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

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

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

[b)] Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115, Germany

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

jianhaibao@tju.edu.cn, p.ren@iam.uni-bonn.de, wangfy@tju.edu.cn

February 13, 2021

Abstract

To characterize the regularity of distribution-path dependent SDEs in the initial distribution which varies in the class of probability measures on the path space, we introduce the intrinsic and Lions derivatives for probability measures on Banach spaces, and prove the chain rule of the Lions derivative for 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 2020: 60J60, 58J65.
Keywords: distribution-path dependent SDEs, Bismut formula, asymptotic Bismut formula, Malliavin calculus, Lions derivative

1 Introduction

To characterize stochastic systems with evolutions affected by both micro environment and history, the distribution-path dependent SDEs have been considered in [21, 30], where the

*Supported in part by NNSFC (11771326, 11831014, 12071340, 11921001), and DFG through the CRC Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications.

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 monographs [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 formula is named as Bismut formula or Bismut-Elworthy-Li formula. 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, [35] and references therein. Due to their wide applications, the Bismut type formulas have been investigated for different models; see, for instance, [10, 26, 32, 33, 40, 42] for SDEs/SPDEs driven by jump processes, [16, 17, 25, 36, 37, 39, 41] 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 [27] 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 distribution, so that the Bismut type formula is invalid. In this case, a weaker derivative formula, called asymptotic Bismut formula, has been established in [23].

The aim of this paper is to establish (asymptotic) derivative formulas for the Lions derivative in the initial distribution of distribution-path dependent SDEs, so that results derived in [4, 23, 27] are generalized and improved. Since the functional solution of a distribution-path 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 will introduce and study derivatives for probability measures on the path space (or more generally, on a Banach 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

$$\|\xi\|_{\mathscr{C}} := \sup_{-r_0\leq\theta\leq 0} |\xi(\theta)|, \quad \xi \in \mathscr{C}.$$

For $t \geq 0$ and $f \in C([-r_0, \infty); \mathbb{R}^d)$, the $\mathscr{C}$-valued function $(f_t)_{t\geq 0}$ defined by

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

is called the segment (or window) process of $(f(t))_{t\geq -r_0}$. 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 $p \in [1, \infty)$, $\mathscr{P}_p(\mathscr{C})$ the set of probability measures on $\mathscr{C}$ with finite p-th moment, i.e.,

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

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

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

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):

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

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 and satisfy 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

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

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

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

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

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

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

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., it only depends on μ rather than the choices of the initial value X_0, the Brownian motion and the reference probability space.

The remainder of this paper is organized as follows. Since $\mathscr{C}$ is a Banach space, in Section 2 we introduce the intrinsic and Lions derivatives for probability measures on Banach spaces, and establish a derivative formula in the distribution of Banach-valued random variables. In Section 3, we prove the well-posedness of (1.1) under assumption **(A)**, which generalizes the corresponding results derived in [21] for $p = 2$ and in [30] for Lipschitz continuous $b(t, \cdot)$. In Sections 4 and 5, we calculate the Malliavin derivative of X_t^μ with respect to 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 the Bismut type formula for the Lions derivative of $(P_t f)(\mu)$ 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 $(P_t f)(\mu)$ in μ in case of $\sigma(t, \xi, \mu) = \sigma(t, \xi)$ (i.e., the diffusion term is path dependent but independent of the measure argument μ).

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 $(P_t f)(\mu)$.

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] concerning mean-field games, where $\mathscr{P}_2(\mathbb{R}^d)$ consists of all probability measures on $\mathbb{R}^d$ with finite second moment. The relation between them has been clarified in the recent paper [28, 29], 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

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

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

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

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

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

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

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

is called the intrinsic derivative of f at μ.

If moreover

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

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 $f \in C_b^1(\mathscr{P}_p(\mathbb{B}))$.

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 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 $L^p(\Omega)$. We assume that either ξ_ε is continuous in $\varepsilon \in [0, 1]$ or the probability space is Polish (i.e., $\mathscr{F}$ is the $\mathbb{P}$-complete Borel σ-field induced by a Polish metric on Ω).*

(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*

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

for some constant $C > 0$*, then*

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

(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*

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

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 $\mathscr{P}_p(\mathbb{B}) = \mathscr{P}_2(\mathbb{R}^d)$ (i.e., $p = 2$ and $\mathbb{B} = \mathbb{R}^d$).

Lemma 2.2. *Let* $\{(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)\}_{i=1,2}$ *be two atomless, Polish complete probability spaces, and let* $X_i, i = 1, 2,$ *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. Moreover, it suffices to prove for the Polish probability space case. Indeed, when ξ_ε is continuous in ε, we may take $\overline{\Omega} = C([0,1]; \mathbb{R}^d)$, let $\overline{\mathbb{P}}$ be the distribution of $\xi_\cdot$, let $\overline{\mathscr{F}}$ be the $\overline{\mathbb{P}}$-complete Borel σ-field on $\overline{\Omega}$ induced by the uniform norm, and consider the coordinate random variable $\bar{\xi}_\cdot(\omega) := \omega, \omega \in \overline{\Omega}$. Then $\mathscr{L}_{\bar{\xi}_\cdot|\widetilde{\mathbb{P}}} = \mathscr{L}_{\xi_\cdot|\mathbb{P}}$, so that $\mathscr{L}_{\bar{\xi}_\varepsilon|\overline{\mathbb{P}}} = \mathscr{L}_{\xi_\varepsilon|\mathbb{P}}$ for any $\varepsilon \in [0,1]$ and $\mathscr{L}_{\bar{\xi}_0'|\overline{\mathbb{P}}} = \mathscr{L}_{\xi_0'|\mathbb{P}}$, hence we have reduced the situation to the Polish setting.

(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 $n \ge 1$ we find measurable maps

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

such that

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

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

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

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

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

Next, (2.4) implies

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

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

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

Combining (2.4)-(2.8) leads to

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

Since $f(\mu)$ is continuous in μ and $D^Lf(\mu_0)(x)$ 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

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

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

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

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

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

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

3 Well-posedness of (1.1)

When $p=2$, the existence and uniqueness of strong solutions to (1.1) follows from [21, Theorem 3.1]; see also [30, 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.

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*

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

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

$$\mathbb{E}\Big(\sup_{0\leq t\leq T}\|X_t - Y_t\|_{\mathscr{C}}^p\Big) \leq c\,\mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^p. \tag{3.2}$$

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 [31, Theorem 2.3], the assumption **(A)** implies the well-posedness of the decoupled SDE with memory: for any $\mu \in C([0,T];\mathscr{P}_p(\mathscr{C}))$ and $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$,

$$\mathrm{d}Y^\mu(t) = b(t, Y_t^\mu, \mu_t)\mathrm{d}t + \sigma(t, Y_t^\mu, \mu_t)\mathrm{d}W(t), \quad t > 0,\ Y_0^\mu = X_0. \tag{3.3}$$

As shown in the proof of [22, 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 use the fixed point theorem in the distribution variable as explained in the proof of [20, Theorem 3.3]. For fixed $T > 0$, define

$$\mathscr{D}_T = \big\{\mu \in C([0,T];\mathscr{P}_p(\mathscr{C})) :\ \mu_0 = \mathscr{L}_{X_0}\big\},$$

which is a Polish space under the metric

$$\mathbb{W}_{p,\lambda}(\mu,\nu) := \sup_{0\leq t\leq T}\big(\mathrm{e}^{-\lambda t}\mathbb{W}_p(\mu_t,\nu_t)\big), \quad \lambda > 0.$$

Let

$$(H(\mu))_t := \mathscr{L}_{Y_t^\mu}, \quad t \in [0,T], \mu \in \mathscr{D}_T.$$

By the fixed-point theorem, for the strong existence and uniqueness of (1.1), it is sufficient to prove the contraction of the mapping H under the metric $\mathbb{W}_{p,\lambda}$ for large $\lambda > 0$; that is, we only need to verify

(i) $H : \mathscr{D}_T \to \mathscr{D}_T$,

(ii) There exist constants $\lambda > 0$ and $\alpha \in (0,1)$ such that

$$W_{p,\lambda}(H(\mu), H(\nu)) \leq \alpha W_{p,\lambda}(\mu,\nu), \quad \mu,\nu \in \mathscr{D}_T.$$

Under the assumption **(A)**, (i) follows easily from Itô's formula and BDG's inequality. Below we only prove (ii). For any $\mu,\nu \in \mathscr{D}_T$, let $\Psi(t) = Y^\mu(t) - Y^\nu(t), t \in [-r_0, T]$. By Itô's formula and (3.3), we find a constant $c_1 > 0$ such that

$$\mathrm{d}|\Psi(t)|^p \leq c_1\big\{\|\Psi_t\|_{\mathscr{C}}^p + \mathbb{W}_p(\mu_t,\nu_t)^p\big\}\mathrm{d}t + \mathrm{d}M(t), \tag{3.4}$$

where

$$M(t) := p\int_0^t |\Psi(s)|^{p-2}\big\langle \Psi(s), (\sigma(s,Y_s^\mu,\mu_s)-\sigma(s,Y_s^\nu,\nu_s))\mathrm{d}W(s)\big\rangle.$$

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

$$\|\sigma(s,Y_s^\mu,\mu_s)-\sigma(s,Y_s^\nu,\nu_s)\|_{\mathrm{HS}}^2 \le K\|\Psi_t\|_{\mathscr{C}}^2,$$

we find constants $c_2,c_3>0$ such that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\vee(t-r_0)\le s\le t}|M(s)|^p\Big) &\le c_2\mathbb{E}\bigg(\int_{0\vee(t-r_0)}^t |\Psi(s)|^{2(p-1)}\big\{\|\Psi_s\|_{\mathscr{C}}^2+1_{\{p\ge 2\}}\mathbb{W}_p(\mu_s,\nu_s)^2\big\}\mathrm{d}s\bigg)^{\frac12}\\ &\le \frac12\mathbb{E}\|\Psi_t\|_{\mathscr{C}}^p + c_3\int_0^t \big\{\mathbb{E}\|\Psi_s\|_{\mathscr{C}}^p+\mathbb{W}_p(\mu_s,\nu_s)^p\big\}\mathrm{d}s.\end{aligned}$$

This, together with (3.4) and $Y_0^\mu=Y_0^\nu=X_0$, yields

$$\mathbb{E}\|\Psi_t\|_{\mathscr{C}}^p \le c_4\int_0^t \big\{\mathbb{E}\|\Psi_s\|_{\mathscr{C}}^p+\mathbb{W}_p(\mu_s,\nu_s)^p\big\}\mathrm{d}s,\quad t\in[0,T],$$

for some constant $c_4>0$. Thus, the Gronwall inequality gives

$$\mathbb{E}\|\Psi_t\|_{\mathscr{C}}^p \le c_4\mathrm{e}^{c_4T}\int_0^t \mathbb{W}_p(\mu_s,\nu_s)^p\mathrm{d}s,\quad t\in[0,T],$$

which implies that for any $\lambda>0$,

$$\mathrm{e}^{-\lambda t}\mathbb{E}\|\Psi_t\|_{\mathscr{C}}^p \le c_4\mathrm{e}^{c_4T}\int_0^t \mathrm{e}^{-\lambda(t-s)}\mathrm{e}^{-\lambda s}\mathbb{W}_p(\mu_s,\nu_s)^p\mathrm{d}s \le \frac{c_4\mathrm{e}^{c_4T}}{\lambda}\mathbb{W}_{p,\lambda}(\mu,\nu).$$

Since

$$W_{p,\lambda}(H(\mu),H(\nu)) \le \sup_{0\le t\le T}\big(\mathrm{e}^{-\lambda t}\mathbb{E}\|\Psi_t\|_{\mathscr{C}}^p\big),$$

this implies (ii) for $\alpha=\frac12$ and large enough $\lambda>0$. Therefore, the proof is finished. □

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 bounded on bounded subsets of $[0,\infty)\times\mathscr{C}\times\mathscr{P}_p(\mathscr{C})$, 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) $\{(\nabla_\eta\sigma)(t,\cdot,\mu)\}(\xi)$ 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

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

(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) For any $T>0$ there exists a constant $K>0$ such that for any $t\in[0,T]$,

$$2\langle\xi(0),\{(\nabla_\xi b)(t,\cdot,\mu)\}(\eta)\rangle^+ +\|\{(\nabla_\xi\sigma)(t,\cdot,\mu)\}(\eta)\|_{\mathrm{HS}}^2\le K\|\xi\|_{\mathscr{C}}^2,\ \ \xi,\eta\in\mathscr{C},\mu\in\mathscr{P}_p(\mathscr{C}).$$

(B_4) If $p\in[1,2)$, then $\sigma(t,\xi,\mu)=\sigma(t,\xi)$ depends only on t and ξ, and there exists an increasing function $K:[0,\infty)\to[0,\infty)$ such that

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

Obviously, **(B)** implies **(A)** so that Lemma 3.1 applies. For any $T>0$, set $\mathscr{C}_T:=C([0,T];\mathbb{R}^m)$ and consider the Cameron-Martin space

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

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 $\varepsilon\ge0$, consider the SDE

$$\begin{aligned}&\mathrm{d}X^{h,\varepsilon,\mu}(t)=\big\{b(t,X_t^{h,\varepsilon,\mu},\mu_t)+\varepsilon\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)\dot h(t)\big\}\mathrm{d}t+\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)\mathrm{d}W(t),\\&t\in[0,T],X_0^{h,\varepsilon,\mu}=X_0^\mu,\mu_t:=\mathscr{L}_{X_t^\mu}.\end{aligned}\tag{4.1}$$

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

$$D_hX^\mu(t):=\lim_{\varepsilon\to0}\frac{X^{h,\varepsilon,\mu}(t)-X^\mu(t)}{\varepsilon}$$

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.

Lemma 4.1. *Assume* **(A)** *and let* (B_4) *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*

$$\mathbb{E}\Big(\sup_{0\le t\le T} \|X_t^{h,\varepsilon,\mu} - X_t^\mu\|_{\mathscr{C}}^{2\vee p}\Big) \le c\,\varepsilon^{2\vee p}, \quad \varepsilon \in [0,1]. \tag{4.2}$$

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

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

By (1.1) and (4.1), we have

$$\begin{aligned} \mathrm{d}Z^{h,\varepsilon}(t) =& \Big\{\frac{b(t,X_t^{h,\varepsilon,\mu},\mu_t) - b(t,X_t^\mu,\mu_t)}{\varepsilon} + \sigma(t,X_t^{h,\varepsilon,\mu},\mu_t)\dot{h}(t)\Big\}\mathrm{d}t \\ &+ \frac{\sigma(t,X_t^{h,\varepsilon,\mu},\mu_t) - \sigma(t,X_t^\mu,\mu_t)}{\varepsilon}\mathrm{d}W(t), \quad Z_0^{h,\varepsilon} = \mathbf{0}. \end{aligned} \tag{4.3}$$

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

$$\begin{aligned} |Z^{h,\varepsilon}(t\wedge\tau_n)|^q \le& \frac{q}{2}\int_0^{t\wedge\tau_n} \Big\{\frac{2}{\varepsilon}\langle Z^{h,\varepsilon}(s), b(s,X_s^{h,\varepsilon,\mu},\mu_s) - b(s,X_s^\mu,\mu_s)\rangle \\ &+ \frac{q-1}{\varepsilon^2}\|\sigma(s,X_s^{h,\varepsilon,\mu},\mu_s) - \sigma(s,X_s^\mu,\mu_s)\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s + N^\varepsilon(t) + M^\varepsilon(t) \\ \le& c\int_0^{t\wedge\tau_n} \|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q \mathrm{d}s + N^\varepsilon(t) + M^\varepsilon(t), \end{aligned} \tag{4.4}$$

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_4) when $p \in [1,2)$, we find constants $c_0, c_1 > 0$ such that

$$\begin{aligned} \mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|N^\varepsilon(s)|\Big) \le& q\psi\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^{2(q-1)}\int_0^{t\wedge\tau_n}\|\sigma(s,X_s^{h,\varepsilon,\mu},\mu_s)\|^2\mathrm{d}s\Big)^{1/2} \\ \le& \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^q\Big) + c_0\mathbb{E}\Big(\int_0^t (1+\|X_s^{h,\varepsilon,\mu}\|_{\mathscr{C}}^{2\wedge p})\mathrm{d}s\Big)^{\frac{2\vee p}{2}} \\ \le& \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|Z^{h,\varepsilon}(s)|^2\Big) + c_1, \quad t \in [0,T]. \end{aligned} \tag{4.5}$$

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

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}|M^\varepsilon(s)|\Big) &\le c_2\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\int_0^{t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\Big)^{1/2}\\ &\le \frac{1}{4}\mathbb{E}\Big(\sup_{0\le s\le t\wedge\tau_n}\|Z_s^{h,\varepsilon}\|_{\mathscr{C}}^q\Big) + c_3\int_0^t\mathbb{E}\|Z_{s\wedge\tau_n}^{h,\varepsilon}\|_{\mathscr{C}}^q\mathrm{d}s.\end{aligned}\tag{4.6}$$

Combining (4.4)-(4.6), 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 applying Gronwall's inequality followed by letting $n\to\infty$, we derive (4.2). □

Lemma 4.2. *Assume* **(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*

$$D_hX_t^\mu := \lim_{\varepsilon\downarrow 0}\frac{X_t^{h,\varepsilon,\mu}-X_t^\mu}{\varepsilon},\quad t\in[0,T]\tag{4.7}$$

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

$$\begin{aligned}\mathrm{d}w^h(t) &= \big\{\{(\nabla_{w_t^h}b)(t,\cdot,\mu_t)\}(X_t^\mu) + \sigma(t,X_t^\mu,\mu_t)\dot{h}(t)\big\}\mathrm{d}t\\ &\quad + \{(\nabla_{w_t^h}\sigma)(t,\cdot,\mu_t)\}(X_t^\mu)\mathrm{d}W(t),\ t\in[0,T],\ w_0^h = \mathbf{0}, \mu_t := \mathscr{L}_{X_t^\mu}.\end{aligned}\tag{4.8}$$

Proof. By (B_3) and the boundedness of σ due to (B_1), for any adapted $h\in L^2(\Omega\to\mathcal{H},\mathbb{P})$, the SDE (4.8) has a unique solution in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$ and for some constant $C>0$,

$$\mathbb{E}\Big(\sup_{0\le t\le T}\|w_t^h\|_{\mathscr{C}}^2\Big) \le C\mathbb{E}\|h\|_{\mathcal{H}}^2 < \infty.\tag{4.9}$$

So, it remains to prove that the limit in (4.7) exists in $L^2(\Omega\to C([0,T];\mathscr{C}),\mathbb{P})$, and it solves (4.8). 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

$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\sup_{0\le t\le T}|\Lambda^{h,\varepsilon}(t)|^2\Big) = 0.\tag{4.10}$$

Observe that (4.2) and (4.9) imply

$$\mathbb{E}\Big(\sup_{0\le t\le T}|\Lambda^{h,\varepsilon}(t)|^2\Big) < \infty.\tag{4.11}$$

By (4.3) and (4.8), we have

$$\begin{aligned}\mathrm{d}\Lambda^{h,\varepsilon}(t) &= \big\{\{(\nabla_{\Lambda_t^\varepsilon}b)(t,\cdot,\mu_t)\}(X_t^\mu) + \Gamma_1^\varepsilon(t)\big\}\mathrm{d}t\\ &\quad + \big\{\{(\nabla_{\Lambda_t^\varepsilon}\sigma)(t,\cdot,\mu_t)\}(X_t^\mu) + \Gamma_2^\varepsilon(t)\big\}\mathrm{d}W(t),\end{aligned}\tag{4.12}$$

where

(4.13)
$$\begin{aligned}\Gamma_1^\varepsilon(t) :&= \big(\sigma(t, X_t^{h,\varepsilon,\mu}, \mu_t) - \sigma(t, X_t^\mu, \mu_t)\big)\dot h(t)\\ &\quad + \int_0^1 \big\{\{(\nabla_{Z_t^{h,\varepsilon}} b)(t,\cdot,\mu_t)\}(X_t^\mu + \theta(X_t^{h,\varepsilon,\mu} - X_t^\mu)) - \{(\nabla_{Z_t^{h,\varepsilon}} b)(t,\cdot,\mu_t)\}(X_t^\mu)\big\}\mathrm{d}\theta\\ \Gamma_2^\varepsilon(t) :&= \int_0^1 \big\{\{(\nabla_{Z_t^{h,\varepsilon}} \sigma)(t,\cdot,\mu_t)\}(X_t^\mu + \theta(X_t^{h,\varepsilon,\mu} - X_t^\mu)) - \{(\nabla_{Z_t^{h,\varepsilon}} \sigma)(t,\cdot,\mu_t)\}(X_t^\mu)\big\}\mathrm{d}\theta.\end{aligned}$$

Obviously, when $\sigma(t,\xi,\mu) = \sigma(t,\mu)$ does not depend on ξ, the noise term in (4.12) disappears so that the SDE reduces to an ODE for which we can allow h to be non-adapted. Applying Itô's formula yields

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

Obviously, (B_3) implies

$$\mathbb{E}\Big(\sup_{0\le s\le t} \Upsilon_1^\varepsilon(s)\Big) \le 3K\int_0^t \mathbb{E}\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s, \tag{4.14}$$

while Cauchy-Schwarz's inequality gives

$$\mathbb{E}\Big(\sup_{0\le s\le t} |\Upsilon_2^\varepsilon(s)|\Big) \le \int_0^t \big\{2\mathbb{E}|\Lambda^{h,\varepsilon}(s)|^2 + \mathbb{E}|\Gamma_1^\varepsilon(s)|^2 + 2\,\mathbb{E}\|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s. \tag{4.15}$$

Next, by (B_3) and BDG's inequality, we find constants $c_1, c_2 > 0$ such that

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t} \Upsilon_3^\varepsilon(s)\Big) &\le c_1\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\int_0^t \big\|\{(\nabla_{\Lambda_s^{h,\varepsilon}}\sigma)(s,\cdot,\mu_s)\}(X_s^\mu) + \Gamma_2^\varepsilon(s)\big\|^2\mathrm{d}s\Big)^{1/2}\\ &\le \frac{1}{2}\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big) + c_2\int_0^t \big\{\mathbb{E}\|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2 + \mathbb{E}\|\Gamma_2^\varepsilon(s)\|^2\big\}\mathrm{d}s.\end{aligned} \tag{4.16}$$

Combining (4.14), (4.15) with (4.16), there exists a constant $c_3 > 0$ such that

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

By Gronwall's inequality and (4.11), this implies

$$\mathbb{E}\Big(\sup_{0\le s\le t}|\Lambda^{h,\varepsilon}(s)|^2\Big) \le c_3\mathrm{e}^{c_3 t}\mathbb{E}\int_0^t \big\{|\Gamma_1^\varepsilon(s)|^2 + \|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s. \tag{4.17}$$

Moreover, by (4.13), we have

$$|\Gamma_1^\varepsilon(t)|^2 + \|\Gamma_2^\varepsilon(t)\|_{\mathrm{HS}}^2 \le I_\varepsilon(t)|\dot h(t)|^2 + J_\varepsilon(t)\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2, \tag{4.18}$$

where according to (B_1) and (B_3) we find a constant $c(T)>0$ increasing in T such that

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

By (B_3), and (4.2) and $h \in L^\infty(\Omega\to\mathcal{H},\mathbb{P})$, we obtain

$$\limsup_{\varepsilon\to0} \mathbb{E}\int_0^T I_\varepsilon(t)|\dot h(t)|^2\mathrm{d}t \le 2K\|h\|_{L^\infty(\Omega\to\mathcal{H},\mathbb{P})}^2 \limsup_{\varepsilon\to0}\mathbb{E}\Big[\sup_{t\in[0,T]}\|X_t^{h,\varepsilon,\mu} - X_t^\mu\|_{\mathscr{C}}^2\Big] = 0. \tag{4.19}$$

Below we complete the proof of (4.10) by considering two different cases.

(1) When $p>2$, (3.1) and (4.2) 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

$$\mathbb{E}[\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^2\|X_t^{h,\varepsilon,\mu} - X_t^\mu\|_{\mathscr{C}}^{p-2}] = \varepsilon^{p-2}\mathbb{E}\|Z_t^{h,\varepsilon}\|_{\mathscr{C}}^p \le c\,\varepsilon^{p-2} \to 0 \qquad \text{as } \varepsilon\to0.$$

Then, by the dominated convergence theorem, (4.2) 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 (4.18) and (4.19), implies

$$\lim_{\varepsilon\to0}\mathbb{E}\int_0^T \big\{|\Gamma_1^\varepsilon(t)|^2 + \|\Gamma_2^\varepsilon(t)\|_{\mathrm{HS}}^2\big\}\mathrm{d}t = 0 \tag{4.20}$$

so that (4.10) follows from (4.17).

(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,

$$\mathbb{E}\int_0^t \big\{|\Gamma_1^\varepsilon(s)|^2 + \|\Gamma_2^\varepsilon(s)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s \le \varepsilon_T + 2K\int_0^t \|\Lambda_s^{h,\varepsilon}\|_{\mathscr{C}}^2\mathrm{d}s, \quad t\in[0,T], \tag{4.21}$$

where, by the dominated convergence theorem,

$$\varepsilon_T := \int_0^T \mathbb{E}\big\{I_\varepsilon(t)|\dot h(t)|^2 + J_\varepsilon(t)\|w_t^h\|_{\mathscr{C}}^2\big\}\mathrm{d}t \to 0 \qquad \text{as } \varepsilon\to0.$$

Substituting (4.21) into (4.17) and using Gronwall's lemma, we derive

$$\lim_{\varepsilon\to0}\mathbb{E}\Big(\sup_{0\le t\le T}|\Lambda^{h,\varepsilon}(t)|^2\Big) \le \lim_{\varepsilon\to0}\varepsilon_T \mathrm{e}^{(c_3+2K)T} = 0.$$

Therefore, (4.10) holds. □

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

$$\mathbb{E}[D_h F] = \mathbb{E}[F D^*(h)], \quad F \in \mathscr{D}(D), h \in \mathscr{D}(D^*). \tag{4.22}$$

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

$$D^*(h) = \int_0^T \langle \dot{h}(t), \mathrm{d}W(t) \rangle, \tag{4.23}$$

see, for example, [24].

Proposition 4.3. *Assume* **(B)**. *For any $h \in \mathscr{D}(D^*)$ which is adapted if $\sigma(t, \xi, \mu)$ depends on ξ,* (4.8) *has a unique functional solution satisfying* (4.9) *for some constant $C > 0$, and for any $f \in C_b^1(\mathscr{C})$,*

$$\mathbb{E}\big[(\nabla_{w_T^h} f)(X_T^\mu)\big] = \mathbb{E}\big[f(X_T^\mu) D^*(h)\big]. \tag{4.24}$$

Proof. As explained in the proof of Lemma 4.2, the first assertion follows from assumptions **(A)** and **(B)**. So it suffices to prove (4.24).

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 (4.22), we obtain

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

In general, for adapted $h \in \mathscr{D}(D^*)$, we choose $(h_n)_{n \geq 0} \subset L^\infty(\Omega \to \mathcal{H}, \mathbb{P}) \cap \mathscr{D}(D^*)$ such that

$$\lim_{n \to \infty} \mathbb{E}\big[\|h_n - h\|_{\mathcal{H}}^2 + |D^*(h_n) - D^*(h)|^2\big] = 0. \tag{4.26}$$

In terms of (4.25), we have

$$\mathbb{E}\big[(\nabla_{w_T^{h_n}} f)(X_T^\mu)\big] = \mathbb{E}\big[f(X_T^\mu) D^*(h_n)\big], \quad n \geq 1. \tag{4.27}$$

By **(B)** and (4.8), we find a constant $C > 0$ such that

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

This, together with $f \in C_b^1(\mathscr{C})$ and (4.26), yields the desired formula (4.24) by taking $n \to \infty$ in (4.27). □

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

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

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

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

provided the limit above exists. Obviously,

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

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

Lemma 5.1. *Assume* **(A)**. *For any $T > 0$ and $q \geq p$, there exists a constant $c > 0$ such that*

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

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

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

By **(A)**, 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

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

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

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

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

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

with the initial value $v_0^\xi = \xi$, where, for $t\ge 0$, $\mu_t := \mathscr{L}_{X_t^\mu}$ and

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

Let $p\ge 2$. By **(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

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

Lemma 5.2. *Assume* **(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 $(\Phi_t^{\xi,\varepsilon})_{t\ge 0}$ solves (5.5). To end the proof, it suffices to prove

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

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

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

where

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

By Itô's formula, we obtain

$$
|\Xi^{\xi,\varepsilon}(t)|^2 \le \Theta_1^\varepsilon(t) + \Theta_2^\varepsilon(t) + \Theta_3^\varepsilon(t) + \Theta_4^\varepsilon(t), \quad t \ge 0, \tag{5.10}
$$

where

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

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

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

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

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

Moreover, applying BDG's inequality and using (B_3), we find constants $c_3, c_4 > 0$ such that

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

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

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

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

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

This implies (5.9) by following the argument to deduce (4.10) from (4.17). □

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

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

Proposition 5.3. *Assume* **(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*

$$\Phi(X_0^\mu)=\mathbb{E}(\{\nabla X_T^\mu\}^*\nabla f(X_T^\mu)|X_0^\mu),$$

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.,

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

We aim to investigate 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 [27] 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 [27], 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

Theorem 6.1. *Assume* **(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*

$$w_T^{h_{\xi,\mu}}=\nabla_\xi X_T^\mu, \tag{6.2}$$

where $\nabla_\xi X_T^\mu$ is in (5.2) and $w_T^{h_{\xi,\mu}}$ solves (4.8) for $h=h_{\xi,\mu}$. Moreover, suppose that for some increasing function $\alpha_T:[0,\infty)\to[0,\infty)$ we have

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

Then the following assertions hold.

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

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

(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*

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

Consequently,

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

(3) *If moreover*

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

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

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

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

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

Whence, one has

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

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

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

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

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

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

$$|D^L_\phi(P_Tf)(\mu)|^2\le\alpha_T(\|\mu\|_p)\{\mathbb E\|\phi(X_0^\mu)\|_{\mathscr C}^p\}^{\frac2p}\mathbb E[f^2(X_T^\mu)]=\alpha_T(\|\mu\|_p)\|\phi\|^2_{T_{\mu,p}}(P_Tf^2)(\mu),\ \ \phi\in T_{\mu,p}.$$

Therefore, (6.6) holds true.

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

$$I_\mu(\phi):=\frac{|(P_Tf)(\mu\circ(\mathrm{Id}+\phi)^{-1})-(P_Tf)(\mu)-\gamma_\phi|}{\|\phi\|_{T_{\mu,p}}}\to0\text{ as }\|\phi\|_{T_{\mu,p}}\downarrow0,\tag{6.11}$$

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

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

holds for

$$\begin{aligned}A_\mu(\phi)&:=\frac1{\|\phi\|_{T_{\mu,p}}}\int_0^1\mathbb E\big[\big|\big\{f(X_T^{s\phi(X_0^\mu),\mu})-f(X_T^\mu)\big\}D^*(h_{\phi(X_0^\mu),\mu})\big|\big]\mathrm ds,\\ B_\mu(\phi)&:=\frac{\|f\|_\infty}{\|\phi\|_{T_{\mu,p}}}\int_0^1\big(\mathbb E[|D^*(h_{\phi(X_0^\mu),\mu\circ(\mathrm{Id}+s\phi)^{-1}})-D^*(h_{\phi(X_0^\mu),\mu})|^2]\big)^{\frac12}\mathrm ds.\end{aligned}$$

Since $f\in C_b(\mathscr C)$, and (5.4) implies $\mathbb E\|X_T^{s\phi(X_0^\mu),\mu}-X_T^\mu\|^p_{\mathscr C}\to0$ as $\|\phi\|_{T_{p,\mu}}\to0$, it follows from (6.3) and the dominated convergence theorem that

$$\lim_{\|\phi\|_{T_{\mu,p}}\to0}A_\mu(\phi)=0.$$

Finally, (6.7) implies $\lim_{\|\phi\|_{T_{\mu,p}}\to0}B_\mu(\phi)=0$. Therefore, (6.11) follows from (6.12). □

Remark 6.1 When $r_0=0$ (i.e. without memory), the Bismut formula for the L-derivative has been establish in [27] for all $f\in\mathscr B_b(\mathscr C)$, by applying a 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-degenerate case

Theorem 6.2. *Assume* **(B)** *for some $p\ge2$, 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 $\mu,\nu\in\mathscr P_p(\mathscr C)$,*

$$\begin{aligned}&|(P_Tf)(\mu)-(P_Tf)(\nu)|\\&\le C(T)\|f\|_\infty\Big\{1+(T-r_0)^{-\frac12}+K_2(c(1+\|\mu\|_p+\|\nu\|_p))\\&\quad+(\|\mu\|_p+\|\nu\|_p)^{\frac{p-2}2}\Big\}\mathbb W_p(\mu,\nu).\end{aligned}\tag{6.13}$$

(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*

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

holds for

$$\begin{aligned} H^\phi(t) :&= \big\{(\nabla_{Z_t} b)(t,\cdot,\mu_t)\big\}(X^\mu_t) + (\mathbb{E}[{}_{\mathscr{C}^*}\langle D^L b(t,\xi,\cdot)(\mu_t)(X^\mu_t), Z_t\rangle_{\mathscr{C}}])|_{\xi=X^\mu_t} \\ &\quad + \frac{\phi(X^\mu_0)(0)1_{[0,T-r_0]}(t)}{T-r_0}, \end{aligned}$$

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

$$Z(t) := \begin{cases} \phi(X^\mu_0)(t), & \text{if } t \in [-r_0, 0], \\ \frac{(T-r_0-t)^+}{T-r_0}\phi(X^\mu_0)(0), & \text{if } t \ge 0. \end{cases}$$

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

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

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

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

where

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

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

$$|\dot h_{\xi,\mu}(t)|^2 \le c_1\|\xi\|^2_{\mathscr{C}}\big\{(T-r_0)^{-2}1_{[0,T-r_0]}(t) + \|X^\mu_t\|^{p-2}_{\mathscr{C}} + K_2(\|\mu_t\|_p)^2\big\}, \quad t \in [0,T]. \tag{6.18}$$

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

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

for some constant $c = c(T) > 0$ increasing in T. This, combining (3.1) with (4.23) and (6.18), yields

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

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

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

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

On the other hand, by Lemmas 4.2 and 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). □

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 [27], 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 $d = l+m$:

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

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 $t \geq 0$, $\sigma(t)$ is invertible, $b^{(1)}(t,\cdot) \in C^2(\mathbb{R}^{l+m} \to \mathbb{R}^l)$, $b^{(2)}(t,\xi,\mu)$ 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

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

holds for all $t \geq 0$ and $(\xi,\mu) \in \mathscr{C} \times \mathscr{P}_p(\mathscr{C})$.

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

For any $\mu \in \mathscr{P}_p(\mathscr{C})$, let $(X_t^\mu)_{t\geq 0}$ 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 $(P_T f)(\mu) := \mathbb{E} f(X_T^\mu)$, we shall follow the line of [27, 39], 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

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

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

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

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

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

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

Specific examples for $b^{(1)}$ satisfying **(H2)** are included in [27, Example 2.1]. Let $T > r_0$. According to the proof of [39, 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

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

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 [27], 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

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

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

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

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

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

Furthermore, we have

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

where in the last identity we used

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

see the proof of [27, 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

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

Hence, the uniqueness of this equation implies

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

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 [39, Theorem 1.1] that $h_{\xi,\mu} \in \mathscr{D}(D^*)$ satisfies (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.

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 $T > r_0$, $f \in \mathscr{B}_b(\mathscr{C})$,*

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

(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*

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

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 $\mu \in \mathscr{P}_p(\mathscr{C})$,

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

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 [34], for any $\xi \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ we consider the SDE with memory

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

Then, due to (B_3), 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

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

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

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

which obviously solves (6.29) up to time T.

Theorem 6.4. *Assume* **(B)** *for some* $p \ge 2$. *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* $\mu, \nu \in \mathscr{P}_p(\mathscr{C})$,

$$|(P_T f)(\mu) - (P_T f)(\nu)| \le \frac{C(T)}{\sqrt{T - r_0}}\|f\|_\infty \mathbb{W}_p(\mu, \nu). \tag{6.32}$$

(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*

$$D^L_\phi(P_T f)(\mu) = -\mathbb{E}\bigg(f(X_T^\mu)\int_0^T \langle\{\sigma^*(\sigma\sigma^*)^{-1}\}(t)H^\phi(t), \mathrm{d}W(t)\rangle\bigg), \quad \phi \in T_{\mu,p} \tag{6.33}$$

holds for

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

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

$$\|D^L(P_T f)(\mu)\|_{T^*_{\mu,p}} \le \frac{C(T)}{\sqrt{T - r_0}}\{(P_T f^2)(\mu)\}^{\frac{1}{2}} \tag{6.34}$$

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

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

where

$$G^\xi(t) := \Big\{\big\{\big(\nabla_{U_t^\xi} b\big)(t, \cdot, \mu_t)\big\}(X_t^\mu) + \big(\mathbb{E}_{\mathscr{C}^*}\langle D^L b(t, \eta, \cdot)(\mu_t)(X_t^\mu), U_t^\xi\rangle_{\mathscr{C}}\big)\big|_{\eta = X_t^\mu}\Big\}1_{[T-r_0,T]}(t)$$

$$+\frac{U^\xi(t)}{T-r_0-t}1_{[0,T-r_0)}(t),\quad t\in[0,T].$$

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

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

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 $\nabla_\xi X_t^\mu-w^{h_{\xi,\mu}}(t)$ solve the SDE with memory

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

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 [23] for SDEs with memory to that on the L-derivative for distribution-path dependent SDEs. Coming back to SDEs with memory, our conditions are slightly weaker since we allow the drift terms 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 $t\ge0$ and $\xi\in\mathscr{C}$. For any $\lambda\ge0,\mu\in\mathscr{P}_p(\mathscr{C})$ and $\phi\in T_{\mu,p}$, consider the following SDE with memory

$$\begin{aligned}\mathrm{d}Z^{\mu,\phi,\lambda}(t)=&\big\{\{(\nabla_{Z_t^{\mu,\phi,\lambda}}b)(t,\cdot,\mu_t)\}(X_t^\mu)-\lambda Z^{\mu,\phi,\lambda}(t)\big\}\mathrm{d}t\\&+\{(\nabla_{Z_t^{\mu,\phi,\lambda}}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t),\quad Z_0^{\mu,\phi,\lambda}=\phi(X_0^\mu),\ t\ge0.\end{aligned}\tag{7.1}$$

According to [31, Theorem 2.3], (B_3) implies that (7.1) has a unique functional solution $(Z_t^{\mu,\phi,\lambda})_{t\ge0}$ such that

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

Theorem 7.1. *Assume* **(B)** *for some $p\ge2$ such that (B_3) 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 $f\in C_p^1(\mathscr{C})$,*

$$D_\phi^L(P_Tf)(\mu)=\mathbb{E}\bigg(f(X_T^\mu)\int_0^T\langle\dot h^{\mu,\phi,\lambda}(t),\mathrm{d}W(t)\rangle\bigg)+\mathbb{E}\big(\nabla_{Z_T^{\mu,\phi,\lambda}}f\big)(X_T^\mu),\quad\lambda\geq0,\tag{7.3}$$

where

$$\begin{aligned}h^{\mu,\phi,\lambda}(t):=\int_0^t\big\{\sigma^*(\sigma\sigma^*)^{-1}\big\}(s,X_s^\mu)\Big\{&\big(\mathbb{E}_{\mathscr{C}^*}\langle D^Lb(s,\xi,\cdot)(\mu_s)(X_s^\mu),D_\phi^LX_s^\mu\rangle_{\mathscr{C}}\big)\Big|_{\xi=X_s^\mu}\\&+\lambda Z^{\mu,\phi,\lambda}(s)\Big\}\mathrm{d}s,\quad t\geq0.\end{aligned}\tag{7.4}$$

(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*

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

(2) *If $p\in[2,4]$ and*

$$K<\sup_{\alpha>0}\frac{\alpha}{p(p-1+32p\mathrm{e}^{\alpha r_0})\mathrm{e}^{\alpha r_0}},\tag{7.6}$$

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

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

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*

$$F_\alpha(t)=\int_0^t\mathrm{e}^{-\alpha(t-s)}\mathrm{d}M(s),\quad t\geq0,\quad\alpha>0.$$

Then for any $p>2$, there exists a function $r:[0,\infty)\to[0,\infty)$ with $r_\alpha\to0$ as $\alpha\to\infty$ such that

$$\mathbb{E}\bigg[\sup_{s\in[0,t]}|F_\alpha(s)|^p\bigg]\leq r_\alpha\mathbb{E}\int_0^tg(s)^{\frac{p}{2}}\mathrm{d}s,\quad t\geq0.$$

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

$$\mathbb{E}\bigg[\sup_{s\in[0,t]}\bigg|\int_0^s\mathrm{e}^{-\alpha(s-u)}A(u)\mathrm{d}W(u)\bigg|^p\bigg]\leq d^{p-1}r_\alpha\mathbb{E}\int_0^t\|A(s)\|^p\mathrm{d}s,\quad t\geq0.$$

Lemma 7.3. *Assume* **(B)** *for some $p\geq2$ such that (B_3) 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*

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

(2) *If $p\in[2,4]$ and (7.6) holds, then there exists constants $c,\delta,\lambda_0>0$ such that (7.7) holds.*

Proof. (1) Let $p>4$ and denote by $Z_t^\lambda=Z_t^{\mu,\phi,\lambda}$. Applying Itô's formula for (7.1) and using (B_3), we obtain

$$\begin{aligned}\mathrm{d}|Z^\lambda(t)|^2&=\big\{2\,\langle Z^\lambda(t),\{(\nabla_{Z_t^\lambda}b)(t,\cdot,\mu_t)\}(X_t^\mu)\rangle+\|\{(\nabla_{Z_t^\lambda}\sigma)(t,\cdot)\}(X_t^\mu)\|_{\mathrm{HS}}^2\\&\quad-2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t+\mathrm{d}M^\lambda(t)\\&\le\big\{K\|Z_t^\lambda\|_\infty^2-2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t+\mathrm{d}M^\lambda(t),\end{aligned} \tag{7.8}$$

where

$$\mathrm{d}M^\lambda(t):=2\langle Z^\lambda(t),\{(\nabla_{Z_t^\lambda}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)\rangle. \tag{7.9}$$

Then for $\beta\in(0,\lambda)$ we obtain

$$|Z^\lambda(t)|^2\mathrm{e}^{2\beta t}\le|Z^\lambda(0)|+K\int_0^t\mathrm{e}^{-2(\lambda-\beta)(t-s)}\mathrm{e}^{\beta s}\|Z_s^\lambda\|_\infty^2\mathrm{d}s+\int_0^t\mathrm{e}^{-2(\lambda-\beta)(t-s)}\mathrm{e}^{\beta s}\mathrm{d}M^\lambda(s). \tag{7.10}$$

Obviously,

$$\mathrm{e}^{-\alpha r_0}\sup_{s\in[t-r_0,t]}(\mathrm{e}^{\alpha s}|Z^\lambda(s)|^p)\le G_\alpha(t):=\mathrm{e}^{\alpha(t-r_0)}\|Z_t^\lambda\|_{\mathscr{C}}^p\le\sup_{s\in[t-r_0,t]}(\mathrm{e}^{\alpha s}|Z^\lambda(s)|^p),\quad\alpha>0. \tag{7.11}$$

Combining this with (7.10), Lemma 7.2 and (B_3) and employing Hölder's inequality, for $p>4$ we find a positive function r on $[0,\infty)$ with $r_\alpha\to0$ as $\alpha\to\infty$ such that

$$\mathrm{e}^{-p\beta r_0}\mathbb{E}[G_{p\beta}(t)]\le3^{\frac p2-1}\|\phi\|_{T,\mu}^p+r_{\lambda-\beta}\int_0^t\mathbb{E}G_{p\beta}(s)\mathrm{d}s,\quad t\ge0.$$

Thus, by Gronwall's lemma we derive

$$\mathbb{E}[G_{p\beta}(t)]\le3^{\frac p2-1}\mathrm{e}^{p\beta r_0}\|\phi\|_{T,\mu}^p\exp\big[(r_{\lambda-\beta}\mathrm{e}^{p\beta r_0})t\big],\quad t\ge0.$$

This yields

$$\mathbb{E}[\|Z_t^\lambda\|_{\mathscr{C}}^p]\le3^{\frac p2-1}\mathrm{e}^{2p\beta r_0}\|\phi\|_{T,\mu}^p\exp\big[-(p\beta-r_{\lambda-\beta}\mathrm{e}^{p\beta r_0})t\big],\quad t\ge0.$$

This implies (7.7) by taking $\beta=\delta$ and $p\delta-r_{\lambda-\delta}\mathrm{e}^{p\delta r_0}\ge\delta$ for large λ due to $r_\alpha\to0$ as $\alpha\to\infty$.

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

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

$$+\int_0^t \mathrm{e}^{-(\lambda-\beta)(t-s)}\mathrm{e}^{\beta s}\{(\nabla_{Z_s^\lambda}\sigma)(s,\cdot)\}(X_s^\mu)\mathrm{d}W(s).$$

Combining this with (7.11), the boundedness of $\|\nabla b\|+\|\nabla\sigma\|$ and Lemma 7.2 and applying Hölder's inequality, we find a function $r:[0,\infty)\to[0,\infty)$ with $r_\alpha\to 0$ as $\alpha\to\infty$ such that

$$\mathrm{e}^{-\beta p r_0}\mathbb{E}[G_\beta(t)]\le 3^{p-1}\|\phi\|_{T,\mu}^p+r_{\lambda-\beta}\int_0^t \mathbb{E}[G_\beta(s)]\mathrm{d}s.$$

This, by using Gronwall's inequality, yields

$$\mathrm{e}^{p\beta t}\mathbb{E}[\|Z_t^\lambda\|_{\mathscr{C}}^p]=\mathbb{E}[G_\beta(t)]\le 3^{p-1}\mathrm{e}^{2\beta p r_0}\|\phi\|_{T_{\mu,p}}^p\exp\big[r_{\lambda-\beta}\mathrm{e}^{p\beta r_0}t\big],$$

which implies (7.7) by taking $\beta=2\delta$ and large enough λ such that $\mathrm{e}^{p\beta r_0}r_{\lambda-\beta}\le\delta$ due to $r_\alpha\to 0$ as $\alpha\to\infty$.

(3) Let $p\in[2,4]$ and (7.6). From (7.8), we have

$$\mathrm{d}|Z^\lambda(t)|^2\le\big\{K\|Z_t^\lambda\|_\infty^2-2\lambda|Z^\lambda(t)|^2\big\}\mathrm{d}t+2\langle Z^\lambda(t),\{(\nabla_{Z_t^\lambda}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)\rangle.$$

Then for any $p\in[2,4]$ and $\alpha\in(0,p\lambda)$, by Itô's formula and (B_3), it follows that

$$\begin{aligned}\mathrm{d}(\mathrm{e}^{\alpha t}|Z^\lambda(t)|^p)\le&\mathrm{e}^{\alpha t}\Big\{-(p\lambda-\alpha)|Z^\lambda(t)|^p+\frac{1}{2}Kp(p-1)\|Z_t^\lambda\|_{\mathscr{C}}^p\Big\}\mathrm{d}t\\&+p\mathrm{e}^{\alpha t}|Z^\lambda(t)|^{p-2}\langle Z^\lambda(t),\{(\nabla_{Z_t^\lambda}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)\rangle.\end{aligned}\tag{7.12}$$

Using (7.11) and combining (7.12) with BDG's inequality, we obtain

$$\begin{aligned}&\mathbb{E}[G_\alpha(t)]\le\mathbb{E}\Big[\sup_{s\in[t-r_0,t]}(\mathrm{e}^{\alpha s}|Z^\lambda(s)|^p)\Big]\\&\le\|\phi\|_{T_{\mu,p}}^p+\frac{1}{2}Kp(p-1)\mathrm{e}^{\alpha r_0}\int_0^t\mathbb{E}[\eta_\alpha(s)]\mathrm{d}s+4p\sqrt{2K}\mathbb{E}\Big[\Big(\int_{(t-r_0)^+}^t\mathrm{e}^{2\alpha s}|Z^\lambda(s)|^p\|Z_s^\lambda\|_{\mathscr{C}}^p\mathrm{d}s\Big)^{\frac{1}{2}}\Big]\\&\le\|\phi\|_{T_{\mu,p}}^p+\frac{1}{2}Kp(p-1+32p\mathrm{e}^{\alpha r_0})\mathrm{e}^{\alpha r_0}\int_0^t\eta_\alpha(s)\mathrm{d}s+\frac{1}{2}\mathbb{E}[\eta_\alpha(t)].\end{aligned}$$

Whence, Gronwall's inequality yields

$$\mathbb{E}[G_\alpha(t)]\le 2\|\phi\|_{T_{\mu,p}}^p\mathrm{e}^{\gamma t},\quad\gamma:=Kp(p-1+32p\mathrm{e}^{\alpha r_0})\mathrm{e}^{\alpha r_0}.$$

This, together with (7.11), leads to

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

By (7.6), we may find $\lambda_0>0$ large enough and $\alpha\in(0,p\lambda_0)$ such that $\delta:=\alpha-\gamma>0$, so that (7.7) holds for some constant $c>0$ and all $\lambda\ge\lambda_0$. □

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

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

Therefore, the uniqueness of solutions to (7.1) yields

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

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

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

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 $\mathbb{R}^{l+m} = \mathbb{R}^l \times \mathbb{R}^m$:

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

where $(W(t))_{t\ge0}$ is an m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge0}, \mathbb{P})$, $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ for $\mathscr{C} := C([-r_0, 0]; \mathbb{R}^{l+m})$, and

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

are measurable satisfying one of the following assumptions.

(C1) Let $p \in (2, \infty)$. $b(t, \xi, \mu)$ and $\sigma(t, \xi)$ are bounded on bounded sets, C^1-smooth in $(\xi, \mu) \in \mathscr{C} \times \mathscr{P}_p(\mathscr{C})$ with bounded $\|\{\nabla b^{(2)}(t, \cdot, \mu)\}(\xi)\| + \|\{(\nabla\sigma)(t, \cdot)\}(\xi)\| + \|D^L b(t, \xi, \cdot)(\mu)\|_{T_{p,\mu}^*}$, and there exist constants $\beta, \kappa > 0$ satisfying

$$\kappa^p < 2^{1-\frac{p}{2}} p^p (p-1)^{1-p} \sup_{\alpha\in(0,\beta)} \mathrm{e}^{-p\alpha r_0}(\beta - \alpha) \tag{7.14}$$

such that

$$\langle z^{(1)}(0), \{(\nabla_z b^{(1)})(t, \cdot)\}(\xi)\rangle \le \kappa |z^{(1)}(0)| \cdot \|z\|_{\mathscr{C}} - \beta |z^{(1)}(0)|^2. \tag{7.15}$$

(C2) Let $p \in [2, \infty)$. $b(t, \xi, \mu)$ and $\sigma(t, \xi)$ 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 $K, \beta, \theta > 0$ satisfying

$$\theta^2 < \sup_{\alpha \in (0, \beta p)} \frac{\alpha}{p(p - 1 + 32p\mathrm{e}^{\alpha r_0})\mathrm{e}^{\alpha r_0}} \tag{7.16}$$

such that

$$\begin{aligned} &\langle z^{(1)}(0), \{(\nabla_z b^{(1)})(t, \cdot, \mu)\}(\xi)\rangle \leq K|z^{(2)}| \cdot \|z\|_{\mathscr{C}} + \frac{\theta^2}{2}\|z^{(1)}\|^2_{\mathscr{C}} - \beta|z^{(1)}(0)|^2, \\ &\langle z^{(2)}(0), \{(\nabla_z b^{(2)})(t, \cdot, \mu)\}(\xi)\rangle \leq K|z^{(2)}(0)| \cdot \|z\|_{\mathscr{C}}, \\ &\|\{(\nabla_z \sigma)(t, \cdot)\}(\xi)\| \leq \theta\|z\|_{\mathscr{C}}, \quad t \geq 0, z, \xi \in \mathscr{C}, \mu \in \mathscr{P}_p(\mathscr{C}). \end{aligned} \tag{7.17}$$

Let $\mu_t = \mathscr{L}_{X^\mu_t}$ 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 $\mathbb{R}^{l+m}$

$$\begin{aligned} \mathrm{d}Z(t) =& \big\{\{(\nabla_{Z_t} b)(t, \cdot, \mu_t)\}(X^\mu_t) - \lambda\big(\mathbf{0}, Z^{(2)}(t)\big)\big\}\mathrm{d}t \\ &+ \big(\mathbf{0}, \{(\nabla_{Z_t}\sigma)(t, \cdot)\}(X^\mu_t)\mathrm{d}W(t)\big), \quad Z_0 = \phi(X^\mu_0). \end{aligned} \tag{7.18}$$

By [31, Theorem 2.3], under assumption **(C1)** or **(C2)**, (7.18) has a unique functional solution. We denote the functional solution by $Z^{\mu,\phi,\lambda}_t$ to emphasize the dependence on μ, ϕ and λ. When $\sigma\sigma^*$ is invertible, let

$$\begin{aligned} h^{\mu,\phi,\lambda}(t) = \int_0^t &\{\sigma^*(\sigma\sigma^*)^{-1}\}(s, X^\mu_s)\Big\{\lambda Z^{(2)}(s)) \\ &+ \mathbb{E}\big[{}_{\mathscr{C}^*}\langle D^L b^{(2)}(s, \xi, \cdot)(\mu_s)(X^\mu_s), D^L_\phi X^\mu_s\rangle_{\mathscr{C}}\big]\big|_{\xi = X^\mu_s}\Big\}\mathrm{d}s, \quad t \geq 0. \end{aligned} \tag{7.19}$$

Theorem 7.4. *Assume* **(C1)** *or* **(C2)**, *and let $\sigma\sigma^*$ be invertible with $\|(\sigma\sigma^*)^{-1}\|_\infty < \infty$. Then for any $T > 0$ and $f \in C^1_p(\mathscr{C})$, $(P_T f)(\mu)$ is L-differentiable in $\mu \in \mathscr{P}_p(\mathscr{C})$ such that*

$$D^L_\phi(P_T f)(\mu) = \mathbb{E}\bigg(f(X^\mu_T)\int_0^T \langle \dot{h}^{\mu,\phi,\lambda}(s), \mathrm{d}W(s)\rangle\bigg) + \mathbb{E}(\nabla_{Z_T} f)(X^\mu_T), \quad \mu \in \mathscr{P}_p(\mathscr{C}), \phi \in T_{\mu,p}. \tag{7.20}$$

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

$$\begin{aligned} &\bigg|D^L_\phi(P_T f)(\mu) - \mathbb{E}\bigg(f(X^\mu_T)\int_0^T \langle \dot{h}^{\mu,\phi,\lambda}(s), \mathrm{d}W(s)\rangle\bigg)\bigg| \\ &\leq c\,\mathrm{e}^{-\delta T}\big\{(P_T\|\nabla f\|^{\frac{p}{p-1}})(\mu)\big\}^{\frac{p-1}{p}}\|\phi\|_{T_{\mu,p}}, \quad \lambda \geq \lambda_0, T > 0, \ f \in C^1_p(\mathscr{C}). \end{aligned} \tag{7.21}$$

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

Lemma 7.5. *Assume* **(C1)** *or* **(C2)**. *Then there exist constants $c, \delta, \lambda_0 > 0$ such that for any $\lambda \geq \lambda_0$,*

$$\mathbb{E}[\|Z^{\mu,\phi,\lambda}_t\|^p_{\mathscr{C}}] \leq c\,\mathrm{e}^{-\delta t}\|\phi\|^p_{T_{\mu,p}}, \quad t \geq 0, \ \mu \in \mathscr{P}_p(\mathscr{C}), \ \phi \in T_{\mu,p}. \tag{7.22}$$

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 **(C1)** hold. By (7.18), we have

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

Then, by the boundedness of $\|\nabla b^{(2)}\| + \|\nabla\sigma\|$ 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

$$\mathrm{e}^{p\alpha(t-r_0)}\mathbb{E}\|Z_t^{(2)}\|_{\mathscr{C}}^p \le c_1\|\phi^{(2)}\|_{T_{\mu,p}}^p + r_{\lambda-\alpha}\int_0^t \mathrm{e}^{\alpha ps}\mathbb{E}\|Z_s\|_{\mathscr{C}}^p\mathrm{d}s. \tag{7.23}$$

On the other hand, by (7.15) we have

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

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

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

Hence for any $\varepsilon > 0$ there exists a constant $c_2 > 0$ such that

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

Combining this with (7.23), we arrive at

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

for some constants $c_3 > 0$ with

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

By Gronwall's lemma, we obtain

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

Due to (7.14), we find a constant $\varepsilon > 0$ such that

$$p\alpha > 2^{\frac{p}{2}-1}\kappa^p\Big(\frac{1-1/p}{\beta-\alpha}\Big)^{p-1}(1+\varepsilon)\mathrm{e}^{p\alpha r_0}.$$

This implies

$$\lim_{\lambda\to\infty}\gamma_{\lambda,\varepsilon}=2^{\frac{p}{2}-1}\kappa^p\Big(\frac{1-1/p}{\beta-\alpha}\Big)^{p-1}(1+\varepsilon)\mathrm{e}^{p\alpha r_0}<p\alpha.$$

Hence, we may find constants $\lambda_0,\delta>0$ such that $\alpha p-\gamma_{\lambda,\varepsilon}\geq\delta$ for $\lambda\geq\lambda_0$. Therefore, (7.22) holds.

(2) Let **(C2)** hold. For $\varepsilon\in(0,1)$, set

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

By (7.17) and Itô's formula, for $\lambda\geq 4\beta$, we have

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

where $\|\rho_t\|_{\mathscr{C}}:=\sup_{-r_0\leq\theta\leq 0}|\rho(t+\theta)|$ and

$$\gamma_{\lambda,\varepsilon}:=\max\Big\{\theta^2+\varepsilon\Big(\theta^2+\frac{2K^2(1+\varepsilon)^2}{\lambda\varepsilon^2}\Big),\ \theta^2+\frac{2K^2(1+\varepsilon)^2}{\lambda\varepsilon^2}\Big\}.\tag{7.24}$$

Then, for any $p\geq 2$ and ($\alpha\in(0,p\beta)$), it follows that

$$\begin{aligned}
\mathrm{d}(\mathrm{e}^{\alpha t}|\rho(t)|^p)\leq&\mathrm{e}^{\alpha t}\Big\{-(\beta p-\alpha)|\rho(t)|^p+\frac{1}{2}p(\gamma_{\lambda,\varepsilon}+(p-2)\theta^2)\|\rho_t\|_{\mathscr{C}}^p\Big\}\mathrm{d}t\\
&+\varepsilon p\mathrm{e}^{\alpha t}|\rho(t)|^{p-2}\langle Z^{(2)}(t),\{(\nabla_{Z_t}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)\rangle.
\end{aligned}\tag{7.25}$$

Noting that

$$\mathrm{e}^{-\alpha r_0}\sup_{s\in[t-r_0,t]}(\mathrm{e}^{\alpha s}|\rho(s)|^p)\leq\eta_\alpha(t):=\mathrm{e}^{\alpha(t-r_0)}\|\rho_t\|_{\mathscr{C}}^p\leq\sup_{s\in[t-r_0,t]}(\mathrm{e}^{\alpha s}|\rho(s)|^p),\tag{7.26}$$

and combining (7.25) with BDG's inequality, for any $\alpha\in(0,\beta p]$, we obtain

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

$$\le \|\phi\|^p_{T_{\mu,p}} + \frac{1}{2}p(\gamma_{\lambda,\varepsilon} + (p-2)\theta^2)\mathrm{e}^{\alpha r_0}\int_0^t \eta_\alpha(s)\mathrm{d}s + 4\sqrt{2}p\theta\mathrm{e}^{\alpha r_0}\mathbb{E}\Big[|\eta_\alpha(t)|\Big(\int_0^t \eta_\alpha(s)\mathrm{d}s\Big)^{\frac{1}{2}}\Big]$$
$$\le \|\phi\|^p_{T_{\mu,p}} + \Big(\frac{1}{2}p(\gamma_{\lambda,\varepsilon} + (p-2)\theta^2) + 16p^2\theta^2\mathrm{e}^{\alpha r_0}\Big)\mathrm{e}^{\alpha r_0}\int_0^t \eta_\alpha(s)\mathrm{d}s + \frac{1}{2}\mathbb{E}[\eta_\alpha(t)].$$

By Gronwall's inequality, we arrive at

$$\mathbb{E}[\eta_\alpha(t)] \le 2\|\phi\|^p_{T_{\mu,p}}\mathrm{e}^{c_{\lambda,\varepsilon}(\alpha)t}, \quad c_{\lambda,\varepsilon}(\alpha) := \Big(p(\gamma_{\lambda,\varepsilon} + (p-2)\theta^2) + 32p^2\mathrm{e}^{\alpha r_0}\theta^2\Big)\mathrm{e}^{\alpha r_0}.$$

This and (7.26) yield

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

Note that (7.24) implies

$$\lim_{\varepsilon\downarrow 0}\lim_{\lambda\to\infty} c_{\lambda,\varepsilon}(\alpha) = \mathrm{e}^{\alpha r_0}\Big(p - 1 + 32p\mathrm{e}^{\alpha r_0}\Big)p\theta^2.$$

Then, by (7.16), 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

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

Then (7.22) holds. □

Proof of Theorem 7.4. Since the L-differentiability is implied by Proposition 5.3, while (7.21) follows from Lemma 7.5 and (7.20), it suffices to prove (7.20).

Simply denote $h = h^{\mu,\phi,\lambda}$. By **(C1)** or **(C2)** , there exists a constant $c_1 > 0$ such that

$$\mathbb{E}\|v_t^\phi\|^p_{\mathscr{C}} \le c\,\mathrm{e}^{c_1 t}\|\phi\|^p_{T_{\mu,p}}, \quad t \ge 0, \quad \phi \in T_{\mu,p}. \tag{7.27}$$

This together with (7.19) and (7.22) 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

$$\begin{aligned}\mathrm{d}w^h(t) &= \{(\nabla_{w_t^h} b)(t,\cdot,\mu_t)\}(X_t^\mu)\mathrm{d}t + \big(0, \sigma(t, X_t^\mu)\dot{h}(t)\big)\mathrm{d}t\\ &\quad + (\mathbf{0}, \{(\nabla_{w_t^h}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)),\ t \in [0,T],\ w_0^h = \mathbf{0}.\end{aligned} \tag{7.28}$$

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

$$\begin{aligned}\mathrm{d}v^\phi(t) &= \{(\nabla_{v^\phi(t)} b)(t,\cdot,\mu_t)\}(X_t^\mu)\mathrm{d}t + (\mathbb{E}_{\mathscr{C}^*}\langle D^L b(t,\xi,\cdot)(\mu_t)(X_t^\mu), v_t^\phi\rangle_{\mathscr{C}})\Big|_{\xi = X_t^\mu}\mathrm{d}t\\ &\quad + \big(\mathbf{0}, \{(\nabla_{v_t^\phi}\sigma)(t,\cdot)\}(X_t^\mu)\mathrm{d}W(t)\big), \quad t \in [0,T], \quad v_0^\phi = \phi(X_0^\mu).\end{aligned} \tag{7.29}$$

From (7.28) and (7.29) we see that

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

solves (7.18). In particular, $Z_T = v_T^\phi - w_T^h = D_\phi^L X_T^\mu - D_h X_T^\mu$. Then (7.20) follows from Proposition 4.3. □

References

[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.

[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, *Comm. Math Sci.* **18** (2020), 1863–1890.

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

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

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

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

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

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

[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.

[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.

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

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

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

[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.

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

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

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

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

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

[20] Huang, X., Ren,P., Wang, F.-Y., Distribution dependent stochastic differential equations, *Front. Math. China* (to appear), arXiv:2012.13656.

[21] Huang, X., Röckner, M., Wang, F.-Y., Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs, *Disc. Cont. Dyn. Sys.-A.* **39** (2019), 3017–3035.

[22] Huang, X., Wang, F.-Y., McKean-Vlasov SDEs with drifts discontinuous under Wasserstein distance, *Disc. Cont. Dyn. Sys.-A.* **41** (2021), 1667–1679.

[23] Kulik, A., Scheutzow, M., Well-posedness, stability, and sensitivities for stochastic delay equations: a generalized coupling approach, *Ann. Probab.* **48** (2020), 3041–3076.

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

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

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

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

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

[29] Ren, P., Wang, F.-Y., Stochastic analysis for measure-valued processes, *Sci. China Math.* **2** (2020), 231–252.

[30] Ren, P., Wang, F.-Y., Donsker-Varadhan large deviations for path-distribution dependent SPDEs, *J. Math. Anal. App.*, **499** (2021), 125000.

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

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

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

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

[35] Wang, F.-Y., *Harnack inequalities and applications for stochastic partial differential equations*, Springer, 2013, Berlin.

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

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

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

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

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

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

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