Regularity Estimates and Intrinsic-Lions Derivative Formula for Singular McKean-Vlasov SDEs *

Feng-Yu Wang
Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, United Kingdom
wangfy@tju.edu.cn
July 25, 2022

Abstract

Regularity estimates and Bismut formula of $L^k$ ($k \geq 1$) intrinsic-Lions derivative are presented for singular McKean-Vlasov SDEs, where the noise coefficient belongs to a local Sobolev space, and the drift contains a locally integrable time-space term as well as a time-space-distribution term Lipschitz continuous in the space and distribution variables. The results are new also for classical SDEs.

AMS subject Classification: 60B05, 60B10.
Keywords: McKean-Vlasov SDEs, intrinsic derivative, Lions derivative, Bismut formula.

1 Introduction and main results

Since 1984 when Bismut [4] presented his derivative formula for diffusion semigroups on Riemannian manifolds, this type formula has been widely developed and applied. Recently, Bismut formula was established in [21] for singular SDEs with a locally integrable drift.

On the other hand, as crucial probability models characterizing nonlinear Fokker-Planck equations and mean field games, distribution dependent (also called McKean-Vlasov or mean-field) SDEs have been intensively investigated, see for instance the monographs [17, 8] and the survey [12]. In particular, Bismut type formulas of the Lions derivative have been established in [15, 6, 13] for regular McKean-Vlasov SDEs where the drift is at least Dini continuous in

*Supported in part by NNSFC (11771326, 11831014, 11921001).
the space variable, see also [9] for Bismut formula of the decoupled SDE where the distribution parameter is fixed.

In this paper, we aim to establish Bismut formula of the intrinsic-Lions derivative in the $L^k$-Wasserstein space ($k \geq 1$) for singular McKean-Vlasov SDEs. In the following we first introduce the model considered in the paper, then recall the intrinsic and Lions (i.e. $L$-) derivatives, and finally state the main results of the paper.

1.1 McKean-Vlasov SDE

Let $\mathcal{P}$ be the space of probability measures on $\mathbb{R}^d$. We will use $| \cdot |$ to denote the absolute value or the norm in the Euclidean space, and $\| \cdot \|$ the operator norm for linear operators or matrices. For any $k \in [1, \infty)$, the $L^k$-Wasserstein space

$\mathcal{P}_k := \{ \mu \in \mathcal{P} : \mu(| \cdot |^k) < \infty \}$

is a Polish space under the $L^k$-Wasserstein distance

$W_k(\mu, \nu) := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^k \pi(dx, dy) \right)^{\frac{1}{k}}$, \quad \mu, \nu \in \mathcal{P}_k,$

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

Throughout the paper, we fix $T \in (0, \infty)$ and consider the following McKean-Vlasov SDE on $\mathbb{R}^d$:

$$dX_t = b_t(X_t, L_{X_t})dt + \sigma_t(X_t)dW_t, \quad t \in [0, T],$$

(1.1)

where $W_t$ is an $m$-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$, $L_\xi$ is the distribution (i.e. the law) of a random variable $\xi$, and

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

are measurable. When different probability measures are considered, we denote $L_\xi = L_{\xi|\mathbb{P}}$ to emphasize the distribution of $\xi$ under $\mathbb{P}$.

**Definition 1.1.** (1) A continuous adapted process $(X_t)_{t \in [0, T]}$ is called a solution of (1.1), if with $\mathcal{L}_X \in C([0, T]; \mathcal{P}_k)$,

$$\int_0^T \mathbb{E}\left[ |b_r(X_r, L_{X_r})| + \|\sigma_r(X_r)\|^2 \right] dr < \infty$$

and $\mathbb{P}$-a.s.

$$X_t = X_0 + \int_0^t b_r(X_r, L_{X_r})dr + \int_0^t \sigma_r(X_r)dW_r, \quad t \in [0, T].$$

We call (1.1) strongly well-posed for distributions in $\mathcal{P}_k$, if it has a unique solution for any initial value $X_0 \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P})$.

(2) A couple $(\tilde{X}_t, \tilde{W}_t)_{t \in [0, T]}$ is called a weak solution of (1.1), if $\tilde{W}_t$ is an $m$-dimensional Brownian motion on a complete filtration probability space $(\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0, T]}, \tilde{\mathbb{P}})$ such that $(\tilde{X}_t)_{t \in [0, T]}$ is a
solution of (1.1) for \((\tilde{W}_t, \tilde{P})\) replacing \((W_t, P)\). (1.1) is called weakly well-posed for distributions in \(\mathcal{P}_k\), if for any \(\nu \in \mathcal{P}_k\) it has a weak solution with initial distribution \(\nu\), and for any two weak solutions \((X^i_t, W^i_t)\) under \(P^i\), \(i = 1, 2\), \(\mathcal{L}_{X^i_t}^1 = \mathcal{L}^X_{x^2} = \mathcal{L}^X_{x^2} = \mathcal{L}^X_{x^2} = \mathcal{L}^X_{x^2}\).

(3) We call (1.1) well-posed for distributions in \(\mathcal{P}_k\), if it is both strongly and weakly well-posed for distributions in \(\mathcal{P}_k\).

When the SDE (1.1) is well-posed for distributions in \(\mathcal{P}_k\), we denote \(P_t^* \mu = \mathcal{L}_{X_t}\) for the solution with initial distribution \(\mu \in \mathcal{P}_k\), and define the family of linear operators \(\{P_t\}_{t \in [0, T]}\) from \(\mathcal{B}_b(\mathbb{R}^d)\) to \(\mathcal{B}_b(\mathcal{P}_k)\), where \(\mathcal{B}_b(\cdot)\) stands for the set of all bounded measurable functions on a measurable space:

\[
P_t f(\mu) := (P_t^* \mu)(f) = \int_{\mathbb{R}^d} f d(P_t^* \mu), \quad t \in [0, T], f \in \mathcal{B}_b(\mathbb{R}^d), \mu \in \mathcal{P}_k.
\]

To characterize the singularity of coefficients \(b\) and \(\sigma\) in time-space variables, we recall some functional spaces introduced in [21]. For any \(p \geq 1\), \(L^p(\mathbb{R}^d)\) is the class of measurable functions \(f\) on \(\mathbb{R}^d\) such that

\[
\|f\|_{L^p(\mathbb{R}^d)} := \left( \int_{\mathbb{R}^d} |f(x)|^p dx \right)^{\frac{1}{p}} < \infty.
\]

For any \(p, q > 1\), let \(\tilde{L}^p_q\) denote the class of measurable functions \(f\) on \([0, T] \times \mathbb{R}^d\) such that

\[
\|f\|_{\tilde{L}^p_q} := \sup_{z \in \mathbb{R}^d} \left( \int_0^T \|B(z, t) f\|_{L^p(\mathbb{R}^d)}^q dt \right)^{\frac{1}{q}} < \infty,
\]

where \(B(z, 1) := \{x \in \mathbb{R}^d : |x - z| \leq 1\}\). We denote \(f \in \tilde{H}_q^{2, p}\) if \(|f| + |\nabla f| + |\nabla^2 f| \in \tilde{L}_q^p\). We will take \((p, q)\) from the class

\[
\mathcal{K} := \{(p, q) : p, q \in (2, \infty), \frac{d}{p} + \frac{2}{q} < 1\}.
\]

### 1.2 Intrinsic/L-derivative in \(\mathcal{P}_k\)

The intrinsic derivative for measures was introduced in [1] to construct diffusion processes on configuration spaces over Riemannian manifolds, and used in [16] to study the geometry of dissipative evolution equations, see [2] for analysis and geometry on the Wasserstein space over a metric measure space. Moreover, the stronger notion \(L\)-derivative (i.e. Lions derivative) was introduced in [7] for functions on \(\mathcal{P}_k\) to investigate mean field games. The following general notion of intrinsic/L-derivative was introduced in [6] for functions on the \(L^k\)-Wasserstein space over a Banach space.

For any \(\mu \in \mathcal{P}_k\), the tangent space at \(\mu\) is

\[
T_{\mu, k} := L^k(\mathbb{R}^d \rightarrow \mathbb{R}^d; \mu).
\]

Then \(\mu \circ (id + \phi)^{-1} \in \mathcal{P}_k\) for \(\phi \in T_{\mu, k}\), where \(id\) is the identity map on \(\mathbb{R}^d\).
Definition 1.2. Let $f$ be a continuous function on $\mathcal{P}_k$. It is called intrinsically differentiable at a point $\mu \in \mathcal{P}_k$, if

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

is a well defined bounded linear functional. In this case, the intrinsic derivative is the unique element

$$D^I f(\mu) \in T^*_{\mu,k} := L_{k}^*(\mathbb{R}^d \rightarrow \mathbb{R}^d; \mu), \ k^* := \frac{k}{k-1} (= \infty \text{ if } k = 1)$$

such that

$$\int_{\mathbb{R}^d} \langle D^I f(\mu)(x), \phi(x) \rangle \mu(dx) = D^I \phi f(\mu), \ \phi \in T_{\mu,k}.$$ 

If moreover

$$\lim_{\|\phi\|_{T_{\mu,k}} \downarrow 0} \frac{\|f(\mu \circ (id + \phi)^{-1}) - f(\mu) - D^I_{\phi} f(\mu)\|}{\|\phi\|_{T_{\mu,k}}} = 0,$$

then $f$ is called $L$-differentiable at $\mu$ with the $L$-derivative $D^L f(\mu) := D^I f(\mu)$.

$f$ is called intrinsically (or $L$-) differentiable on $\mathcal{P}_k$, if it is intrinsically (or $L$-) differentiable at any $\mu \in \mathcal{P}_k$.

We will decompose the drift as $b_t(x, \mu) = b_t^{(0)}(x) + b_t^{(1)}(x, \mu)$, where $|b^{(0)}| \in \tilde{L}^{p_0}_{q_0}$ for some $(p_0, q_0) \in \mathcal{K}$ and $b_t^{(1)}$ is in the class $\mathcal{D}_k$ defined as follows.

Definition 1.3. $\mathcal{D}_k$ is the class of continuous functions $g$ on $\mathbb{R}^d \times \mathcal{P}_k$ such that $g(x, \mu)$ is differentiable in $x$, $L$-differentiable in $\mu$, and $D^L g(x, \mu)(y)$ has a version jointly continuous in $(x, y, \mu) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_k$ such that

$$|D^L g(x, \mu)(y)| \leq c(x, \mu)(1 + |y|^{k-1}), \ x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_k$$

holds for some positive function $c$ on $\mathbb{R}^d \times \mathcal{P}_k$.

Typical examples of functions in $\mathcal{D}_k$ are cylindrical functions of type

$$g(x, \mu) := F(x, \mu(h_1), \ldots, \mu(h_n))$$

for some $F \in C^1(\mathbb{R}^d \times \mathbb{R}^n)$ and $\{h_i\}_{1 \leq i \leq n} \subset C^1(\mathbb{R}^d)$ such that

$$\sup_{1 \leq i \leq n} |\nabla h_i(y)| \leq c(1 + |y|^{k-1}), \ y \in \mathbb{R}^d$$

holds for some constant $c > 0$. In this case

$$D^L g(x, \mu)(y) = \sum_{i=1}^{n} \{\partial_i F(x, \cdot)\}(\mu(h_1), \ldots, \mu(h_n)) \nabla h_i(y), \ x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_k.$$
1.3 Main results

To establish Bismut formula of $P_t f$ on $\mathcal{P}_k$, we make the following assumption.

(H) $b_t(x, \mu) = b_t^{(0)}(x) + b_t^{(1)}(x, \mu)$ such that the following conditions hold.

1. $a := \sigma \sigma^*$ is invertible with $\|a\|_\infty + \|a^{-1}\|_\infty < \infty$, where $\sigma^*$ is the transposition of $\sigma$, and

$$\lim_{\varepsilon \to 0} \sup_{|x-y| \leq \varepsilon, t \in [0,T]} \|a_t(x) - a_t(y)\| = 0.$$  

2. $|b^{(0)}| \in \hat{L}^{p_0}_{q_0}$ for some $(p_0, q_0) \in \mathcal{K}$. Moreover, $\sigma_t$ is weakly differentiable such that

$$\|\nabla \sigma\| \leq \sum_{i=1}^{l} f_i$$

holds for some $l \in \mathbb{N}$ and $0 \leq f_i \in \hat{L}^{p_i}_{q_i}$ with $(p_i, q_i) \in \mathcal{K}, 1 \leq i \leq l$.

3. $b_t^{(1)} \in \mathcal{D}_k$ such that

$$\sup_{(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_k} \left\{ |b_t^{(1)}(0, \delta_0)| + \|\nabla b_t^{(1)}(x, \mu)\| + \|D^{L} b_t^{(1)}(x, \mu)\|_{L^k(\mu)} \right\} < \infty,$$

where $\delta_0$ is the Dirac measure at $0 \in \mathbb{R}^d$, $\nabla$ is the gradient in the space variable $x \in \mathbb{R}^d$, and $D^L$ is the $L$-derivative in the distribution variable $\mu \in \mathcal{P}_k$.

We will show that (H) implies the well-posedness of (1.1) for distributions in $\mathcal{P}_k$. To calculate the intrinsic derivative $D^f P_t f(\mu)$, for any $\varepsilon \in [0, 1]$ and $\phi \in T_{\mu,k}$, we consider the following SDE:

$$dX_t^{\mu,\varepsilon,\phi} = b_t(X_t^{\mu,\varepsilon,\phi}, \mathcal{L}X_t^{\mu,\varepsilon,\phi})dt + \sigma_t(X_t^{\mu,\varepsilon,\phi})dW_t,$$

$$t \in [0,T], \quad X_0^{\mu,\varepsilon,\phi} = X_0^{\mu} + \varepsilon \phi(X_0^{\mu}).$$

We will prove that the derivative process

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

exists in $L^k(\Omega \to C([0,T]; \mathbb{R}^d), \mathbb{P})$. We also need the derivative of the decoupled SDE

$$dX_t^{\mu,x} = b_t(X_t^{\mu,x}, P_t^{\ast} \mu)dt + \sigma_t(X_t^{\mu,x})dW_t,$$

$$t \in [0,T], X_0^{\mu,x} = x, x \in \mathbb{R}^d, \mu \in \mathcal{P}_k.$$

By Theorem 2.1 (H) implies the well-posedness of (1.6) and that for any $v \in \mathbb{R}^d$,

$$\nabla_v X_t^{\mu,x} := \lim_{\varepsilon \downarrow 0} \frac{X_t^{\mu,x+\varepsilon v(x)} - X_t^{\mu,x}}{\varepsilon}, \quad t \in [0,T]$$

exists in $L^k(\Omega \to C([0,T]; \mathbb{R}^d), \mathbb{P})$.

Our first result present some continuity estimates in terms of the initial data.
Theorem 1.1. Assume (H). Then the following assertions hold.

1. \( (1.1) \) is well-posed for distributions in \( \mathcal{P}_k \), and for any \( j \geq 1 \) there exists a constant \( c > 0 \) such that any solution \( X_t \) satisfies

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |X_t|^j \right] \leq c \left\{ 1 + (\mathbb{E}[|X_0|^k])^\frac{j}{k} + |X_0|^j \right\}.
\]

In particular, there exists a constant \( c > 0 \) such that

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |X_t|^k \right] \leq c(1 + \mathbb{E}[|X_0|^k]).
\]

2. For any \( j \geq 1 \) there exists a constant \( c > 0 \) such that for any two solutions \( X^1_t, X^2_t \) of \( (1.1) \) with initial distributions in \( \mathcal{P}_k \),

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |X^1_t - X^2_t|^j \right] \leq c \left\{ (\mathbb{E}[|X^1_0 - X^2_0|^k])^\frac{j}{k} + |X^1_0 - X^2_0|^j \right\}.
\]

In particular, there exists a constant \( c > 0 \) such that

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |X^1_t - X^2_t|^k \right] \leq c\mathbb{E}[|X^1_0 - X^2_0|^k].
\]

3. There exists a constant \( c > 0 \) such that

\[
\|P_t^\mu \mu - P_t^\nu \nu\|_{\text{var}} := \sup_{|f| \leq 1} |P_t f(\mu) - P_t f(\nu)| \leq \frac{c}{\sqrt{t}} \mathbb{W}_k(\mu, \nu), \quad t \in (0,T], \mu, \nu \in \mathcal{P}_k.
\]

By taking \( X^1_0 \) and \( X^2_0 \) such that

\[
\mathcal{L}X^1_0 = \mu, \quad \mathcal{L}X^2_0 = \nu, \quad \mathbb{E}[|X^1_0 - X^2_0|^k] = \mathbb{W}_k(\mu, \nu)^k,
\]
we deduce from \( (1.11) \) that

\[
\mathbb{W}_k(P_t^\mu \mu, P_t^\nu \nu) \leq c\mathbb{W}_k(\mu, \nu), \quad t \in [0,T], \mu, \nu \in \mathcal{P}_k
\]
holds for some constant \( c > 0 \). Our next result provides derivative estimates and Bismut formula of \( D_t^i P_t f \) for \( t \in (0,T] \) and \( f \in \mathcal{B}_b(\mathbb{R}^d) \).

Theorem 1.2. Assume (H). Then the following assertions hold.

1. For any \( \mu \in \mathcal{P}_k, \phi \in T_{\mu,k} \) and \( v, x \in \mathbb{R}^d \), \( \nabla_\phi X^\mu_t \) and \( \nabla_\mu X^{\mu,x}_t \) exist in \( L^k(\Omega \to C([0,T];\mathbb{R}^d), \mathbb{P}) \). Moreover, for any \( j \geq 1 \) there exists a constant \( c > 0 \) such that

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |\nabla_\phi X^\mu_t|^j \right] \leq c \left\{ \|\phi\|_{L^k(\mu)}^j + |\phi(X^\mu_0)|^j \right\}, \quad \mu \in \mathcal{P}_k, \phi \in T_{\mu,k},
\]

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} |\nabla_\mu X^{\mu,x}_t|^j \right] \leq c|v|^j, \quad \mu \in \mathcal{P}_k, x, v \in \mathbb{R}^d.
\]
Theorem 1.3. Assume

Let $b$ continuous and $X$ $P$-differentiable on $\mathcal{P}_k$. Moreover, for any $\phi \in T_{\mu,k}$ and $\beta \in C^1([0, t])$ with $\beta_0 = 0$ and $\beta_t = 1$,

$$D_k^2 Pf(\mu) = \int_{\mathbb{R}^d} \mathbb{E} \left[ f(X^\mu_t) \int_0^t \beta_s \left( \nabla_x (X^\mu_s) \nabla_{\phi(x)} X^\mu_s \right) \right] \mu(dx)$$

(1.16)

$$+ \mathbb{E} \left[ f(X^\mu_t) \int_0^t \left( \nabla_x (X^\mu_s) \mathbb{E} \left[ \left( D^T b^{(1)}_s(z, P^\mu_s \mu)(X^\mu_s, \nabla X^\mu_s) \right) \right] \right] \right].$$

Consequently, for any $p > 1$ there exists a constant $c > 0$ such that

$$\|D^T Pf(\mu)\|_{L^k(\mathcal{P})} \leq \frac{c}{\sqrt{t}} \|f(X^\mu_t)\|_{L^k(\mathcal{P})}, \quad t \in (0, T], f \in \mathcal{B}_b(\mathbb{R}^d), \mu \in \mathcal{P}_k.$$

(1.17)

In particular, there exists a constant $c > 0$ such that

$$\|D^T Pf(\mu)\|_{L^k(\mathcal{P})} \leq \frac{c}{\sqrt{t}} \|f(X^\mu_t)\|_{L^k(\mathcal{P})}, \quad t \in (0, T], f \in \mathcal{B}_b(\mathbb{R}^d), \mu \in \mathcal{P}_k.$$

Finally, to prove the $L$-differentiability of $P_t f$, we need the uniform continuity of $\sigma_t(x)$, $\nabla b^{(1)}_t(x, \mu)$ and $D^L b_t(x, \mu)(y)$ in $(x, y, \mu)$:

$$\lim_{\epsilon \downarrow 0} \sup_{t \in [0, T]} \sup_{|x - x'| \leq \epsilon} \{ \| \sigma_t(x) - \sigma_t(x') \| + \| \nabla b^{(1)}_t(x, \mu) - \nabla b^{(1)}_t(x', \nu) \| \} = 0,$$

(1.18)

$$\lim_{\epsilon \downarrow 0} \sup_{t \in [0, T]} \sup_{|x - x'| \leq \epsilon} \| D^L b^{(1)}_t(x, \mu)(y) - D^L b^{(1)}_t(x', \nu)(y') \| = 0.$$

Under this condition and $(H)$, the following result ensures the $L$-differentiability of $P_t f$ in $\mathcal{P}_k$ for $k > 1$, which improves the corresponding result in [13] where $k = 2$, $\sigma_t$ is Lipschitz continuous and $b^{(1)}_t(0)$ is Dini continuous.

**Theorem 1.3.** Assume $(H)$ and (1.18) for $k \in (1, \infty)$. Then for any $t \in (0, T]$ and $f \in \mathcal{B}_b(\mathbb{R}^d)$, $P_t f$ is $L$-differentiable on $\mathcal{P}_k$.

In Section 2, we establish the Bismut formula for singular SDEs where $b_t(x, \mu) = b_t(x)$ is independent of $\mu$. Then we prove the above theorems in Sections 3-5 respectively.

## 2 Bismut formula for singular SDEs

Let $b_t(x, \mu) = b_t(x)$ do not depend on $\mu$, so that (1.1) becomes

$$dX_t = b_t(X_t) dt + \sigma_t(X_t) dW_t, \quad t \in [0, T].$$

Let $X^\mu_t$ solves (2.1) for $X^\mu_0 = x$, and consider

$$P_t f(x) := \mathbb{E}[f(X^\mu_t)], \quad t \geq 0, f \in \mathcal{B}_b(\mathbb{R}^d), x \in \mathbb{R}^d.$$

Under $(H)$ with $l = 1$ and $b^{(1)} = 0$, the following Bismut formula (2.3) is included in Theorem 1.1(iii) of [21] for $f \in C^1_b(\mathbb{R}^d)$. It is reasonable but nontrivial to extend the formula from $f \in C^1_b(\mathbb{R}^d)$ to $f \in \mathcal{B}_b(\mathbb{R}^d)$. The technique we used in step (d) in the proof of Theorem 2.1(2) is due to [15], which will also be used in the proofs of Theorems 1.2 and 1.3.
Theorem 2.1. Let \((H)\) hold for \(b_t^{(1)}(x,\mu) = b_t^{(1)}(x)\). Then \((2.1)\) is well-posed and the following assertions hold.

1. For any \(j \geq 1\) and \(x, v \in \mathbb{R}^d\),
   
   \[
   \nabla_v X_t^x := \lim_{\varepsilon \to 0} \frac{X_t^{x+\varepsilon v} - X_t^x}{\varepsilon}, \quad t \in [0, T]
   \]

   exists in \(L^j(\Omega \to C([0, T]; \mathbb{R}^d), \mathbb{P})\), and there exists a constant \(c(j) > 0\) such that
   
   \[
   \sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0, T]} |\nabla_v X_t^x|^j \right) \leq c(j) |v|^j, \quad v \in \mathbb{R}^d.
   \]

2. For any \(t \in (0, T)\) and \(\beta \in C^1([0, t])\) with \(\beta_0 = 0\) and \(\beta_t = 1\),
   
   \[
   \nabla_v P_t f(x) = \mathbb{E} \left[ f(X_t^x) \int_0^t \beta_s' \left\{ \sigma_s^*(\sigma_s^*)^{-1} \right\} (X_s^x) \nabla_v X_s^x, dW_s \right]
   \]

   holds for any \(x \in \mathbb{R}^d\) and \(f \in \mathcal{B}_b(\mathbb{R}^d)\). Consequently, for any \(p > 1\) there exists a constant \(c(p) > 0\) such that
   
   \[
   |\nabla P_t f| \leq \frac{c(p)}{\sqrt{t}} (P_t |f|^p(x))^{\frac{1}{p}}, \quad t \in (0, T), f \in \mathcal{B}_b(\mathbb{R}^d).
   \]

Proof. The well-posedness follows from Lemma 3.1 in [20]. Below we prove assertions (1) and (2) by using Zvonkin’s transform as in [21].

(a) Zvonkin’s transform. Let \(\nabla_v\) be the directional derivative along \(v\). Consider

   \[
   L_t = \frac{1}{2} \text{tr}\{\sigma_t \sigma_t^* \nabla^2\} + \nabla b_t, \quad t \in [0, T].
   \]

By [23] Theorem 2.1, there exists \(\lambda_0 > 0\), such that when \(\lambda \geq \lambda_0\), the PDE for \(u : [0, T] \to \mathbb{R}^d\):

\[
(\partial_t + L_t)u_t = \lambda u_t - b_t^{(0)}, \quad t \in [0, T], u_T = 0
\]

has a unique solution such that

\[
f_0 := \|\nabla^2 u\| + |(\partial_t + \nabla b^{(0)}) u| \in \tilde{L}^p_{\omega_0}, \quad \|u\|_\infty + \|\nabla u\|_\infty \leq \frac{1}{2}.
\]

Let \(\Theta_t := id + u_t\) and

   \[
   \tilde{b}_t := \{\lambda u_t + b_t^{(1)}\} \circ \Theta_t^{-1}, \quad \tilde{\sigma}_t := \{(\nabla \Theta_t) \sigma_t\} \circ \Theta_t^{-1}, \quad t \in [0, T].
   \]

By (2.5), and Itô’s formula [23] Lemma 3.3, \(Y_t^x := \Theta_t(X_t^x)\) solves

\[
dY_t^x = \tilde{b}_t(Y_t^x)dt + \tilde{\sigma}_t(Y_t^x)dW_t, \quad Y_0^x = \Theta_0(x).
\]
Moreover, by (H), we find a constant $\kappa > 0$ such that

$$
\|\nabla \tilde{b}\|_{\infty} + \|\tilde{\sigma}\|_{\infty} + \| (\tilde{\sigma} \tilde{\sigma}^*)^{-1} \|_{\infty} < \infty, \quad \|\nabla \tilde{\sigma}\| \leq \kappa \sum_{i=0}^{\ell} f_i.
$$

By \cite[Theorem 3.1]{23}, for any $(p, q) \in \mathcal{K}$, there exists a constant $c > 0$ such that

$$
\mathbb{E} \left( \int_s^t f_r(X_r) dr \bigg| \mathcal{F}_s \right) \leq c \|f\|_{L_{q/2}^p(s, t)}, \quad f \in \tilde{L}_{q/2}^p(s, t), \quad 0 \leq s \leq t \leq T,
$$

where $\tilde{L}_{q}^p(s, t)$ is defined as $\tilde{L}_{q}^p$ for $[s, t]$ replacing $[0, T]$. This implies Khasminskii’s estimate (see for instance \cite[Lemma 4.1]{21}): there exists an increasing map $\Psi : (0, \infty) \times (0, \infty) \to (0, \infty)$ such that for any solution $X_t$ of (2.1),

$$
\mathbb{E} \left[ e^{\lambda \int_0^T |f_t|^p(X_t) dt} \right] \leq \Psi(\lambda, \|f\|_{\tilde{L}_{q}^p}), \quad \lambda > 0, \quad f \in \tilde{L}_{q}^p, \quad (p, q) \in \mathcal{K}.
$$

Combining this with (4.12) and $f_i \in \tilde{L}_{q}^p(T)$ for $(p_i, q_i) \in \mathcal{K}$, we obtain

$$
\mathbb{E} \left[ e^{N \int_0^T (\|\nabla \tilde{b}\|_{\infty} + \|\nabla \tilde{\sigma}^t\|_{2}) dt} \right] < \infty, \quad N \geq 1.
$$

So, for any $v, x \in \mathbb{R}^d$, the linear SDE

$$
dv_t = (\nabla_{\tilde{v}} \tilde{b})(Y_t^x) + (\nabla_{\tilde{v}} \tilde{\sigma})(Y_t^x) dW_t, \quad v_0 = v + \nabla_{\tilde{v}} u_0(x)
$$

has a unique solution, and by Itô’s formula and the stochastic Gronwall inequality Lemma 3.7 in \cite[22]{22}, for any $j \geq 1$ there exists a constant $c(j) > 0$ such that

$$
\sup_{x \in \mathbb{R}^d} \mathbb{E} \left[ \sup_{t \in [0, T]} |v_t|^j \right] \leq c(j) |v|^j, \quad j \geq 1.
$$

(b) Proof of assertion (1). Let $Y_t^{x+\varepsilon v} := \Theta_t(X_t^{x+\varepsilon v})$. By (2.11), for the first assertion it suffices to prove

$$
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \frac{Y_t^{x+\varepsilon v} - Y_t^x}{\varepsilon} - v_t^x \right|^j \right] = 0, \quad j \geq 1.
$$

Indeed, by an approximation argument indicated in Remark 2.1 below, see also Remark 2.1 in \cite[23]{23}, we may assume that $\nabla^2 \tilde{b}^{(1)}$ is bounded so that by Lemma 2.3(3) in \cite[23]{23},

$$
|\nabla \Theta_t(x) - \nabla \Theta_t(y)| \leq c|x - y|^{\alpha}, \quad t \in [0, T], \quad x, y \in \mathbb{R}^d
$$

holds for some constants $c > 0$ and $\alpha \in (0, 1)$. Combining this with (2.6) and (2.12), we see that $\nabla_{\tilde{v}} Y_t^x$ exists in $L^j(\Omega \to C([0, T]; \mathbb{R}^d), \mathbb{P})$ with

$$
\nabla_{\tilde{v}} Y_t^x = (\nabla \Theta_t(X_t^x)^{-1}) \nabla_{\tilde{v}} Y_t^x = (\nabla \Theta_t(X_t^x)^{-1}) v_t^x, \quad t \in [0, T].
$$
To prove (2.12), let
\[ v^\varepsilon_r := \frac{Y^x_{r+\varepsilon} - Y^x_r}{\varepsilon}, \quad s \in [0,T], \varepsilon \in (0,1). \]

By (2.8), (2.9), [21] Lemma 2.1, and the stochastic Gronwall inequality [22] Lemma 3.7, as in the proof of [21] (4.8) we have

\[ \sup_{\varepsilon \in (0,1)} \mathbb{E} \left[ \sup_{t \in [0,T]} |\dot{v}^\varepsilon_t| \right] < \infty, \quad j \geq 1. \]  

Write

\[ v^\varepsilon_r = \int_0^r (\nabla v^s \tilde{b}_s)(Y^x_s)ds + \int_0^r (\nabla v^s \tilde{\sigma}_s)(Y^x_s)dW_s + \alpha^\varepsilon_r, \quad r \in [0,t], \]

where

\[ \alpha^\varepsilon_r := \int_0^r \xi^\varepsilon_s ds + \int_0^t \eta^\varepsilon_s dW_s \]

for

\[ \xi^\varepsilon_s := \frac{\tilde{b}_s(Y^x_{s+\varepsilon} - Y^x_s)}{\varepsilon} - (\nabla v^s \tilde{b}_s)(Y^x_s), \]

\[ \eta^\varepsilon_s := \frac{\tilde{\sigma}_s(Y^x_{s+\varepsilon} - Y^x_s)}{\varepsilon} - (\nabla v^s \tilde{\sigma}_s)(Y^x_s). \]

We aim to prove

\[ \lim_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{t \in [0,T]} |\alpha^\varepsilon_t|^n \right] = 0, \quad n \geq 1. \]

Firstly, since \( \nabla \tilde{b}_s \) and \( \nabla \tilde{\sigma}_s \) exists a.e., for a.e. \( x \in \mathbb{R}^d \) we have

\[ \limsup_{\varepsilon \downarrow 0} \sup_{|z| \leq 1} \left\{ \left| \frac{\tilde{b}_s(x+\varepsilon z) - \tilde{b}_s(x)}{\varepsilon} - \nabla \tilde{b}_s(x) \right| + \left| \frac{\tilde{\sigma}_s(x+\varepsilon z) - \tilde{\sigma}_s(x)}{\varepsilon} - \nabla \tilde{\sigma}_s(x) \right| \right\} = 0. \]

Combining this with (2.14) and noting that \( \mathcal{L}_{Y^x}(s \in (0,T]) \) is absolutely continuous with respect to the Lebesgue measure, see for instance Theorem 6.3.1 in [21], we obtain

\[ \lim_{\varepsilon \to 0} \left\{ |\xi^\varepsilon_s| + ||\eta^\varepsilon_s|| \right\} = 0, \quad \mathbb{P}\text{-a.s.}, \quad s \in (0,T]. \]

Next, let

\[ \mathcal{M} f(x) := \sup_{r \in (0,1)} \frac{1}{|B(0,r)|} \int_{B(0,r)} f(x+z)dz, \quad x \in \mathbb{R}^d, 0 \leq f \in L^1_{\text{loc}}(\mathbb{R}^d). \]

Let \( \theta > 1 \) such that \( (\theta^{-1} p_i, \theta^{-1} q_i) \in \mathcal{K}, 0 \leq i \leq l. \) By \( f_i \in \tilde{L}^p_{q_i}, \) Lemma 2.1 in [21] and (2.9) with \( f = f^{\theta}_i \) and \( (p,q) = (\theta^{-1} p_i, \theta^{-1} q_i), \) we obtain

\[ \sup_{\varepsilon \in [0,1]} \mathbb{E} \left[ \left( \int_0^T (\mathcal{M} f^{\theta}_i)(X^x_{t+\varepsilon t})dt \bigg| \mathcal{F}_0 \right)^n \right] \leq K_n, \quad 0 \leq i \leq l \]
for some constant $K_n > 0$. By $(H)$ and Lemma 2.1 in [21], there exists a constant $c_1 > 0$ such that

$$|\xi^\varepsilon_s|^{2\theta} + \|\eta^\varepsilon_s\|^{2\theta} \leq c_1 |\tilde{c}^\varepsilon_t|^2 \left(1 + \sum_{i=0}^l \left\{(\mathcal{M} f^{2\theta}_i(s, \cdot))(X^x_s) + (\mathcal{M} f^{2\theta}_i(s, \cdot))(X^{x+\varepsilon}_s)\right\}\right).$$

Combining this with (2.14) and (2.19), for any $n \geq 1$ we find constants $c_1(n), c_2(n) > 0$ such that

$$E \left[ \left( \int_0^T \{|\xi^\varepsilon_s|^{2\theta} + \|\eta^\varepsilon_s\|^{2\theta}\} ds \right) \right]^{n} \leq c_1(n) E \left[ \left( \sup_{s \in [0,T]} |v^\varepsilon_s|^{4n} \right) \left( \int_0^T \left\{1 + \sum_{i=0}^l \left(\mathcal{M} f^{2\theta}_i(s, \cdot))(X^x_s) + (\mathcal{M} f^{2\theta}_i(s, \cdot))(X^{x+\varepsilon}_s)\right) ds \right)^n \right]^{1/2} \times \left( E \left[ \left( \int_0^T \left\{1 + \sum_{i=0}^l \left(\mathcal{M} f^{2\theta}_i(s, \cdot))(X^x_s) + (\mathcal{M} f^{2\theta}_i(s, \cdot))(X^{x+\varepsilon}_s)\right) ds \right)^{2n} \right] \right)^{1/2} \leq c_2(n) < \infty, \ \varepsilon \in (0, 1].$$

Thus, by (2.18) and the dominated convergence theorem, we derive

$$\lim_{\varepsilon \to 0} E \left[ \left( \int_0^T \{|\xi^\varepsilon_s|^2 + \|\eta^\varepsilon_s\|^2\} ds \right) \right]^{n} = 0, \ n \geq 1.$$

Therefore, (2.16) and BDG’s inequality imply (2.17).

Finally, by (2.8), (2.10), (2.26), and Lemma 2.1 in [21], for any $j \geq 1$, we find a constant $c(j) > 0$ such that

$$d|v^\varepsilon_s - v^\xi_s|^{2j} \leq C(j) \left\{1 + \sum_{i=0}^l f^{2j}_i(s, Y^x_s)\right\}|v^\varepsilon_s - v^\xi_s|^{2j} ds + c(j) \sup_{r \in [0,s]} |\alpha^\varepsilon_r|^{2j} + dM_s, \ s \in [0, t]$$

holds for some local martingale $M_s$. Since $\lim_{\varepsilon \to 0} |v^\varepsilon_t - v^\xi_t| = 0$, by combining this with (2.9), (2.17), and the stochastic Gronwall inequality [22, Lemma 3.7], we prove (2.12).

(c) Proof of (2.3) for $f \in C_{\text{Lip}}(\mathbb{R}^d)$, the space of Lipschitz continuous functions on $\mathbb{R}^d$. Let $t \in (0, T]$ be fixed, and consider

$$h_s = \int_0^s \beta^r \left[\tilde{\sigma}_r \tilde{\sigma}_r^* \right]^{-1} (Y^x_r)v^0_r dr, \ s \in [0, t].$$

By the same reason leading to (2.11), the SDE

$$dw_s = \left\{\nabla w_s \tilde{b}_s(Y^x_s) + \tilde{\sigma}_s (Y^x_s) h^s_t\right\} ds + (\nabla w_s \tilde{\sigma}_s)(Y^x_s) dW_s,$$

$$w_0 = 0, s \in [0, t]$$

(2.21)
has a unique solution satisfying

\[(2.22) \sup_{x \in \mathbb{R}^d} \mathbb{E} \left[ \sup_{t \in [0,T]} |w_s|^j \right] < \infty, \quad j \geq 1.\]

We aim to prove that the Malliavin derivative \( D_h Y_t^x \) of \( Y_t^x \) along \( h \) exists and

\[(2.23) \quad D_h Y_t^x = w_t.\]

For any \( \varepsilon > 0 \), according to the proof of Lemma 3.1 in [20], (2.8), (2.9) and (2.20) imply the well-posedness of the SDE

\[(2.24) \quad dY_{x,\varepsilon}^s = \left\{ \tilde{b}_t(Y_{x,\varepsilon}^s) + \varepsilon \tilde{\sigma}_s(Y_{x,\varepsilon}^s) h_s' \right\} ds + \tilde{\sigma}_s(Y_{x,\varepsilon}^s) dW_s, \quad s \in [0,t], \quad Y_{x,\varepsilon}^0 = Y_0^x.\]

By (2.8), (2.20), Lemma 2.1 in [21] and Itô's formula, for any \( j \geq 1 \) we find a constant \( c_1(j) > 0 \) such that

\[d|Y_{x,\varepsilon}^s - Y_s^x|^2 \leq c_1(j)|Y_{x,\varepsilon}^s - Y_s^x|^{2j} \sum_{i=0}^t \left\{ 1 + M \{ f_i(s, \cdot) \}^2(Y_s^x) + M f_i(s, \cdot) \}^2(Y_s^x) \right\} ds \]

\[+ c_1(j)\varepsilon^{2j}|v_s|^{2j} ds + dM_s, \quad s \in [0,t] \]

holds for some local martingale \( M_s \). Noting that \( Y_{0,\varepsilon}^x - Y_0^x = 0 \), by combining this with the stochastic Gronwall inequality Lemma 3.7 in [22] and Lemma 2.1 in [21], we obtain

\[(2.25) \quad \sup_{\varepsilon \in (0,1]} \mathbb{E} \left[ \sup_{t \in [0,T]} \frac{|Y_{x,\varepsilon}^s - Y_s^x|}{\varepsilon^j} \right] < \infty, \quad j \geq 1.\]

Let \( w_{x,\varepsilon}^s = \frac{Y_{x,\varepsilon}^s - Y_s^x}{\varepsilon} \). Then

\[(2.26) \quad w_{x,\varepsilon}^r = \int_0^r \left\{ (\nabla w_s \tilde{b}_s)(Y_s^x) + \tilde{\sigma}_s(Y_s^x) h_s' \right\} ds \]

\[+ \int_0^r (\nabla w_s \tilde{\sigma}_s)(Y_s^x) dW_s + \tilde{\alpha}_r^\varepsilon, \quad r \in [0,t] \]

holds for

\[\tilde{\alpha}_r^\varepsilon := \int_0^r \left\{ \frac{\tilde{b}_s(Y_{x,\varepsilon}^s) - \tilde{b}_s(Y_s^x)}{\varepsilon} - (\nabla w_s \tilde{b}_s)(Y_s^x) \right\} ds \]

\[+ \int_0^r \left\{ \sigma_s(Y_{x,\varepsilon}^s) - \tilde{\sigma}_s(Y_s^x) \right\} h_s' ds \]

\[+ \int_0^r \left\{ \frac{\tilde{\sigma}_s(Y_{x,\varepsilon}^s) - \tilde{\sigma}_s(Y_s^x)}{\varepsilon} - (\nabla w_s \tilde{\sigma}_s)(Y_s^x) \right\} dW_s.\]

Combining this with (2.21) and using the same argument leading to (2.12), we prove (2.23).
By (2.20) and the SDE (2.10) for \( v_s \), we see that \( \beta_s v_s \) solves (2.21), so that by the uniqueness we obtain
\[
\nabla_v Y^x_t = v_t = w_t = D_h Y^x_t.
\]
For \( f \in C_{Lip}(\mathbb{R}^d) \), \( \nabla f \) exists a.e. and \( \|\nabla f\|_\infty < \infty \). Since \( \mathcal{L}_{X^x_t} \) is absolutely continuous, see for instance Theorem 6.3.1 in [5], we conclude that \( (\nabla f)(X^x_t) \) exists and is bounded. By the integration by parts formula in Malliavin calculus, see for instance [10], \( \nabla_v Y^x_t = D_h Y^x_t \) implies
\[
\begin{align*}
\nabla_v P_t f(x) &= \nabla_v \mathbb{E}[\{f \circ (\Theta_t)^{-1}\}(Y^x_t)] = \mathbb{E}[\langle \nabla (f \circ \Theta_t^{-1})(Y^x_t), \nabla_v Y^x_t \rangle] \\
&= \mathbb{E}[D_h \{(f \circ \Theta_t^{-1})(Y^x_t)\}] = \mathbb{E}[f(X^x_t) \int_0^t \langle h'_s, dW_s \rangle] \\
&= \mathbb{E}[f(X^x_t) \int_0^t \beta'_s \langle \{\bar{\sigma}_s^*(\sigma_s \bar{\sigma}_s^*)^{-1}\}(Y^x_s)v_s, dW_s \rangle], \quad f \in L_{Lip}(\mathbb{R}^d).
\end{align*}
\]
By \( v_t = \nabla Y^x_t \), \( Y^x_t = \Theta_t(X^x_t) \) and \( \bar{\sigma}_t = \{(\nabla \Theta_t) \sigma_t\} \circ \Theta_t^{-1} \), we obtain
\[
\begin{align*}
\{\bar{\sigma}_s^*(\sigma_s \bar{\sigma}_s^*)^{-1}\}(Y^x_s)v_s &= \left[ \sigma_s^*(\sigma_s \sigma_s^*)^{-1} \{ (\nabla \Theta_s) \sigma_s \sigma_s^*(\nabla \Theta_s)^* \}^{-1} \right] (X^x_s) \{ \nabla \Theta_s (X^x_s) \} \nabla X^x_s \\
&= \{ \sigma_s^*(\sigma_s \sigma_s^*)^{-1} \} (X^x_s) \nabla X^x_s, \quad s \in [0, T],
\end{align*}
\]
so that this implies
\[
(2.27) \quad \nabla_v P_t f(x) = \mathbb{E}\left[ f(X^x_t) \int_0^t \beta'_s \langle \{\sigma_s^*(\sigma_s \sigma_s^*)^{-1}\}(X^x_s) \nabla X^x_s, dW_s \rangle \right], \quad f \in C_{Lip}(\mathbb{R}^d).
\]

(d) Proof of (2.4) and (2.3). Let \( P^*_t \delta_x = \mathcal{L}_{X^x_t} \) and let \( \nu_\varepsilon \) be the finite signed measure defined by
\[
\nu_\varepsilon(A) := \int_0^\varepsilon \mathbb{E}\left[ 1_A(X^x_t+rv) \int_0^t \beta'_s \langle \{\sigma_s^*(\sigma_s \sigma_s^*)^{-1}\}(X^x_s+rv) \nabla X^x_s+rv, dW_s \rangle \right] dr
\]
for \( A \in \mathcal{B}(\mathbb{R}^d) \), the Borel \( \sigma \)-algebra on \( \mathbb{R}^d \). Then (2.27) implies
\[
(P^*_t \delta_{x+\varepsilon v} - P^*_t \delta_x)(f) = \nu_\varepsilon(f), \quad f \in C_{Lip}(\mathbb{R}^d),
\]
where \( \nu(f) := \int fd\nu \) for a (signed) measure \( \nu \) and \( f \in L^1(|\nu|) \). Since \( C_{Lip}(\mathbb{R}^d) \) determines measures, we obtain
\[
P^*_t \delta_{x+\varepsilon v} - P^*_t \delta_x = \nu_\varepsilon,
\]
so that for any \( f \in \mathcal{B}_b(\mathbb{R}^d) \),
\[
P_t f(x + \varepsilon v) - P_t f(x) = \int_0^\varepsilon \mathbb{E}\left[ f(X^x_t+rv) \int_0^t \beta'_s \langle \{\sigma_s^*(\sigma_s \sigma_s^*)^{-1}\}(X^x_s+rv) \nabla X^x_s+rv, dW_s \rangle \right] dr.
\]
Combining this with (2.22) and the boundedness of \( \sigma^*(\sigma \sigma^*)^{-1} \), we prove (2.4).

Next, let \( f \in \mathcal{B}_b(\mathbb{R}^d) \). For any \( r \in (0, T) \), let \( (X^x_{r,t})_{t \in [r,T]} \) solve (2.1) from time \( r \) with \( X^x_{r,r} = x \). Let
\[
P_{r,t} f(x) := \mathbb{E}[f(X^x_{r,t})], \quad f \in \mathcal{B}_b(\mathbb{R}^d), x \in \mathbb{R}^d.
\]
Then the well-posedness implies
\[ P_t = P_{r,t}, \quad 0 < r < t \leq T. \]
Moreover, considering the SDE from time \( r \) replacing 0, (2.4) implies
\[ \| \nabla P_{r,t} f \|_\infty < \infty, \quad f \in \mathcal{B}_b(\mathbb{R}^d), 0 < r < t \leq T. \]
So, by (2.27) for \( (P_{r}, \beta_s/\beta_r) \) replacing \( (P_t, \beta_s) \), we obtain
\[ \nabla v P_t f(x) = \frac{1}{\beta_r} \mathbb{E} \left[ P_{r,t} f(X^x_r) \int_0^t \beta_s' \left\{ \sigma^*_s(\sigma_s)^{-1} \right\} (X^x_s) \nabla X^x_s \, dW_s \right] \]
for all \( f \in \mathcal{B}_b(\mathbb{R}^d) \) and \( r \in (0, t) \) such that \( \beta_r > 0 \). Since the Markov property implies
\[ \mathbb{E}[f(X^x_t)|\mathcal{F}_r] = P_{r,t} f(X^x_r), \]
we obtain
\[ \nabla v P_t f(x) = \frac{1}{\beta_r} \mathbb{E} \left[ f(X^x_t) \int_0^t \beta_s' \left\{ \sigma^*_s(\sigma_s)^{-1} \right\} (X^x_s) \nabla X^x_s \, dW_s \right], \]
so that letting \( r \uparrow t \) gives (2.3).

To conclude this section, we make the following remark which enables us to apply Theorem 2.1 to the decoupled SDE (1.6) with estimates uniformly in \( \mu \).

**Remark 2.1.** For fixed \( \sigma \) but may be variable \( b \), the constants \( c(\cdot) \) in Theorem 2.1 are uniformly in \( b = b^{(0)} + b^{(1)} \) satisfying
\[(2.28) \quad \| b^{(0)} \|_{\tilde{L}^{p_0}_t} + \| \nabla b^{(1)} \|_\infty \leq N \]
for a given constant \( N > 0 \). Indeed, letting \( \gamma \) be the standard Gaussian measure and take
\[ \tilde{b}^{(1)}_t(x) := \int_{\mathbb{R}^d} b^{(1)}_t(x + y) \gamma(dy), \quad x \in \mathbb{R}^d, t \in [0, T], \]
we find constant \( c > 0 \) only depending on \( N \) such that (2.28) implies
\[ \| \nabla \tilde{b}^{(1)} \|_\infty + \| \nabla^2 \tilde{b}^{(1)} \|_\infty + \| b^{(1)}_t - \tilde{b}^{(1)}_t \|_\infty \leq c. \]
Then \( \tilde{b}^{(0)} := b^{(0)} + \tilde{b}^{(1)} - b^{(1)} \) satisfies
\[ \| \tilde{b}^{(0)} \|_{\tilde{L}^{p_0}_t} \leq \| b^{(0)} \|_{\tilde{L}^{p_0}_t} + c \| 1 \|_{\tilde{L}^{p_0}_t} =: c'. \]
According to the proofs of [23, Theorem 2.1 and Theorem 3.1] for \( b = \tilde{b}^{(0)} + \tilde{b}^{(1)} \), the constant \( \lambda_0 > 0 \) before (2.5), \( \| \nabla^2 u \|_{\tilde{L}^{p_0}_t} \), and the constant in Krylov’s estimate are uniformly in \( b \) satisfying (2.28). According to the proof of [21] Lemma 4.1, the same is true for Khasminskii’s estimate (2.9). Therefore, in the proof of Theorem 2.1 constants \( c(\cdot) \) can be taken uniformly in \( b \) satisfying (2.28).
3 Proof of Theorem 1.1

We first present a Lipschitz estimate for $L$-differentiable functions on $P_k$.

**Lemma 3.1.** Let $f$ be $L$-differentiable on $P_k$ such that for any $\mu \in P_k$, $D^L f(\mu)(\cdot)$ has a continuous version satisfying

\[
|D^L f(\mu)(x)| \leq c(\mu)(1 + |x|^{k-1}), \quad x \in \mathbb{R}^d
\]

holds for some constant $c(\mu) > 0$, and

\[
K_0 := \sup_{\mu \in P_k} \|D^L f(\mu)\|_{L^{k^*}(\mu)} < \infty.
\]

Then

\[
|f(\mu_1) - f(\mu_2)| \leq K_0 \mathcal{W}_k(\mu_1, \mu_2), \quad \mu_1, \mu_2 \in P_k.
\]

**Proof.** Let $\xi_1, \xi_2$ be two random variables with

\[
\mathcal{L}_{\xi_1} = \mu_1, \quad \mathcal{L}_{\xi_2} = \mu_2, \quad \mathcal{W}_k(\mu_1, \mu_2) = (\mathbb{E}[|\xi_1 - \xi_2|^k])^{\frac{1}{k}}.
\]

Let $\eta$ be a normal random variable independent of $(\xi_1, \xi_2)$. Then

\[
\gamma(\varepsilon)(r) := \varepsilon \eta + r \xi_1 + (1 - r) \xi_2, \quad r \in [0, 1], \varepsilon \in (0, 1]
\]

are absolutely continuous with respect to the Lebesgue measure and hence atomless. By Theorem 2.1 in [6], (3.1) and the continuity of $D^L f(\mu)(\cdot)$ imply

\[
|f(L^\gamma_1) - f(L^\gamma_0)| = \left| \int_0^1 \mathbb{E}\left[(D^L f(L^\gamma_r))((\gamma_\varepsilon(r)), \xi_1 - \xi_2)\right]dr \right|
\]

\[
\leq \left(\mathbb{E}[|\xi_1 - \xi_2|^k]\right)^{\frac{1}{k}} \int_0^1 \|D^L f(L^\gamma_r)\|_{L^{k^*}(L^\gamma_0)}dr
\]

\[
\leq K\mathcal{W}_k(\mu_1, \mu_2), \quad \varepsilon \in (0, 1].
\]

Letting $\varepsilon \to 0$ we prove (3.3). \hfill \Box

In the following, we prove assertions (1) and (2) by using Zvonkin’s transform.

**Proof of Theorem 1.1(1).** By (H), we have $b_t^{(1)} \in \mathcal{B}_k$ with $\|D^L b_t^{(1)}(x, \mu)\|_{L^{k^*}(\mu)} \leq K$ for some constant $K > 0$. Then Lemma 3.1 implies

\[
|b_t^{(1)}(x, \mu) - b_t^{(1)}(x, \nu)| \leq K\mathcal{W}_k(\mu, \nu),
\]

so that the well-posedness of (1.1) follows from Theorem 3.1(2)(ii) in [19] for $D = \mathbb{R}^d$ for which condition (3) can be dropped from \((A_2^L)^\mu\) therein since $\partial D = \emptyset$, it is also implied by Theorem 1.1(2) in [14] where in (9) the condition $\bar{b}^\mu \in L_q^p$ can be weakened as $\bar{b}^\mu \in \tilde{L}_q^p$, since in the proof we may replace $L^p_q$ by $\tilde{L}_q^p$ according to Theorem 2.1 and Theorem 3.1 in [23].
To prove (1.8) and (1.9), we use Zvonkin’s transform. Consider the differential operator
\[ L_t^\mu = \frac{1}{2} \text{tr} \{ \sigma_t \sigma_t^* \nabla^2 \} + \nabla b\langle \cdot, \mu_t \rangle, \quad t \in [0, T]. \]

By [23, Theorem 2.1], (H) implies that for some \( \lambda_0 \) uniformly in \( \mu_0 \), when \( \lambda \geq \lambda_0 \) the PDE
\[ (\partial_t + L_t^\mu)u_t = \lambda u_t - b_t^{(0)}, \quad t \in [0, T], u_T = 0 \]
has a unique solution \( u \in \bar{H}^{2,p_0} \) such that (3.8) holds. Let \( \Theta_t := id + u_t \). By Itô’s formula in Lemma 3.3 of [23],
\[ Y_t := \Theta_t(X_t) = X_t + u_t(X_t) \]
solves the SDE
\[ dY_t = \{ b_t^{(1)}(X_t, \mu_t) + \lambda u_t(X_t) \} dt + \{ (\nabla \Theta_t) \sigma_t \}(X_t)dW_t, \quad Y_0 = \Theta_0(X_0). \]

By (2.6), there exists a constant \( c_1 > 1 \) such that
\[ |X_t| \leq c_1(1 + |Y_t|) \leq c_1^2(1 + |X_t|), \quad t \in [0, T]. \]

For any \( n \geq 1 \), let
\[ \gamma_{t,n} := \sup_{s \in [0,t]} |Y_s|, \quad \tau_n := \inf\{s \geq 0 : |Y_s| \geq n\}, \quad t \in [0, T]. \]

By BDG’s inequality, (H) and (2.6), for any \( j \geq 1 \) there exists a constant \( c(j) > 0 \) such that
\[ \mathbb{E}(\gamma_{t,n}^j | \mathcal{F}_0) \leq 2|Y_0|^j + c(j) \int_0^t \{ \mathbb{E}(\gamma_{s,n}^j | \mathcal{F}_0) + (\mathbb{E}[|Y_s|^k])^{\frac{j}{k}} + 1 \} ds + c(j), \quad n \geq 1, t \in [0, T]. \]

By Gronwall’s inequality,
\[ \mathbb{E}(\gamma_{t,n}^j | \mathcal{F}_0) \leq \left( 2|Y_0|^j + c(j) \int_0^t \{ (\mathbb{E}[|Y_s|^k])^{\frac{j}{k}} + 1 \} ds + c(j) \right) e^{c(j)t}, \quad n \geq 1, t \in [0, T]. \]

Taking expectations with \( j = k \) and letting \( n \to \infty \), we find a constant \( c_2 > 0 \) such that
\[ \mathbb{E}[\gamma_t^k] := \mathbb{E} \left[ \sup_{s \in [0,t]} |Y_s|^k \right] \leq c_2(1 + \mathbb{E}[|Y_0|^k]) + c_2 \int_0^t \mathbb{E}[|Y_s|^k] ds, \quad t \in [0, T]. \]

Noting that \( \sup_{t \in [0,T]} \mathbb{E}[|X_t|^k] < \infty \) as \( X_t \) is the solution of (1.1) for distributions in \( \mathcal{P}_k \), by combining this with (3.8) and \( \mathbb{E}[\gamma_t^k] \geq \mathbb{E}[|Y_s|^k] \) we obtain
\[ \mathbb{E}[\gamma_t^k] := \mathbb{E} \left[ \sup_{s \in [0,t]} |Y_s|^k \right] \leq c_2 + c_2 \int_0^t \mathbb{E}[\gamma_s^k] ds < \infty, \quad t \in [0, T], \]
so that by Gronwall’s inequality and (3.8), we prove (1.9) for some constant \( c > 0 \). Substituting this into (1.42) and letting \( n \to \infty \), we prove (1.8). \( \square \)
Proof of Theorem 1.1(2). Denote $\mu^i_t := \mathcal{L}X^i_t, i = 1, 2, t \in [0, T]$. Let $u$ solve (3.6) for $L^1_t$ replacing $L^2_t$ such that (2.6) holds. Let $\Theta_t = id + u_t$ and

$$Y^i_t = \Theta_t(X^i_t), \quad t \in [0, T], i = 1, 2.$$  

By (3.6) and Itô's formula we obtain

$$dY^1_t = \left\{b^1_t(X^1_t, \mu^1_t) + \lambda u_t(X^1_t)\right\}dt + \left\{(\nabla \Theta_t)\sigma_t\right\}(X^1_t)dW_t,$$

$$dY^2_t = \left\{b^1_t(X^2_t, \mu^2_t) + \lambda u_t(X^2_t) + \nabla b^1_t(X^2_t, \mu^1_t) - b^1_t(X^1_t, \mu^1_t)u_t(X^1_t)\right\}dt$$

$$+ \left\{(\nabla \Theta_t)\sigma_t\right\}(X^2_t)dW_t, \quad t \in [0, T].$$

So, by Itô’s formula, the process

$$v_t := Y^2_t - Y^1_t, \quad t \in [0, T]$$

satisfies the SDE

$$dv_t = \left\{b^1_t(X^2_t, \mu^2_t) + \lambda u_t(X^2_t) - b^1_t(X^1_t, \mu^1_t) + \nabla b^1_t(X^2_t, \mu^1_t)u_t(X^1_t)\right\}dt$$

$$+ \left\{(\nabla \Theta_t)\sigma_t\right\}(X^2_t) - (\nabla \Theta_t)\sigma_t(X^1_t)\}dW_t, \quad v_0 = \Theta_0(X^2_0) - \Theta_0(X^1_0).$$

By (2.6) and (3.4), we obtain

$$|b^1_t(x, \mu^2_t) - b^1_t(x, \mu^1_t)|^k \leq K^k \mathbb{E}[|X^2_t - X^1_t|^k] \leq (2K)^k \mathbb{E}[|Y^2_t - Y^1_t|^k].$$

Combining this with $(H)$, (2.6), Lemma 2.1 in [21], and applying Itô’s formula, for any $j \geq k$ we find a constant $c_1 > 0$ such that

$$|v_t|^{2j} \leq |v_0|^{2j} + c_1 \int_0^t |v_s|^{2j} \left\{1 + \sum_{i=0}^t \mathcal{M} f^2_i(s, X_s)\right\}ds$$

$$+ c_1 \int_0^t \left(\mathbb{E}[|v_s|^k]\right)^{\frac{2j}{k}} ds + M_t, \quad t \in [0, T]$$

holds for some local martingale $M_t$ with $M_0 = 0$. Since (2.6) implies

$$|v_0| \leq 2|X^1_0 - X^2_0|,$$

by stochastic Gronwall’s inequality [22, Lemma 3.7], Lemma 2.1 in [21] and Khasminskii’s estimate (2.9), we find a constant $c_2 > 0$ such that

$$\gamma_t := \sup_{s \in [0, t]} |v_s|, \quad t \in [0, T]$$

satisfies

$$\mathbb{E}[|\gamma_t|^j |\mathcal{F}_0] \leq c_2 \left(|X^1_0 - X^2_0|^{2j} + \int_0^t \left(\mathbb{E}[|v_s|^k]\right)^{\frac{2j}{k}} ds\right)^{\frac{1}{2}}$$

$$\leq c_2 |X^1_0 - X^2_0|^{2j} + \frac{1}{2} \sup_{s \in [0, t]} \left(\mathbb{E}[|v_s|^k]\right)^{\frac{2j}{k}} + \frac{c_2}{2} \int_0^t \left(\mathbb{E}[|v_s|^k]\right)^{\frac{2j}{k}} ds < \infty, \quad t \in [0, T].$$
Noting that \( \sup_{s \in [0,t]} \mathbb{E}[|v_s|^k] \leq \mathbb{E}[|\gamma_t|^k], \) by taking expectation in (3.11) with \( j = k, \) we derive
\[
\mathbb{E}[|\gamma_t|^k] \leq 2c_2 \mathbb{E}[|X_0^4 - X_0^2|^k] + c_2^2 \int_0^t \mathbb{E}[|\gamma_s|^k]ds, \quad t \in [0,T].
\]
Since \( \mathbb{E}[|\gamma_t|^k] < \infty \) due to (1.9), by Gronwall’s inequality we find a constant \( c > 0 \) such that
\[
\sup_{t \in [0,T]} \mathbb{E}[|v_t|^k] \leq \mathbb{E}[|\gamma_T|^k] \leq c \mathbb{E}[|X_0^4 - X_0^2|^k].
\]
Substituting this into (3.11) implies (1.10).

Proof of Theorem 1.1(3). Let \( \nu \in \mathcal{P}_k \) and take \( \mathcal{F}_0 \)-measurable random variables \( X_0, \tilde{X}_0 \) such that
\[
(3.12) \quad \mathcal{L}_{X_0} = \mu, \quad \mathcal{L}_{\tilde{X}_0} = \nu, \quad \mathbb{E}[|X_0 - \tilde{X}_0|^k] = \mathbb{W}_k(\mu, \nu)^k.
\]
Let \( X_t \) and \( \tilde{X}_t \) solve (1.1) with initial values \( X_0 \) and \( \tilde{X}_0 \) respectively, and denote \( \mu_t := P_t^* \mu = \mathcal{L}_{X_t}, \quad \nu_t := P_t^* \nu = \mathcal{L}_{\tilde{X}_t}, \quad t \in [0,T]. \)

Let \( P_t^\mu \) be the semigroup associated with \( X_t^{\mu,x}. \) According to Remark 2.1, (2.4) holds for \( P_t^\mu \) replacing \( P_t \) and some constant \( c > 0 \) independent of \( \mu. \) Then
\[
(3.13) \quad ||P_t^* \mu - (P_t^\mu)^* \nu||_{var} = ||(P_t^\mu)^* \mu - (P_t^\mu)^* \nu||_{var} \leq \frac{c}{\sqrt{t}} \mathbb{W}_1(\mu, \nu).
\]
On the other hand, let
\[
R_t := e^{-\int_0^t \zeta_s(X_s)\{b_s(\bar{X}_s, \nu_s) - b_s(\tilde{X}_s, \mu_s)dW_s\} - \frac{1}{2} \int_0^t \zeta_s(X_s)\{b_s(\bar{X}_s, \nu_s) - b_s(\tilde{X}_s, \mu_s)dW_s\}^2ds}.
\]
By (H) and Girsanov’s theorem, \( Q_t := R_t \mathbb{P} \) is a probability measure under which
\[
\tilde{W}_s := W_s + \int_0^s \zeta_s(X_s)\{b_s(\bar{X}_s, \nu_s) - b_s(\tilde{X}_s, \mu_s)\}ds, \quad r \in [0,t]
\]
is a Brownian motion. Reformulating the SDE for \( \tilde{X}_s \) as
\[
d\tilde{X}_s = b_s(\bar{X}_s, \mu_s)ds + \sigma_s(\tilde{X}_s)d\tilde{W}_s, \quad \mathcal{L}_{\tilde{X}_0} = \nu,
\]
by the uniqueness we obtain \( \mathcal{L}_{\tilde{X}_t|\tilde{Q}_t} = (P_t^\mu)^* \nu, \) so that by Pinsker’s inequality and (H), we find constants \( c_1 > 0 \) such that
\[
||P_t^* \nu - (P_t^\mu)^* \nu||_{var}^2 = \sup_{|f| \leq 1} |\mathbb{E}[f(\tilde{X}_t)(R_t - 1)]|^2 \leq 2\mathbb{E}[R_t \log R_t]
\]
\[
\leq c_1 \mathbb{E}[\tilde{Q}_t \int_0^t \mathbb{W}_k(\mu_s, \nu_s)^2ds = c_1 \int_0^t \mathbb{W}_k(\mu_s, \nu_s)^2ds.
\]
Combining this with (1.13) and (3.13), we prove (1.12) for some constant \( c > 0. \)
4 Proof of Theorem 1.2

A key step of the proof is to calculate $\nabla_\eta X^\mu_t$. In general, let $X^\mu_t$ solve (1.1) for $L^\mu_{X^\mu} = \mu \in P_k$, and for any $\varepsilon \in [0,1]$ and $\mathcal{F}_0$-measurable random variable $\eta$ with $L_\eta \in P_k$, let $X^\varepsilon_t$ solve (1.1) with $X^\varepsilon_0 = X^\mu_0 + \varepsilon \eta$. We intend to calculate

$$(4.1) \quad \nabla_\eta X^\mu_t := \lim_{\varepsilon \downarrow 0} \frac{X^\varepsilon_t - X^\mu_t}{\varepsilon}, \quad t \in [0,T]$$

in $L^k(\Omega \to C([0,T];\mathbb{R}^d),\mathbb{P})$. In particular, taking $\eta := \phi(X^\mu_0)$ for $\phi \in T_{\mu,k}$, we have

$$(4.2) \quad \nabla_\phi X^\mu_t = \nabla_\eta X^\mu_t, \quad t \in [0,T].$$

Choosing general $\eta$ instead of $\phi(X^\mu_0)$ is useful in the proof of Theorem 1.3.

4.1 The SDE for $\nabla_\eta X^\mu_t$

Comparing with $\nabla_v X^\varepsilon_t$ in Section 2, there are two essential differences in the study of $\nabla_\eta X^\mu_t$:

(1) the $L$-derivative of $b^{(1)}(x,\mu)$ will be involved;

(2) since $X^\mu_0$ is a random variable with $\mathbb{E}[|X^\mu_0|^k] < \infty$, in general we do not have

$$\sup_{\varepsilon \in [0,1]} \mathbb{E}[|X^\varepsilon_t|^k] < \infty, \quad j > k,$$

which is important for the dominated convergence theorem as used in the study of $\nabla_v X^\varepsilon_t$.

Point (1) will be managed by a chain rule due to [6] for functions on $P_k$, see [7] and [11] for earlier versions on $P_2$. Point (2) will be treated using the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_0]$ to replace the expectation $\mathbb{E}$, since we can prove $\mathbb{E}[|X^\varepsilon_t|^j|\mathcal{F}_0] < \infty$ for any $j \geq 1$.

Let $u$ solve (3.6) such that (2.6) and (2.13) hold as explained before. Let $\Theta_t = i_d + u_t$ and

$$(4.3) \quad Y^r_t := \Theta_t(X^\mu_r) = X^\mu_t + u_t(X^\mu_r), \quad t \in [0,T], \quad r \in [0,1].$$

By (3.6) and Itô’s formula, see Lemma 3.3 in [23], for any $r \in [0,1]$ we have

$$(4.4) \quad dY^r_t = \left\{b_t^{(1)}(X^\mu_t, \mu_t) + \lambda u_t(X^\mu_t) + \nabla b_t^{(1)}(X^\mu_t, \mu_t)u_t(X^\mu_t)\right\}dt + \left\{(\nabla \Theta_t)\sigma_t\right\}(X^\mu_t)dW_t, \quad Y^r_0 = \Theta_0(X^\mu_0) = X^\mu_0 + r\eta + u_0(X^\mu_0 + r\eta).$$

For any $t \in [0,T]$ and $v \in L^k(\Omega \to \mathbb{R}^d,\mathbb{P})$, let

$$(4.5) \quad \psi_t(v) := \mathbb{E}\left[(DL^b_t)^{(1)}(z,\mu_t)(X^\mu_t), (\nabla \Theta_t(X^\mu_t))^{-1}v)\right]_{z=X^\mu_t}.$$

By (H)(3), there exists a constant $K > 0$ such that for any $v, \bar{v} \in L^k(\Omega \to \mathbb{R}^d,\mathbb{P})$,

$$(4.6) \quad \psi_t(0) = 0, \quad |\psi_t(v) - \psi_t(\bar{v})| \leq K(\mathbb{E}[|v - \bar{v}|^k])^{\frac{1}{k}}, \quad t \in [0,T].$$
exists in \( L^k(\Omega \to C([0, T]; \mathbb{R}^d), \mathbb{P}) \), by (2.6), (2.13) and (4.3) we see that \( \nabla_{\eta}X_t^{\mu} \) exists in the same sense and

\[
\nabla_{\eta}X_t^{\mu} = (\nabla\Theta_t(X_t^{\mu}))^{-1}\nabla_{\eta}Y_t^0 = (\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta,
\]

Combining this with \((H)\), applying the chain rule Theorem 2.1 in [6], and noting that \( \mu_t \) is absolutely continuous due to Theorem 6.3.1 in [5], we obtain

\[
\lim_{\varepsilon \to 0} \frac{b_t^{(1)}(X_t^{\varepsilon}, \mu_t) - b_t^{(1)}(X_t^{\mu}, \mu_t)}{\varepsilon} = \psi_t(v_t^\eta) + \nabla(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta b_t^{(1)}(X_t^{\mu}, \mu_t),
\]

\[
\lim_{\varepsilon \to 0} \frac{\{(\nabla\Theta_t)(X_t^{\varepsilon})\} - \{(\nabla\Theta_t)(X_t^{\mu})\}}{\varepsilon} = \nabla(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta \{(\nabla\Theta_t)(X_t^{\mu})\},
\]

\[
\lim_{\varepsilon \to 0} \frac{u_t(X_t^{\varepsilon}) - u_t(X_t^{\mu})}{\varepsilon} = \nabla(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta u_t(X_t^{\mu}).
\]

Thus, if \( v_t^\eta \) in (4.7) exists, by (4.4) it should solve the SDE

\[
dv_t^\eta = \left\{ \psi_t(v_t^\eta) + \nabla(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta b_t^{(1)}(X_t^{\mu}, \mu_t) + \nabla\psi_t(v_t^\eta) + \lambda(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta u_t(X_t^{\mu}) \right\} dt \\
+ \nabla(\nabla\Theta_t(X_t^{\mu}))^{-1}v_t^\eta \{(\nabla\Theta_t)(X_t^{\mu})\} dW_t, \quad v_0^\eta = \eta + (\nabla\eta)u_0(X_0).
\]

Therefore, in terms of (4.8), to study \( \nabla_{\eta}X_t^{\mu} \) we first consider the SDE (4.10).

Lemma 4.1. Assume \((H)\). For any \( \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}) \), the SDE (4.10) has a unique solution, and for any \( j \geq 1 \) there exists a constant \( c > 0 \) such that

\[
\mathbb{E}\left[ \sup_{t \in [0, T]} |v_t^\eta|^{2j} \right] \leq c\left\{ (\mathbb{E}[|\eta|^k])^{\frac{1}{2}} + |\eta|^j \right\}, \quad \mu \in \mathcal{P}_k, \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}).
\]

Proof. We simply denote \( X_t = X_t^{\mu}, t \in [0, T] \).

(1) Well-posedness of (4.10). Consider the space

\[
\mathcal{C}_k := \left\{ (v_t)_{t \in [0, T]} \text{ is continuous adapted, } v_0 = v_0^\eta, \mathbb{E}\left[ \sup_{t \in [0, T]} |v_t|^k \right] < \infty \right\},
\]

20
which is complete under the metric

$$
\rho_\lambda(v^1, v^2) := \left( \mathbb{E} \left[ \sup_{t \in [0,T]} e^{-\lambda t} |v^1_t - v^2_t|^2 \right] \right)^\frac{1}{2}, \quad v^1, v^2 \in \mathcal{C}_k
$$

for $\lambda > 0$. By $(H)$, $(2.6)$ and $(4.6)$, there exist a constant $K > 0$ and a function $1 \leq f_0 \in \tilde{L}^p_0$, such that for any random variable $v$,

$$
\left| \nabla (\nabla\Theta(X))^{-1} \psi_\lambda(X_t, \mu_t) + \nabla \psi_\lambda(X_t, \mu_t) \right| \leq K|v|, \\
(4.12) \left| \nabla (\nabla\Theta(X))^{-1} \| \nabla\Theta(X) \| \right| \leq K|v| \sum_{i=0}^l f_i(t, X_t), \\
|\psi_\lambda(v)| \leq K(\mathbb{E}[|v|^k])^{\frac{1}{k}}, \quad t \in [0,T].
$$

Let $f = \sum_{i=0}^l f_i$. Let $\theta > 1$ such that $(\theta^{-1} p_i, \theta^{-1} q_i) \in \mathcal{K}, 0 \leq i \leq l$. By Krylov’s estimate Lemma 3.2(1) in $[23]$, we find a constant $c > 0$ such that

$$
\mathbb{E} \int_0^T f_t(X_t)^{2\theta} dt \leq c \sum_{i=0}^l \| f_i \|_{\tilde{L}^{p_i/\theta}_{q_i/\theta}} = c \sum_{i=0}^l \| f_i \|_{\tilde{L}^{p_i}_{q_i}}^{2\theta} < \infty.
$$

So,

$$
\tau_n := T \wedge \inf \left\{ t \geq 0 : \int_0^t |f_t(X_t)|^{2\theta} dt \geq n \right\} \to T \text{ as } n \to \infty.
$$

Thus,

$$
H_t(v) := v^\eta_0 + \int_0^t \left\{ \psi_\lambda(v_s) + \nabla (\nabla\Theta(X_s))^{-1} \psi_\lambda(X_s, \mu_s) \\
+ \nabla \psi_\lambda(X_s) \nabla (\nabla\Theta(X_s))^{-1} \| \nabla\Theta(X_s) \| \right\} ds \\
+ \int_0^t \nabla (\nabla\Theta(X_s))^{-1} \| \nabla\Theta(X_s) \| ds, \quad t \in [0,T]
$$

is an adapted continuous process on $\mathbb{R}^d$, and for any $n \geq 1$,

$$
H_{\tau_n} : \mathcal{C}_{k,n} \to \mathcal{C}_{k,n}, \quad \mathcal{C}_{k,n} := \left\{ (v_{\tau_n}) : \ v \in \mathcal{C}_k \right\}.
$$

So, it remains to prove that $H$ has a unique fixed point $v^\eta \in \mathcal{C}_k$ satisfying $(4.11)$, which is then the unique solution of $(4.10)$. In the following we explain that it suffices to prove

$$
(4.14) \text{ } H_{\tau_n} \text{ has a unique fixed point in } \mathcal{C}_{k,n}, \text{ } n \geq 1.
$$

Indeed, if $(4.14)$ holds, then the unique fixed point $v^\eta_{\tau_n}$ satisfies

$$
v^{\eta_{\tau_n}} = v^{\eta_{\tau_n}+k}, \quad n, k \geq 1,
$$
so that

\[ v^\eta_t := \lim_{n \to \infty} v^{\eta,n}_{t\wedge \tau_n} \]

is a continuous adapted process on \( \mathbb{R}^d \), and

\[ H_{\wedge \tau_n}(v^n) = v^n_{t\wedge \tau_n} \in \mathcal{C}_{k,n}, \quad n \geq 1. \]

By this and (4.12), for any \( j \geq k \) we find a constant \( c > 0 \) such that

\[ d|v^\eta_t|^{2j} \leq c\left( \mathbb{E}[|v^\eta_{t\wedge \tau_n}|^k] \right)^{\frac{2j}{k}} + |v^\eta_t(2j)(1 + f^2_t(X_t))dt + d\tilde{M}_t, \quad t \in [0, \tau_n], \]

holds for some local martingale \( \tilde{M}_t \). By the stochastic Gronwall inequality, we find constants \( k_1, k_2 > 0 \) such that

\[
\mathbb{E}\left[ \sup_{t \in [0,T]} |v^\eta_{t\wedge \tau_n}|^j \bigg| \mathcal{F}_0 \right] \leq k_1 \left( \int_0^T \mathbb{E}[|v^\eta_{t\wedge \tau_n}|^k] \right)^{\frac{2j}{k}} ds + \mathbb{E}[|v^\eta_0(2j)|] \]

\[
\leq k_2 |\eta|^j + k_1 \left( \int_0^T \mathbb{E}[|v^\eta_{t\wedge \tau_n}|^k] \right)^{\frac{j}{k}} ds .
\]

Taking \( j = k \) we obtain

\[
\mathbb{E}\left[ \sup_{t \in [0,T]} |v^\eta_{t\wedge \tau_n}|^k \bigg| \mathcal{F}_0 \right] \leq k_2 |\eta|^k + \frac{k_2^2}{2} \int_0^T \mathbb{E}[|v^\eta_{t\wedge \tau_n}|^k] ds + \frac{1}{2} \mathbb{E}\left[ \sup_{t \in [0,T]} |v^\eta_{t\wedge \tau_n}|^k \right], \quad t \in [0, T].
\]

Taking expectation and applying Gronwall's inequality, we find a constant \( k_3 > 0 \) such that

\[
\sup_{n \geq 1} \mathbb{E}\left[ \sup_{t \in [0,T]} |v^\eta_{t\wedge \tau_n}|^k \right] \leq k_3 \mathbb{E}[|\eta|^k],
\]

so that (4.13) with \( n \to \infty \) implies (4.11), and it is the unique solution of (4.10) in \( \mathcal{C}_k \) since for each \( n \geq 1, v^\eta_{t\wedge \tau_n} \) is the unique fixed point of \( H_{\wedge \tau_n} \) in \( \mathcal{C}_{k,n} \).

(2) We now verify (4.14). By (4.6) and (4.12), we find constants \( c_1, c_2 > 0 \) such that

\[
\rho_\lambda(H_{\wedge \tau_n}(v^1), H_{\wedge \tau_n}(v^2)) = \mathbb{E}\left[ \sup_{t \in [0,\tau_n]} e^{-\lambda t}|H_t(v^1) - H_t(v^2)|^k \right]
\]

\[
\leq c_1 \mathbb{E}\left[ \sup_{t \in [0,\tau_n]} e^{\lambda t}\left\{ \left( \int_0^t |v^1_s - v^2_s|^k + \mathbb{E}[|v^1_s - v^2_s|^k]^{\frac{1}{k}} ds \right) \right\}^k \right] + \left( \int_0^t |v^1_s - v^2_s|^2 f_s(X_s)^2 ds \right)^{\frac{k}{2}} \right],
\]

\[
\leq 2c_1 T^{k-1} \rho_{\lambda,n}(v^1, v^2) \sup_{t \in [0,T]} \int_0^t e^{-\lambda(t-s)} ds \]

\[
+ c_1 \mathbb{E}\left[ \sup_{t \in [0,\tau_n]} \left( e^{\lambda t}|v^1_t - v^2_t|^k \right)^{\frac{k}{2}} \right] \left( \int_0^t e^{2\lambda(t-s)} f_s(X_s)^2 ds \right)^{\frac{k}{4}}
\]

22
Therefore, when $\lambda > 0$ is large enough, $H_{\lambda \tau_0}$ is contractive in $\rho_{\lambda}$ for large $\lambda > 0$, and hence has a unique fixed point on $\mathcal{C}_{k,n}$.

\[ \square \]

### 4.2 Proofs of Theorem 1.2(1)

Theorem 1.2(1) is implied by the following result for $\eta = \phi(X_0^\mu)$.

**Proposition 4.2.** Assume (H). For any $v \in \mathbb{R}^d$ and $\eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P})$, $\nabla \eta X_t$ and $\nabla_v X_t^x$ exist in $L^k(\Omega \to C([0,T]; \mathbb{R}^d), \mathbb{P})$, and for any $j \geq 1$ there exists a constant $c > 0$ such that

\begin{equation}
\mathbb{E} \left[ \sup_{t \in [0,T]} |\nabla \eta X_t^{\mu \mid j}| \mathcal{F}_0 \right] \leq c \left( \{ \mathbb{E}[|\eta|^j] \}^{\frac{\mu}{j}} + |\eta|^j \right), \quad \mu \in \mathcal{P}_k, \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}), \tag{4.16} \end{equation}

\begin{equation}
\mathbb{E} \left[ \sup_{t \in [0,T]} |\nabla_v X_t^{\mu \mid x \mid j}| \right] \leq c |v|^j, \quad x, v \in \mathbb{R}^d, \mu \in \mathcal{P}_k. \tag{4.17} \end{equation}

**Proof.** The existence of $\nabla_v X_t^{\mu \mid x \mid j}$ and (4.17) follow from Theorem 2.1(1) for $b_t(x) := b_t(x, \mu_t)$ where the constant in (2.2) is uniformly in $\mu_t$ according to Remark 2.1. So, it suffices to prove (4.16). We simply denote

\[ X_t = X_t^\mu, \quad v_t = v_t^0, \quad t \in [0,T]. \]

For any $r \in (0,1]$ let $Y_t^r$ be in (4.13). We have $Y_t := Y_t^0 = \Theta_t(X_t)$. Let

\begin{equation}
\tilde{v}_t^\varepsilon := \frac{Y_t^\varepsilon - Y_t}{\varepsilon}, \quad t \in [0,T], \varepsilon \in (0,1). \tag{4.18} \end{equation}

By Theorem 1.1(2) and (2.6), for any $j \geq 1$ there exists $c(j) > 0$ such that

\begin{equation}
\mathbb{E} \left[ \sup_{t \in [0,T]} |\tilde{v}_t^\varepsilon|^j | \mathcal{F}_0 \right] \leq c(j) \left( \{ \mathbb{E}[|\eta|^j] \}^{\frac{\mu}{j}} + |\eta|^j \right), \quad \varepsilon \in (0,1). \tag{4.19} \end{equation}

We claim that it suffices to prove

\begin{equation}
\lim_{\varepsilon \downarrow 0} \mathbb{E} \left[ \sup_{t \in [0,T]} |\tilde{v}_t^\varepsilon - v_t|^j \right] = 0. \tag{4.20} \end{equation}

Indeed, this implies that

\[ \nabla \eta Y_t := \lim_{\varepsilon \downarrow 0} \tilde{v}_t^\varepsilon = v_t \]

exists in $L^k(\Omega \to C([0,T]; \mathbb{R}^d); \mathbb{P})$, so that (2.6), (2.13) and $\Theta_t := id + v_t$ yield

\[ \nabla \eta X_t := \lim_{\varepsilon \downarrow 0} \frac{X_t^\varepsilon - X_t}{\varepsilon} = (\nabla \Theta_t(X_t))^{-1} v_t \]
exists in the same space, and (4.16) follows from (4.11).

Recall that $\mu_\varepsilon = \mathcal{L}_{X_\varepsilon}, \varepsilon \in [0, 1]$. By (4.6) and Itô's formula, we obtain

$$d\bar{v}_t^\varepsilon = \frac{1}{\varepsilon} \left\{ \lambda u(X_t^\varepsilon) + b^{(1)}_t(X_t^\varepsilon, \mu_t) - \lambda u(X_t) - b_t(X_t, \mu_t) \right\} dt$$

$$+ \left\{ \nabla b^{(1)}_t(X_t^\varepsilon, \mu_t) - b^{(1)}_t(X_t^\varepsilon, \mu_t) \Theta_t(X_t^\varepsilon) \right\} dt$$

$$+ \frac{1}{\varepsilon} \left\{ (\nabla \Theta_t) \sigma_t \right\} (X_t^\varepsilon) - (\nabla \Theta_t) \sigma_t \right\} (X_t) dW_t, \quad \bar{v}_0^\varepsilon = \frac{\Theta_0(X_0^\varepsilon) - \Theta_0(X_0)}{\varepsilon}.$$ (4.21)

Then

$$\bar{v}_t^\varepsilon = \bar{v}_0^\varepsilon + \int_0^t \left\{ \nabla (\nabla \Theta_s(X_s))^{-1} (b^{(1)}_s(\cdot, \mu_s) + \lambda u_s) (X_s) + \nabla \psi_s(\varepsilon) \Theta_s(X_s) \right\} ds$$

$$+ \int_0^t \nabla (\nabla \Theta_s(X_s))^{-1} (\nabla \Theta_s) \sigma_s (X_s) dW_s + \alpha_s^\varepsilon, \quad t \in [0, T],$$ (4.22)

where $\psi_t(v)$ is in (4.5), and for $t \in [0, T]$,

$$\alpha_s^\varepsilon := \int_0^t \xi_s^\varepsilon ds + \int_0^t \eta_s^\varepsilon dW_s,$$

$$\xi_s^\varepsilon := \frac{1}{\varepsilon} \left\{ [b^{(1)}_s(\cdot, \mu_s) + \lambda u_s] (X_s^\varepsilon) - [b^{(1)}_s(\cdot, \mu_s) + \lambda u_s] (X_s) + \nabla b^{(1)}_s(X_s^\varepsilon, \mu_s) \Theta_s(X_s^\varepsilon) \right\} - \left\{ \nabla (\nabla \Theta_s(X_s))^{-1} \varepsilon \{ \nabla \Theta_s \} \sigma_s (X_s) \right\},$$

$$\eta_s^\varepsilon := \frac{\{ \nabla \Theta_s \} \sigma_s (X_s^\varepsilon) - \{ \nabla \Theta_s \} \sigma_s (X_s)}{\varepsilon} - \nabla (\nabla \Theta_s(X_s))^{-1} \varepsilon \{ \nabla \Theta_s \} \sigma_s (X_s).$$

We claim

$$\lim_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{t \in [0, T]} |\alpha_t^\varepsilon|^{n} \right| \mathcal{F}_0 = 0, \quad n \geq 1.$$ (4.23)

This can be proved by the argument leading to (2.17), but with the conditional expectation $\mathbb{E}[\cdot | \mathcal{F}_0]$ replacing the expectation.

Firstly, by (4.19), $Y_t^\varepsilon = X_t^\varepsilon + u_t(X_t^\varepsilon)$ and (2.6), for any $j \geq 1$ there exists $c(j) > 0$ such that

$$\sup_{\varepsilon \in (0, 1]} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \frac{X_t^\varepsilon - X_t}{\varepsilon} \right|^{j} \right| \mathcal{F}_0 \leq c(j) \left( \mathbb{E}[|\eta|^k] \right) \frac{1}{j} + |\eta|^j.$$ (4.24)

Since $\{ \nabla \Theta_s \} \sigma_s, \Theta_s^{(1)}(\cdot, \mu_s)$ and $\nabla u_s$ are a.e. differentiable, by the same reason leading to (2.18), (4.24) implies that for any $s \in (0, T]$, $\mathbb{P}$-a.s.

$$\lim_{\varepsilon \to 0} \left| \frac{\{ \nabla \Theta_s \} \sigma_s(X_s^\varepsilon) - \{ \nabla \Theta_s \} \sigma_s(X_s)}{\varepsilon} - \nabla (\nabla \Theta_s(X_s))^{-1} \varepsilon \{ \nabla \Theta_s \} \sigma_s(X_s) \right| = 0,$$

$$\lim_{\varepsilon \to 0} \left| \frac{b^{(1)}_s(\cdot, \mu_s) + \lambda u_s(X_s^\varepsilon)}{\varepsilon} - \frac{b^{(1)}_s(\cdot, \mu_s) + \lambda u_s(X_s)}{\varepsilon} - \nabla (\nabla \Theta_s(X_s))^{-1} \varepsilon \frac{b^{(1)}(\cdot, \mu_s) + \lambda u_s(X_s)}{\varepsilon} \right| = 0.$$
Next, as in (4.9), by the chain rule in Theorem 2.1 of [6] and \( b_t^{(1)} \in \mathcal{D}_k \), we obtain
\[
\lim_{\varepsilon \to 0} \left| \frac{b_s^{(1)}(X_s^\varepsilon, \mu_s^\varepsilon) - b_s^{(1)}(X_s^\varepsilon, \mu_s)}{\varepsilon} - \psi_s(\tilde{v}_s^\varepsilon) \right| = 0, \quad s \in (0, T].
\]
Thus, for any \( s \in (0, T] \), as \( \varepsilon \to 0 \) we have \( \mathbb{P}\text{-a.s.} \)
\[
\lim_{\varepsilon \to 0} \{ |\xi_s^\varepsilon| + ||\eta_s^\varepsilon|| \} = 0.
\]
Moreover, by (H) and Lemma 2.1 in [21], we find a constant \( c > 0 \) such that
\[
|\xi_s^\varepsilon| + ||\eta_s^\varepsilon|| \leq c|\tilde{v}_s^\varepsilon| \left( 1 + \sum_{i=0}^{l} \{ \mathcal{M} f_i(s, \cdot)(X_s) + \mathcal{M} f_i(s, \cdot)(X_s^\varepsilon) \} \right), \quad s \in [0, T].
\]
Finally, let \( \theta > 1 \) be in the proof of (2.17) such that (2.19) holds for \( X_t^\varepsilon \) replacing \( X_t^{\varepsilon+2\varepsilon} \). By (2.19) for \( X_t^\varepsilon \), (1.19), and Lemma 2.1 in [21], for any \( n \geq 1 \) there exist constants \( c_1(n), c_2(n) > 0 \) such that
\[
\mathbb{E} \left[ \left( \int_0^T I_s^{2\theta} \, ds \right)^n \right] \bigg| \mathcal{F}_0 \bigg]
\leq c_1(n) \mathbb{E} \left[ \left( \sup_{s \in [0, T]} |\tilde{v}_s^\varepsilon|^{4\theta n} \right) \left( \int_0^T \left( 1 + \sum_{i=0}^{l} \{ \mathcal{M} f_i^{(2\theta)}(s, X_s) + \mathcal{M} f_i^{(2\theta)}(s, X_s^\varepsilon) \} \right) \, ds \right)^n \right] \bigg| \mathcal{F}_0 \bigg]
\leq c_1(n) \left( \mathbb{E} \left[ \sup_{s \in [0, T]} |\tilde{v}_s^\varepsilon|^{4\theta n} \right] \right)^{\frac{1}{2}}
\times \left( \mathbb{E} \left[ \left( \int_0^T \left( 1 + \sum_{i=0}^{l} \{ \mathcal{M} f_i^{(2\theta)}(s, X_s) + \mathcal{M} f_i^{(2\theta)}(s, X_s^\varepsilon) \} \right) \, ds \right)^{2\theta n} \right] \right)^{\frac{1}{2}}
\leq c_2(n) (1 + |\eta|^{2\theta n}) < \infty.
\]
By BDG’s inequality and the dominated convergence theorem, this and (4.25) imply (4.23).

Now, by (4.10) and (4.22), the argument leading to (3.10) gives
\[
|v_t - \tilde{v}_t^\varepsilon|^{2k} \leq |v_0 - \tilde{v}_0^\varepsilon|^{2k} + \int_0^t \left( |v_s - \tilde{v}_s^\varepsilon|^{2k} \gamma_t + \left( \mathbb{E}[|\tilde{v}_s^\varepsilon|^{2k}] \right)^2 \right) \, dt
+ K \sup_{r \in [0, t]} |\alpha_r^\varepsilon|^{2k} + M_t, \quad t \in [0, T],
\]
where \( K > 0 \) is a constant and \( \gamma_t \) is a positive process satisfying
\[
\mathbb{E} \left[ e^{N \int_0^T \gamma_t \, dt} \right] < \infty, \quad N > 0.
\]
Therefore, by the stochastic Gronwall inequality [22, Lemma 3.7], we find a constant \( c > 0 \) such that
\[
\mathbb{E} \left[ \sup_{s \in [0, t]} |\tilde{v}_s^\varepsilon - v_s|^{k} \right] \bigg| \mathcal{F}_0 \bigg]
\leq c |v_0 - \tilde{v}_0^\varepsilon|^{k} + c \left( \mathbb{E} \left[ \sup_{s \in [0, t]} |\alpha_s^\varepsilon|^{2k} \right] \bigg| \mathcal{F}_0 \bigg) \right)^{\frac{1}{2}} + c \left( \int_0^t \mathbb{E}[|\tilde{v}_s^\varepsilon - v_s|^{k}] \, ds \right)^{\frac{1}{2}}, \quad t \in [0, T].
\]
By (4.23), (4.26) and noting that \( \lim_{\varepsilon \to 0} |v_0 - \tilde{v}_0| = 0 \), we obtain

\[
(4.27) \quad \limsup_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{s \in [0,t]} |\tilde{v}_s^\varepsilon - v_s| \bigg| \mathcal{F}_0 \right] \leq c \limsup_{\varepsilon \to 0} \left( \int_0^t \left( \mathbb{E}[|\tilde{v}_s^\varepsilon - v_s|^k] \right)^{\frac{1}{k}} ds \right)^{\frac{k}{2}}.
\]

Taking \( j = k \) in (4.11) and (4.19) and (4.24) we see that

\[
\left\{ \mathbb{E} \left[ \sup_{t \in [0,T]} \{|\tilde{v}_t^\varepsilon| + |v_t^\varepsilon|\} \bigg| \mathcal{F}_0 \right] : \varepsilon \in (0,1] \right\}
\]

is uniformly integrable with respect to \( \mathbb{P} \), so that by Fatou’s lemma, (4.27) implies

\[
\begin{aligned}
\quad h_t := \limsup_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{s \in [0,t]} |\tilde{v}_s^\varepsilon - v_s|^k \bigg| \mathcal{F}_0 \right] &= \limsup_{\varepsilon \to 0} \mathbb{E} \left[ \mathbb{E} \left[ \sup_{s \in [0,t]} |\tilde{v}_s^\varepsilon - v_s|^k \bigg| \mathcal{F}_0 \right] \right] \\
&\leq \mathbb{E} \left[ \limsup_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{s \in [0,t]} |\tilde{v}_s^\varepsilon - v_s|^k \bigg| \mathcal{F}_0 \right] \right] \leq c \left( \int_0^t h_s^2 ds \right)^{\frac{1}{2}}, \quad t \in [0,T]
\end{aligned}
\]

and \( h_t < \infty \), so that \( h_t = 0 \) for all \( t \in [0,T] \). Therefore, (4.20) holds and hence the proof is finished. \( \square \)

### 4.3 Proof of Theorem 1.2(2)

For any \( \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}) \), \( \mu \in \mathcal{P}_k \), and \( \varepsilon \in [0,1] \), let \( X^\varepsilon_t \) solve (1.1) for \( X^\varepsilon_0 = X^\mu_0 + \varepsilon \eta \). Consider

\[
\Gamma_\eta(f(X^\mu_t)) := \lim_{\varepsilon \to 0} \frac{\mathbb{E}[f(X^\varepsilon_t) - f(X^\mu_t)]}{\varepsilon}, \quad t \in (0,T], f \in \mathcal{B}_b(\mathbb{R}^d).
\]

Theorem 1.2(2) is implied by the following result for \( \eta = \phi(X_0) \).

**Proposition 4.3.** Assume (H). For any \( \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}) \) and \( \mu \in \mathcal{P}_k \), \( D_t^1 P_t f(\mu) \) exists and satisfies the following formula for any \( \beta \in C^1([0,t]) \) with \( \beta_0 = 0 \) and \( \beta_t = 1 \):

\[
\Gamma_\eta(f(X^\mu_t)) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{E} \left[ f(X^\mu_t^\varepsilon) \int_0^t \beta_s^{\varepsilon}(X^\mu_t^\varepsilon) \nabla_v X^\mu_t^\varepsilon, dW_s \right] \mathcal{L}(X^\mu_t^\varepsilon, \eta)(dx, dv) \\
+ \mathbb{E} \left[ f(X^\mu_t^\varepsilon) \int_0^t \beta_s^{\varepsilon}(X^\mu_t^\varepsilon) \mathbb{E} \left[ (D^1 L^{b_1^{(1)}(z, P^*_s \mu)(X^\mu)_s, \nabla \eta X^\mu_s) \bigg|_{z = X^\mu_t^\varepsilon}, dW_s \right] \right].
\]

Consequently, there exists a constant \( c > 0 \) such that

\[
|\Gamma_\eta(f(X^\mu_t))| \leq \frac{c}{t} \left( P_t |f| f^{k^*}(\mu) \right)^{\frac{1}{k}} \left( \mathbb{E}[|\eta|^k] \right)^{\frac{1}{k}}, \quad t \in (0,T], f \in \mathcal{B}_b(\mathbb{R}^d), \mu \in \mathcal{P}_k, \eta \in L^k(\Omega \to \mathbb{R}^d, \mathcal{F}_0, \mathbb{P}).
\]

**Proof.** Let \( X^{\mu,x}_t \) solve (1.6). Since \( X^\mu_t \) solve (1.6) with initial value \( X^\mu_0 \), the pathwise uniqueness implies

\[
(4.30) \quad X^\mu_t = X^{\mu,x}_t, \quad t \in [0,T].
\]
Let \( (P^\mu_{s,t})_{0 \leq s \leq t \leq T} \) be the semigroup associated with (1.6), i.e. for \((X^\mu_{s,t})_{t \in [s,T]}\) solving (1.6) from time \( s \) with \( X^\mu_{s,s} = x \),

\[
P^\mu_{s,t}f(x) := \mathbb{E}[f(X^\mu_{s,t})], \quad t \in [s,T], \ x \in \mathbb{R}^d.
\]

Simply denote \( P^\mu_t = P^\mu_{0,t} \). Then (4.30) implies

\[
P_t f(\mu) = \mathbb{E}[f(X^\mu_t)] = \int_{\mathbb{R}^d} P^\mu_t f(x) \mu(dx), \quad t \in [0,T], f \in \mathcal{B}_b(\mathbb{R}^d).
\]

By Theorem 2.1 (H) implies that for any \( t \in (0,T) \) and \( \beta \in C^1([0,t]) \) with \( \beta_0 = 0 \) and \( \beta_t = 1 \),

\[
\nabla_v P^\mu_t f(x) = \mathbb{E} \left[ f(X^\mu_t) \int_0^t \beta_s \langle \zeta_s(X^\mu_s \nabla_v X^\mu_s, dW_s) \rangle, \quad v \in \mathbb{R}^d, f \in \mathcal{B}_b(\mathbb{R}^d).
\]

Next, denote \( \mu_t = P^\mu_t \mu = \mathcal{L}_{X^\mu_t} \) and let \( \bar{X}_s^\varepsilon \) solve (1.6) for \( \bar{X}_0^\varepsilon = X_0^\varepsilon \), i.e.

\[
d\bar{X}_s^\varepsilon = b_\varepsilon(X^\varepsilon_s, \mu_s)ds + \sigma_s(X^\varepsilon_s)dW_s, \quad s \in [0,t], \bar{X}_0^\varepsilon = X_0^\varepsilon.
\]

We have

\[
\mathbb{E}[f(\bar{X}_t^\varepsilon)] = \int_{\mathbb{R}^d} (P^\mu_t(\mu)) \mathcal{L}_{X^\varepsilon_0}(dx) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla_v P^\mu_t f(x + \varepsilon v) \mathcal{L}_{(X^\varepsilon_0, \eta)}(dx, dv), \quad f \in \mathcal{B}_b(\mathbb{R}^d).
\]

Combining this with (4.32) and (4.33), and applying the dominated convergence theorem, we obtain

\[
\lim_{\varepsilon \to 0} \frac{\mathbb{E}[f(\bar{X}_t^\varepsilon)] - P_t f(\mu)}{\varepsilon} = \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla_v P^\mu_t f(x) \mathcal{L}_{(X^\varepsilon_0, \eta)}(dx, dv)
\]

On the other hand, denote \( \mu_t^\varepsilon = \mathcal{L}_{X^\varepsilon_t} \) and let

\[
R_t^\varepsilon := e^{\int_0^t \zeta_s(X^\varepsilon_s) \{ b_s^{(1)}(X^\varepsilon_s, \mu_s) - b_s^{(1)}(X^\varepsilon_s, \mu_s^\varepsilon) \} ds} + \xi_t^\varepsilon | \zeta_s(X^\varepsilon_s) \{ b_s^{(1)}(X^\varepsilon_s, \mu_s) - b_s^{(1)}(X^\varepsilon_s, \mu_s^\varepsilon) \} |^2 ds.
\]

By (H), \( \zeta_s = \sigma_s^\varepsilon (\sigma_s \sigma_s^\varepsilon)^{-1} \) and Girsanov’s theorem, \( \mathcal{Q}_t^\varepsilon := R_t^\varepsilon \mathbb{P} \) is a probability measure under which

\[
\tilde{W}_r^\varepsilon := W_r - \int_0^r \zeta_s(X^\varepsilon_s) \{ b_s^{(1)}(X^\varepsilon_s, \mu_s) - b_s^{(1)}(X^\varepsilon_s, \mu_s^\varepsilon) \} ds, \quad r \in [0,t]
\]

is a Brownian motion, and

\[
\sup_{r \in [0,T], \varepsilon \in (0,1]} \mathbb{E} \left[ \left| \frac{R_r^\varepsilon - 1}{\varepsilon^{j}} \right| \right] < \infty, \quad j \geq 1.
\]
Reformulate the SDE for $X_s^\varepsilon$ as
\[ dX_s^\varepsilon = b_s(X_s^\varepsilon, \mu_s) + \sigma_s(X_s^\varepsilon) d\tilde{W}_s^\varepsilon, \quad X_0^\varepsilon = \tilde{X}_0^\varepsilon. \]

By the well-posedness we obtain $L_{X_0^\varepsilon|Q_t} = L_{\tilde{X}_0^\varepsilon|P}$, so that
\[ \mathbb{E}[f(\tilde{X}_t^\varepsilon)] = \mathbb{E}[R_t^\varepsilon f(X_t^\varepsilon)], \quad f \in \mathcal{B}_b(\mathbb{R}^d). \]

Thus,
\[ \frac{\mathbb{E}[f(X_t^\varepsilon)] - \mathbb{E}[f(\tilde{X}_t^\varepsilon)]}{\varepsilon} = \frac{\mathbb{E}[f(X_t^\varepsilon)(1 - R_t^\varepsilon)]}{\varepsilon} = I_1(\varepsilon) + I_2(\varepsilon), \]
where
\[ I_1(\varepsilon) := \mathbb{E} \left[ f(X_t^\varepsilon) \frac{1 - R_t^\varepsilon}{\varepsilon} \right], \quad I_2(\varepsilon) := \mathbb{E} \left[ \{f(X_t^\varepsilon) - f(X_t^\varepsilon)\} \frac{1 - R_t^\varepsilon}{\varepsilon} \right]. \]

By (3.4), (1.9) and the dominated convergence theorem, we obtain
\[ \lim_{\varepsilon \to 0} I_1(\varepsilon) = \mathbb{E} \left[ f(X_t^\varepsilon) \int_0^t \langle \zeta_s(X_s^\varepsilon) \mathbb{E}\{D^\varepsilon b_s^{(1)}(z, \mu_s), \nabla_r X_s^\varepsilon)\} \rangle_{\varepsilon=X_t^\varepsilon} dW_s \right]. \]

So, to prove (4.28) it suffices to verify
\[ \lim_{\varepsilon \to 0} I_2(\varepsilon) = 0. \]

By (3.4), we have
\[ \lim_{s \uparrow t} \sup_{\varepsilon \in (0, \varepsilon]} \mathbb{E} \left[ \frac{|R_s^\varepsilon - R_t^\varepsilon|}{\varepsilon} \right] = 0. \]

Since (H) holds for $[r, T]$ replacing $[0, T]$, (1.12) holds for $(r, T)$ replacing $(0, T)$. Similarly, (2.4) holds for $P_{r,t}^{\mu^\varepsilon}$ and $P_{r,t}^{\mu}$ defined in (4.31) replacing $P_{t-r}$. Therefore, by the Markov property,
\[ |\mathbb{E}[f(X_t^\varepsilon) - f(X_t^\mu)|\mathcal{F}_r]| \leq |(P_{r,t}^{\mu^\varepsilon} f)(X_t^\varepsilon) - (P_{r,t}^{\mu} f)(X_t^\mu)| \leq c \|f\|_{\infty} \left( \frac{|X_t^\varepsilon - X_t^\mu|}{\sqrt{t-s}} \wedge 1 \right) + |(P_{r,t}^{\mu^\varepsilon} f)(X_t^\mu) - (P_{r,t}^{\mu} f)(X_t^\mu)|. \]

On the other hand, let $(\tilde{X}_{s,r}^\varepsilon)_{s \in [r,t]}$ solve the SDE
\[ d\tilde{X}_{s,r}^\varepsilon = b_s(\tilde{X}_{s,r}^\varepsilon, \mu_s) ds + \sigma_s(\tilde{X}_{s,r}^\varepsilon) dW_s, \quad \tilde{X}_{r,r}^\varepsilon = X_r^\mu, \quad s \in [r,t]. \]

We have
\[ P_{r,t}^{\mu^\varepsilon} f(X_t^\varepsilon) = \mathbb{E} \left[ f(\tilde{X}_{r,t}^\varepsilon)|\mathcal{F}_r \right], \quad P_{r,t}^{\mu} f(X_t^\mu) = \mathbb{E} \left[ f(X_t^\mu)|\mathcal{F}_r \right]. \]

Noting that (3.4) and (1.13) imply
\[ |b(x, \mu_t^\varepsilon) - b_t(x, \mu_t)| \leq c_1 \mathbb{W}_k(\mu_0^\varepsilon, \mu_0) \leq c_1 \varepsilon (\mathbb{E}[|\eta|^k])^{\frac{1}{k}}. \]

28
for some constant $c_1 > 0$, by Girsanov’s theorem,
\[
R_{t,r}^\varepsilon := e^{\int_r^t \zeta(X_s^\varepsilon)\{b_s(X_s^\varepsilon, \mu_s^\varepsilon) - b_s(X_s^\mu, \mu_s)\} \, dW_s} - \frac{1}{2} \int_r^t \zeta(X_s^\varepsilon)\{b_s(X_s^\varepsilon, \mu_s^\varepsilon) - b_s(X_s^\mu, \mu_s)\}^2 \, ds
\]
is a probability density such that under $\mathbb{Q}_{t,r} := R_{t,r}^\varepsilon \mathbb{P}$,
\[
\tilde{W}_s := W_s - \int_r^s \zeta(X_s^\varepsilon)\{b_s(X_s^\varepsilon, \mu_s) - b_s(X_s^\mu, \mu_s)\} \, d\theta, \quad s \in [r, t]
\]
is a Brownian motion. Reformulating the SDE for $(X_s^\varepsilon)_{s \in [r,t]}$ as
\[
dX_s^\mu = b_s(X_s^\mu, \mu_s^\varepsilon) \, ds + \sigma_s(X_s^\mu) \, d\tilde{W}_s, \quad X_r^\mu = \bar{X}_{r,r}^\varepsilon, \quad s \in [r, t],
\]
by the uniqueness we obtain
\[
P_{r,t}^\mu f(X_t^\mu) = \mathbb{E}\left[ R_{r,t}^\varepsilon f(X_t^\mu) \mid \mathcal{F}_r \right],
\]
so that by Pinsker’s inequality and (4.40), we find constants $c_2 > 0$ such that
\[
\begin{align*}
|P_{r,t}^\mu f(X_t^\mu) - (P_{r,t}^\mu f)(X_t^\mu)|^2 &\leq \|f\|_{\infty}^2 \mathbb{E}[|1 - R_{r,t}^\varepsilon|\mathcal{F}_r]^2 \\
&\leq 2\|f\|_{\infty}^2 \mathbb{E}_{Q_{r,t}}[\log R_{r,t}^\varepsilon]\mathcal{F}_0 \\
&= \|f\|_{\infty}^2 \int_r^t \mathbb{E}_{Q_{r,t}}[\zeta(X_s^\mu)\{b_s(X_s^\mu, \mu_s^\varepsilon) - b_s(X_s^\mu, \mu_s)\}]^2 \mathcal{F}_r \, ds \\
&\leq c_2 \|f\|_{\infty}^2 (t-r)^{\varepsilon^2} \|\eta\|_{L^2(\mathcal{F})}.
\end{align*}
\]
Combining this with (1.11), (4.36) and (4.39), and noting that $(s \wedge 1)^2 \leq s$ for $s \geq 0$, we find constants $c_3, c_4 > 0$ such that
\[
\begin{align*}
\mathbb{E}\left[ \{f(X_t^\varepsilon) - f(X_t)\} \frac{1 - R_t^\varepsilon}{\varepsilon} \right] &\leq \left( \mathbb{E}\left[ \mathbb{E}\left[ f(X_t^\varepsilon) - f(X_t) \mid \mathcal{F}_r \right]^2 \right] \right)^{\frac{1}{2}} \left( \mathbb{E}\left[ \left( \frac{1 - R_t^\varepsilon}{\varepsilon} \right)^2 \right] \right)^{\frac{1}{2}} \\
&\leq c_4 \|f\|_{\infty} \left( \frac{\mathbb{E}[|X_t^\varepsilon - X_t^\mu|]}{\sqrt{t-r}} \right)^{\frac{1}{2}} + c_4 \|f\|_{\infty} \varepsilon \\
&\leq c_5 \sqrt{T} \|f\|_{\infty} \left( \frac{\varepsilon}{t-r} \right)^{\frac{1}{2}}, \quad \varepsilon \in (0, 1], \ t \in [0, T].
\end{align*}
\]
Combining this with (4.38) we obtain
\[
\lim_{\varepsilon \downarrow 0} \int_{\mathcal{L}_1^\varepsilon} I_2(f) \leq \lim_{r \uparrow t} \lim_{\varepsilon \downarrow 0} \left\{ \mathbb{E}\left[ \{f(X_t^\varepsilon) - f(X_t)\} \frac{1 - R_t^\varepsilon}{\varepsilon} \right] + 2\|f\|_{\infty} \mathbb{E}\left[ \frac{R_t^\varepsilon - R_t^\varepsilon}{\varepsilon} \right] \right\} = 0.
\]
Therefore, (4.38) holds.

In remains to prove (4.29). By Jensen’s inequality, it suffices to prove for $p \in (1, 2]$. By (4.28), we have
\[
\begin{align*}
|\Gamma_\eta(f(X_t^\mu))| &\leq \mathbb{E}(|J_1(X_0^\mu, \eta)|) + |J_2|,
\end{align*}
\]
(4.42)
where
\[ \begin{align*}
J_1(x, v) &:= \mathbb{E} \left[ f(X_t^{x,v}) \int_0^t \beta_s \langle \zeta_s(X_{s}^{x,v}) \nabla_v X_{s}^{x,v}, dW_s \rangle \right], \\
J_2 &:= \mathbb{E} \left[ f(X_t^x) \int_0^t \zeta_s(X_{s}^{x}) \mathbb{E} \left[ (D^{L^0}b^1_s(z, P_s^{\mu_s}) X_{s}^{\mu_s}, \nabla_{\eta} X_{s}^{\mu_s}) | z = X_{s}^{\mu_s}, dW_s \right] \right].
\end{align*} \]

Taking \( \beta_s = \frac{s}{t} \), by \( \|\zeta\|_\infty < \infty \), (1.15) and Hölder’s inequality, we find constants \( c_1, c_2 > 0 \) such that
\[ |J_1(x, v)| \leq \frac{c_1}{t} \left( P_t^{\mu} |f|^p(x) \right)^{\frac{1}{p}} \left( \mathbb{E} \left[ \left( \int_0^t |\nabla_v X_{s}^{x,v}|^2 ds \right)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} \leq \frac{c_2 |x|}{\sqrt{t}} \left( P_t^{\mu} |f|^p(x) \right)^{\frac{1}{p}}, \quad t \in (0, T]. \]

Combining this with (4.32) and \( P_t^{\mu} |f|^p(X_0^{\mu}) = \mathbb{E}[|f(X_t^{\mu})|^p | \mathcal{F}_0] \), we derive
\[ \begin{align*}
\mathbb{E}[|J_1(X_0^{\mu}, \eta)|] &\leq \frac{c_2}{\sqrt{t}} \mathbb{E} \left[ |\eta| (P_t^{\mu} |f|^p(X_0^{\mu}))^{\frac{1}{p}} \right] \\
&\leq \frac{c_2 \|\eta\|_{L^k(\mathcal{P})}}{\sqrt{t}} \left( \mathbb{E}[|f(X_0^{\mu})|^p | \mathcal{F}_0] \right)^{\frac{1}{p}}, \quad t \in (0, T].
\end{align*} \]

On the other hand, by (H), Hölder’s inequality and (4.16) for \( j = k \), we find constants \( c_3, c_4 > 0 \) such that
\[ I_s(z) := |\zeta_s(X_{s}^{\mu}) \mathbb{E} \left[ (D^{L^0}b^1_s(z, P_s^{\mu_s}) X_{s}^{\mu_s}, \nabla_{\eta} X_{s}^{\mu_s}) \right]| \leq c_3 \|\nabla_{\eta} X_{s}^{\mu}\|_{L^k(\mathcal{P})} \leq c_4 \|\eta\|_{L^k(\mathcal{P})}, \]
so that
\[ |J_2| \leq \mathbb{E} \left[ \left( \mathbb{E}[|f(X_t^{\mu})|^p | \mathcal{F}_0] \right)^{\frac{1}{p}} \left( \mathbb{E} \left[ \int_0^t I_s(X_{s}^{\mu})^2 ds \right]^{\frac{p}{2}} \right)^{\frac{1}{p}} \right] \leq c_4 \sqrt{t} \|\eta\|_{L^k(\mathcal{P})} \mathbb{E} \left[ \left( \mathbb{E}[|f(X_t^{\mu})|^p | \mathcal{F}_0] \right)^{\frac{1}{p}} \right]. \]

This and (4.43) imply (4.29).

\[ \square \]

### 4.4 Proof of Theorem 1.3

Simply denote \( X_t = X_t^{\mu} \), and for any \( \varepsilon \in [0, 1] \) let \( X_t^{\varepsilon} \) solve (1.11) with \( X_0^{\varepsilon} = X_0 + \varepsilon \phi(X_0) \), \( \mu^{\varepsilon} := \mathcal{L}_{X_0 + \varepsilon \phi(X_0)} \) and \( \mu_t^{\varepsilon} := P_t^{\mu^{\varepsilon}} = \mathcal{L}_{X_t^{\varepsilon}} \). We have
\[ P_t f(\mu \circ (id + \varepsilon \phi^{-1})) = \mathbb{E}[f(X_t^{\varepsilon})]. \]

It suffices to prove
\[ \begin{align*}
\lim_{\varepsilon \downarrow 0} \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \mathbb{E}[f(X_t^{\varepsilon}) - f(X_t)] \right|_{\varepsilon} - D^f_{\phi} f(\mu) \right| = 0.
\end{align*} \]

30
By applying (4.28) with $\beta_s = \frac{\alpha}{t}$ for $(\mu^r, \phi(X_0))$ replacing $(\mu, \eta)$, we obtain
\[
\frac{d}{dr} \mathbb{E}[f(X_t^r)] := \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}[f(X_t^{r+\varepsilon}) - f(X_t^r)]}{\varepsilon} = \Gamma_{\phi(X_0)}(f(X_t^{\mu^r}))
\]
\[
= \frac{1}{t} \int_{\mathbb{R}^d} \mathbb{E} \left[ f(X_t^{r,.x+r\phi(x)}) \int_0^t \left< \zeta_s(X_{s}^{r,.x+r\phi(x)}) \nabla \phi(x) X_s^{r,.x+r\phi(x)} \right., dW_s \right] \mu(dx)
\]
\[
+ \mathbb{E} \left[ f(X_t^{r}) \int_0^t \left< \zeta_s(X_s^{r}) \mathbb{E} \left[ (D^L b_s^{(1)}(z, \mu^r_s)(X_s^{r}), \nabla \phi(X_0) X_s^{r}) \right] \right|_{z=X_t^r}, dW_s \right].
\]
Combining this with (1.16) for $\beta_s = \frac{\alpha}{t}$, we derive
\[
\sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \frac{\mathbb{E}[f(X_t^r) - f(X_t)]}{\varepsilon} - D^f \phi \mu^r \right| = \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \frac{1}{\varepsilon} \int_0^\varepsilon \left\{ \frac{d}{dr} \mathbb{E}[F(X_t^r)] - D^f \phi \mu^r \right\} dr \right| \leq c \int_0^\varepsilon \sum_{i=1}^4 \alpha_i(r) dr
\]
for some constant $c > 0$, where letting
\[
F_\phi(r, x) := \int_0^t \left< \zeta_s(X_s^{r,.x+r\phi(x)}) \nabla \phi(x) X_s^{r,.x+r\phi(x)} \right., dW_s \right>, \quad r \in [0, 1], x \in \mathbb{R}^d,
\]
\[
G_\phi(r) := \int_0^t \left< \zeta_s(X_s^{r}) \mathbb{E} \left[ (D^L b_s^{(1)}(z, \mu^r_s)(X_s^{r}), \nabla \phi(X_0) X_s^{r}) \right] \right|_{z=X_t^r}, dW_s \right>, \quad r \in [0, 1],
\]
we set
\[
\alpha_1(r) := \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \int_{\mathbb{R}^d} \mathbb{E} \left[ f(X_t^{r,.x+r\phi(x)}) \right. \left\{ F_\phi(r, x) - F_\phi(0, x) \right\} \right] \mu(dx) \right|,
\]
\[
\alpha_2(r) := \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \int_{\mathbb{R}^d} \mathbb{E} \left[ \left\{ f(X_t^{r,.x+r\phi(x)}) - f(X_t^{0,.x}) \right\} F_\phi(0, x) \right] \mu(dx) \right|,
\]
\[
\alpha_3(r) := \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \mathbb{E} \left[ \left\{ f(X_t^r) \right\} G_\phi(r) - G_\phi(0) \right] \right|,
\]
\[
\alpha_4(r) := \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \left| \mathbb{E} \left[ \left\{ f(X_t^r) - f(X_t) \right\} G_\phi(0) \right] \right|.
\]
Since $\|f\|_\infty < \infty$, by (H), (1.14) and (1.15), we conclude that $\{\alpha_i\}_{1 \leq i \leq 4}$ are bounded on $[0, 1]$. So, (4.44) follows if
\[
\lim_{r \downarrow 0} \alpha_i(r) = 0, \quad 1 \leq i \leq 4.
\]
To prove these limits, we need the following two lemmas.

**Lemma 4.4.** Assume (H). For any $j \geq 1$ there exists a constant $c > 0$ such that for any $\mu \in \mathcal{P}_k$ and $\phi \in T_{\mu, k}$ with $\|\phi\|_{L^k(\mu)} \leq 1$,
\[
\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^{r,.x+r\phi(x)} - X_t^{0,.x}|^j \right] \leq c r^j (1 + |\phi(x)|^j), \quad r \in [0, 1].
\]
Proof. By (1.15), we have
\[ \mathbb{E}[|X_t^{\mu,x+r\phi(x)} - X_t^{\mu,x}|^j] \leq cr^j|\phi(x)|^j, \quad r \in [0, 1], x \in \mathbb{R}^d. \]

Combining this with \( \mathbb{W}_k(\mu^r, \mu) \leq r\|\phi\|_{L^k(\mu)} \leq r \), we need only to prove
\[ \sup_{x \in \mathbb{R}^d} \mathbb{E}[|X_t^{\mu,x} - X_t^{\nu,x}|^j] \leq c\mathbb{W}_k(\mu, \nu)^j, \quad \mu, \nu \in \mathcal{P}_k \]
for some constant \( c > 0 \), where \( X_t^{\nu,x} \) solves (1.16) for \( \nu_t := \nu^* \nu \) replacing \( \mu_t := \nu^* \mu \). Let \( u \) solve (3.6) such that (2.6) holds. Let \( \Theta_t = id + u_t \) and
\[ Y_t^{\mu,x} := \Theta_t(X_t^{\mu,x}), \quad Y_t^{\nu,x} := \Theta_t(X_t^{\nu,x}), \quad t \in [0, T]. \]

By Itô’s formula we obtain
\[
d(Y_t^{\mu,x} - Y_t^{\nu,x}) = \left\{ (\nabla \Theta_t) \sigma_t \right\} (X_t^{\mu,x}) - (\nabla \Theta_t) \sigma_t \} (X_t^{\nu,x}) \, dW_t \right\} \left\{ b_t^{(2)}(X_t^{\mu,x}, \mu) + \lambda u_t(X_t^{\mu,x}) - b_t^{(2)}(X_t^{\nu,x}, \nu) - \lambda u_t(X_t^{\nu,x}) \right\} dt.
\]

By (H), (2.6), Lemma 2.1 in [21] and Itô’s formula, for any \( j \geq 1 \) we find a constant \( c > 0 \) such that
\[
|Y_t^{\mu,x} - Y_t^{\nu,x}|^2j \leq c \int_0^t |Y_s^{\mu,x} - Y_s^{\nu,x}|^2j \sum_{i=0}^l \left\{ 1 + \mathcal{M}_f^2(x, X_s^{\mu,x}) + \mathcal{M}_f^2(x, X_s^{\nu,x}) \right\} ds
\]
\[ + c \int_0^t \mathbb{W}_k(\mu_s, \nu_s)^{2j} ds + M_t, \quad t \in [0, T] \]
holds for some local martingale \( M_t \) with \( M_0 = 0 \). Since \( \mathbb{W}_k(\mu_s, \nu_s) \leq c\mathbb{W}_k(\mu, \nu) \) due to (1.13), (4.45) follows from the stochastic Gronwall inequality, Lemma 2.1 in [21], and Khasminskii’s estimate (2.9) for \( X_s^{\mu,x} \) and \( X_s^{\nu,x} \) replacing \( X_s \).

**Lemma 4.5.** Assume (H) and (1.18). For any \( j \geq 1 \) there exist a constant \( c > 0 \) and a positive function \( \varepsilon(\cdot) \) on \( [0, 1] \) with \( \varepsilon(r) \downarrow 0 \) as \( r \downarrow 0 \), such that for any \( \phi \in T_{\mu,k} \) with \( \|\phi\|_{L^k(\mu)} \leq 1 \) and \( r \in [0, 1] \),
\[
(4.46) \quad \sup_{|v| \leq 1} \mathbb{E} \left[ \sup_{t \in [0, T]} |\nabla_v X_t^{\mu,x+r\phi(x)} - \nabla_v X_t^{\mu,x}|^j \right] \leq \min \left\{ c, \varepsilon(r)(1 + |\phi(x)|^j) \right\}, \quad x \in \mathbb{R}^d,
\]
\[
(4.47) \quad \mathbb{E} \left[ \sup_{t \in [0, T]} |\nabla\phi(X_0) X_t^{\mu,x} - \nabla\phi(X_0) X_t^{\mu,x}|^j \right] \leq |\phi(X_0)|^j \min \left\{ c, \varepsilon(r)(1 + |\phi(X_0)|^j) \right\}.
\]
Proof. We only prove (4.46) since (4.47) can be proved in the same way by using (4.16) and (1.10) replacing (1.15) and Lemma 4.4 respectively. We simply denote
\[ X_t := X_t^{x}, \quad X_t^{r,x} := X_t^{x+r\phi(x)}, \quad \tilde{v}_t := \nabla_v X_t^{x}, \quad \tilde{v}_t^{r} := \nabla_v X_t^{r,x}. \]

Let \( u \) solve (3.6) such that (2.6) holds. We may also assume that \( u \) satisfies (2.13) as explained before. Let \( \Theta_t = id + u_t \) and denote
\[ Y_t^x := \Theta_t(X_t^x), \quad Y_t^{r,x} := \Theta_t(X_t^{r,x}), \quad v_t := (\nabla \Theta_t(X_t^x))^{-1} \tilde{v}_t, \quad v_t^r := (\nabla \Theta_t(X_t^{r,x}))^{-1} \tilde{v}_t^r. \]

By (1.15) and (2.6), to prove (4.46) it suffices to find \( \varepsilon(r) \downarrow 0 \) as \( r \downarrow 0 \) such that
\[ \sup_{|v| \leq 1} \mathbb{E}\left[ \sup_{t \in [0,T]} |v_t^r - v_t|^j \right] \leq \varepsilon(r)(1 + |\phi(x)|^j), \quad r \in [0,1], \quad x \in \mathbb{R}^d. \]

By Jensen’s inequality, we only need to prove for \( j \geq 4 \).

To calculate \( v_t \) and \( v_t^r \), for any \( \varepsilon \in [0,1] \) we let
\[ Y_t^{r,x}(\varepsilon) := \Theta_t(X_t^{x+r\phi(x)+\varepsilon v}), \quad Y_t^{x}(\varepsilon) := \Theta_t(X_t^{x+\varepsilon v}). \]

Then the argument leading to (2.12) implies that
\[ v_t = \lim_{\varepsilon \downarrow 0} \frac{Y_t^{x}(\varepsilon) - Y_t^{x}(\varepsilon)}{\varepsilon}, \quad v_t^r = \lim_{\varepsilon \downarrow 0} \frac{Y_t^{r,x}(\varepsilon) - Y_t^{r,x}(\varepsilon)}{\varepsilon}. \]

By (3.6) and Itô’s formula, we obtain
\[
\begin{align*}
dY_t^x(\varepsilon) &= \left\{ b_t^{(1)}(X_t^{x+\varepsilon v}, \mu_t) + \lambda u_t(X_t^{x+\varepsilon v}) \right\} dt + \{(\nabla \Theta_t)\sigma_t\}(X_t^{x+\varepsilon v})dW_t, \\
Y_0^x(\varepsilon) &= x + \varepsilon x,
\end{align*}
\]
\[
\begin{align*}
dY_t^{r,x}(\varepsilon) &= \left\{ b_t^{(1)}(X_t^{r,x+\varepsilon v}, \mu_t) + \lambda u_t(X_t^{x+\varepsilon v}) + \nabla b_t^{(1)}(X_t^{r,x+\varepsilon v}, \mu_t)u_t(X_t^{r,x+\varepsilon v}) \right\} dt \\
&\quad + \{(\nabla \Theta_t)\sigma_t\}(X_t^{r,x+\varepsilon v})dW_t, \quad Y_0^{r,x}(\varepsilon) = x + r\phi(x) + \varepsilon x.
\end{align*}
\]

Combining this with (4.48) and (4.51), we conclude that \( v_t \) and \( v_t^r \) solves the SDEs
\[
\begin{align*}
dv_t &= \left\{ \nabla_{\tilde{v}_t} b_t^{(1)}(X_t^x, \mu_t) + \lambda \nabla_{\tilde{v}_t} u_t(X_t^x) \right\} dt + \nabla_{\tilde{v}_t} \{(\nabla \Theta_t)\sigma_t\}(X_t^x)dW_t, \\
v_0 &= (\nabla \Theta_0(x))^{-1} v,
\end{align*}
\]
\[
\begin{align*}
dv_t^r &= \left\{ \nabla_{\tilde{v}_t} b_t^{(1)}(X_t^{r,x}, \mu_t) + \lambda \nabla_{\tilde{v}_t} u_t(X_t^{r,x}) + \nabla_{\tilde{v}_t} \{(\nabla \Theta_t)\sigma_t\}(X_t^{r,x})dW_t, \\
v_0^r &= (\nabla \Theta_0(x + r\phi(x))^{-1} v.
\end{align*}
\]
Therefore, by (4.49),
\[ z^r_t := v^r_t - v_t, \quad t \in [0, T] \]
solves the SDE
\[
dz^r_t = \left\{ \nabla_t (\nabla_t (X^r_t) - 1) b_t^{(1)} (X^r_t, \mu_t) + \mu u_t (X^r_t) + \nabla_t (\nabla_t (X^r_t) - 1) u_t (X^r_t) \right\} dt
+ \nabla_t (\nabla_t (X^r_t) - 1) \{ (\nabla_t \Theta_t) \sigma_t \} (X^r_t) dW_t - \eta^r_t dt - \xi^r_t dW_t,
\]
where
\[
\eta^r_t := \nabla_t (\Theta_t (X^r_t) - 1) b_t^{(1)} (X^r_t, \mu_t) - \nabla_t (\Theta_t (X^r_t) - 1) b_t^{(1)} (X^r_t, \mu_t) - \lambda \nabla_t (\Theta_t (X^r_t) - 1) u_t (X^r_t),
\]
\[
\xi^r_t := \nabla_t (\Theta_t (X^r_t) - 1) \{ (\nabla_t \Theta_t) \sigma_t \} (X^r_t) - \nabla_t (\Theta_t (X^r_t) - 1) \{ (\nabla_t \Theta_t) \sigma_t \} (X^r_t), \quad t \in [0, T].
\]
By (2.2), (H) and Lemma 2.1 in [21], we find a constant \( c_1 > 0 \) such that
\[
|\eta^r_t| + |\xi^r_t| \leq c_1 |v^r_t| \left\{ \| b_t^{(1)} (X^r_t, \mu_t) - b_t^{(1)} (X^r_t, \mu_t) \| + |X^r_t - X^r_t| \sum_{i=0}^t \left[ 1 + M f_i (t, X^r_t) + M f_i (t, X^r_t) \right] \right\}.
\]
By the boundedness of \( \nabla b_t^{(1)} \) and (1.18), we have
\[
\| \nabla b_t^{(1)} (X^r_t, \mu_t) - \nabla b_t^{(1)} (X^r_t, \mu_t) \| \leq n \left\{ \| X^r_t - X^r_t \| + W_k (\mu_t, \mu_t) \right\}^{\frac{1}{2}} + s_n, \quad n \geq 1,
\]
where for \( \varphi (r) := \sup_{|x - x'| = W_k (\mu, \nu) \leq 1} \| b_t^{(1)} (x, \mu) - b_t^{(1)} (x', \nu) \| \),
\[
s_n := \sup_{r \geq 0} \left\{ \varphi (r) - nr \right\} \downarrow 0 \quad \text{as } n \uparrow \infty.
\]
Using the notation (4.48), by combining this with Lemma 4.1 (1.13) and (2.19) for the processes \( X^r_t \) and \( X^r_t \), for any \( j \geq 4 \) we find positive function \( \varepsilon_1 \) with \( \varepsilon_1 (r) \downarrow 0 \) as \( r \downarrow 0 \) such that for \( \| \phi \|_{L^2 (\mu)} \leq 1 \),
\[
\mathbb{E} \left[ \left( \int_0^T \left\{ |\eta^r_s|^2 + |\xi^r_s|^2 \right\} ds \right) \right] \leq \varepsilon_1 (r) (1 + |\phi (x)|^2), \quad r \in [0, 1], x \in \mathbb{R}^d.
\]
Combining this with (4.52), (H) and BDG’s inequality, we find a constant \( c_1 > 0 \) such that
\[
\gamma^r_t := \sup_{s \in [0, t]} |z^r_s|, \quad t \in [0, T]
\]
satisfies
\[
\mathbb{E} [\gamma^r_t] \leq \varepsilon_2 (r) + c_1 \int_0^t \left\{ \gamma^r_s + \gamma^r_s \right\} |\eta^r_s|^2 + \gamma^r_s |\xi^r_s|^2 \} ds, \quad t \in [0, T],
\]
34
where by \((2.13)\) \(\varepsilon_2(r) := \mathbb{E}[|z_0|^2]\rightarrow 0\) as \(r \rightarrow 0\). Since \(st \leq s^{\frac{n}{2}} + t^n\) holds for \(s, t \geq 0\) and \(n \geq 1\), by taking \(n = \frac{4}{2} + j\) for \(j \geq 4\) respectively, we obtain

\[
\int_0^t \left\{ \gamma_s^{j-1} |\eta_s|^2 + \gamma_s^{j-2} |\xi_s|^2 \right\} ds \\
\leq \left( \int_0^t |z_s^{r(j-1)}|^2 ds \right)^{\frac{1}{2}} \left( \int_0^t |\eta_s|^2 ds \right)^{\frac{1}{2}} + \left( \int_0^t |z_s^{r(j-2)}|^2 ds \right)^{\frac{1}{2}} \left( \int_0^t |\xi_s|^2 ds \right)^{\frac{1}{2}} \\
\leq \left( \int_0^t |z_s^{r(j-1)}|^2 ds \right)^{\frac{1}{2j-1}} + \left( \int_0^t |z_s^{r(j-2)}|^2 ds \right)^{\frac{1}{2j-2}} + \alpha^r,
\]

where

\[
\alpha^r := \left( \int_0^T |\eta_s|^2 ds \right)^{\frac{j}{j}} + \left( \int_0^T |\xi_s|^2 ds \right)^{\frac{j}{j}}.
\]

So, there exists a constant \(c_2 > 0\) such that

\[
c_1 \int_0^t \left\{ \gamma_s^{j-1} |\eta_s|^2 + \gamma_s^{j-2} |\xi_s|^2 \right\} ds \\
\leq c_1 |\gamma_r^j|^{\frac{j(j-2)}{2j-1}} \left( \int_0^t |z_s^{r(j-1)}|^2 ds \right)^{\frac{1}{2j-1}} + c_1 |\gamma_r^j|^{\frac{j(j-4)}{2j-2}} \left( \int_0^t |z_s^{r(j-2)}|^2 ds \right)^{\frac{1}{2j-2}} + c_1 \alpha^r \\
\leq \frac{1}{2} |\gamma_r^j|^{\frac{j}{j}} + c_2 \int_0^t |\gamma_r^j|^2 ds + c_1 \alpha^r.
\]

Since \((1.15)\) implies \(\mathbb{E}[|\gamma_r^j|^2] < \infty\), combining this with \((4.54), (4.55)\) and applying Gronwall’s inequality, we prove \((4.50)\) for some positive function \(\varepsilon\) with \(\varepsilon(r) \downarrow 0\) as \(r \downarrow 0\).

We are now ready to prove \(\alpha_i(r) \rightarrow 0\) as \(r \rightarrow 0\) for \(i = 1, 2, 3, 4\) respectively and hence finish the proof of Theorem 1.3.

(a) \(\alpha_1(r) \rightarrow 0\). As in \((4.53)\), by \((H)\) and \((1.18)\) we find a sequence of positive numbers \(s_n \downarrow 0\) as \(n \uparrow \infty\) such that

\[
\sup_{s \in [0,T]} \|\zeta_s(x) - \zeta_s(y)\|^2 \leq n |x - y|^{2(k-1)} + s_n, \quad n \geq 1,
\]

\[
\sup_{s \in [0,T]} \|D^Lb^{(1)}_s(x,\mu)(y) - D^Lb^{(1)}_s(x',\mu)(y')\| \\
\leq n \{ |x - x'| + |y - y'| + W_k(\mu,\nu) \} \frac{1}{n^2} + s_n, \quad n \geq 1.
\]

By \((4.56)\), Lemma 1.4, Lemma 4.3 and \((4.17)\), we find a constant \(c_1 > 0\) such that for any \(\phi \in T_{\mu,k}\) with \(\|\phi\|_{L^1(\mu)} \leq 1\),

\[
\mathbb{E}[|F_\phi(r, x) - F_\phi(0, x)|] \\
\leq \mathbb{E} \left( \int_0^t |\zeta_s(X_{s}^{\mu,x+r\phi(x)} - \zeta_s(X_{s}^{\mu,x})|^2 \cdot |\nabla_\phi(x) X_{s}^{\mu,x+r\phi(x)}|^2 ds \right)^{\frac{1}{2}}.
\]
Therefore, there exists a constant $c$.

Integrating with respect to $\mu(dx)$ and letting first $r \to 0$ then $n \to \infty$, we prove $\alpha_1(r) \to 0$ as $r \to 0$.

(b) $\alpha_2(r) + \alpha_4(r) \to 0$. Let

$$R_\theta := \int_0^\theta \langle \zeta_s(X^\mu,x) \nabla_{\phi(x)}X^\mu,x, dW_s \rangle, \quad \theta \in [0,t].$$

By (1.15), we find a constant $c_1 > 0$ such that

$$\mathbb{E}[|R_t - R_\theta|] \leq c_1 \sqrt{t - \theta} |\phi(x)|, \quad \theta \in [0,t], x \in \mathbb{R}^d. \quad (4.58)$$

On the other hand, as in (4.39) and (4.41), we find a constant $c_2 > 0$ such that for $\|\phi\|_{L^k(\mu)} \leq 1$,

$$\begin{align*}
&\| \mathbb{E}[f(X^\mu,x + r\phi(x)) - f(X^\mu,x) | \mathcal{F}_0] \| \\
&\leq |(P_{\mu,x}^\theta f)(X^\mu,x + r\phi(x)) - (P_{\mu,x}^\theta f)(X^\mu,x)| \\
&\leq |(P_{\mu,x}^\theta f)(X^\mu,x + r\phi(x)) - (P_{\mu,x}^\theta f)(X^\mu,x)| + \| (P_{\mu,x}^\theta f)(X^\mu,x) - (P_{\mu,x}^\theta f)(X^\mu,x) \| \\
&\leq c_2 \| f \| \left[ 1 + \frac{|X^\mu,x + r\phi(x) - X^\mu,x|}{\sqrt{t - \theta}} \right].
\end{align*}$$

Combining this with (4.58) and Lemma 4.4, we find constants $c_3, c_4 > 0$ such that

$$\begin{align*}
&\| \mathbb{E}[\{f(X^\mu,x + r\phi(x)) - f(X^\mu,x)\} F_\phi(0,x)] \| \\
&\leq 2 \| f \| \mathbb{E}[|R_t - R_\theta|] + \mathbb{E}[\mathbb{E}(f(X^\mu,x + r\phi(x)) - f(X^\mu,x) | \mathcal{F}_\theta) R_\theta] \\
&\leq c_3 \| f \| \left[ \sqrt{t - \theta} |\phi(x)| + \left( r + \frac{\min \{1, r(1 + |\phi(x)|)\}}{\sqrt{t - \theta}} \right) (\mathbb{E}[|R_\theta|^2])^{\frac{1}{2}} \right] \\
&\leq c_4 \| f \| \left[ \sqrt{t - \theta} |\phi(x)| + r |\phi(x)| + \frac{n r^{k-1}(1 + |\phi(x)|)^{k-1} + s_n}{\sqrt{t - \theta}} |\phi(x)| \right],
\end{align*}$$

where

$$s_n := \sup_{s > 0} \{ s \wedge 1 - ns^{k-1} \} \downarrow 0 \text{ as } n \uparrow \infty.$$ 

Therefore, there exists a constant $c_5 > 0$ such that

$$\alpha_2(r) \leq \| f \| \left[ c_5 \sqrt{t - \theta} + \frac{n r^{k-1} + s_n}{\sqrt{t - \theta}} + r \right], \quad \theta \in (0,t).$$
By letting first \( r \to 0 \) then \( n \to \infty \) and finally \( \theta \to t \), we prove \( \alpha_2(r) \to 0 \) as \( r \to 0 \).

The proof of \( \alpha_4(r) \to 0 \) is completely similar.

(c) \( \alpha_3(r) \to 0 \). Write

\[
E[|G(\phi)(r) - G(\phi)(0)|] \leq \varepsilon(r, \phi) + \|\phi\|_{\infty}E[\alpha(\mu^r, \mu^s)],
\]

where

\[
\varepsilon(r, \phi) = E \left( \int_0^t (\phi(X_s^r) - \phi(X_s)) \right)^2 \left( E|\phi(x_0)|^k s \right)^{\frac{2}{k}} ds \right)^{\frac{1}{2}}.
\]

By (4.16) for \( j = k \), we obtain

\[
\sup_{r \in [0,1]} E[|\nabla \phi(x_0)X^r_s|^k] \leq c, \quad \|\phi\|_{L^k(\mu)} \leq 1
\]

for some constant \( c > 0 \), so that by (1.11) and (4.56), we find constants \( c_1, c_2 > 0 \) such that

\[
\sup_{\|\phi\|_{L^k(\mu)} \leq 1} E[(\phi(X_s^r) - \phi(X_s)]^{k-1} + s_n] \leq c_2nr^{k-1} + c_1s_n, \quad n \geq 1.
\]

Then \( \sup_{\|\phi\|_{L^k(\mu)} \leq 1} \varepsilon(r, \phi) \to 0 \) as \( r \to 0 \). It remains to prove

\[
\lim_{r \downarrow 0} \sup_{\|\phi\|_{L^k(\mu)} \leq 1} E[\alpha(\mu^r, \mu)] = 0.
\]

By (H), (1.11), (1.13), Lemma 4.5, (4.66) and (4.59), we find constants \( c_3, c_4, c_5 > 0 \) and positive function \( \tilde{\varepsilon}() \) on \([0,1]\) with \( \tilde{\varepsilon}(r) \to 0 \) as \( r \to 0 \), such that when \( \|\phi\|_{L^k(\mu)} \leq 1 \),

\[
E[(\phi(X_s^r) - \phi(X_s)]^{k-1} + s_n] \leq c_3(\|\phi(x_0)\|_{L^k}^k + \|\phi(x_0)\|_{L^k}^k)^{\frac{1}{k}} + \left(\tilde{\varepsilon} + c_4(\|n^{k_r}\{ |z_s - y_s| + |X_s^r - X_s| + r_s\} + s_n)^{\frac{1}{k}} \right)^{\frac{1}{k}} + \left(\tilde{\varepsilon} + c_5\{ |n^{k_r}\{ |z_s - y_s| + |X_s^r - X_s| + r_s\} + s_n\} \right)^{\frac{1}{k}}
\]

\[
\leq \tilde{\varepsilon}(r) + c_5\|n^{k_r}\{ |z_s - y_s| + |X_s^r - X_s| + r_s\} + s_n\}, \quad n \geq 1.
\]

Combining this with (1.11) we find a constant \( c_6 > 0 \) such that

\[
\sup_{\|\phi\|_{L^k(\mu)} \leq 1} E[\alpha(\mu^r, \mu)] \leq c_6\{\tilde{\varepsilon}(r) + nr^{\frac{1}{k}} + s_n\}, \quad n \geq 1.
\]

By letting first \( r \to 0 \) then \( n \to \infty \) we prove (4.60).
References

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

[2] L. Ambrosio, N. Gigli and G. Savare, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lect. in Math., ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[3] J. Bao, P. Ren, F.-Y. Wang, Bismut formulas for Lions derivative of McKean-Vlasov SDEs with memory, J. Diff. Equat. 282(2021), 285–329.

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

[5] V. I. Bogachev, N. V. Krylov, M. Röckner, S.V. Shaposhnikov, Fokker-Planck-Kolmogorov equations, American Math. Soc. 2015.

[6] J. Bao, P. Ren, F.-Y. Wang, Bismut formulas for Lions derivative of McKean-Vlasov SDEs with memory, J. Diff. Equat. 282(2021), 285–329.

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

[8] R. Carmona, F. Delarue, Probabilistic Theory of Mean Field Games with Applications I, Springer 2019.

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

[10] S. Fang, Introduction to Malliavin Calculus, Tsinghua University Press and Springer, Beijing 2004.

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

[12] X. Huang, P. Ren, F.-Y. Wang, Distribution dependent stochastic differential equations, Front. Math. China 16(2021), 257–301.

[13] X. Huang, Y. Song, F.-Y. Wang, Bismut formula for intrinsic/Lions derivatives of distribution dependent SDEs with singular coefficients, arXiv:2105.11116.

[14] X. Huang, F.-Y. Wang, McKean-Vlasov SDEs with drifts discontinuous under Wasserstein distance, Disc. Cont. Dyn. Syst. Ser. A. 4(2021), 1667–1679.

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

[16] F. Otto, The geometry of dissipative evolution equations: The porous medium equation, Comm. Partial Diff. Equat. 26(2001), 101–174.

[17] A.-S. Sznitman, Topics in propagations of chaos, Lecture notes in Math. Vol. 1464, pp. 165–251, Springer, Berlin, 1991.

[18] F.-Y. Wang, Harnack Inequalities and Applications for Stochastic Partial Differential Equations, Springer, 2013, Berlin.
[19] F.-Y. Wang, Distribution dependent reflecting stochastic differential equations, [arXiv:2106.12737](https://arxiv.org/abs/2106.12737).

[20] F.-Y. Wang, Exponential ergodicity for singular reflecting McKean-Vlasov SDEs, [arXiv:2108.03908](https://arxiv.org/abs/2108.03908).

[21] P. Xia, L. Xie, X. Zhang, G. Zhao, $L^q(L^p)$-theory of stochastic differential equations, Stoch. Proc. Appl. 130(2020), 5188–5211.

[22] L. Xie, X. Zhang, Ergodicity of stochastic differential equations with jumps and singular coefficients, Ann. Inst. Henri Poincaré Probab. Stat. 56(2020), 175–229.

[23] C. Yuan, S.-Q. Zhang, A study on Zvonkin’s transformation for stochastic differential equations with singular drift and related applications, J. Diff. Equat. 297(2021), 277–319.