Ergodicity and Feynman-Kac Formula for Space-Distribution Valued Diffusion Processes *

Panpan Ren[c),d)], Michael Röckner[b)], Feng-Yu Wang[a),c)]

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

[b)] Department of Mathematics, Bielefeld University, D-33501 Bielefeld, Germany

[c)] Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK

[d)] Mathematical Institute, University of Oxford, Woodstock Road, OX2 6GG, UK

Panpan.ren@maths.ox.ac.uk; roeckner@math.uni-bielefeld.de; wangfy@tju.edu.cn

May 17, 2019

Abstract

Let $\mathscr{P}_2$ be the space of probability measures μ on $\mathbb{R}^d$ with $\mu(|\cdot|^2) < \infty$. Consider the following time-dependent second order differential operator on $\mathbb{R}^d \times \mathscr{P}_2$:

$$\begin{aligned}\mathbf{L}_t f(x,\mu) := &\frac{1}{2}\big\langle \bar{a}(t,x,\mu), \nabla^2 f(x,\mu)\big\rangle_{HS} + \big\langle \bar{b}(t,x,\mu), \nabla f(x,\mu)\big\rangle \\ &+ \int_{\mathbb{R}^d} \Big[\frac{1}{2}\big\langle a(t,y,\mu), \nabla\{(Df(x,\mu))(\cdot)\}(y)\big\rangle_{HS} + \big\langle b(t,y,\mu), (Df(x,\mu))(y)\big\rangle\Big]\mu(\mathrm{d}y),\end{aligned}$$

where $t \geq 0$, ∇ is the gradient operator in $x \in \mathbb{R}^d$, D is the intrinsic derivative in $\mu \in \mathscr{P}_2$, introduced by Albeverio, Kondratiev and the second named author in 1996. Furthermore,

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

are measurable with a and $\bar{a}$ non-negative definite. We investigate the existence, uniqueness and exponential ergodicity of the diffusion process generated by $\mathbf{L}_t$, and use the diffusion process to solve the following PDE on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$:

$$(\partial_t + \mathbf{L}_t)u + \mathbf{V}u + \mathbf{f} = 0,$$

where $\mathbf{V}$ and $\mathbf{f}$ are functions on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$. Furthermore, the relation of the intrinsic derivative D (which maps functions on $\mathscr{P}_2$ to a section in a measurable tangent bundle over $\mathscr{P}_2$) with the Lions derivative is clarified.

AMS subject Classification: 60J60, 58J65.
Keywords: Markov process, ergodicity, Feynman-Kac formula, intrinsic derivative, L-derivative.

*Supported in part by NNSFC (11771326, 11831014) and the DFG through the CRC 1283.

1 Introduction

Let $\mathscr{P}_2$ be the space of all probability measures μ on $\mathbb{R}^d$ such that

$$\|\mu\|_2 := \left(\int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x)\right)^{\frac{1}{2}} < \infty.$$

It is well known that $\mathscr{P}_2$ is a Polish space under the Wasserstein distance

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

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

In this paper, we investigate the diffusion process generated by the time dependent differential operator $\mathbf{L}_t := \mathbf{L}_t^{(1)} + \mathbf{L}_t^{(2)}$ on $\mathbb{R}^d \times \mathscr{P}_2$:

$$\begin{aligned}
\mathbf{L}_t^{(1)} f(x,\mu) &:= \frac{1}{2}\big\langle \bar{a}(t,x,\mu), \nabla^2 f(x,\mu)\big\rangle_{HS} + \big\langle \bar{b}(t,x,\mu), \nabla f(x,\mu)\big\rangle,\\
\mathbf{L}_t^{(2)} f(x,\mu) &:= \int_{\mathbb{R}^d} \Big[\frac{1}{2}\big\langle a(t,y,\mu), \nabla\{(Df(x,\mu))(\cdot)\}(y)\big\rangle_{HS} + \big\langle b(t,y,\mu), (Df(x,\mu))(y)\big\rangle\Big]\mu(\mathrm{d}y),
\end{aligned}$$

where $t \geq 0$, ∇ is the gradient operator on $\mathbb{R}^d$, D is the intrinsic derivative on $\mathscr{P}_2$,

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

are measurable, $a = \sigma\sigma^*, \bar{a} = \bar{\sigma}\bar{\sigma}^*$ for some measurable maps

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

and

$$\langle \bar{a}, \nabla^2 f\rangle_{HS} := \sum_{i,j=1}^d \bar{a}_{ij}\partial_i\partial_j f, \quad \langle a, \nabla\{Df\}\rangle_{HS} := \sum_{i,j=1}^d a_{ij}\partial_j\{Df\}_i.$$

For readers' convenience, we recall the intrinsic derivative which was first introduced in [1] on the configuration space over manifolds, and present some classes of reference functions which will be used latter on.

Definition 1.1. Let f be a real function on $\mathscr{P}_2$, and let $\mathrm{Id} : \mathbb{R}^d \to \mathbb{R}^d$ be the identity map.

(1) We call f intrinsicaly differentiable at $\mu \in \mathscr{P}_2$, if

$$L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu) \ni \phi \mapsto D_\phi f(\mu) := \lim_{\varepsilon\downarrow 0} \frac{f(\mu\circ\{\mathrm{Id}+\varepsilon\phi\}^{-1}) - f(\mu)}{\varepsilon}$$

is a well defined bounded linear functional. In this case, the intrinsic derivative is the unique element $Df(\mu) \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$ such that

$$\langle Df(\mu), \phi\rangle_{L^2(\mu)} = D_\phi f(\mu), \quad \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu).$$

f is called intrinsically differentiable if it is intrinsically differentiable at any $\mu \in \mathscr{P}_2$.

(2) If f is intrinsically differentiable with $Df(\mu)(y)$ having a jointly continuous version in $(\mu, y) \in \mathscr{P}_2 \times \mathbb{R}^d$, we denote $f \in C^{(1,0)}(\mathscr{P}_2)$; if moreover Df is bounded, we write $f \in C_b^{(1,0)}(\mathscr{P}_2)$.

(3) If $f \in C^{(1,0)}(\mathscr{P}_2)$ such that $Df(\mu)(y)$ is differentiable in y with $\nabla\{Df(\mu)\}(y)$ jointly continuous in $(\mu, y) \in \mathscr{P}_2 \times \mathbb{R}^d$, we denote $f \in C^{(1,1)}(\mathscr{P}_2)$, and write $f \in C_b^{(1,1)}(\mathscr{P}_2)$ if furthermore Df and ∇Df are bounded.

(4) We write $f \in C_{b,Lip_2}^{(1,1)}(\mathscr{P}_2)$ if $f \in C_b^{(1,1)}(\mathscr{P}_2)$ and $Df(\mu)(y), \nabla\{Df(\mu)\}(y)$ are Lipschitz continuous in $(\mu, y) \in \mathscr{P}_2 \times \mathbb{R}^d$.

Remark 1.1.

(*i*) $(\mathscr{P}_2, (L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu))_{\mu \in \mathscr{P}_2})$ can be considered as a measurable "tangent" bundle of the "manifold" $\mathscr{P}_2$ (see [1] and also [18, 19, 2, 3, 20] for related cases/constructions). So, Df above is a section in this bundle.

(*ii*) The present class $C_b^{(1,1)}(\mathscr{P}_2)$ is larger than that defined in [7, 16]: in addition to conditions in Definition 1.1(3) for $f \in C_b^{1,1}(\mathscr{P}_2)$, it was also assumed in [7, 16] that $Df(\mu)(y)$ is Lipschtiz continuous in $(\mu, y) \in \mathscr{P}_2 \times \mathbb{R}^d$. With this additional condition, the class is denoted in [11] by $C_{b,Lip}^{(1,1)}(\mathscr{P}_2)$.

It is clear that f is intrinsically differentiable if and only if for any $\mu \in \mathscr{P}_2$, the map

$$L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu) \ni \phi \mapsto f(\mu \circ \{\mathrm{Id} + \phi\}^{-1}) \tag{1.1}$$

is Gâteaux differentiable at $\phi = 0$. So, the intrinsic differentiability is weaker than the L-differentiability introduced in P.-L. Lions's lectures [8], which, as explained in [21], is equivalent to the Fréchet differentiability of the map in (1.1) which is stronger than Gâteaux differentiability. On the other hand, since a Gâteaux differentiable function with continuous derivative is Fréchet differentiable, a function in $C_b^{(1,0)}(\mathscr{P}_2)$ must be L-differentiable.

Example 1.1. Consider the following class of cylindrical functions

$$\mathscr{F}C_b^2(\mathscr{P}_2) := \big\{F(\mu) := g(\mu(h_1), \cdots, \mu(h_n)) : \ n \geq 1, g \in C_b^1(\mathbb{R}^n), h_i \in C_b^2(\mathbb{R}^d)\big\}. \tag{1.2}$$

It is easy to see that such a function is in the class $C_b^{(1,1)}(\mathscr{P}_2)$ with

$$DF(\mu) = \sum_{i=1}^n (\partial_i g)(\mu(h_1), \cdots, \mu(h_n)) \nabla h_i \in C_b^1(\mathbb{R}^d \to \mathbb{R}^d). \tag{1.3}$$

The remainder of the paper is organized as follows. In Section 2, we characterize the existence and uniqueness of the $\mathbf{L}_t$-diffusion process by using the corresponding SDE on $\mathbb{R}^d \times \mathscr{P}_2$, see (2.2) below. In Section 3, we investigate the exponential ergodicity of the

$\mathbf{L}$-diffusion process where $\mathbf{L}_t = \mathbf{L}$ does not depend on t. Finally, in Section 4 we use this diffusion process to solve the following PDE on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$ for a fixed $T > 0$:

$$\partial_t u(t,x,\mu) + \mathbf{L_t} u(t,\cdot,\cdot)(x,\mu) + (\mathbf{V}u)(t,x,\mu) + \mathbf{f}(t,x,\mu) = 0, \tag{1.4}$$

where $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2$, and $\mathbf{V}, \mathbf{f}$ are measurable functions on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$. When $\mathbf{V} = 0$ and $(\sigma, b) = (\bar{\sigma}, \bar{b})$ are time independent, this PDE has been studied in [7, 10, 11, 13, 16], where in [13, 16] the case with jumps was also considered. We will extend their results for more general coefficients and non-trivial $\mathbf{V}$, such that (1.4) is a Schrödinger type equation on $\mathbb{R}^d \times \mathscr{P}_2$.

2 Existence and uniqueness of the $\mathbf{L}_t$-diffusions

Recall that a Markov transition kernel on $\mathbb{R}^d \times \mathscr{P}_2$ is a family of probability measures

$$\big\{\mathbf{P}_{s,t}(x,\mu;\cdot) : 0 \le s \le t, (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2\big\}$$

on $\mathbb{R}^d \times \mathscr{P}_2$ such that

(C_1) $\mathbf{P}_{s,t}(\cdot; A)$ is measurable in (x,μ) for any $0 \le s \le t$ and measurable set $A \subset \mathbb{R}^d \times \mathscr{P}_2$; and $\mathbf{P}_{s,s}(x,\mu;\cdot) = \delta_{(x,\mu)}$, the Dirac measure at (x,μ), for any $s \ge 0$ and $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$.

(C_2) For any $0 \le s < r < t$ and $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$, the following Chapman-Kolmogorov equation holds:

$$\mathbf{P}_{s,t}(x,\mu;\cdot) = \int_{\mathbb{R}^d \times \mathscr{P}_2} \mathbf{P}_{r,t}(y,\nu;\cdot)\mathbf{P}_{s,r}(x,\mu;\mathrm{d}y,\mathrm{d}\nu).$$

A Markov transition kernel on $\mathbb{R}^d \times \mathscr{P}_2$ determines a family of Markov operators $\{\mathbf{P}_{s,t} : 0 \le s \le t\}$ on $\mathscr{B}_b(\mathbb{R}^d \times \mathscr{P}_2)$, the Banach space of bounded measurable functions on $\mathbb{R}^d \times \mathscr{P}_2$:

$$\mathbf{P}_{s,t} f(x,\mu) := \int_{\mathbb{R}^d \times \mathscr{P}_2} f(\xi)\, \mathbf{P}_{s,t}(x,\mu;\mathrm{d}\xi), \quad (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2, f \in \mathscr{B}_b(\mathbb{R}^d \times \mathscr{P}_2).$$

Conditions (C_1) and (C_2) are equivalent to $\mathbf{P}_{s,s} f = f$ and the semigroup property

$$\mathbf{P}_{s,t} = \mathbf{P}_{s,r}\mathbf{P}_{r,t}, \quad 0 \le s \le r \le t.$$

To define the Markov process generated by $\mathbf{L}_t$, we consider the following class of cylindrical functions on $\mathbb{R}^d \times \mathscr{P}_2$:

$$\mathscr{C} := \Big\{(x,\mu) \mapsto g\big(h_0(x), \mu(h_1), \cdots, \mu(h_n)\big) :\ n \ge 1, g \in C_b^\infty(\mathbb{R}^{n+1}), \{h_i\}_{0 \le i \le n} \subset C_0^\infty(\mathbb{R}^d)\Big\}.$$

We will use the following assumption.

(A) For any $\mu_\cdot \in C([0,\infty) \to \mathscr{P}_2)$, there exists an increasing function $K : [0,\infty) \to (0,\infty)$ such that

$$|b(t,x,\mu_t)| + \|\sigma(t,x,\mu_t)\| + |\bar{b}(t,x,\mu_t)| + \|\bar{\sigma}(t,x,\mu_t)\| \le K(t)(1+|x|), \quad t \ge 0, x \in \mathbb{R}^d.$$

Definition 2.1. We say that $\mathbf{L}_t$ generates the Markov process with Markov transition kernel $\{\mathbf{P}_{s,t}(x,\mu;\cdot) : (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2, 0 \le s \le t\}$, if the associated Markov semigroup $\mathbf{P}_{s,t}$ satisfies the following Fokker-Planck-Kolmogorov equation (also called Kolmogorov forward equation)

$$\partial_t \mathbf{P}_{s,t} f(x,\mu) = \mathbf{P}_{s,t}(\mathbf{L}_t f)(x,\mu), \quad 0 \le s \le t, (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2, f \in \mathscr{C}, \tag{2.1}$$

in the sense (see [6]) that for any $(s,x,\mu) \in [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2$, $t \ge s$ and $f \in \mathscr{C}$, we have

$$\int_s^t \mathrm{d}r \int_{\mathbb{R}^d \times \mathscr{P}_2} |\mathbf{L}_r f(\xi)| \mathbf{P}_{s,r}(x,\mu;\mathrm{d}\xi) < \infty$$

and

$$\mathbf{P}_{s,t} f(x,\mu) = f(x,\mu) + \int_s^t \mathrm{d}r \int_{\mathbb{R}^d \times \mathscr{P}_2} \mathbf{L}_r f(\xi)\, \mathbf{P}_{s,r}(x,\mu;\mathrm{d}\xi).$$

The Markov process generated by $\mathbf{L}_t$ is called unique, if there is at most one Markov transition kernel such that (2.1) holds.

According to Kolmogorov's theorem, for any $(s,x,\mu) \in [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2$, there exists a unique probability measure $\mathbb{P}_s^{x,\mu}$ on $\Omega_s := \{\mathbb{R}^d \times \mathscr{P}_2\}^{[s,\infty)}$ equipped with the product σ-field, such that the coordinate process

$$X_t(\omega) := \omega_t, \quad t \ge s, \omega \in \Omega_s$$

is a Markov process with

$$\mathbb{P}_s^{x,\mu}(X_t \in \cdot) = \mathbf{P}_{s,t}(x,\mu;\cdot), \quad t \ge s.$$

If the measure $\mathbb{P}_s^{x,\mu}$ is supported on the continuous path $C([s,\infty) \to \mathbb{R}^d \times \mathscr{P}_2)$ for any $(s,x,\mu) \in [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2$, we call the associated Markov process a diffusion process.

2.1 Main result

For any $t \ge 0$ and $\mu \in \mathscr{P}_2$, let

$$L_{t,\mu} f(x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t,x,\mu) \partial_{x_i} \partial_{x_j} f(x) + \sum_{i=1}^d b_i(t,x,\mu) \partial_{x_i} f(x).$$

We will identify an $\mathbf{L}_t$-diffusion process with a weak solution to the following SDE for $(X_t, \mu_t) \in \mathbb{R}^d \times \mathscr{P}_2$:

$$\begin{cases} \partial_t \mu_t = L^*_{t,\mu_t} \mu_t, \\ \mathrm{d}X_t = \bar{b}(t, X_t, \mu_t)\mathrm{d}t + \bar{\sigma}(t, X_t, \mu_t)\mathrm{d}W_t, \end{cases} \tag{2.2}$$

where W_t is the m-dimensional Brownian motion on a complete filtered probability space $(\Omega, \{\mathscr{F}_t\}_{t \ge 0}, \mathbb{P})$. We first introduce the notion of weak well-posedness of (2.2).

Definition 2.2. (1) $\mu_\cdot \in C([s,\infty) \to \mathscr{P}_2)$ is called a solution to

$$\partial_t \mu_t = L^*_{t,\mu_t}\mu_t, \quad \mu_s = \mu, \ t \ge s, \tag{2.3}$$

if

$$\int_s^T \mathrm{d}t \int_{\{|x|\le n\}} \big\{|b(t,x,\mu_t)| + \|\sigma(t,x,\mu_t)\|^2\big\}\mu_t(\mathrm{d}x) < \infty, \quad T \in (s,\infty), n \ge 1, \tag{2.4}$$

and

$$\int_{\mathbb{R}^d} h \mathrm{d}\mu_t = \int_{\mathbb{R}^d} h \mathrm{d}\mu + \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} L_{r,\mu_r} h \, \mathrm{d}\mu_r, \quad h \in C_0^\infty(\mathbb{R}^d), t \ge s. \tag{2.5}$$

(2) If for any $s \ge 0$ and $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$, (2.3) has a unique solution $(\mu_t)_{t\ge s}$ with $\mu_s = \mu$, and the SDE

$$\mathrm{d}X^{x,\mu}_{s,t} = \bar{b}(t, X^{x,\mu}_{s,t}, \mu_t)\mathrm{d}t + \bar{\sigma}(t, X^{x,\mu}_{s,t}, \mu_t)\mathrm{d}W_t, \quad t \ge s, X^{x,\mu}_{s,s} = x \tag{2.6}$$

has a unique weak solution, then we call (2.2) weakly well-posed.

When (2.3) has a unique solution, we denote the solution by $\mu_t = P^*_{s,t}\mu$ to emphasize dependence on the initial value μ at time s, and to link it with the distribution dependent SDE (DDSDE for short)

$$\mathrm{d}X_t = b(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \tag{2.7}$$

where $\mathscr{L}_{X_t}$ is the law of X_t under the reference probability space, see Proposition 2.1 below.

Definition 2.3. The DDSDE (2.7) is said to have a weak solution, if for any $s \ge 0$ and $\mu \in \mathscr{P}_2$, there exist an m-dimensional Brownian motion W_t on a standard filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\ge0}, \mathbb{P})$, and an $\mathscr{F}_t$-adapted continuous process $(X_t)_{t\ge s}$ on $\mathbb{R}^d$, such that $\mathscr{L}_{X_s} = \mu$, $\mathscr{L}_{X_\cdot} \in C([s,\infty) \to \mathscr{P}_2)$ and $\mathbb{P}$-a.s.

$$X_t = X_s + \int_s^t b(r, X_r, \mathscr{L}_{X_r})\mathrm{d}r + \int_s^t \sigma(r, X_r, \mathscr{L}_{X_r})\mathrm{d}W_r, \quad t \ge s.$$

If the laws of any two weak solutions with common initial distribution from time s coincide, we call the weak solution unique.

When (2.7) has a unique weak solution, it is called weakly well-posed. In this case, we denote

$$\mathscr{L}_{X_t} = P^*_{s,t}\mu, \quad t \ge s, \mu \in \mathscr{P}_2$$

for a weak solution $(X_t)_{t\ge s}$ with $\mathscr{L}_{X_s} = \mu$. By the uniqueness, $(P^*_{s,t})_{0\le s\le t}$ satisfies the Chapman-Kolmogorov equation:

$$P^*_{s,t} = P^*_{r,t}P^*_{s,r}\mu, \quad 0 \le s \le r \le t, \mu \in \mathscr{P}_2. \tag{2.8}$$

The following result links (2.3) and (2.7).

Proposition 2.1. *Assume* **(A)**.

(*i*) *Let* $s \geq 0$*,* $\mu \in \mathscr{P}_2$*, and* X_s $\mathscr{F}_s$*-measurable. Then there exists a solution* $(\mu_t)_{t\geq s}$ *to* (2.3) *with initial condition* μ *at* $t = s$*, if and only if there exists a weak solution* $(X_t)_{t\geq s}$ *to the SDE in* (2.7) *with initial condition* X_s*.*

(*ii*) (2.7) *is weakly well-posed if and only if* (2.3) *has a unique solution* $(\mu_t)_{t\geq s}$ *for any* $s \geq 0$ *and* $\mu \in \mathscr{P}_2$*, and in this case* $P_{s,t}^*\mu = \mu_t, t \geq s \geq 0$*.*

Proof. (*i*) For fixed $s \geq 0$ and $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_2$, let $(\mu_t)_{t\geq s}$ solve (2.3). As explained in [5, Section 2] (see also [4]) that there exits a weak solution $(X_t)_{t\geq s}$ of (2.7) for an m-dimensional Brownian motion W_t on $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ (a standard filtered probability space), such that $\mathscr{L}_{X_s} = \mu$ and $\mathscr{L}_{X_t} = \mu_t$, for $t \geq s$.

On the other hand, let $(\mu_t = \mathscr{L}_{X_t})_{t\geq s}$ for a weak solution $(X_t)_{t\geq s}$ to (2.7) with $\mathscr{L}_{X_s} = \mu$. By Itô's formula, for any $h \in C_0^\infty(\mathbb{R}^d)$ we have

$$\mathrm{d}h(X_t) = L_{t,\mu_t}h(X_t)\mathrm{d}t + \langle \nabla h(X_t), \sigma(t, X_t, \mu_t)\mathrm{d}W_t\rangle, \quad t \geq s.$$

This together with condition **(A)** implies

$$\begin{aligned}\mu_t(h) = \mathbb{E}h(X_t) &= \mu(h) + \int_s^t \mathbb{E}[L_{r,\mu_r}h(X_r)]\mathrm{d}r \\ &= \mu(h) + \int_s^t \mu_r(L_{r,\mu_r}h)\mathrm{d}r, \quad t \geq s, h \in C_0^\infty(\mathbb{R}^d).\end{aligned}$$

Therefore, μ_t solves (2.3).

(*ii*) This obviously follows from (*i*).

□

We are now ready to state the main result in this section.

Theorem 2.2. *Assume* **(A)**. *Then the following assertions are equivalent:*

(1) $\mathbf{L}_t$ *generates a unique diffusion process on* $\mathbb{R}^d \times \mathscr{P}_2$ *with Markov transition kernel satisfying*

$$\int_{\mathbb{R}^d} \mathbf{P}_{s,t}(x, \mu; \mathbb{R}^d \times \cdot)\mu(\mathrm{d}x) \text{ is a Dirac measure on } \mathscr{P}_2, \quad t \geq s \geq 0, \mu \in \mathscr{P}_2. \tag{2.9}$$

(2) (2.2) *is weakly well-posed.*

Moreover, when (2) *holds, the unique Markov transition kernel is given by*

$$\mathbf{P}_{s,t}(x, \mu; \cdot) = \mathscr{L}_{X_{s,t}^{x,\mu}} \times \delta_{\mu_t} = \mathscr{L}_{X_{s,t}^{x,\mu}} \times \delta_{P_{s,t}^*\mu}, \quad 0 \leq s \leq t, (x, \mu) \in \mathbb{R}^d \times \mathscr{P}_2.$$

Combining Theorem 2.2 and Proposition 2.1 with known results on the well-posedness of DDSDEs, we obtain explicit sufficient conditions for the existence and uniqueness of the $\mathbf{L}_t$-diffusion process. For instance, in addition to **(A)** we assume

(B) There exists an increasing function $K : [0, \infty) \to [0, \infty)$ such that for all $t \geq 0$,

$$\begin{aligned} &2\langle b(t,x,\mu) - b(t,y,\nu), x-y\rangle + \|\sigma(t,x,\mu) - \sigma(t,y,\nu)\|_{HS}^2 \\ &\leq K(t)\big(|x-y| + \mathbb{W}_2(\mu,\nu)\big), \quad x,y \in \mathbb{R}^d, \mu,\nu \in \mathscr{P}_2. \end{aligned}$$

According to [25, Theorem 2.1], (2.7) is well-posed under assumptions **(A)** and **(B)**, so that (2.3) has a unique solution. If **(B)** also holds for $(\bar{b}, \bar{\sigma})$ replacing (b, σ), then (2.6) is well-posed. Therefore, by Theorem 2.2, $\mathbf{L}_t$ generates a unique diffusion process on $\mathbb{R}^d \times \mathscr{P}_2$.

When $\bar{\sigma}$ is invertible with

$$\|\nabla \bar{\sigma}(t, \cdot, \mu_t)(x)\| + |\bar{b}(t, x, \mu_t)| \in L^q_{loc}([0, \infty) \to L^p(\mathbb{R}^d)) \tag{2.10}$$

for some $p, q > 1$ satisfying $\frac{d}{p} + \frac{2}{q} < 1$, [26] (see also [15] for constant $\bar{\sigma}$) ensures the well-posedness of (2.6). Moreover, if $b(t, x, \mu)$ and $\sigma(t, x, \mu)$ are Lipschitz continuous in μ uniformly in $(t, x) \in [0, T] \times \mathbb{R}^d$ for any $T > 0$, and that (2.10) holds for (b, σ) replacing $(\bar{b}, \bar{\sigma})$, then as shown in [14] that (2.7) is well-posed. Therefore, again by Theorem 2.2, these conditions and **(A)** imply that $\mathbf{L}_t$ generates a unique diffusion process on $\mathbb{R}^d \times \mathscr{P}_2$. In this case, the coefficients b and $\bar{b}$ may be discontinuous in the times-space variable (t, x).

2.2 Proof of Theorem 2.2

We first present the following lemma, which follows from [10, Theorem 3.3] or [12, Proposition A.6] under the stronger condition

$$\int_s^T \mathbb{E}\big(|\alpha_t|^2 + \|\beta_t\|^4\big)\mathrm{d}t < \infty, \quad T \in (s, \infty). \tag{2.11}$$

Lemma 2.3. *Let $\alpha : [s, \infty) \to \mathbb{R}^d$ and $\beta : [s, \infty) \to \mathbb{R}^{d\otimes m}$ be progressively measurable with*

$$\mathbb{E}\bigg(\int_s^T |\alpha_t|\mathrm{d}t\bigg)^2 + \mathbb{E}\int_s^T \|\beta_t\|^2\mathrm{d}t < \infty, \quad T \in (s, \infty). \tag{2.12}$$

For $X_s \in L^2(\Omega \to \mathscr{F}_s; \mathbb{P})$, let $\mu_t = \mathscr{L}_{X_t}$ for

$$X_t := X_s + \int_s^t \alpha_r \mathrm{d}r + \int_s^t \beta_r \mathrm{d}W_r, \quad t \geq s.$$

Then $\mu_\cdot \in C([s, \infty) \to \mathscr{P}_2)$ and for any $f \in C_b^{(1,1)}(\mathscr{P}_2)$,

$$\frac{\mathrm{d}f(\mu_t)}{\mathrm{d}t} = \mathbb{E}\int_{\mathbb{R}^d} \Big[\frac{1}{2}\big\langle \beta_t\beta_t^*, \nabla\{Df(\mu_t)(\cdot)\}(y)\big\rangle_{HS} + \big\langle \alpha_t, (Df(\mu_t))(y)\big\rangle\Big]\mu_t(\mathrm{d}y), \quad t \geq s.$$

Proof. Since $\mu_s \in \mathscr{P}_2$ and (2.12) holds, it is easy to see that $\mu_\cdot \in C([s, \infty) \to \mathscr{P}_2)$. For any $n \geq 1$, let $\alpha_t^n = \alpha_t 1_{\{|\alpha_t| \leq n\}}$, $\beta_t^n = \beta_t 1_{\{\|\beta_t\| \leq n\}}$, and let $\mu_t^n = \mathscr{L}_{X_t^n}$ for

$$X_t^n := X_s + \int_s^t \alpha_r^n \mathrm{d}r + \int_s^t \beta_r^n \mathrm{d}W_r, \quad t \geq s.$$

Then

$$\lim_{n\to\infty}\sup_{t\in[s,T]} W_2(\mu_t^n,\mu_t)=0$$

and (2.11) holds for (α_t^n,β_t^n) replacing (α_t,β_t). So, by [10, Theorem 3.3] or [12, Proposition A.6], we obtain

$$f(\mu_t^n)=f(\mu_s)+\mathbb{E}\int_s^t \mathrm{d}r\int_{\mathbb{R}^d}\Big[\frac{1}{2}\big\langle\beta_r^n\{\beta_r^n\}^*,\nabla\{Df(\mu_r^n)(\cdot)\}(y)\big\rangle_{HS}+\big\langle\alpha_r^n,(Df(\mu_r^n)(y)\big\rangle\Big]\mu_r^n(\mathrm{d}y)$$

for $t\ge s$. Since $Df(\mu)(y)$ and $\nabla\{Df(\mu)\}(y)$ are bounded and continuous in (μ,y), by (2.12) we may apply the dominated convergence theorem with $n\to\infty$ to derive

$$f(\mu_t)=f(\mu_s)+\mathbb{E}\int_s^t \mathrm{d}r\int_{\mathbb{R}^d}\Big[\frac{1}{2}\big\langle\beta_r\beta_r^*,\nabla\{Df(\mu_r)(\cdot)\}(y)\big\rangle_{HS}+\big\langle\alpha_r,(Df(\mu_r)(y)\big\rangle\Big]\mu_r(\mathrm{d}y),\ \ t\ge s.$$

Then the proof is finished. □

Proof of Theorem 2.2. (a) Suppose that (2.2) is weakly well-posed. We construct the desired Markov transition kernel generated by $\mathbf{L}_t$ as follows. By Proposition 2.1, the DDSDE (2.7) is weakly well-posed and $P_{s,t}^*\mu:=\mu_t$ satisfies (2.8), where $(\mu_t)_{t\ge s}$ is the unique solution of (2.3). Moreover, let $(X_{s,t}^{x,\mu})_{t\ge s}$ be a weak solution to (2.6) and denote $P_{s,t}^\mu(x,\cdot)=\mathscr{L}_{X_{s,t}^{x,\mu}}$. By the weak uniqueness we have

$$P_{s,t}^\mu(x,\cdot)=\int_{\mathbb{R}^d}P_{r,t}^\mu(y,\cdot)P_{s,r}^\mu(x,\mathrm{d}y),\quad 0\le s\le r\le t. \tag{2.13}$$

Then for any $0\le s\le t$ and $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$,

$$\mathbf{P}_{s,t}(x,\mu;\mathrm{d}y,\mathrm{d}\nu):=P_{s,t}^\mu(x,\mathrm{d}y)\delta_{P_{s,t}^*\mu}=P_{s,t}^\mu(x,\mathrm{d}y)\delta_{\mu_t} \tag{2.14}$$

gives rise to a Markov transition kernel satisfying (2.9), and due to the continuity of the solution to (2.2), the associated Markov process is continuous. We intend to show that this Markov transition kernel is generated by $\mathbf{L}_t$, equivalently, the associated Markov semigroup satisfies (2.1).

By **(A)** and applying Itô's formula to $|X_{s,t}^{x,\mu}|^2$ in (2.6), we obtain

$$\sup_{r\in[s,t]}\int_{\mathbb{R}^d}|y|^2P_{s,r}^\mu(x,\mathrm{d}y)=\sup_{r\in[s,t]}\mathbb{E}|X_{s,r}^{x,\mu}|^2<\infty. \tag{2.15}$$

Combining this with **(A)** and (2.14), we conclude that for any $f\in\mathscr{C}$,

$$\int_s^t\mathrm{d}r\int_{\mathbb{R}^d\times\mathscr{P}_2}|\mathbf{L}_rf(\xi)|\mathbf{P}_{s,r}(x,\mu;\mathrm{d}\xi)=\int_s^t|\mathbf{L}_rf(X_{s,t}^{x,\mu},\mu_r)|\mathrm{d}r<\infty. \tag{2.16}$$

So, it suffices to show

$$\mathbf{P}_{s,t}f(x,\mu)=f(x,\mu)+\int_s^t\mathbf{P}_{s,r}(\mathbf{L}_rf)(x,\mu)\mathrm{d}s,\quad t\ge s. \tag{2.17}$$

This follows from Lemma 2.3 and Itô's formula for $(X_{s,t}^{x,\mu})_{t\ge s}$. Indeed, sine $\mu_t = \mathscr{L}_{X_t}$ for X_t solving (2.7) with $\mathscr{L}_{X_s} = \mu$, $\mu_\cdot \in C([s,\infty) \to \mathscr{P}_2)$ and condition **(A)** imply (2.12) for $\alpha_t := b(t, X_t, \mu_t)$ and $\beta_t := \sigma(t, X_t, \mu_t)$. By Lemma 2.3, for any $z \in \mathbb{R}^d$ and $t \ge s$, we have

$$\frac{\mathrm{d}f(z,\mu_t)}{\mathrm{d}t} = \mathbf{L}_t^{(2)} f(z,\mu_t), \quad t \ge s. \tag{2.18}$$

On the other hand, by (2.6) and Itô's formula, for any $\nu \in \mathscr{P}_2$ we have

$$\mathrm{d}f(X_{s,t}^{x,\mu},\nu) = \mathbf{L}_t^{(1)} f(X_{s,t}^{x,\mu},\nu)\mathrm{d}t + \big\langle \nabla f(\cdot,\nu)(X_{s,t}^{x,\mu}), \bar\sigma(t, X_{s,t}^{x,\mu},\mu_t)\mathrm{d}W_t\big\rangle, \quad t \ge s.$$

Combining this with (2.18), we obtain

$$\mathrm{d}f(X_{s,t}^{x,\mu},\mu_t) = \mathbf{L}_t f(X_{s,t}^{x,\mu},\mu_t)\mathrm{d}t + \big\langle \nabla f(\cdot,\mu_t)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu},\mu_t)\mathrm{d}W_t\big\rangle, \quad t \ge s. \tag{2.19}$$

It then follows from (2.14) and (2.19) that

$$\begin{aligned}\mathbf{P}_{s,t} f(x,\mu) &= \mathbb{E}[f(X_{s,t}^{x,\mu},\mu_t)] = f(x,\mu) + \mathbb{E}\int_s^t \mathbf{L}_r f(X_{s,r}^{x,\mu},\mu_r)\mathrm{d}r\\ &= f(x,\mu) + \int_s^t \mathbf{P}_{s,r}(\mathbf{L}_r f)(x,\mu_r)\mathrm{d}r, \quad t \ge s.\end{aligned}$$

Therefore, (2.1) holds.

(b) Let $\mathbf{L}_t$ generate a Markov transition kernel

$$\{\mathbf{P}_{s,t}(x,\mu;\cdot) : (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2, 0 \le s \le t\}$$

satisfying (2.9) such that the associated Markov process is a diffusion process. It remains to show that

$$\mu_t(\cdot) := \int_{\mathbb{R}^d \times \mathscr{P}_2} \nu(\cdot)\mathbf{P}_{s,t}(x,\mu;\mathbb{R}^d \times \mathrm{d}\nu)\mu(\mathrm{d}x), \quad t \ge s \tag{2.20}$$

solves (2.3), and

$$\mu_t^x := \mathbf{P}_{s,t}(x,\mu;\cdot \times \{\mu_t\}) = \mathscr{L}_{X_t^{x,\mu}}, \quad t \ge s \tag{2.21}$$

holds for a weak solution $(X_t^{x,\mu})_{t\ge s}$ of (2.6).

Firstly, by (2.9) and (2.20), we have

$$\int_{\mathbb{R}^d} \mathbf{P}_{s,t}(x,\mu;\mathbb{R}^d \times \cdot)\mu(\mathrm{d}x) = \delta_{\mu_t}, \quad t \ge s. \tag{2.22}$$

Next, (2.4) is ensured by assumption **(A)**. For any $h \in C_0^\infty(\mathbb{R}^d)$, let $f(x,\mu) = \mu(h)$. By (1.3) we have

$$\mathbf{L}_r f(z,\nu) = \int_{\mathbb{R}^d} \Big[\frac{1}{2}\sum_{i,j=1}^d a_{ij}(r,y,\nu)\partial_{y_i}\partial_{y_j} h(y) + \sum_{i=1}^d b_i(r,y,\nu)\partial_i h(y)\Big]\nu(\mathrm{d}y)$$

$$= \int_{\mathbb{R}^d} L_{r,\nu} h(y)\,\nu(\mathrm{d}y), \quad r \geq s, (z,\nu) \in \mathbb{R}^d \times \mathscr{P}_2,$$

and due to (2.20),

$$\int_{\mathbb{R}^d} \mathbf{P}_{s,t} f(x,\mu)\mu(\mathrm{d}x) = \int_{\mathbb{R}^d \times \mathscr{P}_2} \nu(h)\mathbf{P}_{s,t}(x,\mu;\mathbb{R}^{\mathrm{d}} \times \mathrm{d}\nu)\mu(\mathrm{d}x) = \mu_t(h), \quad t \geq s.$$

Combining these with (2.1) and (2.22) we obtain

$$\begin{aligned}
\mu_t(h) &= \int_{\mathbb{R}^d} \mathbf{P}_{s,t} f(x,\mu)\mu(\mathrm{d}x) \\
&= \int_{\mathbb{R}^d} f(x,\mu)\mu(\mathrm{d}x) + \int_{\mathbb{R}^d} \mu(\mathrm{d}x) \int_s^t \mathbf{P}_{s,r}(\mathbf{L}_r f)(x,\mu)\mathrm{d}r \\
&= \mu(h) + \int_s^t \mathrm{d}r \int_{\mathbb{R}^d \times \mathscr{P}_2} \nu(L_{r,\nu} h)\mathbf{P}_{s,r}(x,\mu;\mathbb{R}^d \times \mathrm{d}\nu)\mu(\mathrm{d}x) \\
&= \mu(h) + \int_s^t \mu_r(L_{r,\mu_r} h)\mathrm{d}r, \quad t \geq s, h \in C_0^\infty(\mathbb{R}^d).
\end{aligned}$$

Therefore, $(\mu_t)_{t \geq s}$ solves (2.3) with $\mu_s = \mu$.

Finally, for any $h \in C_0^\infty(\mathbb{R}^d)$, let $f(x,\mu) = h(x), (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$. Then (2.22) implies

$$\mathbf{P}_{s,t} f(x,\mu) := \int_{\mathbb{R}^d \times \mathscr{P}_2} f(y,\nu)\mathbf{P}_{s,t}(x,\mu;\mathrm{d}y,\mathrm{d}\nu) = \int_{\mathbb{R}^d} h(y)\mathbf{P}_{s,t}(x,\mu;\mathrm{d}y \times \{\mu_t\}) = \mu_t^x(h).$$

Moreover, $\mathbf{L}_r f(y,\nu) = \bar{L}_{r,\nu} h(y), (y,\nu) \in \mathbb{R}^d \times \mathscr{P}_2$, where

$$\bar{L}_{r,\nu} = \frac{1}{2}\sum_{i,j=1}^d (\bar{\sigma}\bar{\sigma}^*)_{ij}(r,\cdot,\nu)\partial_i\partial_j + \sum_{i=1}^d \bar{b}(r,\cdot,\nu)\partial_i.$$

Combining these with (2.1) gives

$$\begin{aligned}
\mu_t^x(h) &= h(x) + \int_s^t \mathbf{P}_{s,r}(\mathbf{L}_r f)(x,\mu)\mathrm{d}r \\
&= h(x) + \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} \bar{L}_{r,\mu_r} h(y)\mathbf{P}_{s,t}(x,\mu;\mathrm{d}y \times \mathscr{P}_2) \\
&= h(x) + \int_s^t \mu_r^x(\bar{L}_{r,\mu_r} h)\mathrm{d}r, \quad t \geq s.
\end{aligned}$$

That is, μ_t^x solves the PDE

$$\partial_t \mu_t^x = \bar{L}_{t,\mu_t}^* \mu_t^x, \quad \mu_s^x = \delta_x, \ t \geq s$$

in the sense of

$$\mu_t^x(h) = h(x) + \int_s^t \mu_r^x(\bar{L}_{r,\mu_r} h)\,\mathrm{d}r, \quad t \geq s, h \in C_0^\infty(\mathbb{R}^d).$$

Then by [24, Theorem 2.5], there exists a solution to the martingale problem of $\bar{L}_{t,\mu_t}$ with initial value x from time s, which, according to a standard result (see [23]), implies that $\mu_t^x = \mathscr{L}_{X_t^{x,\mu}}$ for a weak solution to the SDE (2.6). $\square$

3 Exponential ergodicity of the L-diffusion process

In this part, we assume that $b(t,x,\mu)=b(x,\mu), \bar b(t,x,\mu)=\bar b(x,\mu), \sigma(t,x,\mu)=\sigma(x,\mu)$ and $\bar\sigma(t,x,\mu)=\bar\sigma(x,\mu)$ do not depend on t, and consider the ergodicity of the diffusion processes generated by $\mathbf{L}_t=\mathbf{L}$. Recall that a Markov process is called ergodic, if there exists a unique probability measure μ_0 on the state space such that for any initial distribution the process converges weakly to μ_0 as time goes to infinity.

We will assume **(A)** and that (2.2) is weakly well-posed. So, Theorem 2.2 ensures that $\mathbf{L}$ generates a unique diffusion process satisfying (2.9) with $\mathbf{P}_{s,t}=\mathbf{P}_{t-s}:=P_{0,t-s}$. Let $\mathbf{P}_t$ be the associated Markov semigroup. A probability measure Λ on $\mathbb{R}^d\times\mathscr{P}_2$ is called $\mathbf{P}_t$-invariant if $P_t^*\Lambda=\Lambda$ for $t\ge 0$; i.e.

$$\int_{\mathbb{R}^d\times\mathscr{P}_2}\mathbf{P}_t f\,\mathrm{d}\Lambda=\int_{\mathbb{R}^d\times\mathscr{P}_2} f\,\mathrm{d}\Lambda,\quad t\ge 0, f\in\mathscr{B}_b(\mathbb{R}^d\times\mathscr{P}_2). \tag{3.1}$$

In this case, for $\mu\in\mathscr{P}_2$, let $P_t^*\mu=\mu_t$ solve (2.3) with $s=0$ and $\mu_0=\mu$. As explained in the proof of Theorem 2.2 that $P_t^*\mu$ is the distribution of the unique weak solution to (2.7) with initial distribution μ. Let $X_t^{x,\mu}$ be the (weak) solution of (2.6) with $s=0$. Finally, for any $\mu\in\mathscr{P}_2$, consider the SDE

$$\mathrm{d}Y_t^{x,\mu}=b(Y_t^{x,\mu},\mu)\mathrm{d}t+\sigma(Y_t^{x,\mu},\mu)\mathrm{d}W_t,\quad Y_0^{x,\mu}=x, \tag{3.2}$$

By the Markov property of $Y_t^{x,\mu}$ and noting that (3.2) coincides with (2.7) when the initial value is a random variable with distribution μ, we have

$$P_t^*\mu=\int_{\mathbb{R}^d}\mathscr{L}_{Y_t^{x,\mu}}\mu(\mathrm{d}x). \tag{3.3}$$

The following is a consequence of Theorem 2.2.

Proposition 3.1. *Let $b,\bar b,\sigma$ and $\bar\sigma$ be time-independent satisfying* **(A)** *such that* (2.2) *is weakly well-posed. Let $\mu_0\in\mathscr{P}_2$. Then the following assertions are equivalent:*

(1) $\mathbf{P}_t$ *is ergodic.*

(2) *There exist $\mu_0,\bar\mu_0\in\mathscr{P}_2$ such that for any $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$,*

$$P_t^*\mu\to\mu_0 \text{ and } \mathscr{L}_{X_t^{x,\mu}}\to\bar\mu_0 \text{ weakly as } t\to\infty.$$

Consequently, if $\mathbf{P}_t$ is ergodic then the unique invariant probability measure is given by $\Lambda=\bar\mu_0\times\delta_{\mu_0}$ for some $\mu_0,\bar\mu_0\in\mathscr{P}_2$. In particular, when $(b,\sigma)=(\bar b,\bar\sigma)$, we have $\mu_0=\bar\mu_0$.

Proof. By Theorem 2.2, it is easy to see that (2) implies that $\mathbf{P}_t$ is ergodic with the unique invariant probability measure $\Lambda:=\bar\mu_0\times\delta_{\mu_0}$ for some $\mu_0,\bar\mu_0\in\mathscr{P}_2$.

On the other hand, by Theorem 2.2, (1) implies that there exists a probability measure Λ on $\mathbb{R}^d\times\mathscr{P}_2$ such that for any $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$,

$$\mathscr{L}_{(X_t^{x,\mu},P_t^*\mu)}\to\Lambda \text{ weakly as } t\to\infty,\quad (x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2.$$

This implies that as $t \to \infty$, $P_t^*\mu \to \mu_0$ weakly for some $\mu_0 \in \mathscr{P}_2$. So, μ_0 is P_t^*-invariant. Moreover, $\Lambda(\mathbb{R}^d \times \cdot) = \delta_{\mu_0}$ and

$$\mathscr{L}_{X_t^{x,\mu}} \to \bar{\mu}_0 := \Lambda(\cdot, \{\mu_0\}), \quad (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2. \tag{3.4}$$

In particular, when $(b,\sigma) = (\bar{b}, \bar{\sigma})$, we have $X_t^{x,\mu} = Y_t^{x,\mu}$. Taking $\mu = \mu_0$ in (3.3) and noting that $P_t^*\mu_0 = \mu_0$, we obtain

$$\mu_0 = P_t^*\mu_0 = \int_{\mathbb{R}^d} \mathscr{L}_{Y_t^{x,\mu_0}} \mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d} \mathscr{L}_{X_t^{x,\mu_0}} \mu_0(\mathrm{d}x), \quad t \geq 0.$$

This together with (3.4) implies $\mu_0 = \bar{\mu}_0$ as $t \to \infty$. □

Next, we present below a result on the exponential ergodicity of $\mathbf{P}_t$. Let $\mathbb{W}_2^\rho$ be the L^2-Wasserstein distance induced by the following metric on $\mathbb{R}^d \times \mathscr{P}_2$:

$$\rho((x,\mu),(y,\nu)) := \sqrt{|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2}$$

on $\mathbb{R}^d \times \mathscr{P}_2$; that is, for any two probability measures Λ_1, Λ_2 on $E := \mathbb{R}^d \times \mathscr{P}_2$,

$$\mathbb{W}_2^\rho(\Lambda_1,\Lambda_2)^2 := \inf_{\Pi \in \mathscr{C}(\Lambda_1,\Lambda_2)} \int_{E\times E} \big(|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2\big)\Pi((\mathrm{d}x,\mathrm{d}\mu);(\mathrm{d}y,\mathrm{d}\nu)),$$

where $\mathscr{C}(\Lambda_1,\Lambda_2)$ is the set of all couplings of Λ_1 and Λ_2. Let $\mathbf{P}_t(x,\mu;\cdot) = \mathbf{P}_{0,t}(x,\mu;\cdot)$ be the time-homogenous transition probability kernel of the $\mathbf{L}$-diffusion process, which is the distributions of the $\mathbf{L}$-diffusion (X_t,μ_t) starting at (x,μ).

Theorem 3.2. *Assume* **(A)**. *If there exist constants $\lambda > \kappa \geq 0$ and $\bar{\lambda}, \bar{\kappa} > 0$ such that*

$$2\langle b(x,\mu) - b(y,\nu), x-y\rangle + \|\sigma(x,\mu) - \sigma(y,\nu)\|_{HS}^2 \leq \kappa \mathbb{W}_2(\mu,\nu)^2 - \lambda|x-y|^2 \tag{3.5}$$

and

$$2\langle \bar{b}(x,\mu) - \bar{b}(y,\nu), x-y\rangle + \|\bar{\sigma}(x,\mu) - \bar{\sigma}(y,\nu)\|_{HS}^2 \leq \bar{\kappa} \mathbb{W}_2(\mu,\nu)^2 - \bar{\lambda}|x-y|^2 \tag{3.6}$$

hold for all $x,y \in \mathbb{R}^d$ and $\mu,\nu \in \mathscr{P}_2$, then $\mathbf{L}$ *generates a unique diffusion process satisfying* (2.9)*, and the process is ergodic with unique invariant probability measure $\Lambda = \bar{\mu}_0 \times \delta_{\mu_0}$ for some $\mu_0, \bar{\mu}_0 \in \mathscr{P}_2$, such that for any $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$,*

$$\mathbb{W}_2^\rho(\mathbf{P}_t(x,\mu;\cdot),\Lambda)^2 \leq \mathbb{W}_2(\mu,\mu_0)^2\left(\mathrm{e}^{-(\lambda-\kappa)t} + \frac{\bar{\kappa}(\mathrm{e}^{-(\lambda-\kappa)t} - \mathrm{e}^{-\bar{\lambda}t})}{\kappa + \bar{\lambda} - \lambda}\right) + \mathrm{e}^{-\bar{\lambda}t}\bar{\mu}_0(|x-\cdot|^2), \quad t \geq 0,$$

where when $\kappa + \bar{\lambda} = \lambda$,

$$\frac{\mathrm{e}^{-(\lambda-\kappa)t} - \mathrm{e}^{-\bar{\lambda}t}}{\kappa + \bar{\lambda} - \lambda} := t\mathrm{e}^{-\bar{\lambda}t}, \quad t \geq 0.$$

That is, the $\mathbf{L}$*-diffusion process is exponentially ergodic in $\mathbb{W}_2^\rho$ with rate $\frac{1}{2}\{\bar{\lambda} \wedge (\lambda - \kappa)\} > 0$.*

Proof. Firstly, by [25, Theorem 3.1], **(A)** and (3.5) imply that the DDSDE (2.7) is well-posed and P_t^* has a unique invariant probability measure μ_0 such that

$$\mathbb{W}_2(P_t^*\mu,\mu_0)^2 \le \mathrm{e}^{-(\lambda-\kappa)t}\mathbb{W}_2(\mu,\mu_0)^2,\quad t\ge 0, \mu\in\mathscr{P}_2. \tag{3.7}$$

Moreover, it is standard that **(A)** and (3.6) imply the well-posedness of the SDE (2.6). So, by Theorem 2.2, for any initial value $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$, $\mathbf{L}$ generates a unique diffusion process $(X_t^{x,\mu}, P_t^*\mu)$ satisfying 2.9.

Next, let $\bar{X}_t$ solve the SDE

$$\mathrm{d}\bar{X}_t = \bar{b}(\bar{X}_t,\mu_0)\mathrm{d}t + \bar{\sigma}(X_t,\mu_0)\mathrm{d}W_t.$$

By (3.6), there exist constants $c_1, c_2>0$ such that

$$2\langle \bar{b}(x,\mu_0), x\rangle + \|\bar{\sigma}(x,\mu_0)\|_{HS}^2 \le c_1 - c_2|x|^2,\quad x\in\mathbb{R}^d.$$

It is standard that this implies the existence of an invariant probability measure $\bar{\mu}_0$ of the diffusion process $\bar{X}_t$. In the following we take $\bar{X}_0$ such that $\mathscr{L}_{\bar{X}_0}=\bar{\mu}_0$, so that

$$\mathscr{L}_{\bar{X}_t} = \bar{\mu}_0,\quad t\ge 0. \tag{3.8}$$

Now, by (3.6) and Itô's formula, we obtain

$$\begin{aligned}
&\mathrm{d}|X_t^{x,\mu}-\bar{X}_t|^2\\
&=\Big\{2\big\langle \bar{b}(X_t^{x,\mu},P_t^*\mu)-\bar{b}(\bar{X}_t,\mu_0), X_t^{x,\mu}-\bar{X}_t\big\rangle + \|\bar{\sigma}(X_t^{x,\mu},P_t^*\mu)-\bar{\sigma}(\bar{X}_t,\mu_0)\|_{HS}^2\Big\}\mathrm{d}t+\mathrm{d}M_t\\
&\le \big\{\bar{\kappa}\mathbb{W}_2(P_t^*\mu,\mu_0)^2 - \bar{\lambda}|X_t^{x,\mu}-\bar{X}_t|^2\big\}\mathrm{d}t + \mathrm{d}M_t,\quad t\ge 0,
\end{aligned}$$

where

$$\mathrm{d}M_t := 2\big\langle(\bar{\sigma}(X_t^{x,\mu},P_t^*\mu)-\bar{\sigma}(\bar{X}_t,\mu_0))\mathrm{d}W_t, X_t^{x,\mu}-\bar{X}_t\big\rangle$$

is a martingale. Combining this with (3.7) and (3.8), we get

$$\begin{aligned}
\mathbb{E}|X_t^{x,\mu}-\bar{X}_t|^2 &\le \mathrm{e}^{-\bar{\lambda}t}\bar{\mu}_0(|x-\cdot|^2) + \bar{\kappa}\mathbb{W}_2(\mu,\mu_0)^2\int_0^t \mathrm{e}^{-\bar{\lambda}(t-s)-(\lambda-\kappa)s}\mathrm{d}s\\
&= \mathrm{e}^{-\bar{\lambda}t}\bar{\mu}_0(|x-\cdot|^2) + \frac{\bar{\kappa}\mathbb{W}_2(\mu,\mu_0)^2}{\kappa+\bar{\lambda}-\lambda}\big(\mathrm{e}^{-(\lambda-\kappa)t}-\mathrm{e}^{-\bar{\lambda}t}\big),
\end{aligned} \tag{3.9}$$

where when $\kappa+\bar{\lambda}=\lambda$,

$$\frac{\mathrm{e}^{-(\lambda-\kappa)t}-\mathrm{e}^{-\bar{\lambda}t}}{\kappa+\bar{\lambda}-\lambda} := \mathrm{e}^{-\bar{\lambda}t}\lim_{s\to\bar{\lambda}}\frac{\mathrm{e}^{(\bar{\lambda}-s)t}-1}{\bar{\lambda}-s} = t\mathrm{e}^{-\bar{\lambda}t}.$$

This together with (3.7), (3.8) and $\Lambda:=\bar{\mu}_0\times\delta_{\mu_0}$ gives

$$\begin{aligned}
&\mathbb{W}_2^\rho(\mathbf{P}_t(x,\mu;\cdot),\Lambda)^2 \le \mathbb{W}_2(P_t^*\mu,\mu_0)^2 + \mathbb{E}|X_t^{x,\mu}-\bar{X}_t|^2\\
&\le \mathbb{W}_2(\mu,\mu_0)^2\bigg(\mathrm{e}^{-(\lambda-\kappa)t} + \frac{\bar{\kappa}\big(\mathrm{e}^{-(\lambda-\kappa)t}-\mathrm{e}^{-\bar{\lambda}t}\big)}{\kappa+\bar{\lambda}-\lambda}\bigg) + \mathrm{e}^{-\bar{\lambda}t}\bar{\mu}_0(|x-\cdot|^2),\quad t\ge 0.
\end{aligned}$$

As a consequence, Λ is the unique invariant probability measure for the $\mathbf{L}$-diffusion process. $\square$

4 Feynman-Kac formula

4.1 A known result

We first recall the result [16, Theorem 9.2] for the PDE (1.4) with $\mathbf{V} = 0$ (see Theorem 4.1 below), which generalizes (with jump) and improves (under weaker conditions) the corresponding results in [7, 10, 11]. To this end, we introduce the class $\mathscr{C}_b^{0,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$.

Definition 4.1. Let $T > 0$ and $f \in C([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$.

(1) We write $f \in C_b^{0,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$, if $f(t,\cdot,\mu) \in C^2(\mathbb{R}^d)$, and $f(t,x,\cdot) \in C_b^2(\mathscr{P}_2)$ such that all derivatives

$$\nabla f(t,x,\mu),\ \nabla^2 f(t,x,\mu),\ Df(t,x,\mu)(y),\ \nabla\{Df(t,x,\mu)(\cdot)\}(y)$$

are bounded and jointly continuous in $(t,x,\mu,y) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2 \times \mathbb{R}^d$, and Lipchitz continuous in (x,y,μ) uniformly in $t \in [0,T]$.

(2) If If $f \in C_b^{0,2,2}([0,T] \times \mathscr{P}_2))$ but $f(t,x,\mu)$ does not depend on t, we write $f \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$.

(3) If $f \in C_b^{0,2,2}([0,T] \times \mathscr{P}_2))$ and $\partial_t f(t,x,\mu)$ is bounded and continuous in $(t,x,\mu) \in [0,T] \times \mathscr{P}_2)$, we denote $f \in C_b^{1,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$.

A subclass of $C_b^{0,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$ is the following space of cylindrical functions:

$$\begin{aligned}&\mathscr{F}C_b^2([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)\\ &:= \Big\{(t,x,\mu) \mapsto g(t,x,\mu(h_1),\cdots,\mu(h_n)) :\ n \geq 1, g \in C_b^2([0,T] \times \mathbb{R}^{n+d}), h_i \in C_b^2(\mathbb{R}^d)\Big\}.\end{aligned}$$

The following result is a special case of [16, Theorem 9.2] where a jump term and a more general function $\mathbf{f}$ are included. For simplicity, we denote $b, \sigma \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$ if their components are in this class.

Theorem 4.1 ([16]). *Let $T > 0, \mathbf{f} \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$, and let $b = \bar{b}, \sigma = \bar{\sigma} \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$. Then for any $\Phi \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$, the PDE*

$$(\partial_s \mathbf{L}_s) u(s,\cdot,\cdot)(x,\mu) + \mathbf{f}(x,\mu) = 0, \quad u(T,\cdot,\cdot) = \Phi,\ (s,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2,$$

has a unique solution $u \in C_b^{1,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$, and the solution is given by

$$u(s,x,\mu) = \mathbb{E}\bigg[\Phi(X_{s,T}^{x,\mu}, P_{s,T}^*\mu) + \int_s^T \mathbf{f}(X_{s,r}^{x,\mu}, P_{s,r}^*\mu)\mathrm{d}r\bigg] \tag{4.1}$$

for $(s,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2$, where $(X_{s,t}^{x,\mu})_{t\in[s,T]}$ solves (2.6) with $\mu_t = P_{s,t}^*\mu$.

According to Theorem 2.2, (4.1) is equivalent to

$$u(s,x,\mu) = \mathbf{P}_{T-s}\Phi(x,\mu) + \int_s^T (\mathbf{P}_{r-s}\mathbf{f})(x,\mu)\mathrm{d}r, \quad (s,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2,$$

where $(\mathbf{P}_t)_{t\geq 0}$ is the diffusion semigroup generated by the time independent operator $\mathbf{L}$ with $(\bar{b},\bar{\sigma}) = (b,\sigma)$.

We aim to extend this result to non-trivial $\mathbf{V}$ and with possibly different $(\bar{b},\bar{\sigma})$, and with a slightly weak condition on the time dependent functions $\mathbf{V}$ and $\mathbf{f}$.

4.2 Main result

We will work with the following class $C_b^{0,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$.

Definition 4.2. We write $f\in C_b^{0,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$, if $f(t,x,\mu)$ is continuous in $(t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2$, C^2 in $x\in\mathbb{R}^d$, $C^{(1,1)}$ in $\mu\in\mathscr{P}_2$, such that the derivatives

$$\nabla f(t,x,\mu),\quad \nabla^2 f(t,x,\mu),\quad Df(t,x,\mu)(y),\quad \nabla\{Df(t,x,\mu)(\cdot)\}(y)$$

are bounded and jointly continuous in $(t,x,\mu,y)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2\times\mathbb{R}^d$. If moreover $\partial_t f(t,x,\mu)$ is continuous in $(t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2$, we denote $f\in C_b^{1,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$; and write $f\in C_b^{2,(1,1)}(\mathbb{R}^d\times\mathscr{P}_2)$ if $f(t,x,\mu)$ does not depend on t.

It is easy to see that $C_b^{0,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ is strictly larger than $C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. For instance, let

$$f(t,x,\mu)=f(\mu)=g(\mu(h_1),\cdots,\mu(h_k)),$$

where $k\geq 1$, $\{h_i\}_{1\leq i\leq k}\subset C_b^2(\mathbb{R}^d)$, and $g\in C_b^1(\mathbb{R}^k)$ but ∇g is non-Lipschitzian. Then $f\in C_b^{0,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ but $f\notin C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$.

The main result in this section is the following.

Theorem 4.2. *Assume that $b,\sigma,\bar b,\bar\sigma\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. Then for any $\mathbf{V},\mathbf{f}\in C_b^{0,2,(1,1)}(\mathbb{R}^d\times\mathscr{P}_2)$ where $\mathbf{V}$ is bounded, and for any $\Phi\in C_b^{2,(1,1)}(\mathbb{R}^d\times\mathscr{P}_2)$, the PDE (1.4) with $u(T,\cdot,\cdot)=\Phi$ has a unique solution $u\in C_b^{1,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$, and the solution is given by*

$$\begin{aligned}&u(t,x,\mu)\\ &=\mathbb{E}\bigg[\Phi(X_{t,T}^{x,\mu},P_{t,T}^*\mu)\mathrm{e}^{\int_t^T\mathbf{V}(r,X_{t,r}^{x,\mu},P_{t,r}^*\mu)\mathrm{d}r}+\int_t^T\mathbf{f}(r,X_{t,r}^{x,\mu},P_{t,r}^*\mu)\mathrm{e}^{\int_t^r\mathbf{V}(\theta,X_{t,\theta}^{x,\mu},P_{t,\theta}^*\mu)\mathrm{d}\theta}\mathrm{d}r\bigg]\end{aligned}\tag{4.2}$$

for $(t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2$.

Let $\mathbb{E}_t^{x,\mu}$ be the expectation taken for the $\mathbf{L}_t$-diffusion process $(X_r,\mu_r)_{r\geq t}$ starting from (x,μ) at time t. Then by Theorem 2.2, (4.2) can be reformulated as

$$u(t,x,\mu)=\mathbb{E}_t^{x,\mu}\bigg[\Phi(X_T,\mu_T)\mathrm{e}^{\int_t^T\mathbf{V}(r,X_r,\mu_r)\mathrm{d}r}+\int_t^T\mathbf{f}(r,X_r,\mu_r)\mathrm{e}^{\int_t^r\mathbf{V}(\theta,X_\theta,\mu_\theta)\mathrm{d}\theta}\mathrm{d}r\bigg].$$

4.3 Proof of Theorem 4.2

We first recall the following result taken from [16] (see also [7, 13]), which was proved for time independent coefficients $b=\bar b$ and $\sigma=\bar\sigma$, but the proof obviously works for the present time dependent coefficients, since all calculations therein only rely on the regularity of coefficients in the space-distribution variables (x,μ) but has nothing to do with derivatives in time.

Lemma 4.3. *Let $b,\sigma,\bar b,\bar\sigma \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. Then*

$$\nabla X_{s,t}^{\cdot,\mu}(x),\ \nabla^2 X_{s,t}^{\cdot,\mu}(x),\ DX_{s,t}^{x,\cdot}(\mu)(y),\ \nabla\{DX_{s,t}^{x,\cdot}(\mu)(\cdot)\}(y)$$

are jointly continuous in $(t,x,\mu,y)\in[s,T]\times\mathbb{R}^d\times\mathscr{P}_2\times\mathbb{R}^d$, and there exists a constant $c>0$ such that

$$\mathbb{E}\big(\|\nabla X_{s,t}^{\cdot,\mu}(x)\|^2+\|\nabla^2 X_{s,t}^{\cdot,\mu}(x)\|^2+\|DX_{s,t}^{x,\cdot}(\mu)\|_{L^2(\mu)},\ \|\nabla\{DX_{s,t}^{x,\cdot}(\mu)\|_{L^2(\mu)}\big)\le c,$$

holds for all $0\le s\le T$ and $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$.

Replacing $(\bar b,\bar\sigma)$ in (2.6) by (b,σ), we consider the SDE

$$\mathrm{d}Y_{s,t}^{x,\mu}=b(t,Y_{s,t}^{x,\mu},P_{s,t}^*\mu)\mathrm{d}t+\sigma(t,Y_{s,t}^{x,\mu},P_{s,t}^*\mu)\mathrm{d}W_t,\quad Y_{s,s}^{x,\mu}=x. \tag{4.3}$$

Then the Markov property of the solution implies

$$P_{s,t}^*\mu=\int_{\mathbb{R}^d}\mathscr{L}_{Y_{s,t}^{x,\mu}}\mu(\mathrm{d}x), \tag{4.4}$$

where $P_{s,t}^*\mu:=\mathscr{L}_{X_t}$ for X_t solving the DDSDE (2.7) with $\mathscr{L}_{X_s}=\mu$. Combining this with Lemma 4.3, which applies to $Y_{s,t}^{x,\mu}$ replacing $X_{s,t}^{x,\mu}$ by taking $(\bar b,\bar\sigma)=(b,\sigma)$, we have the following result.

Lemma 4.4. *Let $b,\sigma\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. Then the assertion in Lemma 4.3 holds for $Y_{s,t}^{x,\mu}$ replacing $X_{s,t}^{x,\mu}$, and for any $f\in C_b^{(1,0)}(\mathscr{P}_2)$,*

$$\begin{aligned}Df(P_{s,t}^*\cdot)(\mu)(y)=&\int_{\mathbb{R}^d}\mathbb{E}\Big[\big\{DY_{s,t}^{z,\cdot}(\mu)(y)\big\}(Df)(P_{s,t}^*\mu)(Y_{s,t}^{z,\mu})\Big]\mu(\mathrm{d}z)\\&+\mathbb{E}\Big[\big\{\nabla Y_{s,t}^{\cdot,\mu}(y)\big\}(Df)(P_{s,t}^*\mu)(Y_{s,t}^{y,\mu})\Big],\end{aligned} \tag{4.5}$$

where for vectors $v_1,v_2\in\mathbb{R}^d$,

$$\big\langle\{\nabla Y_{s,t}^{\cdot,\mu}(z)\}v_1,v_2\big\rangle:=\big\langle\nabla_{v_2}Y_{s,t}^{\cdot,\mu}(z),v_1\big\rangle,$$

and $\big\{DY_{s,t}^{z,\cdot}(\mu)(\cdot)\big\}v_1\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$ is defined by

$$\big\langle\big\{DY_{s,t}^{z,\cdot}(\mu)(\cdot)\big\}v_1,\phi\big\rangle_{L^2(\mu)}:=\big\langle D_\phi\big\{Y_{s,t}^{z,\cdot}(\mu)(\cdot)\big\},v_1\big\rangle,\quad \phi\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu).$$

Proof. Let $f\in C_b^{(1,0)}(\mathscr{P}_2)$, i.e. f is L-differentiable with $Df(\mu)(y)$ having a bounded jointly continuous version. For a family of random variables $\{\xi^\varepsilon:\varepsilon\in[0,1)\}$ with $\dot\xi^0:=\frac{\mathrm{d}\xi^\varepsilon}{\mathrm{d}\varepsilon}\big|_{\varepsilon=0}$ existing in $L^1(\mathbb{P})$, we have

$$\lim_{\varepsilon\downarrow0}\frac{f(\mathscr{L}_{\xi(\varepsilon)})-f(\mathscr{L}_{X^0})}{\varepsilon}=\mathbb{E}\langle Df(\mathscr{L}_{\xi(0)})(\xi(0)),\dot\xi^0\rangle. \tag{4.6}$$

See for instance [22, Proposition 3.1], which is slightly extended from [12, Proposition A.2]. Since $P_{s,t}^*\mu = \int_{\mathbb{R}^d} \mathscr{L}_{Y_{s,t}^{z,\mu}}\mu(\mathrm{d}z)$, it follows form (4.6) that for any $\phi \in L^2(\mu)$,

$$\begin{aligned}
D_\phi f(P_{s,t}^*\cdot)(\mu) &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg\{ f\bigg(\int_{\mathbb{R}^d} \mathscr{L}_{Y_{s,t}^{z,\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}}\mu(\mathrm{d}z)\bigg) + \frac{\mathrm{d}}{\mathrm{d}\varepsilon} f\bigg(\int_{\mathbb{R}^d} \mathscr{L}_{Y_{s,t}^{z+\varepsilon\phi(z),\mu}}\mu(\mathrm{d}z)\bigg)\bigg\}\bigg|_{\varepsilon=0} \\
&= \int_{\mathbb{R}^d} \mathbb{E}\big\langle (Df)(P_{s,t}^*\mu)(Y_{s,t}^{z,\mu}), D_\phi Y_{s,t}^{z,\cdot}(\mu) + \nabla_{\phi(z)} Y_{s,t}^{\cdot,\mu}(z)\big\rangle \mu(\mathrm{d}z) \\
&= \mathbb{E}\int_{\mathbb{R}^d\times\mathbb{R}^d} \Big[\big\langle \big\{DY_{s,t}^{z,\cdot}(\mu)(\cdot)\big\}(Df)(P_{s,t}^*\mu)(Y_{s,t}^{z,\mu}), \phi\big\rangle_{L^2(\mu)} \\
&\qquad + \big\langle \{\nabla Y_{s,t}^{\cdot,\mu}(z)\}(Df)(P_{s,t}^*\mu)(Y_{s,t}^{z,\mu}), \phi(z)\big\rangle\Big]\mu(\mathrm{d}z).
\end{aligned}$$

Therefore, $f(P_{s,t}^*\cdot)$ is intrinsic differentiable such that (4.5) holds. $\square$

Lemma 4.5. *Assume that $b, \sigma, \bar{b}, \bar{\sigma} \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. For $\mathbf{V}, \mathbf{f} \in C_b^{0,2,(1,1)}$, let u be in (4.2). Then $u \in C_b^{0,2,(1,1)}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$.*

Proof. The proof is more or less standard, but for completeness we present below a brief proof.

Obviously, $u(t,x,\mu)$ is joint continuous in (t,x,μ). Next, by making derivatives in (x,μ) for u in (4.2), we obtain

$$\begin{aligned}
\nabla^i u(t,\cdot,\mu)(x) =& \mathbb{E}\bigg[\mathrm{e}^{\int_t^T \mathbf{V}(X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r}\bigg(\langle \nabla^i \Phi(\cdot, P_{s,T}^*\mu)(X_{t,T}^{x,\mu}), \nabla^i X_{t,T}^{\cdot,\mu}(x)\rangle \\
&+ \Phi(X_{t,T}^{x,\mu}, P_{s,T}^*\mu)\int_t^T \langle \nabla^i \mathbf{V}(\cdot, P_{s,r}^*\mu)(X_{t,r}^{x,\mu}), \nabla^i X_{t,r}^{\cdot,\mu}(x)\rangle \mathrm{d}r\bigg) \\
&+ \int_t^T \mathrm{e}^{\int_t^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta}\bigg((\langle \nabla^i \mathbf{f}(\cdot, P_{s,r}^*\mu)(X_{t,r}^{x,\mu}), \nabla^i X_{t,r}^{\cdot,\mu}(x)\rangle \mathrm{d}r \\
&+ \mathbf{f}(X_{t,r}^{x,\mu}, P_{s,r}^*\mu)\int_t^r \big\langle \nabla^i \mathbf{V}(\theta,\cdot,P_{t,\theta}^*\mu)(X_{t,\theta}^{x,\mu})\mathrm{d}\theta, \nabla^i X_{t,r}^{\cdot,\mu}(x)\big\rangle\bigg)\mathrm{d}r\bigg], \quad i=1,2.
\end{aligned}$$

and

$$\begin{aligned}
&Du(t,x,\cdot)(\mu)(y) \\
&= \mathbb{E}\bigg[\mathrm{e}^{\int_t^T \mathbf{V}(X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r}\bigg(\big\langle D\Phi(X_{t,T}^{x,\cdot}, P_{s,T}^*\mu)(\mu)(y), DX_{t,T}^{x,\cdot}(\mu)(y)\big\rangle \\
&+ \Phi(X_{t,T}^{x,\mu}, P_{s,T}^*\mu)\int_t^T \big\langle D\mathbf{V}(X_{t,r}^{x,\cdot}, P_{s,r}^*\mu)(\mu)(y), DX_{t,r}^{x,\cdot}(\mu)\big\rangle(y)\mathrm{d}r \\
&+ D\Phi(X_{s,t}^{x,\mu}, P_{s,t}^*\cdot)(\mu)(y) + \Phi(X_{t,T}^{x,\mu}, P_{s,T}^*\mu)\int_t^T D\mathbf{V}(X_{t,r}^{x,\mu}, P_{s,r}^*\cdot)(\mu)(y)\mathrm{d}r\bigg) \\
&+ \int_t^T \mathrm{e}^{\int_t^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}\mu)\mathrm{d}\theta}\bigg(\big\langle D\mathbf{f}(X_{t,r}^{x,\cdot}, P_{s,r}^*\mu)(\mu), DX_{t,r}^{x,\cdot}(\mu)\big\rangle(y) + D\mathbf{f}(r, X_{s,r}^{x,\mu}, P_{s,r}^*\cdot)(\mu)(y) \\
&+ \mathbf{f}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\int_t^r \big\{D\mathbf{V}(\theta, X_{t,\theta}^{x,\cdot}, P_{t,\theta}\mu)(\mu)(y) + D\mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{s,\theta}\cdot)(\mu)(y)\big\}\mathrm{d}\theta\bigg)\mathrm{d}r\bigg].
\end{aligned}$$

Combining these with Lemma 4.4, we finish the proof. $\square$

Moreover, we apply Lemma 2.3 to prove the following Itô's formula for the solution of (2.3) and (2.6), which is known in e.g. [12] under (2.11) for α, β being the coefficients of the DDSDE (2.7).

Lemma 4.6. *Let b, σ satisfy* **(A)**. *Then for any $f \in \bar{C}^{2,(1,1))}(\mathbb{R}^d \times \mathscr{P}_2)$ and $s \in [0, T)$,*

$$\mathrm{d}f(X_{s,t}^{x,\mu}, P_{s,t}^*\mu) = \mathbf{L}_t f(X_{s,t}^{x,\mu}, P_{s,t}^*\mu)\mathrm{d}t + \langle \nabla f(\cdot, P_{s,t}^*\mu)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu}, P_{s,t}^*\mu)\mathrm{d}W_t \rangle, \quad t \in [s, T].$$

Proof. It is easy to see that the assumption **(A)** implies (2.12) for

$$\alpha_t := b(t, X_t, P_{s,t}^*\mu), \quad \beta_t := \sigma(t, X_t, P_{s,t}^*\mu)$$

for X_t solving (2.7) from time s with $\mathscr{L}_{X_s} = \mu$. By Lemma 2.3 and the definition of $\mathbf{L}_t^{(2)}$, for any $z \in \mathbb{R}^d$ we have

$$\mathrm{d}f(z, P_{s,t}^*\mu) = \mathbf{L}_t^{(2)} f(z, P_{s,t}^*\mu)\mathrm{d}t, \quad t \geq s.$$

Combining this with Itô's formula for $X_{s,t}^{x,\mu}$ in (2.6), we obtain

$$\begin{aligned}
\mathrm{d}f(X_{s,t}^{x,\mu}, P_{s,t}^*\mu) &= \big\{\mathrm{d}f(z, P_{s,t}^*\mu)\big\}\big|_{z=X_{s,t}^{x,\mu}} + \big\{\mathrm{d}f(X_{s,t}^{x,\mu}, \nu)\big\}\big|_{\nu=P_{s,t}^*\mu} \\
&= \mathbf{L}_t f(X_{s,t}^{x,\mu}, P_{s,t}^*\mu)\mathrm{d}t + \langle \nabla f(\cdot, P_{s,t}^*\mu)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu}, P_{s,t}^*\mu)\mathrm{d}W_t \rangle.
\end{aligned}$$

$\square$

We are now ready to prove Theorem 4.2 as follows.

Proof of Theorem 4.2. **(a)** We first prove that u in (4.2) solves (1.4). Let $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathscr{P}_2$. For any $\varepsilon \in (0, T - t)$ we have

$$\begin{aligned}
u(t, x, \mu) &:= \mathbb{E}\bigg[\Phi(X_{t,T}^{x,\mu}, P_{t,T}^*\mu)\mathrm{e}^{\int_t^T \mathbf{V}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r} + \int_t^T \mathrm{e}^{\int_t^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta} \mathbf{f}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r\bigg] \\
&= I_1 + I_2,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &:= \mathbb{E}\bigg[\Phi(X_{t,T}^{x,\mu}, P_{t,T}^*\mu)\mathrm{e}^{\int_{t+\varepsilon}^T \mathbf{V}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r} + \int_{t+\varepsilon}^T \mathrm{e}^{\int_{t+\varepsilon}^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta} \mathbf{f}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r\bigg], \\
I_2 &:= \mathbb{E}\bigg[\Phi(X_{t,T}^{x,\mu}, P_{t,T}^*\mu)\Big\{\mathrm{e}^{\int_t^T \mathbf{V}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r} - \mathrm{e}^{\int_{t+\varepsilon}^T \mathbf{V}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r}\Big\} \\
&\quad + \int_t^{t+\varepsilon} \mathrm{e}^{\int_t^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta} \mathbf{f}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r \\
&\quad + \int_{t+\varepsilon}^T \Big\{\mathrm{e}^{\int_t^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta} - \mathrm{e}^{\int_{t+\varepsilon}^r \mathbf{V}(\theta, X_{t,\theta}^{x,\mu}, P_{t,\theta}^*\mu)\mathrm{d}\theta}\Big\} \mathbf{f}(r, X_{t,r}^{x,\mu}, P_{t,r}^*\mu)\mathrm{d}r\bigg].
\end{aligned}$$

By the Markov property of $(X_{t,r}^{x,\mu}, P_{t,r}^*\mu)_{r \in [t,T]}$, we obtain

$$I_1 = \mathbb{E}\bigg\{\mathbb{E}\bigg[\Phi(X_{t+\varepsilon,T}^{y,\nu}, P_{t+\varepsilon,T}^*\nu)\mathrm{e}^{\int_{t+\varepsilon}^T \mathbf{V}(r, X_{t+\varepsilon,r}^{y,\nu}, P_{t+\varepsilon,r}^*\nu)\mathrm{d}r}$$

$$
\left. \left. + \int_{t+\varepsilon}^{T} \mathbf{f}(r, X_{t+\varepsilon,r}^{y,\nu}, P_{t+\varepsilon,r}^{*}\nu) \mathrm{e}^{\int_{t+\varepsilon}^{r} \mathbf{V}(\theta, X_{t+\varepsilon,\theta}^{y,\nu}, P_{t+\varepsilon,\theta}^{*}\nu)\mathrm{d}\theta} \mathrm{d}r \right] \right|_{(y,\nu)=(X_{t,t+\varepsilon}^{x,\mu}, P_{t,t+\varepsilon}^{*}\mu)} \Bigg\}
$$

$$
= \mathbb{E} u(t+\varepsilon, X_{t,t+\varepsilon}^{x,\mu}, P_{t,t+\varepsilon}^{*}\mu).
$$

Combining this with Lemma 4.5 and Lemma 4.6, we arrive at

$$
I_1 = u(t+\varepsilon, x, \mu) + \mathbb{E} \int_{t}^{t+\varepsilon} \mathbf{L}_r u(t+\varepsilon, \cdot, \cdot)(X_{t,r}^{x,\mu}, P_{t,r}^{*}\mu) \mathrm{d}r.
$$

Noting that $u(t,x,\mu) = I_1 + I_2$, b, σ, Φ and $\mathbf{f}$ are continuous with linear growth, $\mathbf{V}$ is continuous and bounded, and $u \in C_b^{0,2,(1,1)}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$, we may apply the dominated convergence theorem to derive

$$
\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \frac{u(t,x,\mu) - u(t+\varepsilon, x, \mu)}{\varepsilon} \\
&= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \mathbb{E} \bigg\{ \int_{t}^{t+\varepsilon} \mathbf{L}_r u(t+\varepsilon, \cdot, \cdot)(X_{t,r}^{x,\mu}, P_{t,r}^{*}\mu) \mathrm{d}r + I_2 \bigg\} \\
&= \mathbf{L}_t u(t, \cdot, \cdot)(x, \mu) + (u\mathbf{V})(t,x,\mu) + \mathbf{f}(t,x,\mu).
\end{aligned}
$$

Therefore, u solves (1.4), and $\partial_t u$ is continuous on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$, so that by Lemma 4.5 and the definition, we have $u \in C_b^{1,2,(1,1)}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$.

(b) Let $u \in C^{1,2,(1,1)}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$ be a solution to (1.4), we prove that it satisfies (4.2). Indeed, let

$$
\eta_t = u(t, X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) \mathrm{e}^{\int_s^t \mathbf{V}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu)\mathrm{d}r} + \int_s^t \mathbf{f}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu) \mathrm{e}^{\int_s^r \mathbf{V}(\theta, X_{s,\theta}^{x,\mu}, P_{s,\theta}^{*}\mu)\mathrm{d}\theta} \mathrm{d}r, \quad t \in [0,T].
$$

By Lemma 4.6 and (1.4), for any $s \in [0,T)$, we have

$$
\begin{aligned}
\mathrm{d}\eta_t = &\Big\{ (\partial_t + \mathbf{L}_t) u(t, \cdot, \cdot)(X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) + (u\mathbf{V})(t, X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) + \mathbf{f}(t, X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) \\
&\quad + \big\langle \nabla u(t, \cdot, P_{s,t}^{*}\mu)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) \mathrm{d}W_t \big\rangle \Big\} \mathrm{e}^{\int_s^t \mathbf{V}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu)\mathrm{d}r} \\
= &\ \mathrm{e}^{\int_s^t \mathbf{V}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu)\mathrm{d}r} \big\langle \nabla u(t, \cdot, P_{s,t}^{*}\mu)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu}, P_{s,t}^{*}\mu) \mathrm{d}W_t \big\rangle, \quad t \in [s,T].
\end{aligned}
$$

Therefore, for any $s \in [0,T]$,

$$
\begin{aligned}
&u(s,x,\mu) = \mathbb{E}\eta_s = \mathbb{E}\eta_T \\
&= \mathbb{E} \bigg\{ u(T, X_{s,T}^{x,\mu}, P_{s,T}^{*}\mu) \mathrm{e}^{\int_s^T \mathbf{V}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu)\mathrm{d}r} + \int_s^T \mathbf{f}(r, X_{s,r}^{x,\mu}, P_{s,r}^{*}\mu) \mathrm{e}^{\int_s^r \mathbf{V}(\theta, X_{s,\theta}^{x,\mu}, P_{s,\theta}^{*}\mu)\mathrm{d}\theta} \mathrm{d}r \bigg],
\end{aligned}
$$

that is, u satisfies (4.2). $\square$

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] S. Albeverio, Y. G. Kondratiev, M. Röckner, *Analysis and geometry on configuration spaces,* J. Funct. Anal. 154(1998), 444–500.

[3] S. Albeverio, Y. G. Kondratiev, M. Röckner, *Analysis and geometry on configuration spaces: the Gibbsian case,* J. Funct. Anal. 157(1998), 242–291.

[4] V. Barbu, M. Röckner, *Probabilistic representation for solutions to nonlinear Fokker-Planck equations,* SIAM J. Math. Anal. 502018, 4246–4260.

[5] V. Barbu, M. Röckner, *From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE,* arXiv:1808.10706v2

[6] V. I. Bogachev, N. V. Krylov, M. Röckner, S.V. Shaposhnikov, *Fokker–Planck–Kolmogorov equations,* Mathematical Surveys and Monographs 207, American Mathematical Society, Providence, RI, 2015.

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

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

[9] P. E. Chaudru de Raynal, *Strong well-posedness of McKean-Vlasov stochastic differential equation with Hölder drift,* arXiv: 1512.08096; 2019: hal-01248324v3.

[10] J.-F. Chassagneux, D. Crisan, F. Delarue, *A Probabilistic approach to classical solutions of the master equation for large population equilibria,* arXiv:1411.3009v2, 2015.

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

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

[13] T. Hao, J. Li, *Mean-field SDEs with jumps and nonlocal integral-PDEs,* Nonlin. Diff. Equat. Appl. 23(2016), 1–51.

[14] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients,* to appear in Stoch. Proc. Appl. arXiv:1805.01682.

[15] N. V. Krylov, M. Röckner, *Strong solutions of stochastic equations with singular time dependent drift,* Probab. Theory Relat. Fields 131(2005), 154–196.

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

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

[18] L. Overbeck, M. Röckner, B. Schmuland, *An analytic approach to Fleming-Viot processes with interactive selection,* Ann. Probab. 23(1995), 1–36.

[19] L. Overbeck, M. Röckner, *Geometric aspects of finite- and infinite-dimensional Fleming-Viot processes,* Random Oper. Stochastic Equations 5(1997), 35–58.

[20] M. Röckner, *Dirichlet forms on infinite-dimensional "manifold-like" state spaces: a survey of recent results and some prospects for the future,* Probability towards 2000 (L. Accardi et al., eds.), Lecture Notes in Statistics, Springer, Berlin, 1998, pp. 287–306.

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

[22] P. Ren, F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications,* arXiv:1809.06068.

[23] D.W. Stroock, S.R.S. Varadhan, *Multidimensional diffusion processes,* Springer-Verlag, New York 1979.

[24] D. Trevisan, *Well-posedness of multidimensional diffusion processes,* Elect. J. Probab. 21(2016), 1–41.

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

[26] X. Zhang, *Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients,* Electron. J. Probab. 16(2011), 1096-1116.