\|_{\mathscr{C}}^2}\Big\}\mathrm{d}s\le 2c_1\int_0^t\mathbb{E}\|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s.$$

Next, there exists a constant $c_2 > 0$ such that

D09 (5.12) $$\mathbb{E}\Big(\sup_{0\le s\le t} \Theta_2^\varepsilon(s)\Big) \le c_2 \int_0^t \big\{\mathbb{E}|\Xi^{\xi,\varepsilon}(s)|^2 + \mathbb{E}|\Upsilon_1^\varepsilon(s)|^2 + \mathbb{E}|\Upsilon_2^\varepsilon(s)|^2\big\}\mathrm{d}s, \quad t\in[0,T].$$

Moreover, applying BDG's inequality and using (4.1), we find constants $c_3, c_4 > 0$ such that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t} \Theta_3^\varepsilon(s)\Big) &\le c_3 \mathbb{E}\bigg(\sup_{0\le s\le t} |\Xi^{\xi,\varepsilon}(s)|^2 \int_0^t \Big\{\|(\nabla_{\Xi_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)(X_s^\mu) \\ &\qquad + (\mathbb{E}_{\mathscr{C}^*}\langle (D^L\sigma(s,\eta,\cdot))(\mu_s)(X_s^\mu), \Xi_s^{\xi,\varepsilon}\rangle_{\mathscr{C}}) + \Upsilon_2^\varepsilon(s)\|_{\mathrm{HS}}^2\Big|_{\eta=X_s^\mu}\Big\}\mathrm{d}s\bigg)^{1/2} \\ &\le \frac{1}{2}\mathbb{E}\Big(\sup_{0\le s\le t} |\Xi^{\xi,\varepsilon}(s)|^2\Big) + c_4\int_0^t \{\mathbb{E}\|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2 + \mathbb{E}\|\Upsilon_2^\varepsilon(s)\|_{\mathrm{HS}}^2\}\mathrm{d}s, \quad t\in[0,T].\end{aligned}$$

Substituting this and (5.11), (5.12) into (5.10), and noting that $\Xi_0^{\xi,\varepsilon} = \mathbf{0}$, we find a constant $c > 0$ such that

$$\mathbb{E}\Big(\sup_{0\le s\le t} \|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2\Big) \le c\int_0^t \mathbb{E}\|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s + c\int_0^t \{\mathbb{E}|\Upsilon_1^\varepsilon(s)|^2 + \mathbb{E}\|\Upsilon_2^\varepsilon(s)\|_{\mathrm{HS}}^2\}\mathrm{d}s, \quad t\in[0,T].$$

Since $\mathbb{E}\big(\sup_{0\le s\le t}\|\Xi_s^{\xi,\varepsilon}\|_{\mathscr{C}}^2\big) < \infty$ due to (5.4) and (5.8), Gronwall's inequality yields

JH3 (5.13) $$\mathbb{E}\Big(\sup_{0\le s\le T} |\Xi^{\xi,\varepsilon}(s)|^2\Big) \le c\,\mathrm{e}^{cT}\int_0^T \{\mathbb{E}|\Upsilon_1^\varepsilon(t)|^2 + \mathbb{E}\|\Upsilon_2^\varepsilon(t)\|_{\mathrm{HS}}^2\}\mathrm{d}s.$$

This implies (5.9) by following the argument to deduce (4.11) from (4.18). □

Let $C_p^1(\mathscr{C})$ be the class of functions $f\in C^1(\mathscr{C})$ such that for some constant $c>0$,

CPP (5.14) $$\|\nabla f(\xi)\| \le c\,(1+\|\xi\|_\infty^{p-1}), \quad \xi\in\mathscr{C}.$$

PNN **Proposition 5.3.** *Assume* **(A)** *and* **(B)** *for some $p\ge 2$. For any $T\ge 0$, $f\in C_p^1(\mathscr{C})$ and $\mu\in\mathscr{P}_p(\mathscr{C})$, $(P_Tf)(\mu)$ is L-differentiable w.r.t. $\mu\in\mathscr{P}_p(\mathscr{C})$ and*

$$D_\phi^L(P_Tf)(\mu) = \mathbb{E}_{\mathscr{C}^*}\langle \nabla f(X_T^\mu), \nabla_{\phi(X_0^\mu)} X_T^\mu\rangle_{\mathscr{C}}.$$

Consequently, letting $\Phi:\mathscr{C}\to\mathscr{C}^*$ be a measurable function such that $\Phi(X_0^\mu) = \mathbb{E}(\nabla f(X_T^\mu)|X_0^\mu)$, we have $D^L(P_Tf)(\mu) = \Phi$.

Proof. Let $X_t^{\phi,\mu} = X_t^{\mu\circ(\mathrm{Id}+\phi)^{-1}}$ be the functional solution to (1.1) with initial value $X_0^\mu + \phi(X_0^\mu)$. For any $f\in C_p^1(\mathscr{C})$, by Lemma 5.2, (5.8) and (5.14), we may apply Taylor's expansion to derive that for small $\|\phi\|_{T_{\mu,p}}$,

$$(P_Tf)(\mu\circ(\mathrm{Id}+\phi)^{-1}) - (P_Tf)(\mu) = \mathbb{E}[f(X_T^{\phi,\mu}) - f(X_T^\mu)] = \mathbb{E}_{\mathscr{C}^*}\langle \nabla f(X_T^\mu), \nabla_{\phi(X_0^\mu)} X_T^\mu\rangle_{\mathscr{C}} + \mathrm{o}(\|\phi\|_{T_{\mu,p}}).$$

This implies the desired assertion. □

6 Bismut formula for the L-derivative

In this section, we consider (1.1) with $\sigma(t,\xi,\mu)=\sigma(t,\xi(0))$ dependent only on t and $\xi(0)$, i.e.,

E11 (6.1)
$$\mathrm{d}X(t)=b(t,X_t,\mathscr{L}_{X_t})\mathrm{d}t+\sigma(t,X(t))\mathrm{d}W(t).$$

We aim to investigate the the intrinsic derivative of $(P_tf)(\mu)$, given by (1.2) associated with X_t^μ.

The main results (Theorems 6.2, 6.3 and 6.4 below) of this part generalize those derived in [4] for SDEs with memory and in [26] for McKean-Vlasov SDEs without memory. Going back to the case $r_0=0$ (i.e. without memory), the conditions in Theorems 6.2 and 6.3 are weaker than the corresponding ones used in [26], since the drift b herein is allowed to be non-Lipschitz continuous w.r.t. the space variables. We will first prove a general result and then apply it to establish the Bismut formula for (1.1) with additive and multiplicative noise, respectively.

6.1 A general result

TNN **Theorem 6.1.** *Assume* **(A)** *and* **(B)** *for some $p\ge 2$, and let $T>r_0$. Suppose that for any $\mu\in\mathscr{P}_p(\mathscr{C})$ and $\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$, there exists $h_{\xi,\mu}\in\mathscr{D}(D^*)$, which is adapted when $\sigma(t,\xi,\mu)$ depends on ξ, such that*

OY1 (6.2)
$$w_T^{h_{\xi,\mu}}=\nabla_\xi X_T^\mu,$$

where $\nabla_\xi X_T^\mu$ is in (5.2) and $w_T^{h_{\xi,\mu}}$ solves (4.9) for $h=h_{\xi,\mu}$. Moreover, suppose that for some increasing function $\alpha_T:[0,\infty)\to[0,\infty)$ we have

OY2 (6.3)
$$\mathbb{E}|D^*(h_{\xi,\mu})|^2\le\alpha_T(\|\mu\|_p)(\mathbb{E}\|\xi\|_{\mathscr{C}}^p)^{\frac{2}{p}},\quad \xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P}),\mu\in\mathscr{P}_p(\mathscr{C}).$$

Then the following assertions hold.

(1) *For any $f\in\mathscr{B}_b(\mathscr{C})$,*

OY0 (6.4)
$$|(P_Tf)(\mu)-(P_Tf)(\nu)|\le\sqrt{\alpha_T(\|\mu\|_p\vee\|\nu\|_p)}\|f\|_\infty\mathbb{W}_p(\mu,\nu),\quad \mu,\nu\in\mathscr{P}_p(\mathscr{C}).$$

(2) *For any $f\in C_b^1(\mathscr{C})$, $(P_Tf)(\mu)$ is intrinsically differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$ such that*

OBS (6.5)
$$D_\phi^L(P_Tf)(\mu)=\mathbb{E}\big[f(X_T^\mu)D^*(h_{\phi(X_0^\mu),\mu})\big],\quad \phi\in T_{\mu,p}.$$

Consequently,

OY (6.6)
$$\|D^L(P_Tf)(\mu)\|_{T_{\mu,p}^*}^2\le\alpha_T(\|\mu\|_p)(P_Tf^2)(\mu),\quad \mu\in\mathscr{P}_p(\mathscr{C}).$$

(3) *If moreover*

OY2' (6.7)
$$\lim_{\mathbb{W}_p(\nu,\mu)\to0}\ \sup_{\mathbb{E}\|\xi\|_{\mathscr{C}}^p\in(0,1)}\frac{\mathbb{E}|D^*(h_{\xi,\nu})-D^*(h_{\xi,\mu})|^2}{(\mathbb{E}\|\xi\|_{\mathscr{C}}^p)^{\frac{2}{p}}}=0,\quad \mu\in\mathscr{P}_p(\mathscr{C}),$$

then for any $f\in C_b(\mathscr{C})$, $(P_Tf)(\mu)$ is L-differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$ and (6.6) holds.

Proof. (1) We first consider $f \in C_b^1(\mathscr{C})$. Recall that $X_t^{\varepsilon\xi,\mu}$ is the functional solution to (1.1) with $X_0^{\varepsilon\xi,\mu} := \varepsilon\xi + X_0^\mu$, and $\mu_t^{\xi,\varepsilon} = \mathscr{L}_{X_t^{\varepsilon\xi,\mu}}$. Then, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}f(X_T^{s\xi,\mu}) := \lim_{\varepsilon\to 0}\frac{\mathbb{E}f(X_T^{(s+\varepsilon)\xi,\mu}) - \mathbb{E}f(X_T^{s\xi,\mu})}{\varepsilon} = \nabla_\xi(P_Tf)(\mu^{\xi,s}), \quad s \in [0,1].$$

Then, by applying (6.2) with μ replaced by $\mu^{\xi,s}$ and using Proposition 4.3, we obtain

OYY

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}f(X_T^{s\xi,\mu}) &= \mathbb{E}\big[{}_{\mathscr{C}^*}\langle \nabla f(X_T^{s\xi,\mu}), \nabla_\xi X_T^{s\xi,\mu}\rangle_{\mathscr{C}}\big] \\ &= \mathbb{E}\big[{}_{\mathscr{C}^*}\langle \nabla f(X_T^{s\xi,\mu}), w_T^{h_{\xi,\mu^{\xi,s}}}\rangle_{\mathscr{C}}\big] = \mathbb{E}\big[f(X_T^{s\xi,\mu})D^*(h_{\xi,\mu^{\xi,s}})\big].\end{aligned} \tag{6.8}$$

Whence, one has

LB

$$\begin{aligned}(P_Tf)(\mathscr{L}_{X_0^\mu+\xi}) - (P_Tf)(\mu) &= \mathbb{E}f(X_T^{\xi,\mu}) - \mathbb{E}f(X_T^\mu) = \int_0^1 \Big\{\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}f(X_T^{s\xi,\mu})\Big\}\mathrm{d}s \\ &= \int_0^1 \mathbb{E}\big[f(X_T^{s\xi,\mu})D^*(h_{\xi,\mu^{\xi,s}})\big]\mathrm{d}s, \quad f \in C_b^1(\mathscr{C}).\end{aligned} \tag{6.9}$$

Let

$$\widetilde{\mu}_T(A) = \int_0^1 \mathbb{E}[1_A(X_T^{s\xi,\mu})D^*(h_{\xi,\mu^{\xi,s}})\big]\mathrm{d}s, \quad A \in \mathscr{B}(\mathscr{C}).$$

Since $C_b^1(\mathscr{C})$ is dense in $L^1(\mathscr{L}_{X_T^{\xi,\mu}} + \mathscr{L}_{X_T^\mu} + \widetilde{\mu}_T) \supset \mathscr{B}_b(\mathscr{C})$, (6.9) implies

OY4

$$(P_Tf)(\mathscr{L}_{X_0^\mu+\xi}) - (P_Tf)(\mu) = \int_0^1 \mathbb{E}\big[f(X_T^{s\xi,\mu})D^*(h_{\xi,\mu^{\xi,s}})\big]\mathrm{d}s, \quad f \in \mathscr{B}_b(\mathscr{C}). \tag{6.10}$$

Now, for any $\nu \in \mathscr{P}_p(\mathscr{C})$, let $\xi \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_{X_0^\mu+\xi} = \nu$ and

$$\mathbb{W}_p(\mu,\nu) = \{\mathbb{E}\|\xi\|_{\mathscr{C}}^p\}^{\frac{1}{p}}.$$

We deduce form (6.10) that

$$\begin{aligned}|(P_Tf)(\mu) - (P_Tf)(\nu)| &\le \|f\|_\infty \sup_{s\in[0,1]}\big(\mathbb{E}|D^*(h_{\xi,\mu^{\xi,s}})|^2\big)^{\frac{1}{2}} \\ &\le \|f\|_\infty \mathbb{W}_p(\mu,\nu)\sup_{s\in[0,1]}\sqrt{\alpha_T(\|\mu^{\xi,s}\|_p)}.\end{aligned}$$

Combining this with

$$\begin{aligned}\|\mu^{\xi,s}\|_p &= \{\mathbb{E}\|X_0^\mu + s\xi\|_{\mathscr{C}}^p\}^{\frac{1}{p}} = \{\mathbb{E}\|s(X_0^\mu+\xi) + (1-s)X_0^\mu\|_{\mathscr{C}}^p\}^{\frac{1}{p}} \\ &\le (1-s)\{\mathbb{E}\|X_0^\mu\|_{\mathscr{C}}^p\}^{\frac{1}{p}} + s\{\mathbb{E}\|X_0^\mu+\xi\|_{\mathscr{C}}^p\}^{\frac{1}{p}} \le \|\mu\|_p \vee \|\nu\|_p, \quad s \in [0,1],\end{aligned}$$

we prove (6.4).

(2) Let $f \in C_b^1(\mathscr{C})$, $\mu \in \mathscr{P}_p(\mathscr{C})$ and $\phi \in T_{\mu,p}$. Applying (6.8) with $\xi = \phi(X_0^\mu)$ and $s = 0$, we obtain (6.5), which, together with (6.3), implies

$$|D_\phi^L(P_Tf)(\mu)|^2 \le \alpha_T(\|\mu\|_p)\{\mathbb{E}\|\phi(X_0^\mu)\|_{\mathscr{C}}^p\}^{\frac{2}{p}}\mathbb{E}[f^2(X_T^\mu)] = \alpha_T(\|\mu\|_p)\|\phi\|_{T_{\mu,p}}^2(P_Tf^2)(\mu), \quad \phi \in T_{\mu,p}.$$

Therefore, (6.6) holds true.

(3) Let $f \in C_b(\mathscr{C})$. To prove that $(P_T f)$ is L-differentiable, it suffices to verify

OOY (6.11)
$$I_\mu(\phi) := \frac{|(P_T f)(\mu \circ (\mathrm{Id} + \phi)^{-1}) - (P_T f)(\mu) - \gamma_\phi|}{\|\phi\|_{T_{\mu,p}}} \to 0 \text{ as } \|\phi\|_{T_{\mu,p}} \downarrow 0, \tag{6.11}$$

where

$$\gamma_\phi := \mathbb{E}\big[f(X_T^\mu) D^*(h_{\phi(X_0^\mu),\mu})\big], \quad \phi \in T_{\mu,p}.$$

By (6.10) and the definition of γ_ϕ, it is easy to see that

SCR (6.12)
$$I_\mu(\phi) \le A_\mu(\phi) + B_\mu(\phi) \tag{6.12}$$

holds for

$$\begin{aligned}
A_\mu(\phi) &:= \frac{1}{\|\phi\|_{T_{\mu,p}}} \int_0^1 \mathbb{E}\big[\big|\big\{f(X_T^{s\phi(X_0^\mu),\mu}) - f(X_T^\mu)\big\} D^*(h_{\phi(X_0^\mu),\mu})\big|\big] \mathrm{d}s, \\
B_\mu(\phi) &:= \frac{\|f\|_\infty}{\|\phi\|_{T_{\mu,p}}} \int_0^1 \big(\mathbb{E}[|D^*(h_{\phi(X_0^\mu),\mu\circ(\mathrm{Id}+s\phi)^{-1}}) - D^*(h_{\phi(X_0^\mu),\mu})|^2]\big)^{\frac{1}{2}} \mathrm{d}s.
\end{aligned}$$

Since $f \in C_b(\mathscr{C})$, and (5.4) implies $\mathbb{E}\|X_T^{s\phi(X_0^\mu),\mu} - X_T^\mu\|_{\mathscr{C}}^p \to 0$ as $\|\phi\|_{T_{p,\mu}} \to 0$, it follows from (6.3) and the dominated convergence theorem that

$$\lim_{\|\phi\|_{T_{\mu,p}} \to 0} A_\mu(\phi) = 0.$$

Finally, (6.7) implies $\lim_{\|\phi\|_{T_{\mu,p}} \to 0} B_\mu(\phi) = 0$. Therefore, (6.11) follows from (6.12). $\square$

Remark 6.1 When $r_0 = 0$ (i.e. without memory), the Bismut formula for the L-derivative has been establish in [26] for all $f \in \mathscr{B}_b(\mathscr{C})$, by applying the a prior formula like (6.10) for small $\varepsilon > 0$ replacing T. However, in the present case (6.10) is available merely for $T > r_0$, so that this technique is invalid. So, in Theorem 6.1 we only establish the Bismut formula of the L-derivative for $f \in C_b(\mathscr{C})$.

6.2 Additive noise: non-degenrate case

TN1 **Theorem 6.2.** *Assume* **(A)** *and* **(B)** *for some $p \ge 2$, and consider* (1.1) *with $\sigma(t, \xi, \mu) = \sigma(t)$ independent of (ξ, μ) such that $(\sigma\sigma^*)(t)$ is invertible with $(\sigma\sigma^*)^{-1}(t)$ locally bounded in t.*

(1) *There exist an increasing function $C : [r_0, \infty) \to [0, \infty)$ and a constant $c > 0$ such that for any $T > r_0$, $f \in \mathscr{B}_b(\mathscr{C})$, and $\mu, \nu \in \mathscr{P}_p(\mathscr{C})$,*

DL1 (6.13)
$$\begin{aligned}
|(P_T f)(\mu) - (P_T f)(\nu)| \le C(T)\|f\|_\infty \Big\{ & 1 + (T - r_0)^{-1} + K_2(c(1 + \|\mu\|_p + \|\nu\|_p)) \\
& + (\|\mu\|_p + \|\nu\|_p)^{\frac{(p-2)^+}{2}} \Big\} \mathbb{W}_p(\mu, \nu).
\end{aligned} \tag{6.13}$$

(2) *For any $T > r_0$ and $f \in C_b(\mathscr{C})$, $(P_T f)(\mu)$ is L-differentiable in $\mu \in \mathscr{P}_p(\mathscr{C})$ such that*

WW1 (6.14)
$$D_\phi^L(P_T f)(\mu) = -\mathbb{E}\bigg(f(X_T^\mu) \int_0^T \langle \{\sigma^*(\sigma\sigma^*)^{-1}\}(t) H^\phi(t), \mathrm{d}W(t) \rangle \bigg), \quad \phi \in T_{\mu,p} \tag{6.14}$$

holds for

$$H^{\phi}(t) := (\nabla_{Z_t} b)(t,\cdot,\mu_t)(X_t^\mu) + (\mathbb{E}[{}_{\mathscr{C}^*}\langle D^L b(t,\xi,\cdot)(\mu_t)(X_t^\mu), Z_t\rangle_{\mathscr{C}}])|_{\xi=X_t^\mu} + \frac{\phi(X_0^\mu)(0)1_{[0,T-r_0]}(t)}{T-r_0},$$

where $\mu_t := \mathscr{L}_{X_t^\mu}$ *and* $(Z_t)_{t\geq 0}$ *is the segment of* $(Z(t))_{t\geq -r_0}$ *given by*

$$Z(t) := \begin{cases} \phi(X_0^\mu)(t), & \text{if } t\in[-r_0,0],\\ \frac{(T-r_0-t)^+}{T-r_0}\phi(X_0^\mu)(0), & \text{if } t\geq 0.\end{cases}$$

Consequently, there exist an increasing function $C:[r_0,\infty)\to(0,\infty)$ *and a constant* $c>0$ *such that*

NM

$$\|D^L(P_Tf)(\mu)\|_{T^*_{\mu,p}} \leq C(T)\big\{1+(T-r_0)^{-1}+K_2(c(1+\|\mu\|_p))+\|\mu\|_p^{\frac{p-2}{2}}\big\}\{(P_Tf^2)(\mu)\}^{\frac12} \tag{6.15}$$

holds for all $T>r_0$, $f\in C_b(\mathscr{C})$ *and* $\mu\in\mathscr{P}_p(\mathscr{C})$.

Proof. To apply Theorem 6.1, for any $\mu\in\mathscr{P}_p(\mathscr{C})$ and $\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$, let

AAP

$$h_{\xi,\mu}(t) := -\int_0^t \{\sigma^*(\sigma\sigma^*)^{-1}\}(s)H^{\xi,\mu}(s)\mathrm{d}s,\quad t\in[0,T], \tag{6.16}$$

where

APP@

$$\begin{aligned} H^{\xi,\mu}(t) &:= (\nabla_{Z_t^\xi} b)(t,\cdot,\mu_t)(X_t^\mu) + (\mathbb{E}[{}_{\mathscr{C}^*}\langle D^L b(t,\eta,\cdot)(\mu_t)(X_t^\mu), Z_t^\xi\rangle_{\mathscr{C}}])|_{\eta=X_t^\mu}\\ &\qquad + \frac{\xi(0)1_{[0,T-r_0]}(t)}{T-r_0},\\ Z^\xi(t) &:= \xi(t)1_{[-r_0,0]}(t) + \frac{(T-r_0-t)^+}{T-r_0}\xi(0)1_{(0,\infty)}(t). \end{aligned} \tag{6.17}$$

By **(B)**, the boundedness of $(\sigma\sigma^*)^{-1}(t)$ in $t\in[0,T]$, and the definition of $H^{\xi,\mu}(t)$, we find constants $c_1=c_1(T)>0$ increasing in T such that

*DLP

$$|\dot h_{\xi,\mu}(t)|^2 \leq c_1\|\xi\|_{\mathscr{C}}^2\big\{(T-r_0)^{-2}+\|X_t^\mu\|_{\mathscr{C}}^{p-2}+K_2(\|\mu_t\|_p)^2\big\},\quad t\in[0,T]. \tag{6.18}$$

Note that (3.1) and $\mu\in\mathscr{P}_p(\mathscr{C})$ imply

$$\sup_{t\in[0,T]}\|\mu_t\|_p \leq c\,(1\vee\|\mu\|_p)$$

for some constant $c=c(T)>0$ increasing in T. This, combining (3.1), (4.24) with (6.18), yields

*DD2

$$\begin{aligned} \mathbb{E}|D^*(h_{\xi,\mu})|^2 &= \mathbb{E}\int_0^T |\dot h_{\xi,\mu}(t)|^2\mathrm{d}t\\ &\leq c_2(\mathbb{E}\|\xi\|_{\mathscr{C}}^p)^{\frac2p}\big\{(T-r_0)^{-2}+(\mathbb{E}\|X_t^\mu\|_{\mathscr{C}}^p)^{(p-2)/p}+K_2(c\,(1\vee\|\mu\|_p))^2\big\}\\ &\leq c_3(\mathbb{E}\|\xi\|_{\mathscr{C}}^p)^{\frac2p}\big\{1+(T-r_0)^{-2}+\|\mu\|_p^{p-2}+K_2(c\,(1\vee\|\mu\|_p))^2\big\}<\infty \end{aligned} \tag{6.19}$$

for some constants $c_2,c_3>0$ increasing in T.

Note that $(Z_t^\xi)_{t\in[0,T]}$ is the functional solution to the SDE with memory

PL
$$\begin{aligned}\mathrm{d}Z^\xi(t) =&\Big\{(\nabla_{Z_t^\xi} b)(t,\cdot,\mu_t)(X_t^\mu) + \sigma(t)\dot h_{\xi,\mu}(t)\\ &+ (\mathbb{E}[_{\mathscr{C}^*}\langle D^L b(t,\eta,\cdot)(\mu_t)(X_t^\mu), Z_t^\xi\rangle_{\mathscr{C}}])|_{\eta=X_t^\mu}\Big\}\mathrm{d}t, \quad t\in[0,T],\ Z_0^\xi=\xi.\end{aligned}\tag{6.20}$$

On the other hand, by Lemma 4.2 and Lemma 5.2, the process

$$\nabla_\xi X^\mu(t) - w^{h_{\xi,\mu}}(t),\quad t\in[0,T]$$

also solves (6.20) with the same initial value ξ. By the uniqueness of (6.20) and $Z_T^\xi=\mathbf{0}$, we derive $\nabla_\xi X_T^\mu = w_T^{h_{\xi,\mu}}$, that is, (6.2) holds. Moreover, (3.2) implies

$$\mathbb{W}_p(\mu_t,\nu_t)\le c\,\mathbb{W}_p(\mu,\nu),\quad t\in[0,T]$$

for some constant $c>0$, where $\nu_t := \mathscr{L}_{X_t^\nu}$, so that (6.16), (6.17) and the continuity of $b(t,\xi,\mu)$ in μ imply (6.7). Therefore, the desired assertions follow from Theorem 6.1 and (6.19). $\square$

6.3 Additive noise: a degenerate case

As generalizations to the stochastic Hamiltonian system [17] and the counterpart with memory [5] as well as the distribution dependent model [26], we consider the following distribution-path dependent stochastic Hamiltonian system for $X(t)=(X^{(1)}(t),X^{(2)}(t))$ on $\mathbb{R}^{l+m}:=\mathbb{R}^l\times\mathbb{R}^m$, which goes back to (1.1) for $d=l+m$:

E5
$$\begin{cases}\mathrm{d}X^{(1)}(t) = b^{(1)}(t,X(t))\mathrm{d}t,\\ \mathrm{d}X^{(2)}(t) = b^{(2)}(t,X_t,\mathscr{L}_{X_t})\mathrm{d}t + \sigma(t)\mathrm{d}W_t,\end{cases}\tag{6.21}$$

where $(W(t))_{t\ge0}$ is an m-dimensional Brownian motion on a complete filtration probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbb{P})$, for each $t\ge0$, $\sigma(t)$ is an invertible $m\times m$-matrix, and

$$b=(b^{(1)},b^{(2)}):[0,\infty)\times\mathscr{C}\times\mathscr{P}_p(\mathscr{C})\to\mathbb{R}^{l+m}$$

is measurable with $b^{(1)}(t,\xi,\mu)=b^{(1)}(t,\xi(0))$ dependent only on t and $\xi(0)$. Let $\nabla=(\nabla^{(1)},\nabla^{(2)})$ be the gradient operator on $\mathbb{R}^{l+m}$, where $\nabla^{(i)}$ stands for the gradient operator w.r.t. the i-th component, $i=1,2$. Let $\nabla^2=\nabla\nabla$ denote the Hessian operator on $\mathbb{R}^{l+m}$. We assume

(H1) For every $t\ge0$, $b^{(1)}(t,\cdot)\in C^2(\mathbb{R}^{l+m}\to\mathbb{R}^l)$, $b^{(2)}(t,\xi,\mu)$ is C^1 in both $\xi\in\mathscr{C}$ and $\mu\in\mathscr{P}_p(\mathscr{C})$, and there exists an increasing function $K:[0,\infty)\to[0,\infty)$ such that

$$\begin{aligned}\|(\nabla b^{(1)})(t,\cdot,\mu)(\xi(0))\| + \|(\nabla^2 b^{(1)})(t,\cdot)(\xi(0))\| + \|(\nabla b^{(2)})(t,\cdot,\mu)(\xi)\|\\ + \|D^L b^{(2)}(t,\xi,\cdot)(\mu)\|_{T_{\mu,p}^*}\le K(t)\end{aligned}$$

holds for all $t\ge0$ and $(\xi,\mu)\in\mathscr{C}\times\mathscr{P}_p(\mathscr{C})$.

Obviously, the assumption **(H1)** implies **(A)** and **(B)** for the SDE (6.21).

For any $\mu\in\mathscr{P}_p(\mathscr{C})$, let $(X_t^\mu)_{t\ge0}$ be the functional solution to (6.21) with $\mathscr{L}_{X_0^\mu}=\mu$, and denote $\mu_t=\mathscr{L}_{X_t^\mu}$ as before. To establish the Bismut formula for the L-derivative of $(P_Tf)(\mu):=\mathbb{E}f(X_T^\mu)$, we shall follow the line of [26, 37], where the case without memory was investigated. To establish the Bismut formula, we need the following assumption **(H2)**, which implies the hypoellipticity.

(H2) There exist an $l \times m$-matrix B and some constant $\varepsilon \in (0,1)$ such that

B (6.22)
$$\langle ((\nabla^{(2)} b^{(1)})(t,\cdot) - B) B^* a, a\rangle \geq -\varepsilon |B^* a|^2, \quad \forall a \in \mathbb{R}^l.$$

Moreover, there exists an increasing function $\theta_\cdot \in C([0, T-r_0]; \mathbb{R}_+)$ such that

B2 (6.23)
$$\int_0^t s(T-r_0-s) K_{T-r_0,s} B B^* K^*_{T-r_0,s} \mathrm{d}s \geq \theta_t I_{l\times l}, \quad t \in [0, T-r_0],$$

where, for any $s \geq 0$, $(K_{t,s})_{t\geq s}$ solves the following linear random ODE on $\mathbb{R}^l \otimes \mathbb{R}^l$:

Eq1 (6.24)
$$\frac{\mathrm{d}}{\mathrm{d}t} K_{t,s} = (\nabla^{(1)} b^{(1)})(t, X(t)) K_{t,s}, \quad t \geq s, K_{s,s} = I_{l\times l}$$

with $I_{l\times l}$ being the $l \times l$ identity matrix.

Specific examples for $b^{(1)}$ satisfying **(H2)** are included in [26, Example 2.1]. Let $T > r_0$. According to the proof of [37, Theorem 1.1], **(H2)** implies that the $l \times l$ matrices

$$Q_t := \int_0^t s(T-r_0-s) K_{T-r_0,s} (\nabla^{(2)} b^{(1)})(s, X^\mu(s)) B^* K^*_{T-r_0,s} \mathrm{d}s, \quad t \in (0, T-r_0]$$

are invertible with

Q (6.25)
$$\|Q_t^{-1}\| \leq \frac{1}{(1-\varepsilon)\theta(t)}, \quad t \in (0, T-r_0].$$

To apply Theorem 6.1, for any $\xi = (\xi^{(1)}, \xi^{(2)}) \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$, we need to construct $h_{\xi,\mu} \in \mathscr{D}(D^*)$ such that (6.2) holds. To this end, as in [26], where $r_0 = 0$ is concerned, we take the $\mathscr{C}$-valued process $\alpha_t = (\alpha_t^{(1)}, \alpha_t^{(2)})$, which is the segment of $\alpha(t)$ defined by $\alpha(t) = \xi(t)$ for $t \in [-r_0, 0]$ and

aa (6.26)
$$\begin{aligned}
\alpha^{(2)}(t) &:= \frac{(T-r_0-t)^+}{T-r_0} \xi^{(2)}(0) - \frac{t(T-r_0-t)^+ B^* K^*_{T-r_0,t}}{\int_0^{T-r_0} \theta_s^2 \mathrm{d}s} \int_t^{T-r_0} \theta_s^2 Q_s^{-1} K_{T-r_0,0} \xi^{(1)}(0) \mathrm{d}s \\
&\quad - t(T-r_0-t)^+ B^* K^*_{T-r_0,t} Q^{-1}_{T-r_0} \int_0^{T-r_0} \frac{T-r_0-s}{T-r_0} K_{T-r_0,s} \Big(\nabla^{(2)}_{\xi^{(2)}(0)} b^{(1)}\Big)(s, X^\mu(s)) \mathrm{d}s, \\
\alpha^{(1)}(t) &:= 1_{[0,T-r_0]}(t) \bigg(K_{t,0} \xi^{(1)}(0) + \int_0^t K_{t,s} \Big(\nabla^{(2)}_{\alpha^{(2)}(s)} b^{(1)}\Big)(s,\cdot)(X^\mu(s))\, \mathrm{d}s \bigg), \quad t \geq 0.
\end{aligned}$$

Now, let $(h_{\xi,\mu}(t), w^{h_{\xi,\mu}}(t))_{t\in[0,T]}$ be the unique solution to the random ODEs

B00 (6.27)
$$\begin{aligned}
\dot h_{\xi,\mu}(t) &:= \frac{\mathrm{d} h_{\xi,\mu}(t)}{\mathrm{d}t} = \sigma(t)^{-1} \Big\{ (\nabla_{\alpha_t} b^{(2)})(t,\cdot,\mu_t)(X_t^\mu) - \dot\alpha^{(2)}(t) \\
&\qquad + \big(\mathbb{E}_{\mathscr{C}^*} \langle D^L b^{(2)}(t,\eta,\cdot)(\mu_t)(X_t^\mu), \alpha_t + w_t^{h_{\xi,\mu}} \rangle_{\mathscr{C}}\big)\big|_{\eta = X_t^\mu} \Big\}, \\
\frac{\mathrm{d} w^{h_{\xi,\mu}}(t)}{\mathrm{d}t} &= \Big(\Big(\nabla_{w^{h_{\xi,\mu}}(t)} b^{(1)}\Big)(t, X^\mu(t)), \Big(\nabla_{w_t^{h_{\xi,\mu}}} b^{(2)}\Big)(t,\cdot,\mu_t)(X_t^\mu) + \sigma(t) \dot h_{\xi,\mu}(t) \Big), \\
h_{\xi,\mu}(0) &= \mathbf{0} \in \mathbb{R}^m, \quad w_0^{h_{\xi,\mu}} = \mathbf{0} \in \mathscr{C}.
\end{aligned}$$

Let $u^\xi(t)=((u^\xi)^{(1)}(t),(u^\xi)^{(2)}(t))=\alpha(t)+w^{h_{\xi,\mu}}(t), t\geq -r_0$. Then, (6.27) implies

$$(u^\xi)^{(2)}(t)=\alpha^{(2)}(0)+\int_0^t\Big\{(\nabla_{u_t^\xi}b^{(2)})(s,\cdot,\mu_s)(X_s^\mu)+\big(\mathbb{E}_{\mathscr{C}^*}\langle D^Lb^{(2)}(s,\eta,\cdot)(\mu_s)(X_s^\mu),v_s^\xi\rangle_{\mathscr{C}}\big)\big|_{\eta=X_s^\mu}\Big\}\mathrm{d}s.$$

Furthermore, we have

$$\begin{aligned}(u^\xi)^{(1)}(t)&=\alpha^{(1)}(t)+\int_0^t(\nabla_{w^{h_{\xi,\mu}}(s)}b^{(1)})(s,X^\mu(s))\mathrm{d}s\\&=\alpha^{(1)}(t)-\int_0^t(\nabla_{\alpha(s)}b^{(1)})(s,X^\mu(s))\mathrm{d}s+\int_0^t(\nabla_{u^\xi(s)}b^{(1)})(s,X^\mu(s))\mathrm{d}s\\&=\alpha^{(1)}(0)+\int_0^t(\nabla_{u^\xi(s)}b^{(1)})(s,X^\mu(s))\mathrm{d}s,\end{aligned}$$

where in the last identity we used

$$\mathrm{d}\alpha^{(1)}(t)=(\nabla_{\alpha(s)}b^{(1)})(t,X^\mu(t))\mathrm{d}t,$$

see the proof of [26, Theorem 2.3] for more details. Moreover, the equation (5.7) for $v^\xi(t)=((v^\xi)^{(1)}(t),(v^\xi)^{(2)}(t))$ associated with the present SDE (6.21) becomes

$$\begin{aligned}&\frac{d}{\mathrm{d}t}(v^\xi)^{(2)}(t)=(\nabla_{v_t^\xi}b^{(2)})(t,\cdot,\mu_t)(X_t^\mu)+\big(\mathbb{E}_{\mathscr{C}^*}\langle D^Lb^{(2)}(t,\eta,\cdot)(\mu_t)(X_t^\mu),v_t^\xi\rangle_{\mathscr{C}}\big)\big|_{\eta=X_t^\mu},\\&\frac{d}{\mathrm{d}t}(v^\xi)^{(1)}(t)=(\nabla_{v^\xi(t)}b^{(1)})(t,\cdot)(X^\mu(t)),\quad v_0^\xi=\xi.\end{aligned}$$

Hence, the uniqueness of this equation implies

bb2 (6.28)
$$v^\xi(t)=w^{h_{\xi,\mu}}(t)+\alpha(t),\quad t\geq 0.$$

Obviously, $\alpha^{(2)}(t)=\mathbf{0}$ for $t\geq T-r_0$. On the other hand, inserting the expression of $\alpha^{(2)}(t)$ into $\alpha^{(1)}(T-r_0)$, taking the definition of Q_t and changing the order of integral yields $\alpha^{(1)}(T-r_0)=\mathbf{0}$, which further implies $\alpha^{(1)}(t)=\mathbf{0}$, $t\geq T-r_0$, according to the definition of $\alpha^{(1)}$. Hence, we arrive at $\alpha(t)=\mathbf{0}$ for $t\geq T-r_0$. This, combining Lemma 5.2 with (6.28), leads to

$$\nabla_\xi X_T^\mu=v_T^\xi=w_T^{h_{\xi,\mu}},$$

that is, (6.2) holds. Moreover, as shown in the proof of [37, Theorem 1.1] that $h_{\xi,\mu}\in\mathscr{D}(D^*)$ satisfying (6.7), and for small $T-r_0>0$, $\mathbb{E}|D^*(h_{\xi,\mu})|^2$ has the same order as $\mathbb{E}\int_0^{T-r_0}|\dot h_{\xi,\mu}(t)|^2\mathrm{d}t$, so that according to the construction of $h_{\xi,\mu}$ we have

$$\mathbb{E}|D^*(h_{\xi,\mu})|^2\leq\frac{C(T)(T-r_0)^4}{\int_0^{T-r_0}\theta_s^2\mathrm{d}s},\quad T>0,\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P}),\mu\in\mathscr{P}_p(\mathscr{C})$$

for some increasing function $C:[r_0,\infty)\to[0,\infty)$. Therefore, by Theorem 6.1, we have the following result.

T4.2 **Theorem 6.3.** *Assume* **(H1)** *and* **(H2)** *for some* $p\geq 2$.

(1) *There exists an increasing function $C:[r_0,\infty)\to[0,\infty)$ such that for any $T>r_0$, $f\in\mathscr{B}_b(\mathscr{C})$,*

$$|(P_Tf)(\mu)-(P_Tf)(\nu)|\le C(T)(T-r_0)^2\bigg(\int_0^{T-r_0}\theta_s^2\mathrm{d}s\bigg)^{-\frac{1}{2}}\|f\|_\infty\mathbb{W}_p(\mu,\nu),\quad \mu,\nu\in\mathscr{P}_p(\mathscr{C}).$$

(2) *For any $T>r_0$ and $f\in C_b(\mathscr{C})$, $(P_Tf)(\mu)$ is L-differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$ such that*

$$D_\phi^L(P_Tf)(\mu)=-\mathbb{E}\big[f(X_T^\mu)D^*(h_{\phi(X_0^\mu),\mu})\big],\quad \phi\in T_{\mu,p},$$

and there exists an increasing function $C:[r_0,\infty)\to(0,\infty)$ such that for any $f\in C_b(\mathscr{C}),T>r_0$ and $\mu\in\mathscr{P}_p(\mathscr{C})$,

$$\|D^L(P_Tf)(\mu)\|_{T_{\mu,p}^*}\le C(T)(T-r_0)^2\bigg(\int_0^{T-r_0}\theta_s^2\mathrm{d}s\bigg)^{-\frac{1}{2}}\{(P_Tf^2)(\mu)\}^{\frac{1}{2}}.$$

6.4 Multiplicative noise

In this subsection, we assume $\sigma(t,\xi,\mu)=\sigma(t,\xi(0))$. Following the line of [4] due to the idea of [32], for any $\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$ we consider the SDE with memory

WW3 (6.29)
$$\begin{aligned}\mathrm{d}U^\xi(t)=&\Big\{(\nabla_{U_t^\xi}b)(t,\cdot,\mu_t)(X_t^\mu)+(\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(t,\eta,\cdot)(\mu_t)(X_t^\mu),U_t^\xi\rangle_{\mathscr{C}})\big|_{\eta=X_t^\mu}\\&-\frac{U^\xi(t)}{T-r_0-t}\Big\}1_{[0,T-r_0)}(t)\mathrm{d}t+\{(\nabla_{U^\xi(t)}\sigma)(t,\cdot)(X^\mu(t))\}\mathrm{d}W(t),\quad U_0^\xi=\xi.\end{aligned}$$

Then, due to (4.1), the SDE (6.29) has a unique solution for $t<T-r_0$. By repeating the proofs of [4, Lemma 2.1 and Theorem 1.2(1)], we have

BJH (6.30)
$$\int_0^{T-r_0}\frac{\mathbb{E}|U^\xi(t)|^2}{(T-r_0-t)^2}\mathrm{d}t+\mathbb{E}\Big(\sup_{t\in[0,T-r_0)}\|U_t^\xi\|_{\mathscr{C}}^p\Big)\le\frac{C(T)}{T-r_0}\{\mathbb{E}\|\xi\|_{\mathscr{C}}^p\}^{\frac{2}{p}}$$

for some increasing function $C:[r_0,\infty)\to[0,\infty)$, so that we may extend $U^\xi(t)$ for $t\in[0,T]$ by setting

BJH2 (6.31)
$$U^\xi(t)=\mathbf{0},\quad t\in[T-r_0,T],$$

which obviously solves (6.29) up to time T.

TN2 **Theorem 6.4.** *Assume* **(A)** *and* **(B)** *for some $p\ge2$. Let $\sigma(t,\xi,\mu)=\sigma(t,\xi(0))$ depend only on t and $\xi(0)$ such that, for each $x\in\mathbb{R}^d$, $(\sigma\sigma^*)(t,x)$ is invertible with $\sup_{x\in\mathbb{R}^d}\|(\sigma\sigma^*)^{-1}\|(t,x)$ locally bounded in t. Then,*

(1) *There exists an increasing function $C:[r_0,\infty)\to[0,\infty)$ such that for any $T>r_0$, $f\in\mathscr{B}_b(\mathscr{C})$, and $\mu,\nu\in\mathscr{P}_p(\mathscr{C})$,*

DL1 (6.32)
$$|(P_Tf)(\mu)-(P_Tf)(\nu)|\le C(T)\|f\|_\infty\mathbb{W}_p(\mu,\nu).$$

(2) *For any $T>r_0$ and $f\in C_b(\mathscr{C})$, $(P_Tf)(\mu)$ is L-differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$ such that*

$$\boxed{\texttt{WW1}}\quad (6.33)\qquad D^L_\phi(P_Tf)(\mu) = -\mathbb{E}\bigg(f(X^\mu_T)\int_0^T\langle\{\sigma^*(\sigma\sigma^*)^{-1}\}(t)H^\phi(t),\mathrm{d}W(t)\rangle\bigg),\quad \phi\in T_{\mu,p}$$

holds for

$$\begin{aligned}H^\phi(t) :=& \Big\{(\nabla_{U^\xi_t}b)(t,\cdot,\mu_t)(X^\mu_t) + (\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(t,\eta,\cdot)(\mu_t)(X^\mu_t),U^\xi_t\rangle_{\mathscr{C}})\big|_{\eta=X^\mu_t}\Big\}1_{[T-r_0,T]}(t)\\ &+\frac{U^\xi(t)}{T-r_0-t}1_{[0,T-r_0)}(t),\quad t\in[0,T].\end{aligned}$$

Consequently, there exists an increasing function $C:[r_0,\infty)\to(0,\infty)$ such that

$$\boxed{\texttt{NM}}\quad (6.34)\qquad \|D^L(P_Tf)(\mu)\|_{T^*_{\mu,p}}\le C(T)\{(P_Tf^2)(\mu)\}^{\frac12}$$

holds for all $T>r_0,\ f\in C_b(\mathscr{C})$ and $\mu\in\mathscr{P}_p(\mathscr{C})$.

Proof. To apply Theorem 6.1, for any $\mu\in\mathscr{P}_p(\mathscr{C})$ and $\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$, let

$$\boxed{\texttt{AP}}\quad (6.35)\qquad h_{\xi,\mu}(t)=\int_0^t\{\sigma^*(\sigma\sigma^*)^{-1}\}(s,X^\mu(s))G^\xi(s)\mathrm{d}s,\quad t\in[0,T],$$

where

$$\begin{aligned}G^\xi(t) :=& \Big\{(\nabla_{U^\xi_t}b)(t,\cdot,\mu_t)(X^\mu_t) + (\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(t,\eta,\cdot)(\mu_t)(X^\mu_t),U^\xi_t\rangle_{\mathscr{C}})\big|_{\eta=X^\mu_t}\Big\}1_{[T-r_0,T]}(t)\\ &+\frac{U^\xi(t)}{T-r_0-t}1_{[0,T-r_0)}(t),\quad t\in[0,T].\end{aligned}$$

Then, h is adapted and, by (6.30), we find some increasing function $C:[r_0,\infty)\to(0,\infty)$ such that

$$\boxed{\texttt{HBJ}}\quad (6.36)\qquad \mathbb{E}\int_0^T|\dot h_{\xi,\mu}(t)|^2\mathrm{d}t\le\frac{C(T)}{T-r_0}\{\mathbb{E}\|\xi\|^p_{\mathscr{C}}\}^{\frac2p},\quad T>r_0,\mu\in\mathscr{P}_p(\mathscr{C}),\xi\in L^p(\Omega\to\mathscr{C},\mathscr{F}_0,\mathbb{P})$$

so that (6.3) holds true. Moreover, by the regularities of b and σ ensured by **(B)**, the condition (6.7) holds. Therefore, according to Theorem 6.1, it remains to verify (6.2). By (6.29), Lemma 4.2 and Lemma 5.2, we see that both $U^\xi(t)$ and $\nabla_\xi X^\mu_t - w^{h_{\xi,\mu}}(t)$ solve the SDE with memory

$$\begin{aligned}\mathrm{d}Z(t) =& \Big\{(\nabla_{Z_t}b)(t,\cdot,\mu_t)(X^\mu_t)-\sigma(t,X^\mu(t))\dot h(t)\Big\}\mathrm{d}t+\{(\nabla_{Z(t)}\sigma)(t,\cdot)(X^\mu(t))\}\mathrm{d}W(t)\\ &+\Big\{(\mathbb{E}[_{\mathscr{C}^*}\langle D^Lb(t,\eta,\cdot)(\mu_t)(X^\mu_t),Z_t\rangle_{\mathscr{C}}])\big|_{\eta=X^\mu_t}\Big\}\mathrm{d}t,\quad Z_0=\xi,t\in[0,T].\end{aligned}$$

By the uniqueness of solution to this equation and (6.31), we obtain (6.2) and hence finish the proof. □

7 Asymptotic Bismut formula for the L-derivative

In this section, we aim to extend the asymptotic Bismut formula derived in [22] for SDEs with memory to that on the L-derivative for distribution-path dependent SDEs. Even coming back to SDEs with memory, our conditions are slightly weaker since we allow the drift term to be non-Lipschitz continuous.

7.1 The non-degenerate setup

In this subsection, we assume that $\sigma(t,\xi,\mu)=\sigma(t,\xi)$ depends only on $t\ge 0$ and $\xi\in\mathscr{C}$, i.e., the diffusion σ is path-dependent but independent of the measure variable μ. For any $\lambda\ge 0,\mu\in\mathscr{P}_p(\mathscr{C})$ and $\phi\in T_{\mu,p}$, consider the following SDE with memory

E10
$$\begin{aligned}&\mathrm{d}Z^{\mu,\phi,\lambda}(t)=\{(\nabla_{Z_t^{\mu,\phi,\lambda}}b)(t,\cdot,\mu_t)(X_t^\mu)-\lambda Z^{\mu,\phi,\lambda}(t)\}\mathrm{d}t+(\nabla_{Z_t^{\mu,\phi,\lambda}}\sigma)(t,\cdot)(X_t^\mu)\mathrm{d}W(t),\\&Z_0^{\mu,\phi,\lambda}=\phi(X_0^\mu),\ t\ge 0.\end{aligned}\tag{7.1}$$

According to [29, Theorem 2.3], (4.1) implies that (7.1) has a unique functional solution $(Z_t^{\mu,\phi,\lambda})_{t\ge0}$ such that

B1
$$\mathbb{E}\Big(\sup_{0\le s\le t}\|Z_s^{\mu,\phi,\lambda}\|_{\mathscr{C}}^p\Big)<\infty,\quad t>0,\ \phi\in T_{\mu,p},\ \lambda>0.\tag{7.2}$$

th2 **Theorem 7.1.** *Assume* **(A)** *and* **(B)** *for some $p\ge 2$ such that (4.1) holds for some constant K uniformly in $T>0$. Moreover, suppose that $(\sigma\sigma^*)(t,\xi)$ is invertible with $\sup_{\xi\in\mathscr{C}}\|(\sigma\sigma^*)^{-1}\|(t,\xi)$ locally bounded in t.*

(1) *For any $T>0$ and $f\in C_p^1(\mathscr{C})$, $(P_Tf)(\mu)$ is L-differentiable in $\mu\in\mathscr{P}_p(\mathscr{C})$, such that for any $\mu\in\mathscr{P}_p(\mathscr{C}),\phi\in T_{\mu,p}$ and $f\in C_p^1(\mathscr{C})$,*

J1
$$D_\phi^L(P_Tf)(\mu)=\mathbb{E}\bigg(f(X_T^\mu)\int_0^T\langle\dot h^{\mu,\phi,\lambda}(t),\mathrm{d}W(t)\rangle\bigg)+\mathbb{E}(\nabla_{Z_T^{\mu,\phi,\lambda}}f)(X_T^\mu),\quad\lambda\ge 0,\tag{7.3}$$

where

0F0
$$\begin{aligned}h^{\mu,\phi,\lambda}(t):=\int_0^t&\big\{\sigma^*(\sigma\sigma^*)^{-1}\big\}(s,X_s^\mu)^{-1}\Big\{(\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(s,\xi,\cdot)(\mu_s)(X_s^\mu),D_\phi^LX_s^\mu\rangle_{\mathscr{C}})\Big|_{\xi=X_s^\mu}\\&+\lambda Z^{\mu,\phi,\lambda}(s)\Big\}\mathrm{d}s,\quad t\ge 0.\end{aligned}\tag{7.4}$$

(2) *If either $p>4$ or $p>2$ but $\|\nabla b(t,\cdot,\mu)(\xi)\|$ is bounded, then for any $\delta>0$ there exist constants $c,\lambda_0>0$ such that*

J4
$$\begin{aligned}&\bigg|D_\phi^L(P_Tf)(\mu)-\mathbb{E}\bigg(f(X_T^\mu)\int_0^T\langle\dot h^{\mu,\phi,\lambda}(s),\mathrm{d}W(s)\rangle\bigg)\bigg|\\&\le c\mathrm{e}^{-\delta T}\big\{(P_T\|\nabla f\|^{\frac{p}{p-1}})(\mu)\big\}^{\frac{p-1}{p}}\|\phi\|_{T_{\mu,p}},\quad\lambda\ge\lambda_0,T>0,\quad f\in C_p^1(\mathscr{C}),\end{aligned}\tag{7.5}$$

(2) *If $p\in[1,2]$ and $\|\nabla\sigma(t,\cdot)(\xi)\|\le\theta$ for some constant $\theta>0$ and all $(t,\xi)\in[0,\infty)\times\mathscr{C}$ such that*

DFF
$$\theta^2<\sup_{\alpha>0}\frac{\alpha}{p\mathrm{e}^{\alpha r_0}+4p^2\mathrm{e}^{2\alpha r_0}},\tag{7.6}$$

then there exists constants $c,\lambda_0>0$ such that (7.5) holds.

To prove this result, we present the following two lemmas, where the first is due to [13, Lemma 2.2], (7.8).

LBJJ **Lemma 7.2.** *Let M_t be a continuous real martingale with $\mathrm{d}\langle M\rangle_t = g(t)\mathrm{d}t$, and let*

$$F_\alpha(t) := \int_0^t \mathrm{e}^{-\alpha(t-s)}\mathrm{d}M_s, \quad t\ge 0, \alpha>0.$$

Then for any $p>2$, there exists a function $h:[0,\infty)\to[0,\infty)$ with $\kappa_p(\alpha)\to 0$ as $\alpha\to\infty$ such that

$$\mathbb{E}\bigg[\sup_{s\in[0,t]}|F_\alpha(s)|^p\bigg] \le r_\alpha \mathbb{E}\int_0^t g(s)^{\frac{p}{2}}\mathrm{d}s, \quad t\ge 0.$$

Consequently, for any progress measurable process $A(t)$ on $\mathbb{R}^d\otimes\mathbb{R}^m$,

$$\mathbb{E}\bigg[\sup_{s\in[0,t]}\bigg|\int_0^t \mathrm{e}^{-\alpha(t-s)}A(s)\mathrm{d}W(s)\bigg|^p\bigg] \le d^{p-1} r_\alpha \mathbb{E}\int_0^t \|A(s)\|^p\mathrm{d}s, \quad t\ge 0.$$

Lem6 **Lemma 7.3.** *Assume* **(A)** *and* **(B)** *for some $p\ge 2$ such that (4.1) holds for some constant K uniformly in $T>0$.*

(1) *If either $p>4$ or $p>2$ but $\|\nabla b(t,\cdot,\mu)(\xi)\|$ is bounded, then for any $\delta>0$, there exist constants $c,\lambda_0>0$ such that*

F01 (7.7) $$\mathbb{E}[\|Z_t^{\mu,\phi,\lambda}\|_{\mathscr{C}}^p] \le c\mathrm{e}^{-\delta t}\|\phi\|_{T_{\mu,p}}^p, \quad t\ge 0,\ \mu\in\mathscr{P}_p(\mathscr{C}),\ \phi\in T_{\mu,p}, \lambda\ge\lambda_0.$$

(2) *If $p\in[1,2]$ and $\|\nabla\sigma(t,\cdot)(\xi)\|\le\theta$ for some constant $\theta>0$ and all $(t,\xi)\in[0,\infty)\times\mathscr{C}$ such that (7.6) holds, then there exists constants $c,\lambda_0>0$ such that (7.7) holds.*

Proof. (1) Let $p>4$. Let $\lambda>c>0$ be two constants, and simply denote $Z_t^\lambda = Z_t^{\mu,\phi,\lambda}$. By Itô's formula for (7.1) and applying (4.1), we obtain

$$\begin{aligned}\mathrm{d}(|Z^\lambda(t)|^2) &= \big\{2\,\langle Z^\lambda(t), (\nabla_{Z_t^\lambda}b)(t,\cdot,\mu_t)(X_t^\mu)\rangle + \|(\nabla_{Z_t^\lambda}\sigma)(t,\cdot)(X_t^\mu)\|_{\mathrm{HS}}^2 - 2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t + 2\mathrm{d}M^\lambda(t)\\ &\le \big\{K\|Z_t^\lambda\|_\infty^2 - 2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t + 2\mathrm{d}M^\lambda(t),\end{aligned}$$

where

BJJ0 (7.8) $$\mathrm{d}M^\lambda(t) := \langle Z^\lambda(t), (\nabla_{Z_t^\lambda}\sigma)(t,\cdot)(X_t^\mu)\mathrm{d}W(t)\rangle.$$

Then for any constants $\beta>0$, when $\lambda>\beta$ we obtain

E3 (7.9) $$|Z^\lambda(t)|^2\mathrm{e}^{2\beta t} \le |Z^\lambda(0)| + K\int_0^t \mathrm{e}^{-(2\lambda-2\beta)(t-s)}\mathrm{e}^{\beta s}\|Z_s^\lambda\|_\infty^2\mathrm{d}s + \int_0^t \mathrm{e}^{-(2\lambda-2\beta)(t-s)}\mathrm{e}^{\beta s}\mathrm{d}M^\lambda(s).$$

Obviously,

JJD (7.10) $$\sup_{s\in[t-r_0,t]}\big\{\mathrm{e}^{\beta s}|Z^\lambda(s)|\big\}^p \le G_\beta(t) := \big\{\mathrm{e}^{\beta t}\|Z_t^\lambda\|_{\mathscr{C}}\big\}^p \le \mathrm{e}^{\beta p r_0}\sup_{s\in[t-r_0,t]}\big\{\mathrm{e}^{\beta s}|Z^\lambda(s)|\big\}^p.$$

Combining this with (7.9), Lemma 7.2 and (4.1), when $p>4$ we find a constant $c_1>0$ and a positive function h on $[0,\infty)$ with $r_\alpha\to 0$ as $\alpha\to\infty$ such that

$$\mathrm{e}^{-p\beta r_0}\mathbb{E}[G_\beta(t)] \le c_1\|\phi\|_{T,\mu}^p + r_{\lambda-\beta}\int_0^t G_\beta(s)\mathrm{d}s, \quad t\ge 0.$$

Thus, by Gronwall's lemma we derive

$$\mathbb{E}[G_\beta(t)] \le c_1 \mathrm{e}^{p\beta r_0} \|\phi\|_{T,\mu}^p \exp\big[(r_{\lambda-\beta}\mathrm{e}^{p\beta r_0})t\big], \quad t \ge 0.$$

This and (7.10) yield

$$\mathbb{E}[\|Z_t^\lambda\|_{\mathscr{C}}^p] \le c_2 \mathrm{e}^{2p\beta r_0} \|\phi\|_{T,\mu}^p \exp\big[-(p\beta - r_{\lambda-\beta}\mathrm{e}^{p\beta r_0})t\big], \quad t \ge 0.$$

This implies (7.7) by taking $\beta = \delta$ and $p\delta - r_{\lambda-\delta} \ge \delta$ for large λ since $r_\alpha \to 0$ as $\alpha \to \infty$.

(2) Let $p > 2$ with $\|\nabla b(t,\cdot,\mu)(\xi)\|$ bounded. By (7.1), for any $\beta \in (0,\lambda)$ we have

$$\begin{aligned} Z^\lambda(t)\mathrm{e}^{\beta t} =& Z^\lambda(0)\mathrm{e}^{-(\lambda-\beta)t} + \int_0^t \mathrm{e}^{-(\lambda-\beta)(t-s)}\mathrm{e}^{\beta s}(\nabla_{Z_s^\lambda} b)(s,\cdot,\mu_s)(X_s^\mu)\mathrm{d}s \\ &+ \int_0^t \mathrm{e}^{-(\lambda-\beta)(t-s)}\mathrm{e}^{\beta s}(\nabla_{Z_s^\lambda}\sigma)(s,\cdot,\mu_s)(X_s^\mu)\mathrm{d}W(s). \end{aligned}$$

Combining this with (7.10), the boundedness of $\|\nabla b\| + \|\nabla\sigma\|$ and Lemma 7.2, we find a function $h : [0,\infty) \to [0,\infty)$ with $r_\alpha \to 0$ as $\alpha \to \infty$ such that

$$\mathrm{e}^{-\beta p r_0}\mathbb{E}[G_\beta(t)] \le \mathbb{E}\bigg[\sup_{s\in[t-r_0,t]}\big\{\mathrm{e}^{\beta s}|Z^\lambda(s)|\big\}^p\bigg] \le 2^{d-1}\|\phi\|_{T,\mu}^p + r_{\lambda-\beta}\int_0^t \mathbb{E}[G_\beta(s)]\mathrm{d}s.$$

By Gronwall's inequality and using (7.10), we obtain

$$\mathrm{e}^{\beta(t-r_0)}\mathbb{E}[\|Z_t^\lambda\|_{\mathscr{C}}^p] \le \mathbb{E}[G_\beta(t)] \le 2^{p-1}\|\phi\|_{T_{\mu,p}}^p \exp\big[r_{\lambda-\beta}\mathrm{e}^{\beta r_0}t\big].$$

This implies (7.7) by taking $\beta = 2\delta$ and large enough λ such that $\mathrm{e}^{\beta r_0}r_{\lambda-\beta} \le \delta$.

(3) Let $p \in [2,4]$ and $\|\nabla\sigma(t,\cdot)(\xi)\| \le \theta$. We find a constant $c_1 > 0$ such that

$$\mathrm{d}|Z^\lambda(t)|^2 \le \big\{2c_1|Z^\lambda(t)|^{\|}Z_t^\lambda\|_{\mathscr{C}}| + \theta^2\|Z_t^\lambda\|_{\mathscr{C}}^2 - 2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t + 2\langle Z^{(2)}(t), \{\nabla_{Z_t}\sigma(X_t)\}\mathrm{d}W(t)\rangle.$$

Then for any $p \ge 2$, there exists a constant $c_2 > 0$ such that for any $\varepsilon \in (0,1)$,

PL09 (7.11)
$$\begin{aligned} \mathrm{d}|Z^\lambda(t)|^p \le& \Big\{(c_2\varepsilon^{-1} - p\lambda)|Z^\lambda(t)|^p + c(\varepsilon)\|Z_t^\lambda\|_{\mathscr{C}}^p\Big\}\mathrm{d}t \\ &+ p|Z^\lambda(t)|^{p-2}\langle Z^\lambda(t), \{\nabla_{Z_t}\sigma(X_t)\}\mathrm{d}W(t)\rangle. \end{aligned}$$

holds for

$$c(\varepsilon) := \frac{p\theta^2}{2} + (p-2)\varepsilon^{p-2}, \quad \varepsilon \in (0,1).$$

Noting that for any $\alpha > 0$,

099 (7.12)
$$\eta_\alpha(t) := \mathrm{e}^{\alpha(t-r_0)}\|Z_t^\lambda\|_{\mathscr{C}}^p = \mathrm{e}^{\alpha(t-r_0)}\sup_{s\in[t-r_0,t]}|Z^\lambda(s)|^p \le \sup_{s\in[t-r_0,t]}\mathrm{e}^{\alpha s}|Z_s^\lambda|^p,$$

combining (7.11) with BDG's inequality, we obtain

$$\mathbb{E}[\eta_\alpha(t)] \le \mathbb{E}\bigg[\sup_{s\in[t-r_0,t]}\mathrm{e}^{\alpha s}|Z^\lambda(s)|^p\bigg]$$

$$\leq \mathbb{E}[\|\phi(X_0)\|_{\mathscr{C}}^p] + c(\varepsilon)\int_0^t \mathbb{E}[\mathrm{e}^{\alpha s}\|Z_s^\lambda\|_{\mathscr{C}}^p]\mathrm{d}s + 2p\mathbb{E}\bigg[\bigg(\int_{(t-r_0)^+}^t \theta^2\mathrm{e}^{2\alpha s}|Z^\lambda(s)|^p\|Z_s^\lambda\|_{\mathscr{C}}^p\mathrm{d}s\bigg)^{\frac{1}{2}}\bigg]$$
$$\leq \|\phi\|_{T_{\mu,p}}^p + \Big(c(\varepsilon) + 2p^2\theta^2\mathrm{e}^{\alpha r_0}\Big)\mathrm{e}^{\alpha r_0}\int_0^t \eta_\alpha(s)\mathrm{d}s + \frac{1}{2}\mathbb{E}[\eta_\alpha(t)],$$
$$\varepsilon \in (0, c_2/(p\lambda)), \alpha \in (0, p\lambda - c_2\varepsilon^{-1}).$$

By Gronwall's inequality we arrive at

$$\mathbb{E}[\eta_\alpha(t)] \leq 2\|\phi\|_{T_{\mu,p}}^p \mathrm{e}^{\gamma(\varepsilon)t}, \quad \gamma(\varepsilon) := 2\Big(c(\varepsilon) + 2p^2\theta^2\mathrm{e}^{\alpha r_0}\Big)\mathrm{e}^{\alpha r_0}.$$

This and (7.17) yields

$$\mathbb{E}[\|Z_t^\lambda\|_{\mathscr{C}}^p] \leq 2\mathrm{e}^{\alpha r_0}\|\phi\|_{T_{\mu,p}}^p \mathrm{e}^{-\{\alpha-\gamma(\varepsilon)\}t}, \quad t \geq 0, \varepsilon \in (0, c_2/(p\lambda)), \alpha \in (0, p\lambda - c_2\varepsilon^{-1}).$$

Noting that (7.6) implies

$$\lim_{\varepsilon\downarrow 0}\gamma(\varepsilon) = 4p^2\theta^2\mathrm{e}^{2\alpha r_0} + p\theta^2\mathrm{e}^{\alpha r_0},$$

by (7.6) we may find large enough $\lambda_0 > 0$, small enough $\varepsilon \in (0,1)$ and $\alpha \in (0, p\lambda_0 - c_2\varepsilon^{-1})$ such that $\delta := \alpha - g(\varepsilon) > 0$, so that (7.7) holds for some constant $c > 0$n and all $\lambda \geq \lambda_0$. □

The proof of Theorem 7.1. The L-differentiability is implied by Proposition 5.3. So, it suffices to prove (7.3) and (7.5). For simplicity, let $h^\lambda(t) = h^{\mu,\phi,\lambda}(t)$, which was given in (7.4). By **(B)**, (5.8) and (7.7), $h^\lambda \in L^2(\Omega \to \mathcal{H}; \mathbb{P})$ is adapted. According to Lemmas 4.2 and 5.2, the process $Z(t) := \nabla_{\phi(X_0^\mu)}X^\mu(t) - D_{h^\lambda}X^\mu(t)$ solves the SDE with memory

$$\mathrm{d}Z(t) = \{(\nabla_{Z_t}b)(t,\cdot,\mu_t)(X_t^\mu) - \lambda Z(t)\}\mathrm{d}t + (\nabla_{Z_t}\sigma)(t,\cdot)(X_t^\mu)\mathrm{d}W(t),\ t \geq 0, \quad Z_0 = \phi(X_0^\mu).$$

Therefore, the uniqueness of solutions to (7.1) yields

$$Z(t) = \nabla_{\phi(X_0^\mu)}X^\mu(t) - D_{h^\lambda}X^\mu(t), \quad t \geq -r_0.$$

Combining this with the chain rule and the integration by parts formula for the Malliavin derivative, we derive

$$\begin{aligned} D_\phi^L(P_tf)(\mu) &= \mathbb{E}[D_\phi^L f(X_t^\cdot)(\mu)] = \mathbb{E}({}_{\mathscr{C}^*}\langle\nabla f(X_t^\mu), \nabla_{\phi(X_0^\mu)}X_t^\mu\rangle_{\mathscr{C}}) \\ &= \mathbb{E}({}_{\mathscr{C}^*}\langle\nabla f(X_t^\mu), Z_t + D_{h^\lambda}X_t^\mu\rangle_{\mathscr{C}}) = \mathbb{E}(D_{h^\lambda}f(X_t^\mu)) + \mathbb{E}((\nabla_{Z_t}f)(X_t^\mu)) \\ &= \mathbb{E}\bigg(f(X_t^\mu)\int_0^t \big\langle\dot{h}^\lambda(s), \mathrm{d}W(s)\big\rangle\bigg) + \mathbb{E}((\nabla_{Z_t}f)(X_t^\mu)), \quad t \geq 0, \end{aligned}$$

i.e. (7.3) holds. Finally, by Lemma 7.3 and Hölder's inequality, we deduce (7.5) from (7.3). □

7.2 A degenerate setup

In this subsection, we consider the following distribution-path dependent stochastic Hamiltonian system for $X(t) = (X^{(1)}(t), X^{2)}(t))$ on $\mathbb{R}^{l+m} = \mathbb{R}^l \times \mathbb{R}^m$:

d1 (7.13)
$$\begin{cases} \mathrm{d}X^{(1)}(t) = b^{(1)}(t, X_t)\mathrm{d}t, \\ \mathrm{d}X^{(2)}(t) = b^{(2)}(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W(t), \end{cases}$$

where $(W(t))_{t\geq 0}$ is an m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ for $\mathscr{C} := C([-r_0, 0]; \mathbb{R}^{l+m})$, and

$$b := (b^{(1)}, b^{(2)}) : [0,\infty) \times C \times \mathscr{P}_p(\mathscr{C}) \to \mathbb{R}^{l+m},\ \sigma : [0,T] \times \mathscr{C} \to \mathbb{R}^m \otimes \mathbb{R}^m$$

are measurable satisfying the following assumption.

(C) Let $p \in [2,\infty)$. $b(t,\xi,\mu)$ and $\sigma(t,\xi,\mu)$ are bounded on bounded sets, C^1-smooth in $(\xi,\mu) \in \mathscr{C} \times \mathscr{P}_p(\mathscr{C})$ with bounded $\|D^L b(t,\xi,\cdot)(\mu)\|_{T^*_{p,\mu}}$, and there exist constants $\beta, K, \theta > 0$ such that

$$\begin{aligned}&\langle z^{(1)}(0), \{\nabla_z b^{(1)}(t,\cdot)\}(\xi)\rangle + \langle z^{(2)}(0), \{\nabla_z b^{(2)}(t,\cdot,\mu)\}(\xi)\rangle\\ &\quad\leq K|z^{(2)}(0)|\cdot\|z\|_{\mathscr{C}} + \frac{\theta^2}{2}\|z^{(1)}\|^2_{\mathscr{C}} - \beta|z^{(1)}(0)|^2,\\ &\|\nabla_z\sigma(t,\xi)\| \leq \theta\|z\|_{\mathscr{C}},\quad t\geq 0, \xi\in\mathscr{C}, \mu\in\mathscr{P}_p(\mathscr{C}).\end{aligned}$$

Let $\mu_t = \mathscr{L}_{X_t^\mu}$ with $\mathscr{L}_{X_0} = \mu \in \mathscr{P}_p(\mathscr{C})$, and let $\phi \in T_{\mu,p}$. . For any $\lambda > 0$, consider the linear SDE with memory for $Z(t) = (Z^{(1)}(t), Z^{(2)}(t))$ on $\mathbb{R}^{l+m}$

d4
$$\begin{aligned}\mathrm{d}Z(t) =& \big[\{\nabla_{Z_t} b(t,\cdot,\mu_t)\}(X_t^\mu) - \lambda(\mathbf{0}, Z^{(2)}(t))\}\big]\mathrm{d}t\\ &+ (\mathbf{0}, \{\nabla_{Z_t}\sigma(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)),\quad Z_0 = \phi(X_0^\mu).\end{aligned}\tag{7.14}$$

By [29, Theorem 2.3], under assumption **(C)** this equation has has a unique functional solution. We denote the functional solution by $Z_t^{\mu,\phi,\lambda}$ to emphasize the dependence on μ, ϕ and λ. When $\sigma\sigma^*$ is invertible, let

d8
$$\begin{aligned}h^{\mu,\phi,\lambda}(t) :=& \int_0^t \{\sigma^*(\sigma\sigma^*)^{-1}\}(s, X_s^\mu)\Big\{\lambda Z^{(2)}(s))\\ &+ \mathbb{E}_{\mathscr{C}^*}\big[\langle D^L b^{(2)}(\xi,\cdot)(\mu_s)(X_s^\mu), D_\phi^L X_s^\mu\rangle_{\mathscr{C}}\big]\big|_{\xi = X_s^\mu}\Big\}\mathrm{d}s,\quad t\geq 0.\end{aligned}\tag{7.15}$$

L00 **Theorem 7.4.** *Assume* **(C)** *for some $p \geq 2$, and let $\sigma\sigma^*$ be invertible with $\|(\sigma\sigma^*)^{-1}\|_\infty < \infty$. Then for any $T > 0$ and $f \in C_p^1(\mathscr{C})$, $(P_T f)(\mu)$ is L-differentiable in $\mu \in \mathscr{P}_p(\mathscr{C})$ such that*

d5
$$D_\phi^L(P_T f)(\mu) = \mathbb{E}\bigg(f(X_T^\mu)\int_0^T \langle \dot{h}^{\mu,\phi,\lambda}(s), \mathrm{d}W(s)\rangle\bigg) + \mathbb{E}(\nabla_{Z_T} f)(X_T^\mu),\quad \mu\in\mathscr{P}_p(\mathscr{C}), \phi\in T_{\mu,p}.\tag{7.16}$$

Consequently, there exists constants $c, \lambda_0 > 0$ such that

099
$$\begin{aligned}&\bigg|D_\phi^L(P_T f)(\mu) - \mathbb{E}\bigg(f(X_T^\mu)\int_0^T \langle \dot{h}^{\mu,\phi,\lambda}(s), \mathrm{d}W(s)\rangle\bigg)\bigg|\\ &\leq c\mathrm{e}^{-\delta T}\big\{(P_T\|\nabla f\|^{\frac{p}{p-1}})(\mu)\big\}^{\frac{p-1}{p}}\|\phi\|_{T_{\mu,p}},\quad \lambda\geq\lambda_0, T>0,\ f\in C_p^1(\mathscr{C}),\end{aligned}\tag{7.17}$$

provided one of the following conditions holds:

(i) $\theta^2 < \sup_{\alpha\in(0,\beta p)} \frac{\alpha}{p\mathrm{e}^{\alpha r_0} + 4p^2\mathrm{e}^{2\alpha r_0}}$.

(ii) $p > 2$, $\|\nabla b^{(2)}(t, \cdot, \mu)(\xi)\|$ *is bounded, and there exist constants* $\kappa, \beta > 0$*, such that*

$$(p-1)\kappa^p < p^2 2^{\frac{p}{2}} \sup_{\alpha \in (0,\beta)} \mathrm{e}^{-\alpha r_0}(\beta - \alpha),$$
$$\langle \xi(0), \{\nabla_\xi b^{(1)}(t, \cdot)\}(\xi)\rangle \le \kappa \|\xi\|_{\mathscr{C}} |\xi(0)| - \beta|\xi(0)|^2, \quad t \ge 0, \xi \in \mathscr{C}.$$

To prove this result, we first present the following lemma.

L7.3 **Lemma 7.5.** *Assume* **(C)** *for some* $p \ge 2$*. If one of conditions* (i) *and* (ii) *holds, then there exist constants* $c, \delta, \lambda_0 > 0$ *such that for any* $\lambda \ge \lambda_0$,

bb3 (7.18)
$$\mathbb{E}[\|Z_t^{\mu,\phi,\lambda}\|_{\mathscr{C}}^p] \le c\mathrm{e}^{-\delta t}\|\phi\|_{T_{\mu,p}}^p, \quad t \ge 0,\ \mu \in \mathscr{P}_p(\mathscr{C}),\ \phi \in T_{\mu,p}.$$

Proof. We denote $X^\mu = X, Z^{\mu,\phi,\lambda} = Z = (Z^{(1)}, Z^{(2)})$, and $\|Z_t^{(i)}\|_{\mathscr{C}} = \sup_{s\in[t-r_0,t]} |Z^{(i)}(t)|, i = 1, 2$.

(1) Let (i) hold. For $\varepsilon \in (0, 1)$, we take

$$\rho(t) = \sqrt{|Z^{(1)}(t)|^2 + \varepsilon|Z^{(2)}(t)|^2}, \quad t \ge 0.$$

By **(C)** and Itô's formula, when $\lambda \ge 4\beta\varepsilon^{-1} + 2K$, we have

$$\begin{aligned}
\mathrm{d}|\rho(t)|^2 =& \big[2\langle Z^{(1)}(t), \{\nabla_{Z(t)} b^{(1)}(t, \cdot, \mu_t)\}(X(t))\rangle + 2\varepsilon\langle Z^{(2)}(t), \{\nabla_{Z_t} b^{(2)}(t, \cdot, \mu_t)\}(X_t)\rangle \\
&+ \varepsilon\|\nabla_{Z_t}\sigma(X_t)\|_{HS}^2 - \varepsilon\lambda|Z^{(2)}(t)|^2\big]\mathrm{d}t + 2\varepsilon\langle Z^{(2)}(t), \{\nabla_{Z_t}\sigma(X_t)\}\mathrm{d}W(t)\rangle \\
\le& \Big\{ -2\beta|Z^{(1)}(t)|^2 - \varepsilon\Big(\frac{\lambda}{2} - K\Big)|Z^{(2)}(t)|^2 + \theta^2\|Z_t^{(1)}\|_{\mathscr{C}}^2 + \varepsilon\big(K^2\lambda^{-1} + \theta^2\big)\|Z_t\|_{\mathscr{C}}^2\Big\}\mathrm{d}t \\
&+ 2\varepsilon\langle Z^{(2)}(t), \{\nabla_{Z_t}\sigma(t, \cdot)\}(X_t)\mathrm{d}W(t)\rangle \\
\le& \big\{ -2\beta|\rho(t)|^2 + \gamma_{\lambda,\varepsilon}\rho_t^2\big\}\mathrm{d}t + 2\varepsilon\langle Z^{(2)}(t), \{\nabla_{Z_t}\sigma(t, \cdot)\}(X_t)\mathrm{d}W(t)\rangle,
\end{aligned}$$

where

GGG (7.19)
$$\gamma_{\lambda,\varepsilon} := \max\big\{\theta^2 + \varepsilon(K^2\lambda^{-1} + \theta^2),\ 2K^2\lambda^{-1} + \theta^2\big\}.$$

Then for any $p \ge 2$, there exists a constant $c_1 > 0$ such that

PLO (7.20)
$$\mathrm{d}|\rho(t)|^p \le \Big\{ -\beta p|\rho(t)|^p + \frac{p\gamma}{2}\rho_t^2\Big\}\mathrm{d}t + \varepsilon p|\rho(t)|^{p-2}\langle Z^{(2)}(t), \{\nabla_{Z_t}\sigma(X_t)\}\mathrm{d}W(t)\rangle.$$

Noting that

09 (7.21)
$$\eta_\alpha(t) := \mathrm{e}^{\alpha(t-r_0)}\rho_t^p = \mathrm{e}^{\alpha(t-r_0)} \sup_{s\in[t-r_0,t]} |\rho(s)|^p \le \sup_{s\in[t-r_0,t]} \mathrm{e}^{\alpha s}|\rho(s)|^p,$$

combining (7.20) with BDG's inequality, for any $\alpha \in (0, \beta p]$, we obtain

$$\begin{aligned}
\mathbb{E}[\eta_\alpha(t)] &\le \mathbb{E}\Big[\sup_{s\in[t-r_0,t]} \mathrm{e}^{\alpha s}|\rho(s)|^p\Big] \\
&\le \mathbb{E}[\|\phi(X_0)\|_{\mathscr{C}}^p] + \Big(\frac{p\gamma_{\lambda,\varepsilon}}{2} + \frac{p\theta^2}{2} + c_1\theta^2\varepsilon\Big)\int_0^t \mathbb{E}[\mathrm{e}^{\alpha s}\rho_s^p]\mathrm{d}s \\
&\quad + 2p\mathbb{E}\Big[\Big(\int_{(t-r_0)^+}^t \varepsilon^2\theta^2\mathrm{e}^{2\alpha s}|\rho(s)|^{2p-4}|Z^{(2)}(s)|^2\|Z_s\|_{\mathscr{C}}^2\mathrm{d}s\Big)^{\frac{1}{2}}\Big]
\end{aligned}$$

$$\le \|\phi\|_{T_{\mu,p}}^p + \Big(\frac{p\gamma_{\lambda,\varepsilon}}{2} + c_1\theta^2\varepsilon\Big)\mathrm{e}^{\alpha r_0}\int_0^t \eta_\alpha(s)\mathrm{d}s + 2p\theta\mathrm{e}^{\alpha r_0}\mathbb{E}\Big[|\eta_\alpha(t)|\Big(\int_0^t \eta_\alpha(s)\mathrm{d}s\Big)^{\frac{1}{2}}\Big]$$
$$\le \|\phi\|_{T_{\mu,p}}^p + \Big(\frac{p\gamma_{\lambda,\varepsilon}}{2} + c_1\theta^2\varepsilon + 2p^2\mathrm{e}^{\alpha r_0}\theta^2\Big)\mathrm{e}^{\alpha r_0}\int_0^t \eta_\alpha(s)\mathrm{d}s + \frac{1}{2}\mathbb{E}[\eta_\alpha(t)].$$

By Gronwall's inequality we arrive at

$$\mathbb{E}[\eta_\alpha(t)] \le 2\|\phi\|_{T_{\mu,p}}^p \mathrm{e}^{c_{\lambda,\varepsilon}(\alpha)t}, \quad c_{\lambda,\varepsilon}(\alpha) := \Big(p\gamma_{\lambda,\varepsilon} + 2c_1\theta^2\varepsilon + 4p^2\mathrm{e}^{\alpha r_0}\theta^2\Big)\mathrm{e}^{\alpha r_0}.$$

This and (7.21) yields

$$\mathbb{E}[\rho_t^p] \le 2\mathrm{e}^{\alpha r_0}\|\phi\|_{T_{\mu,p}}^p \mathrm{e}^{-\{\alpha - c_{\lambda,\varepsilon}(\alpha)\}t}, \quad t \ge 0.$$

Noting that (7.19) implies

$$\lim_{\lambda\to\infty}\lim_{\varepsilon\downarrow 0} c_{\lambda,\varepsilon}(\alpha) = \mathrm{e}^{\alpha r_0}\Big(4p^2\mathrm{e}^{\alpha r_0}\theta^2 + p\theta^2\Big),$$

by (i), we may find $\alpha \in (0, \beta p)$, small enough $\varepsilon > 0$ and large enough $\lambda_0 > 0$ such that $\delta := \alpha - c_{\lambda_0,\varepsilon}(\alpha) > 0$, so that

$$\mathbb{E}[\|Z_t\|_{\mathscr{C}}^p] \le \varepsilon^{-p}\mathbb{E}[\rho_t^p] \le 2\varepsilon^{-p}\mathrm{e}^{\alpha r_0}\|\phi\|_{T_{\mu,p}}^p \mathrm{e}^{-\delta t}, \quad t \ge 0, \lambda \ge \lambda_0.$$

Then (7.18) holds.

(2) Let (ii) hold. By $(p-1)\kappa^p < p^2 2^{\frac{p}{2}} \sup_{\alpha\in(0,\beta)} \mathrm{e}^{-\alpha r_0}(\beta - \alpha)$, we find a constant $\varepsilon > 0$ such that

NK1 (7.22)
$$p\alpha > \frac{\kappa^p(p-1)}{p(\beta-\alpha)}(1+\varepsilon)\mathrm{e}^{p\alpha r_0}.$$

By (7.14), we have

$$Z^{(2)}(t)\mathrm{e}^{\alpha t} = \phi^{(2)}(X_0)\mathrm{e}^{-(\lambda-\alpha)t} + \int_0^t \mathrm{e}^{-(\lambda-\alpha)(t-s)}\mathrm{e}^{\alpha s}\{\nabla_{Z_s} b^{(2)}(s,\cdot,\mu_s)\}(X_s)\mathrm{d}s$$
$$+ \int_0^t \mathrm{e}^{-(\lambda-\alpha)(t-s)}\mathrm{e}^{\alpha s}\{\nabla_{Z_s}\sigma^{(2)}(s,\cdot,\mu_s)\}(X_s)\mathrm{d}W(s).$$

Then by the boundedness of $\|\nabla b\| + \|\nabla\sigma\|$ and applying Lemma 7.2, we find a constant $c_1 > 0$ and a function $r : [0,\infty) \to [0,\infty)$ with $r_s \to 0$ as $s \to \infty$ such that

SH1 (7.23)
$$\mathbb{E}\Big[\sup_{s\in[t-r_0,t]}\{|Z^{(2)}(s)|\mathrm{e}^{\alpha s}\}^p\Big] \le c_1\|\phi^{(2)}\|_{T_{\mu,p}}^p + r_{ll-\alpha}\int_0^t \|Z_s\|_{\mathscr{C}}^p \mathrm{e}^{\alpha p s}\mathrm{d}s.$$

On the other hand, by (ii) we have

$$\mathrm{d}|Z^{(1)}(t)| \le \{\kappa\|Z_t\|_{\mathscr{C}} - \beta|Z^{(1)}(t)|\}\mathrm{d}t,$$

so that for $\alpha \in (0,\beta)$,

$$\mathrm{e}^{\alpha t}|Z^{(1)}(t) \le \|\phi(X_0)\|_{\mathscr{C}}\mathrm{e}^{-(\beta-\alpha)(t-s)} + \kappa\int_0^t \mathrm{e}^{\alpha s-(\beta-\alpha)(t-s)}\|Z_s\|_{\mathscr{C}}\mathrm{d}s,$$

so that for some constants $c_2 > 0$ we have

$$\begin{aligned}
\mathrm{e}^{(t-r_0)p\alpha}\mathbb{E}[\|Z_t^{(1)}\|_{\mathscr{C}}^p] &\le \mathbb{E}\Big[\sup_{s\in[t-r_0,t]}\{|Z^{(1)}(s)|\mathrm{e}^{\alpha s}\}^p\Big] \\
&\le c_2\|\phi\|_{T_{\mu,p}}^p + \frac{\kappa^p(p-1)}{p(\beta-\alpha)}(1+\varepsilon)\int_0^t \mathbb{E}[\mathrm{e}^{p\alpha s}\|Z_s\|_{\mathscr{C}}^p]\mathrm{d}s.
\end{aligned}$$

Combining this with (7.23) we arrive at

$$\mathrm{e}^{p\alpha t}\mathbb{E}[\|Z_t\|_{\mathscr{C}}^p] \le 2^{\frac{p}{2}}\mathbb{E}[\|Z_t^{(1)}\|_{\mathscr{C}}^p + \|Z_t^{(2)}\|_{\mathscr{C}}^p] \le c_3\|\phi\|_{T_{\mu,p}}^p + \gamma\int_0^t \mathrm{e}^{p\alpha s}\mathbb{E}[\|Z_s\|_{\mathscr{C}}^p]\mathrm{d}s$$

for some constants $c_3 > 0$ and

$$\gamma := 2^{\frac{p}{2}}\Big(\frac{\kappa^p(p-1)}{p(\beta-\alpha)}(1+\varepsilon)\mathrm{e}^{p\alpha r_0} + r_{ll-\alpha}\Big).$$

By Gronwall's lemma, we obtain

$$\mathbb{E}[\|Z_t\|_{\mathscr{C}}^p] \le c_3\|\phi\|_{T_{\mu,p}}^p \exp\big[-(\gamma-p\alpha)t\big].$$

Since (7.22) implies $\lim_{\lambda\to\infty,\varepsilon\to 0}\gamma = 2^{\frac{p}{2}}\frac{\kappa^p(p-1)}{p(\beta-\alpha)}(1+\varepsilon)\mathrm{e}^{p\alpha r_0} < p\alpha$, we may find constats $\lambda_0, \delta > 0$ such that $\alpha p - \gamma \ge \delta$ for $\lambda \ge \lambda_0$. Therefore, (7.18) holds.

□

Proof of Theorem 7.4. Since the L-differentiability is implied by Proposition 5.3, while (7.17) follows from Lemma 7.5 and (7.16), it suffices to prove (7.16).

Simply denote $h = h^{\mu,\phi,\lambda}$. By **(C)**, there exists a constant $c_1 > 0$ such that

d9

$$\mathbb{E}\|v_t^\phi\|_{\mathscr{C}}^p \le c\,\mathrm{e}^{c_1 t}\|\phi\|_{T_{\mu,p}}^p, \quad t\ge 0, \quad \phi\in T_{\mu,p}. \tag{7.24}$$

This together with (7.15) and (7.18) implies that $h \in L^2(\Omega\to\mathcal{H},\mathbb{P})$ is adapted. Let $w_t^h = (w_t^{h,1}, w_t^{h,2})$ be the unique functional solution to the following SDE with memory

d10

$$\begin{aligned}
\mathrm{d}w^h(t) &= \{\nabla_{w_t^h} b(t,\cdot,\mu_t)\}(X_t^\mu)\mathrm{d}t + \big(0,\sigma(t,X_t^\mu)\dot h(t)\big)\mathrm{d}t \\
&\quad + (\mathbf{0},\{\nabla_{w_t^h}\sigma(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)),\ t\in[0,T],\ w_0^h = \mathbf{0}.
\end{aligned} \tag{7.25}$$

By Lemma 4.2, we have $w_t^h = D_h X_t^\mu$. Next, according to Lemma 5.2, $v_t^\phi = (v_t^{\phi,1}, v_t^{\phi,2}) := D_\phi^L X_t^\mu$ exists in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$ and is the unique solution to

d11

$$\begin{aligned}
\mathrm{d}v^\phi(t) &= \{\nabla_{v^\phi(t)} b(t,\cdot,\mu_t)\}(X_t^\mu)\mathrm{d}t + (\mathbb{E}_{\mathscr{C}^*}\langle\{D^L b(t,\xi,\cdot)\}(\mu_t)(X_t^\mu), v_t^\phi\rangle_{\mathscr{C}})\Big|_{\xi=X_t^\mu}\mathrm{d}t \\
&\quad + (\mathbf{0},\{\nabla_{v_t^\phi}\sigma(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)),\quad t\in[0,T],\quad v_0^\phi = \phi(X_0^\mu).
\end{aligned} \tag{7.26}$$

From (7.25) and (7.26) we see that

$$Z(t) := v^\phi(t) - w^{h^{\mu,\phi,\lambda}}(t)$$

solves (7.14). In particular, $Z_T = v_T^\phi - w_T^h = D_\phi^L X_T^\mu - D_h X_T^\mu$. Then (7.16) follows from Proposition 4.3.

□

References

AKR [1] Albeverio, S., Kondratiev, Y. G., Röckner, M., *Differential geometry of Poisson spaces,* C. R. Acad. Sci. Paris Sér. I Math. 323(1996), 1129–1134.

ACHP [2] Amine, O., Coffie, E., Harang, F., Proke, F., A Bismut-Elworthy-Li formula for singular SDE's driven by a fractional Brownian motion and applications to rough volatility modeling, arXiv:1805.11435v1.

Banos [3] Banos, D., The Bismut-Elworthy-Li formula for mean-field stochastic differential equations, *Ann. Inst. Heri Poincaré Probab. Stat.*, **54** (2018), 220–233.

BWY [4] Bao, J., Wang, F.-Y., Yuan, C., Bismut formulae and applications for functional SPDEs, *Bull. Sci. Math.*, **137** (2013), 509–522.

BWY13 [5] J. Bao, F.-Y. Wang, C. Yuan, *Derivative formula and Harnack inequality for degenerate functional SDEs,* Stochastics and Dynamics. 13 (2013), 1250013, 22 pp.

BYY [6] Bao, J., Yin, G., Yuan, C., *Asymptotic Analysis for Functional Differential Equations,* Springer, 2016.

Bismut [7] Bismut, J. M., Large deviation and Malliavin Calculus, Birkhäuser, 1984.

Card [8] Cardaliaguet, P., *Notes on mean field games,* P.-L. Lions lectures at College de France. Online at https://www.ceremade.dauphine.fr/∼cardaliaguet/MFG20130420.pdf.

CDLL19 [9] Cardaliaguet, P., Delarue, F., Lasry, J.-M., Lions, P.-L., *The Master Equation and the Convergence Problem in Mean Field Games,* Princeton University Press, 2019.

BF [10] Cass, T. R., Friz, P. K., The Bismut-Elworthy-Li formula for jump-diffusions and applications to Monte Carlo pricing in finance, arXiv:math/0604311.

CM [11] Crisan, D., McMurray, E., Smoothing properties of McKean-Vlasov SDEs, *Probab. Theory Relat. Fields*, **171** (2018), 97–148.

EL [12] Elworthy, K. D., Li, X.-M., Formulae for the derivatives of heat semigroups, *J. Funct. Anal.*, **125** (1994), 252–286.

E09 [13] Es-Sarhir, A., Scheutzow, M., van Gaans, O., Invariant measures for stochastic functional differential equations with superlinear drift term, *Differential Integral Equations*, **23** (2010), 189–200.

FR [14] Fan, X., Ren, Y., Bismut formulas and applications for stochastic (functional) differential equations driven by fractional Brownian motions, *Stoch. Dyn.*, **17** (2017), 1750028, 19pp.

Fan [15] Fan, X., Derivative formulas and applications for degenerate stochastic differential equations with fractional noises, *J. Theor. Probab.*, **32** (2019), 1360–1381.

Fuh [16] Fuhrman, M., Smoothing properties of nonlinear stochastic equations in Hilbert spaces, *NoDEA*, **3** (1996), 445–464.

GW [17] Guillin, A., Wang, F.-Y., Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality, *J. Differential Equations*, **253** (2012), 20–40.

HSS [18] Hammersley, W., Šiška, D., Szpruch, L., *McKean-Vlasov SDE under measure dependent Lyapunov conditions,* arXiv:1802.03974v1

Ham [19] Halmos, P. R., *Measure Theory,* Springer 1950.

HRW [20] Huang, X., Röckner, M., Wang, F.-Y., Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, *Discrete Contin. Dyn. Syst.*, **39** (2019), 3017–3035.

HW20 [21] Huang, X., Wang, F.-Y., McKean-Vlasov SDEs with drifts discontinuous under Wasserstein distance, arXiv:2002.06877.

KS [22] Kulik, A., Scheutzow, M., Well-posedness, stability, and sensitivities for stochastic delay equations: a generalized coupling approach, arXiv:1808.06050v1.

David [23] Nualart, D., The Malliavin calculus and related topics, vol. 1995, Springer, 2006.

Priola [24] Priola, E., Formulae for the derivatives of degenerate diffusion semigroup, *J. Evol. Equ.*, **6** (2006), 577–600.

PZ [25] Priola, E., Zabczyk, J., Structural properties of semilinear SPDEs driven by cylindrical stable processes, *Probab. Theory Relat. Fields*, **149** (2011), 97–137.

RW19 [26] Ren, P., Wang, F.-Y., Bismut formula for lions derivative of distribution dependent SDEs and applications, *J. Differential Euqations*, **267** (2019), 4745–4777.

RW19b [27] Ren, P., Wang, F.-Y., Derivative formulas in measure on Riemannian manifolds, arXiv:1908.03711.

RW20 [28] Ren, P., Wang, F.-Y., Donsker-Varadhan large deviations for path-distribution dependent SPDEs, arXiv:2002.08652.

VS [29] von Renesse, M.-K., Scheutzow, M., Existence and uniqueness of solutions of stochastic functional differential equations, *Random Oper. Stoch. Equ.*, **18** (2010), 267–284.

Song [30] Song, Y., Gradient estimates and coupling property for semilinear SDEs driven by jump processes, *Sci. China Math.*, **58** (2015), 447–458.

Take [31] Takeuchi, A., The Bismut-Elworthy-Li type formulae for stochastic differential equations with jumps, *J. Theoret. Probab.*, **23** (2010), 576–604.

W07 [32] F.-Y. Wang, Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on non-convex manifolds,*Ann. Probab.*, **39** (2011), 1449–1467.

W13 [33] Wang, F.-Y., *Harnack inequalities and Applications for Stochastic Partial Differential Equations*, Springer, 2013, Berlin.

Wang14 [34] Wang, F.-Y., Derivative formula and gradient estimates for Gruschine type semigroups, *J. Theoret. Probab.*, **27** (2014), 80–95.

Wang16 [35] Wang, F.-Y., Derivative formulas and Poicaré inequality for Kohn-Laplacian type semigroups, *Sci. China Math.*, **59** (2016), 261–280.

Wang [36] Wang, F.-Y., Distribution dependent SDEs for Landau type equations, *Stoch. Process. Appl.*, **128** (2018), 595–621.

WZ13 [37] Wang, F.-Y., Zhang, X., Derivative formula and applications for degenerate semigroups, *J. Math. Pures Appl.*, **99** (2013), 726–740.

WXZ [38] Wang, L., Xie, L., Zhang, X., Derivative formulae for SDEs driven by multiplicative α-stable-like processes, *Stoch. Process. Appl.*, **125** (2015), 867–885.

Zhang [39] Zhang, X., Stochastic flows and Bismut formulas for stochastic Hamiltonian systems, *Stoch. Process. Appl.*, **120** (2010), 1929–1949.

Zhang13 [40] Zhang, X., Derivative formulas and gradient estimates for SDEs driven by α-stable processes, *Stoch. Process. Appl.*, **123** (2013), 1213–1228.