Image Dependent Conditional McKean-Vlasov SDEs for Measure-Valued Diffusion Processes*

Feng-Yu Wang[a),b)]

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

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

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

October 17, 2020

Abstract

We consider a special class of mean field SDEs with common noise which depend on the image of the solution (i.e. the conditional distribution given noise). The strong well-posedness is derived under a monotone condition which is weaker than those used in the literature of mean field games, the Feynman-Kac formula is established to solve Schrördinegr type PDEs on $\mathscr{P}_2$, and the ergodicity is proved for a class of measure-valued diffusion processes.

AMS subject Classification: 60J60, 58J65.
Keywords: Image dependent SDE, measure-valued diffusion process, ergodicity, Feynman-Kac formula, intrinsic/Lions derivative.

1 Introduction

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

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

where $|\cdot|$ is the norm in $\mathbb{R}^d$. We will use $\|\cdot\|$ to denote the operator norm of a matrix or linear operator, and use $\|\cdot\|_{HS}$ to stand for the Hilbert-Schmidt norm. It is well known that

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

$\mathscr{P}_2$ is a Polish space under the Wasserstein distance

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

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

Since 1996 when Albeverio, Kondratiev and Röckner [1] introduced the intrinsic derivative on the configuration space over manifolds, diffusion processes on the space of discrete Radon measures have been investigated by using Dirichlet forms, see [15] and references within. This derivative provides a natural Riemannian structure on the Wasserstein space $(\mathscr{P}_2,\mathbb{W}_2)$, see Subsection 1.2 below.

To develop stochastic analysis and applications on this space, we intend to construct diffusion processes generated by second order differentiable operators and solve the associated PDEs on $\mathscr{P}_2$. Below we first recall the intrinsic/Lions derivative on $\mathscr{P}_2$.

According to [1], let $L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$ be the tangent space of $\mathscr{P}_2$ at point $\mu\in\mathscr{P}_2$, and define the directional derivative by

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

When $\phi\mapsto D_\phi f(\mu)$ is a bounded linear functional on $L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$, or equivalently the map

*D0
$$L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)\ni\phi\mapsto f(\mu\circ(\mathrm{Id}+\phi)^{-1}) \tag{1.1}$$

is Gateaux differentiable at $\phi=0$, there exists a unique element $Df(\mu)\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$ such that

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

In this case, we call f intrinsically differentiable at μ with derivative $Df(\mu)$. According to Lions (see [4]), if $Df(\mu)$ exists and

*D1
$$\lim_{\mu(|\phi|^2)\to 0}\frac{f(\mu\circ(\mathrm{Id}+\phi)^{-1})-f(\mu)-D_\phi f(\mu)}{\sqrt{\mu(|\phi|^2)}}=0, \tag{1.2}$$

i.e. the map in (1.1) is Fréchet differentiable at $\phi=0$, we call f L-differentiable at $\mu\in\mathscr{P}_2$. If f is L-differentiable at any $\mu\in\mathscr{P}_2$, we call it L-differentiable. Note that $Df(\mu)$ is a μ-a.e. defined $\mathbb{R}^d$-valued function. Let $\{Df(\mu)\}_i$ be its i-th component for $1\le i\le d$.

In this paper, we investigate diffusion processes and applications on the Wasserstein space $\mathscr{P}_2$. Let $m\ge 1$, and let

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

be measurable such that $|b(t,\cdot,\mu)|+\|\sigma(t,\cdot,\mu)\|_{HS}^2\in L^1(\mu)$ for any $(t,\mu)\in[0,\infty)\times\mathscr{P}_2$. We consider the following time-dependent second order differential operators on $\mathscr{P}_2$:

HA1
$$\begin{aligned}\mathscr{A}_t f(\mu) := {}&\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\big\langle\sigma(t,y,\mu)\sigma(t,z,\mu)^*, D^2f(\mu)(y,z)\big\rangle\mu(\mathrm{d}y)\mu(\mathrm{d}z)\\ &+\int_{\mathbb{R}^d}\Big(\frac{1}{2}\big\langle(\sigma\sigma^*)(t,y,\mu),\nabla\{Df(\mu)\}(y)\big\rangle+\big\langle b(t,y,\mu),Df(\mu)(y)\big\rangle\Big)\mu(\mathrm{d}y),\end{aligned} \tag{1.3}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{R}^d$ or $\mathbb{R}^d \otimes \mathbb{R}^d$. We also consider the following extension of $\mathscr{A}_t$ on $\mathbb{R}^d \times \mathscr{P}_2$:

HA2 (1.4)
$$\begin{aligned} \tilde{\mathscr{A}_t} f(x,\mu) &:= \mathscr{A}_t f(x,\cdot)(\mu) + \frac{1}{2}\big\langle \sigma(t,x,\mu)\sigma(t,x,\mu)^*, \nabla^2 f(x,\mu)\big\rangle + \big\langle b(t,x,\mu), \nabla f(x,\mu)\big\rangle \\ &\quad + \int_{\mathbb{R}^d} \big\langle (D\nabla f)(x,\mu)(y), \sigma(t,y,\mu)\sigma(t,x,\mu)^*\big\rangle \mu(\mathrm{d}y). \end{aligned} \tag{1.4}$$

To present reasonable pre-domains of $\mathscr{A}_t$ and $\tilde{\mathscr{A}_t}$, we introduce below some classes of L-differentiable functions.

(1) We write $f \in C^1(\mathscr{P}_2)$, if f is L-differentiable and the derivative has a μ-version $Df(\mu)(x)$ which is jointly continuous in $(\mu,x) \in \mathscr{P}_2 \times \mathbb{R}^d$. If moreover $Df(\mu)(x)$ is bounded in $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$, we denote $f \in C_b^1(\mathscr{P}_2)$.

(2) We write $f \in C^{(1,1)}(\mathscr{P}_2)$, if $f \in C_b^1(\mathscr{P}_2)$ and $Df(\mu)(x)$ is differentiable in x such that the $\mathbb{R}^d \otimes \mathbb{R}^d$-valued function

$$\nabla\{Df(\mu)\}(x) := \big(\partial_{x_j}\{Df(\mu)(x)\}_i\big)_{1\le i,j\le d}$$

is jointly continuous in $(\mu,x) \in \mathscr{P}_2 \times \mathbb{R}^d$. If moreover $Df(\mu)(x)$ and $\nabla\{Df(\mu)\}(x)$ are bounded in $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$, we denote $f \in C_b^{(1,1)}(\mathscr{P}_2)$.

(3) We write $f \in C^2(\mathscr{P}_2)$, if $f \in C^{(1,1)}(\mathscr{P}_2)$ and $Df(\mu)(x)$ is L-differentiable in μ such that the $\mathbb{R}^d \otimes \mathbb{R}^d$-valued function

$$D^2 f(\mu)(x,y) := \big(\big\{D[Df(\mu)(x)]_i(y)\big\}_j\big)_{1\le i,j\le d}$$

is jointly continuous in $(\mu,x,y) \in \mathscr{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d$. If moreover $f \in C_b^{(1,1)}(\mathscr{P}_2)$ and $D^2 f(\mu)(x,y)$ is bounded in $(x,y,\mu) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathscr{P}_2$, we denote $f \in C_b^2(\mathscr{P}_2)$.

(4) We write $f \in C^{2,2}(\mathbb{R}^k \times \mathscr{P}_2)$ for some $k \ge 1$, if f is a continuous function on $\mathbb{R}^k \times \mathscr{P}_2$ such that $f(\cdot,\mu) \in C^2(\mathbb{R}^k)$ for $\mu \in \mathscr{P}_2$, $f(x,\cdot) \in C^2(\mathscr{P}_2)$ for $x \in \mathbb{R}^k$,

$$(D\nabla f)(x,\mu)(y) := \big(\big\{D[\partial_{x_i} f(x,\mu)]\big\}_j\big)_{1\le i,j\le d} \in \mathbb{R}^d \otimes \mathbb{R}^d$$

exists, and the derivatives

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

are bounded and jointly continuous in the corresponding arguments.

Example 1.1. For any $p \ge 1$, consider the following class of cylindrical functions

FC (1.5)
$$\begin{aligned} \mathscr{F}C_b^p(\mathscr{P}_2) := \big\{ f(\mu) &:= g(\mu(h_1), \cdots, \mu(h_n)) : \\ & n \ge 1, g \in C_b^p(\mathbb{R}^n), h_i \in C_b^p(\mathbb{R}^d), 1 \le i \le n \big\}. \end{aligned} \tag{1.5}$$

When $p=2$, such a function is in the class $C_b^2(\mathscr{P}_2)$ with

LDD

$$\begin{aligned} Df(\mu)(x) &= \sum_{i=1}^n (\partial_i g)(\mu(h_1),\cdots,\mu(h_n))\nabla h_i(x),\\ D^2 f(\mu)(x,y) &= \sum_{i,j=1}^n (\partial_i\partial_j g)(\mu(h_1),\cdots,\mu(h_n))\{\nabla h_i(x)\}\otimes\{\nabla h_j(y)\}, \end{aligned} \tag{1.6}$$

where $\{\nabla h_i(x)\}\otimes\{\nabla h_j(y)\}\in\mathbb{R}^d\otimes\mathbb{R}^d$ is defined as

$$\big(\{\nabla h_i(x)\}\otimes\{\nabla h_j(y)\}\big)_{kl} = \{\partial_k h_i(x)\}\partial_l h_j(y),\quad 1\le k,l\le d, x,y\in\mathbb{R}^d.$$

Moreover, $f\in C^{2,2}(\mathbb{R}^d\times\mathscr{P}_2)$ if $f(x,\mu)=g(x,\mu(h_1),\cdots,\mu(h_n))$ for some $n\ge 1$, $g\in C_b^2(\mathbb{R}^{n+d})$ and $\{h_i\}_{1\le i\le n}\subset C_b^2(\mathbb{R}^d)$.

We will construct the $\mathscr{A}_t$-diffusion process by solving the following SDE on $\mathbb{R}^d$:

E1

$$\mathrm{d}X_{s,t}^{x,\mu} = b(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)\mathrm{d}t + \sigma(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)\mathrm{d}W_t,\ \ \Lambda_{s,t}^\mu := \mu\circ(X_{s,t}^{\cdot,\mu})^{-1},\ \ t\ge s, X_{s,s}^{x,\mu}=x, \tag{1.7}$$

where W_t is the m-dimensional Brownian motion on a complete filtration probability space $(\Omega,\{\mathscr{F}_t\}_{t\ge0},\mathbb{P})$, $(s,x,\mu)\in[0,\infty)\times\mathbb{R}^d\times\mathscr{P}_2$. Since this SDE depends on the image of solutions, we call it image dependent SDE.

It turns out that the solution of (1.7) for $s=0$ gives rise to a strong solution to the following conditional McKean-Vlasov SDE arising from mean field games:

E''

$$\mathrm{d}X_t = b(t,X_t,\mathscr{L}_{X_t|W})\mathrm{d}t + \sigma(t,X_t,\mathscr{L}_{X_t|W})\mathrm{d}W_t,\quad \mathscr{L}_{X_0}=\mu\in\mathscr{P}_2, \tag{1.8}$$

where $\mathscr{L}_\xi$ and $\mathscr{L}_{\xi|W}$ denote the distribution and the conditional distribution given $\{W_t: t\ge 0\}$ for a random variable ξ. More precisely, $X_t = X_{0,t}^{X_0,\mu}$, see the proof of Corollary 2.2 below. So, the measure-valued process $\Lambda_{0,t}$ in (1.7) is indeed the conditional distribution of X_t given the noise W, and we may call (1.7) an image dependent conditional McKean-Vlasov SDE as in the title of the paper.

The weak solution to (1.8) has been investigated by using mean filed games with common noise. More precisely, let $\{x_i\}_{i\ge}$ be a sequence of points in $\mathbb{R}^d$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \delta_{x_i} = \mu\ \text{ weakly},$$

consider the SDEs

$$\mathrm{d}X_t^{n,i} = b\Big(t,X_t^{n,i},\frac{1}{n}\sum_{j=1}^n\delta_{X_t^{n,j}}\Big)\mathrm{d}t + \sigma\Big(t,X_t^{n,i},\frac{1}{n}\sum_{j=1}^n\delta_{X_t^{n,j}}\Big)\mathrm{d}W_t,\quad X_0^{n,i}=x_i, 1\le i\le n.$$

Then under reasonable conditions, when $n\to\infty$ the law of $(X_t^{n,1})_{t\ge0}$ converges weakly to a probability measure on the path space $C([0,\infty);\mathbb{R}^d)$ which solves (1.8) weakly. See [5, 6, 7, 8, 10] for the study of a more general model than (1.8) where an additional independent Brownian noise W_t^0 is included:

E'1
$$\mathrm{d}X_t = b(t, X_t, \mathscr{L}_{X_t|W})\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t|W})\mathrm{d}W_t + \sigma^0(t, X_t, \mathscr{L}_{X_t|W})\mathrm{d}W_t^0 \tag{1.9}$$

for $\mathscr{L}_{X_0} = \mu \in \mathscr{P}_2$, where σ^0 takes values in $\mathbb{R}^d \otimes \mathbb{R}^l$ if W_t^0 is l-dimensional. The study of this type SDE using mean field games goes back to the poineering works of Lasry and Lions [17, 18, 19] and Huang, Malhamé and Caines [11, 12], see the nice monograph [5] for a theory of mean field games with common noises and applications. In this paper, we will study the strong solutions and applications of (1.7) (hence, (1.8)) in a straightforward way under reasonably weaker conditions on the coefficients.

In the remainder of this section, we first summarize the main results of the paper, then present a link of the present model to the Brownian motion on $\mathscr{P}_2$ for further study, and finally introduce some previous work for analysis on the Wasserstein space.

1.1 Summary of main results

Existence and uniqueness. Under a monotone condition, Theorem 2.1 ensures the existence, uniqueness and moment estimates of solutions to the image SDE (1.7), and that the unique solution is the diffusion processes generated by $\mathscr{A}_t$ on $\mathscr{P}_2$ and $\tilde{\mathscr{A}_t}$ on $\mathbb{R}^d \times \mathscr{P}_2$ respectively. As a consequence, the strong well-posedness is derived for the conditional distribution dependent SDE (1.8). Our monotone condition is weaker than those in [5, 6] but incomparable with those of [10], see Remark 2.1 below for details.

Feynman-Kac formula. By using the diffusion process $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)$, Theorem 3.1 solves the following PDE for U on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$:

PDE
$$\begin{aligned} &\partial_t U(t,x,\mu) + \tilde{\mathscr{A}_t} U(t,x,\cdot)(\mu) + (VU)(t,x,\mu) + F(t,x,\mu) = 0, \\ &U(T,x,\mu) = \Phi(x,\mu), \quad (t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2, \end{aligned} \tag{1.10}$$

where $T > 0$ is a fixed time, Φ is a function on $\mathbb{R}^d \times \mathscr{P}_2$, and V, F are functions on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$. When Φ, F and V do not depend on $x \in \mathbb{R}^d$, this PDE reduces to

PDE2
$$\begin{aligned} &\partial_t U(t,\mu) + \mathscr{A}_t U(t,\cdot)(\mu) + (VU)(t,\mu) + F(t,\mu) = 0, \\ &U(T,\mu) = \Phi(\mu), \quad (t,\mu) \in [0,T] \times \mathscr{P}_2. \end{aligned} \tag{1.11}$$

When $V = 0$ these two SPDEs are included as a special case by the Master equations studied in the literature of mean field games under stronger assumptions on b and σ, see Remark 3.1 below for details.

Exponential ergodicity and structure of invariant probability measures. Let b and σ do not depend on t. Under a dissipativity condition, Theorem 4.1 provides the exponential convergence rate of the diffusion process $(X_t^{x,\mu}, \Lambda_t^\mu) := (X_{0,t}^{x,\mu}, \Lambda_{0,t}^\mu)$ to its unique invariant probability measure $\tilde{\Pi}$. Consequently, the diffusion process Λ_t^μ converges at the same rate to the invariant probability measure $\Pi := \tilde{\Pi}(\mathbb{R}^d \times \cdot)$.

Moreover, let $b_0(x) = b(x, \delta_x), \sigma_0(x) = \sigma(x, \delta_x)$, and let μ_0 be the unique invariant probability measure for the classical SDE

E0
$$\mathrm{d}X_t = b_0(X_t)\mathrm{d}t + \sigma_0(X_t)\mathrm{d}W_t. \tag{1.12}$$

By Theorem 4.2, $\tilde{\Pi}$ and Π have the representations

E01
$$\tilde{\Pi}(\mathrm{d}x, \mathrm{d}\mu) = \mu_0(\mathrm{d}x)\delta_{\delta_x}(\mathrm{d}\mu), \quad \Pi = \int_{\mathbb{R}^d} \delta_{\delta_x}\mu_0(\mathrm{d}x), \tag{1.13}$$

where δ_{δ_x} is the Dirac measure at point $\delta_x \in \mathscr{P}_2$. This structure describes an asymptotic collision property of the diffusion process Λ_t^μ: starting from any probability measure $\mu \in \mathscr{P}_2$, the measure-valued process eventually decays to a Dirac random variable, for which the whole mass focus on a single random point.

1.2 Some related studies

Brownian motion on $\mathscr{P}_2$. A Riemannian structure has been introduced in [2] on the Wasserstein space $(\mathscr{P}_2, \mathbb{W}_2)$. With the intrinsic/Lions derivative, this space is an infinite-dimensional Riemannian manifold with gradient D and Riemannian metric $\langle \cdot, \cdot \rangle_{L^2(\mu)}$ on the tangent space $L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$; that is, $\mathbb{W}_2$ is the Riemannian distance induced by D.

As in the finite-dimensional Riemannian setting, we introduce the square field

$$\Gamma(f, g)(\mu) := \int_{\mathbb{R}^d} \langle Df(\mu)(x), Dg(\mu)(x) \rangle \mu(\mathrm{d}x), \quad f, g \in C_b^1(\mathscr{P}_2),$$

and the Laplace operator

$$\Delta f(\mu) := \int_{\mathbb{R}^d} \mathrm{tr}\big\{D^2 f(\mu)(x, x)\big\}\mu(\mathrm{d}x), \quad f \in C^2(\mathscr{P}_2).$$

Then by the chain rule we have

$$\Gamma(f, g) = \frac{1}{2}\big\{\Delta(fg) - f\Delta g - g\Delta f\big\}, \quad f, g \in C^2(\mathscr{P}_2).$$

This structure can be easily extended to the Wasserstein space $\mathscr{P}_2(M)$ over a Riemannian manifold M. Note that when M is compact we have $\mathscr{P}_2(M) = \mathscr{P}(M)$, the space of all probability measures on M.

To develop stochastic analysis on $\mathscr{P}_2$, it is interesting to construct the Brownian motion, i.e. the diffusion process generated by $\frac{1}{2}\Delta$; or more generally, to construct diffusion processes on $\mathscr{P}_2$ with square field Γ. This is the main motivation of [23] introduced in the next subsection.

Below we explain that when $\sigma\sigma^* = \mathrm{Id}$ and $\mu = \delta_x$ is a Dirac measure at some point $x \in \mathbb{R}^d$, the process $(\Lambda_{0,t}^\mu)_{t \geq 0}$ is such a diffusion process. Indeed, it is easy to check that the square field of the $\mathscr{A}_t$-diffusion process is

$$\Gamma_t(f, g)(\mu) := \big\{\mathscr{A}_t(fg)(\mu) - f\mathscr{A}_t g - g\mathscr{A}_t\big\}(\mu)$$

$$= \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \sigma(t,x,\mu)^* Df(\mu)(x), \sigma(t,y,\mu)^* Df(\mu)(y)\rangle \mu(\mathrm{d}x)\mu(\mathrm{d}y) \quad f,g\in C_b^2(\mathscr{P}_2), \mu\in\mathscr{P}_2.$$

In particular, when $\sigma\sigma^* = \mathrm{Id}$, we have

$$\Gamma_t(f,g)(\mu) = \Gamma(f,g)(\mu), \quad \mu\in\mathscr{P}_2^0 := \{\delta_x : x\in\mathbb{R}^d\}.$$

Since when $\mu=\delta_x$ for some $x\in\mathbb{R}^d$, $\Lambda_{s,t}^\mu = \delta_{X_{s,t}^{x,\delta_x}}$ is a diffusion process on $\mathscr{P}_2^0$, Theorem 2.1(2) below implies that $(\Lambda_{s,t}^\mu)_{t\geq s}$ for $\mu\in\mathscr{P}_2^0$ is a diffusion process with square field Γ. However, this does not hold for $\mu\notin\mathscr{P}_2^0$.

Measure-valued diffusion processes. Measure-valued diffusion processes have been constructed using Dirichlet forms. Let $\mathscr{P}(\mathbb{S}^1)$ be the space of all probability measures on the unit circle $\mathbb{S}^1$. A family of probability measures $\{\mathbb{P}_\beta\}_{\beta>0}$ on $\mathscr{P}(\mathbb{S}^1)$, called "entropic measures" with inverse temperature $\beta>0$, have been constructed by von Renesse and Sturm [23] such that for each $\beta>0$, the bilinear form

$$\mathscr{E}(f,g) := \int_{\mathscr{P}_2(\mathbb{S}^1)} \langle Df(\mu), Dg(\mu)\rangle_{L^2(\mu)} \mathbb{P}_\beta(\mathrm{d}\mu)$$

gives a symmetric Dirichlet form on $L^2(\Pi_\beta)$, which refers to a $\mathbb{P}_\beta$-a.e. starting diffusion process on $\mathscr{P}(\mathbb{S}^1)$. See also [24] for a different Dirichlet form on $\mathscr{P}([0,1])$ with square field Γ. The construction of Dirichlet forms in these papers heavily relies on the one-dimensional property. See also [15, 22, 27, 28] and references within for the study of different type measure-valued diffusion processes using Dirichlet forms.

Next, following the idea of Konarovskyi (see e.g.[16]), [13, 20] constructed another type of diffusion process on $\mathscr{P}([0,1])$ by taking the limit as $N\to\infty$ of a system with N coalescing and mass-carrying particles. The generator $\mathscr{L}_t$ of the process has the formulation [20, Theorem A.3]

$$\mathscr{L}_t f(\mu) = \frac{1}{2}\mathbb{E}\int_0^1 \Big[\frac{\{Df(\mu)\}'(\xi_t(u))}{m_t(u)} + D^2 f(\mu)(\xi_t(u),\xi_t(u))\Big]\mathrm{d}u, \quad f\in C_b^2(\mathscr{P}([0,1])), t\geq 0,$$

where $\{(\xi_t(u))_{t\geq 0} : u\in[0,1]\}$ is a family of continuous martingales with $\xi\in D((0,1);C([0,\infty))$ and quadratic variations [20, Proposition 5.7]

$$\langle \xi(u),\xi(u)\rangle_t = \int_0^t \frac{\mathrm{d}s}{m_s(u)}, \quad m_s(u) = \int_0^1 1_{\{\xi_s(u)\neq\xi_s(v)\}}\mathrm{d}v.$$

Recently, the study of this type Wasserstein diffusion processes have been used in [21] to solve a class of Fokker-Planck equations on the interval driven by an infinite-dimensional noise.

2 Image dependent SDE and diffusion processes on $\mathscr{P}_2$

We will construct the $\mathscr{A}_t$-diffusion process by solving the image SDE (1.7). In general, we allow the coefficients

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

to be random but progressively measurable with respect to the filtration $\mathscr{F}_t$. We first present the definition of solution.

Definition 2.1. Let $(s,\mu)\in[0,\infty)\times\mathscr{P}_2$. A family of adapted processes $\{(X_{s,t}^{x,\mu})_{t\ge s}:x\in\mathbb{R}^d\}$ is called a solution to (1.7), if the following conditions hold $\mathbb{P}$-a.s.:

(a) $X_{s,t}^{x,\mu}$ is continuous in $t\in[s,\infty)$ and measurable in $x\in\mathbb{R}^d$;

(b) $\Lambda_{s,t}^{\mu}:=\mu\circ(X_{s,t}^{\cdot,\mu})^{-1}\in\mathscr{P}_2$ is continuous in $t\ge s$;

(c) $\mathbb{E}\int_s^t(|b(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})|+\|\sigma(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\|_{HS}^2)\mathrm{d}r<\infty$ and

$$X_{s,t}^{x,\mu}=x+\int_s^t b(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{d}r+\int_s^t\sigma(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{d}W_r,\quad t\ge s,x\in\mathbb{R}^d.$$

The image SDE (1.7) is called well-posed, if it has a unique solution for any $(s,\mu)\in[0,\infty)\times\mathscr{P}_2$.

To ensure the well-posedness of (1.7), we make the following assumption on b and σ.

(A) The progressively measurable coefficients $b(t,x,\mu)$ and $\sigma(t,x,\mu)$ are continuous in $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$, there exists $K\in L^q_{loc}([0,\infty)\to[0,\infty))$ for some $q>1$ such that $\mathbb{P}$-a.s. for any $t\ge0$,

A1 (2.1) $$|b(t,x,\mu)|^2+\|\sigma(t,x,\mu)\|_{HS}^2\le K(t)\big(1+|x|^2+\|\mu\|_2^2\big),\quad(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2,$$

A2 (2.2) $$\begin{aligned}&2\langle b(t,x,\mu)-b(t,y,\nu),x-y\rangle^++\|\sigma(t,x,\mu)-\sigma(t,y,\nu)\|_{HS}^2\\&\le K(t)\big(|x-y|^2+\mathbb{W}_2(\mu,\nu)^2\big),\quad(x,\mu),(y,\nu)\in\mathbb{R}^d\times\mathscr{P}_2.\end{aligned}$$

T2.1 **Theorem 2.1.** *Assume* **(A)**. *Then the image SDE* (1.7) *is well-posed, and the unique solution $X_{s,t}^{x,\mu}$ is jointly continuous in $(t,x)\in[s,\infty)\times\mathbb{R}^d$. Moreover:*

(1) *For any $p\ge1$, there exists an increasing function $C_p:[0,\infty)\to[0,\infty)$ such that*

EST (2.3) $$\mathbb{E}\sup_{r\in[s,t]}\big\{|X_{s,r}^{x,\mu}|^{2p}+\mu(|X_{s,r}^{\cdot,\mu}|^2)^p\big\}\le C_p(t)(1+|x|^{2p}+\|\mu\|_2^{2p}),$$

UPP (2.4) $$\mathbb{E}\sup_{r\in[s,t]}\big\{|X_{s,r}^{x,\mu}-X_{s,r}^{y,\nu}|^{2p}+\mathbb{W}_2(\Lambda_{s,r}^{\mu},\Lambda_{s,r}^{\nu})^{2p}\big\}\le C_p(t)(|x-y|^{2p}+\mathbb{W}_2(\mu,\nu)^{2p})$$

hold for all $0\le s\le t$, $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_2$. Consequently, $X_{s,t}^{x,\mu}$ is jointly continuous in $(t,x)\in[s,\infty)\times\mathbb{R}^d$.

(2) *When (b,σ) is deterministic, $\{(\Lambda_{s,t}^{\mu})_{t\ge s}:\mu\in\mathscr{P}_2\}$ is a diffusion process on $\mathscr{P}_2$ generated by $\mathscr{A}_t$; i.e. it is a continuous strong Markov process such that for any $\mu\in\mathscr{P}_2$ and any $f\in C_b^2(\mathscr{P}_2)$,*

$$f(\Lambda_{s,t}^{\mu})-f(\mu)-\int_s^t\mathscr{A}_rf(\Lambda_{s,r}^{\mu})\mathrm{d}r,\quad t\ge s$$

is a martingale.

(3) *When (b,σ) is deterministic, $\{(X^{x,\mu}_{s,t}, \Lambda^{\mu}_{s,t})_{t\ge s} : \mu \in \mathscr{P}_2\}$ is a diffusion on $\mathbb{R}^d \times \mathscr{P}_2$ generated by $\tilde{\mathscr{A}_t}$; i.e. it is a continuous strong Markov process such that for any $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$ and any $f \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$,*

$$f(X^{x,\mu}_{s,t}, \Lambda^{\mu}_{s,t}) - f(x,\mu) - \int_s^t \tilde{\mathscr{A}_r} f(X^{x,\mu}_{s,t}, \Lambda_{s,r})\mathrm{d}r, \quad t \ge s$$

is a martingale.

C2.2 **Corollary 2.2.** *Assume* **(A)**. *Then for any $\mathscr{F}_0$-measurable random variable X_0 with $\mu := \mathscr{L}_{X_0} \in \mathscr{P}_2$, the conditional distribution dependent SDE* (1.8) *has a unique solution which is given by $X_t = X^{X_0,\mu}_{0,t}$.*

Proof. By the independence of W and $\mathscr{F}_0$ and that X_0 is $\mathscr{F}_0$-measurable with distribution μ, it is easy to show $\mathscr{L}_{X^{X_0,\mu}_{0,t}|W} = \mu \circ (X^{\cdot,\mu}_{0,t})^{-1}$, which implies that $X_t := X^{X_0,\mu}_{0,t}$ solves (1.8). On the other hand, the uniqueness of (1.8) can be easily proved by using Itô's formula and condition (2.2). □

Remark 2.1. Consider the conditional Mckean-Vlasov SDE (or mean field SDE with common noise) (1.9). According to [5, Theorem 3.13 and Theorem 3.35] (see also [6, Theorem 3.2] for the weak existence), if b, σ, σ^0 are jointly continuous on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$, and there exists a constant $K > 0$ such that

$$|b(t,x,\mu)| + \|(\sigma,\sigma^0)(t,x,\mu)\| \le K(1+|x|+\|\mu\|_2), \quad \|\nabla_x b(t,x,\mu)\| + \|\nabla_x(\sigma,\sigma^0)(t,x,\mu)\| \le K$$

holds for any $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2$, then for any initial distribution in $\mathscr{P}_2$, the SDE (1.9) has a weak solution up to time T; if moreover $\sigma > 0, \sigma^0$ are constant and $b(t,x,\mu) = b^0(t,\mu) + cx$ for some constant c and $b^0(t,\mu)$ being bounded and Lipschitz continuous in μ uniformly in $t \in [0,T]$, then the weak solution is unique, hence the SDE is strongly well-posed. Obviously, these conditions are stronger than our assumption **(A)**.

Next, according to [10, Theorem 2.5 and Theorem 2.7], (1.9) has a weak solution provided one of the following assumptions hold:

(i) $\mathbb{E}|X_0|^p < \infty$ for some $p > 2$, b, σ, σ^0 are bounded and jointly continuous on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$;

(ii) b, σ, σ^0 are of the integral type

$$f(t,x,\mu) = \int_{\mathbb{R}^d} \tilde{f}(t,x,y)\mu(\mathrm{d}y)$$

for bounded measurable $\tilde{f}$, $(\tilde{\sigma},\tilde{\sigma}^0)(\tilde{\sigma},\tilde{\sigma}^0)^* \ge \lambda \mathrm{Id}$ for some constant $\lambda > 0$.

Moreover, by [10, Theorem 3.3], if σ and σ^0 do not depend on the distribution term such that the SDE

$$\mathrm{d}X^0_t = \sigma(t,X^0_t)\mathrm{d}W_t + \sigma^0(t,X^0_t)\mathrm{d}W^0_t$$

is well-posed, σ is invertible such that $\{\sigma^{-1}b\}(t,x,\mu)$ is bounded and Lipchitz continuous in μ with respect to the total variation norm uniformly in $(t,x)\in[0,T]\times\mathbb{R}^d$, then the weak (hence strong) solution of (1.9) is unique. Consequently, for the well-posedness these conditions only apply to the non-degenerate case with bounded $\sigma^{-1}b$, but the advantage is that the drift can be non-continuous in $(x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2$.

In the following two subsections, we prove Theorem 2.1(1) and (2)-(3) respectively.

2.1 Proof of Theorem 2.1(1)

Obviously, the uniqueness follows from (2.4). Below we prove (2.3), (2.4), joint continuity and the existence of the solution respectively.

(I) Estimate (2.3)**.** Let $(X_{s,t}^{x,\mu})_{x\in\mathbb{R}^d,t\ge s}$ be a solution of (1.7). We have

LLST
$$\|\Lambda_{s,t}^\mu\|_2^2=\|\mu\circ(X_{s,t}^{\cdot,\mu})^{-1}\|_2^2=\mu(|X_{s,t}^{\cdot,\mu}|^2),\quad t\ge s. \tag{2.5}$$

So, by (2.1) and Itô's formula, we may find out $\kappa\in L^1_{loc}([0,\infty)\to[0,\infty))$ such that

LLS
$$\mathrm{d}|X_{s,t}^{x,\mu}|^2\le\kappa(t)\big(1+|X_{s,t}^{x,\mu}|^2+\mu(|X_{s,t}^{\cdot,\mu}|^2)\big)\mathrm{d}t+2\big\langle X_{s,t}^{x,\mu},\sigma(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)\mathrm{d}W_t\big\rangle,\quad t\ge s. \tag{2.6}$$

Let $\gamma_t^x=2\sigma(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)^*X_{s,t}^{x,\mu}$. Since $(\Lambda_{s,t}^\mu)_{t\ge s}$ is an adapted continuous process on $\mathscr{P}_2$ and due to (2.1), $\sigma(t,x,\mu)$ has linear growth in x, there exists an increasing function $c:[0,\infty)\to[0,\infty)$ such that

$$\mu(|\gamma_t^\cdot|)\le c(t)\big\{1+\mu(|X_{s,t}^{\cdot,\mu}|^2)\big\}=c(t)\big\{1+\|\Lambda_{s,t}^\mu\|_2^2\big\}<\infty.$$

So, integrating (2.6) with respect to $\mu(\mathrm{d}x)$ leads to

LLS2
$$\mathrm{d}\mu(|X_{s,t}^{\cdot,\mu}|^2)\le\kappa(t)\big(1+2\mu(|X_{s,t}^{\cdot,\mu}|^2)\big)\mathrm{d}t+\langle\mu(\gamma_t^\cdot),\mathrm{d}W_t\rangle,\quad t\ge s. \tag{2.7}$$

Let $h_{s,t}:=\mathrm{e}^{2\int_s^t\kappa(r)\mathrm{d}r}$ and

$$\tau_n=\inf\big\{t\ge s:\mu(|X_{s,t}^{\cdot,\mu}|^2)+|X_{s,t}^{x,\mu}|^2\ge n\big\},\quad n\ge1.$$

Then (2.7) implies

OPP
$$\mu(|X_{s,t\wedge\tau_n}^{\cdot,\mu}|^2)\le h_{s,t}\|\mu\|_2^2+\int_s^t h_{r,t}\kappa(r)\mathrm{d}r+\int_s^{t\wedge\tau_n}h_{r,t}\langle\mu(\gamma_r^\cdot),\mathrm{d}W_r\rangle,\quad t\ge s, \tag{2.8}$$

so that by (2.6),

PL
$$\begin{aligned}&|X_{s,t\wedge\tau_n}^{x,\mu}|^2\le|x|^2+\int_s^{t\wedge\tau_n}\langle\mu(\gamma_r^\cdot),\mathrm{d}W_r\rangle\\&+\int_s^{t\wedge\tau_n}\kappa(r)\bigg\{1+|X_{s,r}^{x,\mu}|^2+h_{s,r}\|\mu\|_2^2+h_{s,r}\int_s^r\kappa(\theta)\mathrm{d}\theta+\int_s^r h_{\theta,r}\langle\gamma_\theta^x,\mathrm{d}W_\theta\rangle\bigg\}\mathrm{d}r\end{aligned} \tag{2.9}$$

holds for $t \geq s$. Moreover, (2.1) implies

XP
$$|\gamma_t^x|^2 = |2\sigma(t, X_{s,t}^{x,\mu}, \Lambda_{s,t}^{\mu})^* X_{s,t}^{x,\mu}|^2 \leq 4K(t)|X_{s,t}^{x,\mu}|^2\big(1 + |X_{s,t}^{x,\mu}|^2 + \mu(|X_{s,t}^{\cdot,\mu}|^2)\big). \tag{2.10}$$

This together with the Schwarz inequality gives

XP'
$$|\mu(\gamma_t^{\cdot})|^2 \leq 4K(t)\mu(|X_{s,t}^{\cdot,\mu}|^2)\big(1 + 2\mu(|X_{s,t}^{\cdot,\mu}|^2)\big). \tag{2.11}$$

Then for any $p \geq 1$ and $\varepsilon > 0$, there exists a constant $c = c(p, \varepsilon) > 0$ such that

$$\bigg(\int_s^{t\wedge\tau_n} |\mu(\gamma_r^{\cdot})|^2 \mathrm{d}r\bigg)^{\frac{p}{2}} \leq \varepsilon \sup_{r\in[s,t\wedge\tau_n]} \big\{\mu(|X_{s,r}^{\cdot,\mu}|^2)\big\}^p + c\int_s^{t\wedge\tau_n} K(r)\big(1 + \big\{\mu(|X_{s,r}^{\cdot,\mu}|^2)\big\}^p\big)\mathrm{d}r.$$

Combining this with (2.8) and using the BDG inequality, we may find an increasing function $C_0 : [0,\infty) \to [0,\infty)$ such that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} \big\{\mu(|X_{s,r}^{\cdot,\mu}|^2)\big\}^p\Big] \\
&\leq \frac{1}{2}\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} \big\{\mu(|X_{s,r}^{\cdot,\mu}|^2)\big\}^p\Big] + \frac{C_0(t)}{2}\bigg(1 + \|\mu\|_2^{2p} + \mathbb{E}\int_s^t \big\{\mu(|X_{s,r\wedge\tau_n}^{\cdot,\mu}|^2)\big\}^p \mathrm{d}r\bigg).
\end{aligned}$$

By Gronwall's inequality, this implies

OPP'
$$\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} \big\{\mu(|X_{s,r}^{\cdot,\mu}|^2)\big\}^p\Big] \leq C_0(t)\mathrm{e}^{\int_s^t C_0(r)\mathrm{d}r}(1 + \|\mu\|_2^{2p}). \tag{2.12}$$

Similarly, by (2.9)-(2.12) and the BDG inequality, we conclude that for any $p \geq 1$ there exist increasing functions $C_1, C_2 : [0,\infty) \to [0,\infty)$ such that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} |X_{s,r}^{x,\mu}|^{2p}\Big] &\leq C_1(t)\big(1 + |x|^{2p} + \|\mu\|_2^{2p}\big) + C_1(t)\mathbb{E}\bigg(\int_s^{t\wedge\tau_n} \kappa(r)|X_{s,t}^{x,\mu}|^2\mathrm{d}r\bigg)^p \\
&\quad + C_1(t)\mathbb{E}\bigg(\int_s^{t\wedge\tau_n} \big\{|\mu(\gamma_r^{\cdot})|^2 + |\gamma_r^x|^2\big\}\mathrm{d}r\bigg)^{\frac{p}{2}} \\
&\leq \frac{1}{2}\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} |X_{s,r}^{x,\mu}|^{2p}\Big] + C_2(t)\big(1 + |x|^{2p} + \|\mu\|_2^{2p}\big) + C_2(t)\mathbb{E}\int_s^t \kappa(r)|X_{s,r}^{x,\mu}|^{2p}\mathrm{d}r, \quad t \geq s.
\end{aligned}$$

By Grownwall's lemma, there exists an increasing function $Q : [0,\infty) \to (0,\infty)$ such that

$$\mathbb{E}\Big[\sup_{r\in[s,t\wedge\tau_n]} |X_{s,r}^{x,\mu}|^{2p}\Big] \leq Q(t)(1 + |x|^{2p} + \|\mu\|_2^{2p}), \quad t \geq s.$$

By letting $n \to \infty$ in this inequality and (2.12), we prove (2.3) for some increasing function $C_p : [0,\infty) \to [0,\infty)$.

(II) Estimate (2.4). Let $\pi \in \mathscr{C}(\mu,\nu)$ such that

PP1
$$\mathbb{W}_2(\mu,\nu)^2 = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\pi(\mathrm{d}x,\mathrm{d}y). \tag{2.13}$$

Then $\pi_{s,t} := \pi\circ(X_{s,t}^{\cdot,\mu}, X_{s,t}^{\cdot,\nu})^{-1} \in \mathscr{C}(\Lambda_{s,t}^\mu, \Lambda_{s,t}^\nu)$, so that

PP2
$$\begin{aligned} \mathbb{W}_2(\Lambda_{s,t}^\mu,\Lambda_{s,t}^\nu)^2 &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\pi_{s,t}(\mathrm{d}x,\mathrm{d}y)\\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |X_{s,t}^{x,\mu}-X_{s,t}^{y,\nu}|^2\pi(\mathrm{d}x,\mathrm{d}y) =: \ell_{s,t},\quad t\ge s. \end{aligned} \tag{2.14}$$

Thus, by (2.2) and Itô's formula, we obtain

UO
$$\begin{aligned} \mathrm{d}|X_{s,t}^{x,\mu}-X_{s,t}^{y,\nu}|^2 &\le K(t)\big\{|X_{s,t}^{x,\mu}-X_{s,t}^{y,\nu}|^2+\ell_{s,t}\big\}\mathrm{d}t\\ &\quad + 2\big\langle X_{s,t}^{x,\mu}-X_{s,t}^{y,\nu}, \{\sigma(r,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)-\sigma(r,X_{s,t}^{y,\nu},\Lambda_{s,t}^\nu)\}\mathrm{d}W_t\big\rangle,\quad t\ge s. \end{aligned} \tag{2.15}$$

Integrating both sides with respect to $\pi_{s,t}(\mathrm{d}x,\mathrm{d}y)$, and letting

$$\eta_t = 2\int_{\mathbb{R}^d\times\mathbb{R}^d} \{\sigma(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)-\sigma(t,X_{s,t}^{y,\nu},\Lambda_{s,t}^\nu)\}^*(X_{s,t}^{x,\mu}-X_{s,t}^{y,\nu})\,\pi(\mathrm{d}x,\mathrm{d}y),$$

we arrive at

$$\mathrm{d}\ell_{s,t} \le 2K(t)\ell_{s,t}\mathrm{d}t + \langle \eta_t, \mathrm{d}W_t\rangle,\quad t\ge s.$$

This together with $\ell_{s,s} = \mathbb{W}_2(\mu,\nu)^2$ implies

IP
$$\ell_{s,t} \le \mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{2\int_s^t K(r)\mathrm{d}r} + \int_s^t \mathrm{e}^{2\int_r^t K(\theta)\mathrm{d}\theta}\langle \eta_r, \mathrm{d}W_r\rangle,\quad t\ge s. \tag{2.16}$$

Moreover, **(A)** and the Schwarz inequality yield

EAT
$$\begin{aligned} |\eta_r|^2 &\le 4K(r)\ell_{s,r}\int_{\mathbb{R}^d\times\mathbb{R}^d}\big\{|X_{s,r}^{x,\mu}-X_{s,r}^{y,\nu}|^2+\mathbb{W}_2(\Lambda_{s,r}^\mu,\Lambda_{s,r}^\nu)^2\big\}\pi(\mathrm{d}x,\mathrm{d}y)\\ &\le 8K(r)\ell_{s,r}^2,\quad r\ge s. \end{aligned} \tag{2.17}$$

For given $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_2$, let

$$\tau_n = \inf\big\{t\ge s: \|\Lambda_{s,t}^\mu\|_2 + \|\Lambda_{s,t}^\nu\|_2 + |X_{s,t}^{x,\mu}| + |X_{s,t}^{y,\nu}| \ge n\big\}.$$

By (2.16), (2.17) and using the Hölder and BDG inequalities, we may find out increasing functions $c_1, c_2 : [0,\infty)\to[0,\infty)$ such that

$$\begin{aligned} \mathbb{E}\Big[\sup_{r\in[s,t]}\ell_{s,r\wedge\tau_n}^p\Big] &\le c_1(t)\mathbb{W}_2(\mu,\nu)^{2p} + c_1(t)\mathbb{E}\bigg(\int_s^{t\wedge\tau_n}|\eta_r|^2\mathrm{d}r\bigg)^{\frac{p}{2}}\\ &\le c_1(t)\mathbb{W}_2(\mu,\nu)^{2p} + c_2(t)\int_s^t \mathbb{E}\ell_{s,r\wedge\tau_n}^p\mathrm{d}r + \frac{1}{2}\mathbb{E}\Big[\sup_{r\in[s,t]}\ell_{s,r\wedge\tau_n}^p\Big],\quad t\ge s. \end{aligned}$$

Then it follows from Gronwall's lemma that

$$\mathbb{E}\Big[\sup_{r\in[s,t]}\ell^p_{s,r\wedge\tau_n}\Big]\le 2c_1(t)\mathrm{e}^{2tc_2(t)}\mathbb{W}_2(\mu,\nu)^{2p},\quad t\ge s.$$

By letting $n\to\infty$ and using Fatou's lemma, we obtain

ETB (2.18)
$$\mathbb{E}\Big[\sup_{r\in[s,t]}\ell^p_{s,r}\Big]\le 2c(t)\mathrm{e}^{2tc_p(t)}\mathbb{W}_2(\mu,\nu)^{2p},\quad t\ge s.$$

Similarly, by (2.15), (2.18), assumption **(A)** and using the Hölder and BDG inequality, for any $p\ge 1$ we find out increasing functions $K_1,K_2:[0,\infty)\to[0,\infty)$ such that

$$\begin{aligned}\mathbb{E}\Big[\sup_{r\in[s,t]}|X^{x,\mu}_{s,r\wedge\tau_n}-X^{y,\nu}_{s,r\wedge\tau_n}|^{2p}\Big]&\le|x-y|^{2p}+K_1(t)\mathbb{E}\int_s^{t\wedge\tau_n}K(r)\big\{|X^{x,\mu}_{s,r}-X^{y,\nu}_{s,r}|^{2p}+\ell^p_{s,r}\big\}\mathrm{d}r\\&\le|x-y|^{2p}+K_2(t)\mathbb{E}\int_s^t K(r)|X^{x,\mu}_{s,r\wedge\tau_n}-X^{y,\nu}_{s,r\wedge\tau_n}|^{2p}\,\mathrm{d}r+K_2(t)\mathbb{W}_2(\mu,\nu)^{2p},\quad t\ge s.\end{aligned}$$

Therefore, by Grownwall's lemma, there exists an increasing function $C:[0,\infty)\to(0,\infty)$ such that

$$\mathbb{E}\Big[\sup_{r\in[s,t]}|X^{x,\mu}_{s,r\wedge\tau_n}-X^{y,\nu}_{s,r\wedge\tau_n}|^{2p}\Big]\le C(t)\big(|x-y|^{2p}+\mathbb{W}_2(\mu,\nu)^{2p}\big),\quad t\ge s.$$

Letting $n\to\infty$ and using Fatou's lemma, we arrive at

$$\mathbb{E}\Big[\sup_{r\in[s,t]}|X^{x,\mu}_{s,r}-X^{y,\nu}_{s,r}|^{2p}\Big]\le C(t)\big(|x-y|^{2p}+\mathbb{W}_2(\mu,\nu)^{2p}\big),\quad t\ge s.$$

Combining this with (2.14) and (2.18), we prove (2.4) for some increasing function $C_p:[0,\infty)\to[0,\infty)$.

(III) Joint continuity of $X^{x,\mu}_{s,t}$ in (t,x). Let $K\in L^q_{loc}([0,\infty)\to[0,\infty))$ for some $q>1$. By (2.1), (2.3) and (2.4), for any $n,p\ge1$, there exist constants $C_1,C_2>0$ such that for any $n\ge t\ge r\ge s$, and $|x|,|y|\le n$,

PPO (2.19)
$$\begin{aligned}&\mathbb{E}\big(|X^{x,\mu}_{s,t}-X^{y,\mu}_{s,r}|^{2p}\big)\le 2^{2p-1}\big(\mathbb{E}|X^{x,\mu}_{s,t}-X^{y,\mu}_{s,t}|^{2p}+\mathbb{E}|X^{y,\mu}_{s,t}-X^{y,\mu}_{s,r}|^{2p}\big)\\&\le C_1|x-y|^{2p}+C_1\mathbb{E}\bigg|\int_r^t K(\theta)\sqrt{1+|X^{y,\mu}_{s,\theta}|^2+\mu(|X^{y,\mu}_{s,\theta}|^2)}\mathrm{d}\theta\bigg|^{2p}\\&\quad+C_1\mathbb{E}\bigg(\int_r^t K(\theta)\big\{1+|X^{y,\mu}_{s,\theta}|^2+\mu(|X^{y,\mu}_{s,\theta}|^2)\big\}\mathrm{d}\theta\bigg)^p\\&\le C_1|x-y|^{2p}+C_1\bigg(\int_r^t K(\theta)^q\mathrm{d}\theta\bigg)^{\frac{2p}{q}}\mathbb{E}\bigg|\int_r^t\big(1+|X^{y,\mu}_{s,\theta}|^2+\mu(|X^{y,\mu}_{s,\theta}|^2)\big)^{\frac{q}{2(q-1)}}\mathrm{d}\theta\bigg|^{\frac{2p(q-1)}{q}}\\&\quad+C_1\bigg(\int_r^t K(\theta)^q\bigg)^{\frac{p}{q}}\mathbb{E}\bigg(\int_r^t\big\{1+|X^{y,\mu}_{s,\theta}|^2+\mu(|X^{y,\mu}_{s,\theta}|^2)\big\}^{\frac{q}{q-1}}\mathrm{d}\theta\bigg)^{\frac{p(q-1)}{q}}\\&\le C_2\Big\{|x-y|^{2p}+(t-r)^{\frac{p(q-1)}{q}}\Big\}.\end{aligned}$$

By Kolmogorov's continuity criterion, for large enough $p>1$ this implies that $X^{x,\mu}_{s,t}$ has a $\mathbb{P}$-version jointly continuous in $(t,x)\in[s,n]\times\{x\in\mathbb{R}^d:|x|\le n\}$. Since $n\ge1$ is arbitrary, $X^{x,\mu}_{s,t}$ has a version jointly continuous in $(t,x)\in[s,\infty)\times\mathbb{R}^d$.

(IV) Existence of solution. It suffices to construct a solution up to an arbitrarily fixed time $T > 0$. To this end, we adopt an iteration argument as in [25].

(1) For fixed $(s, \mu) \in [0, T] \times \mathscr{P}_2$, let $\Lambda^{0,\mu}_{s,t} = \mu$ and $X^{0,x,\mu}_{s,t} = x$ for all $x \in \mathbb{R}^d$ and $t \geq s$.

(2) Assume that for some $n \in \mathbb{Z}_+$ we have constructed adapted $(X^{n,x,\mu}_{s,t})_{t\geq s, x\in\mathbb{R}^d}$ which is jointly continuous in $(t, x) \in [s, \infty) \times \mathscr{P}_2$, and satisfies

ESTN
$$\mathbb{E}\bigg[\sup_{r\in[s,t]} |X^{n,x,\mu}_{s,r}|^2\bigg] \leq c(t)(1 + |x|^2 + \|\mu\|_2^2), \quad t \geq s, x \in \mathbb{R}^d \tag{2.20}$$

for some increasing $c : [0, \infty) \to [0, \infty)$. Consequently, $\Lambda^{n,\mu}_{s,t} := \mu \circ (X^{n,\cdot,\mu}_{s,t})^{-1} \in \mathscr{P}_2$ is continuous in $t \geq s$. Indeed, by the Fubini theorem, (2.20) implies

$$\mathbb{E}\bigg[\mu\Big(\sup_{r\in[s,t]} |X^{n,\cdot,\mu}_{s,r}|^2\Big)\bigg] \leq c(t)(1 + 2\|\mu\|_2^2) < \infty, \quad t \geq s,$$

so that $\mathbb{P}$-a.s

$$\mu\Big(\sup_{r\in[s,t]} |X^{n,\cdot,\mu}_{s,r}|^2\Big) < \infty, \quad t \geq s.$$

Then by the dominated convergence theorem and the continuity of $X^{n,x,\mu}_{s,t}$ in $t \geq s$, we obtain $\mathbb{P}$-a.s.

$$\lim_{r\to t} \mathbb{W}_2(\Lambda^{n,\mu}_{s,r\vee s}, \Lambda^{n,\mu}_{s,t})^2 \leq \lim_{r\to t} \mu\big(|X^{n,\cdot,\mu}_{s,r\vee s} - X^{n,\cdot,\mu}_{s,t}|^2\big) = 0, \quad t \geq s.$$

(3) Let $(X^{n+1,x,\mu}_{s,t})_{t\geq s}$ solve the SDE

$$\mathrm{d}X^{n+1,x,\mu}_{s,t} = b(t, X^{n+1,x,\mu}_{s,t}, \Lambda^{n,\mu}_{s,t})\mathrm{d}t + \sigma(t, X^{n+1,x,\mu}_{s,t}, \Lambda^{n,\mu}_{s,t})\mathrm{d}W_t, \quad t \geq s, X^{n+1,x,\mu}_{s,s} = x.$$

By **(A)** and (2.20), it is easy to see that this SDE is well-posed, and when x varies the inequality (2.20) holds for $X^{n+1,x,\mu}_{s,t}$ replacing $X^{n,x,\mu}_{s,t}$ with possibly a different function $c : [0, \infty) \to [0, \infty)$. Moreover, as in **(III)**, **(A)** and (2.20) also imply the joint continuity of $X^{n+1,x,\mu}_{s,t}$ in $(t, x) \in [s, \infty) \times \mathbb{R}^d$. Consequently, as shown in step (2) that $\Lambda^{n+1,\mu}_{s,t} := \mu \circ (X^{n+1,\cdot,\mu}_{s,t})^{-1} \in \mathscr{P}_2$ is continuous in $t \geq s$.

Therefore, we have constructed a sequence $\{(X^{n,x,\mu}_{s,t}, \Lambda^{n,\mu}_{s,t})_{t\geq s, x\in\mathbb{R}^d}\}_{n\geq 0}$, which satisfies (2.20), $X^{n,x,\mu}_{s,t}$ is jointly continuous in $(t, x) \in [s, \infty) \times \mathbb{R}^d$, and $\mathbb{P}$-a.s.

AL
$$X^{n+1,x,\mu}_{s,t} = x + \int_s^t b(r, X^{n+1,x,\mu}_{s,r}, \Lambda^{n,\mu}_{s,r})\mathrm{d}r + \int_s^t \sigma(r, X^{n+1,x,\mu}_{s,r}, \Lambda^{n,\mu}_{s,r})\mathrm{d}W_r, \quad t \geq s, x \in \mathbb{R}^d. \tag{2.21}$$

The following lemma gives a constant $t_0 > 0$ independent of $(s, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathscr{P}_2$, such that $\{X^{n,x,\mu}_{s,\cdot}\}_{n\geq 1}$ is a Cauchy sequence in $L^2(\Omega \to C([s, s + t_0] \to \mathbb{R}^d); \mathbb{P})$.

LL1 **Lemma 2.3.** *Assume* **(A)**. *For fixed $T > 0$, there exists a constant $t_0 > 0$ such that*

$$\lim_{n,m\to\infty} \sup_{(s,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2} \frac{\mathbb{E}\sup_{t\in[s,s+t_0]} |X^{m,x,\mu}_{s,t} - X^{n,x,\mu}_{s,t}|^2}{1 + |x|^2 + \|\mu\|_2^2} = 0.$$

Proof. As in (2.14), we have $\mathbb{W}_2(\Lambda^{n,\mu}_{s,t},\Lambda^{n-1,\mu}_{s,t})^2 \le \mu(|X^{n,\cdot,\mu}_{s,t}-X^{n-1,\cdot,\mu}_{s,t}|^2)$ for $n\ge 1$. Combining this with (2.2) and Itô's formula, we obtain

$$\begin{aligned}
\mathrm{d}|X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t}|^2 &\le K(t)\Big\{|X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t}|^2+\mu(|X^{n,\cdot,\mu}_{s,t}-X^{n-1,\cdot,\mu}_{s,t}|^2)\Big\}\mathrm{d}t\\
&\quad+2\big\langle X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t},\{\sigma(t,X^{n+1,x,\mu}_{s,t},\Lambda^{n,\mu}_{s,t})-\sigma(t,X^{n,x,\mu}_{s,t},\Lambda^{n-1,\mu}_{s,t})\}\mathrm{d}W_t\big\rangle,\quad t\ge s.
\end{aligned}$$

So, by (2.2) and using the BDG inequality, we may find out constants $c_1,c_2>0$ such that

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[s,s+t_0]}|X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t}|^2\Big]\\
&\le\int_s^t K(r)\mathbb{E}\big[|X^{n+1,x,\mu}_{s,r}-X^{n,x,\mu}_{s,r}|^2+\mu(|X^{n,\cdot,\mu}_{s,r}-X^{n-1,\cdot,\mu}_{s,t}|^2)\big]\mathrm{d}r\\
&\quad+c_1\mathbb{E}\bigg(\int_s^t K(r)|X^{n+1,x,\mu}_{s,r}-X^{n,x,\mu}_{s,r}|^2\big\{|X^{n+1,x,\mu}_{s,r}-X^{n,x,\mu}_{s,t}|^2+\mu(|X^{n,\cdot,\mu}_{s,r}-X^{n-1,\cdot,\mu}_{s,t}|^2)\big\}\mathrm{d}r\bigg)^{\frac12}\\
&\le\frac{c_2}{2}\int_s^t K(r)\mathbb{E}\big[|X^{n+1,x,\mu}_{s,r}-X^{n,x,\mu}_{s,r}|^2+\mu(|X^{n,\cdot,\mu}_{s,r}-X^{n-1,\cdot,\mu}_{s,t}|^2)\big]\mathrm{d}r\\
&\quad+\frac12\mathbb{E}\Big[\sup_{t\in[s,s+t_0]}|X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t}|^2\Big],\quad t\ge s.
\end{aligned}$$

Since (2.20) holds for all n, this and Grownwall's inequality imply

TTT (2.22)
$$\mathbb{E}\sup_{r\in[s,t]}|X^{n+1,x,\mu}_{s,r}-X^{n,x,\mu}_{s,r}|^2\le c_2\int_s^t \mathrm{e}^{c_2\int_r^t K(\theta)\mathrm{d}\theta}\mathbb{E}\mu(|X^{n,\cdot,\mu}_{s,r}-X^{n-1,\cdot,\mu}_{s,r}|^2)\mathrm{d}r,\quad t\ge s \tag{2.22}$$

for all $(s,x)\in[0,T]\times\mathbb{R}^d$. Taking integral with respect to $\mu(\mathrm{d}x)$ leads to

$$\sup_{r\in[s,t]}\mathbb{E}\mu(|X^{n+1,\cdot,\mu}_{s,r}-X^{n,\cdot,\mu}_{s,r}|^2)\le c_2(t-s)\mathrm{e}^{c_2\int_s^t K(r)\mathrm{d}r}\sup_{r\in[s,t]}\mathbb{E}\mu(|X^{n,\cdot,\mu}_{s,r}-X^{n-1,\cdot,\mu}_{s,r}|^2),\quad t\ge s.$$

Now, taking $t_0>0$ such that

TVV (2.23)
$$\varepsilon:=c_2t_0\mathrm{e}^{c_2\int_0^{T+t_0}K(r)\mathrm{d}r}<1, \tag{2.23}$$

by iterating in n we arrive at

$$\begin{aligned}
&\sup_{s\in[0,T],t\in[s,s+t_0]}\mathbb{E}\mu(|X^{n+1,\cdot,\mu}_{s,t}-X^{n,\cdot,\mu}_{s,t}|^2)\le\varepsilon\sup_{s\in[0,T],t\in[s,s+t_0]}\mathbb{E}\mu(|X^{n,\cdot,\mu}_{s,t}-X^{n-1,\cdot,\mu}_{s,t}|^2)\\
&\le\cdots\le\varepsilon^n\sup_{s\in[0,T],t\in[s,s+t_0]}\mathbb{E}\mu(|X^{1,\cdot,\mu}_{s,t}-X^{0,\cdot,\mu}_{s,t}|^2)=c(x,\mu)\varepsilon^n<\infty,
\end{aligned}$$

where due to (2.20),

$$c(x,\mu):=\sup_{s\in[0,T]}\sup_{t\in[s,s+t_0]}\mathbb{E}\mu(|X^{1,\cdot,\mu}_{s,t}-x|^2)\le c(1+|x|^2+\|\mu\|_2^2)$$

for some constant $c>0$. Substituting this into (2.22) and using (2.23), we get

$$\sup_{s\in[0,T]}\mathbb{E}\sup_{t\in[s,s+t_0]}|X^{n+1,x,\mu}_{s,t}-X^{n,x,\mu}_{s,t}|^2\le c(1+|x|^2+\|\mu\|_2^2)\varepsilon^n,\quad n\ge 1.$$

This finishes the proof. $\square$

By Lemma 2.3, there exist a constant $t_0 > 0$ depending on $T > 0$, such that for any $s \in [0, T)$ we have a family of continuous processes

$$\{(X_{s,t}^{x,\mu})_{t\in[s,s+t_0]} : x \in \mathbb{R}^d, \mu \in \mathscr{P}_2\}$$

which are measurable in x and

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{r\in[s,s+t_0]} \big(|X_{s,r}^{n,x,\mu} - X_{s,t}^{x,\mu}|^2 + \mu(|X_{s,r}^{n,\cdot,\mu} - X_{s,t}^{\cdot,\mu}|^2)\big)\Big] = 0.$$

Letting $\Lambda_{s,t}^{\mu} = \mu \circ (X_{s,t}^{\cdot,\mu})^{-1}$, by this and (2.14) we obtain

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{r\in[s,s+t_0]} \mathbb{W}_2(\Lambda_{s,t}^{n,\mu}, \Lambda_{s,t}^{\mu})^2\Big] \le \mathbb{E}\Big[\sup_{r\in[s,s+t_0]} \mu(|X_{s,r}^{n,\cdot,\mu} - X_{s,t}^{\cdot,\mu}|^2)\Big] = 0.$$

Thus, the continuity of $\Lambda_{s,t}^{n,\mu}$ in $t \in [s, s+t_0]$ implies that of $\Lambda_{s,t}^{\mu}$; due to (2.20) we may find out a constant $c_1 > 0$ such that

AL2
$$\mathbb{E}\Big[\sup_{t\in[s,s+t_0]} \big\{\mu(|X_{s,t}^{\cdot,\mu}|^2) + |X_{s,t}^{x,\mu}|^2\big\}\Big] \le c_1\big(1 + |x|^2 + \|\mu\|_2^2\big), \quad (s, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathscr{P}_2; \tag{2.24}$$

and finally, by assumption **(A)** we may let $n \to \infty$ in (2.21) to derive

$$X_{s,t}^{x,\mu} = x + \int_s^t b(r, X_{s,r}^{x,\mu}, \Lambda_{s,r}^{\mu})\mathrm{d}r + \int_s^t \sigma(r, X_{s,r}^{x,\mu}, \Lambda_{s,r}^{\mu})\mathrm{d}W_r, \quad t \in [s, s+t_0], x \in \mathbb{R}^d.$$

So, when $T \le s + t_0$ we have solved the SDE up to time T.

In the case that $T > s + t_0$, let $\bar{s} = s + t_0, \bar{x} = X_{s,s+t_0}^{x,\mu}$ and $\bar{\mu} = \Lambda_{s,s+t_0}^{\mu}$. Since given $\mathscr{F}_{s+t_0}$ the process $(W_t - W_{\bar{s}})_{t\ge\bar{s}}$ is an m-dimensional Brownian motion, and $(\bar{x}, \bar{\mu})$ is given as well, as in above we may construct a solution $(X_{\bar{s},t}^{\bar{x},\bar{\mu}}, \Lambda_{\bar{s},t}^{\bar{\mu}})_{t\in[\bar{s},\bar{s}+t_0]}$ for (1.7) with $\bar{s}$ replacing s. Then extending $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^{\mu})$ to $t \in [\bar{s}, \bar{s} + t_0]$ by letting

$$X_{s,t}^{x,\mu} = X_{\bar{s},t}^{\bar{x},\bar{\mu}}, \quad \Lambda_{s,t}^{\mu} = \Lambda_{\bar{s},t}^{\bar{\mu}}, \quad t \in [\bar{s}, \bar{s} + t_0],$$

we see that $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^{\mu})_{t\in[s,s+2t_0]}$ solves (1.7) up to time $\bar{s}+t_0 = s+2t_0$. Runing this procedure for k times until $s + kt_0 \ge T$, we construct a solution to (1.7) up to time T.

2.2 Proof of Theorem 2.1(2)-(3)

We first establish Itô's formula for the diffusion process $(\Lambda_{s,t}^{\mu})_{t\ge s}$. To this end, we need the following chain rule for the L-derivative, which is essentially due to [4, Theorem 6.5] where the reference probability space is Polish, see also [9, Proposition A.2] for general probability space but bounded random variables $\{\xi_s\}_{s\in[0,\varepsilon]}$ (note that D_k therein is compact).

L2.0 **Lemma 2.4.** *Let $\{\xi_s\}_{s\in[0,\varepsilon]}$ for some $\varepsilon > 0$ be a family of square integrable random variables on $\mathbb{R}^d$ with respect to a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{L}_{\xi_s}$ denote the law of ξ_s. If*

$$\xi_0' := \lim_{s\downarrow 0} \frac{\xi_s - \xi_0}{s}$$

exists in $L^2(\Omega \to \mathbb{R}^d; \mathbb{P})$, then for any $f \in C^1(\mathscr{P}_2)$,

$$\lim_{s\downarrow 0} \frac{f(\mathscr{L}_{\xi_s}) - f(\mathscr{L}_{\xi_0})}{s} = \mathbb{E}\langle Df(\mathscr{L}_{\xi_0})(\xi_0), \xi_0'\rangle.$$

Proof. By a standard extension argument, we may and do assume that $(\Omega, \mathscr{F}, \mathbb{P})$ is atomless. For instance, we enlarge $(\Omega, \mathscr{F}, \mathbb{P})$ by $(\Omega \times [0,1], \mathscr{F} \times \mathscr{B}([0,1]), \mathbb{P} \times \mathrm{d}r)$ and use $\tilde{\xi}_s$ to replace ξ_s, where $\tilde{\xi}_s(\omega, r) := \xi_s(\omega)$ for $(\omega, r) \in \Omega \times [0,1]$, so that $\mathscr{L}_{\tilde{\xi}_s}$ under $\mathbb{P} \times \mathrm{d}r$ coincides with $\mathscr{L}_{\xi_s}$ under $\mathbb{P}$. Then the proof is completely similar to that of [**?**, Proposition 3.1] for ξ_s replacing $X + sY$. $\square$

L2.1 **Lemma 2.5** (Itô's formula)**.** *Assume* **(A)** *and let* $\{\Lambda^\mu_{s,t} = \mu \circ (X^{\cdot,\mu}_{s,t})^{-1}\}_{t \geq s}$ *for the solution to* (1.7). *Then for any* $f \in C^2_b(\mathscr{P}_2)$,

$$\mathrm{d}f(\Lambda^\mu_{s,t}) = (\mathscr{A}_t f)(\Lambda^\mu_{s,t})\mathrm{d}t + \bigg\langle \int_{\mathbb{R}^d} \big\{\sigma(t, x, \Lambda^\mu_{s,t})^*(Df)(\Lambda^\mu_{s,t})(x)\big\}\mu(\mathrm{d}x), \mathrm{d}W_t \bigg\rangle, \quad t \geq s.$$

Proof. For any $t \geq s$ and small $\varepsilon > 0$, let

$$\xi_r = (1-r)X^{\cdot,\mu}_{s,t} + rX^{\cdot,\mu}_{s,t+\varepsilon} : \mathbb{R}^d \to \mathbb{R}^d, \quad r \in [0,1].$$

Then $\mu \circ \xi_r^{-1}$ is the law of ξ_r on the probability space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \mu)$. By (2.3),

$$\sup_{r \in [0,1]} \mathbb{E}\|\mu \circ \xi_r^{-1}\|_2^2 \leq \mathbb{E}\Big[\sup_{r \in [0,1]} \mu(|\xi_r|^2)\Big] < \infty, \quad t \geq s.$$

Moreover, $\xi_r' := \frac{\mathrm{d}}{\mathrm{d}r}\xi_r = X^{\cdot,\mu}_{s,t+\varepsilon} - X^{\cdot,\mu}_{s,t}$ exists in $L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$. So, Lemma 2.4 implies

BG1 (2.25)
$$\begin{aligned} f(\Lambda^\mu_{s,t+\varepsilon}) - f(\Lambda^\mu_{s,t}) &= f(\mu \circ \xi_1^{-1}) - f(\mu \circ \xi_0^{-1}) = \int_0^1 \Big(\frac{\mathrm{d}}{\mathrm{d}r} f(\mu \circ \xi_r^{-1})\Big)\mathrm{d}r \\ &= \int_{\mathbb{R}^d \times [0,1]} \big\langle Df(\mu \circ \xi_r^{-1})(\xi_r^x), X^{x,\mu}_{s,t+\varepsilon} - X^{x,\mu}_{s,t}\big\rangle \mu(\mathrm{d}x)\mathrm{d}r \\ &= \int_{\mathbb{R}^d} I_1(x)\mu(\mathrm{d}x) + \int_{\mathbb{R}^d \times [0,1]} I_2(x,r)\mu(\mathrm{d}x)\mathrm{d}r + \int_{\mathbb{R}^d \times [0,1]} I_3(x,r)\mu(\mathrm{d}x)\mathrm{d}r, \end{aligned}$$

where, since $\mu \circ \xi_0^{-1} = \Lambda^\mu_{s,t}$,

$$\begin{aligned} I_1(x) &:= \langle Df(\Lambda^\mu_{s,t})(X^{x,\mu}_{s,t}), X^{x,\mu}_{s,t+\varepsilon} - X^{x,\mu}_{s,t}\rangle, \\ I_2(x,r) &:= \big\langle Df(\mu \circ \xi_r^{-1})(\xi_r^x) - Df(\mu \circ \xi_0^{-1})(\xi_r^x), X^{x,\mu}_{s,t+\varepsilon} - X^{x,\mu}_{s,t}\big\rangle, \\ I_3(x,r) &:= \big\langle Df(\Lambda^\mu_{s,t})(\xi_r^x) - Df(\Lambda^\mu_{s,t})(\xi_0^x), X^{x,\mu}_{s,t+\varepsilon} - X^{x,\mu}_{s,t}\big\rangle. \end{aligned}$$

Below, we calculate $I_1(x), I_2(x)$ and $I_3(x)$ respectively.

Firstly, by (1.7) and $f \in C^2_b(\mathscr{P}_2)$, we have

BG2 (2.26)
$$\begin{aligned} I_1(x) &= \int_t^{t+\varepsilon} \big\langle (Df)(\Lambda^\mu_{s,t})(X^{x,\mu}_{s,t}), \mathrm{d}X^{x,\mu}_{s,u}\big\rangle = \int_t^{t+\varepsilon} \big\langle (Df)(\Lambda^\mu_{s,u})(X^{x,\mu}_{s,u}), \mathrm{d}X^{x,\mu}_{s,u}\big\rangle + \mathrm{o}(\varepsilon) \\ &= \int_t^{t+\varepsilon} \big\langle (Df)(\Lambda^\mu_{s,u})(X^{x,\mu}_{s,u}), b(u, X^{x,\mu}_{s,u}, \Lambda^\mu_{s,u})\big\rangle \mathrm{d}u \\ &\quad + \int_t^{t+\varepsilon} \big\langle (Df)(\Lambda^\mu_{s,u})(X^{x,\mu}_{s,u}), \sigma(u, \mathrm{d}X^{x,\mu}_{s,u}, \Lambda^\mu_{s,u})\mathrm{d}W_u\big\rangle + \mathrm{o}(\varepsilon), \end{aligned}$$

where and in the following, $\mathrm{o}(\varepsilon)$ means ε-dependent (real, vector or matrix valued) random variables satisfying $\lim_{\varepsilon\to 0}\varepsilon^{-1}|\mathrm{o}(\varepsilon)|=0$.

Next, (1.7) implies

**0

$$(X^{x,\mu}_{s,t+\varepsilon}-X^{x,\mu}_{s,t})\otimes(X^{y,\mu}_{s,t+\varepsilon}-X^{y,\mu}_{s,t})=\int_t^{t+\varepsilon}\sigma(u,X^{x,\mu}_{s,u},\Lambda^\mu_{s,u})\sigma(u,X^{y,\mu}_{s,u},\Lambda^\mu_{s,u})^*\mathrm{d}u+\mathrm{o}(\varepsilon). \tag{2.27}$$

Combining this with $f\in C_b^2(\mathscr{P}_2)$, we deduce from Lemma 2.4 and $\xi'_\theta=X^{\cdot,\mu}_{s,t+\varepsilon}-X^{\cdot,\mu}_{s,t}$ that up to an error term $\mathrm{o}(\varepsilon)$,

BG3

$$\begin{aligned}
I_2(x,r)&=\int_0^r\mathrm{d}\theta\int_{\mathbb{R}^d}\big\langle(D^2f)(\mu\circ\xi_\theta^{-1})(\xi_r^x,\xi_\theta^y),(X^{x,\mu}_{s,t+\varepsilon}-X^{x,\mu}_{s,t})\otimes(\xi^y_\theta)'\big\rangle\mu(\mathrm{d}y)\\
&=\int_0^r\mathrm{d}\theta\int_t^{t+\varepsilon}\mathrm{d}u\int_{\mathbb{R}^d}\big\langle(D^2)f(\mu\circ\xi_\theta^{-1})(\xi_r^x,\xi_\theta^y),\sigma(u,X^{x,\mu}_{s,u},\Lambda^\mu_{s,u})\sigma(u,X^{y,\mu}_{s,u},\Lambda^\mu_{s,u})^*\big\rangle\mu(\mathrm{d}y)\\
&=r\int_t^{t+\varepsilon}\mathrm{d}u\int_{\mathbb{R}^d}\big\langle(D^2f)(\Lambda^\mu_{s,u})(X^{x,\mu}_{s,u},X^{y,\mu}_{s,u}),\sigma(u,X^{x,\mu}_{s,u},\Lambda^\mu_{s,u})\sigma(u,X^{y,\mu}_{s,u},\Lambda^\mu_{s,u})^*\big\rangle\mu(\mathrm{d}y).
\end{aligned}\tag{2.28}$$

Similarly, by using (2.27) with $x=y$, we obtain that up to an error term $\mathrm{o}(\varepsilon)$,

$$\begin{aligned}
I_3(x,r)&=\big\langle(Df)(\mu\circ\xi_0^{-1})(\xi_r^x)-(Df)(\mu\circ\xi_0^{-1})(\xi_0^x),X^{x,\mu}_{s,t+\varepsilon}-X^{x,\mu}_{s,t}\big\rangle\\
&=\int_0^r\big\langle\nabla\{(Df)(\Lambda^\mu_{s,t})\}(\xi^x_\theta),(X^{x,\mu}_{s,t+\varepsilon}-X^{x,\mu}_{s,t})\otimes(X^{x,\mu}_{s,t+\varepsilon}-X^{x,\mu}_{s,t})\big\rangle\mathrm{d}r\\
&=r\int_t^{t+\varepsilon}\big\langle\nabla\{(Df)(\Lambda^\mu_{s,u})\}(X^{x,\mu}_{s,u}),(\sigma\sigma^*)(t,X^{x,\mu}_{s,u},\Lambda^\mu_{s,u})\big\rangle\mathrm{d}u.
\end{aligned}$$

Combining this with (2.25)-(2.28), we arrive at

$$\begin{aligned}
&\mathrm{d}f(\Lambda^\mu_{s,t})-\int_{\mathbb{R}^d}\big\langle(Df)(\Lambda^\mu_{s,t})(x),\sigma(t,x,\Lambda^\mu_{s,t})\mathrm{d}W_t\big\rangle\mu(\mathrm{d}x)\\
&=\bigg(\int_{\mathbb{R}^d}\big\langle(Df)(\Lambda^\mu_{s,t})(X^{x,\mu}_{s,t}),b(t,X^{x,\mu}_{s,t},\Lambda^\mu_{s,t})\big\rangle\mu(\mathrm{d}x)\bigg)\mathrm{d}t\\
&\quad+\bigg(\frac12\int_{\mathbb{R}^d}\big\langle\nabla\{(Df)(\Lambda^\mu_{s,t})\}(X^{x,\mu}_{s,t}),(\sigma\sigma^*)(t,X^{x,\mu}_{s,t},\Lambda^\mu_{s,t})\big\rangle\Big\}\mu(\mathrm{d}x)\bigg)\mathrm{d}t\\
&\quad+\bigg(\frac12\int_{\mathbb{R}^d\times\mathbb{R}^d}\big\langle(D^2f)(\Lambda^\mu_{s,t})(X^{x,\mu}_{s,t},X^{y,\mu}_{s,t}),\sigma(t,X^{x,\mu}_{s,t},\Lambda^\mu_{s,t})\sigma(t,X^{y,\mu}_{s,t},\Lambda^\mu_{s,t})^*\big\rangle\mu(\mathrm{d}x)\mu(\mathrm{d}y)\bigg)\mathrm{d}t\\
&=(\mathscr{A}_tf)(\Lambda^\mu_{s,t})\mathrm{d}t.
\end{aligned}$$

Then the proof is finished. □

Remark 2.2. We note that under a moment condition the the Iô's formula for the conditional distribution has been established in [5]. More precisely, consider the Itô process

$$\mathrm{d}X_t=B_t\mathrm{d}t+\Sigma_t\mathrm{d}W_t+\Sigma_t^0\mathrm{d}W_t^0,$$

where B_t, Σ_t and Σ_t^0 are progressively measurable such that

$$\boxed{\texttt{4M}} \quad (2.29) \qquad \mathbb{E}\int_0^T \big\{|B_t|^2 + \|\Sigma_t\|^4 + \|\Sigma_t^0\|^4\big\}\mathrm{d}t < \infty,$$

then for any $f \in C_b^2(\mathscr{P}_2)$, $\mu_t := \mathscr{L}_{X_t|W}$ satisfies the Itô formula

$$\begin{aligned} f(\mu_t) - f(\mu_0) &= \int_0^t \mathbb{E}\big[\langle Df(\mu_s)(X_s), B_s\rangle|W\big]\mathrm{d}s + \int_0^t \big\langle \mathbb{E}\big[Df(\mu_s)(X_s)\big|W\big], \Sigma_s \mathrm{d}W_s\big\rangle \\ &+ \frac{1}{2}\int_0^t \mathbb{E}\big[\mathrm{tr}\big\{\nabla\big(Df(\mu_s)\big)(X_s)\big(\Sigma_s\Sigma_s^* + \Sigma_s^0(\Sigma_s^0)^*\big)\big|W\big]\mathrm{d}s \\ &+ \frac{1}{2}\int_0^t \mathrm{d}s \int_{\mathbb{R}^d} \mathbb{E}\big[\mathrm{tr}\big\{D^2 f(\mu_s)(X_s, z)\Sigma_s\Sigma_s^*\big\}\big|W\big]\mu_s(\mathrm{d}z), \quad t \in [0,T]. \end{aligned}$$

So, if (2.29) holds for $B_t := b(t, X_t, \mathscr{L}_{X_t|W}), \Sigma_t := \sigma(t, X_t, \mathscr{L}_{X_t|W})$ and $\Sigma_t^0 = 0$, then Lemma 2.5 follows. However, our condition **(A)** is not enough to ensure (2.29) unless the initial value satisfies $\mathbb{E}|X_0|^4 < \infty$ (as σ has linear growth) and $K \in L^2([0,T])$.

Proof of Theorem 2.1(2)-(3). By the uniqueness result in Theorem 2.1, we have the flow property

$$\boxed{\texttt{MK}} \quad (2.30) \qquad X_{s,t}^{x,\mu} = X_{r,t}^{X_{s,r}^{x,\mu}, \Lambda_{s,r}^{\mu}}, \quad \Lambda_{s,t}^{\mu} = \Lambda_{r,t}^{\Lambda_{s,r}^{\mu}}, \quad 0 \le s \le r \le t,$$

which implies that both $(\Lambda_{s,t}^\mu)_{t\ge s}$ and $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)_{t\ge s}$ are Markov processes.

Next, by (2.4), these two Markov processes are Feller and hence, strong Markovian. Therefore, Theorem 2.1(2) follows from Lemma 2.5.

Finally, for any $f \in C_b^{2,2}(\mathbb{R}^d, \mathscr{P}_2)$, Lemma 2.5 and the classical Itô's formula for the semimartingale $(X_{s,t}^{x,\mu})_{t\ge s}$ imply

$$\boxed{\texttt{ITO}} \quad (2.31) \qquad \begin{aligned} \mathrm{d}f(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu) &= (\tilde{\mathscr{A}}_t f)(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)\mathrm{d}t + \big\langle \nabla f(\cdot, \Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu}), \sigma(t, X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)\mathrm{d}W_t\big\rangle \\ &+ \int_{\mathbb{R}^d} \big\langle Df(X_{s,t}^{x,\mu}, \cdot)(\Lambda_{s,t}^\mu)(x), \sigma(t, x, \Lambda_{s,t}^\mu)\mathrm{d}W_t\big\rangle \mu(\mathrm{d}x), \quad t \ge s. \end{aligned}$$

This proves Theorem 2.1(3). □

3 Feynman-Kac formula for PDEs on $\mathbb{R}^d \times \mathscr{P}_2$

In this section, we solve the PDEs (1.10) and (1.11) by using $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)_{0\le s\le t\le T}$. As mentioned in Abstract that when $V = 0$ they are included by the Master equations investigated in the literature of mean filed games with common noise.

A function on U on $[0,T] \times \mathbb{R}^d \times \mathscr{P}_2$ is called a solution to (1.10), if $U(t,x,\mu)$ is differentiable in t and $U(t,\cdot,\cdot) \in C^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$ such that (1.10) holds. If moreover $U(t,x,\mu)$ does not depend on x, it is called a solution to (1.11). We first introduce the following class $C_b^{0,2,2}([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$.

Definition 3.1. Let f be a real, vector or matrix valued function on $[0,T]\times\mathbb{R}^k\times\mathscr{P}_2$ for some $k\geq 1$. We write $f\in C_b^{0,2,2}([0,T]\times\mathbb{R}^k\times\mathscr{P}_2)$, if f is jointly continuous, $f(t,\cdot,\cdot)\in C_b^{2,2}(\mathbb{R}^k\times\mathscr{P}_2)$ for every $t\in[0,T]$, and all derivatives

$$\begin{aligned}&\nabla f(t,x,\mu),\quad \nabla^2 f(t,x,\mu),\quad Df(t,x,\mu)(y),\\ &D\{\nabla f(t,x,\mu)\}(y),\quad \nabla\{Df(t,x,\mu)(\cdot)\}(y),\quad D^2 f(t,x,\mu)(y,z)\end{aligned}$$

are bounded and jointly continuous in corresponding arguments. If moreover $f(t,x,\mu)$ does not depend on x, we denote $f\in C_b^{0,2}([0,T]\times\mathscr{P}_2)$.

T4.1 **Theorem 3.1.** *Assume that $b,\sigma\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ are deterministic.*

(1) *For any $\Phi\in C_b^{2,2}(\mathbb{R}^d\times\mathscr{P}_2)$, $F\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$, and bounded $V\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$,*

$$U(t,x,\mu):=\mathbb{E}\bigg[\Phi(X_{t,T}^{x,\mu},\Lambda_{t,T}^{\mu})\mathrm{e}^{\int_t^T V(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})dr}+\int_t^T F(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\mathrm{e}^{\int_t^r V(\theta,X_{t,\theta}^{x,\mu},\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta}\mathrm{d}r\bigg]$$

is the unique solution of (1.10) *in the class $C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ with $\partial_t U\in C([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$.*

(2) *For any $\Phi\in C_b^{2}(\mathbb{R}^d\times\mathscr{P}_2)$, $F\in C_b^{0,2}([0,T]\times\mathscr{P}_2)$, and bounded $V\in C_b^{0,2}([0,T]\times\mathscr{P}_2)$,*

$$U(t,\mu):=\mathbb{E}\Big[\Phi(\Lambda_{t,T}^{\mu})\mathrm{e}^{\int_t^T V(r,\Lambda_{t,r}^{\mu})\mathrm{d}r}+\int_t^T F(r,\Lambda_{t,r}^{\mu})\mathrm{e}^{\int_t^r V(\theta,\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta}\mathrm{d}r\Big]$$

is the unique solution of (1.11) *in the class $C_b^{0,2}([0,T]\times\mathscr{P}_2)$ with $\partial_t U\in C([0,T]\times\mathscr{P}_2)$*

Remark 3.1. When σ is constant, $b(t,x,\mu)$ and $\nabla_x b(t,x,\mu)$ are in the class $C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ and $V=0$, [5, Theorem 5.45] implies that $U(t,x,\mu)$ given in Theorem 3.1(1) solves the Master equation (1.10) with $V=0$. These conditions are stronger than those in Theorem 3.1.

Proof of Theorem 3.1. Since $\tilde{\mathscr{A}_t}F(x,\mu)=\mathscr{A}_tF(\mu)$ holds for $F\in C_b^2(\mathscr{P}_2)$, (2) follows from (1). So, it suffices to prove Theorem 3.1(1).

If $U\in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ is a solution of (1.10), then (2.31) yields

$$\begin{aligned}\mathrm{d}U(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^{\mu})&=(\partial_t+\tilde{\mathscr{A}_t})U(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^{\mu})\mathrm{d}t+\mathrm{d}M_t\\ &=\mathrm{d}M_t-\big(VU+F\big)(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^{\mu})\mathrm{d}t,\quad t\in[s,T]\end{aligned}$$

for some martingale $(M_t)_{t\in[s,T]}$. Thus, the process

$$\eta_t:=U(t,X_{s,t}^{x,\mu},\Lambda_{s,t}^{\mu})\mathrm{e}^{\int_s^t V(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{d}r}+\int_s^t F(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{e}^{\int_s^r V(\theta,X_{s,\theta}^{x,\mu},\Lambda_{s,\theta}^{\mu})\mathrm{d}\theta}\mathrm{d}r,\quad t\in[s,T]$$

satisfies

$$\mathrm{d}\eta_t=\mathrm{e}^{\int_s^t V(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{d}r}\mathrm{d}M_t,\quad t\in[s,T].$$

So,

$$
\begin{aligned}
U(s,x,\mu) &= \mathbb{E}\eta_s = \mathbb{E}\eta_T \\
&= \mathbb{E}\bigg[\Phi(X_{s,T}^{x,\mu},\Lambda_{s,T}^{\mu})\mathrm{e}^{\int_s^T V(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})dr} + \int_s^T F(r,X_{s,r}^{x,\mu},\Lambda_{s,r}^{\mu})\mathrm{e}^{\int_s^r V(\theta,X_{s,\theta}^{x,\mu},\Lambda_{s,\theta}^{\mu})\mathrm{d}\theta}\mathrm{d}r\bigg]
\end{aligned}
$$

as claimed in Theorem 3.1(1).

On the other hand, let U be given in Theorem 3.1(1). For any $t \in [0,T)$ and $\varepsilon \in (0,T-t)$, by (2.30) and the formula of $U(t,x,\mu)$ in Theorem 3.1(1),

$$
U(t,x,\mu) - \mathbb{E}\big[U(t+\varepsilon, X_{t,t+\varepsilon}^{x,\mu}, \Lambda_{t,t+\varepsilon}^{x})\big] = I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon)
$$

holds for

$$
\begin{aligned}
I_1(\varepsilon) &:= \mathbb{E}\Big[\Phi(X_{t+\varepsilon,T}^{X_{t,t+\varepsilon}^{x,\mu},\Lambda_{t,t+\varepsilon}^{\mu}}, \Lambda_{t+\varepsilon,T}^{\Lambda_{t,t+\varepsilon}^{\mu}})\big(\mathrm{e}^{\int_t^T V(r,X_{r,T}^{x,\mu},\Lambda_{r,T}^{\mu})dr} - \mathrm{e}^{\int_{t+\varepsilon}^T V(r,X_{r,T}^{x,\mu},\Lambda_{r,T}^{\mu})dr}\big)\Big], \\
I_2(\varepsilon) &:= \mathbb{E}\bigg[\int_t^{t+\varepsilon} F(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\mathrm{e}^{\int_t^r V(\theta,X_{t,\theta}^{x,\mu},\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta}\mathrm{d}r\bigg], \\
I_3(\varepsilon) &:= \mathbb{E}\bigg[\int_{t+\varepsilon}^T F(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\big(\mathrm{e}^{\int_t^r V(\theta,X_{t,\theta}^{x,\mu},\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta} - \mathrm{e}^{\int_{t+\varepsilon}^r V(\theta,X_{t,\theta}^{x,\mu},\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta}\big)\mathrm{d}r\bigg].
\end{aligned}
$$

Therefore,

LMT

$$
\begin{aligned}
&\lim_{\varepsilon\to 0}\frac{U(t,x,\mu) - \mathbb{E}[U(t+\varepsilon, X_{t,t+\varepsilon}^{x,\mu}, \Lambda_{t,t+\varepsilon}^{x})]}{\varepsilon} \\
&= V(t,x,\mu)\mathbb{E}\big[\Phi(X_{t,T}^{x,\mu},\Lambda_{t,T}^{\mu})\mathrm{e}^{\int_t^T V(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\mathrm{d}r}\big] + F(t,x,\mu) \\
&\quad + V(t,x,\mu)\mathbb{E}\bigg[\int_t^T F(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\mathrm{e}^{\int_t^r V(\theta,X_{t,\theta}^{x,\mu},\Lambda_{t,\theta}^{\mu})\mathrm{d}\theta}\bigg] \\
&= (VU+F)(t,x,\mu).
\end{aligned} \tag{3.1}
$$

By Proposition 3.2 below, $U \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$ and $\tilde{\mathscr{A}}_t U(t,x,\mu)$ is continuous in (t,x,μ). Then (2.31) implies

$$
\mathbb{E}[U(t+\varepsilon, X_{t,t+\varepsilon}^{x,\mu}, \Lambda_{t,t+\varepsilon}^{x})] = U(t+\varepsilon,x,\mu) + \mathbb{E}\int_t^{t+\varepsilon}\tilde{\mathscr{A}}_r U(r,X_{t,r}^{x,\mu},\Lambda_{t,r}^{\mu})\mathrm{d}r.
$$

Combining this with (3.1) we arrive at

$$
-\partial_t U(t,x,\mu) = \lim_{\varepsilon\to 0}\frac{U(t,x,\mu) - U(t+\varepsilon,x,\mu)}{\varepsilon} = \tilde{\mathscr{A}}_t U(t,x,\mu) + (UV+F)(t,x,\mu).
$$

Therefore, U solves (1.10) with continuous $\tilde{\mathscr{A}}_t U$. $\square$

The remainder of this section devotes to the proof of the following result.

P4.2 **Proposition 3.2.** *Under conditions of Theorem* 3.1 *and let* U *be given in Theorem* 3.1(1)*. Then* $U \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$*, so that* $\tilde{\mathscr{A}}_t U$ *is continuous on* $[0,T]\times\mathbb{R}^d\times\mathscr{P}_2$.

We first introduce some notations which will be used in calculations.

(a) For $f \in C^2(\mathbb{R}^d)$,

$$(\nabla f(x))v_1 := \langle \nabla f(x), v_1 \rangle = \nabla_{v_1} f(x), \ \ (\nabla^2 f(x))(v_1, v_2) := \mathrm{Hess}_f(v_1, v_2), \ \ x, v_1, v_2 \in \mathbb{R}^d.$$

(b) For $f \in C^2(\mathscr{P}_2)$,

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

(c) Derivatives of vector or matrix valued functions are given by those of component functions. For instance, for $f = (f_{ij}) \in C^1(\mathbb{R}^d \times \mathscr{P}_2 \to \mathbb{R}^l \otimes \mathbb{R}^k)$,

$$\nabla_v f(x, \mu) := \big(\langle \nabla f_{ij}(x, \mu), v \rangle\big), \quad D_\phi f(x, \mu) := \big(D_\phi f_{ij}(x, \mu)\big),$$

where $x, v \in \mathbb{R}^d, \mu \in \mathscr{P}_2$ and $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$.

We will also need the following notion of uniform boundedness and continuity.

Definition 3.2. Let $\mathbb{B}$ be a Banach space, and let E be a topological space. The family

$$\big\{\eta(x) \in L^1(\Omega \to \mathbb{B}; \mathbb{P}) : x \in E\big\}$$

is called $L^{\infty-}(\mathbb{P})$ bounded continuous, if for any $p \geq 1$,

$$\sup_{x \in E} \mathbb{E}\|\eta(x)\|^p < \infty, \quad \lim_{y \to x} \mathbb{E}\|\eta(x) - \eta(y)\|^p = 0, \quad x \in E.$$

Let $\mathscr{L}(\mathbb{B}_1 \to \mathbb{B}_2)$ denote the space of all bounded linear operators from a Banach space $\mathbb{B}_1$ to the other one $\mathbb{B}_2$. When $\mathbb{B}_1$ and $\mathbb{B}_2$ are finite-dimensional Hilbert spaces, we regard $\mathscr{L}(\mathbb{B}_1 \to \mathbb{B}_2)$ as Euclidean space. The following lemma can be easily proved by using Itô's formula, so we omit the proof to save space.

LP **Lemma 3.3.** *Let $k, l \geq 1$, and let*

$$\begin{aligned}
&B_1 : \Omega \times [0,T] \times \mathbb{R}^l \times \mathscr{P}_2 \to \mathbb{R}^k, \quad \Sigma_1 : \Omega \times [0,T] \times \mathbb{R}^l \times \mathscr{P}_2 \to \mathbb{R}^k \otimes \mathbb{R}^m, \\
&B_2 : \Omega \times [0,T] \times \mathbb{R}^l \times \mathscr{P}_2 \to \mathbb{R}^k \otimes \mathbb{R}^k, \quad \Sigma_2 : \Omega \times [0,T] \times \mathbb{R}^l \times \mathscr{P}_2 \to \mathscr{L}(\mathbb{R}^k \to \mathbb{R}^k \otimes \mathbb{R}^m)
\end{aligned}$$

be progressively measurable. If $\{B_2, \Sigma_2\}$ are uniformly bounded and continuous in $(t, x, \mu) \in [0,T] \times \mathbb{R}^l \times \mathscr{P}_2$, and $\{B_1(t,x,\mu), \Sigma_1(t,x,\mu)\}$ are $L^{\infty-}(\mathbb{P})$ bounded continuous, then for any $e \in \mathbb{R}^k$ and $(x, \mu) \in \mathbb{R}^l \times \mathscr{P}_2$, the solution $(\eta_{s,t}^{x,\mu})_{t \in [s,T]}$ for the SDE

$$\mathrm{d}\eta_{s,t}^{x,\mu} = \big\{B_1(t,x,\mu) + B_2(t,x,\mu)\eta_t^{x,\mu}\big\}\mathrm{d}t + \big\{\Sigma_1(t,x,\mu) + \Sigma_2(t,x,\mu)\eta_t^{x,\mu}\big\}\mathrm{d}W_t, \ \eta_{s,s}^{x,\mu} = e, t \in [s,T]$$

is $L^{\infty-}(\mathbb{P})$ bounded continuous.

In the following subsections, we calculate the first and second order derivatives of $(X_{s,t}^{x,\mu}, \Lambda_{s,t}^\mu)$ in x and μ respectively, which will be used in the proof of Proposition 3.2.

3.1 Formulas for $\nabla X_{s,t}^{x,\mu}$ and $\nabla^2 X_{s,t}^{x,\mu}$

Let $\{e_i\}_{1\le i\le d}$ be the canonical orthonormal basis of $\mathbb{R}^d$. Given $(\Lambda_{s,t}^\mu)_{t\ge s}$, the SDE (1.7) becomes the classical one with random coefficients of bounded and continuous first and second order derivatives in x. So, when $\nabla b(t,x,\mu)$ and $\nabla\sigma(t,x,\mu)$ are $L^{\infty-}(\mathbb{P})$ bounded continuous, by taking ∂_{x_i} to $X_{s,t}^{x,\mu}$ in (1.7), we see that for any $1\le i\le d$,

$$v_{s,t}^{i,x,\mu} := \partial_{x_i} X_{s,t}^{x,\mu}, \quad t\ge s$$

solves the linear SDE

GRT (3.2)
$$\begin{aligned}&\mathrm{d}v_{s,t}^{i,x,\mu} = \Big[\big\{\nabla b(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}v_{s,t}^{i,x,\mu}\Big]\mathrm{d}t + \Big[\big\{\nabla\sigma(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}v_{s,t}^{i,x,\mu}\Big]\mathrm{d}W_t,\\ &t\ge s,\ v_{s,s}^{i,x,\mu}=e_i.\end{aligned}\tag{3.2}$$

If moreover $\nabla^2 b(t,x,\mu)$ and $\nabla^2\sigma(t,x,\mu)$ are $L^{\infty-}(\mathbb{P})$ bounded continuous, then by taking ∂_j to the SDE (3.2), we see that for $1\le j\le d$

$$v_{s,t}^{i,j,x,\mu} := \partial_{x_i}\partial_{x_j} X_{s,t}^{x,\mu}, \quad t\ge s$$

solves the SDEs

$$\begin{aligned}\mathrm{d}v_{s,t}^{i,j,x,\mu} &= \Big[\big\{\nabla b(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}v_{s,t}^{i,j,x,\mu} + \big\{\nabla^2 b(t,\cdot,,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}(v_{s,t}^{i,x,\mu}, v_{s,t}^{j,x,\mu})\Big]\mathrm{d}t\\ &+ \Big[\big\{\nabla\sigma(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}v_{s,t}^{i,j,x,\mu} + \big\{\nabla^2\sigma(t,\cdot,,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}(v_{s,t}^{i,x,\mu}, v_{s,t}^{j,x,\mu})\Big]\mathrm{d}W_t, \quad v_{s,s}^{i,j,x,\mu}=0.\end{aligned}$$

Combining these with Lemma 3.3, we obtain the following result.

LP-1 **Lemma 3.4.** *Assume* **(A)** *and that $\nabla b(t,x,\mu), \nabla^2 b(t,x,\mu), \nabla\sigma(t,x,\mu)$ and $\nabla^2\sigma(t,x,\mu)$ are $L^{\infty-}(\mathbb{P})$ bounded continuous, then so are $\nabla X_{s,t}^{x,\mu}$ and $\nabla^2 X_{s,t}^{x,\mu}$.*

3.2 Formula for $DX_{s,t}^{x,\mu}$

We will establish the SDE for $DX_{s,t}^{x,\mu}(y)$ under the following condition **(C)** on b and σ.

(C) Assume that b and σ are progressively measurable such that the derivatives

$$\nabla b(t,x,\mu),\quad \nabla\sigma(t,x,\mu),\quad Db(t,x,\mu)(y),\quad D\sigma(t,x,\mu)(y)$$

are uniformly bounded and continuous in $(x,\mu,y)\in\mathbb{R}^d\times\mathscr{P}_2\times\mathbb{R}^d$.

LP-2 **Lemma 3.5.** *Assume* **(C)**. *Then for any $(x,\mu,y)\in\mathbb{R}^d\times\mathscr{P}_2\times\mathbb{R}^d$, $w_{s,t}^{x,\mu}(y) := (DX_{s,t}^{x,\mu})(y)$ for $t\in[s,T]$ exists and solves the SDE*

$$\begin{aligned}\mathrm{d}w_{s,t}^{x,\mu}(y) = \bigg[&\big\{w_{s,t}^{x,\mu}(y)\big\}^*\nabla b(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu}) + (\nabla X_{s,t}^{y,\mu})^*\{Db(t,X_{s,t}^{x,\mu},\cdot)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{y,\mu})\\ &+\int_{\mathbb{R}^d}\big\{w_{s,t}^{z,\mu}(y)\big\}^*\big\{Db(t,X_{s,t}^{x,\mu},\cdot)(\Lambda_{s,t}^\mu)\big\}(X_{s,t}^{z,\mu})\mu(\mathrm{d}z)\bigg]\mathrm{d}t\end{aligned}$$

$$
\begin{aligned}
&+\Big[\big\{w_{s,t}^{x,\mu}(y)\big\}^*\big\{\nabla\sigma(t,\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\big\}+(\nabla X_{s,t}^{y,\mu})^*\{D\sigma(t,X_{s,t}^{x,\mu},\cdot)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{y,\mu})\\
&\quad+\int_{\mathbb{R}^d}\big\{w_{s,t}^{z,\mu}(y)\big\}^*\big\{D\sigma(t,X_{s,t}^{x,\mu},\cdot)(\Lambda_{s,t}^\mu)\big\}(X_{s,t}^{z,\mu})\mu(\mathrm{d}z)\Big]\mathrm{d}W_t,\quad w_{s,s}^{x,\mu,y}=0,
\end{aligned}
$$

where $\big\{w_{s,t}^{x,\mu}(y)\big\}^*$ *is the transposition of the matrix* $w_{s,t}^{x,\mu}(y)$*. Consequently,* $(DX_{s,t}^{x,\mu})(y)$ *is* $L^{\infty-}(\mathbb{P})$ *bounded continuous.*

To prove the existence of $DX_{s,t}^{x,\mu}$, for fixed $\phi\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$, let $\mu_\varepsilon=\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}$ and consider

$$\xi_{s,t}^{x,\varepsilon}:=\frac{X_{s,t}^{x,\mu_\varepsilon}-X_{s,t}^{x,\mu}}{\varepsilon},\quad \varepsilon\in(0,1),t\in[s,T].$$

We first establish the SDE for $D_\phi X_{s,t}^{x,\mu}:=\lim_{\varepsilon\downarrow0}\xi_{s,t}^{x,\varepsilon}$. To this end, we need the following lemma.

B1 **Lemma 3.6.** *Assume* **(A)** *and let* $\tilde\xi_{s,t}^{x,\varepsilon}:=\frac{X_{s,t}^{x+\varepsilon\phi(x),\mu_\varepsilon}-X_{s,t}^{x,\mu_\varepsilon}}{\varepsilon}$*. Then for any* $f\in C^{1,1}(\mathbb{R}^d\times\mathscr{P}_2)$ *with*

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

the process

$$
\begin{aligned}
\Xi_{s,t}^{x,\varepsilon}(f):=\;&\frac{f(X_{s,t}^{x,\mu_\varepsilon},\Lambda_{s,t}^{\mu_\varepsilon})-f(X_{s,t}^{x,\mu},\Lambda_{s,t}^\mu)}{\varepsilon}-\nabla_{\xi_{s,t}^{x,\varepsilon}}f(\cdot,\Lambda_{s,t}^\mu)(X_{s,t}^{x,\mu})\\
&-\int_{\mathbb{R}^d}\big\langle\xi_{s,t}^{z,\varepsilon}+\tilde\xi_{s,t}^{z,\varepsilon},\{Df(X_{s,t}^{z,\mu},\cdot)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{z,\mu})\big\rangle\mu(\mathrm{d}z),\quad t\in[s,T]
\end{aligned}
$$

satisfies

AD1 (3.3) $$\big|\Xi_{s,t}^{x,\varepsilon}(f)\big|^2\le 8K_f\big(|\xi_{s,t}^{x,\varepsilon}|^2+\mu(|\xi_{s,t}^{\cdot,\varepsilon}+\tilde\xi_{s,t}^{\cdot,\varepsilon}|^2)\big),\quad t\in[s,T],$$

AD2 (3.4) $$\lim_{\varepsilon\downarrow0}\mathbb{E}\big|\Xi_{s,t}^{x,\varepsilon}(f)\big|^2=0.$$

Proof. Let $\eta_r^x=X_{s,t}^{x,\mu}+r(X_{s,t}^{x+\varepsilon\phi(x),\mu_\varepsilon}-X_{s,t}^{x,\mu}),\ r\in[0,1]$. Then $\eta_0^x=X_{s,t}^{x,\mu},\eta_1^x=X_{s,t}^{x+\varepsilon\phi(x),\mu_\varepsilon}$, so that

$$\mathscr{L}_{\eta_0|\mu}:=\mu\circ(X_{s,t}^{\cdot,\mu})^{-1}=\Lambda_{s,t}^\mu,\quad \mathscr{L}_{\eta_1|\mu}:=\mu\circ(X_{s,t}^{\cdot+\varepsilon\phi,\mu_\varepsilon})^{-1}=\mu_\varepsilon\circ(X_{s,t}^{\cdot,\mu_\varepsilon})^{-1}=\Lambda_{s,t}^{\mu_\varepsilon}.$$

Moreover, $\frac{\mathrm{d}}{\mathrm{d}r}\eta_r^x=\xi_{s,t}^{x,\varepsilon}+\tilde\xi_{s,t}^{x,\varepsilon}$. Then by Lemma 2.4, we have

$$
\begin{aligned}
&\frac{d}{\mathrm{d}r}f(y,\mathscr{L}_{\eta_r|\mu})=\Big\langle Df(y,\cdot)(\mathscr{L}_{\eta_r|\mu})(\eta_r),\frac{\mathrm{d}}{\mathrm{d}r}\eta_r\Big\rangle_{L^2(\mu)}\\
&=\varepsilon\int_{\mathbb{R}^d}\big\langle Df(y,\cdot)(\mathscr{L}_{\eta_r|\mu})(\eta_r^z),\xi_{s,t}^{z,\varepsilon}+\tilde\xi_{s,t}^{z,\varepsilon}\big\rangle\mu(\mathrm{d}z),\quad r\in[0,1],y\in\mathbb{R}^d.
\end{aligned}
$$

So, letting $\zeta_r^x = (1-r)X_{s,t}^{x,\mu} + rX_{s,t}^{x,\mu_\varepsilon}$, we obtain

$$\frac{f(X_{s,t}^{x,\mu_\varepsilon}, \Lambda_{s,t}^{\mu_\varepsilon}) - f(X_{s,t}^{x,\mu}, \Lambda_{s,t}^{\mu})}{\varepsilon} = \frac{1}{\varepsilon}\int_0^1 \Big\{\frac{\mathrm{d}}{\mathrm{d}r} f(\zeta_r^x, \mathscr{L}_{\eta_r|\mu})\Big\}\mathrm{d}r$$
$$= \int_0^1 \Big\{\big\langle \nabla f(\cdot, \mathscr{L}_{\eta_r|\mu})(\zeta_r^x), \xi_{s,t}^{x,\varepsilon}\big\rangle + \int_{\mathbb{R}^d} \big\langle Df(\zeta_r^x, \cdot)(\mathscr{L}_{\eta_r|\mu})(\eta_r^z), \xi_{s,t}^{z,\varepsilon} + \tilde{\xi}_{s,t}^{z,\varepsilon}\big\rangle \mu(\mathrm{d}z)\Big\}\mathrm{d}r.$$

This together with the definition of $\Xi_{s,t}^{x,\varepsilon}(f)$ gives

P0 (3.5)
$$\begin{aligned}\big|\Xi_{s,t}^{x,\varepsilon}(f)\big|^2 = &\ \bigg|\int_0^1 \Big\{\big\langle \nabla f(\cdot, \mathscr{L}_{\eta_r|\mu})(\zeta_r^x) - \nabla f(\cdot, \Lambda_{s,t}^{\mu})(X_{s,t}^{x,\mu}), \xi_{s,t}^{x,\varepsilon}\big\rangle \\ &+ \int_{\mathbb{R}^d} \big\langle Df(\eta_r^x, \cdot)(\mathscr{L}_{\eta_r|\mu})(\zeta_r^z) - Df(X_{s,t}^{x,\mu}, \cdot)(\Lambda_{s,t}^{\mu})(X_{s,t}^{z,\mu}), \xi_{s,t}^{z,\varepsilon} + \tilde{\xi}_{s,t}^{z,\varepsilon}\big\rangle \mu(\mathrm{d}z)\Big\}\mathrm{d}r\bigg|^2 \\ \le &\, 8K_f\big(|\xi_{s,t}^{x,\varepsilon}|^2 + \mu(|\xi_{s,t}^{\cdot,\varepsilon} + \tilde{\xi}_{s,t}^{\cdot,\varepsilon}|^2)\big),\end{aligned}$$

which implies (3.3). On the other hand, it is easy to see that (2.4) implies

PG1 (3.6)
$$\sup_{x\in\mathbb{R}^d, \varepsilon\in(0,1)} \mathbb{E}\Big[\sup_{t\in[s,T]} \big\{|\xi_{s,t}^{x,\varepsilon}|^2 + \mu(|\tilde{\xi}_{s,t}^{\cdot,\varepsilon}|^2)\big\}\Big] \le c\mu(|\phi|^2), \quad \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$$

for some constant $c > 0$. Combining this with the facts that $(\nabla f, Df)$ is bounded continuous, $\lim_{r\to 0} \zeta_r^z = X_{s,t}^{z,\mu}$, and $\lim_{r\to 0} \mathscr{L}_{\eta_r|\mu} = \Lambda_{s,t}^{\mu}$, we may apply the dominated convergence theorem to deduce (3.4) from the first equality in (3.5) with $\varepsilon \downarrow 0$. $\square$

B2 **Lemma 3.7.** *Assume* **(C)**. *For any $(s, x, \mu) \in [0, T] \times \mathbb{R}^d, \times \mathscr{P}_2$ and $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d)$, $w_{s,t}^{x,\mu,\phi} := D_\phi X_{s,t}^{x,\mu}$ for $t \in [s, T]$ exists in $L^2(\Omega \to C([s, T] \to \mathbb{R}^d); \mathbb{P})$, and there exists a constant $C > 0$ such that*

NN1 (3.7)
$$\mathbb{E}\Big[\sup_{s\le t\le T} |w_{s,t}^{x,\mu,\phi}|^2\Big] \le C\mu(|\phi|^2), \quad (s, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathscr{P}_2.$$

Moreover, for any $t \in [s, T]$,

NN1' (3.8)
$$\begin{aligned} w_{s,t}^{x,\mu,\phi} = &\int_s^t \Big\{\nabla_{w_{s,r}^{x,\mu,\phi}} b(r, \cdot, \Lambda_{s,r}^{\mu})(X_{s,r}^{x,\mu})\Big\}\mathrm{d}r + \int_s^t \Big\{\nabla_{w_{s,t}^{x,\mu,\phi}} \sigma(t, \cdot, \Lambda_{s,t}^{\mu})(X_{s,t}^{x,\mu})\Big\}\mathrm{d}W_r \\ &+ \int_s^t \bigg(\int_{\mathbb{R}^d} \big\langle \{Db(r, X_{s,r}^{x,\mu}, \cdot)(\Lambda_{s,r}^{\mu})\}(X_{s,r}^{z,\mu}), w_{s,r}^{z,\mu,\phi} + \nabla_{\phi(z)} X_{s,t}^{z,\mu}\big\rangle \mu(\mathrm{d}z)\bigg)\mathrm{d}r \\ &+ \int_s^t \bigg(\int_{\mathbb{R}^d} \big\langle \{D\sigma(r, X_{s,r}^{x,\mu}, \cdot)(\Lambda_{s,t}^{\mu})\}(X_{s,r}^{z,\mu}), w_{s,r}^{z,\mu,\phi} + \nabla_{\phi(z)} X_{s,t}^{z,\mu}\big\rangle \mu(\mathrm{d}z)\bigg)\mathrm{d}W_r. \end{aligned}$$

Proof. To prove the existence of $w_{s,t}^{x,\mu,\phi} := D_\phi X_{s,t}^{x,\mu}$ in $L^2(\Omega \to C([s, T] \to \mathbb{R}^d); \mathbb{P})$, it suffices to show

N1 (3.9)
$$\lim_{\varepsilon,\delta\downarrow 0} \mathbb{E}\Big[\sup_{t\in[s,T]} |\xi_{s,t}^{x,\varepsilon} - \xi_{s,t}^{x,\delta}|^2\Big] = 0.$$

By the definition of $\xi^{x,\varepsilon}_{s,t}$ and letting

$$\Xi^{x,\varepsilon}_{s,t}(b) = \big(\Xi^{x,\varepsilon}_{s,t}(b_i)\big)_{1\le i\le d},\quad \Xi^{x,\varepsilon}_{s,t}(\sigma) = \big(\Xi^{x,\varepsilon}_{s,t}(\sigma_{i,j})\big)_{1\le i\le d, 1\le j\le m},$$

we obtain

N0

$$\begin{aligned}
\xi^{x,\varepsilon}_{s,t} &= \frac{1}{\varepsilon}\int_s^t \big\{b(r, X^{x,\mu_\varepsilon}_{s,r}, \Lambda^{\mu_\varepsilon}_{s,r}) - b(r, X^{x,\mu}_{s,r}, \Lambda^{\mu}_{s,r})\big\}\mathrm{d}r \\
&\qquad\qquad + \frac{1}{\varepsilon}\int_s^t \big\{\sigma(r, X^{x,\mu_\varepsilon}_{s,r}, \Lambda^{\mu_\varepsilon}_{s,r}) - \sigma(r, X^{x,\mu}_{s,r}, \Lambda^{\mu}_{s,r})\big\}\mathrm{d}W_r \\
&= \int_s^t \Big\{\Xi^{x,\varepsilon}_{s,r}(b) + \nabla_{\xi^{x,\varepsilon}_{s,r}} b(r,\cdot,\Lambda^{\mu}_{s,r})(X^{x,\mu}_{s,r})\Big\}\mathrm{d}r \\
&+ \int_s^t \Big\{\int_{\mathbb{R}^d} \big\langle \{Db(r, X^{x,\mu}_{s,r},\cdot)(\Lambda^{\mu}_{s,t})\}(X^{z,\mu}_{s,t}), \xi^{z,\varepsilon}_{s,r} + \tilde{\xi}^{z,\varepsilon}_{s,r}\big\rangle \mu(\mathrm{d}z)\Big\}\mathrm{d}r \\
&+ \int_s^t \Big\{\Xi^{x,\varepsilon}_{s,r}(\sigma) + \nabla_{\xi^{x,\varepsilon}_{s,r}} \sigma(r,\cdot,\Lambda^{\mu}_{s,r})(X^{x,\mu}_{s,r})\Big\}\mathrm{d}W_r \\
&+ \int_s^t \Big\{\int_{\mathbb{R}^d} \big\langle \{D\sigma(r, X^{x,\mu}_{s,r},\cdot)(\Lambda^{\mu}_{s,t})\}(X^{z,\mu}_{s,t}), \xi^{z,\varepsilon}_{s,r} + \tilde{\xi}^{z,\varepsilon}_{s,r}\big\rangle \mu(\mathrm{d}z)\Big\}\mathrm{d}W_r.
\end{aligned}\tag{3.10}$$

Combining this with **(C)** and using the BDG inequality, we may find out a constant $C>0$ such that for any $t\in[s,T]$,

N2

$$\begin{aligned}
\mathbb{E}\Big[\sup_{r\in[s,t]}|\xi^{x,\varepsilon}_{s,r} - \xi^{x,\delta}_{s,r}|^2\Big] \le C\mathbb{E}\int_s^t \Big\{&|\Xi^{x,\varepsilon}_{s,r}(b) - \Xi^{x,\delta}_{s,r}(b)|^2 + \|\Xi^{x,\varepsilon}_{s,r}(\sigma) - \Xi^{x,\delta}_{s,r}(\sigma)\|^2 \\
&+ |\xi^{x,\varepsilon}_{s,r} - \xi^{x,\delta}_{s,r}|^2 + \mu(|\xi^{\cdot,\varepsilon}_{s,r} - \xi^{\cdot,\delta}_{s,r}|^2 + |\tilde{\xi}^{\cdot,\varepsilon}_{s,r} - \tilde{\xi}^{\cdot,\delta}_{s,r}|^2)\Big\}\mathrm{d}r.
\end{aligned}\tag{3.11}$$

Integrating both sides with respect to $\mu(\mathrm{d}x)$, we obtain

$$\begin{aligned}
\mathbb{E}\mu(|\xi^{\cdot,\varepsilon}_{s,t} - \xi^{\cdot,\delta}_{s,t}|^2) &\le C\mathbb{E}\int_s^t \mu\big(|\Xi^{\cdot,\varepsilon}_{s,r}(b) - \Xi^{\cdot,\delta}_{s,r}(b)|^2 + \|\Xi^{\cdot,\varepsilon}_{s,r}(\sigma) - \Xi^{\cdot,\delta}_{s,r}(\sigma)\|^2 + |\tilde{\xi}^{\cdot,\varepsilon}_{s,r} - \tilde{\xi}^{\cdot,\delta}_{s,r}|^2\big)\mathrm{d}r \\
&\quad + 2C\int_s^t \mathbb{E}\mu(|\xi^{\cdot,\varepsilon}_{s,r} - \xi^{\cdot,\delta}_{s,r}|^2)\mathrm{d}r,\quad t\in[s,T].
\end{aligned}$$

Then by Grownwall's inequality, (3.4), (3.6), and the existence of

$$\lim_{\varepsilon\downarrow 0}\tilde{\xi}^{\cdot,\varepsilon}_{s,r} = \nabla_\phi X^{\cdot,\mu}_{s,t} \text{ in } L^2(\mathbb{P})$$

as explained in Subsection 4.1, which implies $\lim_{\varepsilon,\delta\downarrow 0}\mathbb{E}|\tilde{\xi}^{\cdot,\varepsilon}_{s,r} - \tilde{\xi}^{\cdot,\delta}_{s,r}|^2 = 0$, we derive

$$\begin{aligned}
&\lim_{\varepsilon,\delta\downarrow 0}\sup_{t\in[s,T]}\mathbb{E}\mu(|\xi^{\cdot,\varepsilon}_{s,t} - \xi^{\cdot,\delta}_{s,t}|^2) \\
&\le C\mathrm{e}^{2CT}\lim_{\varepsilon,\delta\downarrow 0}\mathbb{E}\int_s^T \mu\big(|\Xi^{\cdot,\varepsilon}_{s,r}(b) - \Xi^{\cdot,\delta}_{s,r}(b)|^2 + \|\Xi^{\cdot,\varepsilon}_{s,r}(\sigma) - \Xi^{\cdot,\delta}_{s,r}(\sigma)\|^2 + |\tilde{\xi}^{\cdot,\varepsilon}_{s,r} - \tilde{\xi}^{\cdot,\delta}_{s,r}|^2\big)\mathrm{d}r = 0.
\end{aligned}$$

Substituting this into (3.11) and using Gronwall's inequality again, we arrive at

$$
\begin{aligned}
&\lim_{\varepsilon,\delta\downarrow 0}\mathbb{E}\Big[\sup_{t\in[s,T]}|\xi_{s,t}^{x,\varepsilon}-\xi_{s,t}^{x,\delta}|^2\Big]\\
&\le C\mathrm{e}^{CT}\lim_{\varepsilon,\delta\downarrow 0}\mathbb{E}\int_s^T\Big\{|\Xi_{s,r}^{x,\varepsilon}(b)-\Xi_{s,r}^{x,\delta}(b)|^2+\|\Xi_{s,r}^{x,\varepsilon}(\sigma)-\Xi_{s,r}^{x,\delta}(\sigma)\|^2\\
&\qquad\qquad+\mu\big(|\xi_{s,r}^{\cdot,\varepsilon}-\xi_{s,r}^{\cdot,\delta}|^2+|\tilde{\xi}_{s,r}^{\cdot,\varepsilon}-\tilde{\xi}_{s,r}^{\cdot,\delta}|^2\big)\Big\}\mathrm{d}r=0.
\end{aligned}
$$

Therefore, (3.9) holds, so that

$$
w_{s,t}^{x,\mu,\phi}:=D_\phi X_{s,t}^{x,\mu}=\lim_{\varepsilon\downarrow 0}\xi_{s,t}^{x,\varepsilon},\quad t\in[s,T]
$$

exists in $L^2(\Omega\to C([s,T]\to\mathbb{R}^d);\mathbb{P})$, and (3.7) follows from (3.6). Moreover, by **(C)** and Lemma 3.6, we may let $\varepsilon\downarrow 0$ in (3.10) to derive the desired equation for $w_{s,t}^{x,\mu,\phi}$. □

Proof of Lemma 3.5. By (3.7), $(DX_{s,t}^{x,\mu})_{t\in[s,T]}$ exists with

*W1 (3.12)
$$
\langle DX_{s,t}^{x,\mu},\phi\rangle_{L^2(\mu)}=D_\phi X_{s,t}^{x,\mu}=w_{s,t}^{x,\mu,\phi},\quad \phi\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu).
$$

On the other hand, let $w_{s,t}^{x,\mu}(y)$ solve the SDE in Lemma 3.5. Then $\tilde{w}_{s,t}^{x,\mu,\phi}:=\langle w_{s,t}^{x,\mu},\phi\rangle_{L^2(\mu)}$ solves the SDE in Lemma 3.7 for $w_{s,t}^{x,\mu,\phi}$. By the uniqueness, we have $w_{s,t}^{x,\mu,\phi}=\tilde{w}_{s,t}^{x,\mu,\phi}$. Combining this with (3.12), we obtain μ-a.e. $w_{s,t}^{x,\mu}=DX_{s,t}^{x,\mu}$. Then the proof is finished. □

3.3 Some other derivatives

We first present a formula for $Df(\Lambda_{s,t}^\mu)$.

LPP **Lemma 3.8.** *Assume* **(C)**. *For any* $f\in C_b^1(\mathscr{P}_2)$,

YP (3.13)
$$
\begin{aligned}
&\{Df(\Lambda_{s,t}^{\cdot})(\mu)\}(y)\\
&=\big(\nabla X_{s,t}^{y,\mu}\big)^*\big\{(Df)(\Lambda_{s,t}^\mu)\big\}(X_{s,t}^{y,\mu})+\int_{\mathbb{R}^d}\big(DX_{s,t}^{x,\mu}\big)^*(y)\{(Df)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{x,\mu})\mu(\mathrm{d}x).
\end{aligned}
$$

Proof. Let $\phi\in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$. Since $\Lambda_{s,t}^\mu=\mu\circ(X_{s,t}^{\cdot,\mu})^{-1}$, for any $\varepsilon>0$ we have

$$
\begin{aligned}
&\int_{\mathbb{R}^d}h(z)\big(\Lambda_{s,t}^{\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}\big)(\mathrm{d}z)=\int_{\mathbb{R}^d}h\big(X_{s,t}^{x,\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}\big)\big(\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}\big)(\mathrm{d}x)\\
&=\int_{\mathbb{R}^d}h\big(X_{s,t}^{x+\varepsilon\phi(x),\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}\big)\mu(\mathrm{d}x),\quad h\in\mathscr{B}_b(\mathbb{R}^d).
\end{aligned}
$$

So, $\Lambda_{s,t}^{\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}$ is the law of

$$
x\mapsto X_{s,t}^{x+\varepsilon\phi(x),\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}
$$

on the probability space $(\mathbb{R}^d,\mathscr{B}(\mathbb{R}^d),\mu)$. Therefore, by Lemmas 2.4 and 3.5, we obtain

$$
\langle Df(\Lambda_{s,t}^{\cdot})(\mu),\phi\rangle_{L^2(\mu)}:=\frac{\mathrm{d}}{\mathrm{d}\varepsilon}f(\Lambda_{s,t}^{\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}})\Big|_{\varepsilon=0}
$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \Big\langle \{(Df)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{x,\mu}), \frac{\mathrm{d}}{\mathrm{d}\varepsilon} X_{s,t}^{x+\varepsilon\phi(x),\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}}\Big|_{\varepsilon=0}\Big\rangle \mu(\mathrm{d}x)\\
&= \int_{\mathbb{R}^d} \big\langle \{(Df)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{x,\mu}), \nabla_{\phi(x)} X_{s,t}^{x,\mu} + D_\phi X_{s,t}^{x,\mu}\big\rangle \mu(\mathrm{d}x)\\
&= \int_{\mathbb{R}^d} \big\langle (\nabla X_{s,t}^{x,\mu})^* \{(Df)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{x,\mu}), \phi(x)\big\rangle \mu(\mathrm{d}x)\\
&\quad + \int_{\mathbb{R}^d\times\mathbb{R}^d} \big\langle (DX_{s,t}^{x,\mu})^*(y) \{(Df)(\Lambda_{s,t}^{x,\mu})\}(X_{s,t}^{x,\mu}), \phi(y)\big\rangle \mu(\mathrm{d}x)\mu(\mathrm{d}y)\\
&= \Big\langle \big(\nabla X_{s,t}^{\cdot,\mu}\big)^* \big\{(Df)(\Lambda_{s,t}^\mu)\big\}(X_{s,t}^{\cdot,\mu}) + \int_{\mathbb{R}^d} \big(DX_{s,t}^{x,\mu}\big)^*(\cdot)\{(Df)(\Lambda_{s,t}^\mu)\}(X_{s,t}^{x,\mu})\mu(\mathrm{d}x), \phi\Big\rangle_{L^2(\mu)}.
\end{aligned}
$$

Therefore, (3.13) holds. □

Next, when $b, \sigma \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$, by making derivatives to the SDE for $w_{s,t}^{x,\mu}(y)$ presented in Lemma 3.5, we derive the following result.

LP-3 **Lemma 3.9.** *Assume that $b, \sigma \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$. Then all derivatives*

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

are $L^{\infty-}(\mathbb{P})$ bounded continuous.

Proof. (a) We first consider $\{D\nabla X_{s,t}^{x,\mu}\}(y)$. Since $b, \sigma \in C_b^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$, by (3.2) and Lemmas 3.4-3.5, $v_{s,t}^{x,\mu} := \nabla_v X_{s,t}^{x,\mu}$ for $v \in \mathbb{R}^d$ solves the SDE

$$\mathrm{d}v_{s,t}^{x,\mu} = Z_1(t,x,\mu)v_{s,t}^{x,\mu}\mathrm{d}t + \{Z_2(t,x,\mu)v_{s,t}^{x,\mu}\}\mathrm{d}W_t,\quad v_{s,s}^{x,\mu} = v,$$

where

$$Z_1 : [0,T]\times\mathbb{R}^d\times\mathscr{P}_2 \to \mathbb{R}^d,\quad Z_2 : [0,T]\times\mathbb{R}^d\times\mathscr{P}_2 \to \mathbb{R}^d\otimes\mathbb{R}^d$$

are progressively measurable and satisfy

- **(D)** $Z_1(t,x,\mu)$ and $Z_2(t,x,\mu)$ are uniformly bounded and continuous in $(t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}_2$; $DZ_1(t,x,\mu)(y)$ and $DZ_2(t,x,\mu)(y)$ are $L^{\infty-}(\mathbb{P})$ bounded continuous.

Then for any $\phi \in L^2(\mathbb{R}^d\to\mathbb{R}^d;\mu)$ and $\mu_\varepsilon := \mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}$ for small $\varepsilon>0$, $\gamma_{s,t}^\varepsilon := \frac{v_{s,t}^{x,\mu_\varepsilon}-v_{s,t}^{x,\mu}}{\varepsilon}$ solves the SDE

$$
\begin{aligned}
&\mathrm{d}\gamma_{s,t}^\varepsilon = \{Z_1(t,x,\mu)\gamma_{s,t}^\varepsilon\}\mathrm{d}t + \{Z_2(t,x,\mu)\gamma_{s,t}^\varepsilon\}\mathrm{d}W_t\\
&+ \frac{\{Z_1(t,x,\mu^\varepsilon)-Z_1(t,x,\mu)\}v_{s,t}^{x,\mu_\varepsilon}}{\varepsilon}\mathrm{d}t + \frac{\{Z_2(t,x,\mu_\varepsilon)-Z_2(t,x,\mu)\}v_{s,t}^{x,\mu_\varepsilon}}{\varepsilon}\mathrm{d}W_t,\quad \eta_{s,s}^\varepsilon = 0.
\end{aligned}
$$

By **(D)**, we may repeat the proof of Lemma 3.7 to conclude that $D_\phi v_{s,t}^{x,\mu} := \lim_{\varepsilon\downarrow 0}\eta_{s,t}^\varepsilon$ exists and solves the SDE

$$
\begin{aligned}
\mathrm{d}\{D_\phi v_{s,t}^{x,\mu}\} &= \big\{Z_1(t,x,\mu)D_\phi v_{s,t}^{x,\mu} + (D_\phi Z_1(t,x,\mu))v_{s,t}^{x,\mu}\big\}\mathrm{d}t\\
&\quad + \big\{Z_2(t,x,\mu)D_\phi v_{s,t}^{x,\mu} + (D_\phi Z_2(t,x,\mu))v_{s,t}^{x,\mu}\big\}\mathrm{d}W_t,\quad D_\phi v_{s,s}^{x,\mu} = 0.
\end{aligned}
$$

Hence, $Dv^{x,\mu}_{s,t}(y)$ solves the SDE

$$\begin{aligned}\mathrm{d}\{Dv^{x,\mu}_{s,t}(y)\} = &\big\{Z_1(t,x,\mu)Dv^{x,\mu}_{s,t}(y) + (DZ_1(t,x,\mu)(y))v^{x,\mu}_{s,t}\big\}\mathrm{d}t\\ &+ \big\{Z_2(t,x,\mu)Dv^{x,\mu}_{s,t}(y) + (DZ_2(t,x,\mu)(y))v^{x,\mu}_{s,t}\big\}\mathrm{d}W_t, \quad Dv^{x,\mu}_{s,s}(y) = 0.\end{aligned}$$

Therefore, by Lemma 3.4 and **(D)**, Lemma 3.3 yields that $\{D\nabla X^{x,\mu}_{s,t}\}(y)$ is $L^{\infty-}(\mathbb{P})$ bounded continuous.

(b) To calculate $\nabla\{DX^{\cdot,\mu}_{s,t}(y)\}(x), \nabla\{DX^{x,\mu}_{s,t}(\cdot)\}(y)$ and $D^2X^{x,\mu}_{s,t}(y,z) := D\{DX^{x,\mu}_{s,t}(y)\}(z)$, we reformulate the SDE in Lemma 3.5 for $w^{x,\mu}_{s,t}(y) := DX^{x,\mu}_{s,t}(y)$ as

$$\mathrm{d}w^{x,\mu}_{s,t} = \big\{A_1(t,x,\mu)w^{x,\mu}_{s,t} + A_2(t,x,\mu)\big\}\mathrm{d}t + \big\{B_1(t,x,\mu)w^{x,\mu}_{s,t} + B_2(t,x,\mu)\big\}\mathrm{d}W_t, \quad w^{x,\mu}_{s,s} = 0,$$

where, due to Lemmas 3.4-3.5 and (a), $\{A_i, B_i\}_{i=1,2}$ are progressively measurable maps such that

- A_1 and B_1 are uniformly bounded and continuous in $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_2$;
- $\{A_i, B_i, \nabla A_i, \nabla B_i, DA_i, DB_i\}_{i=1,2}$ are $L^{\infty-}(\mathbb{P})$ bounded continuous in corresponding arguments.

So, as explained in (a), by taking derivatives $\partial_{x_i}, \partial_{y_i}$ and D_ϕ to this SDE respectively and applying Lemma 3.3, we prove that $\partial_{y_i}DX^{x,\mu}_{s,t}(y)$ and $D^2X^{x,\mu}_{s,t}(y,z)$ are $L^{\infty-}(\mathbb{P})$ bounded continuous in related arguments. We omit the details to save space. □

3.4 Proof of Proposition 3.2

Since $b, \sigma \in C^{0,2,2}_b([0,T] \times \mathbb{R}^d \times \mathscr{P}_2)$, assertions in Lemmas 3.4, 3.5, and 3.9 hold. Then it is straightforward to show that U given in Theorem 3.1(1) is in the class $C^{0,2,2}([0,T] \times \mathscr{P}_2)$.

Firstly, for any $1 \le i \le d$, by taking derivative ∂_{x_i} to the formula of U, we obtain

$$\begin{aligned}\partial_{x_i}U(t,x,\mu) &= \mathbb{E}\Big[\big\langle \nabla\Phi(\cdot,\Lambda^\mu_{t,T})(X^{x,\mu}_{t,T}), \partial_{x_i}X^{x,\mu}_{t,T}\big\rangle \mathrm{e}^{\int_t^T V(r,X^{x,\mu}_{t,r},\Lambda^\mu_{t,r})\mathrm{d}r}\Big]\\ &+ \mathbb{E}\Big[\Phi(X^{x,\mu}_{t,T},\Lambda^\mu_{t,T})\mathrm{e}^{\int_t^T V(r,X^{x,\mu}_{t,r},\Lambda^\mu_{t,r})\mathrm{d}r}\int_t^T \big\langle \nabla V(r,\cdot,\Lambda^\mu_{t,r})(X^{x,\mu}_{t,r}), \partial_{x_i}X^{x,\mu}_{t,r}\big\rangle \mathrm{d}r\Big]\\ &+ \mathbb{E}\int_t^T \big\langle \nabla F(r,\cdot,\Lambda^\mu_{t,r})(X^{x,\mu}_{t,r}), \partial_{x_i}X^{x,\mu}_{t,r}\big\rangle \mathrm{e}^{\int_t^r V(\theta,X^{x,\mu}_{t,\theta},\Lambda^\mu_{t,\theta})\mathrm{d}\theta}\,\mathrm{d}r\\ &+ \mathbb{E}\int_t^T \Big\{F(r,X^{x,\mu}_{t,r},\Lambda^\mu_{t,r})\mathrm{e}^{\int_t^r V(\theta,X^{x,\mu}_{t,\theta},\Lambda^\mu_{t,\theta})\mathrm{d}\theta}\int_t^r \big\langle \nabla V(\theta,\cdot,\Lambda^\mu_{t,\theta})(X^{x,\mu}_{t,\theta}), \partial_{x_i}X^{x,\mu}_{t,\theta}\big\rangle \mathrm{d}\theta\Big\}\,\mathrm{d}r.\end{aligned}$$

By assumptions on Φ, V, F and Lemmas 3.4, 3.5 and 3.9, this formula implies that $\nabla U(t,x,\mu)$ is bounded and continuous. Moreover, by taking derivatives ∂_{x_j} and D to the formula, we conclude that $\nabla^2U(t,x,\mu)$ and $D\{\nabla X^{x,\mu}_{s,t}\}(y)$ are bounded and continuous as well.

Similarly, we may prove the assertion for $DU(t,x,\mu)(y), \partial_{x_i}\{DU(t,x,\mu)(y)\}, \partial_{y_i}\{DU(t,x,\mu)(y)\}$ and $D^2U(t,x,\mu)(y,z)$. For simplicity, we only consider the case for $V = F = 0$, for the general case the formulation is only more complicated due to derivatives to F and V, but there is no any essential difference for the proof. For $V = F = 0$ the formula for U becomes

$$U(t,x,\mu) = \mathbb{E}\Phi(X^{x,\mu}_{t,T}, \Lambda^\mu_{s,t}).$$

Then by (3.13) and the chain rule we obtain

$$\begin{aligned}DU(t,x,\mu)(y) = \mathbb{E}\bigg[&\big\{\nabla\Phi(\cdot,\Lambda^{\mu}_{s,t})(X^{x,\mu}_{s,t})\big\}(DX^{x,\mu}_{s,t})(y) + (\nabla X^{y,\mu}_{s,t})^*\big\{(D\Phi(t,X^{x,\mu}_{s,t},\cdot)(\Lambda^{\mu}_{s,t})\big\}(X^{y,\mu}_{s,t})\\&+\int_{\mathbb{R}^d}(DX^{z,\mu}_{s,t})^*(y)\big\{D\Phi(X^{x,\mu}_{s,t},\cdot)(\Lambda^{\mu}_{s,t})\big\}(X^{z,\mu}_{s,t})\mu(\mathrm{d}z)\bigg].\end{aligned}$$

Since $\Phi \in C_b^{2,2}(\mathbb{R}^d \times \mathscr{P}_2)$, by Lemmas 3.4, 3.5 and 3.9 we deduce from this formula that $DU(t,x,\mu)(y)$ is bounded and continuous. Moreover, by taking derivatives $\partial_{x_i}, \partial_{y_i}, D$ to this formula, we conclude that $\partial_{x_i}\{DU(t,x,\mu)(y)\}, \partial_{y_i}\{DU(t,x,\mu)(y)\}$ and $D^2U(t,x,\mu)(y,z)$ are bounded and continuous as well. In conclusion, $U \in C^{0,2,2}([0,T]\times\mathbb{R}^d\times\mathscr{P}_2)$.

4 Ergodicity and structure of invariant measures

In this part, we assume that $b(t,x,\mu) = b(x,\mu)$ and $\sigma(t,x,\mu) = \sigma(x,\mu)$ are deterministic, and consider the ergodicity of the diffusion processes generated by $\mathscr{A}$ and $\tilde{\mathscr{A}}$.

Recall that a Markov process is called ergodic, if for any initial distribution, when $t \to \infty$ the process converges weakly to the unique invariant probability measure. For square integrable Markov processes, the weak convergence is equivalent to the convergence under the Wasserstein distance. To estimate the Wasserstein distance for solutions to the image SDE (1.7), we take the following hypothesis:

(H) $b(t,x,\mu) = b(x,\mu)$ and $\sigma(t,x,\mu) = \sigma(x,\mu)$ are deterministic, continuous in (x,μ) and do not depend on t. There exist constants $\lambda \in \mathbb{R}$ and $\kappa, \delta, K \geq 0$ such that

$$\begin{aligned}&2\langle b(x,\mu)-b(y,\nu), x-y\rangle + \|\sigma(x,\mu)-\sigma(y,\nu)\|_{HS}^2 \leq \kappa\mathbb{W}(\mu,\nu)^2 - \lambda|x-y|^2,\\&\|\sigma(x,\mu)-\sigma(y,\nu)\|_{HS}^2 \leq K\big\{\mathbb{W}(\mu,\nu)^2 + |x-y|^2\big\},\\&|b(x,\mu)|^2 + \|\sigma(x,\mu)\|_{HS}^2 \leq \delta(1+|x|^2+\|\mu\|_2^2),\quad x,y\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_2.\end{aligned}$$

By Theorem 2.1, **(H)** implies the well-posedness of (1.7). In the present time-homogenous case, we only consider the solution from time $s=0$, i.e. $(X^{x,\mu}_t, \Lambda^\mu_t) := (X^{x,\mu}_{0,t}, \Lambda^\mu_{0,t})$ for $t \geq 0$.

Let $P_t(\mu;\cdot)$ and $\tilde{P}_t(x,\mu;\cdot)$ denote the laws of Λ^μ_t and $(X^{x,\mu}_t, \Lambda^\mu_t)$ respectively. Then the associated Markov semigroups P_t and $\tilde{P}_t$ are given by

$$\begin{aligned}&P_t f(\mu) := \mathbb{E}f(\Lambda^\mu_t) = \int_{\mathscr{P}_2} f(\nu)P_t(\mu;\mathrm{d}\nu),\quad f\in\mathscr{B}_b(\mathscr{P}_2),\\&\tilde{P}_t g(x,\mu) := \mathbb{E}g(X^{x,\mu}_t,\Lambda^\mu_t) = \int_{\mathbb{R}^d\times\mathscr{P}_2} g(y,\nu)\tilde{P}_t(x,\mu;\mathrm{d}y,\mathrm{d}\nu),\quad g\in\mathscr{B}_b(\mathbb{R}^d\times\mathscr{P}_2).\end{aligned}$$

Let $\mathscr{P}_2(\mathscr{P}_2)$ (resp. $\mathscr{P}_2(\mathbb{R}^d\times\mathscr{P}_2)$) be the set of probability measures on $\mathscr{P}_2$ (resp. $\mathbb{R}^d\times\mathscr{P}_2$) with finite second moments, and let $\mathbf{W}_2^{\mathscr{P}_2}$ be the L^2-Warsserstein distance on $\mathscr{P}_2(\mathscr{P}_2)$ induced by $\mathbb{W}_2$, while $\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}$ be that on $\mathscr{P}_2(\mathbb{R}^d\times\mathscr{P}_2)$ induced by the metric

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

For any $Q \in \mathscr{P}_2(\mathscr{P}_2)$ and $\tilde{Q} \in \mathscr{P}_2(\mathbb{R}^d \times \mathscr{P}_2)$, let

$$QP_t = \int_{\mathscr{P}_2} P_t(\mu;\cdot)Q(\mathrm{d}\mu), \quad \tilde{Q}\tilde{P}_t = \int_{\mathbb{R}^d\times\mathscr{P}_2} \tilde{P}_t(x,\mu;\cdot)\tilde{Q}(\mathrm{d}x,\mathrm{d}\mu).$$

In the following two subsections, we first investigate the exponential ergodicity of the diffusion processes generated by $\mathscr{A}$ and $\tilde{\mathscr{A}}$, then figure out the structure of the invariant probability measures.

4.1 Exponential ergodicity

T3.1 **Theorem 4.1.** *Assume* **(H)**. *Then for any* $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$,

LLPT0
$$\mathbb{E}\mathbb{W}_2(\Lambda_t^\mu, \Lambda_t^\nu)^2 \le \mathbb{W}_2(\mu,\nu)^2 \mathrm{e}^{-(\lambda-\kappa)t}, \quad t \ge 0, \tag{4.1}$$

LLPT1
$$\mathbb{E}|X_t^{x,\mu} - X_t^{y,\nu}|^2 \le |x-y|^2\mathrm{e}^{-\lambda t} + \mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{-(\lambda-\kappa)t}, \quad t \ge 0. \tag{4.2}$$

Consequently, if $\lambda > \kappa$ *then:*

(1) $\tilde{P}_t$ *has a unique invariant probability measure* $\tilde{\Pi} \in \mathscr{P}_2(\mathbb{R}^d \times \mathscr{P}_2)$ *such that for any* $\tilde{Q} \in \mathscr{P}_2(\mathbb{R}^d \times \mathscr{P}_2)$,

ES1
$$\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q}\tilde{P}_t, \tilde{\Pi})^2 \le 2\mathrm{e}^{-(\lambda-\kappa)t}\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q},\tilde{\Pi})^2, \quad t \ge 0; \tag{4.3}$$

(2) $\Pi := \tilde{\Pi}(\mathbb{R}^d \times \cdot)$ *is the unique invariant probability measure of* P_t *such that for any* $Q \in \mathscr{P}_2(\mathscr{P}_2)$,

ES2
$$\mathbf{W}_2^{\mathscr{P}_2}(QP_t(\mu;\cdot), \Pi)^2 \le \mathrm{e}^{-(\lambda-\kappa)t}\mathbf{W}_2^{\mathscr{P}_2}(Q,\Pi)^2, \quad t \ge 0. \tag{4.4}$$

Proof. (a) We first prove (4.1) and (4.2). Let $\pi \in \mathscr{C}(\mu,\nu)$ such that

$$\mathbb{W}_2(\mu,\nu)^2 = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\pi(\mathrm{d}x,\mathrm{d}y).$$

Then for any $t \ge 0$,

$$\pi_t := \pi \circ (X_t^{\cdot,\mu}, X_t^{\cdot,\nu})^{-1} \in \mathscr{C}(\Lambda_t^\mu, \Lambda_t^\nu),$$

so that

DR1
$$\mathbb{W}_2(\Lambda_t^\mu,\Lambda_t^\nu)^2 \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\pi_t(\mathrm{d}x,\mathrm{d}y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} |X_t^{x,\mu} - X_t^{y,\nu}|^2\pi(\mathrm{d}x,\mathrm{d}y) =: \ell_t. \tag{4.5}$$

Combining this with **(H)** and Itô's formula, we obtain

$$\mathrm{d}|X_t^{x,\mu} - X_t^{y,\nu}|^2 \le \big\{\kappa\ell_t - \lambda|X_t^{x,\mu} - X_t^{y,\nu}|^2\big\}\mathrm{d}t + \mathrm{d}M_t$$

for some martingale M_t, which implies

DR2
$$\mathrm{e}^{\lambda t}\mathbb{E}|X_t^{x,\mu} - X_t^{y,\nu}|^2 \le |x-y|^2 + \kappa\int_0^t \mathrm{e}^{\lambda s}\mathbb{E}\ell_s\mathrm{d}s, \quad t \ge 0. \tag{4.6}$$

Integrating with respect to $\pi(\mathrm{d}x, \mathrm{d}y)$ gives

$$\mathrm{e}^{\lambda t}\mathbb{E}\ell_t \le \mathbb{W}_2(\mu,\nu)^2 + \kappa \int_0^t \mathrm{e}^{\ell s}\mathbb{E}\ell_s \mathrm{d}s, \quad t \ge 0,$$

which together with Grownwall's lemma and (4.5) leads to

$$\mathbb{E}\mathbb{W}_2(\Lambda_t^\mu, \Lambda_t^\nu)^2 \le \mathbb{E}\ell_t \le \mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{-(\lambda-\kappa)t}, \quad t \ge 0.$$

Thus, (4.1) holds. Substituting (4.1) into (4.6) we arrive at

$$\begin{aligned}\mathbb{E}|X_t^{x,\mu} - X_t^{y,\nu}|^2 &\le \mathrm{e}^{-\lambda t}|x-y|^2 + \kappa\mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{-\lambda t}\int_0^t \mathrm{e}^{\kappa s}\mathrm{d}s \\ &\le \mathrm{e}^{-\lambda t}|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{-(\lambda-\kappa)t}.\end{aligned}$$

Hence, (4.2) holds.

(b) Existence of invariant probability measures. Consider, for instance $(X_t^{0,\delta_0}, \Lambda_t^{\delta_0})$, where δ_0 is the Dirac measure at $0 \in \mathbb{R}^d$. Let $\tilde{\Pi}_t = \tilde{P}_t(0,\delta_0;\cdot)$ be the law of $(X_t^{0,\delta_0}, \Lambda_t^{\delta_0})$. By the completeness of the Wasserstein space, if

INV (4.7)
$$\lim_{s,t\to\infty} \mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{\Pi}_t, \tilde{\Pi}_s)^2 = 0,$$

then there exists a probability measure $\tilde{\Pi}$ on $\mathbb{R}^d \times \mathscr{P}_2$ with $\|\tilde{\Pi}\|_2^2 := \tilde{\Pi}(\rho^2) < \infty$ such that $\lim_{t\to\infty} \mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{\Pi}_t, \tilde{\Pi}) = 0$. Consequently, $\tilde{\Pi}$ is an invariant probability measure for $\tilde{P}_t$. Moreover, since the law of $\Lambda_t^{\delta_0}$ is $\Pi_t(\mathbb{R}^d \times \cdot)$, which converges to $\Pi := \tilde{\Pi}(\mathbb{R}^d \times \cdot)$ weakly as $t \to \infty$, we see that Π is an invariant probability measure of P_t.

To prove (4.7), let $t > s \ge 0$. By the Markov property we have

$$\tilde{\Pi}_t = P_t(0,\delta_0;\cdot) = \int_{\mathbb{R}^d\times\mathscr{P}_2} P_s(x,\mu;\cdot)\tilde{\Pi}_{t-s}(\mathrm{d}x,\mathrm{d}\mu).$$

Combining this with (4.1) and (4.2) we obtain

$$\begin{aligned}&\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}\big(\tilde{\Pi}_t, \tilde{\Pi}_s\big)^2 \le \int_{\mathbb{R}^d\times\mathscr{P}_2} \mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}\big(\tilde{P}_s(x,\mu;\cdot), \tilde{P}_s(0,\delta_0;\cdot)\big)^2\,\tilde{\Pi}_{t-s}(\mathrm{d}y,\mathrm{d}\nu) \\ &\le \int_{\mathbb{R}^d\times\mathscr{P}_2} \big\{\mathbb{E}|X_s^{0,\delta_0} - X_s^{x,\mu}|^2 + \mathbb{W}_2(\Lambda_s^\mu, \Lambda_s^{\delta_0})^2\big\}\tilde{\Pi}_{t-s}(\mathrm{d}x,\mathrm{d}\mu) \\ &\le \int_{\mathbb{R}^d\times\mathscr{P}_2} \big\{|x|^2\mathrm{e}^{-\lambda s} + 2\mathbb{W}_2(\delta_0,\mu)^2\mathrm{e}^{-(\lambda-\kappa)s}\big\}\tilde{\Pi}_{t-s}(\mathrm{d}x,\mathrm{d}\mu) \\ &= \mathrm{e}^{-\lambda s}\mathbb{E}|X_{t-s}^{0,\delta_0}|^2 + 2\mathrm{e}^{-(\lambda-\kappa)s}\mathbb{E}\mathbb{W}_2(\delta_0,\Lambda_{t-s}^{\delta_0})^2 = (\mathrm{e}^{-\lambda s} + 2\mathrm{e}^{-(\lambda-\kappa)s})\mathbb{E}|X_{t-s}^{0,\delta_0}|^2.\end{aligned}$$

So, to prove (4.7) it remains to show that

DR4 (4.8)
$$\sup_{t\ge 0} \mathbb{E}|X_t^{0,\delta_0}|^2 < \infty.$$

By assumption **(H)** with $\lambda > \kappa$, for any $\lambda > \lambda' > \kappa' > \kappa$ there exists a constant $c > 0$ such that

$$2\langle b(x,\mu), x\rangle + \|\sigma(x,\mu)\|_{HS}^2 \le c + \kappa'\|\mu\|_2^2 - \lambda'|x|^2, \quad (x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2.$$

Combining this with Itô's formula, and noting that $\|\Lambda_t^{\delta_0}\|_2^2 = \delta_0(|X_t^{\cdot,\delta_0}|^2) = |X_t^{0,\delta_0}|^2$, we obtain

$$\mathrm{d}|X_t^{0,\delta_0}|^2 \le \big\{c + (\kappa' - \lambda')|X_t^{0,\delta_0}|^2\big\}\mathrm{d}t + \mathrm{d}M_t$$

for some martingale M_t. This implies

$$\mathbb{E}|X_t^{0,\delta_0}|^2 \le c\int_0^t \mathrm{e}^{-(\lambda'-\kappa')s}\mathrm{d}s, \quad t \ge 0.$$

Since $\lambda' > \kappa'$, we derive (4.8) and hence finish the proof of the existence of invariant probability measures. Moreover, the invariant probability measure $\tilde{\Pi}$ satisfies

$$\int_{\mathbb{R}^d\times\mathscr{P}_2} (|x|^2 + \|\mu\|_2^2)\tilde{\Pi}(\mathrm{d}x,\mathrm{d}\mu) \le \lim_{t\to\infty} \mathbb{E}|X_t^{0,\delta_0}|^2 \le \frac{c}{\lambda'-\kappa'} < \infty.$$

Hence, $\tilde{\Pi} \in \mathscr{P}_2(\mathbb{R}^d \times \mathscr{P}_2)$.

(c) It is easy to see that (4.3) follows from (4.1) and (4.2). Indeed, letting $\Gamma \in \mathscr{C}(\tilde{Q}, \tilde{\Pi})$ such that

$$\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q},\tilde{\Pi})^2 = \int_{(\mathbb{R}^d\times\mathscr{P}_2)^2} \rho^2 \mathrm{d}\Gamma,$$

we deduce from (4.1), (4.2) and $\tilde{\Pi} = \tilde{\Pi}\tilde{P}_t$ that

$$\begin{aligned}
&\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q}\tilde{P}_t,\tilde{\Pi})^2 = \mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q}\tilde{P}_t,\tilde{\Pi}\tilde{P}_t)^2\\
&\le \int_{(\mathbb{R}^d\times\mathscr{P}_2)^2} \mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{P}_t(x,\mu;\cdot),\tilde{P}_t(y,\nu;\cdot))^2\Gamma(\mathrm{d}x,\mathrm{d}\mu;\mathrm{d}y,\mathrm{d}\nu)\\
&\le \int_{(\mathbb{R}^d\times\mathscr{P}_2)^2} \mathbb{E}\big\{|X_t^{x,\mu} - X_t^{y,\nu}|^2 + \mathbb{W}_2(\Lambda_t^\mu,\Lambda_t^\nu)^2\big\}\Gamma(\mathrm{d}x,\mathrm{d}\mu;\mathrm{d}y,\mathrm{d}\nu)\\
&\le \int_{(\mathbb{R}^d\times\mathscr{P}_2)^2} \big\{|x-y|^2\mathrm{e}^{-\lambda t} + 2\mathbb{W}_2(\mu,\nu)^2\mathrm{e}^{-(\lambda-\kappa)t}\big\}\Gamma(\mathrm{d}x,\mathrm{d}\mu;\mathrm{d}y,\mathrm{d}\nu)\\
&\le 2\mathrm{e}^{-(\lambda-\kappa)t}\mathbf{W}_2^{\mathbb{R}^d\times\mathscr{P}_2}(\tilde{Q},\tilde{\Pi})^2, \quad t \ge 0.
\end{aligned}$$

In particular, $\tilde{\Pi}$ is the unique invariant probability measure of P_t.

(d) As shown in (b) and (c), (4.1) for $\lambda > \kappa$ implies that P_t has a unique invariant probability measure Π satisfying the estimate (4.4). Noting that $P_t(\mu;\cdot) = \tilde{P}_t(x,\mu;\mathbb{R}^d\times\cdot)$ holds for all $(x,\mu) \in \mathbb{R}^d \times \mathscr{P}_2$, we have $\Pi = \tilde{\Pi}(\mathbb{R}^d \times \cdot)$. □

4.2 Structure of invariant probability measures

Under condition **(H)**, let $b_0(x) = b(x,\delta_x)$ and $\sigma_0(x) = \sigma(x,\delta_x)$. Then the SDE (1.12) is well-posed. Let P_t^0 be the associated Markov semigroup.

T4.2 **Theorem 4.2.** *Assume* **(H)**. *If P_t^0 has an invariant probability measure μ_0, then*

$$\tilde\Pi_0(\mathrm{d}x,\mathrm{d}\mu) := \mu_0(\mathrm{d}x)\delta_{\delta_x}(\mathrm{d}\mu)$$

is an invariant probability measure of $\tilde P_t$. Consequently, $\Pi_0 := \tilde\Pi_0(\mathbb{R}^d\times\cdot) = \int_{\mathbb{R}^d}\delta_{\delta_x}\mu_0(\mathrm{d}x)$ is an invariant probability measure of P_t, and when $\lambda>\kappa$, the unique invariant probability measures $\tilde\Pi$ and Π in Theorem 4.1 *satisfy* (1.13).

Proof. Recall that $(X_t^{x,\mu},\Lambda_t^\mu)$ solve the SDE

$$\mathrm{d}X_t^{x,\mu} = b(X_t^{x,\mu},\Lambda_t^\mu)\mathrm{d}t + \sigma(X_t^{x,\mu},\Lambda_t^\mu)\mathrm{d}W_t,\quad X_0^{x,\mu}=x,$$

where $\Lambda_t^\mu := \mu\circ(X_t^{\cdot,\mu})^{-1}$. Then, when $\mu=\delta_x$ we have $\Lambda_t^\mu=\delta_{X_t^{x,\delta_x}}$, so that $(X_t^{x,\delta_x})_{t\ge0}$ solves the SDE (1.12). By the uniqueness of this SDE and that μ_0 is an invariant probability measure of P_t^0, we obtain

$$\int_{\mathbb{R}^d}\big[\mathbb{E}g(X_t^{x,\delta_x})\big]\mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d}P_t^0 g(x)\mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d}g(x)\mu_0(\mathrm{d}x),\quad t\ge0, g\in\mathscr{B}_b(\mathbb{R}^d).$$

Combining this with $\tilde P_t f(x,\delta_x) = \mathbb{E}f(X_t^{x,\delta_x},\delta_{X_t^{x,\delta_x}})$ for $f\in\mathscr{B}_b(\mathbb{R}^d\times\mathscr{P}_2)$, and taking $g(x)=f(x,\delta_x)$, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\mathscr{P}_2}\tilde P_t f(x,\mu)\tilde\Pi_0(\mathrm{d}x,\mathrm{d}\mu) = \int_{\mathbb{R}^d}\tilde P_t f(x,\delta_x)\mu_0(\mathrm{d}x)\\
&= \int_{\mathbb{R}^d}\big[\mathbb{E}f(X_t^{x,\delta_x},\delta_{X_t^{x,\delta_x}})\big]\mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d}\big[\mathbb{E}g(X_t^{x,\delta_x})\big]\mu_0(\mathrm{d}x)\\
&= \int_{\mathbb{R}^d}g(x)\mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d}f(x,\delta_x)\mu_0(\mathrm{d}x) = \int_{\mathbb{R}^d\times\mathscr{P}_2}f(x,\mu)\tilde\Pi_0(\mathrm{d}x,\mathrm{d}\mu).
\end{aligned}$$

Therefore, $\tilde\Pi_0$ is an invariant probability measure of $\tilde P_t$. In particular, by taking $f(x,\mu)=f(\mu)$, we see that Π_0 is an invariant probability measure of P_t.

Finally, if $\lambda>\kappa$, by Theorem 4.1, Π and $\tilde\Pi$ are the unique invariant probability measures of P_t and $\tilde P_t$ respectively. So, $\tilde\Pi=\tilde\Pi_0$ and $\Pi=\Pi_0$; that is, (1.13) holds. □

Acknowledgement. The author is grateful to the referee for valuable suggestions and to Professor Renming Song for helpful conversations.

References

AKR [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.

AS [2] L. Ambrosio, G. Savaré, *Gradient flows of probability measures,* in “Handbook of Differential Equations: Evolutionary Equations” (C.M. Dafermos and E. Feireisl Eds.), vol. 3, pp. 1–136, Elsevier, 2007.

BCCdRH20 [3] P. Briand, P. Cardaliaguet, P.-E. Chaudru de Raynal, Y. Hu, *Forward and backward stochastic dierential equations with normal constraints in law,* Stoch. Proc. Appl. 2020 online published.

Card [4] 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.

CD18 [5] R. Carmona, F. Delarue, *Probabilistic Theory of Mean Field Games with Applications II. Mean-field Games with Common Noise and Master Equations,* Springer, 2018.

CDL16 [6] R. Carmona, F. Delarue, D. Lacker, *Mean field games with common noise,* Ann. Probab. 44(2016), 3740–3803.

CDLL19 [7] P. Cardaliaguet, F. Delarue, J.-M. Lasry, P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games,* Princeton University Press, 2019.

DLR19 [8] F. Delarue, D. Lacker, K. Ramanan, *From the master equation to mean field game limit theory: A central limit theorem,* Electr. J. of Probab. 24(2019), 1–54.

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

HSS19 [10] W. R. P. Hammersley, D. Šiška, L. Szpruch, *Weak existence and uniqueness for McKean-Vlasov SDEs with common noise,* arXiv:1908.00955

HMC06a [11] M. Huang, R. Malhamé, P. Caines, *Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle,* Communications in Information & Systems, 2006.

HMC06b [12] M. Huang, R. Malhamé, P. Caines, *Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and de-centralized Nash equilibria,* IEEE transactions on automatic control, 2007.

KvR19 [13] V. Konarovskyi and M.-K. von Renesse, *Modied massive arratia flow and Wasserstein diffusion,* Comm. Pure Appl. Math. 72(2019),

KT19 [14] V. N. Kolokoltsov, M. Troeva, *On mean field games with common noise and McKean-Vlasov SPDEs,* Stoch. Anal. Appl. 37(2019), 522–549.

KLV [15] Y. Kondratiev, E. Lytvynov, A. Vershik, *Laplace operators on the cone of Radon measures,* J. Funct. Anal. 269(2015), 2947–2976.

[12] [16] V. Konarovskyi, *A system of coalescing heavy diffusion particles on the real line,* Ann. Probab. 45(2017), 3293–3335.

LL06a [17] J.-M. Lasry, P.-L. Lions, *Jeux á champ moyen (I): Le cas stationnaire,* Comptes Rendus Math. Acad. des Sci. Ser. I 343(2006) 619–625.

LL06b [18] J.-M. Lasry, P.-L. Lions, *Jeux á champ moyen (II): Horizon finiet contrôle optimal,* Comptes Rendus Math. Acad. des Sci. Ser. I 343(2006), 679–684.

LL07 [19] J.-M. Lasry, P.-L. Lions, *Mean field games,* Japan. J. Math. 2(2007), 229–260.

M18 [20] V. Marx, *A new approach for the construction of a Wasserstein diffusion,* Electr. J. Probab. 23(2018), 1–54.

M20 [21] V. Marx, *Innite-dimensional regularization of McKean-Vlasov equation with a Wasserstein diffusion,* arXiv:2002.10157.

RW19 [22] P. Ren, F.-Y. Wang, *Spectral gap for measure-valued diffusion processes,* J. Math. Anal. Appl. 483(2020), 123624.

RS [23] M.-K. von Renesse, K.-T. Sturm, *Entropic measure and Wasserstein diffusion,* Ann. Probab. 37(2009), 1114–1191.

Shao [24] J. Shao, *A new probability measure-valued stochastic process with Ferguson-Dirichlet process as reversible measure,* Electr. J. Probab. 16(2011), 271–292.

W18 [25] F.-Y. Wang, *Distribution dependent SDEs for Landau type equations,* Stoch. Proc. Appl. 128(2018), 595–621.

VII [26] C. Villani, *Optimal Transport Old and New,* Springer-Verlag, Berlin, 2009.

W19 [27] F.-Y. Wang, *Functional inequalities for weighted Gamma distribution on the space of finite measures,* Elect. J. Probab. 25(2020), 1-27.

WZ19 [28] F.-Y. Wang, W. Zhang, *Nash inequality for diffusion processes associated with Dirichlet distributions,* Front. Math. China 14(2019), 1317–1338.