Bismut Formula for Lions Derivative of Distribution Dependent SDEs and Applications*

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

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

[b)] Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, United Kingdom

673788@swansea.ac.uk, wangfy@tju.edu.cn, F.-Y.Wang@swansea.ac.uk

November 16, 2018

Abstract

By using Malliavin calculus, Bismut type formulas are established for the Lions derivative of $P_t f(\mu) := \mathbb{E} f(X_t^\mu)$, where $t > 0$, f is a bounded measurable function, and X_t^μ solves a distribution dependent SDE with initial distribution μ. As applications, explicit estimates are derived for the Lions derivative and the total variational distance between distributions of solutions with different initial data. Both degenerate and non-degenerate situations are considered. Due to the lack of the semigroup property and the invalidity of the formula $P_t f(\mu) = \int P_t f(x)\mu(\mathrm{d}x)$, essential difficulties are overcome in the study.

AMS subject Classification: 60J60, 58J65.
Keywords: Distribution dependent SDEs, Bismut formula, Warsserstein distance, L-derivative.

1 Introduction

The Bismut formula introduced in [3], also called Bismut-Elworthy-Li formula due to [12], is a powerful tool in characterising the regularity of distribution for SDEs and SPDEs. A plenty of results have been derived for this type formulas and applications by using stochastic analysis and coupling methods, see for instance [24] and references therein.

*Supported in part by NNSFC (11771326, 11831014, 11431014,11726627).

On the other hand, because of crucial applications in the study of nonlinear PDEs and environment dependent financial systems, the distribution dependent SDEs (also called McKean-Vlasov or mean filed SDEs) have received increasing attentions, see [10, 11, 13, 14, 18, 22, 23] and references therein. Recently, this type SDEs have been applied in [5, 9, 17, 20] to characterize PDEs involving the Lions derivative (L-derivative for short) introduced by P.-L. Lions in his lectures [6]. Moreover, Harnack inequality, gradient estimates and exponential ergodicity have been investigated in [27] and [21]. In this paper, we aim to establish Bismut type L-derivative formula for distribution dependent SDEs with possibly degenerate noise.

To introduce our main results, we first recall the L-derivative. Let $\mathscr{P}(\mathbb{R}^d)$ be the space of all probability measures on $\mathbb{R}^d$, and let

$$\mathscr{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathscr{P}(\mathbb{R}^d) : \ \mu(|\cdot|^2) := \int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x) < \infty \right\}.$$

Then $\mathscr{P}_2(\mathbb{R}^d)$ is a Polish space under the Wasserstein distance

$$\mathbb{W}_2(\mu, \nu) := \inf_{\pi \in \mathscr{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{2}}, \quad \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d),$$

where $\mathscr{C}(\mu, \nu)$ is the set of couplings for μ and ν; that is, $\pi \in \mathscr{C}(\mu, \nu)$ is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\cdot \times \mathbb{R}^d) = \mu$ and $\pi(\mathbb{R}^d \times \cdot) = \nu$. We will use $\mathbf{0}$ to denote vectors with components 0, or the constant map taking value $\mathbf{0}$.

Definition 1.1. Let $f : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$, and let $g : M \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ for a differentiable manifold M.

(1) f is called L-differentiable at $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, if the functional

$$L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu) \ni \phi \mapsto f(\mu \circ (\mathrm{Id} + \phi)^{-1})$$

is Fréchet differentiable at $\mathbf{0} \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$; that is, there exists (hence, unique) $\gamma \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that

$$\lim_{\mu(|\phi|^2) \to 0} \frac{f(\mu \circ (\mathrm{Id} + \phi)^{-1}) - f(\mu) - \mu(\langle \gamma, \phi \rangle)}{\sqrt{\mu(|\phi|^2)}} = 0. \tag{1.1}$$

In this case, we denote $D^L f(\mu) = \gamma$ and call it the L-derivative of f at μ.

(2) If the L-derivative $D^L f(\mu)$ exists for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, then f is called L-differentiable. If, moreover, for every $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a μ-version $D^L f(\mu)(\cdot)$ such that $D^L f(\mu)(x)$ is jointly continuous in $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$, we denote $f \in C^{(1,0)}(\mathscr{P}_2(\mathbb{R}^d))$.

(3) g is called differentiable on $M \times \mathscr{P}_2(\mathbb{R}^d)$, if for any $(x, \mu) \in M \times \mathscr{P}_2(\mathbb{R}^d)$, $g(\cdot, \mu)$ is differentiable at x and $g(x, \cdot)$ is L-differentiable at μ. If, moreover, $\nabla g(\cdot, \mu)(x)$ and $D^L g(x, \cdot)(\mu)(y)$ are joint continuous in $(x, y, \mu) \in M^2 \times \mathscr{P}_2(\mathbb{R}^d)$, where ∇ is the gradient operator on M, we write $g \in C^{1,(1,0)}(M \times \mathscr{P}_2(\mathbb{R}^d))$.

As indicated in [20] that for any $n \geq 1$, $g \in C^1(\mathbb{R}^n)$ and $h_1, \cdots, h_n \in C_b^1(\mathbb{R}^d)$, the cylindrical function

$$\mu \mapsto g(\mu(h_1), \cdots, \mu(h_n))$$

is in $C^{(1,0)}(\mathscr{P}_2(\mathbb{R}^d))$ with

$$D^L g(\mu)(x) = \sum_{i=1}^n \big(\partial_i g(\mu(h_1), \cdots, \mu(h_n))\big) \nabla h_i(x), \quad (x, \mu) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d).$$

Obviously, if f is L-differentiable at μ, then

$$(1.2) \quad D_\phi^L f(\mu) := \lim_{\varepsilon \downarrow 0} \frac{f(\mu \circ (\mathrm{Id} + \varepsilon\phi)^{-1}) - f(\mu)}{\varepsilon} = \mu\big(\langle D^L f(\mu), \phi \rangle\big), \ \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu).$$

We may call D_ϕ^L the directional L-derivative along ϕ, which was introduced in [**?**, **?**].

When $D_\phi^L f(\mu)$ is a bounded linear functional of $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$, there exists a unique $\xi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that $D_\phi^L f(\mu) = \mu(\langle \xi, \phi \rangle)$ holds for all $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$. In this case, $\phi \mapsto f(\mu \circ (\mathrm{Id} + \phi)^{-1})$ is Gâteaux differentiable at $\mathbf{0}$, and we say that f is weakly L-differentiable at μ, since the Gâteaux differentiability is weaker than the Fréchet one.

By (1.2), for an L-differentiable function f on $\mathscr{P}_2(\mathbb{R}^d)$, we have

$$(1.3) \qquad \|D^L f(\mu)\| := \|D^L f(\mu)(\cdot)\|_{L^2(\mu)} = \sup_{\mu(|\phi|^2) \leq 1} |D_\phi^L f(\mu)|.$$

For a vector-valued function $f = (f_i)$, or a matrix-valued function $f = (f_{ij})$ with L-differentiable components, we write

$$D_\phi^L f(\mu) = (D_\phi^L f_i(\mu)), \text{ or } D_\phi^L f(\mu) = (D_\phi^L f_{ij}(\mu)), \quad \mu \in \mathscr{P}_2(\mathbb{R}^d).$$

Let W_t be a d-dimensional Brownian motion on the natural filtered probability space $(\Omega^0, \mathscr{F}^0, \{\mathscr{F}_t^0\}_{t \geq 0}, \mathbb{P})$. To ensure that for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ there exists a random variable X on $\mathbb{R}^d$ with distribution μ, let μ^0 be a probability measure on $\mathbb{R}^d$ which is equivalent to the Lebesgue measure, and enlarge the probability space as

$$(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \geq 0}, \mathbb{P}) := (\Omega^0 \times \mathbb{R}^d, \mathscr{F}^0 \times \mathscr{B}(\mathbb{R}^d), \{\mathscr{F}_t^0 \times \mathscr{B}(\mathbb{R}^d)\}_{t \geq 0}, \ \mathbb{P}^0 \times \mu^0).$$

Then

$$W_t(\omega) := W_t(\omega^0), \quad t \geq 0, \omega := (\omega^0, x) \in \Omega$$

is a d-dimensional Brownian motion on $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \geq 0}, \mathbb{P})$. Let $\mathscr{L}_\xi$ denote the distribution of a random variable on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In case different probability spaces are concerned, we write $\mathscr{L}_{\xi|\mathbb{P}}$ instead of $\mathscr{L}_\xi$ to emphasize the reference probability measure $\mathbb{P}$.

Consider the following distribution dependent SDE on $\mathbb{R}^d$:

$$(1.4) \qquad \mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \quad X_0 \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P}),$$

where

$$\sigma : [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\otimes d}, \quad b : [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

are continuous such that for some increasing function $K : [0,\infty) \to [0,\infty)$ there holds

$$\begin{aligned} &|b_t(x,\mu) - b_t(y,\nu)| + \|\sigma_t(x,\mu) - \sigma_t(y,\nu)\| \\ &\le K(t)\big(|x-y| + \mathbb{W}_2(\mu,\nu)\big), \quad t \ge 0, x, y \in \mathbb{R}^d, \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d) \end{aligned} \tag{1.5}$$

and

$$\|\sigma_t(\mathbf{0},\delta_{\mathbf{0}})\| + |b_t(\mathbf{0},\delta_{\mathbf{0}})| \le K(t), \quad t \ge 0, \tag{1.6}$$

where and in what follows, for $x \in \mathbb{R}^d$ we denote by δ_x the Dirac measure at x, and $\|\cdot\|$ is the operator norm. For any $t \ge 0$, let $L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_t, \mathbb{P})$ be the class of $\mathscr{F}_t$-measurable square integrable random variables on $\mathbb{R}^d$. By (1.5) and (1.6), for any $s \ge 0$ and $X_s \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_s, \mathbb{P})$, (1.4) has a unique solution $(X_{s,t})_{t\ge s}$ with $X_{s,s} = X_s$ and

$$\mathbb{E}\Big[\sup_{t\in[s,T]} |X_{s,t}|^2\Big] < \infty, \quad T \ge s, \tag{1.7}$$

see, for instance [27], where gradient estimates and Harnack inequalities are also derived for the associated nonlinear semigroup. See also [16, 18] for weaker conditions ensuring the existence and uniqueness of solutions to (1.4). For any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $s \ge 0$, let $(X^\mu_{s,t})_{t\ge s}$ be the solution to (1.4) with $\mathscr{L}_{X_{s,s}} = \mu$. Denote

$$P^*_{s,t}\mu = \mathscr{L}_{X^\mu_{s,t}}, \quad t \ge s, \mu \in \mathscr{P}_2(\mathbb{R}^d). \tag{1.8}$$

Let

$$(P_{s,t}f)(\mu) = (P^*_{s,t}\mu)(f) := \int_{\mathbb{R}^d} f \mathrm{d}(P^*_{s,t}\mu) = \mathbb{E} f(X^\mu_{s,t}), \quad t \ge s, f \in \mathscr{B}_b(\mathbb{R}^d), \mu \in \mathscr{P}_2(\mathbb{R}^d). \tag{1.9}$$

Then for any $0 \le s \le t$, $P_{s,t}$ is a linear operator from $\mathscr{B}_b(\mathbb{R}^d)$ to $\mathscr{B}_b(\mathscr{P}_2(\mathbb{R}^d))$.

In this paper, we aim to establish the Bismut type formula for the L-derivative of $P_{s,t}f$ for $t > s$. By considering the SDE for $\tilde{X}_t := X_{t+s}, t \ge 0$, without loss of generality we may and do assume $s = 0$. So, for simplicity, below we only establish the derivative formula for $P_t f := P_{0,t} f, t > 0$. More precisely, for any $T > 0$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$, we aim to construct an integrable random variable $M^{\mu,\phi}_T$ such that

$$D^L_\phi(P_T f)(\mu) = \mathbb{E}\big[f(X^\mu_T) M^{\mu,\phi}_T\big], \quad f \in \mathscr{B}_b(\mathbb{R}^d), \tag{1.10}$$

which in turn implies the L-differentiability of $P_T f$. Note that the derivative formula for $(P_T f)(x) := (P_T f)(\delta_x)$ along a vector $v \in \mathbb{R}^d$ is derived in [2], which is the special case of (1.10) with $\mu = \delta_x$ and $\phi \equiv v$. Moreover, formulas of the L-derivative and integration by parts have been presented in [8] for the following de-coupled SDE:

$$\mathrm{d}X^{x,\mu}_t = b(t, X^{x,\mu}_t, P^*_t\mu)\mathrm{d}t + \sigma(t, X^{x,\mu}_t, P^*_t\mu)\mathrm{d}W_t, \quad X^{x,\mu}_t = x,$$

which is different from the original SDE (1.4) but has important applications in solving PDEs with Lions' derivatives, see [5, 17, 20] and references within.

When the SDE (1.4) is distribution independent, i.e. $b_t(x,\mu) = b_t(x)$ and $\sigma_t(x,\mu) = \sigma_t(x)$ do not depend on μ, the Bismut type formula

$$\nabla P_T f(x) = \mathbb{E}\big[f(X_T^x) M_T^x\big], \quad x \in \mathbb{R}^d, f \in \mathscr{B}_b(\mathbb{R}^d) \tag{1.11}$$

has been well studied in the literature, where M_T^x is an integrable random variable on $\mathbb{R}^d$, which is measurable in $x \in \mathbb{R}^d$ when it varies, see for instance [1, 15, 25, 26, 28] and references within. Since the coefficients are distribution independent, we have

$$(P_T f)(\mu) = \int_{\mathbb{R}^d} (P_T f)(x)\, \mu(\mathrm{d}x), \tag{1.12}$$

so that $P_T f$ is L-differentiable with $D^L(P_T f)(\mu) = \nabla P_T f$. Hence, by (1.11) and (1.12) we obtain

$$\begin{aligned} D^L_\phi(P_T f)(\mu) = \mu(\langle D^L P_T f, \phi\rangle) &= \int_{\mathbb{R}^d} \mathbb{E}\big[f(X_T^x)\langle M_T^x, \phi(x)\rangle\big]\mu(\mathrm{d}x) \\ &= \mathbb{E}\big[f(X_T^\mu)\langle M_T^{X_0^\mu}, \phi(X_0^\mu)\rangle\big]. \end{aligned}$$

Therefore, (1.10) holds for $M_T^{\mu,\phi} = \langle M_T^{X_0^\mu}, \phi(X_0^\mu)\rangle$.

However, when the SDE is distribution dependent, as explained in [27] that in general (1.12) does not hold, so it is non-trivial to establish the Bismut type formula (1.10).

The remainder of the paper is organized as follows. In section 2, we state our main results on Bismut formulas of $D^L_\phi P_T f$ and applications, for both non-degenerate and degenerate distribution dependent SDEs. To establish the Bismut formula using Malliavin calculus, we make necessary preparations in Section 3 concerning partial derivatives in the initial value, and Malliavin derivative for solutions of (1.4). Finally, complete proofs of the main results are addressed in Section 4.

2 Main results

Let $|\cdot|$ denote the norm in $\mathbb{R}^d$, and $\|\cdot\|$ denote the operator norm for matrices or more generally linear operators. We make the following assumption.

(H) For any $t \geq 0$, $b_t, \sigma_t \in C^{1,(1,0)}(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$. Moreover, there exists a continuous function $K : [0,\infty) \to [0,\infty)$, such that (1.6) holds and

$$\begin{aligned} &\max\Big\{\|\nabla b_t(\cdot,\mu)(x)\|, \|D^L b_t(x,\cdot)(\mu)\|, \frac{1}{2}\|\nabla\sigma_t(\cdot,\mu)(x)\|^2, \frac{1}{2}\|D^L\sigma_t(x,\cdot)(\mu)\|^2\Big\} \\ &\leq K_t, \quad t \geq 0, x \in \mathbb{R}^d, \mu \in \mathscr{P}_2(\mathbb{R}^d), \end{aligned}$$

where as in (1.3), $\|D^L f(\mu)\| := \|D^L f(\mu)(\cdot)\|_{L^2(\mu)}$ for an L-differentiable function f at μ.

Obviously, **(H)** implies (1.5) and (1.6), so that the SDE (1.4) has a unique solution for any initial value $X_0 \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$.

In the following two subsections, we state our main results for non-degenerate and degenerate cases respectively.

2.1 The non-degenerate case

Due to technical reason, the following result Theorem 2.1 only works for distribution independent σ_t. But some other results (for instance Proposition 3.2) apply to the general setting. So, in addition to **(H)** we also assume

(2.1)
$$\sigma_t(x,\mu) = \sigma_t(x) \text{ is invertible with } \|\sigma_t(x,\mu)^{-1}\| \le \lambda_t \text{ for some } \lambda \in C([0,\infty) \to (0,\infty)).$$

Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, and let X_t solve (1.4) for $X_0 \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$ with $\mathscr{L}_{X_0} = \mu$. Given $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$, consider the following SDE for v_t^ϕ on $\mathbb{R}^d$:

$$\begin{aligned} \mathrm{d}v_t^\phi = &\Big\{\nabla_{v_t^\phi} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), v_t^\phi\rangle\big)\big|_{y=X_t}\Big\}\mathrm{d}t \\ &+ \Big\{\nabla_{v_t^\phi}\sigma_t(\cdot, \mathscr{L}_{X_t})(X_t)\Big\}\mathrm{d}W_t, \quad v_0^\phi = \phi(X_0). \end{aligned} \tag{2.2}$$

By **(H)**, this linear SDE is well-posed with $\sup_{t\in[0,T]} \mathbb{E}|v_t^\phi|^2 \le C\mu(|\phi|^2)$ for some constant $C = C(T) > 0$, see (4.21) below. Denote $g_s' = \frac{\mathrm{d}}{\mathrm{d}s} g_s$ for a differentiable function g of $s \in \mathbb{R}$.

Theorem 2.1. *Assume* **(H)** *and* (2.1). *Then for any $f \in \mathscr{B}_b(\mathbb{R}^d), \mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $T > 0$, $P_T f$ is L-differentiable at μ such that for any $g \in C^1([0,T])$ with $g_0 = 0$ and $g_T = 1$,*

$$D_\phi^L(P_T f)(\mu) = \mathbb{E}\bigg[f(X_T)\int_0^T \langle \zeta_t^\phi, \ \mathrm{d}W_t\rangle\bigg], \ \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu), \tag{2.3}$$

where X_t solves (1.4) *for $\mathscr{L}_{X_0} = \mu$, and*

$$\zeta_t^\phi := \sigma_t(X_t)^{-1}\Big\{g_t' v_t^\phi + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), g_t v_t^\phi\rangle\big)\big|_{y=X_t}\Big\}, \quad t \in [0,T].$$

Moreover, the limit

$$D_\phi^L P_T^*\mu := \lim_{\varepsilon\downarrow 0} \frac{P_T^*\mu \circ (\mathrm{Id} + \varepsilon\phi)^{-1} - P_T^*\mu}{\varepsilon} = \psi P_T^*\mu \tag{2.4}$$

exists in the total variational norm, where ψ is the unique element in $L^2(\mathbb{R}^d \to \mathbb{R}, P_T^*\mu)$ such that $\psi(X_T) = \mathbb{E}\big(\int_0^T \langle \zeta_t^\phi, \ \mathrm{d}W_t\rangle \big| X_T\big)$, and $(\psi P_T^*\mu)(A) := \int_A \psi \mathrm{d}P_T^*\mu, \ A \in \mathscr{B}(\mathbb{R}^d)$.

Remark 2.1. When $f \in C_b^1(\mathbb{R}^d)$, (2.3) can be proved as in the distribution independent case by constructing a proper random variable h on the Cameron-Martin space such that $D_h X_T = \nabla_\phi X_T$. However, for the L-differentiability of $P_T f$, one has to construct $\gamma \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that (1.1) holds for $P_T f$ replacing f, which is non-trivial.

Moreover, comparing with the classical case where (2.3) for $f \in C_b^1(\mathbb{R}^d)$ can be easily extended to $f \in \mathscr{B}_b(\mathbb{R}^d)$, there is essential difficulty to do this in the distribution dependent setting. More precisely, when b_t and σ_t do not depend on the distribution, we have the semigroup property $P_T f(\mu) = P_t(P_{t,T}f)(\mu)$ for $t \in (0,T)$, where $P_{t,T}f(x) := P_{t,T}f(\delta_x)$ for the Dirac measure δ_x at point x. In many cases, we have $P_{t,T}f \in C_b^1(\mathbb{R}^d)$ for $f \in \mathscr{B}_b(\mathbb{R}^d)$. Then for any $f \in \mathscr{B}_b(\mathbb{R}^d)$, one may apply the derivative formula (2.3) with $(P_t, P_{t,T}f)$ replacing (P_T, f) to derive a derivative formula for $P_T f$. However, in the distribution dependent case, due to the lack of (1.12) we no longer have $P_T f(\mu) = P_t(P_{t,T}f)(\mu)$, so that this argument becomes invalid. To overcome this difficulty we will make a new approximation argument, see step (a) in the proof of Theorem 2.1 for details.

As applications of Theorem 2.1, the following result consists of estimates on the L-derivative and the total variational distance between distributions of solutions with different initial data.

Corollary 2.2. *Assume* **(H)** *and* (2.1) *for some increasing functions K and continuous function λ.*

(1) *For any $f \in \mathscr{B}_b(\mathbb{R}^d)$ and $T > 0$,*

$$\begin{aligned}\|D^L(P_Tf)(\mu)\|^2 &:= \sup_{\mu(|\phi|^2)\le 1} |D^L_\phi(P_Tf)(\mu)|^2 \\ &\le \big\{(P_Tf^2)(\mu) - (P_tf(\mu))^2\big\} \int_0^T \Big(\frac{1}{T} + K_t\Big)^2 \lambda_t^2 \mathrm{e}^{8K_t t}\mathrm{d}t.\end{aligned} \tag{2.5}$$

(2) *For any $T > 0$,*

$$\begin{aligned}&|P_Tf(\mu) - P_Tf(\nu)|^2 \\ &\le \|f\|_\infty^2 \mathbb{W}_2(\mu,\nu)^2 \int_0^T \Big(\frac{1}{T} + K_t\Big)^2 \lambda_t^2 \mathrm{e}^{8K(t)t}\mathrm{d}t, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d), f \in \mathscr{B}_b(\mathbb{R}^d).\end{aligned} \tag{2.6}$$

Consequently, for any $T > 0$ and $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$,

$$\begin{aligned}\|P_T^*\mu - P_T^*\nu\|_{var}^2 &:= \sup_{A\in\mathscr{B}(\mathbb{R}^d)} |(P_T^*\mu)(A) - (P_T^*\nu)(A)|^2 \\ &\le \mathbb{W}_2(\mu,\nu)^2 \int_0^T \Big(\frac{1}{T} + K_t\Big)^2 \lambda_t^2 \mathrm{e}^{8K(t)t}\mathrm{d}t.\end{aligned} \tag{2.7}$$

2.2 Stochastic Hamiltonian systems

Consider the following distribution dependent stochastic Hamiltonian system for $X_t = (X_t^{(1)}, X_t^{(2)})$ on $\mathbb{R}^{m+d} = \mathbb{R}^m \times \mathbb{R}^d$:

$$\begin{cases}\mathrm{d}X_t^{(1)} = b_t^{(1)}(X_t)\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = b_t^{(2)}(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t \mathrm{d}W_t,\end{cases} \tag{2.8}$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Brownian motion as before, and for each $t \geq 0$, σ_t is an invertible $d \times d$-matrix,

$$b_t = (b_t^{(1)}, b_t^{(2)}) : \mathbb{R}^{m+d} \times \mathscr{P}_2(\mathbb{R}^{m+d}) \to \mathbb{R}^{m+d}$$

is measurable with $b_t^{(1)}(x,\mu) = b_t^{(1)}(x)$ independent of the distribution μ. Let $\nabla = (\nabla^{(1)}, \nabla^{(2)})$ be the gradient operator on $\mathbb{R}^{m+d} = \mathbb{R}^m \times \mathbb{R}^d$, where $\nabla^{(i)}$ is the gradient in the i-th component, $i = 1, 2$. Let $\nabla^2 = \nabla\nabla$ denote the Hessian operator on $\mathbb{R}^{m+d}$. We assume

(H1) For every $t \geq 0$, $b_t^{(1)} \in C_b^2(\mathbb{R}^{m+d} \to \mathbb{R}^m)$, $b_t^{(2)} \in C^{1,(1,0)}(\mathbb{R}^{m+d} \times \mathscr{P}_2(\mathbb{R}^{m+d}) \to \mathbb{R}^d)$, and there exists an increasing function $K : [0,\infty) \to [0,\infty)$ such that (1.6) and

$$\|\nabla b_t(\cdot,\mu)(x)\| + \|D^L b_t^{(2)}(x,\cdot)(\mu)\| + \|\nabla^2 b_t^{(1)}(\cdot,\mu)(x)\| \leq K(t)$$

hold for all $t \geq 0, (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$.

Obviously, this assumption implies **(H)** for the SDE (2.8). We aim to establish the derivative formula of type (1.10) with P_t and P_t^* being defined by (1.8) and (1.9) for the SDE (2.8). To follow the line of [28] where the distribution independent model was investigated, we need the following assumption **(H2)**.

For any $s \geq 0$, let $\{K_{t,s}\}_{t\geq s}$ solve the following linear random ODE on $\mathbb{R}^{m\otimes m}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} K_{t,s} = (\nabla^{(1)} b^{(1)})(X_t) K_{t,s}, \quad t \geq s, K_{s,s} = I_{m\times m}, \tag{2.9}$$

where $I_{m\times m}$ is the $m \times m$-order identity matrix.

(H2) There exists $B \in \mathscr{B}_b([0,T] \to \mathbb{R}^{m\otimes d})$ such that

$$\langle (\nabla^{(2)} b_t^{(1)} - B_t) B_t^* a, a \rangle \geq -\varepsilon |B_t^* a|^2, \quad \forall a \in \mathbb{R}^m \tag{2.10}$$

holds for some constant $\varepsilon \in [0,1)$. Moreover, there exists an increasing function $\theta \in C([0,T])$ with $\theta_t > 0$ for $t \in (0,T]$ such that

$$\int_0^t s(T-s) K_{T,s} B_s B_s^* K_{T,s}^* \mathrm{d}s \geq \theta_t I_{m\times m}, \quad t \in (0,T]. \tag{2.11}$$

Example 2.1. Let

$$b_t^{(1)}(x) = Ax^{(1)} + Bx^{(2)}, \quad x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^{m+d}$$

for some $m \times m$-matrix A and $m \times d$-matrix B. If the Kalman's rank condition

$$\mathrm{Rank}[B, AB, \cdots, A^k B] = m$$

holds for some $k \geq 1$, then **(H2)** is satisfied with $\theta_t = c_T t$ for some constant $c_T > 0$, see the proof of [28, Theorem 4.2]. In general, **(H2)** remains true under small perturbations of this $b_t^{(1)}$.

According to the proof of [28, Theorem 1.1], **(H2)** implies that the matrices

$$Q_t := \int_0^t s(T-s)K_{T,s}\nabla^{(2)} b_s^{(1)}(X_s)B_s^* K_{T,s}^* \mathrm{d}s, \quad t \in (0,T]$$

are invertible with

$$\|Q_t^{-1}\| \le \frac{1}{(1-\varepsilon)\theta_t}, \quad t \in (0,T]. \tag{2.12}$$

For $(X_t)_{t\in[0,T]}$ solving (2.8) with $\mathscr{L}_{X_0} = \mu$ and $\phi = (\phi^{(1)}, \phi^{(2)}) \in L^2(\mathbb{R}^{m+d} \to \mathbb{R}^{m+d}, \mu)$, let

$$\begin{aligned} \alpha_t^{(2)} =& \frac{T-t}{T}\phi^{(2)}(X_0) - \frac{t(T-t)B_t^* K_{T,t}^*}{\int_0^T \theta_s^2 \mathrm{d}s} \int_t^T \theta_s^2 Q_s^{-1} K_{T,0}\phi^{(1)}(X_0)\mathrm{d}s \\ & - t(T-t)B_t^* K_{T,t}^* Q_T^{-1} \int_0^T \frac{T-s}{T} K_{T,s} \nabla^{(2)}_{\phi^{(2)}(X_0)} b_s^{(1)}(X_s)\mathrm{d}s, \quad t \in [0,T], \end{aligned} \tag{2.13}$$

and

$$\alpha_t^{(1)} = K_{t,0}\phi^{(1)}(X_0) + \int_0^t K_{t,s}\nabla^{(2)}_{\alpha_s^{(2)}} b_s^{(1)}(X_s(x))\,\mathrm{d}s, \quad t \in [0,T]. \tag{2.14}$$

Moreover, let $(h_t^\alpha, w_t^\alpha)_{t\in[0,T]}$ be the unique solution to the random ODEs

$$\begin{aligned} \frac{\mathrm{d}h_t^\alpha}{\mathrm{d}t} &= \sigma_t^{-1}\Big\{\nabla_{\alpha_t} b_t^{(2)}(X_t, \mathscr{L}_{X_t}) - (\alpha_t^{(2)})' \\ &\qquad + \big(\mathbb{E}\langle D^L b_t^{(2)}(y,\cdot)(\mathscr{L}_{X_t})(X_t), \alpha_t + w_t^\alpha\rangle\big)\big|_{y=X_t}\Big\}, \\ \frac{\mathrm{d}w_t^\alpha}{\mathrm{d}t} &= \nabla_{w_t^\alpha} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + (\mathbf{0}, \sigma_t(h_t^\alpha)'), \qquad h_0^\alpha = w_0^\alpha = 0. \end{aligned} \tag{2.15}$$

Let $(D^*, \mathscr{D}(D^*))$ be the Malliavin divergence operator associated with the Brownian motion $(W_t)_{t\in[0,T]}$, see Subsection 3.2 below for details. Then the main result in this part is the following.

Theorem 2.3. *Assume* **(H1)** *and* **(H2)**. *Then $h^\alpha \in \mathscr{D}(D^*)$ with $\mathbb{E}|D^*(h^\alpha)|^p < \infty$ for all $p \in [1,\infty)$. Moreover, for any $f \in \mathscr{B}_b(\mathbb{R}^{m+d})$ and $T > 0$, $P_T f$ is L-differentiable at μ such that*

$$D^L_\phi (P_T f)(\mu) = \mathbb{E}\big[f(X_T)\, D^*(h^\alpha)\big]. \tag{2.16}$$

Consequently:

1. (2.4) *holds for the unique $\psi \in L^2(\mathbb{R}^{m+d} \to \mathbb{R}, P_T^*\mu)$ such that $\psi(X_T) = \mathbb{E}(D^*(h^\alpha)|X_T)$.*
2. *There exists a constant $c \ge 0$ such that for any $T > 0$,*

$$\|D^L(P_T f)(\mu)\| \le c\sqrt{P_T|f|^2(\mu) - (P_T f)^2(\mu)}\,\frac{\sqrt{T}(T^2+\theta_T)}{\int_0^T \theta_s^2 \mathrm{d}s}, \quad f \in \mathscr{B}_b(\mathbb{R}^{m+d}), \tag{2.17}$$

$$\|P_T^*\mu - P_T^*\nu\|_{var} \le c\mathbb{W}_2(\mu,\nu)\frac{\sqrt{T}(T^2+\theta_T)}{\int_0^T \theta_s^2 \mathrm{d}s}, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d). \tag{2.18}$$

3 Preparations

We first introduce a formula of the L-derivative re-organized from [6, Theorem 6.5] and [9, Proposition A.2], then investigate the partial derivatives of X_t in the initial value, and the Malliavin derivatives of X_t with respect to the Brownian motion W_t.

3.1 A formula of L-derivative

The following result is essentially due to [6, Theorem 6.5] for $f \in C^{(1,0)}(\mathscr{P}_2(\mathbb{R}^d))$, and [9, Proposition A.2] for bounded X and Y. We include a complete proof for readers' convenience.

Proposition 3.1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be an atomless probability space, and let $X, Y \in L^2(\Omega \to \mathbb{R}^d, \mathbb{P})$ with $\mathscr{L}_X = \mu$. If either X and Y are bounded and f is L-differentiable at μ, or $f \in C^{(1,0)}(\mathscr{P}_2(\mathbb{R}^d))$, then*

$$\lim_{\varepsilon \downarrow 0} \frac{f(\mathscr{L}_{X+\varepsilon Y}) - f(\mu)}{\varepsilon} = \mathbb{E}\langle D^L f(\mu)(X), Y\rangle. \tag{3.1}$$

Consequently,

$$\Big|\lim_{\varepsilon \downarrow 0} \frac{f(\mathscr{L}_{X+\varepsilon Y}) - f(\mu)}{\varepsilon}\Big| = \big|\mathbb{E}\langle D^L f(\mu)(X), Y\rangle\big| \le \|D^L f(\mu)\| \sqrt{\mathbb{E}|Y|^2}. \tag{3.2}$$

Proof. It is easy to see that (3.2) follows from (1.3) and (3.1). Indeed, letting $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that $\phi(X) = \mathbb{E}(Y|X)$, we have

$$\begin{aligned} &\big|\mathbb{E}\langle D^L f(\mu)(X), Y\rangle\big| = \big|\mathbb{E}\langle D^L f(\mu)(X), \phi(X)\rangle\big| = \big|\mu(\langle D^L f(\mu), \phi\rangle)\big| \\ &\le \|D^L f(\mu)\| \cdot \|\phi\|_{L^2(\mu)} = \|D^L f(\mu)\| \big(\mathbb{E}|\mathbb{E}(Y|X)|^2\big)^{\frac{1}{2}} \le \|D^L f(\mu)\| \sqrt{\mathbb{E}|Y|^2}. \end{aligned}$$

Below we prove (3.1) for the stated two situations respectively.

(1) Assume that X and Y are bounded. For any $\mathbb{R}^d$-valued random variable ξ, let $F(\xi) = f(\mathscr{L}_\xi)$. Next, let $(\bar\Omega, \bar{\mathscr{F}}, \bar{\mathbb{P}})$ be an atomless Polish probability space, and let $\bar X \in L^2(\bar\Omega \to \mathbb{R}^d, \bar{\mathbb{P}})$ with $\mathscr{L}_{\bar X|\bar{\mathbb{P}}} = \mu$, where $\mathscr{L}_{\cdot|\bar{\mathbb{P}}}$ denotes the distribution of a random variable under $\bar{\mathbb{P}}$. According to [9, Proposition A.2(iii)], if

$$\bar F(\bar Y) := f(\mathscr{L}_{\bar Y|\bar{\mathbb{P}}}), \quad \bar Y \in L^2(\bar\Omega \to \mathbb{R}^d, \bar{\mathbb{P}})$$

is Fréchet differentiable at $\bar X$ with derivative $D\bar F(\bar X) = D^L f(\mu)(\bar X)$, then

$$\lim_{\varepsilon \downarrow 0} \frac{f(\mathscr{L}_{X+\varepsilon Y}) - f(\mathscr{L}_X) - \varepsilon \mathbb{E}\langle D^L f(\mu)(X), Y\rangle}{\varepsilon} = 0. \tag{3.3}$$

Equivalently, (3.1) holds. Below we construct the desired $\bar X$ and $(\bar\Omega, \bar{\mathscr{F}}, \bar{\mathbb{P}})$ such that $D\bar F(\bar X) = D^L f(\mu)(\bar X)$.

A natural choice of $(\bar\Omega, \bar{\mathscr F}, \bar{\mathbb P})$ is $(\mathbb R^d, \mathscr B(\mathbb R^d), \mu)$, but to ensure the atomless property, we take $(\bar\Omega, \bar{\mathscr F}, \bar{\mathbb P}) = (\mathbb R^d \times \mathbb R, \mathscr B(\mathbb R^d \times \mathbb R), \mu \times \lambda)$, where λ is the standard Gaussian measure on $\mathbb R$. Then $(\bar\Omega, \bar{\mathscr F}, \bar{\mathbb P})$ is an atomless Polish probability space. Let

$$\bar X(\bar\omega) = x, \quad \bar\omega = (x, r) \in \mathbb R^d \times \mathbb R.$$

We have $\mathscr L_{\bar X} = \mu$. Moreover, let

$$\tilde f(\tilde\mu) = f(\tilde\mu(\cdot \times \mathbb R)), \quad \tilde\mu \in \mathscr P_2(\mathbb R^d \times \mathbb R).$$

It is easy to see that the L-differentiability of f at μ implies that of $\tilde f$ at $\mu \times \delta_0$ with

$$D^L \tilde f(\mu \times \delta_0)(x, r) = (D^L f(\mu)(x), 0), \quad (x, r) \in \mathbb R^d \times \mathbb R. \tag{3.4}$$

Finally, on the probability space $(\Omega, \mathscr F, \mathbb P)$ we have

$$F(Y) := f(\mathscr L_Y) = \tilde f(\mathscr L_{\tilde Y}), \quad \tilde Y := (Y, 0) \in L^2(\Omega \to \mathbb R^d \times \mathbb R, \mathscr F, \mathbb P). \tag{3.5}$$

Letting $\tilde X = (X, 0) \in L^2(\Omega \to \mathcal T^d \times \mathbb R, \mathscr F, \mathbb P)$, by [9, Proposition A.2(iii)], the formula (3.3) holds for $(\tilde X, \tilde Y, \tilde f, \mu \times \delta_0)$ replacing (X, Y, f, μ), i.e.

$$\lim_{\varepsilon \downarrow 0} \frac{\tilde f(\mathscr L_{\tilde X + \varepsilon \tilde Y}) - \tilde f(\mathscr L_{\tilde X}) - \mathbb E\langle D^L \tilde f(\mu \times \delta_0), \varepsilon \tilde Y\rangle}{\varepsilon} = 0.$$

Combining this with (3.4) and (3.5), we prove (3.3). Therefore, (3.1) holds.

(2) Let $f \in C^{(1,0)}(\mathscr P_2(\mathbb R^d))$ and let $\mu \in \mathscr P_2(\mathbb R^d)$ and $X \in L^2(\Omega \to \mathbb R^d, \mathbb P)$ with $\mathscr L_X = \mu$. For any $n \geq 1$, let

$$x_n = \frac{x}{\sqrt{1 + n^{-1}|x|^2}}, \quad x \in \mathbb R^d.$$

By (3.1) for bounded X and Y, for any $n \geq 1$ we have

$$\begin{aligned} f(\mathscr L_{X_n + \varepsilon Y_n}) - f(\mathscr L_{X_n}) &= \int_0^\varepsilon \frac{\mathrm d}{\mathrm d s} f(\mathscr L_{X_n + sY_n})\, \mathrm d s \\ &= \int_0^\varepsilon \mathbb E\langle D^L f(\mathscr L_{X_n + sY_n})(X_n + sY_n), Y_n\rangle\, \mathrm d s. \end{aligned} \tag{3.6}$$

Since $f \in C^{(1,0)}(\mathscr P_2(\mathbb R^d))$, it follows that

$$\sup_{n \geq 1, s \in [0, \varepsilon]} \|D^L f(\mathscr L_{X_n + sY_n})\| < \infty, \quad \lim_{n \to \infty} \{f(\mathscr L_{X_n + \varepsilon Y_n}) - f(\mathscr L_{X_n})\} = f(\mathscr L_{X + \varepsilon Y}) - f(\mathscr L_X),$$

and for any $s \in [0, \varepsilon]$,

$$\lim_{n \to \infty} \mathbb E\big(|X - X_n|^2 + |Y - Y_n|^2 + |D^L f(\mathscr L_{X_n + sY_n})(X_n + sY_n) - D^L f(\mathscr L_{X + sY})(X + sY)|^2\big) = 0.$$

Then letting $n \to \infty$ in (3.6) we arrive at

$$f(\mathscr L_{X + \varepsilon Y}) - f(\mathscr L_X) = \int_0^\varepsilon \mathbb E\langle D^L f(\mathscr L_{X + sY})(X + sY), Y\rangle\, \mathrm d s, \quad \varepsilon > 0. \tag{3.7}$$

This implies (3.1). More precisely, it is easy to see that $\{\mathscr{L}_{X+sY}\}$ is compact in $\mathscr{P}_2(\mathbb{R}^d)$. So, $f \in C^{(1,0)}(\mathscr{P}_2(\mathbb{R}^d))$ implies

$$A := \sup_{s\in[0,1]} \sqrt{\mathbb{E}|D^L f(\mathscr{L}_{X+sY})(X+sY)|^2} = \sup_{s\in[0,1]} \|D^L f(\mathscr{L}_{X+sY})\|_{L^2(\mathscr{L}_{X+sY})} < \infty. \tag{3.8}$$

Combining this with the continuity property of $D^L f$ on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$, we conclude that

$$\lim_{\varepsilon\downarrow 0} D^L f(\mathscr{L}_{X+sY})(X+sY) = D^L f(\mathscr{L}_X)(X) \text{ weakly in } L^2(\Omega \to \mathbb{R}^d, \mathbb{P}).$$

In particular,

$$\lim_{\varepsilon\downarrow 0} \mathbb{E}\langle D^L f(\mathscr{L}_{X+sY})(X+sY), Y\rangle = \mathbb{E}\langle D^L f(\mathscr{L}_X)(X), Y\rangle. \tag{3.9}$$

Moreover, (3.8) implies

$$\sup_{s\in[0,1]} \mathbb{E}\big|\langle D^L f(\mathscr{L}_{X+sY})(X+sY), Y\rangle\big| \le A\sqrt{\mathbb{E}|Y|^2} < \infty.$$

Due to this, (3.7) and (3.9), the dominated convergence theorem gives

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} \frac{f(\mathscr{L}_{X+\varepsilon Y}) - f(\mathscr{L}_X)}{\varepsilon} &= \lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\langle D^L f(\mathscr{L}_{X+sY})(X+sY), Y\rangle\,\mathrm{d}s \\ &= \mathbb{E}\langle D^L f(\mathscr{L}_X)(X), Y\rangle.\end{aligned}$$

□

3.2 Partial derivative in initial value

For any $T > 0$, let $\mathscr{C}_T = C([0,T] \to \mathbb{R}^d)$ be the path space over $\mathbb{R}^d$ with time interval $[0,T]$, and let $X_0, \eta \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$. For any $\varepsilon \ge 0$, let $(X_t^\varepsilon)_{t\ge 0}$ solve the SDE

$$\mathrm{d}X_t^\varepsilon = b_t(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon})\mathrm{d}t + \sigma_t(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon})\mathrm{d}W_t, \quad X_0^\varepsilon = X_0 + \varepsilon\eta. \tag{3.10}$$

Obviously, $X_t = X_t^0$ solves (1.4) with initial value X_0. Consider the following linear SDE for v_t^η on $\mathbb{R}^d$:

$$\begin{aligned}\mathrm{d}v_t^\eta = &\Big\{\nabla_{v_t^\eta} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), v_t^\eta\rangle\big)\big|_{y=X_t}\Big\}\mathrm{d}t \\ &+ \Big\{\nabla_{v_t^\eta} \sigma_t(\cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L \sigma_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), v_t^\eta\rangle\big)\big|_{y=X_t}\Big\}\mathrm{d}W_t, \quad v_0^\eta = \eta.\end{aligned} \tag{3.11}$$

The main result of this part is the following.

Proposition 3.2. *Assume* **(H)**. *Then for any $T > 0$, the limit*

$$\nabla_\eta X_t := \lim_{\varepsilon\downarrow 0} \frac{X_t^\varepsilon - X_t}{\varepsilon}, \quad t \in [0,T] \tag{3.12}$$

exists in $L^2(\Omega \to \mathscr{C}_T, \mathbb{P})$. Moreover, $(v_t^\eta := \nabla_\eta X_t)_{t\in[0,T]}$ is the unique solution to the linear SDE (3.11).

To prove the existence of $\nabla_\eta X_t$ in (3.12), it suffices to show that when $\varepsilon \downarrow 0$

$$\xi^\varepsilon(t) := \frac{X_t^\varepsilon - X_t}{\varepsilon}, \quad t \in [0,T] \tag{3.13}$$

is a Cauchy sequence in $L^2(\Omega \to \mathscr{C}_T, \mathbb{P})$, i.e.

$$\lim_{\varepsilon,\delta\downarrow 0} \mathbb{E}\Big[\sup_{t\in[0,T]} |\xi^\varepsilon(t) - \xi^\delta(t)|^2\Big] = 0. \tag{3.14}$$

To this end, we need the following two lemmas.

Lemma 3.3. *Assume* **(H)**. *Then*

$$\sup_{\varepsilon\in(0,1]} \mathbb{E}\Big[\sup_{t\in[0,T]} |\xi^\varepsilon(t)|^2\Big] < \infty.$$

Proof. By **(H)**, there exists a constant $C_1 > 0$ such that

$$\begin{aligned}
&\mathrm{d}|X_t^\varepsilon - X_t|^2 \\
&= \big\{2\langle b_t(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}) - b_t(X_t, \mathscr{L}_{X_t}), X_t^\varepsilon - X_t\rangle + \|\sigma_t(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}) - \sigma_t(X_t, \mathscr{L}_{X_t})\|_{HS}^2\big\}\mathrm{d}t + \mathrm{d}M_t \\
&\le C_1\big\{|X_t^\varepsilon - X_t|^2 + \mathbb{W}_2(\mathscr{L}_{X_t^\varepsilon}, \mathscr{L}_{X_t})^2\big\}\mathrm{d}t + \mathrm{d}M_t,
\end{aligned}$$

where

$$\mathrm{d}M_t := 2\Big\langle X_t^\varepsilon - X_t, (\sigma_t(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}) - \sigma_t(X_t, \mathscr{L}_{X_t}))\mathrm{d}W_t\Big\rangle$$

satisfies

$$\mathrm{d}\langle M\rangle_t \le C_1^2\big\{|X_t^\varepsilon - X_t|^2 + \mathbb{W}_2(\mathscr{L}_{X_t^\varepsilon}, \mathscr{L}_{X_t})^2\big\}^2\mathrm{d}t. \tag{3.15}$$

Then by the BDG inequality, and noting that $\mathbb{W}_2(\mathscr{L}_\xi, \mathscr{L}_\eta)^2 \le \mathbb{E}|\xi - \eta|^2$ for two random variables ξ, η, we may find out a constant $C_2 > 0$ such that

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |X_s^\varepsilon - X_s|^2\Big] \le \varepsilon^2|\eta|^2 + 2C_1\int_0^t \mathbb{E}|X_s^\varepsilon - X_s|^2\mathrm{d}s + C_2\mathbb{E}\sqrt{\langle M\rangle_t}. \tag{3.16}$$

Noting that $\mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon}, \mathscr{L}_{X_s})^2 \le \mathbb{E}|X_s^\varepsilon - X_s|^2$, (3.15) yields

$$\begin{aligned}
&C_2\mathbb{E}\sqrt{\langle M\rangle_t} \le C_1C_2\mathbb{E}\bigg(\int_0^t \big\{|X_s^\varepsilon - X_s|^2 + \mathbb{W}_2(\mathscr{L}_{X_s^\varepsilon}, \mathscr{L}_{X_s})^2\big\}^2\mathrm{d}s\bigg)^{\frac{1}{2}} \\
&\le C_1C_2\mathbb{E}\bigg(\sup_{s\in[0,t]}\big\{|X_s^\varepsilon - X_s|^2 + \mathbb{E}|X_s^\varepsilon - X_s|^2\big\}\int_0^t \big\{|X_s^\varepsilon - X_s|^2 + \mathbb{E}|X_s^\varepsilon - X_s|^2\big\}\mathrm{d}s\bigg)^{\frac{1}{2}} \\
&\le \frac{1}{2}\mathbb{E}\Big[\sup_{s\in[0,t]} |X_s^\varepsilon - X_s|^2\Big] + \frac{C_3}{2}\int_0^t \mathbb{E}|X_s^\varepsilon - X_s|^2\,\mathrm{d}s
\end{aligned}$$

for some constant $C_3 > 0$. Combining this with (3.16) and noting that due to (1.7)

$$\mathbb{E}\Big[\sup_{s\in[0,t]}|X_s^\varepsilon - X_s|^2\Big] < \infty,$$

we arrive at

$$\mathbb{E}\Big[\sup_{s\in[0,t]}|X_s^\varepsilon - X_s|^2\Big] \le 2\varepsilon^2|\eta|^2 + C_3\int_0^t \mathbb{E}|X_s^\varepsilon - X_s|^2\mathrm{d}s, \quad t\in[0,T], \varepsilon>0.$$

Therefore, Gronwall's inequality gives

$$\sup_{\varepsilon\in(0,1]}\mathbb{E}\Big[\sup_{t\in[0,T]}|\xi^\varepsilon(t)|^2\Big] = \sup_{\varepsilon\in(0,1]}\frac{1}{\varepsilon^2}\mathbb{E}\Big[\sup_{s\in[0,T]}|X_s^\varepsilon - X_s|^2\Big] \le 2\mathrm{e}^{C_3 T}\mathbb{E}|\eta|^2 < \infty.$$

□

For any differentiable (real, vector, or matrix valued) function f on $\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)$, let

$$\begin{aligned}\Xi_f^\varepsilon(t) = &\frac{f(X_t^\varepsilon, \mathscr{L}_{X_t^\varepsilon}) - f(X_t,\mathscr{L}_{X_t})}{\varepsilon} - \nabla_{\xi^\varepsilon(t)}f(\cdot,\mathscr{L}_{X_t})(X_t)\\ &- \big\{\mathbb{E}\langle D^L f(y,\cdot)(\mathscr{L}_{X_t})(X_t), \xi^\varepsilon(t)\rangle\big\}\big|_{y=X_t}, \quad t\in[0,T],\varepsilon>0.\end{aligned} \tag{3.17}$$

Lemma 3.4. *Assume* **(H)**. *For any (real, vector, or matrix valued) $C^{1,(1,0)}$-function f on $\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)$ with*

$$K_f := \sup_{(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)}\big(|\nabla f(\cdot,\mu)(x)|^2 + \|D^L f(x,\cdot)(\mu)\|^2_{L^2(\mu)}\big) < \infty, \tag{3.18}$$

there holds

$$\big|\Xi_f^\varepsilon(t)\big|^2 \le 4K_f\big(\mathbb{E}|\xi^\varepsilon(t)|^2 + |\xi^\varepsilon(t)|^2\big) \quad \textit{and} \quad \lim_{\varepsilon\downarrow 0}\mathbb{E}\big|\Xi_f^\varepsilon(t)\big|^2 = 0, \quad t\in[0,T]. \tag{3.19}$$

Proof. Let $X_t^\varepsilon(s) = X_t + s(X_t^\varepsilon - X_t)$, $s\in[0,1]$. By the chain rule and (3.1), we have

$$\begin{aligned}&\frac{f(X_t^\varepsilon,\mathscr{L}_{X_t^\varepsilon}) - f(X_t,\mathscr{L}_{X_t})}{\varepsilon} = \frac{1}{\varepsilon}\int_0^1\Big\{\frac{\mathrm{d}}{\mathrm{d}s}f\big(X_t^\varepsilon(s),\mathscr{L}_{X_t^\varepsilon(s)}\big)\Big\}\mathrm{d}s\\ &= \int_0^1\Big\{\nabla_{\xi^\varepsilon(t)}f(\cdot,\mathscr{L}_{X_t^\varepsilon(s)})(X_t^\varepsilon(s)) + \big(\mathbb{E}\big\langle D^L f(y,\cdot)(\mathscr{L}_{X_t^\varepsilon(s)})(X_t^\varepsilon(s)),\xi^\varepsilon(t)\big\rangle\big)\big|_{y=X_t^\varepsilon(s)}\Big\}\mathrm{d}s.\end{aligned}$$

Combining this with (3.18) we obtain

$$\begin{aligned}\big|\Xi_f^\varepsilon(t)\big|^2 \le\ & 2\int_0^1\Big|\nabla_{\xi^\varepsilon(t)}\big\{f(\cdot,\mathscr{L}_{X_t^\varepsilon(s)})(X_t^\varepsilon(s)) - f(\cdot,\mathscr{L}_{X_t})(X_t)\big\}\Big|^2\mathrm{d}s\\ &+ 2\int_0^1\Big|\big(\mathbb{E}\big\langle D^L f(y,\cdot)(\mathscr{L}_{X_t^\varepsilon(s)})(X_t^\varepsilon(s)),\xi^\varepsilon(t)\big\rangle\big)\big|_{y=X_t^\varepsilon(s)}\\ &\qquad - \big(\mathbb{E}\big\langle D^L f(y,\cdot)(\mathscr{L}_{X_t})(X_t),\xi^\varepsilon(t)\big\rangle\big)\big|_{y=X_t}\Big|^2\mathrm{d}s\\ \le\ & 8K_f\big(|\xi^\varepsilon(t)|^2 + \mathbb{E}|\xi^\varepsilon(t)|^2\big).\end{aligned} \tag{3.20}$$

So, the first inequality in (3.19) holds. Moreover, Lemma 3.3 implies

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}\bigg[\sup_{s\in[0,1]}|X_t^\varepsilon(s)-X_t|^2\bigg]\le \lim_{\varepsilon\downarrow 0}\mathbb{E}|X_t^\varepsilon-X_t|^2=0.$$

Thus, the $C^{1,(1,0)}$-property of f, Lemma 3.3 and the first inequality in (3.20) yield that $\Xi_f^\varepsilon(t)\to 0$ in probability as $\varepsilon\to 0$. Combining this with the first inequality in (3.19), Lemma 3.3, and using the dominated convergence theorem, we derive $\lim_{\varepsilon\downarrow 0}\mathbb{E}\big|\Xi_f^\varepsilon(t)\big|^2=0$. □

Proof of Proposition 3.2. Let $(\Xi_b^\varepsilon(t),K_{b_t})$ and $(\Xi_\sigma^\varepsilon(t),K_{\sigma_t})$ be defined as in (3.17) and (3.18) for b_t and σ_t replacing f respectively. By **(H)**, there exists a constant $C_1>0$ such that

$$\sup_{t\in[0,T]}\big(K_{b_t}+K_{\sigma_t}\big)\le C_1<\infty.$$

Then Lemma 3.4 gives

$$\begin{aligned}&\big|\Xi_b^\varepsilon(t)\big|^2+\big|\Xi_\sigma^\varepsilon(t)\big|^2\le 4C\big(|\xi^\varepsilon(t)|^2+\mathbb{E}|\xi^\varepsilon(t)|^2\big),\\&\lim_{\varepsilon\downarrow 0}\mathbb{E}\big(\big|\Xi_b^\varepsilon(t)\big|^2+\big|\Xi_\sigma^\varepsilon(t)\big|^2\big)=0,\quad t\in[0,T].\end{aligned}\tag{3.21}$$

By (3.10), (3.13), and (3.17) for b_t and σ_t replacing f, we have

$$\begin{aligned}\xi^\varepsilon(t)=&\int_0^t\Big\{\Xi_b^\varepsilon(s)+\nabla_{\xi^\varepsilon(s)}b_s(\cdot,\mathscr{L}_{X_s})(X_s)+\big(\mathbb{E}\langle D^Lb_s(y,\cdot)(\mathscr{L}_{X_s})(X_s),\xi^\varepsilon(s)\rangle\big)\big|_{y=X_s}\Big\}\mathrm{d}s\\&+\int_0^t\Big\langle\Xi_\sigma^\varepsilon(s)+\nabla_{\xi^\varepsilon(s)}\sigma_s(\cdot,\mathscr{L}_{X_s})(X_s)+\big(\mathbb{E}\langle D^L\sigma_s(y,\cdot)(\mathscr{L}_{X_s})(X_s),\xi^\varepsilon(s)\rangle\big)\big|_{y=X_s},\ \mathrm{d}W_s\Big\rangle\end{aligned}$$

for $t\in[0,T]$. So, for any $\varepsilon,\delta\in(0,1]$, $\xi^{\varepsilon,\delta}(t):=\xi^\varepsilon(t)-\xi^\delta(t)$ satisfies

$$\begin{aligned}|\xi^{\varepsilon,\delta}(t)|^2\le\ &4\int_0^t\big|\Xi_b^\varepsilon(s)-\Xi_b^\delta(s)\big|^2\mathrm{d}s+4\bigg|\int_0^t\big\langle\Xi_\sigma^\varepsilon(s)-\Xi_\sigma^\delta(s),\mathrm{d}W_s\big\rangle\bigg|^2\\&+4T\int_0^t\Big|\nabla_{\xi^{\varepsilon,\delta}(s)}b_s(\cdot,\mathscr{L}_{X_s})(X_s)+\big(\mathbb{E}\langle D^Lb_s(y,\cdot)(\mathscr{L}_{X_s})(X_s),\xi^{\varepsilon,\delta}(s)\rangle\big)|_{y=X_s}\Big|^2\mathrm{d}s\\&+4\bigg|\int_0^t\Big\langle\nabla_{\xi^{\varepsilon,\delta}(s)}\sigma_s(\cdot,\mathscr{L}_{X_s})(X_s)+\big(\mathbb{E}\langle D^L\sigma_s(y,\cdot)(\mathscr{L}_{X_s})(X_s),\xi^{\varepsilon,\delta}(s)\rangle\big)|_{y=X_s},\ \mathrm{d}W_s\Big\rangle\bigg|^2.\end{aligned}$$

Combining this with **(H)** and using the BDG inequality, we find out a constant $C_2>0$ such that

$$\begin{aligned}\mathbb{E}\bigg[\sup_{s\in[0,t]}\xi^{\varepsilon,\delta}(s)\bigg]\le\ &C_2\int_0^T\mathbb{E}\Big(\big|\Xi_b^\varepsilon(s)-\Xi_b^\delta(s)\big|^2+\big|\Xi_\sigma^\varepsilon(s)-\Xi_\sigma^\delta(s)\big|^2\Big)\mathrm{d}s\\&+C_2\int_0^t\mathbb{E}|\xi^{\varepsilon,\delta}(s)|^2\ \mathrm{d}s,\quad t\in[0,T].\end{aligned}$$

Since Lemma 3.3 ensures that $\mathbb{E}\big[\sup_{s\in[0,t]}\xi^\varepsilon(s)\big]<\infty$, by Gronwall's lemma this yields

$$\mathbb{E}\bigg[\sup_{s\in[0,T]}\xi^{\varepsilon,\delta}(s)\bigg]\le C_2\mathrm{e}^{C_2T}\int_0^T\mathbb{E}\Big(\big|\Xi_b^\varepsilon(s)-\Xi_b^\delta(s)\big|^2+\big|\Xi_\sigma^\varepsilon(s)-\Xi_\sigma^\delta(s)\big|^2\Big)\mathrm{d}s.$$

Combining this with (3.21) and Lemma 3.3, and applying the dominated convergence theorem, we prove the first assertion in Proposition 3.2.

Finally, by (3.10), (3.12), (3.21) and (3.17) for b_t,σ_t replacing f, we conclude that $v_t^\eta:=\nabla_\eta X_t$ solves the SDE (3.11). Since this SDE is linear, the uniqueness is trivial. Then the proof is finished. $\square$

3.3 Malliavin derivative

Consider the Cameron-Martin space

$$\mathbb{H}=\bigg\{h\in C([0,T]\to\mathbb{R}^d):h_0=\mathbf{0},h_t'\text{ exists a.e. }t,\|h\|_{\mathbb{H}}^2:=\int_0^T|h_t'|^2\mathrm{d}t<\infty\bigg\}.$$

Let $\eta\in L^2(\Omega\to\mathbb{R}^d,\mathscr{F}_0,\mathbb{P})$ with $\mathscr{L}_\eta=\mu$, and let μ_T be the distribution of $W_{[0,T]}:=\{W_t\}_{t\in[0,T]}$, which is a probability measure (i.e. Wiener measure) on the path space $\mathscr{C}_T:=C([0,T]\to\mathbb{R}^d)$. For $F\in L^2(\mathbb{R}^d\times\mathscr{C}_T,\mu\times\mu_T)$, $F(\eta,W_{[0,T]})$ is called Malliavin differentiable along direction $h\in\mathbb{H}$, if the directional derivative

$$D_hF(\eta,W_{[0,T]}):=\lim_{\varepsilon\to0}\frac{F(\eta,W_{[0,T]}+\varepsilon h)-F(\eta,W_{[0,T]})}{\varepsilon}$$

exists in $L^2(\Omega,\mathbb{P})$. If the map $\mathbb{H}\ni h\mapsto D_hF\in L^2(\Omega,\mu)$ is bounded, then there exists a unique $DF(\eta,W_{[0,T]})\in L^2(\Omega\to\mathbb{H},\mathbb{P})$ such that $\langle DF(\eta,W_{[0,T]}),h\rangle_{\mathbb{H}}=D_hF(\eta,W_{[0,T]})$ holds in $L^2(\Omega,\mathbb{P})$ for all $h\in\mathbb{H}$. In this case, we write $F(\eta,W_{[0,T]})\in\mathscr{D}(D)$ and call $DF(\eta,W_{[0,T]})$ the Malliavin gradient of $F(\eta,W_{[0,T]})$. It is well known that $(D,\mathscr{D}(D))$ is a closed linear operator from $L^2(\Omega,\mathscr{F}_T,\mathbb{P})$ to $L^2(\Omega\to\mathbb{H},\mathscr{F}_T,\mathbb{P})$. The adjoint operator $(D^*,\mathscr{D}(D^*))$ of $(D,\mathscr{D}(D))$ is called Malliavin divergence. For simplicity, in the sequel we denote $F(\eta,W_{[0,T]})$ by F. Then we have the integration by parts formula

$$\mathbb{E}\big(D_hF\big|\mathscr{F}_0\big)=\mathbb{E}\big(FD^*(h)\big|\mathscr{F}_0\big),\quad F\in\mathscr{D}(D),h\in\mathscr{D}(D^*).\tag{3.22}$$

It is well known that for adapted $h\in L^2(\Omega\to\mathbb{H},\mathbb{P})$, one has $h\in\mathscr{D}(D^*)$ with

$$D^*(h)=\int_0^T\langle h_t',\mathrm{d}W_t\rangle.\tag{3.23}$$

For more details and applications on Malliavin calculus one may refer to [19] and references therein.

To calculate the Malliavian derivative of X_t with $\mathscr{L}_{X_0}=\mu\in\mathscr{P}_2(\mathbb{R}^d)$, we write $X_t=F_t(W.)$ as a functional of the Brownian motion $\{W_s\}_{s\in[0,t]}$. Then by definition, for an adapted $h\in L^2(\Omega\to\mathbb{H},\mathbb{P})$,

$$D_hX_t=\lim_{\varepsilon\downarrow0}\frac{F_t(W.+\varepsilon h.)-F_t(W.)}{\varepsilon},\quad 0\le t\le T.$$

On the other hand, by the pathwise uniqueness of (1.4), $X_t^{h,\varepsilon} := F_t(W_\cdot + \varepsilon h_\cdot)$ solves the SDE

$$\mathrm{d}X_t^{h,\varepsilon} = b_t(X_t^{h,\varepsilon}, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(X_t^{h,\varepsilon}, \mathscr{L}_{X_t})\mathrm{d}(W_t + \varepsilon h_t), \quad X_0^{h,\varepsilon} = X_0, \tag{3.24}$$

which is well-posed due to **(H)** and $h'_\cdot \in L^2(\Omega \times [0,T], \mathbb{P} \times \mathrm{d}t)$. When $\sigma_t(x,\mu)$ does not depend (x,μ), this SDE reduces to a random ODE for $Y_t^{h,\varepsilon} := X_t^{h,\varepsilon} - \sigma_t W_t$, which is well-posed also for non-adapted h like h^α in Theorem 2.3. The main result of this part is the following which is well known by regarding (1.4) as the classical SDE, since in (3.24) the distribution $\mathscr{L}_{X_t}$ does not depend on the variable ε.

Proposition 3.5. *Assume* **(H)**. *Let $h \in L^2(\Omega \to \mathbb{H}, \mathbb{P})$, which is adapted if $\sigma_t(x,\mu)$ depends on x or μ. Then the limit*

$$D_h X_t := \lim_{\varepsilon \downarrow 0} \frac{X_t^{h,\varepsilon} - X_t}{\varepsilon}, \quad t \in [0,T] \tag{3.25}$$

exists in $L^2(\Omega \to \mathscr{C}_T, \mathbb{P})$. Moreover, $(w_t^h := D_h X_t)_{t \in [0,T]}$ is the unique solution to the SDE

$$\begin{aligned} \mathrm{d}w_t^h = &\Big\{\nabla_{w_t^h}\sigma_t(\cdot, \mathscr{L}_{X_t})(X_t)\Big\}\mathrm{d}W_t \\ &+ \Big\{\nabla_{w_t^h} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \sigma_t(\cdot, \mathscr{L}_{X_t})(X_t)h_t'\Big\}\mathrm{d}t, \quad w_0^h = \mathbf{0}. \end{aligned} \tag{3.26}$$

4 Proofs of main results

We first present an integration by parts formula for $\nabla_\eta X_T$ with $\eta \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$, then prove Theorem 2.1, Corollary 2.2 and Theorem 2.3 respectively.

4.1 An integration by parts formula

Theorem 4.1. *Assume* **(H)** *and* (2.1). *Then for any $f \in C_b^1(\mathbb{R}^d)$, $\eta \in L^2(\Omega \to \mathbb{R}^d, \mathbb{P})$, and any $0 \le r < T$ and $g \in C^1([r,T])$ with $g_r = 0$ and $g_T = 1$,*

$$\mathbb{E}\big(\langle \nabla f(X_T), \nabla_\eta X_T\rangle \big| \mathscr{F}_r\big) = \mathbb{E}\bigg(f(X_T)\int_r^T \langle \zeta_t^\eta,\ \mathrm{d}W_t\rangle \bigg| \mathscr{F}_r\bigg) \tag{4.1}$$

holds for

$$\zeta_t^\eta := \sigma_t(X_t)^{-1}\Big\{g_t' v_t^\eta + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), g_t v_t^\eta\rangle\big)\big|_{y=X_t}\Big\}, \quad t \in [0,T].$$

Proof. Having Propositions 3.2 and 3.5 in hands, the proof is more or less standard. For v_t^η solving (3.11), we take

$$h_t = \int_{t\wedge r}^t 1_{\{s \ge r\}}\zeta_s \,\mathrm{d}s, \quad t \in [0,T]. \tag{4.2}$$

By **(H)**, (2.1), and that $h \in L^2(\Omega \to \mathbb{H}, \mathbb{P})$ is adapted, Proposition 3.5 applies. Let $\tilde{v}_t = g_t v_t^\eta$ for $t \in [r, T]$. Then (3.11) and (4.2) imply

$$\begin{aligned}
\mathrm{d}\tilde{v}_t &= \Big\{\nabla_{\tilde{v}_t} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L b_t(y, \cdot)(\mathscr{L}_{X_t})(X_t), \tilde{v}_t\rangle\big)\big|_{y=X_t} + g_t' v_t^\eta\Big\}\mathrm{d}t \\
&\quad + \Big\{\nabla_{\tilde{v}_t}\sigma_t(\cdot, \mathscr{L}_{X_t})(X_t)\Big\}\mathrm{d}W_t \\
&= \Big\{\nabla_{\tilde{v}_t} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \sigma_t(X_t, \mathscr{L}_{X_t})h_t'\Big\}\mathrm{d}t + \Big\{\nabla_{\tilde{v}_t}\sigma_t(X_t)\Big\}\mathrm{d}W_t, \quad t \geq r, \ \tilde{v}_r = \mathbf{0}.
\end{aligned}$$

So, $(\tilde{v}_t)_{t \geq r}$ solves the SDE (3.26) with $\tilde{v}_r = \mathbf{0}$. On the other hand, by (4.2) we have $h_t' = 0$ for $t < r$, so that the solution to (3.26) with $w_0^h = 0$ satisfies $w_r^h = 0$. So, the uniqueness of this SDE from time r implies $\tilde{v}_t = w_t^h$ for all $t \geq r$. Combining this with Propositions 3.2 and 3.5, we obtain

$$\nabla_\eta X_T = v_T^\eta = g_T v_T^\eta = \tilde{v}_T = w_T^h = D_h X_T.$$

Thus, by the chain rule and the integration by parts formula (3.22), for any bounded $\mathscr{F}_r$-measurable $G \in \mathscr{D}(D)$, we have

$$\begin{aligned}
&\mathbb{E}\big(G\langle \nabla f(X_T), \nabla_\eta X_T\rangle\big) = \mathbb{E}\big(G\langle \nabla f(X_T), D_h X_T\rangle\ \big) = \mathbb{E}\big(G D_h f(X_T)\big) \\
&= \mathbb{E}\big(D_h\{G f(X_T)\} - f(X_T) D_h G) = \mathbb{E}(G f(X_T) D^*(h)\big),
\end{aligned}$$

where in the last step we have used $D_h G = 0$ since G is $\mathscr{F}_r$-measurable but $h_t' = 0$ for $t \leq r$. Noting that the class of bounded $\mathscr{F}_r$-measurable $G \in \mathscr{D}(D)$ is dense in $L^2(\Omega, \mathscr{F}_r, \mathbb{P})$, this implies

$$\mathbb{E}\big(\langle \nabla f(X_T), \nabla_\eta X_T\rangle \big| \mathscr{F}_r\big) = \mathbb{E}(f(X_T) D^*(h) \big| \mathscr{F}_r\big).$$

Combining this with

$$D^*(h) = \int_r^T \langle h_t', \mathrm{d}W_t\rangle = \int_r^T \big\langle \zeta_t^\eta, \mathrm{d}W_t\big\rangle$$

due to (3.23) and (4.2), we prove (4.1). □

4.2 Proof of Theorem 2.1

Let $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. We first establish (2.3) for $f \in \mathscr{B}_b(\mathbb{R}^d)$, then construct $\gamma \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that

$$\lim_{\mu(|\phi|^2) \to 0} \frac{|(P_T f)(\mu \circ (\mathrm{Id} + \phi)^{-1}) - (P_T f)(\mu) - \mu(\langle \phi, \gamma\rangle)|}{\sqrt{\mu(|\phi|^2)}} = 0, \tag{4.3}$$

which, by definition, implies that $P_T f$ is L-differentiable at μ with $D^L P_T f(\mu) = \gamma$.

(a) Proof of (2.3) for $f \in \mathscr{B}_b(\mathbb{R}^d)$. When $f \in C_b^1(\mathbb{R}^d)$, (2.3) follows from (4.1) for $\eta = \phi(X_0)$. Below we extend the formula to $f \in \mathscr{B}_b(\mathbb{R}^d)$. For $s \in [0, 1]$, let $X_t^{\phi,s}$ solve (1.4) for $X_0^{\phi,s} = X_0 + s\phi(X_0)$. We have $\mu^{\phi,s} := \mathscr{L}_{X_0^{\phi,s}} = \mu \circ (\mathrm{Id} + s\phi)^{-1}$, and by the definition of

$\nabla_\eta X_T$ for $\eta = \phi(X_0)$,

$$
(4.4)\qquad \begin{aligned}
&(P_T f)(\mu^{\phi,\varepsilon}) - (P_T f)(\mu) = \mathbb{E}[f(X_T^{\phi,\varepsilon}) - f(X_T)] = \int_0^\varepsilon \frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[f(X_T^{\phi,s})]\,\mathrm{d}s \\
&= \int_0^\varepsilon \mathbb{E}\langle (\nabla f)(X_T^{\phi,s}), \nabla_{\phi(X_0)} X_T^{\phi,s}\rangle\,\mathrm{d}s, \quad f \in C_b^1(\mathbb{R}^d).
\end{aligned}
$$

Next, let $(v_t^{\phi,s})_{t\in[0,T]}$ solve (3.11) for $\eta = \phi(X_0)$ and X_t^s replacing X_t, i.e.

$$
(4.5)\qquad \begin{aligned}
\mathrm{d}v_t^{\phi,s} &= \Big\{\nabla_{v_t^{\phi,s}} b_t(\cdot, \mathscr{L}_{X_t^{\phi,s}})(X_t^{\phi,s}) + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t^{\phi,s}})(X_t^{\phi,s}), v_t^{\phi,s}\rangle\big)\big|_{y=X_t^{\phi,s}}\Big\}\mathrm{d}t \\
&\quad + \Big\{\nabla_{v_t^{\phi,s}} \sigma_t(X_t^{\phi,s})\Big\}\mathrm{d}W_t, \quad v_0^{\phi,s} = \phi(X_0).
\end{aligned}
$$

Let

$$
\zeta_t^{\phi,s} := \sigma_t(X_t^{\phi,s})^{-1}\Big\{g_t' v_t^{\phi,s} + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t^{\phi,s}})(X_t^{\phi,s}), g_t v_t^{\phi,s}\rangle\big)\big|_{y=X_t^{\phi,s}}\Big\}, \quad t \in [0,T].
$$

Then (4.4) and (4.1) imply

$$
(4.6)\qquad (P_T f)(\mu^{\phi,\varepsilon}) - (P_T f)(\mu) = \int_0^\varepsilon \mathbb{E}\bigg[f(X_T^{\phi,s}) \int_0^T \big\langle \zeta_t^{\phi,s},\ \mathrm{d}W_t\big\rangle\bigg]\,\mathrm{d}s, \quad f \in C_b^1(\mathbb{R}^d),
$$

By a standard approximation argument, we may extend this formula to all $f \in \mathscr{B}_b(\mathbb{R}^d)$. Indeed, let

$$
\nu_\varepsilon(A) = \int_0^\varepsilon \mathbb{E}\bigg[1_A(X_T^{\phi,s}) \int_0^T \big\langle \zeta_t^{\phi,s},\ \mathrm{d}W_t\big\rangle\bigg]\,\mathrm{d}s, \quad A \in \mathscr{B}(\mathbb{R}^d).
$$

Then ν_ε is a finite signed measure on $\mathbb{R}^d$ with

$$
\int_{\mathbb{R}^d} f \mathrm{d}\nu_\varepsilon = \int_0^\varepsilon \mathbb{E}\bigg[f(X_T^{\phi,s}) \int_0^T \big\langle \zeta_t^{\phi,s},\ \mathrm{d}W_t\big\rangle\bigg]\,\mathrm{d}s, \quad f \in \mathscr{B}_b(\mathbb{R}^d).
$$

So, (4.6) is equivalent to

$$
(4.7)\qquad \int_{\mathbb{R}^d} f \mathrm{d}P_T^*\mu^{\phi,\varepsilon} - \int_{\mathbb{R}^d} f \mathrm{d}P_T^*\mu = \int_{\mathbb{R}^d} f \mathrm{d}\nu_\varepsilon, \quad f \in C_b^1(\mathbb{R}^d).
$$

Since $\nu_{T,\varepsilon} := P_T^*\mu^{\phi,\varepsilon} + P_T^*\mu + |\nu_\varepsilon|$ is a finite measure on $\mathbb{R}^d$, $C_b^1(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d, \nu_{T,\varepsilon})$. Hence, (4.7) holds for all $f \in \mathscr{B}_b(\mathbb{R}^d) \subset L^1(\mathbb{R}^d, \nu_{T,\varepsilon})$. Consequently, (4.6) holds for all $f \in \mathscr{B}_b(\mathbb{R}^d)$. Thus,

$$
(4.8)\qquad \frac{(P_T f)(\mu^{\phi,\varepsilon}) - (P_T f)(\mu)}{\varepsilon} = \frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\bigg[f(X_T^{\phi,s}) \int_0^T \big\langle \zeta_t^{\phi,s},\ \mathrm{d}W_t\big\rangle\bigg]\,\mathrm{d}s, \quad f \in \mathscr{B}_b(\mathbb{R}^d).
$$

It is easy to see from **(H)** that

$$
\lim_{s\to 0} \sup_{t\in[0,T]} \mathbb{E}\big(|X_t^{\phi,s} - X_t|^2 + |v_t^{\phi,s} - v_t^\phi|^2\big) = 0.
$$

So,

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\bigg|\int_0^T \big\langle \zeta_t^{\phi,s}-\zeta_t^\phi, \mathrm{d}W_t\big\rangle\bigg| = 0. \tag{4.9}$$

Combining this with (4.8), we see that (2.3) for $f\in\mathscr{B}_b(\mathbb{R}^d)$ follows from

$$\lim_{\varepsilon\downarrow 0}\mathbb{E}\bigg[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}\int_0^T \big\langle \zeta_t^\phi, \mathrm{d}W_t\big\rangle\bigg]=0,\quad f\in\mathscr{B}_b(\mathbb{R}^d). \tag{4.10}$$

To prove this equality, we denote

$$I_r:=\int_0^r \big\langle \zeta_t^\phi,\mathrm{d}W_t\big\rangle,\quad r\in(0,T).$$

Applying (4.1) with $g_t:=\frac{t-r}{T-r}$ for $t\in[r,T]$ and using **(H)**, we may find out a constant $C(T,r)>0$ such that

$$\begin{aligned}
&\big|\mathbb{E}[I_r\{f(X_T^{\phi,\varepsilon})-f(X_T)\}]\big| = \bigg|\mathbb{E}\bigg[I_r\int_0^\varepsilon \langle\nabla f(X_T^{\phi,s}),\nabla_{\phi(X_0)}X_T^{\phi,s}\rangle \mathrm{d}s\bigg]\bigg|\\
&\le \mathbb{E}\bigg[|I_r|\cdot\bigg|\int_0^\varepsilon \mathbb{E}\big(\langle\nabla f(X_T^{\phi,s}),\nabla_{\phi(X_0)}X_T^{\phi,s}\rangle\big|\mathscr{F}_r\big)\mathrm{d}s\bigg|\bigg]\\
&\le \frac{C(T,r)}{T-r}\|f\|_\infty\int_0^\varepsilon \mathbb{E}\bigg[|I_r|\bigg(\int_r^T \big|v_t^{\phi,s}\big|^2\mathrm{d}t\bigg)^{\frac12}\bigg]\mathrm{d}s,\quad f\in C_b^1(\mathbb{R}^d).
\end{aligned}$$

By the argument extending (4.6) from $f\in C_b^1(\mathbb{R}^d)$ to $f\in\mathscr{B}_b(\mathbb{R}^d)$, we conclude from this that for any $r\in(0,T)$,

$$\lim_{\varepsilon\downarrow 0}\sup_{\|f\|_\infty\le 1}\big|\mathbb{E}[I_r\{f(X_T^{\phi,\varepsilon})-f(X_T)\}]\big|=0.$$

Therefore,

$$\begin{aligned}
&\limsup_{\varepsilon\downarrow 0}\sup_{\|f\|_\infty\le 1}\bigg|\mathbb{E}\bigg[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}\int_0^T\big\langle\zeta_t^\phi,\mathrm{d}W_t\big\rangle\bigg]\bigg|\\
&=\limsup_{\varepsilon\downarrow 0}\sup_{\|f\|_\infty\le 1}\bigg|\mathbb{E}\bigg[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}\int_r^T\big\langle\zeta_t^\phi,\mathrm{d}W_t\big\rangle\bigg]\bigg|\\
&\le 2\bigg(\mathbb{E}\int_r^T|\zeta_t^\phi|^2\mathrm{d}t\bigg)^{\frac12},\quad r\in(0,T).
\end{aligned} \tag{4.11}$$

By letting $r\uparrow T$ we prove (4.10).

(b) For any $f\in\mathscr{B}_b(\mathbb{R}^d)$, we intend to find out $\gamma\in L^2(\mathbb{R}^d\to\mathbb{R}^d,\mu)$ such that

$$\mathbb{E}\bigg[f(X_T)\int_0^T\big\langle\zeta_t^\phi,\mathrm{d}W_t\big\rangle\bigg]=\mu(\langle\phi,\gamma\rangle),\quad \phi\in L^2(\mathbb{R}^d\to\mathbb{R}^d,\mu). \tag{4.12}$$

When $f \in C_b(\mathbb{R}^d)$, in step (c) we will deduce from this and (2.3) that $\gamma = D^L P_T f(\mu)$. To construct the desired γ, consider the SDE

$$\mathrm{d}X_t^\phi = b_t(X_t^\phi, \mathscr{L}_{X_t^\phi})\mathrm{d}t + \sigma_t(X_t^\phi)\mathrm{d}W_t, \quad X_0^\phi = X_0 + \phi(X_0),$$

and let v_t^ϕ solve (2.2). Since (2.2) is a linear equation for v_t^ϕ with initial value $\phi(X_0) \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$, the functional

$$L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu) \ni \phi \mapsto L\phi := \mathbb{E}\bigg[f(X_T)\int_0^T \big\langle \zeta_t^\phi, \mathrm{d}W_t\big\rangle\bigg]$$

is linear, and by **(H)** and (2.1), there exists a constant $C(T) > 0$ such that

$$|L\phi|^2 \le C(T)\,\mathbb{E}|\phi(X_0)|^2 = C(T)\,\mu(|\phi|^2), \quad \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu).$$

Then L is a bounded linear functional on the Hilbert space $L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$. By Riesz's representation theorem, there exists a unique $\gamma \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$ such that

$$L\phi = \mu(\langle \gamma, \phi\rangle), \quad \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu).$$

Therefore, (4.12) holds.

(c) Now, for $f \in \mathscr{B}_b(\mathbb{R}^d)$, we intend to verify (4.3) for γ in (4.12), so that $P_T f$ is L-differentiable with $D^L(P_T f)(\mu) = \gamma$. By (4.8) for $\varepsilon = 1$, we have

$$(P_T f)(\mu^1) - (P_T f)(\mu) = \int_0^1 \mathbb{E}\bigg[f(X_T^{\phi,s})\int_0^T \big\langle \zeta_t^{\phi,s},\ \mathrm{d}W_t\big\rangle\bigg], \quad f \in \mathscr{B}_b(\mathbb{R}^d). \tag{4.13}$$

For $\mathbb{R}^d$ random variables X, v, let

$$N_t(X, v) = \sigma_t(X)^{-1}\Big\{g_t' v + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_X)(X), g_t v\rangle\big)\big|_{y=X}\Big\}, \quad t \in [0,T].$$

Then $\zeta_t^{\phi,s} = N_t(X_t^{\phi,s}, v^{\phi,s})$ and $\zeta_t^\phi = N_t(X_t, v^\phi)$. Combining this with (4.12) and (4.13), and noting that $\mu^1 = \mu\circ(\mathrm{Id}+\phi)^{-1})$, we arrive at

$$\frac{|(P_T f)(\mu\circ(\mathrm{Id}+\phi)^{-1})) - (P_T f)(\mu) - \mu(\langle \phi, \gamma\rangle)|}{\sqrt{\mu(|\phi|^2)}} \le\ \varepsilon_1(\phi) + \varepsilon_2(\phi) + \varepsilon_3(\phi), \tag{4.14}$$

where

$$\begin{aligned}
\varepsilon_1(\phi) &:= \frac{1}{\sqrt{\mu(|\phi|^2)}}\int_0^1 \mathbb{E}\bigg|\big(f(X_T^{\phi,s}) - f(X_T)\big)\int_0^T \langle \zeta_t^{\phi,s}, \mathrm{d}W_t\rangle\bigg|\mathrm{d}s,\\
\varepsilon_2(\phi) &:= \frac{\|f\|_\infty}{\sqrt{\mu(|\phi|^2)}}\int_0^1 \mathbb{E}\bigg|\int_0^T \big\langle N_t(X_t^{\phi,s}, v^\phi) - N_t(X_t, v^\phi), \mathrm{d}W_t\big\rangle\bigg|\mathrm{d}s,\\
\varepsilon_3(\phi) &:= \frac{\|f\|_\infty}{\sqrt{\mu(|\phi|^2)}}\int_0^1 \mathbb{E}\bigg|\int_0^T \big\langle N_t(X_t^{\phi,s}, v^{\phi,s}) - N_t(X_t^{\phi,s}, v^\phi), \mathrm{d}W_t\big\rangle\bigg|\mathrm{d}s.
\end{aligned}$$

It is easy to deduce from **(H)** that for any $p \geq 2$ there exists a constant $c(p) > 0$ such that

$$\sup_{t\in[0,T],s\in[0,1]} \mathbb{E}\big(|X_t^{\phi,s} - X_t|^p + |v_t^{\phi,s}|^p\big|\mathscr{F}_0\big) \leq c(p)|\phi(X_0)|^p. \tag{4.15}$$

Combining this with the continuity of $\sigma_t(x)$ in x uniformly in $t \in [0,T]$, we conclude that

$$\lim_{\mu(|\phi|^2)\to 0} \varepsilon_2(\phi) = 0. \tag{4.16}$$

Next, by the argument deducing (2.3) from (4.8), it is easy to see that (4.15) implies

$$\lim_{\mu(|\phi|^2)\to 0} \varepsilon_1(\phi) = 0. \tag{4.17}$$

Moreover, by the SDEs for $v_t^{\phi,s}$ and v_t^ϕ we have

$$\mathrm{d}(v_t^{\phi,s} - v_t^\phi) = \big\{A_t(v_t^{\phi,s} - v_t^\phi) + \tilde{A}_t v_t^{\phi,s}\big\}\mathrm{d}t + \big\{B_t(v_t^{\phi,s} - v_t^\phi) + \tilde{B}_t v_t^\phi\big\}\mathrm{d}W_t,$$

where for a square integrable random variable v on $\mathbb{R}^d$,

$$\begin{aligned}
A_t v &:= \nabla_v b_t(\cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), v\rangle\big)\big|_{y=X_t},\\
\tilde{A}_t v &:= \nabla_v b_t(\cdot, \mathscr{L}_{X_t^{\phi,s}})(X_t^{\phi,s}) + \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t^{\phi,s}})(X_t^{\phi,s}), v\rangle\big)\big|_{y=X_t^{\phi,s}}\\
&\qquad - \nabla_v b_t(\cdot, \mathscr{L}_{X_t})(X_t) - \big(\mathbb{E}\langle D^L b_t(y,\cdot)(\mathscr{L}_{X_t})(X_t), v\rangle\big)\big|_{y=X_t},\\
B_t v &:= \nabla_v \sigma_t(X_t), \quad \tilde{B}_t v := \nabla_v \sigma_t(X_t^{\phi,s}) - \nabla_v \sigma_t(X_t).
\end{aligned}$$

Combining this with (4.15) and **(H)**, there exists a constant $c > 0$ such that

$$\mathrm{d}|v_t^{\phi,s} - v_t^\phi|^2 \leq c|v_t^{\phi,s} - v_t^\phi|^2\mathrm{d}t + c\big(\|\tilde{A}_t\|^2 + \|\tilde{B}_t\|^2\big)\big(|v_t^{\phi,s}|^2 + |v_t^\phi|^2\big)\mathrm{d}t + \mathrm{d}M_t, \ \ |v_0^{\phi,s} - v_0^\phi| = 0 \tag{4.18}$$

holds for some martingale M_t, and that

$$\|\tilde{A}_t\|^2 + \|\tilde{B}_t\|^2 \leq c, \quad \lim_{\mu(|\phi|^2)\to 0} \big(\|\tilde{A}_t\|^2 + \|\tilde{B}_t\|^2\big) = 0, \quad t \in [0,T], s \in [0,1]. \tag{4.19}$$

By (4.18) and (4.15) for $p = 4$, there exists a constant $c' > 0$ such that

$$\begin{aligned}
&\mathbb{E}(|v_t^{\phi,s} - v_t^\phi|^2|\mathscr{F}_0)\\
&\leq c\int_0^t \mathbb{E}(|v_r^{\phi,s} - v_r^\phi|^2|\mathscr{F}_0)\mathrm{d}r + 2c\int_0^T \sqrt{\mathbb{E}(\|\tilde{A}_t\|^4 + \|\tilde{B}_t\|^4|\mathscr{F}_0)} \cdot \sqrt{\mathbb{E}(|v_t^{\phi,s}|^4 + |v_t^\phi|^4|\mathscr{F}_0)}\,\mathrm{d}t\\
&\leq c\int_0^t \mathbb{E}(|v_r^{\phi,s} - v_r^\phi|^2|\mathscr{F}_0)\mathrm{d}r + c'\varepsilon(\phi)|\phi(X_0)|^2, \quad s \in [0,1], t \in [0,T],
\end{aligned}$$

where

$$\varepsilon(\phi) := \int_0^T \sqrt{\mathbb{E}(\|\tilde{A}_t\|^4 + \|\tilde{B}_t\|^4|\mathscr{F}_0)}\,\mathrm{d}t.$$

Then Gronwall's lemma and (4.19) yield

$$\sup_{s\in[0,T]} \mathbb{E}(|v_t^{\phi,s}-v_t^\phi|^2|\mathscr{F}_0) \le c'\mathrm{e}^{cT}\varepsilon(\phi)|\phi(X_0)|^2,$$
$$\lim_{\mu(|\phi|^2)\to 0} \mathbb{E}\varepsilon(\phi)=0.$$

Combining this with the definition of $\varepsilon_3(\phi)$, **(H)**, and Jensen's inequality for the conditional expectation $\mathbb{E}(\cdot|\mathscr{F}_0)$, we may find out constants $C_1, C_2>0$ depending on $\|f\|_\infty$ and T such that

$$\begin{aligned}
&\lim_{\mu(|\phi|^2)\to 0}\varepsilon_3(\phi) \le \lim_{\mu(|\phi|^2)\to 0}\frac{C_1}{\sqrt{\mu(|\phi|^2)}}\int_0^1\mathbb{E}\bigg(\int_0^T|v_t^{\phi,s}-v_t^\phi|^2\mathrm{d}t\bigg)^{\frac{1}{2}}\mathrm{d}s\\
&\le \lim_{\mu(|\phi|^2)\to 0}\frac{C_1}{\sqrt{\mu(|\phi|^2)}}\int_0^1\mathbb{E}\bigg(\int_0^T\mathbb{E}(|v_t^{\phi,s}-v_t^\phi|^2|\mathscr{F}_0)\mathrm{d}t\bigg)^{\frac{1}{2}}\mathrm{d}s\\
&\le \lim_{\mu(|\phi|^2)\to 0}\frac{C_2}{\sqrt{\mu(|\phi|^2)}}\int_0^1\mathbb{E}\big(|\phi(X_0)|\sqrt{\varepsilon(\phi)}\big)\mathrm{d}s\\
&\le \lim_{\mu(|\phi|^2)\to 0}\frac{C_2\sqrt{(\mathbb{E}|\phi(X_0)|^2)\mathbb{E}\varepsilon(\phi)}}{\sqrt{\mu(|\phi|^2)}} = \lim_{\mu(|\phi|^2)\to 0}C_2\sqrt{\mathbb{E}\varepsilon(\phi)}=0.
\end{aligned}$$

This, together with (4.14), (4.16) and (4.17), implies (4.3). Therefore, P_Tf is L-differentiable at μ with $D^L(P_Tf)(\mu)=\gamma$.

(d) Finally, (2.3) and (4.8) imply

$$\begin{aligned}
&\Big|\frac{P_T^*\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}-P_T^*\mu}{\varepsilon}(f)-(\psi P_T^*\mu)(f)\Big|\\
&=\bigg|\frac{(P_Tf)(\mu^{\phi,\varepsilon})-(P_Tf)(\mu)}{\varepsilon}-\mathbb{E}\bigg[f(X_T)\int_0^T\langle\zeta_t^\phi,\mathrm{d}W_t\rangle\bigg]\bigg|\\
&\le\frac{\|f\|_\infty}{\varepsilon}\int_0^\varepsilon\mathbb{E}\bigg|\int_0^T\langle\zeta_t^{\phi,s}-\zeta_t^\phi,\mathrm{d}W_t\rangle\bigg|\mathrm{d}s\\
&\quad+\frac{1}{\varepsilon}\bigg|\mathbb{E}\bigg[\big\{f(X_T^{\phi,\varepsilon})-f(X_T)\big\}\int_0^T\big\langle\zeta_t^\phi,\mathrm{d}W_t\big\rangle\bigg]\bigg|\mathrm{d}s.
\end{aligned}$$

Combining this with (4.9) and (4.10) we prove (2.4).

4.3 Proof of Corollary 2.2

Proof of (1). By **(H)** and (2.2), there exists a martingale M_t such that

$$\mathrm{d}|v_t^\phi|^2\le 4K_t|v_t^\phi|(|v_t^\phi|+\mathbb{E}|v_t^\phi|)\mathrm{d}t+\mathrm{d}M_t,\quad |v_0^\phi|^2=|\phi(X_0)|^2, \tag{4.20}$$

where $K(t)$ is increasing in $t\ge 0$. Then

$$\mathbb{E}|v_t^\phi|^2\le\mathbb{E}|\phi(X_0)|^2+4K_t\int_0^t\big\{\mathbb{E}|v_s^\phi|^2+(\mathbb{E}|v_s^\phi|)^2\big\}\mathrm{d}s\le\mu(|\phi|^2)+8K_t\int_0^t\mathbb{E}|v_s^\phi|^2\mathrm{d}s.$$

By Gronwall's inequality this implies

$$\mathbb{E}|v_t^\phi|^2 \le \mathrm{e}^{8K_t t}\mu(|\phi|^2), \quad t\in[0,T]. \tag{4.21}$$

Next, since $\mathbb{E}\int_0^T \langle \xi_t^\phi,\ \mathrm{d}W_t\rangle = 0$, (2.3) is equivalent to

$$D_\phi^L(P_Tf)(\mu) = \mathbb{E}\bigg[\big\{f(X_T) - P_Tf(\mu)\big\}\int_0^T \big\langle \zeta_t^\phi,\ \mathrm{d}W_t\big\rangle\bigg].$$

Combining this with (4.21) and using Jensen's inequality, when $\mu(|\phi|^2)\le 1$ we have

$$\begin{aligned}|D_\phi^L(P_Tf)(\mu)|^2 &\le \big\{(P_Tf^2)(\mu) - (P_Tf(\mu))^2\big\}\int_0^T \mathbb{E}\big|\zeta_t^\phi\big|^2\mathrm{d}t\\ &\le \big\{(P_Tf^2)(\mu) - (P_Tf(\mu))^2\big\}\int_0^T \big(|g_t'| + K(t)|g_t|\big)^2\lambda_t^2\mathrm{e}^{8tK_t}\mathrm{d}t\end{aligned}$$

for any $g\in C^1([0,T])$ with $g_0=0$ and $g_T=1$. Taking $g_t=\frac{t}{T},\ t\in[0,T]$, we prove the estimate (2.5). $\square$

Proof of (2). Let $f\in\mathscr{B}_b(\mathbb{R}^d)$ with $\|f\|_\infty\le 1$. By Theorem 2.1, P_Tf is L-differentiable. Moreover, by Theorem 4.1, P_Tf is Lipschitz continuous on $\mathscr{P}_2(\mathbb{R}^d)$. Indeed, for any $\mu_1,\mu_2\in\mathscr{P}_2(\mathbb{R}^d)$, let $X_1,X_2\in L^2(\Omega\to\mathbb{R}^d,\mathscr{F}_0,\mathbb{P})$ such that $\mathscr{L}_{X_i}=\mu_i, 1\le i\le 2$, and $\mathbb{E}|X_1-X_2|^2 = \mathbb{W}_2(\mu_1,\mu_2)^2$. Let X_t^s be the solution to (1.4) with $X_0 = X_1+s(X_2-X_1), s\in[0,1]$. Then Theorem 4.1 implies

$$\begin{aligned}|P_Tf(\mu_1) - P_Tf(\mu_2)|^2 &= |\mathbb{E}f(X_T^0) - \mathbb{E}f(X_T^1)|^2 = \bigg|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}f(X_T^s)\,\mathrm{d}s\bigg|^2\\ &= \bigg|\int_0^1 \mathbb{E}\langle\nabla f(X_T^s), \nabla_{X_2-X_1}X_T^s\rangle\mathrm{d}s\bigg|^2 \le c\mathbb{E}|X_2-X_1|^2 = c\mathbb{W}_2(\mu_1,\mu_2)^2\end{aligned}$$

for some constant $c>0$.

To apply Proposition 3.1, we take $\{\mu_n,\nu_n\}_{n\ge1}\subset\mathscr{P}_2(\mathbb{R}^d)$ which have compact supports and are absolutely continuous with respect to the Lebesgue measure, such that

$$\lim_{n\to\infty}\big\{\mathbb{W}_2(\mu,\mu_n) + \mathbb{W}_2(\nu,\nu_n)\big\} = 0. \tag{4.22}$$

According to [4], see also [6, Theorem 5.8], for any $n\ge1$ there exists a unique map $\phi_n\in L^2(\mathbb{R}^d\to\mathbb{R}^d,\mu)$ such that

$$\nu_n = \mu_n\circ(\mathrm{Id}+\phi_n)^{-1}, \quad \mathbb{W}_2(\mu_n,\nu_n)^2 = \mu_n(|\phi_n|^2). \tag{4.23}$$

Let $X_n\in L^2(\Omega\to\mathbb{R}^d,\mathscr{F}_0,\mathbb{P})$ such that $\mathscr{L}_{X_n}=\mu_n$. By Proposition 3.1, (2.5) and (4.23), we obtain

$$|(P_Tf)(\mu_n) - (P_Tf)(\nu_n)|^2 = \bigg|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}s}(P_Tf)(\mathscr{L}_{X_n+s\phi_n(X_n)})\,\mathrm{d}s\bigg|^2$$

$$= \left| \int_0^1 \mathbb{E}\big\langle D^L (P_T f)(\mathscr{L}_{X_n + s\phi_n(X_n)})(X_n + s\phi_n(X_n)), \phi_n(X_n)\big\rangle \,\mathrm{d}s \right|^2$$
$$\le \frac{\|f\|_\infty^2 \mu_n(|\phi_n|^2)}{\int_0^T \lambda_t^{-2} \mathrm{e}^{-8tK_t}\mathrm{d}t} = \frac{\|f\|_\infty^2 \mathbb{W}_2(\mu_n, \nu_n)^2}{\int_0^T \lambda_t^{-2}\mathrm{e}^{-8tK_t}\mathrm{d}t}.$$

By the continuity of $P_T f$ and (4.22), by letting $n \to \infty$ we prove

$$|(P_T f)(\mu) - (P_T f)(\nu)|^2 \le \frac{\mathbb{W}_2(\mu,\nu)^2}{\int_0^T \lambda_t^{-2}\mathrm{e}^{-8tK_t}\mathrm{d}t}, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d), \quad f \in \mathscr{B}_b(\mathbb{R}^d), \|f\|_\infty \le 1.$$

Therefore, (2.6) and (2.7) hold. $\square$

4.4 Proof of Theorem 2.3

Let $T > r \ge 0, \mu \in \mathscr{P}_2(\mathbb{R}^{m+d})$ and let X_t solve (2.8) with $\mathscr{L}_{X_0} = \mu$. To realize the procedure in the proof of Theorem 2.1 for the present degenerate setting, we first extend Theorem 4.1 using $D^*(h^\alpha_{r,\cdot})$ to replace $\int_r^T \langle \zeta_t^\eta, \mathrm{d}W_t\rangle$, where for a $C^1([r,T] \to \mathbb{R}^{m+d})$-valued random variable $\alpha_\cdot = (\alpha_\cdot^{(1)}, \alpha_\cdot^{(2)})$, let $(h^\alpha_{r,t}, w^\alpha_{r,t})_{t\in[r,T]}$ be the unique solutio to the random ODEs

$$\begin{aligned} \frac{\mathrm{d}h^\alpha_{r,t}}{\mathrm{d}t} &= \sigma_t^{-1}\Big\{\nabla_{\alpha_t} b_t^{(2)}(X_t, \mathscr{L}_{X_t}) - (\alpha_t^{(2)})' \\ &\qquad + \big(\mathbb{E}\langle D^L b_t^{(2)}(y,\cdot)(\mathscr{L}_{X_t})(X_t), \alpha_t + w^\alpha_{r,t}\rangle\big)\big|_{y=X_t}\Big\}, \\ \frac{\mathrm{d}w^\alpha_{r,t}}{\mathrm{d}t} &= \nabla_{w^\alpha_{r,t}} b_t(\cdot, \mathscr{L}_{X_t})(X_t) + (\mathbf{0}, \sigma_t (h^\alpha_{r,t})'), \qquad h^\alpha_{r,r} = 0, w^\alpha_{r,r} = 0. \end{aligned} \tag{4.24}$$

Theorem 4.2. *Assume* (**H1**). *Let $T > r \ge 0$, $\eta \in L^2(\Omega \to \mathbb{R}^{m+d}, \mathscr{F}_0, \mathbb{P})$, and let X_t solve* (2.8) *with $\mathscr{L}_{X_0} = \mu \in \mathscr{P}_2(\mathbb{R}^{m+d})$. If there exists a $C^1([r,T] \to \mathbb{R}^{m+d})$-valued random variable $\alpha_\cdot = (\alpha_\cdot^{(1)}, \alpha_\cdot^{(2)})$ such that $\alpha_r = \nabla_\eta X_r, \alpha_T = \mathbf{0}$,*

$$(\alpha_t^{(1)})' = \nabla_{\alpha_t} b_t^{(1)}(X_t), \quad t \in [r,T], \tag{4.25}$$

and $h^\alpha_{r,\cdot} \in \mathscr{D}(D^*)$, then for any $f \in C_b^1(\mathbb{R}^{m+d})$,

$$\mathbb{E}\big(\langle \nabla f(X_T), \nabla_\eta X_T\rangle \big| \mathscr{F}_r\big) = \mathbb{E}\big(f(X_T)\, D^*(h^\alpha_{r,\cdot})\big|\mathscr{F}_r\big). \tag{4.26}$$

Proof. Letting $w_t = w^\alpha_{r,t} 1_{\{t>r\}}$, Proposition 3.5 implies that $w_t = D_{h^\alpha_{r,\cdot}} X_t,\ t \in [0,T]$. By (4.24), we have

$$w_t = \int_{t\wedge r}^t \Big\{\nabla_{w_s} b_s(\cdot, \mathscr{L}_{X_s})(X_s) + \Big(\mathbf{0}, \sigma_s (h^\alpha_{r,s})'\Big\}\mathrm{d}s, \quad t \in [0,T].$$

Extending α_t with $\alpha_t := \nabla_\eta X_t$ for $t \in [0,r)$, and letting $v_t = w_t + \alpha_t$ for any $t \in [0,T]$, we obtain

$$\begin{aligned} v_t = \alpha_t + \int_{t\wedge r}^t &\Big\{\nabla_{v_s} b_s(\cdot, \mathscr{L}_{X_s})(X_s) + \Big(\mathbf{0}, \big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), v_s\rangle\big)\big|_{y=X_s}\Big) \\ &+ (\mathbf{0}, \sigma_s (h^\alpha_s)' - \big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), w_s + \alpha_s\rangle\big)\big|_{y=X_s}) - \nabla_{\alpha_s} b_s(\cdot, \mathscr{L}_{X_s})(X_s)\Big\}\mathrm{d}s. \end{aligned} \tag{4.27}$$

By (4.25),

$$\int_{t\wedge r}^t \nabla_{\alpha_s} b_s^{(1)}(\cdot,\mathscr{L}_{X_s})(X_s)\,\mathrm{d}s = 1_{\{t>r\}}\big(\alpha_t^{(1)} - \nabla_\eta X_r^{(1)}\big),$$

while the definition of $h_{r,s}^\alpha$ implies

$$\begin{aligned}&\int_{t\wedge r}^t \Big\{\sigma_s(h_s^\alpha)' - \big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), w_s+\alpha_s\rangle\big)\big|_{y=X_s} - \nabla_{\alpha_s} b_s^{(2)}(\cdot,\mathscr{L}_{X_s})(X_s)\Big\}\mathrm{d}s\\ &= -\int_{t\wedge r}^t (\alpha_s^{(2)})'\mathrm{d}s = 1_{\{t>r\}}\big(\nabla_\eta X_r^{(2)} - \alpha_t^{(2)}\big).\end{aligned}$$

Combining these with (4.27) and Proposition 3.2 leads to

$$\begin{aligned}v_t &= \nabla_\eta X_r + \int_{t\wedge r}^t \Big\{\nabla_{v_s} b_s(\cdot,\mathscr{L}_{X_s})(X_s) + \Big(\mathbf{0}, \big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), v_s\rangle\big)\big|_{y=X_s}\Big)\Big\}\mathrm{d}s\\ &= \eta + \int_0^t \Big\{\nabla_{v_s} b_s(\cdot,\mathscr{L}_{X_s})(X_s) + \Big(\mathbf{0}, \big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), v_s\rangle\big)\big|_{y=X_s}\Big)\Big\}\mathrm{d}s,\quad t\in[0,T].\end{aligned}$$

That is, v_t solves (3.11) so that by Proposition 3.2 we obtain $v_t := w_t + \alpha_t = \nabla_\eta X_t$. Since $\alpha_T = 0$, this implies $D_{h_{r,\cdot}^\alpha} X_T = \nabla_\eta X_T$. Thus, for any bounded $\mathscr{F}_r$-measurable $G\in\mathscr{D}(D)$,

$$\begin{aligned}&\mathbb{E}\big[G\langle\nabla f(X_T), \nabla_\eta X_T\rangle\big] = \mathbb{E}\big[G D_{h_{r,\cdot}^\alpha} f(X_T)\big]\\ &= \mathbb{E}\big[D_{h_{r,\cdot}^\alpha}\{Gf(X_T)\} - f(X_T) D_{h_{r,\cdot}^\alpha} G\big] = \mathbb{E}\big[G f(X_T) D^*(h_{r,\cdot}^\alpha)\big],\end{aligned}\tag{4.28}$$

where in the last step we have used the integration by parts formula (3.22) and $D_{h_{r,\cdot}^\alpha} G = 0$ since G is $\mathscr{F}_r$-measurable but

$$D_{h_{r,\cdot}^\alpha} G = \int_0^T (h_{r,\cdot}^\alpha)'(s)\cdot\{(DG)_\cdot\}'(s)\mathrm{d}s = 0,$$

$(h_{r,\cdot}^\alpha)'(s) = 0$ for $s\le r$. Noting that the class of bounded $\mathscr{F}_r$-measurable functions $G\in\mathscr{D}(D)$ is dense in $L^2(\Omega,\mathscr{F}_r,\mathbb{P})$, (4.28) implies (4.26). □

Proof of Theorem 2.3. With Theorem 4.2 in hands, the proof is completely similar to that of Theorem 2.1. Let

$$v_t^\phi = ((v_t^\phi)^{(1)}, (v_t^\phi)^{(2)}) = (\nabla_{\phi(X_0)} X_t^{(1)}, \nabla_{\phi(X_0)} X_t^{(2)}) = \nabla_{\phi(X_0)} X_t,\quad t\in[0,T].$$

For any $0\le r<T$, let

$$\begin{aligned}\alpha_{r,t}^{(2)} =& \frac{T-t}{T-r}(v_t^\phi)^{(2)} - \frac{(t-r)(T-t)B_t^* K_{T,t}^*}{\int_0^T \theta_s^2\mathrm{d}s}\int_t^T \theta_s^2 Q_s^{-1} K_{T,r}(v_t^\phi)^{(1)}\mathrm{d}s\\ &- (t-r)(T-t)B_t^* K_{T,t}^* Q_T^{-1}\int_0^T \frac{T-s}{T} K_{T,s}\nabla^{(2)} b_s^{(1)}(X_s)\phi^{(2)}(X_0)\mathrm{d}s,\quad t\in[r,T],\end{aligned}\tag{4.29}$$

and

$$\alpha_{r,t}^{(1)} = K_{t,r}(v_t^{\phi})^{(1)} + \int_r^t K_{t,s}\nabla^{(2)}_{\alpha_s^{(2)}} b_s^{(1)}(X_s(x))\,\mathrm{d}s, \quad t\in[r,T]. \tag{4.30}$$

Then $\alpha_{r,\cdot} := (\alpha_{r,t}^{(1)}, \alpha_{r,t}^{(2)})$ satisfies

$$\alpha_{r,r} = \nabla_{\phi(X_0)}X_r, \quad \alpha_{r,T}=0,$$

and by (2.9) and Duhamel's formula, (4.30) implies

$$(\alpha_{r,\cdot}^{(1)})'(t) = \nabla_{\alpha_{r,t}} b_t^{(1)}(X_t), \quad t\in[r,T].$$

Moreover, let $h_{r,\cdot}^{\alpha_{r,\cdot}}$ be defined in (4.24) for $\alpha_{r,\cdot}$ replacing α. Noting that **(H1)** and **(H2)** imply [28, (H)] for $l_1=l_2=0$, the proof of [28, Theorem 1.1] with $\phi(s):=(s-r)(T-s)$ for $s\in[r,T]$ ensures that $h_{r,\cdot}^{\alpha_{r,\cdot}}\in\mathscr{D}(D^*)$ with $D^*(h_{r,\cdot}^{\alpha_{r,\cdot}})\in L^p(\mathbb{P})$ for all $p\in(1,\infty)$. So, by Theorem 2.3 with $\eta=\phi(X_0)$ we obtain

$$\mathbb{E}(\langle\nabla f(X_T),\nabla_{\phi(X_0)}X_T\rangle|\mathscr{F}_r) = \mathbb{E}(f(X_T)D^*(h_{r,\cdot}^{\alpha_{r,\cdot}})|\mathscr{F}_r), \quad f\in C_b^1(\mathbb{R}^d), r\in[0,T). \tag{4.31}$$

In particular, taking $r=0$ we obtain $D^*(h)\in L^p(\mathbb{P})$ for all $p\in(1,\infty)$ and

$$D_\phi^L P_T f(\mu) = \mathbb{E}(\langle\nabla f(X_T),\nabla_{\phi(X_0)}X_T\rangle) = \mathbb{E}(f(X_T)D^*(h^\alpha)|\mathscr{F}_r), \quad f\in C_b^1(\mathbb{R}^d). \tag{4.32}$$

Basing on these two formulas, by repeating the proof of Theorem 2.1 with $I_r:=\mathbb{E}(D^*(h^\alpha)|\mathscr{F}_r)$, we prove (2.16) and the L-differentiability of P_Tf for $f\in\mathscr{B}_b(\mathbb{R}^{m+d})$. Finally, the estimates (2.17) and (2.18) follows from (2.16) as in the proof of Theorem 2.1, together with the corresponding estimate on $\mathbb{E}|D^*(h^\alpha)|^2$ as in the proof of [28, Theorem 1.1]. For instance, below we outline the proof of (2.16).

Firstly, for $s\in(0,1)$ let X_t^s solve (2.8) with $X_0^{\phi,s}=X_0+s\phi(X_0)$, let $\mu^{\phi,s}=\mathscr{L}_{X_0^{\phi,s}}=\mu\circ(\mathrm{Id}+\phi)^{-1}$, and let $\alpha_{r,t}^{\phi,s}$ be defined as $\alpha_{r,t}$ with $X_t^{\phi,s}$ replacing X_t. Then as in (4.4) and (4.7), (4.32) implies

$$\begin{aligned}(P_Tf)(\mu^{\phi,\varepsilon})-(P_Tf)(\mu) &= \int_0^\varepsilon \mathbb{E}\langle(\nabla f)(X_T^{\phi,s}),\nabla_{\phi(X_0)}X_T^{\phi,s}\rangle\,\mathrm{d}s\\ &= \int_0^\varepsilon \mathbb{E}\big[f(X_T^{\phi,s})D^*(h^{\alpha^{\phi,s}})\big], \quad f\in C_b^1(\mathbb{R}^{m+d}),\end{aligned} \tag{4.33}$$

where $h^{\alpha^{\phi,s}}:=h_{0,\cdot}^{\alpha_{0,\cdot}^{\phi,s}}$ satisfies

$$\lim_{s\to0}\mathbb{E}|D^*(h^{\alpha^{\phi,s}})-D^*(h)|^2=0. \tag{4.34}$$

By the argument leading to (4.8), (4.33) yields

$$\frac{(P_Tf)(\mu^{\phi,\varepsilon})-(P_Tf)(\mu)}{\varepsilon} = \frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\big[f(X_T^{\phi,s})D^*(h^{\alpha^{\phi,s}})\big]\,\mathrm{d}s, \quad f\in\mathscr{B}_b(\mathbb{R}^{m+d}).$$

Combining this with (4.34), we prove (2.16) provided

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\big[\{f(X_T^{\phi,s})-f(X_T)\}D^*(h^\alpha)\big]\,\mathrm{d}s=0. \tag{4.35}$$

For any $r\in(0,T)$, let $I_r=\mathbb{E}(D^*(h^\alpha)|\mathscr{F}_r)$. By (4.33) we obtain

$$\begin{aligned}&\mathbb{E}\big[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}I_r\big]=\mathbb{E}\big[I_r\mathbb{E}(f(X_T^{\phi,\varepsilon})-f(X_T)|\mathscr{F}_r)\big]\\&=\mathbb{E}\bigg[I_r\int_0^\varepsilon \mathbb{E}\big(\langle\nabla f(X_T^{\phi,s}),\nabla X_T^{\phi,s}\rangle\big|\mathscr{F}_r\big)\,\mathrm{d}s\bigg]=\mathbb{E}\bigg[I_r\int_0^\varepsilon \mathbb{E}\big(f(X_T^{\phi,s})D^*(h_{r,\cdot}^{\alpha_{r,\cdot}})\big|\mathscr{F}_r\big)\,\mathrm{d}s\bigg]\\&=\int_0^\varepsilon \mathbb{E}\big[I_rf(X_T^{\phi,s})D^*(h_{r,\cdot}^{\alpha_{r,\cdot}})\big]\,\mathrm{d}s,\quad f\in C_b^1(\mathbb{R}^d).\end{aligned}$$

Combining this with the argument extending (4.8) from $f\in C_b^1(\mathbb{R}^d)$ to $f\in\mathscr{B}_b(\mathbb{R}^d)$, we obtain

$$\mathbb{E}\big[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}I_r\big]=\int_0^\varepsilon \mathbb{E}\big[I_rf(X_T^{\phi,s})D^*(h_{r,\cdot}^{\alpha_{r,\cdot}})\big]\,\mathrm{d}s,\quad f\in\mathscr{B}_b(\mathbb{R}^d).$$

Consequently,

$$\lim_{\varepsilon\to 0}\mathbb{E}\big[\{f(X_T^{\phi,\varepsilon})-f(X_T)\}I_r\big]=0,\quad f\in\mathscr{B}_b(\mathbb{R}^d),r\in(0,T).$$

Then for any $r\in(0,T)$,

$$\begin{aligned}&\limsup_{\varepsilon\downarrow 0}\bigg|\frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\big[\{f(X_T^{\phi,s})-f(X_T)\}D^*(h^\alpha)\big]\,\mathrm{d}s\bigg|\\&=\limsup_{\varepsilon\downarrow 0}\bigg|\frac{1}{\varepsilon}\int_0^\varepsilon \mathbb{E}\big[\{f(X_T^{\phi,s})-f(X_T)\}\cdot\{D^*(h^\alpha)-I_r\}\big]\,\mathrm{d}s\bigg|\\&\le 2\|f\|_\infty\mathbb{E}|D^*(h^\alpha)-\mathbb{E}(D^*(h^\alpha)|\mathscr{F}_r)|.\end{aligned}$$

Letting $r\uparrow T$ we derive (4.35), and hence prove (2.16) as explained above. □

Acknowledgement. Financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" is acknowledged. The authors would like to thanks Professor Yulin Song for helpful comments.

References

[1] M. Arnaudon, A. Thalmaier, *The differentiation of hypoelliptic diffusion semigroups,* Illinois J. Math. 54(2010), 1285–1311.

[2] D. Baños, *The Bismut-Elworthy-Li formula for mean-field stochastic differential equations,* Ann. l'Inst. H. Poinc. Probab. Statis. 54(2018), 220–233.

[3] J. M. Bismut, *Large Deviations and the Malliavin Calculus,* Boston: Birkhäuser, MA, 1984.

[4] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions,* Comm. Pure Appl. Math. 44, 4 (1991), 375–417.

[5] R. Buckdahn, J. Li, S. Peng, C. Rainer, *Mean-field stochastic differential equations and associated PDEs,* Annal. Probab. 2(2017), 824-878.

[6] P. Cardaliaguet, *Notes on mean field games,* P.-L. Lions lectures at College de France. https://www.researchgate.net/publication/228702832_Notes_on_Mean_Field_Games

[7] J.-F. Chassagneux, D. Crisan, F. Delarue, *Classical solutions to the master equation for large population equilibria,* arXiv:1411.3009, 2014.

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

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

[10] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part I : existence, uniqueness and smothness,* Comm. Part. Diff. Equat. 25(2000), 179–259.

[11] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part II: H-Theorem and Applications,* Comm. Part. Diff. Equat. 25(2000), 261–298.

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

[13] N. Fournier, A. Guillin, *From a Kac-like particle system to the Landau equation for hard potentials and Maxwell molecules,* Ann. Sci. l'ENS 50(2017), 157–199.

[14] H. Guérin, *Existence and regularity of a weak function-solution for some Landau equations with a stochastic approach,* Stoch. Proc. Appl. 101(2002), 303–325.

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

[16] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients,* arXiv:1805.01682.

[17] J. Li, *Mean-field forward and backward SDEs with jumps and associated nonlocal quasi-linear integral-PDEs,* Stoch. Proc. Appl. 128(2018), 3118–3180.

[18] Yu. S. Mishura, A. Yu. Veretennikov, *Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations,* arXiv:1603.02212v4

[19] D. Nualart, *The Malliavin Calculus and Related Topics,* Springer, Berlin, 2006.

[20] P. Ren, F.-Y. Wang, *Space-distribution PDEs for path independent additive functionals of McKean-Vlasov SDEs,* arXiv:1805.10841.

[21] Y. Song, *Gradient estimates and exponential ergodicity for mean-field SDEs with jumps,* to appear in J. Theort. Probab.

[22] A.-S. Sznitman, *Topics in propagation of chaos,* In "École d'Été de Probabilités de Sain-Flour XIX-1989", Lecture Notes in Mathematics 1464, p. 165–251, Springer, Berlin, 1991.

[23] C. Villani, *On the spatially homogeneous Landau equation for Maxwellian Mocecules,* Math. Mod. Meth. Appl. Sci. 8(1998), 957–983.

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

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

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

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

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