Nonlinear Fokker–Planck equations for Probability Measures on Path Space and Path-Distribution Dependent SDEs*

Xing Huang[a)], **Michael Röckner**[b)], **Feng-Yu Wang**[a),c)]

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

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

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

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

September 05, 2017

Abstract

By investigating path-distribution dependent stochastic differential equations, the following type of nonlinear Fokker–Planck equations for probability measures $(\mu_t)_{t\geq 0}$ on the path space $\mathscr{C} := C([-r_0, 0]; \mathbb{R}^d)$, is analyzed:

$$\partial_t \mu(t) = L^*_{t,\mu_t} \mu_t, \quad t \geq 0,$$

where $\mu(t)$ is the image of μ_t under the projection $\mathscr{C} \ni \xi \mapsto \xi(0) \in \mathbb{R}^d$, and

$$L_{t,\mu}(\xi) := \frac{1}{2} \sum_{i,j=1}^d a_{ij}(t, \xi, \mu) \frac{\partial^2}{\partial_{\xi(0)_i} \partial_{\xi(0)_j}} + \sum_{i=1}^d b_i(t, \xi, \mu) \frac{\partial}{\partial_{\xi(0)_i}}, \quad t \geq 0, \xi \in \mathscr{C}, \mu \in \mathscr{P}^{\mathscr{C}}.$$

Under reasonable conditions on the coefficients a_{ij} and b_i, the existence, uniqueness, Lipschitz continuity in Wasserstein distance, total variational norm and entropy, as well as derivative estimates are derived for the martingale solutions.

AMS subject Classification: 60J75, 47G20, 60G52.
Keywords: Nonlinear PDE for probability measures, path-distribution dependent SDEs, Wasserstein distance, Harnack inequality, coupling by change of measure.

*Supported in part by NNSFC (11771326, 11431014). Financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" is acknowledged.

1 Introduction

In this paper, we investigate nonlinear PDEs for probability measures on the path space using path-distribution dependent SDEs. To explain the motivation of the study, let us start from the following classical PDE on $\mathscr{P}(\mathbb{R}^d)$, the set of probability measures on $\mathbb{R}^d$ equipped with the weak topology:

$$\partial_t \mu(t) = L^* \mu(t). \quad t \geq 0, \tag{1.1}$$

for a second-order differential operator

$$L := \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \partial_i \partial_j + \sum_{i=1}^{d} b_i \partial_i,$$

where $a = (a_{ij}) : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $b = (b_i) : \mathbb{R}^d \to \mathbb{R}^d$ are locally integrable. (1.1) is just the (linear) Fokker–Planck–Kolmogorov equation (FRKE) associated to the operator L in the sense of [2]. We call $\mu \in C(\mathbb{R}_+; \mathscr{P}(\mathbb{R}^d))$ a solution of (1.1), if

$$\int_{\mathbb{R}^d} f \mathrm{d}\mu(t) = \int_{\mathbb{R}^d} f \mathrm{d}\mu(0) + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} (Lf) \mathrm{d}\mu(s), \quad t \geq 0, f \in C_0^\infty(\mathbb{R}^d).$$

To construct and analyze solutions of (1.1) using the time marginal distributions of Markov processes as proposed by A. N. Kolmogorov [10], K. Itô developed the theory of stochastic differential equations (SDEs), see e.g.[9]. Let σ be a matrix-valued function such that $a = \sigma\sigma^*$, and let $W(t)$ be a d-dimensional Brownian motion. Consider the following Itô SDE

$$\mathrm{d}X(t) = b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}W(t). \tag{1.2}$$

By Itô's formula, the time marginals $\mu(t) := \mathscr{L}_{X(t)}$ = the law of $X(t)$ for t ≥ 0, solve the equation (1.1). This enables one to investigate FPKEs using a probabilistic approach.

Obviously, (1.1) is a linear equation. In applications, many important PDEs for probability measures (or probability densities) are nonlinear, see, for instance, [4, 5, 6, 7, 8, 15] and references within for the study of Landau type equations. Such PDEs are also of Fokker–Planck type, but are non-linear (see Sections 6.7 and 9.8 (v) in [2]). To analyze non-linear FPKEs for probability measures, the following distribution-dependent version of (1.2) has been studied in the recent paper [23] by the third named author:

$$\mathrm{d}X(t) = b(t, X(t), \mathscr{L}_{X(t)})\mathrm{d}t + \sigma(t, X(t), \mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{1.3}$$

where

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d, \quad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable. For any $t \geq 0$ and $\mu \in \mathscr{P}(\mathbb{R}^d)$, consider the second order differential operator

$$L_{t,\mu} := \frac{1}{2} \sum_{i,j=1}^{d} (\sigma\sigma^*)_{ij}(t, \cdot, \mu) \partial_i \partial_j + \sum_{i=1}^{d} b_i(t, \cdot, \mu) \partial_i.$$

Under reasonable integrability conditions on σ and b, by Itô's formula we see that for a solution $X(t)$ of (1.3), $\mu(t) := \mathscr{L}_{X(t)}$ solves the nonlinear FPKE

$$\mathrm{d}\mu(t) = L^*_{t,\mu(t)}\mu(t) \tag{1.4}$$

in the sense that

$$\int_{\mathbb{R}^d} f \mathrm{d}\mu(t) = \int_{\mathbb{R}^d} f \mathrm{d}\mu(0) + \int_0^t \mathrm{d}s \int_{\mathbb{R}^d} (L_{s,\mu(s)} f)\mathrm{d}\mu(s), \quad t \geq 0, f \in C_0^\infty(\mathbb{R}^d).$$

In [23], by investigating existence, uniqueness, exponential convergence, and gradient-Harnack type inequalities for the distribution dependent SDE (1.3), the existence of a class of regular solutions to the nonlinear FPKE (1.4) is proved.

In the above two situations, the stochastic systems are Markovian (or memory-free); i.e. the evolution of the system does not depend on its past. However, many real-world models, in particular those arising from mathematical finance and biology, are with memory, so that the associated evolution equations are path dependent. In this case, the distributions of the solution solve non-linear FPKEs for probability measures on path space. In this paper, we investigate such a class of FPKEs by using path-distribution dependent SDEs.

In Section 2, we introduce the framework of the study and the main results on nonlinear FPKEs for probability measures on path space. To prove these results, we investigate the corresponding path-distribution dependent SDEs in Sections 3-5, where strong/weak existence and uniqueness of solutions as well as Harnack type inequalities are derived respectively. We will mainly follow the ideas of [23], but substantial additional efforts have to be made in order to generalize the results in there to the case, where the coefficients do not only depend on the time marginals, but are also on the distribution of the path.

2 Nonlinear PDEs for measures on path space

Throughout the paper, we fix $r_0 > 0$ and consider the path space $\mathscr{C} := C([-r_0, 0]; \mathbb{R}^d)$ equipped with the uniform norm $\|\xi\|_\infty := \sup_{\theta \in [-r_0.0]} |\xi(\theta)|$. Let $\mathscr{P}_2^{\mathscr{C}}$ be the class of probability measures on $\mathscr{C}$ of finite second-order moment, i.e. $\mu(\|\cdot\|_\infty^2) := \int_{\mathscr{C}} \|\xi\|_\infty^2 \mu(\mathrm{d}\xi) < \infty$. Then $\mathscr{P}_2^{\mathscr{C}}$ is a Polish space under the Wasserstein distance

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

where $\mathscr{C}(\mu, \nu)$ denotes the class of couplings for μ and ν. It is well known that $(\mathscr{P}_2^{\mathscr{C}}, \mathbb{W}_2)$ is a Polish space and the $\mathbb{W}_2$-metric is consistent with the weak topology. We will study non-linear FPKEs on $\mathscr{P}_2^{\mathscr{C}}$.

Let

$$b : \mathbb{R}_+ \times \mathscr{C} \times \mathscr{P}_2^{\mathscr{C}} \to \mathbb{R}^d; \quad \sigma : \mathbb{R}_+ \times \mathscr{C} \times \mathscr{P}_2^{\mathscr{C}} \to \mathbb{R}^d \otimes \mathbb{R}^d \tag{2.1}$$

be measurable. For any $t \geq 0, \mu \in \mathscr{P}_2^{\mathscr{C}}$, consider the following differential operator $L_{t,\mu}$ from $C_0^\infty(\mathbb{R}^d)$ to the set of all $\mathscr{B}(\mathscr{C})$-measurable functions: for $f \in C_0^\infty(\mathbb{R}^d)$,

$$(L_{t,\mu}f)(\xi) := \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^*)_{ij}(t,\xi,\mu)(\partial_i\partial_j f)(\xi(0)) + \sum_{i=1}^d b_i(t,\xi,\mu)(\partial_i f)(\xi(0)), \quad \xi \in \mathscr{C}.$$

Then the associated nonlinear FPKE for probability measures $(\mu_t)_{t\geq 0}$ on the path space $\mathscr{C}$ is

$$\partial_t \mu(t) = L^*_{t,\mu_t}\mu_t, \tag{2.2}$$

where $\mu(t)$ is the marginal distribution of μ_t at $\theta = 0$; i.e.

$$\{\mu(t)\}(\mathrm{d}x) := \mu_t(\{\xi \in \mathscr{C} : \xi(0) \in \mathrm{d}x\}).$$

A continuous functional $\mu_\cdot : \mathbb{R}_+ \to \mathscr{P}_2^{\mathscr{C}}$ is called a solution to (2.2), if $\int_0^t \mathrm{d}s \int_{\mathscr{C}} |L_{s,\mu_s}f|\mathrm{d}\mu_s < \infty$ for $f \in C_0^\infty(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} f\mathrm{d}\mu(t) = \int_{\mathbb{R}^d} f\mathrm{d}\mu(0) + \int_0^t \mathrm{d}s \int_{\mathscr{C}} (L_{s,\mu_s}f)\mathrm{d}\mu_s, \quad t \geq 0, \ f \in C_0^\infty(\mathbb{R}^d). \tag{2.3}$$

We will investigate martingale solutions of (2.2) which are realized by marginals of probability measures on the infinite-time path space $\mathscr{C}_\infty := C([-r_0,\infty);\mathbb{R}^d)$. For a probability measure μ^∞ on $\mathscr{C}_\infty$, consider its marginal distributions

$$\mu^\infty(t) := \mu^\infty \circ \{\pi(t)\}^{-1} \in \mathscr{P}(\mathbb{R}^d), \quad \mu_t^\infty := \mu^\infty \circ \pi_t^{-1} \in \mathscr{P}(\mathscr{C}), \quad t \geq 0,$$

where $\pi(t) : C([-r_0,\infty);\mathbb{R}^d) \to \mathbb{R}^d$ and $\pi_t : C([-r_0,\infty);\mathbb{R}^d) \to \mathscr{C}$ are projection operators defined by

$$\pi(t)\xi = \xi(t) \in \mathbb{R}^d, \quad \pi_t\xi = \xi_t \in \mathscr{C} \text{ with } \xi_t(\theta) := \xi(t+\theta) \text{ for } \theta \in [-r_0, 0].$$

Definition 2.1. A solution $(\mu_t)_{t\geq 0}$ of (2.2) is called a martingale solution, if there exists a probability measure μ^∞ on $\mathscr{C}_\infty$ such that

(1) $\mu_t = \mu_t^\infty$ for all $t \geq 0$.

(2) For any $f \in C_0^\infty(\mathbb{R}^d)$, the family of functionals

$$M^f(t) := f(\pi(t)\cdot) - \int_0^t (L_{s,\mu_s}f)(\pi_s\cdot)\mathrm{d}s, \quad t \geq 0$$

on $\mathscr{C}_\infty$ is a μ^∞-martingale; that is,

$$\int_A M^f(t_2)\mathrm{d}\mu^\infty = \int_A M^f(t_1)\mathrm{d}\mu^\infty, \quad t_2 > t_1 \geq 0, \ A \in \sigma(\pi(s) : s \leq t_1),$$

where $\sigma(\pi(s) : s \leq t_1)$ is the σ-field on $\mathscr{C}_\infty$ induced by the projections $\pi(s)$ for $s \in [-r_0, t_1]$.

To construct the martingale solutions of (2.2) using path-distribution dependent SDEs, we need the following assumptions.

$(H1)$ (Continuity) For every $t \ge 0$, $b(t,\cdot,\cdot)$ is continuous on $\mathscr{C} \times \mathscr{P}_2^{\mathscr{C}}$, and there exist locally bounded functions $\alpha_1, \alpha_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\|\sigma(t,\xi,\mu) - \sigma(t,\eta,\nu)\|^2 \le \alpha_1(t)\|\xi - \eta\|_\infty^2 + \alpha_2(t)\mathbb{W}_2(\mu,\nu)^2, \quad t \ge 0; \xi, \eta \in \mathscr{C}; \mu, \nu \in \mathscr{P}_2^{\mathscr{C}}.$$

$(H2)$ (Monotonicity) There exist a constant $\kappa \ge 0$ and locally bounded functions $\beta_1, \beta_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\begin{aligned} &2\langle b(t,\xi,\mu) - b(t,\eta,\nu), \xi(0) - \eta(0)\rangle + \|\sigma(t,\xi,\mu) - \sigma(t,\eta,\nu)\|_{HS}^2 \\ &\le \beta_1(t)\|\xi - \eta\|_\infty^2 + \beta_2(t)\mathbb{W}_2(\mu,\nu)^2 - \kappa|\xi(0) - \eta(0)|^2, \quad t \ge 0; \xi, \eta \in \mathscr{C}; \mu, \nu \in \mathscr{P}_2^{\mathscr{C}}. \end{aligned}$$

$(H3)$ (Growth) b is bounded on bounded sets in $[0,\infty) \times \mathscr{C} \times \mathscr{P}_2^{\mathscr{C}}$, and there exists a locally bounded function $K : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$|b(t,0,\mu)|^2 + \|\sigma(t,0,\mu)\|^2 \le K(t)\big\{1 + \mu(\|\cdot\|_\infty^2)\big\}, \quad t \ge 0, \ \mu \in \mathscr{P}_2^{\mathscr{C}}.$$

The following result characterizes the martingale solutions of (2.2) with $\mathbb{W}_2$-Lipschitz estimate.

Theorem 2.1. *Assume* $(H1)$-$(H3)$. *Then for any* $\mu_0 \in \mathscr{P}_2^{\mathscr{C}}$, *there exists a unique martingale solution* $(\mu_t)_{t\ge0}$ *of* (2.2). *Moreover,*

(1) $\mu_t(\|\cdot\|_\infty^2)$ *is locally bounded in* t.

(2) *For any two martingale solutions* $(\mu_t)_{t\ge0}$ *and* $(\nu_t)_{t\ge0}$ *of* (2.2),

$$\begin{aligned} \mathbb{W}_2(\mu_t,\nu_t)^2 \le \inf_{\varepsilon\in[0,1]} \Bigg\{ &\frac{\mathbb{W}_2(\mu_0,\nu_0)^2}{1-\varepsilon} \\ &\times \inf_{\delta\in[0,\kappa]} \exp\left[(r_0 - t)\delta + \frac{e^{\delta r_0}}{1-\varepsilon}\int_0^t \Big\{\frac{4(\alpha_1(r)+\alpha_2(r))}{\varepsilon} + \beta_1(r) + \beta_2(r)\Big\}dr\right]\Bigg\} \end{aligned}$$

holds for all $t \ge 0$ *and* $\varepsilon \in (0,1)$.

From now on, for any $\nu_0, \mu_0 \in \mathscr{P}_2^{\mathscr{C}}$, we denote μ_t and ν_t the martingale solutions of (2.2) staring at μ_0 and ν_0 respectively.

To estimate the continuity of μ_t in μ_0 with respect to entropy and total variational norm, we make the following stronger assumption.

(A) $\sigma(t,x)$ is invertible, and there exist increasing functions $\kappa_0, \kappa_1, \kappa_2, \lambda : \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $t \ge 0, x, y \in \mathbb{R}^d, \xi, \eta \in \mathscr{C}$ and $\mu, \nu \in \mathscr{P}_2^{\mathscr{C}}$,

$$\begin{aligned} &|b(t,0,\mu)|^2 + \|\sigma(t,x)\|^2 \le \kappa_0(t)(1 + |x|^2 + \mu(\|\cdot\|_\infty^2)), \\ &\|\sigma(t,\cdot)^{-1}\|_\infty \le \lambda(t), \quad \|\sigma(t,x) - \sigma(t,y)\|_{HS}^2 \le \kappa_1(t)|x-y|^2, \\ &|b(t,\xi,\mu) - b(t,\eta,\nu)| \le \kappa_2(t)(\|\xi - \eta\|_\infty + \mathbb{W}_2(\mu,\nu)). \end{aligned}$$

Recall that for any two probability measures μ, ν on some measurable space $(E, \mathscr{F})$, the entropy and variational norm are defined as follows:

$$\mathrm{Ent}(\nu|\mu) := \begin{cases} \int (\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}) \mathrm{d}\nu, & \text{if } \nu \text{ is absolutely continuous with respect to } \mu, \\ \infty, & \text{otherwise;} \end{cases}$$

and

$$\|\mu - \nu\|_{var} := \sup_{A \in \mathscr{F}} |\mu(A) - \nu(A)|.$$

By Pinsker's inequality (see [3, 11]),

$$\|\mu - \nu\|_{var}^2 \le \frac{1}{2} \mathrm{Ent}(\nu|\mu), \quad \mu, \nu \in \mathscr{P}(E). \tag{2.4}$$

Then (2.6) below implies

$$\|\mu_t - \nu_t\|_{var}^2 \le \frac{\psi(t)}{2(t - r_0)} \mathbb{W}_2(\mu_0, \nu_0)^2, \quad t > r_0, \ \mu_0, \nu_0 \in \mathscr{P}_2^{\mathscr{C}}, \tag{2.5}$$

for some $\psi \in C(\mathbb{R}_+; \mathbb{R}_+)$. There are a lot of examples where $\mathbb{W}_2(\mu_n, \mu_0) \to 0$ but μ_n is singular with respect to μ_0 such that $\mathrm{Ent}(\mu_n|\mu_0) = \infty$ and $\|\mu_n - \mu_0\|_{var} = 1$. So, both (2.5) and (2.6) are non-trivial. Indeed, these estimates correspond to the log-Harnack inequality for the associated semigroups, see Theorem 4.1 below for details.

Theorem 2.2. *Assume* **(A)**.

(1) *There exists* $\psi \in C(\mathbb{R}_+; \mathbb{R}_+)$ *such that*

$$\mathrm{Ent}(\nu_t|\mu_t) \le \frac{\psi(t)}{t - r_0} \mathbb{W}_2(\mu_0, \nu_0)^2, \quad t > r_0, \mu_0, \nu_0 \in \mathscr{P}_2^{\mathscr{C}}. \tag{2.6}$$

(2) *If there exists an increasing function* $\kappa_3 : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$\|\sigma(t, x) - \sigma(t, y)\| \le \kappa_3(t)(1 \wedge |x - y|), \quad t \ge 0, x, y \in \mathbb{R}^d, \tag{2.7}$$

then there exists a positive continuous function H *defined on the domain*

$$D := \{(p, t) : t \ge 0, p > (1 + \kappa_3(t)\lambda(t))^2\},$$

such that

$$\int_{\mathscr{C}} \left(\frac{\mathrm{d}\nu_t}{\mathrm{d}\mu_t} \right)^{\frac{1}{p}} \mathrm{d}\nu_t \le \inf_{\pi \in \mathscr{C}(\mu_0, \nu_0)} \int_{\mathscr{C} \times \mathscr{C}} \mathrm{e}^{H(p,t)\left(1 + \frac{|\xi(0) - \eta(0)|^2}{t - r_0} + \|\xi - \eta\|_\infty^2\right)} \mathrm{d}\pi$$

holds for all $t > r_0$ *and* $p > (1 + \kappa_3(t)\lambda(t))^2$.

Remark 2.1. According to Theorem 2.1(2), if there exists a constant $\varepsilon \in (0,1)$ such that

$$\limsup_{t\to\infty} \frac{1}{t}\int_0^t \Big(\frac{4(\alpha_1(s)+\alpha_2(s))}{\varepsilon_t(1-\varepsilon_t)} + \frac{\beta_1(s)+\beta_2(s)}{1-\varepsilon_t}\Big)\mathrm{d}s < \sup_{\delta\in[0,\kappa]} \delta \mathrm{e}^{-\delta r_0}, \tag{2.8}$$

then

$$\mathbb{W}_2(\mu_t,\nu_t)^2 \le c\mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu_0,\nu_0)^2,\quad t\ge 0, \tag{2.9}$$

holds for some constants $c,\lambda>0$; i.e. the solution to (2.2) has exponential contraction in $\mathbb{W}_2$. If $\sigma(t,\cdot,\cdot)$ and $b(t,\cdot,\cdot)$ do not depend on t, i.e. the equation is time-homogenous, we $\mu_t = P_t^*\mu_0$. By the uniqueness we see that P_t^* is a semigroup, i.e. $P_{t+s}^* = P_t^*P_s^*, s,t\ge 0$. Then (2.8) implies that P_t^* has a unique invariant probability measure $\mu\in\mathscr{P}_2^{\mathscr{C}}$. Combining (2.9) with the semigroup property of P_t^* and (2.5)-(2.6), we conclude that (2.8) also implies the exponential convergence in entropy and total variational norm:

$$\max\{\mathrm{Ent}(\nu_t|\mu), \|\mu-\nu_t\|_{var}^2\} \le c_1\mathbb{W}_2(\mu,\nu_{t-1})^2 \le c_2\mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu,\nu_0)^2,\quad t\ge 1, \nu_0\in\mathscr{P}_2^{\mathscr{C}}$$

for some constants $c_1,c_2>0$.

Finally, we investigate the shift quasi-invariance and differentiability of μ_t along Cameron–Martin vectors in $\mathbb{H}^1 := \{\xi\in\mathscr{C} : \int_{-r_0}^0 |\xi'(s)|^2\mathrm{d}s<\infty\}$. For $\eta\in\mathscr{C}$ and a probability measure μ on $\mathscr{C}$, we say that μ is differentiable along ξ if for any $A\in\mathscr{B}(\mathscr{C})$, $\partial_\xi\mu(A) := \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mu(A+\varepsilon\xi)\big|_{\varepsilon=0}$ exists and $\partial_\xi\mu(\cdot)$ is a signed measure on $\mathscr{C}$.

Theorem 2.3. *Assume* **(A)** *and let $b(t,\cdot,\mu)$ be differentiable on $\mathscr{C}$, $\sigma(t,x)=\sigma(t)$ be independent of x. Then for any $t>r_0, \eta\in\mathbb{H}^1$ and $\mu_0\in\mathscr{P}_2^{\mathscr{C}}$, μ_t is differentiable along η, both $\partial_\eta\mu_t$ and $\mu_t(\cdot+\eta)$ are absolutely continuous with respect to μ_t, and for some $\Psi\in C(\mathbb{R}_+;\mathbb{R}_+)$*

$$\begin{aligned}
&\int_{\mathscr{C}} \Big(\log\frac{\mathrm{d}\mu_t(\cdot+\eta)}{\mathrm{d}\mu_t}\Big)\mathrm{d}\mu_t(\cdot+\eta) \le \Psi(t)\Big(\frac{|\eta(-r_0)|^2}{t-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big),\\
&\int_{\mathscr{C}} \Big(\frac{\mathrm{d}\mu_t(\cdot+\eta)}{\mathrm{d}\mu_t}\Big)^{\frac{1}{p}}\mathrm{d}\mu_t(\cdot+\eta) \le \exp\Big[\Psi(t)\Big(\frac{|\eta(-r_0)|^2}{t-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big)\Big],\quad p>1,\\
&\int_{\mathscr{C}} \Big|\frac{\mathrm{d}\partial_\eta\mu_t}{\mathrm{d}\mu_t}\Big|^2\mathrm{d}\mu_t \le \Psi(t)\Big(\frac{|\eta(-r_0)|^2}{t-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big).
\end{aligned}$$

Proof of Theorems 2.1-2.3. For $\mu_0\in\mathscr{P}_2^{\mathscr{C}}$, take a $\mathscr{F}_0$-measurable random variable X_0 on $\mathscr{C}$ such that $\mathscr{L}_{X_0}=\mu_0$. According to Theorem 3.1, Corollary 4.2, Corollary 5.2 and (2.4), $\mu_t := \mathscr{L}_{X_t}$ satisfies the estimates in Theorems 2.1-2.3 under the corresponding assumptions. So, it suffices to show that $(\mathscr{L}_{X_t})_{t\ge 0}$ is the unique martingale solution of (2.2).

Let $\mu^\infty = \mathscr{L}_{\{X(s)\}_{s\in[-r_0.\infty)}}$. We have $\mathscr{L}_{X_t} = \mu_t^\infty$. By (3.1) and Itô's formula, for any $f\in C_0^\infty(\mathbb{R}^d)$, $(M^f(t))_{t\ge 0}$ is a μ^∞-martingale such that $\mu_t := \mathscr{L}_{X_t}$ satisfies

$$\int_{\mathbb{R}^d} f\mathrm{d}\mu(t) = \mathbb{E}f(X(t)) = \mathbb{E}f(X(0)) + \int_0^t \mathbb{E}(L_{s,\mu_s}f)(X_s)\mathrm{d}s$$

$$= \int_{\mathbb{R}^d} f \mathrm{d}\mu(0) + \int_0^t \mathrm{d}s \int_{\mathscr{C}} (L_{s,\mu_s} f) \mathrm{d}\mu_s, \quad t \geq 0, f \in C_0^\infty(\mathbb{R}^d).$$

Therefore, $\mathscr{L}_{\{X(s)\}_{s\in[-r_0.\infty)}}$ is a martingale solution of (2.2). When the coefficients are distribution-free, it is well known that the weak solution of (3.1) is equivalent to the martingale solution, so that the uniqueness of the martingale solutions of (2.2) follows from Theorem 3.1(3) below. In the following, we explain that the same is true for the present distribution dependent case.

Let $\mu_t = \mu_t^\infty$, for some probability measure μ^∞ on $\mathscr{C}_\infty$, be a martingale solution of (2.2). We intend to prove $\mu^\infty = \mathscr{L}_{\{X(s)\}_{s\in[-r_0.\infty)}}$, so that the martingale solution is unique. Let $\bar{\Omega} := \mathscr{C}_\infty, \bar{\mathscr{F}}_t$ for $t \geq 0$ be the completion of $\sigma(\pi(s) : s \leq t)$ with respect to μ^∞, and $\bar{\mathbb{P}} := \mu^\infty$. By Theorem 3.1(3) below, it suffices to prove that the coordinate process

$$\bar{X}(t)(\omega) := \omega(t), \quad t \geq 0, \omega \in \bar{\Omega}$$

is a weak solution to (3.1). To this end, for the given $(\mu_t)_{t\geq 0}$, define

$$\bar{\sigma}(t,\xi) := \sigma(t,\xi,\mu_t), \quad \bar{b}(t,\xi) = b(t,\xi,\mu_t), \quad t \geq 0, \ \xi \in \mathscr{C},$$

and consider the corresponding operator

$$(\bar{L}_t f)(\xi) := \frac{1}{2} \sum_{i,j=1}^d (\bar{\sigma}\bar{\sigma}^*)_{ij}(t,\xi)(\partial_i \partial_j f)(\xi(0)) + \sum_{i=1}^d \bar{b}_i(t,\xi)(\partial_i f)(\xi(0)), \quad t \geq 0, \xi \in \mathscr{C}$$

for $f \in C_0^\infty(\mathbb{R}^d)$. Since $(\mu_t)_{t\geq 0}$ is a martingale solution of (2.2), for any $f \in C_0^\infty(\mathbb{R}^d)$, the process

$$M^f(t) := f(\bar{X}(t)) - f(\bar{X}(0)) - \int_0^t (\bar{L}_s f)(\bar{X}_s) \mathrm{d}s, \quad t \geq 0$$

is a martingale on the probability space $(\bar{\Omega}, (\bar{\mathscr{F}}_t)_{t\geq 0}, \bar{\mathbb{P}})$. By $(H1)$-$(H3)$, the martingale property also holds for f being polynomials of order 2. In particular, by taking $f(x) = x$ we see that

$$M(t) := M^f(t) = \bar{X}(t) - \bar{X}(0) - \int_0^t \bar{b}(s, \bar{X}_s) \mathrm{d}s \tag{2.10}$$

is a $\mathbb{R}^d$-valued martingale, and with $f(x) := x_i x_j$ we conclude that

$$\langle M_i, M_j \rangle(t) = \int_0^t (\bar{\sigma}\bar{\sigma}^*)_{ij}(s, \bar{X}_s) \mathrm{d}s, \quad 1 \leq i,j \leq d.$$

Then according to Stroock–Varadhan (see, for example, Theorems 4.5.1 and 4.5.2 in [13]), we may construct a d-dimensional Brownian motion $\tilde{W}(t)$ on a product probability space of $(\tilde{\Omega}, \tilde{\mathscr{F}}_t, \tilde{\mathbb{P}})$ with $(\bar{\Omega}, \bar{\mathscr{F}}_t, \bar{\mathbb{P}})$ as a marginal space, and when σ is invertible these two spaces coincide, such that

$$M(t) = \int_0^t \bar{\sigma}(s, \bar{X}_s) \mathrm{d}\tilde{W}(s), \quad t \geq 0.$$

Combining this with (2.10), we see that $\bar{X}(t)$ solves the stochastic functional differential equation

$$\mathrm{d}\bar{X}(t) = \bar{b}(t, \bar{X}_t)\mathrm{d}t + \bar{\sigma}(t, \bar{X}_t)\mathrm{d}\tilde{W}(t) \tag{2.11}$$

with $\mathscr{L}_{\bar{X}_0}|_{\tilde{\mathbb{P}}} = \mathscr{L}_{\bar{X}_0}|_{\bar{\mathbb{P}}} = \mu_0$. Since, by definition, $\mu_t = \mathscr{L}_{\bar{X}_t}|_{\bar{\mathbb{P}}} = \mathscr{L}_{\bar{X}_t}|_{\tilde{\mathbb{P}}}$, $\bar{X}(t)$ solves the path-distribution dependent SDE

$$\mathrm{d}\bar{X}(t) = b(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t}|_{\tilde{\mathbb{P}}})\mathrm{d}t + \sigma(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t}|_{\tilde{\mathbb{P}}})\mathrm{d}\tilde{W}(t),$$

i.e. $(\bar{X}, \tilde{W})$ is a weak solution of (3.1). Noting that $\mu^\infty := \mathscr{L}_{\bar{X}|\bar{\mathbb{P}}} = \mathscr{L}_{\bar{X}|\tilde{\mathbb{P}}}$, by the weak uniqueness of (3.1) due to Theorem 3.1(3) below, we obtain $\mu^\infty = \mathscr{L}_{\{X(s)\}_{s\in[-r_0,\infty)}}$ as desired. □

3 Path-distribution dependent SDEs

Recall that for $\gamma(\cdot) \in C([-r_0, \infty); \mathbb{R}^d)$, the segment functional $\gamma_\cdot \in C(\mathbb{R}_+; \mathscr{C})$ is defined by

$$\gamma_t(\theta) := \gamma(t + \theta), \quad \theta \in [-r_0, 0], t \geq 0.$$

For σ, b in (2.1), consider the following path-distribution dependent SDE on $\mathbb{R}^d$:

$$\mathrm{d}X(t) = b(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}W(t), \tag{3.1}$$

where $W = (W(t))_{t\geq 0}$ is a d-dimensional standard Brownian motion with respect to a complete filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, $\mathscr{L}_{X_t}$ is the distribution of X_t. We investigate the strong solutions of (3.1) and determine properties, of their distributions.

We first recall the definition of the strong and weak solutions, see for instance [23, Definition 1.1] in the path independent setting. For simplicity, we will only consider square integrable solutions.

Definition 3.1. (1) For any $s \geq 0$, a continuous adapted process $(X_{s,t})_{t\geq s}$ on $\mathscr{C}$ is called a (strong) solution of (3.1) from time s, if

$$\mathbb{E}\|X_{s,t}\|_\infty^2 + \int_s^t \mathbb{E}\big\{|b(r, X_{s,r}, \mathscr{L}_{X_{s,r}})| + \|\sigma(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\|^2\big\}\mathrm{d}r < \infty, \quad t \geq s,$$

and $(X_{s,}(t) := X_{s,t}(0))_{t\geq s}$ satisfies $\mathbb{P}$-a.s.

$$X_{s,}(t) = X_s(s) + \int_s^t b(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\mathrm{d}r + \int_s^t \sigma(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\mathrm{d}W(r), \quad t \geq s.$$

We say that (3.1) has (strong or pathwise) existence and uniqueness, if for any $s \geq 0$ and $\mathscr{F}_s$-measurable random variable $X_{s,s}$ with $\mathbb{E}\|X_{s,s}\|_\infty^2 < \infty$, the equation from time s has a unique solution $(X_{s,t})_{t\geq s}$. When $s = 0$ we simply denote $X_{0,} = X$; i.e. $X_{0,}(t) = X(t), X_{0,t} = X_t, t \geq 0$.

(2) A couple $(\tilde X_{s,t}, \tilde W(t))_{t\ge s}$ is called a weak solution to (3.1) from time s, if $\tilde W(t)$ is a d-dimensional Brownian motion a complete filtered probability space $(\tilde\Omega, \{\tilde{\mathscr F}_t\}_{t\ge s}, \tilde{\mathbb P})$, and $\tilde X_{s,t}$ solves

$$\mathrm{d}\tilde X_{s,}(t) = b(t, \tilde X_{s,t}, \mathscr L_{\tilde X_{s,t}}|_{\tilde{\mathbb P}})\mathrm{d}t + \sigma(t, \tilde X_{s,t}, \mathscr L_{\tilde X_{s,t}}|_{\tilde{\mathbb P}})\mathrm{d}\tilde W(t), \quad t \ge s. \tag{3.2}$$

(3) (3.1) is said to satisfy weak uniqueness, if for any $s\ge 0$, the distribution of a weak solution $(X_{s,t})_{t\ge s}$ to (3.1) from $s \ge 0$ is uniquely determined by $\mathscr L_{X_{s,s}}$.

When (3.1) has strong existence and uniqueness, the solution $(X_t)_{t\ge0}$ is a Markov process in the sense that for any $s\ge 0$, $(X_t)_{t\ge s}$ is determined by solving the equation from time s with initial state X_s. More precisely, letting $\{X^\xi_{s,t}\}_{t\ge s}$ denote the solution of the equation from time s with initial state $X_{s,s}=\xi$, the existence and uniqueness imply

$$X^\xi_{s,t} = X^{X^\xi_{s,u}}_{u,t}, \quad t\ge u\ge s\ge 0, \xi \text{ is } \mathscr F_s\text{-measurable with } \mathbb E\|\xi\|_\infty^2 < \infty. \tag{3.3}$$

When (3.1) also has weak uniqueness, we may define a semigroup $(P^*_{s,t})_{t\ge s}$ on $\mathscr P_2^{\mathscr C}$ by letting $P^*_{s,t}\mu = \mathscr L_{X_{s,t}}$ for $\mathscr L_{X_{s,s}} = \mu \in \mathscr P_2^{\mathscr C}$. Indeed, by (3.3) we have

$$P^*_{s,t} = P^*_{u,t}P^*_{s,u}, \quad t\ge u\ge s\ge 0. \tag{3.4}$$

For simplicity we set $P^*_t = P^*_{0,t},\ t\ge 0$.

Theorem 3.1. *Assume (H1)-(H3).*

(1) *For any $s \ge 0$ and $X_{s,s}\in L^2(\Omega\to\mathscr C;\mathscr F_s)$, (3.1) has a unique strong solution $(X_{s,t})_{t\ge s}$ with*

$$\mathbb E \sup_{t\in[s,T]} \|X_{s,t}\|_\infty^2 \le H(T)(1+\mathbb E\|X_{s,s}\|_\infty^2), \quad T\ge t\ge s\ge 0 \tag{3.5}$$

for some increasing function $H:\mathbb R_+\to\mathbb R_+$.

(2) *For any two solutions $X_{s,t}$ and $Y_{s,t}$ of (3.1) with $\mathscr L_{X_{s,s}}, \mathscr L_{Y_{s,s}} \in \mathscr P_2^{\mathscr C}$,*

$$\begin{aligned}\mathbb E\|X_{s,t}-Y_{s,t}\|_\infty^2 \le \inf_{\varepsilon\in(0,1)}\bigg\{&\frac{\mathbb E\|X_{s,s}-Y_{s,s}\|_\infty^2}{1-\varepsilon}\\ \times \inf_{\delta\in[0,\kappa],\varepsilon\in(0,1)} \exp\bigg[&(r_0+s-t)\delta + \frac{\mathrm e^{\delta r_0}}{1-\varepsilon}\int_s^t \Big\{\frac{4(\alpha_1(r)+\alpha_2(r))}{\varepsilon} + \beta_1(r)+\beta_2(r)\Big\}\mathrm{d}r\bigg]\bigg\}.\end{aligned}$$

(3) *(3.1) satisfies weak uniqueness, and for any $t\ge 0$,*

$$\begin{aligned}\mathbb W_2(P_t^*\mu_0, P_t^*\nu_0)^2 \le \inf_{\varepsilon\in(0,1)}\bigg\{&\frac{\mathbb W_2(\mu_0,\nu_0)^2}{1-\varepsilon}\\ \times \inf_{\delta\in[0,\kappa],\varepsilon\in(0,1)} \exp\bigg[&(r_0-t)\delta + \frac{\mathrm e^{\delta r_0}}{1-\varepsilon}\int_0^t \Big\{\frac{4(\alpha_1(r)+\alpha_2(r))}{\varepsilon} + \beta_1(r)+\beta_2(r)\Big\}\mathrm{d}r\bigg].\end{aligned}$$

We will prove this result by using the argument of [23]. For fixed $s \ge 0$ and $\mathscr{F}_s$-measurable $\mathscr{C}$-valued random variable $X_{s,s}$ with $\mathbb{E}\|X_{s,s}\|_\infty^2 < \infty$, we construct the solution of (3.1) by iterating in distribution as follows. Firstly, let

$$X_{s,t}^{(0)}(\theta) = X_{s,s}\big(0 \wedge (t-s+\theta)\big) \text{ for } \theta \in [-r_0, 0],\ \mu_{s,t}^{(0)} = \mathscr{L}_{X_{s,t}^{(0)}}, \quad t \ge s.$$

For any $n \ge 1$, let $(X_{s,t}^{(n)})_{t \ge s}$ solve the classical path-dependent SDE

$$\mathrm{d}X_{s,}^{(n)}(t) = b(t, X_{s,t}^{(n)}, \mu_{s,t}^{(n-1)})\mathrm{d}t + \sigma(t, X_{s,t}^{(n)}, \mu_{s,t}^{(n-1)})\,\mathrm{d}W(t), \quad X_{s,s}^{(n)} = X_{s,s}, t \ge s, \tag{3.6}$$

where $\mu_{s,t}^{(n-1)} := \mathscr{L}_{X_{s,t}^{(n-1)}}$ and $X_{s,t}^{(n)}(\theta) := X_{s,}^{(n)}(t-s+\theta)$ for $\theta \in [-r_0, 0]$.

Lemma 3.2. *Assume (H1)-(H3). For every $n \ge 1$, the path-dependent SDE (3.6) has a unique strong solution $X_{s,t}^{(n)}$ with*

$$\mathbb{E} \sup_{t \in [s-r_0, T]} |X_{s,}^{(n)}(t)|^2 < \infty, \quad T > s, n \ge 1. \tag{3.7}$$

Moreover, for any $T > 0$, there exists $t_0 > 0$ such that for all $s \in [0,T]$ and $X_{s,s} \in L^2(\Omega \to \mathscr{C}; \mathscr{F}_s)$,

$$\mathbb{E} \sup_{t \in [s, s+t_0]} |X_{s,}^{(n+1)}(t) - X_{s,}^{(n)}(t)|^2 \le 4\mathrm{e}^{-n} \mathbb{E} \sup_{t \in [s, s+t_0]} |X_{s,}^{(1)}(t)|^2, \quad s \in [0,T], n \ge 1. \tag{3.8}$$

Proof. The proof is similar to that of [23, Lemma 2.1]. Without loss of generality, we may assume that $s = 0$ and simply denote $X_{0,}(t) = X(t), X_{0,t} = X_t, t \ge 0$.

(1) We first prove that the SDE (3.6) has a unique strong solution and (3.7) holds.

For $n = 1$, let

$$\bar{b}(t,\xi) = b(t,\xi,\mu_t^{(0)}), \quad \bar{\sigma}(t,\xi) = \sigma(t,\xi,\mu_t^{(0)}), \quad t \ge 0, \xi \in \mathscr{C}.$$

Then (3.6) reduces to

$$\mathrm{d}X^{(1)}(t) = \bar{b}(t, X_t^{(1)})\mathrm{d}t + \bar{\sigma}(t, X_t^{(1)})\mathrm{d}W(t), \quad X_0^{(1)} = X_0, t \ge 0. \tag{3.9}$$

By (H1)-(H3), the coefficients $\bar{b}$ and $\bar{\sigma}$ satisfy the standard monotonicity condition which imply strong existence, uniqueness and non-explosion for the stochastic functional differential equation (3.9), see e.g. [18, Corollary 4.1.2] with $D = \mathbb{R}^d$ and $u_n = 1$. It is also standard to prove (3.7) using Itô's formula

$$\begin{aligned}
\mathrm{d}|X^{(1)}(t)|^2 &= 2\big\langle \sigma(t, X_t^{(1)}, \mu_t^{(0)})\mathrm{d}W(t), X^{(1)}(t)\big\rangle \\
&\quad + \big\{2\big\langle b(t, X_t^{(1)}, \mu_t^{(0)}), X^{(1)}(t)\big\rangle + \|\sigma(t, X_t^{(1)}, \mu_t^{(0)})\|_{HS}^2\big\}\mathrm{d}t.
\end{aligned}$$

By (H1)-(H3), there exists an increasing function $H : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$2\big\langle b(t,\xi,\mu_t^{(0)}), \xi(0)\big\rangle + \|\sigma(t,\xi,\mu_t^{(0)})\|_{HS}^2$$

$$
\begin{aligned}
&\le 2\big\langle b(t,\xi,\mu_t^{(0)}) - b(t,0,\mu_t^{(0)}), \xi(0)\big\rangle + 2|b(t,0,\mu_t^{(0)})| \cdot |\xi(0)| \\
&+ 2\|\sigma(t,\xi,\mu_t^{(0)}) - \sigma(t,0,\mu_t^{(0)})\|_{HS}^2 + 2\|\sigma(t,0,\mu_t^{(0)})\|_{HS}^2 \\
&\le H(t)\big\{1 + \|\xi\|_\infty^2 + \mu_t^{(0)}(\|\cdot\|_\infty^2)\big\}, \quad t \ge 0, \xi \in \mathscr{C}.
\end{aligned}
$$

Combining this with $(H3)$ and applying the BDG inequality for $p = 1$, for any $N \in [1,\infty)$ and $\tau_N := \inf\{t \ge 0 : |X^{(1)}(t)| \ge N\}$, we have

$$
\begin{aligned}
&\mathbb{E} \sup_{s\in[-r_0,t\wedge\tau_N]} |X^{(1)}(s)|^2 \le 4\mathbb{E}\|X_0^{(1)}\|_\infty^2 + 2H(t)\mathbb{E}\int_0^{t\wedge\tau_N} \big(1 + \|X_s^{(1)}\|_\infty^2 + \mu_s^{(0)}(\|\cdot\|_\infty^2)\big)\mathrm{d}s \\
&\qquad + 4H(t)\mathbb{E}\bigg(\int_0^{t\wedge\tau_N} |X^{(1)}(s)|^2\big(1 + \|X_s^{(1)}\|_\infty^2 + \mu_s^{(0)}(\|\cdot\|_\infty^2)\big)\mathrm{d}s\bigg)^{\frac{1}{2}} \\
&\le 4\mathbb{E}\|X_0^{(1)}\|_\infty^2 + \frac{1}{2}\mathbb{E}\sup_{s\in[-r_0,t\wedge\tau_N]} |X^{(1)}(s)|^2 \\
&+ \{2H(t) + 8H(t)^2\}\mathbb{E}\int_0^{t\wedge\tau_N} \big(1 + \|X_s^{(1)}\|_\infty^2 + \mu_s^{(0)}(\|\cdot\|_\infty^2)\big)\mathrm{d}s, \quad t \ge 0.
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\mathbb{E}\sup_{s\in[-r_0,t\wedge\tau_N]} |X_s^{(1)}|^2 \le 8\mathbb{E}\|X_0^{(1)}\|_\infty^2 \\
&+ \{4H(t) + 16H(t)^2\}\int_0^t \Big\{1 + \mathbb{E}\sup_{r\in[-r_0,s\wedge\tau_N]} |X^{(1)}(r)|^2 + \mu_s^{(0)}(\|\cdot\|_\infty^2)\Big\}\mathrm{d}s, \quad t \ge 0.
\end{aligned}
$$

By first applying Gronwall's Lemma then letting $N \to \infty$, we arrive at

$$
\mathbb{E}\sup_{s\in[-r_0,t]} |X^{(1)}(s)|^2 < \infty, \quad t \ge 0.
$$

Therefore, (3.7) holds for $n = 1$.

Now, assuming that the assertion holds for $n = k$ for some $k \ge 1$, we intend to prove it for $n = k+1$. This can be done by repeating the above argument with $(X_\cdot^{(k+1)}, \mu_\cdot^{(k)}, X_\cdot^{(k)})$ replacing $(X_\cdot^{(1)}, \mu_\cdot^{(0)}, X_\cdot^{(0)})$, so, we omit the proof.

(2) To prove (3.8), let

$$
\begin{aligned}
&\xi^{(n)}(t) = X^{(n+1)}(t) - X^{(n)}(t), \\
&\Lambda_t^{(n)} = \sigma(t, X_t^{(n+1)}, \mu_t^{(n)}) - \sigma(t, X_t^{(n)}, \mu_t^{(n-1)}), \\
&B_t^{(n)} = b(t, X_t^{(n+1)}, \mu_t^{(n)}) - b(t, X_t^{(n)}, \mu_t^{(n-1)}).
\end{aligned}
$$

By $(H2)$ and Itô's formula, there exists an increasing function $K_1 : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$
\mathrm{d}|\xi^{(n)}(t)|^2 \le 2\langle \Lambda_t^{(n)}\mathrm{d}W(t), \xi^{(n)}(t)\rangle + K_1(t)\big\{\|\xi_t^{(n)}\|_\infty^2 + \mathbb{W}_2(\mu_t^{(n)}, \mu_t^{(n-1)})^2\big\}\mathrm{d}t.
$$

By the BDG inequality for $p = 1$ and since $\mathbb{W}_2(\mu_s^{(n)}, \mu_s^{(n-1)})^2 \le \mathbb{E}\|\xi_s^{(n)}\|_\infty^2$, we obtain

$$
\mathbb{E}\sup_{s\in[0,t]} |\xi^{(n)}(s)|^2 \le 2\mathbb{E}\sup_{s\in[0,t]}\int_0^s \langle \Lambda_r^{(n)}\mathrm{d}W(r), \xi^{(n)}(r)\rangle
$$

$$
\begin{aligned}
&\quad + K_1(t)\int_0^t \Big\{\mathbb{E}\|\xi_s^{(n)}\|_\infty^2 + \mathbb{W}_2(\mu_s^{(n)},\mu_s^{(n-1)})^2\Big\}\mathrm{d}s\\
&\le 4\mathbb{E}\bigg(\int_0^t \{|\xi^{(n)}(s)|^2\|\Lambda_s^{(n)}\|^2\}\mathrm{d}s\bigg)^{\frac12} + K_1(t)\int_0^t \Big\{\mathbb{E}\|\xi_s^{(n)}\|_\infty^2 + \mathbb{W}_2(\mu_s^{(n)},\mu_s^{(n-1)})^2\Big\}\mathrm{d}s\\
&\le \frac12\mathbb{E}\sup_{s\in[0,t]}|\xi^{(n)}(s)|^2 + 8\int_0^t \mathbb{E}\|\Lambda_s^{(n)}\|^2\mathrm{d}s + K_1(t)\int_0^t \Big\{\mathbb{E}\|\xi_s^{(n)}\|_\infty^2 + \mathbb{W}_2(\mu_s^{(n)},\mu_s^{(n-1)})^2\Big\}\mathrm{d}s.
\end{aligned}
$$

Combining this and $(H1)$ we deduce that

$$
\mathbb{E}\sup_{s\in[0,t]}|\xi^{(n)}(s)|^2 \le K_2(t)\int_0^t \Big\{\mathbb{E}\sup_{r\in[0,s]}|\xi^{(n)}(r)|^2 + \mathbb{W}_2(\mu_s^{(n)},\mu_s^{(n-1)})^2\Big\}\mathrm{d}s,\quad t\ge 0
$$

for some increasing function $K_2:\mathbb{R}_+\to\mathbb{R}_+$. By Gronwall's Lemma, we obtain

$$
\begin{aligned}
\mathbb{E}\sup_{s\in[0,t]}|\xi^{(n)}(s)|^2 &\le tK_2(t)\mathrm{e}^{tK_2(t)}\sup_{s\in[0,t]}\mathbb{W}_2(\mu_s^{(n)},\mu_s^{(n-1)})^2\\
&\le tK_2(t)\mathrm{e}^{tK_2(t)}\mathbb{E}\sup_{s\in[0,t]}|\xi^{(n-1)}(s)|^2,\quad t\ge 0.
\end{aligned}
$$

Taking $t_0>0$ such that $t_0K_2(T)\mathrm{e}^{t_0K_2(T)}\le \mathrm{e}^{-1}$, we arrive at

$$
\mathbb{E}\sup_{s\in[0,t_0]}|\xi^{(n)}(s)|^2 \le \mathrm{e}^{-1}\mathbb{E}\sup_{s\in[0,t_0]}|\xi^{(n-1)}(s)|^2,\quad n\ge 1.
$$

Since

$$
\mathbb{E}\sup_{s\in[0,t_0]}|\xi^{(0)}(s)|^2 \le 2\mathbb{E}\Big\{|X(0)|^2 + \sup_{s\in[0,t_0]}|X^{(1)}(s)|^2\Big\} \le 4\mathbb{E}\sup_{s\in[0,t_0]}|X^{(1)}(s)|^2,
$$

we obtain (3.8). □

Proof of Theorem 3.1. Without loss of generality, we only consider $s=0$ and simply denote $X_{0,}=X$; i.e. $X_{0,}(t)=X(t), X_{0,t}=X_t, t\ge 0$.

(1) Since the uniqueness follows from Theorem 3.1(2), which will be proved in the next step, in this step we only prove existence and estimate (3.5). By Lemma 3.2, there exists a unique adapted continuous process $(X_t)_{t\in[0,t_0]}$ such that

$$
\lim_{n\to\infty}\sup_{t\in[0,t_0]}\mathbb{W}_2(\mu_t^{(n)},\mu_t)^2 \le \lim_{n\to\infty}\mathbb{E}\sup_{t\in[0,t_0]}|X^{(n)}(t)-X(t)|^2 = 0, \tag{3.10}
$$

where μ_t is the distribution of X_t. By (3.6),

$$
X^{(n)}(t) = X(0) + \int_0^t b(s,X_s^{(n)},\mu_s^{(n-1)})\mathrm{d}s + \int_0^t \sigma(s,X_s^{(n)},\mu_s^{(n-1)})\mathrm{d}W(s).
$$

Then (3.10), $(H1)$, $(H3)$ and the dominated convergence theorem imply that $\mathbb{P}$-a.s.

$$
X(t) = X(0) + \int_0^t b(s,X_s,\mu_s)\mathrm{d}s + \int_0^t \sigma(s,X_s,\mu_s)\mathrm{d}W(s),\quad t\in[0,t_0].
$$

Therefore, $(X_t)_{t\in[0,t_0]}$ solves (3.1) up to time t_0, and (3.10) implies $\mathbb{E}\sup_{s\in[0,t_0]}|X(s)|^2<\infty$. The same holds for $(X_{s,t})_{t\in[s,(s+t_0)\wedge T]}$ and $s\in[0,T]$. So, by solving the equation piecewise in time, and using the arbitrariness of $T>0$, we conclude that (3.1) has a strong solution $(X_t)_{t\geq 0}$ with

$$\mathbb{E}\sup_{s\in[0,t]}|X(s)|^2<\infty,\quad t\geq 0. \tag{3.11}$$

(2) By Itô's formula and $(H2)$, we have

$$\begin{aligned}\mathrm{d}\{\mathrm{e}^{\kappa t}|X(t)-Y(t)|^2\}\leq &2\mathrm{e}^{\kappa t}\big\langle X(t)-Y(t),\{\sigma(t,X_t,\mathscr{L}_{X_t})-\sigma(t,Y_t,\mathscr{L}_{Y_t})\}\mathrm{d}W(t)\big\rangle\\ &+\mathrm{e}^{\kappa t}\big\{\beta_1(t)\|X_t-Y_t\|_\infty^2+\beta_2(t)\mathbb{W}_2(\mathscr{L}_{X_t},\mathscr{L}_{Y_t})^2\big\}\mathrm{d}t.\end{aligned}$$

Noting that $\mathbb{W}_2(\mathscr{L}_{X_t},\mathscr{L}_{Y_t})^2\leq\mathbb{E}\|X_t-Y_t\|_\infty^2$, we see that $\gamma_t:=\sup_{s\in[-r_0,t]}\mathrm{e}^{\kappa s^+}|X(s)-Y(s)|^2$ satisfies

$$\begin{aligned}\mathbb{E}\gamma_t\leq&\mathbb{E}\|X_0-Y_0\|_\infty^2+\mathbb{E}\int_0^t(\beta_1+\beta_2)(r)\mathrm{e}^{\kappa r}\|X_r-Y_r\|_\infty^2\mathrm{d}s\\ &+2\mathbb{E}\sup_{s\in[0,t]}\int_0^s\mathrm{e}^{\kappa r}\big\langle X(r)-Y(r),\{\sigma(r,X_r,\mathscr{L}_{X_r})-\sigma(r,Y_r,\mathscr{L}_{Y_r})\}\mathrm{d}W(r)\big\rangle.\end{aligned} \tag{3.12}$$

By $(H1)$, the BDG inequality for $p=1$ and since $\mathbb{W}_2(\mathscr{L}_{X_r},\mathscr{L}_{Y_r})^2\leq\mathbb{E}\|X_r-Y_r\|_\infty^2$, we have

$$\begin{aligned}&2\mathbb{E}\sup_{s\in[0,t]}\int_0^s\mathrm{e}^{\kappa r}\big\langle X(r)-Y(r),\{\sigma(r,X_r,\mathscr{L}_{X_r})-\sigma(r,Y_r,\mathscr{L}_{Y_r})\}\mathrm{d}W(r)\big\rangle\\ &\leq 4\mathbb{E}\bigg(\int_0^t\mathrm{e}^{2\kappa s}|X(s)-Y(s)|^2\big(\alpha_1(s)\|X_s-Y_s\|_\infty^2+\alpha_s(s)W_2(\mathscr{L}_{X_r},\mathscr{L}_{Y_r})^2\big)\mathrm{d}s\bigg)^{\frac{1}{2}}\\ &\leq\varepsilon\mathbb{E}\gamma_t+\frac{4}{\varepsilon}\int_0^t(\alpha_1(s)+\alpha_2(s))\mathbb{E}[\mathrm{e}^{\kappa s}\|X_s-Y_s\|_\infty^2]\mathrm{d}s\\ &\leq\varepsilon\mathbb{E}\gamma_t+\frac{4}{\varepsilon}\mathrm{e}^{\kappa r_0}\int_0^t(\alpha_1(s)+\alpha_2(s))\mathbb{E}\gamma_s\mathrm{d}s.\end{aligned}$$

Combining this with (3.12) we obtain

$$\mathbb{E}\gamma_t\leq\frac{\mathbb{E}\|X_0-Y_0\|_\infty^2}{1-\varepsilon}+\frac{\mathrm{e}^{\kappa r_0}}{1-\varepsilon}\int_0^t\Big\{\frac{4}{\varepsilon}(\alpha_1(s)+\alpha_2(s))+\beta_1(s)+\beta_2(s)\Big\}\mathbb{E}\gamma_s\mathrm{d}s,\quad t\geq s.$$

So, Gronwall's Lemma implies

$$\mathbb{E}\gamma_t\leq\frac{\mathbb{E}\|X_0-Y_0\|_\infty^2}{1-\varepsilon}\exp\bigg[\frac{\mathrm{e}^{\kappa r_0}}{1-\varepsilon}\int_0^t\Big\{\frac{4}{\varepsilon}(\alpha_1(s)+\alpha_2(s))+\beta_1(s)+\beta_2(s)\Big\}\mathrm{d}s\bigg].$$

Noting that $\mathbb{E}\gamma_t\geq\mathrm{e}^{(t-r_0)\kappa}\mathbb{E}\|X_t-Y_t\|_\infty^2$, this implies

$$\mathbb{E}\|X_t-Y_t\|_\infty^2\leq\frac{\mathbb{E}\|X_0-Y_0\|_\infty^2}{1-\varepsilon}$$

$$\times \exp\bigg[(r_0-t)\kappa + \frac{\mathrm{e}^{\kappa r_0}}{1-\varepsilon}\int_0^t \Big\{\frac{4}{\varepsilon}(\alpha_1(s)+\alpha_2(s))+\beta_1(s)+\beta_2(s)\Big\}\mathrm{d}s\bigg].$$

Since $(H2)$ remains true if κ is replaced by a smaller constant δ, this estimate also holds for $\delta\in[0,\kappa]$ replacing κ. Therefore, the estimate in Theorem 3.1(2) holds.

(3) Let $(X_t)_{t\ge 0}$ solve (3.1) with $\mathscr{L}_{X_0}=\mu_0$, and let $(\tilde X_t,\tilde W(t))$ on $(\tilde\Omega,\{\tilde{\mathscr{F}}_t\}_{t\ge0},\tilde{\mathbb{P}})$ be a weak solution of (3.1) such that $\mathscr{L}_{X_0}|_{\mathbb{P}}=\mathscr{L}_{\tilde X_0}|_{\tilde{\mathbb{P}}}=\mu_0$, i.e. $\tilde X_t$ solves

$$\mathrm{d}\tilde X(t)=b(t,\tilde X_t,\mathscr{L}_{\tilde X_t}|_{\tilde{\mathbb{P}}})\mathrm{d}t+\sigma(t,\tilde X_t,\mathscr{L}_{\tilde X_t}|_{\tilde{\mathbb{P}}})\mathrm{d}\tilde W(t),\quad \mathscr{L}_{\tilde X_0}=\mu_0. \tag{3.13}$$

We aim to prove $\mathscr{L}_X|_{\mathbb{P}}=\mathscr{L}_{\tilde X}|_{\tilde{\mathbb{P}}}$. Let $\mu_t=\mathscr{L}_{X_t}|_{\mathbb{P}}$ and

$$\bar b(t,\xi)=b(t,\xi,\mu_t),\quad \bar\sigma(t,\xi)=\sigma(t,\xi,\mu_t),\quad t\ge0,\xi\in\mathscr{C}.$$

By $(H1)$-$(H3)$, the stochastic functional differential equation

$$\mathrm{d}\bar X(t)=\bar b(t,\bar X_t)\mathrm{d}t+\bar\sigma(t,\bar X_t)\mathrm{d}\tilde W(t),\quad \bar X_0=\tilde X_0 \tag{3.14}$$

has a unique solution. According to Yamada–Watanabe, it also satisfies weak uniqueness. Noting that

$$\mathrm{d}X(t)=\bar b(t,X_t)\mathrm{d}t+\bar\sigma(t,X_t)\mathrm{d}W(t),\quad \mathscr{L}_{X_0}|_{\mathbb{P}}=\mathscr{L}_{\tilde X_0}|_{\tilde{\mathbb{P}}},$$

the weak uniqueness of (3.14) implies

$$\mathscr{L}_{\bar X}|_{\tilde{\mathbb{P}}}=\mathscr{L}_X|_{\mathbb{P}}. \tag{3.15}$$

So, (3.14) reduces to

$$\mathrm{d}\bar X(t)=b(t,\bar X_t,\mathscr{L}_{\bar X_t}|_{\tilde{\mathbb{P}}})\mathrm{d}t+\sigma(t,\bar X_t,\mathscr{L}_{\bar X_t}|_{\tilde{\mathbb{P}}})\mathrm{d}\tilde W(t),\quad \bar X_0=\tilde X_0.$$

Since the strong uniqueness of (3.13) is ensured by Step (1), we obtain $\bar X=\tilde X$. Therefore, (3.15) implies $\mathscr{L}_{\tilde X}|_{\tilde{\mathbb{P}}}=\mathscr{L}_X|_{\mathbb{P}}$ as wanted.

Finally, since $\mathscr{C}$ is a Polish space, for any $\mu_0,\nu_0\in\mathscr{P}_2^{\mathscr{C}}$, we can take $\mathscr{F}_0$-measurable random variables X_0, Y_0 such that $\mathscr{L}_{X_0}=\mu_0,\mathscr{L}_{Y_0}=\nu_0$ and $\mathbb{W}_2(\mu_0,\nu_0)^2=\mathbb{E}\|X_0-Y_0\|_\infty^2$. Combining this with $\mathbb{W}_2(P_t^*\mu_0,P_t^*\nu_0)^2\le\mathbb{E}\|X_t-Y_t\|_\infty^2$, we deduce the estimate in Theorem 3.1(3) from that in Theorem 3.1(2). $\square$

4 Harnack inequality and applications

To prove Theorem 2.2, we investigate Harnack inequalities of the operator P_t defined by

$$(P_tf)(\mu_0)=\int_{\mathscr{C}} f\mathrm{d}(P_t^*\mu_0),\quad f\in\mathscr{B}_b(\mathscr{C}),t\ge0,\mu_0\in\mathscr{P}_2^{\mathscr{C}}. \tag{4.1}$$

We will consider the Harnack inequality with a power $p>1$ introduced in [16], and the log-Harnack inequality developed in [12, 19], where classical SDEs on $\mathbb{R}^d$ and manifolds are considered. To establish these inequalities for the present path-distribution dependent SDEs, we will adopt coupling by change of measures introduced in [1, 17]. We refer to [18] for a general theory on this method and applications.

To construct the desired coupling for the segment solution X_t, we need to assume that $\sigma(t,\xi,\mu)=\sigma(t,\xi(0))$; that is, we consider the following simpler version of (3.1):

$$\mathrm{d}X(t)=b(t,X_t,\mathscr{L}_{X_t})\mathrm{d}t+\sigma(t,X(t))\mathrm{d}W(t). \tag{4.2}$$

Theorem 4.1. *Assume* **(A)**. *Then there exists $H_1\in C(\mathbb{R}_+;\mathbb{R}_+)$ such that for any $\mu_0,\nu_0\in\mathscr{P}_2^{\mathscr{C}}$, $\mathscr{F}_0$-measurable random variables X_0,Y_0 with $\mathscr{L}_{X_0}=\mu_0,\mathscr{L}_{Y_0}=\nu_0$, and $f\in\mathscr{B}_b^+(\mathscr{C})$,*

$$(P_T\log f)(\nu_0)\le\log(P_Tf)(\mu_0)+H_1(T)\mathbb{E}\bigg(\frac{|X(0)-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\bigg),\quad T>r_0. \tag{4.3}$$

If moreover (2.7) *holds for some increasing $\kappa_3:\mathbb{R}_+\to\mathbb{R}_+$, then there exists $H_2\in C(D;\mathbb{R}_+)$, where D is as in Theorem 2.2, such that*

$$(P_Tf)(\nu_0)\le(P_Tf^p)^{\frac{1}{p}}(\mu_0)\mathbb{E}\Big(\mathrm{e}^{H_2(p,T)\big(1+\frac{|X(0)-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\big)}\Big),\quad T>r_0,(p,T)\in D \tag{4.4}$$

holds for μ_0,ν_0 and X_0,Y_0 as above.

As a consequence of Theorem 4.1, we have the following result, see, for instance, the proof of [22, Prposition 3.1].

Corollary 4.2. *Assume* **(A)** *and let $T>r_0$. For any $\mu_0,\nu_0\in\mathscr{P}_2$, $P_T^*\mu_0$ and $P_T^*\nu_0$ are equivalent and the Radon-Nykodim derivative satisfies the entropy estimate*

$$\int_{\mathscr{C}}\bigg(\log\frac{\mathrm{d}P_T^*\nu_0}{\mathrm{d}P_T^*\mu_0}\bigg)\mathrm{d}P_T^*\nu_0\le\inf_{\mathscr{L}_{X_0}=\mu_0,\mathscr{L}_{Y_0}=\nu_0}\mathbb{E}\bigg[H_1(T)\bigg(\frac{|X(0)-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\bigg)\bigg],\quad T>r_0.$$

If (2.7) *holds, then for any $T>r_0$ and $p>(1+\kappa_3(T)\lambda(T))^2$,*

$$\int_{\mathscr{C}}\bigg(\frac{\mathrm{d}P_T^*\nu_0}{\mathrm{d}P_T^*\mu_0}\bigg)^{\frac{1}{p}}\mathrm{d}(P_T^*\nu_0)\le\inf_{\mathscr{L}_{X_0}=\mu_0,\mathscr{L}_{Y_0}=\nu_0}\mathbb{E}\Big(\mathrm{e}^{H_2(p,T)\big(1+\frac{|X(0)-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\big)}\Big).$$

Proof of Theorem 4.1. For $\mu_t:=P_t^*\mu_0$ and $\nu_t:=P_t^*\nu_0$, we may rewrite (4.2) as

$$\mathrm{d}X(t)=\bar{b}(t,X_t)\mathrm{d}t+\sigma(t,X(t))\mathrm{d}\bar{W}(t),\quad \mathscr{L}_{X_0}=\mu_0, \tag{4.5}$$

where

$$\begin{aligned}&\bar{b}(t,\xi):=b(t,\xi,\nu_t),\quad \mathrm{d}\bar{W}(t):=\mathrm{d}W(t)+\bar{\gamma}(t)\mathrm{d}t,\\&\bar{\gamma}(t):=\sigma^{-1}(t,X(t))[b(t,X_t,\mu_t)-b(t,X_t,\nu_t)].\end{aligned}$$

By assumption **(A)** and Theorem 3.1(3), we have

$$|\bar{\gamma}(t)|\le\lambda(t)\kappa_2(t)\mathbb{W}_2(\mu_t,\nu_t)\le K(t)\mathbb{W}_2(\mu_0,\nu_0),\quad t\in[0,T] \tag{4.6}$$

for some increasing function $K : \mathbb{R}_+ \to \mathbb{R}_+$. Let

$$\bar{R}_t = \exp\left\{ -\int_0^t \langle \bar{\gamma}(s), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t |\bar{\gamma}(s)|^2 \mathrm{d}s \right\}, \quad t \in [0,T]. \tag{4.7}$$

By Girsanov's theorem, $\{\bar{W}(t)\}_{t\in[0,T]}$ is a d-dimensional Brownian motion under the probability measure $\bar{\mathbb{P}}_T := \bar{R}_T\mathbb{P}$.

Next, according to the proof of [18, Theorem 4.3.1] or [22, Theorem 1.1], we can construct an adapted process $\tilde{\gamma}(t)$ on $\mathbb{R}^d$ such that

(a) Under the probability measure $\bar{\mathbb{P}}_T$,

$$\tilde{R}_t := \exp\left\{ -\int_0^t \langle \tilde{\gamma}(s), \mathrm{d}\bar{W}(s)\rangle - \frac{1}{2}\int_0^t |\tilde{\gamma}(s)|^2 \mathrm{d}s \right\}, \quad t \in [0,T]$$

is a martingale, such that $\tilde{\mathbb{P}}_T := \tilde{R}_T\bar{\mathbb{P}}_T = \tilde{R}_T\bar{R}_T\mathbb{P}$ is a probability measure under which

$$\tilde{W}(t) := \bar{W}(t) + \int_0^t \tilde{\gamma}(s)\mathrm{d}s = W(t) + \int_0^t \big(\bar{\gamma}(s) + \tilde{\gamma}(s)\big)\mathrm{d}s, \quad t \in [0,T]$$

is a d-dimensional Brownian motion.

(b) Letting $Y(t)$ solve the following stochastic functional differential equation under the probability measure $\tilde{\mathbb{P}}_T$ with the given initial value Y_0:

$$\mathrm{d}Y(t) = \bar{b}(t, Y_t)\mathrm{d}t + \sigma(t, Y(t))\mathrm{d}\tilde{W}(t), \tag{4.8}$$

we have $\mathscr{L}_{Y_0|\tilde{\mathbb{P}}} = \mathscr{L}_{Y_0} = \nu_0$ and $X_T = Y_T$ $\tilde{\mathbb{P}}_T$-a.s.

(c) There exists $C \in C(\mathbb{R}_+;\mathbb{R}_+)$ such that

$$\mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T |\tilde{\gamma}(s)|^2\mathrm{d}s \le C(T)\mathbb{E}\Big(\frac{|X(0)-Y(0)|^2}{T-r_0} + \|X_0 - Y_0\|_\infty^2\Big).$$

By the definition of $\bar{b}$ we see that $(Y_t, \tilde{W}(t))$ is a weak solution to the equation (4.5) with initial distribution ν_0, so that by the weak uniqueness, $\mathscr{L}_{Y_t}|_{\tilde{\mathbb{P}}_T} = \nu_t, t \in [0,T]$. Combining this with (b) we obtain

$$(P_T f)(\nu_0) = \mathbb{E}_{\tilde{\mathbb{P}}_T}[f(Y_T)] = \mathbb{E}_{\tilde{P}_T}[f(X_T)] = \mathbb{E}[\bar{R}_T\tilde{R}_T f(X_T)], \quad f \in \mathscr{B}_b^+(\mathscr{C}).$$

Letting $R_T = \bar{R}_T\tilde{R}_T$, by Young's inequality and Hölder's inequality respectively, we obtain

$$(P_T \log f)(\nu_0) \le \mathbb{E}[R_T \log R_T] + \log \mathbb{E}[f(X_T)] = \mathbb{E}[R_T \log R_T] + \log(P_T f)(\mu_0), \tag{4.9}$$

and

$$(P_T f(\nu_0))^p \le (\mathbb{E}R_T^{\frac{p}{p-1}})^{p-1}(\mathbb{E}f^p(X_T)) = (\mathbb{E}R_T^{\frac{p}{p-1}})^{p-1}P_T f^p(\mu_0), \quad p > 1. \tag{4.10}$$

We are now ready to prove assertions (1) and (2) as follows.

By (4.6) , (c) and since $\mathbb{W}_2(\mu_0,\nu_0)^2 \le \mathbb{E}\|X_0-Y_0\|_\infty^2$,

$$\begin{aligned}
\mathbb{E}[R_T\log R_T] &\le \frac{1}{2}\mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T |\bar\gamma(s)+\tilde\gamma(s)|^2\mathrm{d}s\\
&\le \mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T |\tilde\gamma(s)|^2\mathrm{d}s+\int_0^T |\bar\gamma(s)|^2\mathrm{d}s\\
&\le \mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T |\tilde\gamma(s)|^2\mathrm{d}s+\int_0^T \lambda(t)^2\kappa_2(t)^2\mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t\\
&\le H_1(T)\mathbb{E}\bigg(\frac{|X(0)-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\bigg),\quad T>r_0
\end{aligned}\tag{4.11}$$

holds for some $H_1\in C(\mathbb{R}_+;\mathbb{R}_+)$. Combining this with (4.9) we obtain (4.3).

Finally, according to the proof of [22, Theorem 4.1], there exists $C\in C(D;\mathbb{R}_+)$ such that

$$(\mathbb{E}_{\bar{\mathbb{P}}_T}\tilde R_T^{\frac{p}{p-1}})^{\frac{p-1}{p}} \le \mathbb{E}\Big(\mathrm{e}^{C(p,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\big)}\Big),\quad T>r_0,(p,T)\in D.$$

For any $p>p(T):=(1+\kappa_3(T)\lambda(T))^2$, by applying this estimate for $p_1:=\frac{1}{2}(p+(p(T))$ and combining with $R_T=\tilde R_T\bar R_T$, (4.6), (4.7) and $\mathbb{W}_2(\mu_0,\nu_0)^2\le\mathbb{E}\|X_0-Y_0\|_\infty^2$, we arrive at

$$\begin{aligned}
\Big(\mathbb{E}R_T^{\frac{p}{p-1}}\Big)^{\frac{p-1}{p}} &= \Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\tilde R_T^{\frac{p}{p-1}}\bar R_T^{\frac{1}{p-1}}\Big)^{\frac{p-1}{p}} \le \Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\tilde R_T^{\frac{p_1}{p_1-1}}\Big)^{\frac{p_1-1}{p_1}}\Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\bar R_T^{\frac{p_1}{p-p_1}}\Big)^{\frac{p-p_1}{pp_1}}\\
&\le \mathbb{E}\Big(\mathrm{e}^{C(p_1,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\big)}\Big)\Big(\mathbb{E}\bar R_T^{\frac{p}{p-p_1}}\Big)^{\frac{p-p_1}{pp_1}}\\
&\le \mathbb{E}\Big(\mathrm{e}^{H_2(p,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r_0}+\|X_0-Y_0\|_\infty^2\big)}\Big),\quad T>r_0,(p,T)\in D
\end{aligned}$$

for some $H_2\in C(\mathbb{R}_+;\mathbb{R}_+)$. Therefore, (4.4) follows from (4.10). $\square$

5 Shift Harnack inequality and integration by parts formula

To prove Theorem 2.3, we investigate the shift Harnack inequality and integration by parts formula introduced in [21]. Assume that $\sigma(t,\xi,\mu)=\sigma(t)$ is invertible. Then the path-distribution dependent SDE (3.1) becomes

$$\mathrm{d}X(t)=b(t,X_t,\mathscr{L}_{X_t})\mathrm{d}t+\sigma(t)\mathrm{d}W(t),\quad \mathscr{L}_{X_0}=\mu_0.$$

To apply the existing shift Harnack inequality and integration by parts formula, we let

$$\bar b(t,\xi):=b(t,\xi,\mu_t),\quad \mu_t:=\mathscr{L}_{X_t}=P_t^*\mu_0,\quad t\ge0,\xi\in\mathscr{C}$$

and rewrite this equation as

$$\mathrm{d}X(t)=\bar b(t,X_t)\mathrm{d}t+\sigma(t)\mathrm{d}W(t),\quad \mathscr{L}_{X_0}=\mu_0.$$

Then the following result follows from [18, Theorem 4.2.3].

Theorem 5.1. *Let $\sigma : [0,\infty) \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $b : [0,\infty) \times \mathscr{C} \times \mathscr{P}_2^{\mathscr{C}} \to \mathbb{R}^d$ satisfy* **(A)**, *and assume that for any $(t,\mu) \in \mathbb{R}_+ \times \mathscr{P}_2^{\mathscr{C}}$, $b(t,\cdot,\mu)$ is differentiable. Then*

$$\Lambda(T) := \sup_{t\in[0,T]} \|\sigma(t)^{-1}\|^2 < \infty, \quad K(T) := \sup_{t\in[0,T],\mu\in\mathscr{P}_2^{\mathscr{C}}} \|\nabla b(t,\cdot,\mu)\|_\infty^2 < \infty, \quad T \geq 0.$$

Moreover:

(1) *For any $p > 1, T > r_0, \mu_0 \in \mathscr{P}_2^{\mathscr{C}}, \eta \in \mathbb{H}^1$ and $f \in \mathscr{B}_b^+(\mathscr{C})$,*

$$\begin{aligned}(P_T f)^p(\mu_0) \leq & (P_T f^p(\eta + \cdot))(\mu_0) \\ & \times \exp\left[\frac{p\,\Lambda(T)\,(1+T^2K(T))\left(\frac{|\eta(-r_0)|^2}{T-r_0} + \|\eta\|_{\mathbb{H}^1}^2\right)}{(p-1)^2}\right], \quad p > 1,\end{aligned}$$

and

$$(P_T \log f)(\mu_0) \leq \log(P_T f(\eta+\cdot))(\mu_0) + \Lambda(T)\left(1+T^2K(T)\right)\left(\frac{|\eta(-r_0)|^2}{T-r_0} + \|\eta\|_{\mathbb{H}^1}^2\right).$$

(2) *For any $T > r_0$, let*

$$\begin{aligned}\Phi(t) &= 1_{[0,T-r_0]}(t)\frac{\eta(-r_0)}{T-r_0} + 1_{(T-r_0,T]}(t)\eta'(t-T), \\ \Theta(t) &= \int_0^{t^+} \Phi(s)\mathrm{d}s, \quad t \in [-r_0, T].\end{aligned}$$

Then for any $f \in C^1(\mathscr{C})$, $\eta \in \mathbb{H}^1$ and $\mu_0 \in \mathscr{P}_2^{\mathscr{C}}$,

$$\mathbb{E}(\nabla_\eta f)(X_T) = \mathbb{E}\left[f(X_T)\int_0^T \left\langle \sigma(t)^{-1}(\Phi(t) - \nabla_{\Theta_t} b(t,\cdot,P_t^*\mu_0)(X_t)),\ \mathrm{d}W(t)\right\rangle\right].$$

As consequence of Theorem 5.1 we have the following result.

Corollary 5.2. *In the situation of Theorem 5.1. For any $\mu_0 \in \mathscr{P}_2^{\mathscr{C}}, \eta \in \mathbb{H}^1$ and $T > r_0$, $\mu_T := P_T^*\mu_0$ satisfies*

$$\begin{aligned}&\int_{\mathscr{C}} \Big(\log\frac{\mathrm{d}\mu_T(\cdot+\eta)}{\mathrm{d}\mu_T}\Big)\mathrm{d}\mu_T(\cdot+\eta) \leq \Lambda(T)(1+T^2K(T))\Big(\frac{|\eta(-r_0)|^2}{T-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big), \\ &\int_{\mathscr{C}} \Big(\frac{\mathrm{d}\mu_T(\cdot+\eta)}{\mathrm{d}\mu_T}\Big)^{\frac{1}{p}}\mathrm{d}\mu_T(\cdot+\eta) \leq \exp\left[\frac{\Lambda(T)(1+T^2K(T))}{(p-1)^2}\Big(\frac{|\eta(-r_0)|^2}{T-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big)\right], \quad p > 1, \\ &\int_{\mathscr{C}} \Big|\frac{\mathrm{d}\partial_\eta\mu_T}{\mathrm{d}\mu_T}\Big|^2\mathrm{d}\mu_T \leq \Lambda(T)\big(1+K(T)T^2\big)\Big(\frac{|\eta(-r_0)|^2}{T-r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big).\end{aligned}$$

Proof. The first two estimates follow from Theorem 5.1(1), see [21] or [18, §1.4]. As the last estimate is not explicitly given in these references, we present a brief proof below. It is easy to see that

$$M(T) := \int_0^T \left\langle \sigma(t)^{-1}(\Phi(t) - \nabla_{\Theta_t} b(t,\cdot,P_t^*\mu_0)(X_t)),\ \mathrm{d}W(t)\right\rangle$$

satisfies

$$\mathbb{E}M(T)^2 \le C(T) := \Lambda(T)\big(1 + K(T)T^2\big)\Big(\frac{|\eta(-r_0)|^2}{T - r_0} + \|\eta\|_{\mathbb{H}^1}^2\Big).$$

Then, Theorem 5.1(2) implies that

$$C^1(\mathscr{C}) \ni f \mapsto (\partial_\eta \mu_T)(f) := \left(\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_{\mathscr{C}} f \mathrm{d}\mu_T(\cdot + \varepsilon\eta)\right)\bigg|_{\varepsilon=0}$$

is a densely defined bounded linear functional on $L^2(\mu_T)$ with

$$\big|(\partial_\eta \mu_T)(f)\big|^2 \le \mu_T(f^2)\mathbb{E}M(T)^2 \le C(T)\mu_T(f^2).$$

By the Riesz Representation Theorem, it uniquely extends to a bounded linear functional

$$(\partial_\eta \mu_T)(f) := \int_{\mathscr{C}} f g \mathrm{d}\mu_T, \quad f \in L^2(\mu_T)$$

for some $g \in L^2(\mu_T)$ with $\mu_T(g^2) \le C(T)$. Consequently, μ_T is differentiable along η with $(\partial_\eta \mu_T)(A) = \int_A g \mathrm{d}\mu_T, A \in \mathscr{B}(\mathscr{C})$, and $\partial_\eta \mu_T$ is absolutely continuous with respect to μ_T such that

$$\int_{\mathscr{C}} \Big(\frac{\mathrm{d}\partial_\eta \mu_T}{\mathrm{d}\mu_T}\Big)^2 \mathrm{d}\mu_T = \int_{\mathscr{C}} g^2 \mathrm{d}\mu_T \le C(T).$$

□

References

[1] M. Arnaudon, A. Thalmaier, F.-Y. Wang, *Harnack inequality and heat kernel estimate on manifolds with curvature unbounded below,* Bull. Sci. Math. 130(2006), 223–233.

[2] V. Bogachev, A. Krylov, M. Röckner, S. Shaposhnikov, *Fokker-Planck-Kolmogorov equations,* Monograph, AMS(2015).

[3] I. Csiszár, J. Körne, *Information Theory: Coding Theorems for Discrete Memory- less Systems,* Academic Press, New York, 1981.

[4] K. Carrapatoso, *Exponential convergence to equilibrium for the homogeneous Landau equation with hard potentials,* Bull. Sci. Math. 139(2015), 777–805.

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

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

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

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

[9] I. Ikeda, S. Watanabe, *Stochastic differential equations and Diffusion processes,* North-Holland, Amsterdam, 1981.

[10] A. N. Kolmogoroff, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung,* Math. Ann. 104(1931), 415–458.

[11] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes,* Holden-Day, San Francisco, 1964.

[12] M. Röckner, F.-Y. Wang, *Log-harnack inequality for stochastic differential equations in Hilbert spaces and its consequences*, Infin. Dim. Anal. Quat. Probab. Relat. Top. 13(2010), 27–37.

[13] D.W. Stroock, S.R.S. Varadhan, *Multidimensional Diffusion Processes,* Springer, New York, 1979.

[14] C. Villani, *Optimal Transport, Old and New,* Springer, 2009.

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

[16] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds,* Probab. Theory Related Fields, 109(1997), 417–424.

[17] F.-Y. Wang, *Harnack inequality and applications for stochastic generalized porous media equations,* Ann. Probab. 35(2007), 1333–1350.

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

[19] F.-Y.Wang, *Harnack inequalities on manifolds with boundary and applications,* J. Math. Pures Appl. 94(2010), 304–321.

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

[21] F.-Y.Wang, *Integration by parts formula and shift Harnack inequality for stochastic equations,* Ann. Probab. 42(2014), 994–1019.

[22] F.-Y. Wang, C. Yuan, *Harnack inequalities for functional SDEs with multiplicative noise and applications,* Stoch. Proc. Appl. 121(2011), 2692–2710.

[23] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations,* arXiv:1606.05843. To appear in Stoch. Proc. Appl.

[24] T. Yamada, S. Watanabe *On the uniqueness of solutions of stochastic differential equations I & II,* J. Math. Kyoto Univ. 11(1971), 155–167; 553–563.