Bismut Formula and Gradient Estimates for Dirichlet Semigroups with Application to Singular Killed DDSDEs

Feng-Yu Wang, Xiao-Yu Zhao

Center for Applied Mathematics and KL-AAGDM, Tianjin University, 300072, China

wangfy@tju.edu.cn, zhxy_0628@tju.edu.cn

March 27, 2026

Abstract

By establishing a local version of Bismut formula for Dirichlet semigroups on a regular domain, gradient estimates are derived for killed SDEs with singular drifts. As an application, the total variation distance between two solutions of killed DDSDEs is bounded above by the truncated 1-Wasserstein distance of initial distributions, in the regular and singular cases respectively.

MSC Classification: 60B05,60B10.

Keywords: Bismut formula, gradient estimate, Dirichlet semigroup, killed DDSDE.

1 Introduction

The Bismut formula, first established in Bismut [3] for the heat semigroup on Riemannian manifolds by using Malliavin calculus, has been intensively studied and applied for SDEs and SPDEs. In order to derive gradient estimates on diffusion semigroups by using local geometry quantities, a local version of Bismut formula is established in [12] by refining the martingale argument developed in [4], which is then applied in [13] to derive explicit gradient estimates on Dirichlet semigroups.

In recent years, the Bismut formula has been established in [16] for singular SDEs where the drifts only satisfy a local integrability condition. When the drifts also depend on the distribution of the solution, the SDEs are known as distribution dependent SDEs (DDSDEs for short). In this case, the Bismut formula has been established in [2, 7, 10, 15] for the intrinsic/Lions derivative with respective to the initial distribution, as well as in [9] for the extrinsic derivative.

To characterize nonlinear Dirichlet problem, the well-posedness and Lipschitz continuity in initial distribution have been derived in [14] for a class of singular DDSDEs in a domain with killed boundary. As far as we know, there is no any results on Bismut formula and applications for killed distribution dependent SDEs. In this paper, we intend to establish the Bismut formula for Dirichlet semigroups for killed SDEs, and apply to derive gradient estimates on DDSDEs.

Before moving on, let us recall the gradient estimate derived in [13] for the Dirichlet semigroup on manifold, and formulate the equivalent estimate on distributions of the associated killed diffusions. Consider the Dirichlet semigroup P_t^D generated by $\mathscr{L} := \Delta + Z$ on $\bar{D}$, where D is a regular open domain in a Riemannian manifold, Δ is the Laplacian and Z is a C^1 vector field on $\bar{D}$. We have

$$P_t^D f(x) = \mathbb{E}[\mathbf{1}_{\{t<\tau(X^x)\}} f(X_t^x)], \quad t \geq 0, \ x \in \bar{D}, \ f \in \mathscr{B}_b(\bar{D}),$$

where $\mathscr{B}_b(\bar{D})$ is the space of all bounded measurable functions on $\bar{D}$, X_t^x is the diffusion process generated by $\mathscr{L}$ with $X_0^x = x$, and

$$\tau(X^x) := \inf\{t \geq 0 : X_t^x \in \partial D\}$$

is the first hitting time of X_t^x to the boundary ∂D. Here and in the following, we set $\inf \emptyset = \infty$. By using the local version of Bismut formula established in [12], the following type gradient estimate is derived in [13]:

$$|\nabla P_t^D f(x)| \leq \frac{c\|f\|_\infty}{\sqrt{t}\,(\rho_\partial(x) \wedge 1)}, \quad t \in (0,T], \ x \in D, \ f \in \mathscr{B}_b(\bar{D}), \tag{1.1}$$

where ρ_∂ is the Riemannian distance function to the boundary ∂D, and $c > 0$ is a constant explicitly determined by local geometry quantities of D, including the dimension, curvature bounds of the generator in D, and the second fundamental form on ∂D. Below we formulate an equivalent inequality (1.3) to (1.1), where the point $x \in D$ is replaced by sub-probability measures on D, which is essential for killed DDSDEs.

For any probability measure μ on $\bar{D}$, its restriction $\mu|_D$ to D is a sub probability measure on D, i.e.

$$\mu|_D \in \mathscr{P}_D := \big\{\nu : \ \nu \text{ is a measure on } D, \ \nu(D) \leq 1\big\}.$$

Let $P_t^{D*}\mu \in \mathscr{P}_D$ be defined as

$$(P_t^{D*}\mu)(A) := \mu(P_t^D \mathbf{1}_A) = \int_D P_t^D \mathbf{1}_A \mathrm{d}\mu = \int_D P_t^D \mathbf{1}_A \mathrm{d}\mu|_D, \quad A \in \mathscr{B}_D, \ t \in (0,T],$$

where $\mathscr{B}_D$ is the Borel σ-field on D. Since $P_t^{D*}\mu$ only depends on $\mu|_D$, P_t^* is well-defined on $\mathscr{P}_D$, i.e. for any $\mu \in \mathscr{P}_D$, $P_t^{D*}\mu$ is identified as $P_t^{D*}\bar{\mu}$ for a probability measure $\bar{\mu}$ on $\bar{D}$ with $\bar{\mu}|_D = \mu$. So,

$$(P_t^{D*}\mu)(f) := \int_D f \mathrm{d}(P_t^{D*}\mu) = \int_D P_t^D f \mathrm{d}\mu, \quad t \in (0,T], \ \mu \in \mathscr{P}_D, \ f \in \mathscr{B}_b(\bar{D}). \tag{1.2}$$

Let

$$\rho(x,y) := \big\{|f(x) - f(y)| : \ f \in C^1(D), \ |(\rho_\partial \wedge 1)\nabla f| \leq 1\big\},$$

which is the intrinsic distance in D induced by the Riemannian metric with weight $(\rho_\partial \wedge 1)^{-2}$, see [13] for details. The associated truncated 1-Wasserstein distance is defined as

$$\hat{\mathbb{W}}_1^\rho(\mu,\nu) := \inf_{\pi \in \mathscr{C}_D(\mu,\nu)} \int_{\bar{D}\times\bar{D}} [\rho(x,y) \wedge 1]\pi(\mathrm{d}x, \mathrm{d}y),$$

where $\pi \in \mathscr{C}_D(\mu,\nu)$ means that π is a probability measure on $\bar{D}\times\bar{D}$ such that

$$\pi(A\times\bar{D})=\mu(A),\ \ \pi(\bar{D}\times A)=\nu(A),\quad A\in\mathscr{B}_D.$$

Then the gradient estimate (1.1) is equivalent to

$$|P_t^D f(x)-P_t^D f(y)|\le \frac{c\rho(x,y)}{\sqrt{t}}\|f\|_\infty,\quad t\in(0,T],\ f\in\mathscr{B}_b(\bar{D}),\ x,y\in\bar{D}.$$

This together with $|P_tf(x)-P_tf(y)|\le 2\|f\|_\infty$ implies that for some constant $c'>0$,

$$|P_t^D f(x)-P_t^D f(y)|\le \frac{c'[1\wedge\rho(x,y)]}{\sqrt{t}}\|f\|_\infty,\quad t\in(0,T],\ f\in\mathscr{B}_b(\bar{D}),\ x,y\in\bar{D}.$$

Hence, (1.1) implies

$$\|P_t^{D*}\mu-P_t^{D*}\nu\|_{var}\le\frac{c'}{\sqrt{t}}\hat{\mathbb{W}}_1^\rho(\mu,\nu),\quad t>0,\quad \mu,\nu\in\mathscr{P}_D. \tag{1.3}$$

On the other hand, it is trivial that (1.3) implies (1.1) for $c=c'$. So, we may regard (1.3) as gradient estimate on P_t^D. This is essential for killed DDSDE, since in this case the distributions for the solutions with deterministic initial values are not enough to determine those with random initial values.

From now on, let D be a connected, possibly unbounded, open domain in $\mathbb{R}^d$ with boundary ∂D, let $\bar{D}:=D\cup\partial D$. For any random variable ξ on $\bar{D}$, its restricted distribution to D is defined as

$$\mathscr{L}_\xi^D:=\mathbb{P}(\xi\in\cdot)\in\mathscr{P}_D.$$

When different probability spaces are considered, we denote $\mathscr{L}_\xi^D$ by $\mathscr{L}_{\xi|\mathbb{P}}^D$ to emphasize the underline probability $\mathbb{P}$.

According to [5], for the solution of a regular killed SDE on $\bar{D}$, the associated Dirichlet semigroup P_t^D on $\mathscr{B}_b(\bar{D})$ satisfies (see Lemma 2.3 below)

$$\|\nabla P_t^D f\|_\infty\le\frac{c}{\sqrt{t}}\|f\|_\infty,\ \ t\in(0,T],\ f\in\mathscr{B}_b(\bar{D}). \tag{1.4}$$

By letting $P_t^D f|_{D^c}=0$ for $t>0$, this is equivalent to

$$|P_t^D f(x)-P_t^D f(y)|\le\frac{c\|f\|_\infty}{\sqrt{t}}|x-y|,\quad t\in(0,T],\ f\in\mathscr{B}_b(\bar{D}),\ x,y\in\bar{D},$$

and hence implies that for some possibly larger constant $c'>0$

$$\|P_t^{D*}\mu-P_t^{D*}\nu\|_{var}\le\frac{c'}{\sqrt{t}}\hat{\mathbb{W}}_1(\mu,\nu),\ \ t\in(0,T],\ \mu,\nu\in\mathscr{P}_D, \tag{1.5}$$

where $\hat{\mathbb{W}}_1$ is defined as $\hat{\mathbb{W}}_1^\rho$ with $\rho(x,y)$ replaced by $|x-y|$:

$$\hat{\mathbb{W}}_1(\mu,\nu):=\inf_{\pi\in\mathscr{C}_D(\mu,\nu)}\int_{\bar{D}\times\bar{D}}[|x-y|\wedge 1]\pi(\mathrm{d}x,\mathrm{d}y).$$

On the other hand, since $\rho_\partial \wedge 1 \le 1$, it is clear that $\rho(x,y) \ge |x-y|$. So, $\hat{\mathbb{W}}_1 \le \hat{\mathbb{W}}_1^\rho$, hence (1.5) is stronger than (1.3).

In this paper, we intend to establish the estimates (1.3) and (1.5) for the following type killed DDSDE on $\bar{D}$, in the regular and singular cases respectively:

$$\mathrm{d}X_t = \mathbf{1}_{\{t<\tau(X)\}} \left\{ b_t(X_t, \mathscr{L}^D_{X_t})\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t \right\}, \quad t \in [0,T], \tag{1.6}$$

where $T > 0$ is a fixed time, W_t is the m-dimensional Brownian motion on a probability base (complete filtration probability space) $(\Omega, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$,

$$\tau(X) := \inf\{t \in [0,T] : X_t \in \partial D\},$$

and

$$b : [0,T] \times D \times \mathscr{P}_D \to \mathbb{R}^d, \quad \sigma : [0,T] \times D \to \mathbb{R}^d \otimes \mathbb{R}^m$$

are measurable.

Definition 1.1. (1) A continuous adapted process $(X_t)_{t\in[0,T]}$ on $\bar{D}$ is called a solution of (1.6), if $\mathbb{P}$-a.s.

$$\int_0^{T\wedge\tau(X)} \left\{ |b_t(X_t, \mathscr{L}^D_{X_t})| + \|\sigma_t(X_t)\|^2 \right\}\mathrm{d}t < \infty,$$

and

$$X_t = X_0 + \int_0^{t\wedge\tau(X)} \left\{ b_s(X_s, \mathscr{L}^D_{X_s})\mathrm{d}s + \sigma_s(X_s)\mathrm{d}W_s \right\}, \quad t \in [0,T].$$

We call (1.6) strongly well-posed, if for any $\mathscr{F}_0$-measurable random variable X_0 on $\bar{D}$, the SDE has a unique solution.

(2) A couple $(X_t, W_t)_{t\in[0,T]}$ is called a weak solution of (1.6), if there exists a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$ such that W_t is the m-dimensional Brownian motion and X_t solves (1.6). We call (1.6) weakly well-posed, if for any $\mu \in \mathscr{P}_D$, there exists a weak solution $(X_t, W_t)_{t\in[0,T]}$ of (1.6) with respect to a probability base $(\Omega, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$ such that $\mathscr{L}^D_{X_0|\mathbb{P}} = \mu$, and if there is another weak solution $(\tilde{X}_t, \tilde{W}_t)_{t\in[0,T]}$ with respect to probability base $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}})$ such that $\mathscr{L}^D_{X_0|\tilde{\mathbb{P}}} = \mu$, then $\mathscr{L}^D_{X_t|\mathbb{P}} = \mathscr{L}^D_{\tilde{X}_t|\tilde{\mathbb{P}}}$ for $t \in [0,T]$.

(3) We call (1.6) well-posed, if it is both strongly and weakly well-posed.

When (1.6) is well-posed, we let $P_t^{D*}\mu = \mathscr{L}^D_{X_t}$ for the solution with $\mathscr{L}^D_{X_0} = \mu$, and study the the regularity of the maps

$$\mathscr{P}_D \ni \mu \mapsto P_t^{D*}\mu \in \mathscr{P}_D, \quad t \in (0,T]$$

by establishing the estimate (1.3) and (1.5) for P_t^{D*} in place of P_t^*.

To measure the singularity, we recall locally integrable functional spaces introduced in [16]. For any $t > s \ge 0$ and $p, q \in (1,\infty)$, we write $f \in \tilde{L}_p^q([s,t])$ if $f : [s,t] \times \mathbb{R}^d \to \mathbb{R}$ is measurable with

$$\|f\|_{\tilde{L}_p^q([s,t])} := \sup_{z\in\mathbb{R}^d} \left\{ \int_s^t \left(\int_{B(z,1)} |f(u,x)|^p \mathrm{d}x \right)^{\frac{q}{p}} \mathrm{d}u \right\}^{\frac{1}{q}} < \infty,$$

where $B(z,1) := \{x \in \mathbb{R}^d : |x-z| \le 1\}$ is the unit ball centered at point z. When $s=0$, we simply denote

$$\tilde{L}_p^q(t) = \tilde{L}_p^q([0,t]), \quad \|f\|_{\tilde{L}_p^q(t)} = \|f\|_{\tilde{L}_p^q([0,t])}.$$

We will take (p,q) from the space

$$\mathscr{K} := \Big\{(p,q) \in (2,\infty)\times(2,\infty) : \ \frac{d}{p} + \frac{2}{q} < 1\Big\}.$$

In Section 2, we study the killed SDE where $b_t(x,\mu) = b_t(x)$ does not depend on the distribution variable μ, by first establishing Bismut formula in the regular case, then deriving gradient estimate in the singular case. As applications, in Section 3, the estimate (1.3) is presented for singular killed DDSDEs, and the stronger estimate (1.5) for regular killed DDSDEs. It is worthy indicating that the Bismut formula derived in Section 2 is new, since the existing formula in [12, 13] only works for time homogeneous models with diffusion coefficient at least C^2-smooth, so that the required curvature condition is well-defined.

2 Bismut formula and gradient estimate

In this section, we let $b_t(x,\mu) = b_t(x)$ be independent of μ, and consider the classical SDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b_t(X_t)\,\mathrm{d}t + \sigma_t(X_t)\,\mathrm{d}W_t, \quad t \in [0,T]. \tag{2.1}$$

Under the following assumption, the well-posedness and Bismut formula has been derived in [6] for this SDE, see [15, 16] for earlier results where $k_1 = 1$.

(A) *Let $a := \sigma\sigma^*$, $b = b_t^{(0)} + \sum_{i=1}^{k_1} b_t^{(i)}$ for some $k_1 \in \mathbb{N}$ and measurable maps*

$$b^{(i)} : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d, \quad 0 \le i \le k_1,$$

such that the following conditions hold.

(1) $\|a_t(x)\| + \|a_t(x)^{-1}\| + \|\nabla b_t^{(0)}(x)\| + |b_t^{(0)}(0)|$ *is uniformly bounded in* $(t,x) \in [0,T]\times\mathbb{R}^d$, *and*

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

(2) *There exist* $k_2 \in \mathbb{N}$, $\{(p_i,q_i),(p_j',q_j')\}_{1\le i\le k_1,\ 1\le j\le k_2} \subset \mathscr{K}$, *and* $\big\{0 \le f_j \in \tilde{L}_{q_j'}^{p_j'}(T)\big\}_{1\le j\le k_2}$ *such that*

$$\sup_{1\le i\le k_1} \|b^{(i)}\|_{\tilde{L}_{q_i}^{p_i}(T)} < \infty, \quad \|\nabla a\| \le \sum_{j=1}^{k_2} f_j.$$

According to [6, Proposition 5.2], which extends [11, Theorem 1.1.3] with $k_1 = 1$, **(A)** implies that the SDE (2.1) is well-posed, i.e. for any $x \in \mathbb{R}^d$, it has a unique solution X_t^x with $X_0^x = x$. Moreover, for any $x, v \in \mathbb{R}^d$, the derivative process

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

exists in $L^k(\Omega \to C([0,T];\mathbb{R}^d),\mathbb{P})$ for any $k \in [1,\infty)$, and there exists a constant $c(k) > 0$ such that

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\Big[\sup_{t\in[0,T]} |\nabla_v X_t^x|^k\Big] \le c(k)|v|^k, \quad x, v \in \mathbb{R}^d. \tag{2.2}$$

Now, let

$$\tau(x) = \tau(X^x) := \inf\big\{t \in [0,T]:\ X_t^x \in \partial D\big\}, \quad x \in \bar{D}.$$

Then the associated Dirichlet semigroup is defined as

$$P_t^D f(x) := \mathbb{E}\big[\mathbf{1}_{\{t<\tau(x)\}} f(X_t^x)\big], \quad t \in [0,T],\ x \in \bar{D},\ f \in \mathscr{B}_b(\bar{D}).$$

In general, for any $(s,x) \in [0,T) \times \mathbb{R}^d$, let $(X_{s,t}^x)_{t\in[s,T]}$ be the unique solution to (2.1) from time s with $X_{s,s}^x = x$, and define

$$\begin{aligned} &\tau(s,x) := \inf\big\{t \in [s,T]:\ X_{s,t}^x \in \partial D\big\},\\ &P_{s,t}^D f(x) := \mathbb{E}\big[\mathbf{1}_{\{t<\tau(s,x)\}} f(X_{s,t}^x)\big], \quad t \in [s,T],\ x \in \bar{D},\ f \in \mathscr{B}_b(\bar{D}). \end{aligned} \tag{2.3}$$

Then $P_t^D = P_{0,t}^D$ for $t \in [0,T]$, and the Markov property of the solution implies the flow property

$$P_{s,t}^D = P_{s,r}^D P_{r,t}^D, \quad 0 \le s \le r \le t \le T. \tag{2.4}$$

In the following, we first establish a local version Bismut formula for P_t^D in the regular case, then derive gradient estimate on P_t^D in the singular case.

2.1 Bismut formula in the regular case

We first recall some notations. Let M be a Riemannian manifold possibly with a boundary ∂M. For any $\alpha \in (0,1)$ and $k \in \mathbb{Z}_+$, let $C_{loc}^{k+\alpha}(M)$ be the set of all real functions g on M such that $\nabla^l g$ exists for $0 \le l \le k$, where $\nabla^0 g := g$, and $\nabla^k g$ is locally α-Hölder continuous. We write $g \in C_b^{k+\alpha}(M)$, if $\nabla^l g$ is bounded for any $0 \le l \le k$, and $\nabla^k g$ is globally α-Hölder continuous on M.

Let $I \subset [0,T]$ and $M \subset \mathbb{R}^d$. For any $\alpha_1, \alpha_2 \in (0,1]$, we write $f \in C_b^{\alpha_1,\alpha_2}(I \times M)$, if $f \in C_b(I \times M)$ such that for a constant $c > 0$

$$|f_t(x) - f_s(y)| \le c\big(|t-s|^{\alpha_1} + |x-y|^{\alpha_2}\big), \quad t, s \in I,\ x, y \in M.$$

We write $f \in C_{loc}^{\alpha_1,\alpha_2}(I \times M)$ if $f \in C_b^{\alpha_1,\alpha_2}(I' \times M')$ holds for any compact sets $I' \times M' \subset I \times M$. Moreover, for any $k_1, k_2 \in \mathbb{Z}_+$, we write $f \in C_{loc}^{k_1+\alpha_1,k_2+\alpha_2}(I \times M)$ if $\partial_t^{l_1} f, \nabla^{l_2} f \in C_{loc}^{\alpha_1,\alpha_2}(I \times M)$ holds for any integers $0 \le l_1 \le k_1$ and $0 \le l_2 \le k_2$, and $f \in C_b^{k_1+\alpha_1,k_2+\alpha_2}(I \times M)$ if $\partial_t^{k_1} f, \nabla^{k_2} f \in C_b^{\alpha_1,\alpha_2}(I \times M)$ holds for any integers $0 \le l_1 \le k_1$ and $0 \le l_2 \le k_2$.

For any $k \in \mathbb{Z}_+$ and $\alpha \in (0,1)$, we say that ∂D is uniformly of class $C^{k+\alpha}$, if it has a unique inward unit normal vector field N such that for some constant $r_0 > 0$, the polar coordinates

$$\partial D \times [0,r_0] \ni (z,r) \mapsto z + rN(z) \in D_{r_0} := \big\{x \in \bar{D}:\ \rho_\partial(x) := \mathrm{dist}(x,\partial D) \le r_0\big\}$$

is bijective, and each component belongs to $C_b^{k+\alpha}(\partial D \times [0,r_0])$. In this case, we have $\rho_\partial \in C_b^{k+\alpha}(D_{r_0})$.

To establish a local version Bismut formula for P_t^D, we make the following assumption.

(B) *Let $a := \sigma\sigma^*$. There exist constants $\alpha \in (0,1)$ and $K > 0$ such that*

(1) *∂D is uniformly of class $C^{3+\alpha}$, or uniformly of class $C^{2+\alpha}$ and $(a_t, b_t) = (a, b)$ does not depend on t.*

(2) *Either $(a_t, b_t) = (a, b)$ do not depend on t, or*

$$\|1_{D_{r_0}} b\|_\infty \le K \quad \text{for some} \quad r_0 > 0. \tag{2.5}$$

(3) *All components of b and a belong to $C^{\alpha/2,1}_{loc}([0,T] \times \mathbb{R}^d)$ such that*

$$\begin{aligned} &\|a\|_\infty + \|a^{-1}\|_\infty + \|\nabla\sigma\|_\infty^2 \le K, \\ &\langle \nabla_v b_t(x), v\rangle \le K|v|^2, \quad (t,x,v) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d. \end{aligned} \tag{2.6}$$

We have the following result.

Theorem 2.1. *Assume* **(B)**. *For $(t,x) \in (0,T] \times D$, let $(\beta_s)_{s\in[0,t]}$ be an adapted absolutely continuous real process such that*

$$\beta_0 = 1, \quad \beta_s = 0 \text{ for } s \ge \tau(x) \wedge t, \tag{2.7}$$

and for some $\varepsilon > 0$

$$C_{\beta,\varepsilon,t} := \mathbb{E}\left(\int_0^{t\wedge\tau(x)} |\beta_s'|^2 \,\mathrm{d}s\right)^{\frac{1+\varepsilon}{2}} < \infty, \tag{2.8}$$

then for any $v \in \mathbb{R}^d$ and $f \in \mathscr{B}_b(\bar{D})$,

$$\nabla_v P_t^D f(x) = -\mathbb{E}\left[\mathbf{1}_{\{t<\tau(x)\}} f(X_t^x) \int_0^t \beta_s' \Big\langle (\sigma^* a^{-1})_s(X_s^x)\nabla_v X_s^x, \mathrm{d}W_s \Big\rangle\right]. \tag{2.9}$$

Consequently, for any $p \in (1+\varepsilon^{-1}, \infty]$, there exists a constant $c(p,\varepsilon) > 0$ such that for any $t \in (0,T]$, $x \in D$ and $f \in \mathscr{B}_b(\bar{D})$,

$$\big|\nabla P_t^D f(x)\big| := \limsup_{y\to x} \frac{|P_t^D f(y) - P_t^D f(x)|}{|x-y|} \le c(p,\varepsilon) C_{\beta,\varepsilon,t}^{\frac{1}{1+\varepsilon}} \left(P_t^D |f|^p(x)\right)^{\frac{1}{p}}. \tag{2.10}$$

Remark 2.2. *We would like to make comments on the local Bismut formula (2.9) and the condition (2.5).*

(1) *When $(a_t, b_t) = (a, b)$ does not depend on t and is C^2-smooth, the local Bismut formula of type (2.9) was first established and applied in [12, 13]. The above result extends the main results in these papers to time dependent and less regular coefficients, which will be further applied to the singular case for coefficients satisfying assumption* **(A)**, *see Theorem 2.4 below.*

(2) *The condition* (2.5) *is used to ensure the gradient estimate* (2.12) *below due to [1, Theorem 5.3]. According to [5, Theorem 1.3], this estimate holds when* $(a_t, b_t) = (a, b)$ *does not depend on* t*, since* (2.6) *implies the Lyapunov condition*

$$\frac{1}{2}\mathrm{tr}\{a\nabla^2 V\} + b \cdot \nabla V \le \lambda_0 V \tag{2.11}$$

on $\bar{D}$ *for some* $\lambda_0 > 0$ *and* $V(x) = |x|^2$*.*

To prove this theorem, we first present the following lemma.

Lemma 2.3. *Assume* **(B)***.*

(1) *There exists a constant* $c > 0$ *depending only on* d, T, r_0, K *and* ∂D *such that*

$$\|\nabla P_{s,t}^D f\|_\infty \le \frac{c\|f\|_\infty}{\sqrt{t-s}}, \quad 0 \le s < t \le T,\ f \in \mathscr{B}_b(\bar{D}). \tag{2.12}$$

(2) *For any* $f \in \mathscr{B}_b(\bar{D})$ *and* $t \in (0, T]$*, we have* $P_{\cdot,t}^D f \in C_{loc}^{1+\frac{\alpha}{2}, 2+\alpha}([0, t) \times \bar{D})$*, and*

$$\partial_s P_{s,t}^D f = -L_s P_{s,t}^D f, \quad s \in [0, t), \tag{2.13}$$

where

$$L_s := \mathrm{tr}\{a_s \nabla^2\} + b_s \cdot \nabla, \quad s \in [0, t).$$

Proof. (1) Let $P_t^D f|_{D^c} = 0$ for $t > 0$. The gradient estimate (2.12) is equivalent to

$$|P_{s,t}^D f(x) - P_{s,t}^D f(y)| \le \frac{c|x-y|}{\sqrt{t-s}}\|f\|_\infty, \quad 0 \le s < t \le T,\ f \in \mathscr{B}_b(\bar{D}),\ x, y \in \bar{D}.$$

So, by the monotone class theorem, we only need to prove (2.12) for $f \in C_b(\bar{D})$.

If, in addition to **(B)**, all components of a and b belong to $C_{loc}^{\alpha/2, 1+\alpha}([0, T] \times \mathbb{R}^d)$, then (2.12) is known. Indeed, when ∂D is uniformly of class $C^{3+\alpha}$, (2.12) is derived in [1, Theorem 5.3] for $f \in C_b(\bar{D})$; when b_t and a_t do not depend on t and ∂D is uniformly of class $C^{2+\alpha}$, (2.12) can be found in [5, Theorem 1.3] where the constant c only depends on K, T and ∂D, see also [8].

In general, (2.12) can be derived by using mollifier approximations as follows. Let $0 \le g \in C_0^\infty(\mathbb{R}^{d+1})$ with support contained in $\{(r, x) \in \mathbb{R}\times\mathbb{R}^d : |(r, x)| \le 1\}$ such that $\int_{\mathbb{R}^{d+1}} g(r, x)\,\mathrm{d}r\mathrm{d}x = 1$. For any $n \ge 1$, let $g_n(r, x) = n^{d+1} g(nr, nx)$ and define

$$\begin{aligned} a_t^n(x) &:= \sigma_t^n(x)\sigma_t^n(x)^*, \quad \sigma_t^n(x) := \int_{\mathbb{R}\times\mathbb{R}^d} \sigma_{(t+s)^+\wedge T}(x+y) g_n(s, y)\,\mathrm{d}s\,\mathrm{d}y, \\ b_t^n(x) &:= \int_{\mathbb{R}\times\mathbb{R}^d} b_{(t+s)^+\wedge T}(x+y) g_n(s, y)\,\mathrm{d}s\,\mathrm{d}y, \quad (t, x) \in \mathbb{R} \times \mathbb{R}^d. \end{aligned}$$

Then $b^n, a^n \in C^\infty([0, T] \times \mathbb{R}^d)$, (2.6) holds for (b^n, a^n) in place of (b, a) with the same constant K, and when $n \ge n_0$ for a larger enough constant $n_0 \ge 1$, (2.5) holds for $(b^n, r_0/2)$ in place of

(b, r_0) with the same constant K. So, there exists a constant $c > 0$ depending only on d, T, r_0, K such that

$$\|\nabla P_{s,t}^{D,n} f\|_\infty \le \frac{c\|f\|_\infty}{\sqrt{t-s}}, \quad 0 \le s < t \le T,\ f \in \mathscr{B}_b(\bar{D}),\ n \ge n_0,$$

where $P_{s,t}^{D,n}$ is defined as in (2.3) for the solution to $X^{n,x}$ to the following SDE in place of X^x:

$$\mathrm{d}X_{s,t}^{n,x} = b_t^n(X_{s,t}^{n,x})\mathrm{d}t + \sigma_t^n(X_{s,t}^{n,x})\mathrm{d}W(t), \quad 0 \le s \le t \le T, \quad X_{s,s}^{n,x} = x \in D.$$

Noting that

$$\lim_{n\to\infty} P_{s,t}^{D,n} f = P_{s,t}^D f, \quad 0 \le s < t \le T,\ f \in C_b(\bar{D}),$$

see the proof of (2.32) in a more general setting where b may be singular, we derive (2.12).

(2) For any fixed $t \in (0, T]$ and $s \in [0, t)$, since $P_{r,t}^D f \in C_b(\bar{D})$ for $f \in \mathscr{B}_b(\bar{D})$ and $r \in (s, t)$ due to (2.12), by the flow property (2.4) we only need to prove (2.13) for $f \in C_b(\bar{D})$. In this case, by [1, Theorem 3.4], the following PDE has a unique bounded classical solution $u \in C_{loc}^{1+\alpha/2, 2+\alpha}((0, t] \times \bar{D})$:

$$\begin{aligned} &\partial_r u_r(x) = L_{t-r} u_r(x), \quad (r, x) \in (0, t] \times D, \\ &u_0 = f,\ u_r|_{\partial D} = 0, \quad r \in (0, t]. \end{aligned} \tag{2.14}$$

By Itô's formula,

$$\mathrm{d}u_{t-r}(X_{s,r}^x) = \big\langle \nabla u_{t-r}(X_{s,r}^x), \sigma_r(X_{s,r}^x)\mathrm{d}W_r \big\rangle, \quad x \in \bar{D},\ r \in [s, t) \cap [s, \tau(s, x)].$$

Combining this and the fact that for any $(s, x) \in [0, T) \times \bar{D}$,

$$u_{t-t\wedge\tau(s,x)}(X_{s,t\wedge\tau(s,x)}^x) = u_{t-\tau(s,x)}(X_{s,\tau(s,x)}^x) = 0 \text{ if } t \ge \tau(s, x),$$

we derive

$$\begin{aligned} u_{t-s}(x) &= \mathbb{E}[u_{t-s}(X_{s,s}^x)] = \mathbb{E}[u_{t-t\wedge\tau(s,x)}(X_{s,t\wedge\tau(s,x)}^x)] \\ &= \mathbb{E}[1_{\{t<\tau(s,x)\}} f(X_{s,t}^x)] = P_{s,t}^D f(x), \quad x \in \bar{D},\ 0 \le s \le t. \end{aligned}$$

Then (2.14) implies (2.13). □

Proof of Theorem 2.1. For any fixed $x \in D$, we simply denote $X_t^x = X_t$ and $\tau(x) = \tau$. Let $f \in \mathscr{B}_b(\bar{D})$.

(a) Below, we prove that (2.10) can be deduced from (2.9). By (2.6), for any $v \in \mathbb{R}^d$, the derivative process $v_t := \nabla_v X_t$ solves the SDE

$$\mathrm{d}v_t = (\nabla_{v_t} b_t)(X_t)\mathrm{d}t + (\nabla_{v_t}\sigma)(X_t)\mathrm{d}W_t, \quad v_0 = v,\ t \in [0, T],$$

and

$$\mathrm{d}|v_t|^2 = \big\{2\langle(\nabla_{v_t} b_t)(X_t), v_t\rangle + \|\nabla_{v_t}\sigma\|^2(X_t)\big\}\mathrm{d}t + \mathrm{d}M_t \le 3K|v_t|^2\mathrm{d}t + \mathrm{d}M_t, \quad t \in [0, T],$$

where $\mathrm{d}M_t := 2\big\langle v_t, (\nabla_{v_t}\sigma)(X_t)\mathrm{d}W_t\big\rangle$ satisfies

$$\mathrm{d}\langle M\rangle_t \le 4K|v_t|^2\mathrm{d}t.$$

By the Burkholder-Davis-Gundy inequality, for any $k \in [1,\infty)$, we find a constant $c(k) > 0$ such that (2.2) holds. Combining this with Hölder's inequality, for any $t \in [0,T]$, $p > 1+\varepsilon^{-1}$ and $q = \frac{(p-1)(1+\varepsilon)}{\varepsilon p-1-\varepsilon}$, we obtain

$$\begin{aligned}
A_t &:= \mathbb{E}\bigg[\bigg(\int_0^{t\wedge\tau} \big|\beta_s'(\sigma^*a^{-1})_s(X_s^x)\nabla_v X_s\big|^2\mathrm{d}s\bigg)^{\frac{p}{2(p-1)}}\bigg]\\
&\le \|\sigma^*a^{-1}\|_\infty^{\frac{p}{p-1}}\mathbb{E}\bigg[\bigg(\Big(\sup_{s\in[0,t]}|\nabla_v X_s|^{\frac{p}{p-1}}\Big)\bigg(\int_0^{t\wedge\tau}\big|\beta_s'\big|^2\mathrm{d}s\bigg)^{\frac{p}{2(p-1)}}\bigg]\\
&\le \|\sigma^*a^{-1}\|_\infty^{\frac{p}{p-1}}\bigg(\mathbb{E}\Big[\sup_{s\in[0,t]}|\nabla_v X_s|^{\frac{pq}{p-1}}\Big]\bigg)^{\frac{1}{q}}\bigg[\mathbb{E}\bigg(\int_0^{t\wedge\tau}|\beta_r'|^2\mathrm{d}r\bigg)^{\frac{1+\varepsilon}{2}}\bigg]^{\frac{p}{(p-1)(1+\varepsilon)}} < \infty
\end{aligned} \tag{2.15}$$

due to (2.2), (2.6) and (2.8). By (2.9), Hölder's inequality and the Burkholder-Davis-Gundy inequality, we find a constant $c_p > 0$ such that for any $t \in (0,T]$,

$$\begin{aligned}
&|\nabla_v P_t f(x)| = \bigg|\mathbb{E}\bigg[\mathbf{1}_{\{t<\tau\}}f(X_t)\int_0^{t\wedge\tau}\beta_s'\Big\langle(\sigma^*a^{-1})_s(X_s)\nabla_v X_s, \mathrm{d}W_s\Big\rangle\bigg]\bigg|\\
&\le c_p\big(P_t^D|f|^p(x)\big)^{\frac{1}{p}}A_t^{\frac{p-1}{p}}\\
&\le c_p\|\sigma^*a^{-1}\|_\infty\big(P_t^D|f|^p(x)\big)^{\frac{1}{p}}\mathbb{E}\bigg(\sup_{s\in[0,t]}|\nabla_v X_s|^{\frac{pq}{p-1}}\bigg)^{\frac{p-1}{pq}}\bigg[\mathbb{E}\bigg(\int_0^{t\wedge\tau}|\beta_r'|^2\mathrm{d}r\bigg)^{\frac{1+\varepsilon}{2}}\bigg]^{\frac{1}{1+\varepsilon}}.
\end{aligned} \tag{2.16}$$

So, (2.10) follows from (2.2) and (2.6).

(b) We now establish (2.9). For any fixed $t \in (0,T]$, by (2.7) and (2.15),

$$M_s := \int_0^s \beta_r'\big\langle(\sigma^*a^{-1})_r(X_r)\nabla_v X_r, \mathrm{d}W_r\big\rangle = \int_0^{s\wedge\tau}\beta_r'\big\langle(\sigma^*a^{-1})_r(X_r)\nabla_v X_r, \mathrm{d}W_r\big\rangle, \quad s\in[0,t]$$

is a uniformly integrable martingale. By (2.12), (2.13) and Itô's formula,

$$P_{s,t}^D f(X_s) - P_t^D f(x) = \int_0^s \big\langle\nabla P_{r,t}^D f(X_r), \sigma_r(X_r)\mathrm{d}W_r\big\rangle =: N_s, \quad s\in[0,t)\cap[0,\tau] \tag{2.17}$$

is a martingale. So,

$$I(s) := P_{s,t}^D f(X_s)\int_0^s \beta_r'\big\langle(\sigma^*a^{-1})_r(X_r)\nabla_v X_r, \mathrm{d}W_r\big\rangle, \quad s\in[0,t)\cap[0,\tau]$$

is a uniformly integrable process with

$$I(s) = (N_s + P_t^D f(x))M_s = I_1(s) + I_2(s) + I_3(s), \quad s\in[0,t)\cap[0,\tau], \tag{2.18}$$

where

$$I_1(s) := \int_0^s \beta_r' \langle \nabla_v X_r, \nabla P_{r,t}^D f(X_r)\rangle \mathrm{d}r, \quad s \in [0,t),$$
$$I_2(s) := \int_0^s P_{r,t}^D f(X_r) \beta_r' \langle (\sigma^* a^{-1})_r(X_r)\nabla_v X_r, \mathrm{d}W_r\rangle, \quad s \in [0,t),$$
$$I_3(s) := \int_0^s M_r \mathrm{d}N_r, \qquad s \in [0,t) \cap [0,\tau].$$

By (2.2), (2.6), (2.13) and the condition on β_s, we see that $I_2(s)$ and $I_3(s)$ are martingales for $s \in [0,t) \cap [0,\tau]$. Moreover, by $\beta_0 = 1$, the chain rule, integration by parts formula, and (2.17), we obtain

$$\begin{aligned} I_1(s) &= \int_0^s \beta_r' \nabla_v \big[P_{r,t}^D f(X_r)\big] \mathrm{d}r \\ &= \beta_s \nabla_v \big[P_{s,t}^D f(X_s)\big] - \nabla_v P_t^D f(x) - \int_0^s \beta_r \nabla_v \big[\mathrm{d}P_{r,t}^D f(X_r)\big] \\ &= \beta_s \nabla_v \big[P_{s,t}^D f(X_s)\big] - \nabla_v P_t^D f(x) - \tilde{M}_s, \quad s \in [0,t), \end{aligned} \tag{2.19}$$

where, by Lemma 2.3 and the boundness of β_r and $\nabla\sigma$,

$$\tilde{M}_s := \int_0^s \beta_r \Big[\mathrm{Hess}_{P_{r,t}^D f(X_r)}(\nabla_v X_r, \sigma_r(X_r)\mathrm{d}W_r) + \big\langle \nabla P_{r,t}^D f(X_r), \nabla_{\nabla_v X_r}(\sigma_r)(X_r)\mathrm{d}W_r\big\rangle\Big], \quad s \in [0,t)$$

is a local martingale, i.e. for

$$\hat{\tau}_n := \inf\{t \geq 0 : \ |X_t| \geq n\}, \quad n \geq 1,$$

we have $\hat{\tau}_n \to \infty$ as $n \to \infty$, and for each $n \geq 1$,

$$[0,t) \ni s \to \tilde{M}_{s\wedge\hat{\tau}_n}$$

is a martingale. Combining (2.18) with (2.19), and noting that $I_2(s) + I_3(s)$ is a martingale for $s \in [0,t) \cap [0,\tau]$, we conclude that for any $n \geq 1$ and $s < t$,

$$\begin{aligned} 0 = I(0) &= \mathbb{E}[I(s\wedge\tau\wedge\hat{\tau}_n) - I_1(s\wedge\tau\wedge\hat{\tau}_n)] \\ &= \mathbb{E}\Big(I(s\wedge\tau\wedge\hat{\tau}_n) + \nabla_v P_t^D f(x) - \beta_{s\wedge\tau\wedge\hat{\tau}_n} \nabla_v\big[P_{s\wedge\tau\wedge\hat{\tau}_n,t} f(X_{s\wedge\tau\wedge\hat{\tau}_n})\big]\Big). \end{aligned}$$

Combining this with the uniform integrability of $\{I(s\wedge\tau)\}_{s\in[0,t)}$, the boundness of β, and noting that (2.12) implies

$$\sup_{r\in[0,s]} \|\nabla P_{r,t}^D f\|_\infty < \infty, \quad s \in [0,t),$$

it follows from dominated convergence theorem that

$$\begin{aligned} \nabla_v P_t^D f(x) &= \lim_{s\uparrow t, n\uparrow\infty} \Big[\mathbb{E}\Big(-I(s\wedge\tau\wedge\hat{\tau}_n) + \beta_{s\wedge\tau\wedge\hat{\tau}_n} \nabla_v\big[P_{s\wedge\tau\wedge\hat{\tau}_n,t} f(X_{s\wedge\tau\wedge\hat{\tau}_n})\big]\Big)\Big] \\ &= \lim_{s\uparrow t} \Big[\mathbb{E}\Big(-I(s\wedge\tau) + \beta_{s\wedge\tau} \nabla_v\big[P_{s\wedge\tau,t} f(X_{s\wedge\tau})\big]\Big)\Big]. \end{aligned} \tag{2.20}$$

Noting that $\mathbb{P}(\tau=t)=0$ and $P^D_{\tau,t}f(X_\tau)=0$ for $\tau<t$, we have $\mathbb{P}$-a.e.,

$$P^D_{t\wedge\tau,t}f(X_{t\wedge\tau})=1_{\{t<\tau\}}f(X_t)+1_{\{t>\tau\}}P^D_{\tau,t}f(X_\tau)=1_{\{t<\tau\}}f(X_t),$$

so that, by the uniform integrability of $\{I(s\wedge\tau)\}_{s\in[0,t)}$ and the dominated convergence theorem,

$$\lim_{s\uparrow t}\mathbb{E}[I(s\wedge\tau)]=\mathbb{E}[I(t\wedge\tau)]=\mathbb{E}\bigg[1_{\{t<\tau\}}f(X_t)\int_0^t\beta_r'\langle\zeta_r(X_r)\nabla_vX_r,\mathrm{d}W_r\rangle\bigg]. \tag{2.21}$$

By (2.2), (2.12) and $\|f\|_\infty<\infty$, we find constants $c_0,c_1>0$ independent of t and $s\in[0,t)$, such that

$$\begin{aligned}&\mathbb{E}\big|\beta_{s\wedge\tau}\nabla_vP^D_{s\wedge\tau,t}f(X_{s\wedge\tau})\big|=\mathbb{E}\big|\beta_{s\wedge\tau}\langle\nabla P^D_{s\wedge\tau,t}f(X_{s\wedge\tau}),\nabla_vX_{s\wedge\tau}\rangle\big|\\&\le\frac{c_0}{\sqrt{t-s}}\Big(\mathbb{E}\big|\beta_{s\wedge\tau}\big|^{1+\varepsilon}\Big)^{\frac{1}{1+\varepsilon}}\Big(\mathbb{E}|\nabla_vX_{s\wedge\tau}|^{\frac{1+\varepsilon}{\varepsilon}}\Big)^{\frac{\varepsilon}{1+\varepsilon}}\le\frac{c_1|v|}{\sqrt{t-s}}\Big(\mathbb{E}\big|\beta_{s\wedge\tau}\big|^{1+\varepsilon}\Big)^{\frac{1}{1+\varepsilon}}.\end{aligned} \tag{2.22}$$

Moreover, by $\beta_{t\wedge\tau}=0$ and Hölder's inequality, we have

$$|\beta_{s\wedge\tau}|^{1+\varepsilon}=\bigg|\int_{s\wedge\tau}^{t\wedge\tau}\beta_r'\,\mathrm{d}r\bigg|^{1+\varepsilon}\le(t-s)^{\frac{1+\varepsilon}{2}}\bigg(\int_{s\wedge\tau}^{t\wedge\tau}|\beta_r'|^2\,\mathrm{d}r\bigg)^{\frac{1+\varepsilon}{2}},\quad s\in[0,t]$$

so that (2.8) implies

$$\frac{1}{\sqrt{t-s}}\Big(\mathbb{E}\big|\beta_{s\wedge\tau}\big|^{1+\varepsilon}\Big)^{\frac{1}{1+\varepsilon}}\le\bigg(\mathbb{E}\bigg[\bigg(\int_{s\wedge\tau}^{t\wedge\tau}|\beta_r'|^2\,\mathrm{d}r\bigg)^{\frac{1+\varepsilon}{2}}\bigg]\bigg)^{\frac{1}{1+\varepsilon}}\to0\text{ as }s\uparrow t.$$

This together with (2.20), (2.21) and (2.22) implies (2.9) for $f\in\mathscr{B}_b(\bar D)$. □

2.2 Gradient estimate in the singular case

(C) *Assume that* **(A)** *and* **(B)**(1)-(2) *hold for some* $r_0>0$.

Theorem 2.4. *Assume* **(C)**, *and let* P_t^D *be associated with the SDE* (2.1). *Then for any* $p>2$ *there exists a constant* $c(p)>0$ *such that for any* $t\in(0,T]$, $x\in D$ *and* $f\in\mathscr{B}_b(\bar D)$,

$$|\nabla P_t^Df(x)|\le\frac{c(p)}{\sqrt{t}\,(\rho_\partial(x)\wedge1)}\big(P_t^D|f|^p(x)\big)^{\frac1p}. \tag{2.23}$$

Proof. By induction in $k_1\in\mathbb{N}$ as in the proof of [6, Proposition 5.2], we only need to prove for $k_1=1$.

(a) To apply (2.10), we first make regular approximation of the SDE. Let $0\le g\in C_0^\infty(\mathbb{R}^{d+1})$ with support contained in $\{(r,x)\in\mathbb{R}\times\mathbb{R}^d:|(r,x)|\le1\}$ such that $\int_{\mathbb{R}^{d+1}}g(r,x)\,\mathrm{d}r\mathrm{d}x=1$. For any $n\ge1$, let $g_n(r,x)=n^{d+1}g(nr,nx)$ and

$$b_t^n(x)=(b^{(0)}+b^{1,n})_t(x),\quad b_t^{1,n}(x):=\int_{\mathbb{R}\times\mathbb{R}^d}b^{(1)}_{(t+s)^+\wedge T}(x+y)g_n(s,y)\,\mathrm{d}s\,\mathrm{d}y,$$

$$\sigma_t^n(x)=\int_{\mathbb{R}\times\mathbb{R}^d}\sigma_{(t+s)^+\wedge T}(x+y)g_n(s,y)\,\mathrm{d}s\,\mathrm{d}y,\quad a_t^n(x):=\sigma_t^n(x)\sigma_t^n(x)^*,\quad (t,x)\in\mathbb{R}\times\mathbb{R}^d.$$

For any $n\geq 1$ and any $x\in D$, denote $\tau_n(x)=\tau(X^{n,x})$, and

$$P_t^{D,n}f(x):=\mathbb{E}\left[\mathbf{1}_{\{t<\tau(X^{n,x})\}}f(X_t^{n,x})\right],\quad t\in[0,T],\ f\in\mathscr{B}_b(\bar{D}),$$

where $X_t^{n,x}$ solves

$$\mathrm{d}X_t^{n,x}=b_t^n(X_t^{n,x})\mathrm{d}t+\sigma_t^n(X_t^{n,x})\mathrm{d}W_t,\quad t\in[0,T],\quad X_0^{n,x}=x\in D.$$

Under **(C)**, for any $n\geq 1$, **(B)** holds for (b^n,a^n) in place of (a,b). By (2.16) with $\varepsilon=1$ for $P_t^{D,n}$ in place of P_t^D, for any $p>2$, we have

$$\left|\nabla_v P_t^{D,n}f(x)\right|\leq(P_t^{D,n}|f|^p(x))^{\frac{1}{p}}\|(\sigma^n)^*(a^n)^{-1}\|_\infty C_{\beta,1,t}^{\frac{1}{2}}\left(\mathbb{E}\sup_{s\in[0,t]}|\nabla_v X_s^{n,x}|^{\frac{2p}{p-2}}\right)^{\frac{p-2}{2p}}\tag{2.24}$$

for any $x\in D$, $v\in\mathbb{R}^d$, $t\in(0,T]$, $f\in\mathscr{B}_b(\bar{D})$ and $n\geq 1$, where β_s is an adapted real process with $\beta_0=1$ and $\beta_s=0$ for $s\geq t\wedge\tau_n$ will be constructed such that (2.8) holds for $\varepsilon=1$.

(b) Let $f_i^n:=\int_{\mathbb{R}\times\mathbb{R}^d}f_i(x+y)g_n(s,y)\,\mathrm{d}s,\ 1\leq i\leq k_2$. Then

$$\|\nabla a^n\|\leq\sum_{i=1}^{k_2}f_i^n,$$

and

$$\begin{aligned}
&\lim_{n\to\infty}\left\{\|b^{1,n}-b^{(1)}\|_{\tilde{L}_{q_1}^{p_1}(T)}+\sum_{i=1}^{k_2}\|f_i^n-f_i\|_{\tilde{L}_{q_i}^{p_i}(T)}\right\}=0,\\
&\lim_{n\to\infty}\sup_{t\in[0,T]}\sup_{x\in\mathbb{R}^d}\|\sigma_t^n(x)-\sigma_t(x)\|_{HS}=0,\\
&\sup_{n\geq 1}\left\{\|a^n\|_\infty+\|(a^n)^{-1}\|_\infty\right\}\leq\|a\|_\infty+\|a^{-1}\|_\infty,\\
&\sup_{n\geq 1}\left\{\|b^{1,n}\|_{\tilde{L}_{q_1}^{p_1}(T)}+\sum_{i=1}^{k_2}\|f_i^n\|_{\tilde{L}_{q_i'}^{p_i'}(T)}\right\}\leq\|b^{(1)}\|_{\tilde{L}_{q_1}^{p_1}(T)}+\sum_{i=1}^{k_2}\|f_i\|_{\tilde{L}_{q_i'}^{p_i'}(T)}.
\end{aligned}\tag{2.25}$$

By the proof of [17, Theorem 3.9], we have

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^{n,x}-X_t^x|\right]=0,\quad x\in\mathbb{R}^d.\tag{2.26}$$

In addition, as shown in the proof of [15, Theorem 2.1(1)], for any $k\in[1,\infty)$, one finds a constant $c(k)>0$ such that

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

This together with (2.24) implies that for for any $p>2$, there exists a constant $c_1>0$ uniformly in n such that

$$\left|\nabla P_t^{D,n} f(x)\right| \leq c_1 (P_t^{D,n}|f|^p(x))^{\frac{1}{p}} \left(\mathbb{E}\int_0^t |\beta_s'|^2 \mathrm{d}s\right)^{\frac{1}{2}} \tag{2.27}$$

holds for any $t\in(0,T]$, $x\in D$, $f\in \mathscr{B}_b(\bar D)$ and $n\geq 1$.

(c) To derive explicit gradient estimate from (2.27), we construct the process β by following the line of [13]. Since ∂D is uniformly of class $C^{2+\alpha}$, we find a constant $r_1\in(0,r_0)$ such that $\rho_\partial\in C_b^2(D_{r_1})$, see [5, Proposition B.3]. Let $h\in C_b^\infty([0,\infty))$ such that

$$0\leq h\leq 1,\quad h'\geq 0,\quad h(r)=r \text{ for } r\leq \frac{r_1}{2},\quad h(r)=1 \text{ for } r\geq r_1. \tag{2.28}$$

Then $g:=h\circ\rho_\partial\in C_b^2(\bar D)$ and there exists a constant $C_0>1$ such that

$$C_0^{-1}\leq \frac{g}{\rho_\partial\wedge 1}(x)\leq C_0,\quad x\in\bar D. \tag{2.29}$$

By **(C)**, $|b|+\|a\|+\|\nabla^2\rho_\partial\|$ is bounded in D_{r_0}. Then there exists large enough $n_0\in\mathbb{N}$ such that the family $\{|b^n|+\|a^n\|\}_{n\geq n_0}$ is uniformly bounded on $[0,T]\times D_{r_1}$. So, there exists a constant $C_1>0$ such that

$$L_t^n g(x) := \frac{1}{2}\mathrm{tr}\{a_t^n(x)\nabla^2 g(x)\} + b_t^n(x)\cdot\nabla g(x)\leq C_1,\quad (t,x)\in[0,T]\times D_{r_1},\ n\geq n_0. \tag{2.30}$$

Now, for any fixed $x\in D$, simply denote $X_t^{n,x}=X_t^n, \tau_n(x)=\tau_n$ and define

$$S_n(t)=\int_0^t g^{-2}(X_{s\wedge\tau}^n)\mathrm{d}s,\qquad s_n(t)=\inf\{s\in[0,T]: S_n(s)\geq t\},\quad t\in[0,T]\ , n\geq 1.$$

By $0\leq g\leq 1$, (2.29) and (2.30), the same argument in [13, Section 4] implies that $S_n(t)$ is increasing with $S_n(t)=\infty$ for $t>\tau_n$, $s_n(t)\leq t\wedge\tau_n$, $S_n(s_n(t))=t$ for $t\leq S_n(\tau_n)$, $s_n(S_n(t))=t$ for $t\leq\tau_n$. Hence, for any $t\in(0,T]$,

$$\beta_s := 1-\frac{1}{t}\int_0^{s\wedge s_n(t)} g^{-2}(X_r^n)\mathrm{d}r,\quad s\in[0,t]$$

satisfies $\beta_0=1$, and when $s\geq\tau_n$ $(\geq s_n(t))$,

$$\beta_s = 1-\frac{1}{t}\int_0^{s_n(t)} g^{-2}(X_r^n)\,\mathrm{d}r = 1-\frac{1}{t}S_n(s_n(t))=0.$$

Moreover, because of (2.28), $g:=h\circ\rho_\partial$ and (2.30), by repeating the calculations on page 118 in [13] for (L_t^n, g) in pace of (L,f) therein, we find a constant $c_3>0$ such that

$$\mathbb{E}\left[\int_0^{t\wedge\tau_n}|\beta_s'|^2\,\mathrm{d}s\right] = \frac{1}{t^2}\mathbb{E}\left[\int_0^{s_n(t)} g^{-4}(X_s^n)\,\mathrm{d}s\right]\leq \frac{c_3}{tg^2(x)},\quad n\geq n_0.$$

Combining this with (2.27) and (2.29), for any $p > 2$ we find a constant $c > 0$ such that

$$\left|\nabla P_t^{D,n} f(x)\right| \le \frac{c}{\sqrt{t}(\rho_\partial(x) \wedge 1)} (P_t^{D,n}|f|^p(x))^{\frac{1}{p}}, \quad n \ge n_0,\ f \in \mathscr{B}_b(\bar{D}). \tag{2.31}$$

(d) Let $P_t^{D,n} f|_{D^c} = 0$ for $t > 0$, so that $\nabla P_t^{D,n} f|_{D^c} = 0$. Then (2.31) is equivalent to

$$\begin{aligned}
&\left|P_t^{D,n} f(x) - P_t^{D,n} f(y)\right| \\
&\le \frac{c|x-y|}{\sqrt{t}} \int_0^1 \frac{1}{\rho_\partial(sx + (1-s)y) \wedge 1} \big(P_t^{D,n}|f|^p(sx + (1-s)y)\big)^{\frac{1}{p}} \mathrm{d}s
\end{aligned}$$

for any $n \ge n_0$, $t \in (0, T]$, $x, y \in \bar{D}$ and $f \in \mathscr{B}_b(\bar{D})$. Combining with the flow property (2.4), to deduce the desired gradient estimate on $P_t^D f$, it remains to show

$$P_t^D f = \lim_{n\to\infty} P_t^{D,n} f, \quad t \in (0, T],\ f \in C_b(\bar{D}). \tag{2.32}$$

For any $t \in (0, T]$, $x \in D$ and $n \ge 1$,

$$\begin{aligned}
\mathbb{P}\big(t < \tau, t \ge \tau_n\big) &= \mathbb{P}\left(\inf_{s\in[0,t]} \rho_\partial(X_s) > 0,\ \inf_{s\in[0,t]} \rho_\partial(X_s^n) = 0\right) \\
&\le \mathbb{P}\left(\sup_{s\in[0,t]} |X_s^n - X_s| \ge \varepsilon\right) + \mathbb{P}\left(0 < \inf_{s\in[0,t]} \rho_\partial(X_s) < \varepsilon\right).
\end{aligned}$$

By first letting $n \uparrow \infty$ and then $\varepsilon \downarrow 0$, it follows from (2.26) that

$$\lim_{n\uparrow\infty} \mathbb{P}\big(t < \tau, t \ge \tau_n\big) = 0, \quad t \in (0, T],\ x \in D.$$

Similarly, by $\mathbb{P}(\tau = t) = 0$ and (2.26), letting

$$\rho(x, D) := \inf_{y\in D} |x - y|, \quad x \in \bar{D},$$

we have

$$\begin{aligned}
&\mathbb{P}\big(t \ge \tau, t < \tau_n\big) = \mathbb{P}\big(t > \tau, t < \tau_n\big) \\
&= \mathbb{P}\left(\sup_{s\in[0,t]} \rho(X_s, D) > 0,\ \sup_{s\in[0,t]} \rho(X_s^n, D) = 0\right) \\
&\le \mathbb{P}\left(\sup_{s\in[0,t]} |X_s^n - X_s| \ge \varepsilon\right) + \mathbb{P}\left(0 < \sup_{s\in[0,t]} \rho(X_s, D) < \varepsilon\right)
\end{aligned}$$

for any $t \in (0, T]$, $x \in D$ and $n \ge 1$, which tends to 0 by letting first $n \uparrow \infty$ then $\varepsilon \downarrow 0$. Hence,

$$\mathbf{1}_{\{t<\tau_n\}} \xrightarrow{\text{a.s.}} \mathbf{1}_{\{t<\tau\}} \text{ as } n \to \infty, \quad t \in (0, T],\ x \in D.$$

By combining this with (2.26) and the dominated convergence theorem, we derive (2.32). Therefore, the proof is finished. $\square$

3 Gradient estimate for killed DDSDEs

In this part, we establish the estimates (1.3) and (1.5) for the killed DDSDE (1.6) in the regular and singular cases respectively.

3.1 Well-posedness of (1.6)

The well-posedness of (1.6) has been derived in [14], where the noise may also depend on distribution, and coefficients are Lispchitz continuous in the distribution variable with respect to the (truncated) 1-Wasserstein distance. Below we present a result ensuring the well-posedness for the case that the noise is distribution-free and the drift is Lipschitz continuous in the distribution variable with respect to the total variation distance. This situation is not included in [14], although the proof is more or less standard by using the fixed point argument in distribution parameter.

For any $\mu \in C^w([0,T];\mathscr{P}_D)$, the set of all weakly continuous maps from $[0,T]$ to $\mathscr{P}_D$, let $a := \sigma\sigma^*$ and

$$b_t^\mu(x) := b_t(x,\mu_t), \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

(D) *There exists $0 \le \psi \in L^2([0,T])$ such that*

$$|b_t(x,\mu) - b_t(x,\nu)| \le \psi_t \|\mu-\nu\|_{var}, \quad t \in [0,T],\ x \in \mathbb{R}^d,\ \mu,\nu \in \mathscr{P}_D. \tag{3.1}$$

Moreover, one of the following conditions hold for any $\mu \in C^w([0,T];\mathscr{P}_D)$:

- **(A)** *holds for b^μ in place of b with constant K possibly depend on μ, or*
- *$\|a\|_\infty + \|a^{-1}\|_\infty < \infty$, and σ_t and b_t^μ are continuous on $\mathbb{R}^d$ such that for some $0 \le h \in L^1([0,T])$ possibly depend on μ,*

$$\langle b_t^\mu(x) - b_t^\mu(y), x-y\rangle + \frac{1}{2}\|\sigma_t(x)-\sigma_t(y)\|_{HS}^2 \le h_t|x-y|^2, \quad t \in [0,T],\ x,y \in \mathbb{R}^d.$$

Theorem 3.1. *Assume* **(D)**. *Then the SDE* (1.6) *is well-posed, and there exists a constant $c > 0$ such that the associated map $P_t^{D*} : \mathscr{P}_D \to \mathscr{P}_D$ satisfies*

$$\|P_t^{D*}\gamma - P_t^{D*}\tilde\gamma\|_{var} \le c\|\gamma-\tilde\gamma\|_{var}, \quad t \in [0,T],\ \gamma,\tilde\gamma \in \mathscr{P}_D. \tag{3.2}$$

Proof. (a) Let X_0 be a fixed $\mathscr{F}_0$-measurable random variable on $\bar D$ with $\gamma := \mathscr{L}_{X_0}^D \in \mathscr{P}_D$. Under **(D)**, for any

$$\mu \in \mathscr{C}_T^\gamma := \big\{\mu \in C^w([0,T];\mathscr{P}_D):\ \mu_0 = \gamma\big\},$$

the SDE

$$\mathrm{d}X_t^\mu = b_t^\mu(X_t^\mu)\mathrm{d}t + \sigma_t(X_t^\mu)\mathrm{d}W_t, \quad X_0^\mu = X_0,\ t \in [0,T] \tag{3.3}$$

is well-posed, which follows from [6, Proposition 5.2] when **(A)** holds for b^μ in place of μ, and is well-known in the other monotone case. Let

$$\Phi_t\mu := \mathscr{L}_{X_t^\mu}^D, \quad t \in [0,T].$$

Then $\Phi:\mathscr{C}_T^\gamma\to\mathscr{C}_T^\gamma$, and for the well-posedness it suffices to show that Φ has a unique fixed point $\bar\mu\in\mathscr{C}_T^\gamma$, and in this case $X_t:=X^{\bar\mu}_{t\wedge\tau(X^{\bar\mu})}$ becomes the unique solution of (1.6), where

$$\tau(X^{\bar\mu}):=\inf\{t\ge 0:\ X_t^{\bar\mu}\in\partial D\}.$$

To this end, we will find $\lambda>0$ such that Φ is contractive under the following complete metric on $\mathscr{C}_T^\gamma$:

$$\|\mu-\nu\|_\lambda:=\sup_{t\in[0,T]}\mathrm{e}^{-\lambda t}\|\mu_t-\nu_t\|_{var},\quad \mu,\nu\in\mathscr{C}_T^\gamma.$$

For a different element $\nu\in\mathscr{C}_T^\gamma$, we reformulate (3.3) as

$$\mathrm{d}X_t^\mu=b_t^\nu(X_t^\mu)\mathrm{d}t+\sigma_t(X_t^\mu)\mathrm{d}\tilde W_t,\ \ X_0^\mu=X_0,\ t\in[0,T],$$

where, by (3.1) and Girsanov's theorem,

$$\tilde W_t:=W_t-\int_0^t(\sigma_s^*a_s^{-1})(X_s^\mu)\big\{b_s^\nu(X_s^\mu)-b_s^\mu(X_s^\mu)\big\}\mathrm{d}s,\ \ t\in[0,T]$$

is an m-dimensional Brownian motion under the weighted probability $\mathrm{d}\mathbb{Q}:=R_T\mathrm{d}\mathbb{P}$, for the exponential martingale

$$R_t:=\mathrm{e}^{\int_0^t\langle\xi_s,\mathrm{d}W_s\rangle-\frac12\int_0^t|\xi_s|^2\mathrm{d}s},\quad \xi_s:=(\sigma_s^*a_s^{-1})(X_s^\mu)\big\{b_s^\nu(X_s^\mu)-b_s^\mu(X_s^\mu)\big\},\quad 0\le s\le t\le T.$$

So, by the weak uniqueness of (3.3),

$$\|\Phi_t\mu-\Phi_t\nu\|_{var}=\sup_{\|f\|_\infty\le1}\big|\mathbb{E}\big[1_{\{t<\tau(X^\mu)\}}(R_t-1)f(X_t^\mu)\big]\big|\le\mathbb{E}[|R_t-1|],\ \ t\in[0,T].$$

This together with Pinsker's inequality and (3.1) implies that

$$\|\Phi_t\mu-\Phi_t\nu\|_{var}\le\sqrt{2\mathbb{E}[R_t\log R_t]}\le\bigg(c_0\int_0^t\psi_s^2\|\mu_s-\nu_s\|_{var}^2\mathrm{d}s\bigg)^{\frac12},\ \ t\in[0,T]\tag{3.4}$$

holds for some constant $c_0>0$. Thus, when $\lambda>0$ is large enough,

$$\begin{aligned}\|\Phi\mu-\Phi\nu\|_\lambda&:=\sup_{t\in[0,T]}\mathrm{e}^{-\lambda t}\|\Phi_t\mu-\Phi_t\nu\|_{var}\\&\le\|\mu-\nu\|_\lambda\sup_{t\in[0,T]}\bigg(c_0\int_0^t\psi_s^2\mathrm{e}^{-2(t-s)}\mathrm{d}s\bigg)^{\frac12}\le\frac12\|\mu-\nu\|_\lambda,\quad\mu,\nu\in\mathscr{C}_T^\gamma.\end{aligned}$$

Therefore, Φ has a unique fixed point in $\mathscr{C}_T^\gamma$.

(b) Let $\tilde\gamma\in\mathscr{P}_D$ be different from γ, and choose

$$\mu_t=P_t^{D*}\gamma,\quad \nu_t=P_t^{D*}\tilde\gamma,\quad t\in[0,T].$$

Let X_0 and Y_0 be $\mathscr{F}_0$-measurable random variables on $\bar D$ such that

$$\mathscr{L}_{X_0}^D=\gamma,\ \ \mathscr{L}_{Y_0}^D=\tilde\gamma,\quad \mathbb{P}\big(\mathbf{1}_{\{X_0\in D\}}X_0\ne\mathbf{1}_{\{Y_0\in D\}}Y_0\big)\le2\|\gamma-\tilde\gamma\|_{var}.$$

Let $\tilde X_t^\mu$ and $\tilde X_t^\nu$, with $\tilde X_0^\mu = X_0$ and $\tilde X_0^\nu = Y_0$, solve the following SDE

$$\mathrm{d}\tilde X_t = b_t^\nu(\tilde X_t)\mathrm{d}t + \sigma_t(\tilde X_t)\mathrm{d}W_t, \quad t \in [0,T]. \tag{3.5}$$

Then

$$P_t^{D*}\gamma = \Phi_t\mu := \mathscr{L}^D_{X_t^\mu}, \quad \Phi_t\nu = \mathscr{L}^D_{\tilde X_t^\mu}, \quad P_t^{D*}\tilde\gamma = \mathscr{L}^D_{\tilde X_t^\nu}, \quad t \in [0,T].$$

Moreover, the pathwise uniqueness of (3.5) implies

$$\big\{1_{\{X_0\in D\}}X_0 = \mathbf{1}_{\{Y_0\in D\}}Y_0\big\} \subset \big\{1_{\{t<\tau(\tilde X^\mu)\}}\tilde X_t^\mu = 1_{\{t<\tau(\tilde X^\nu)\}}\tilde X_t^\nu\big\},$$

so that

$$\|\Phi_t\nu - P_t^{D*}\tilde\gamma\|_{var} = \|\mathscr{L}^D_{\tilde X_t^\mu} - \mathscr{L}^D_{\tilde X_t^\nu}\|_{var} \le \|\gamma-\tilde\gamma\|_{var}.$$

Combining this with (3.4), we obtain

$$\begin{aligned}\|P_t^{D*}\gamma - P_t^{D*}\tilde\gamma\|_{var} &\le \|\Phi_t\mu - \Phi_t\nu\|_{var} + \|\Phi_t\nu - P_t^{D*}\tilde\gamma\|_{var}\\ &\le \bigg(c_0\int_0^t \psi_s^2\|P_s^{D*}\gamma - P_s^{D*}\tilde\gamma\|_{var}^2\mathrm{d}s\bigg)^{\frac12} + \|\gamma-\tilde\gamma\|_{var}, \quad t\in[0,T].\end{aligned}$$

Taking square for both sides and applying Gronwall's inequality, we derive (3.2) for some constant $c > 0$.

□

3.2 Gradient estimate in the regular case

Theorem 3.2. *If* **(B)** *holds for b^μ in place of b uniformly in $\mu \in C^w([0,T];\mathscr{P}_D)$, and there exists a constant $K > 0$ such that*

$$|b_t(x,\mu) - b_t(x,\nu)| \le K\|\mu-\nu\|_{var}, \quad t\in[0,T],\ x\in\mathbb{R}^d,\ \mu,\nu\in\mathscr{P}_D. \tag{3.6}$$

Then there exists a constant $c > 0$ such that the map $P_t^{D*}: \mathscr{P}_D \to \mathscr{P}_D$ associated with (1.6) *satisfies*

$$\|P_t^{D*}\mu_0 - P_t^{D*}\nu_0\|_{var} \le \frac{c}{\sqrt t}\hat{\mathbb{W}}_1(\mu_0,\nu_0), \quad t\in(0,T],\ \mu_0,\nu_0\in\mathscr{P}_D. \tag{3.7}$$

Proof. For $\mu_0,\nu_0\in\mathscr{P}_D$, define

$$\mu_t := P_t^{D*}\mu_0, \quad \nu_t := P_t^{D*}\nu_0, \quad t\in[0,T],$$

and let $P_{s,t}^{D,\mu}$ (respectively $P_{s,t}^{D,\nu}$) be the Dirichlet semigroup $P_{s,t}^D$ defined in (2.3) for b^μ (respectively b^ν) in place of b. By Lemma 2.3, there exists a constant $c_1 > 0$ such that

$$\|\nabla P_{s,t}^{D,\mu}f\|_\infty \le \frac{c_1\|f\|_\infty}{\sqrt{t-s}}, \quad 0\le s<t\le T,\ f\in\mathscr{B}_b(\bar D), \tag{3.8}$$

and for any $t\in(0,T]$, $f\in\mathscr{B}_b(\bar D)$,

$$\partial_s P_{s,t}^{D,\mu}f = -\Big\{\frac12\mathrm{tr}(a_s\nabla^2) + b_s^\mu\cdot\nabla\Big\}P_{s,t}^{D,\mu}f, \quad s\in[0,t). \tag{3.9}$$

For any $\gamma \in \mathscr{P}_D$ and $t \in [0,T]$, let $P_t^{D,\mu*}\gamma \in \mathscr{P}_D$ be defined by

$$(P_t^{D,\mu*}\gamma)(A) := \mu\big(P_t^{D,\mu}1_A\big), \quad A \in \mathscr{B}_D,$$

where $P_t^{D,\mu} := P_{0,t}^{D,\mu}$. By the same reason deducing (1.5) from (1.4), (3.8) implies

$$\|P_t^{D,\mu*}\mu_0 - P_t^{D,\mu*}\nu_0\|_{var} \le \frac{c_2}{\sqrt{t}}\hat{\mathbb{W}}_1(\mu_0,\nu_0), \quad t \in (0,T] \tag{3.10}$$

for some constant $c_2 > 0$ uniformly in μ_0, ν_0. On the other hand, let X_s^ν solve the SDE

$$\mathrm{d}X_s^\nu = b_s^\nu(X_s^\nu)\mathrm{d}s + \sigma_s(X_s^\nu)\mathrm{d}W_s, \quad \mathscr{L}_{X_0^\nu}^D = \nu_0, \ s \in [0,T].$$

For any $f \in \mathscr{B}_b(\bar{D})$, by (3.9) and Itô's formula, we obtain

$$\mathrm{d}P_{s,t}^{D,\mu}f(X_s^\nu) = \big\langle b_s^\nu(X_s^\nu) - b_s^\mu(X_s^\nu), \nabla P_{s,t}^{D,\mu}f(X_s^\nu)\big\rangle \mathrm{d}s + \mathrm{d}M_s, \quad 0 \le s < t \le T,$$

where by (3.8) and $\|\sigma\|_\infty < \infty$,

$$\mathrm{d}M_s := \big\langle \nabla P_{s,t}^{D,\mu}f(X_s^\nu), \sigma_s(X_s^\nu)\mathrm{d}W_s\big\rangle, \quad 0 \le s < t \le T$$

is a martingale. Combining this with

$$P_{s\wedge\tau(X^\nu),t}^{D,\mu}f(X_{s\wedge\tau(X^\nu)}^\nu) = 0 \text{ if } t > s \ge \tau(X^\nu),$$

we derive the Duhamel formula

$$\begin{aligned}
\nu_0(P_t^{D,\nu}f) &= \lim_{s\uparrow t}\mathbb{E}\big[\mathbf{1}_{\{s<\tau(X^\nu)\}}P_{s,t}^{D,\mu}f(X_s^\nu)\big] = \lim_{s\uparrow t}\mathbb{E}\big[P_{s\wedge\tau(X^\nu),t}^{D,\mu}f(X_{s\wedge\tau(X^\nu)}^\nu)\big]\\
&= \mathbb{E}[P_{0,t}^{D,\mu}f(X_0^\nu)] + \mathbb{E}\int_0^{t\wedge\tau(X^\nu)} \big\langle b_s^\nu - b_s^\mu, \nabla P_{s,t}^{D,\mu}f\big\rangle(X_s^\nu)\mathrm{d}s\\
&= \nu_0(P_t^{D,\mu}f) + \int_0^t \nu_0\Big(P_s^{D,\nu}\big\langle b_s^\nu - b_s^\mu, \nabla P_{s,t}^{D,\mu}f\big\rangle\Big)\mathrm{d}s, \quad t \in [0,T], \ f \in \mathscr{B}_b(\bar{D}).
\end{aligned}$$

Combining this with (3.6), (3.8) and noting that

$$\mu_t := P_t^{D*}\mu_0 = P_t^{D,\mu*}\mu_0, \quad \nu_t := P_t^{D*}\nu_0 = P_t^{D,\nu*}\nu_0, \quad t \in [0,T],$$

we obtain

$$\|P_t^{D,\mu*}\nu_0 - \nu_t\|_{var} = \sup_{\|f\|_\infty \le 1}\big|\nu_0(P_t^{D,\mu}f - P_t^{D,\nu}f)\big| \le c_2K\int_0^t \frac{\|\mu_s - \nu_s\|_{var}}{\sqrt{t-s}}\mathrm{d}s, \quad t \in [0,T].$$

Combining this with (3.10) we arrive at

$$\begin{aligned}
\|\mu_t - \nu_t\|_{var} &\le \|\nu_t - P_t^{D,\mu*}\nu_0\|_{var} + \|P_t^{D,\mu*}\nu_0 - P_t^{D,\mu*}\mu_0\|_{var}\\
&\le c_2K\int_0^t \frac{\|\mu_s - \nu_s\|_{var}}{\sqrt{t-s}}\mathrm{d}s + \frac{c_2}{\sqrt{t}}\hat{\mathbb{W}}_1(\mu_0,\nu_0), \quad t \in (0,T].
\end{aligned} \tag{3.11}$$

It is trivial that

$$\phi_\lambda := \sup_{t\in[0,T]} \mathrm{e}^{-\lambda t}\sqrt{t}\|\mu_t-\nu_t\|_{var} < \infty, \quad \lambda\in(0,\infty).$$

By (3.11), when $\lambda>0$ is large enough,

$$\phi_\lambda \le c_2\hat{\mathbb{W}}_1(\mu_0,\nu_0) + c_2K\phi_\lambda \sup_{t\in(0,T]}\int_0^t \frac{\mathrm{e}^{-\lambda(t-s)}\sqrt{t}}{\sqrt{s(t-s)}}\mathrm{d}s \le c_2\hat{\mathbb{W}}_1(\mu_0,\nu_0)+\frac{1}{2}\phi_\lambda,$$

so that $\phi_\lambda \le 2c_2\hat{\mathbb{W}}_1(\mu_0,\nu_0)$. Consequently,

$$\|P_t^{D*}\mu_0 - P_t^{D*}\nu_0\|_{var} = \|\mu_t-\nu_t\|_{var} \le \frac{2c_2\mathrm{e}^{\lambda T}}{\sqrt{t}}\hat{\mathbb{W}}_1(\mu_0,\nu_0), \quad t\in(0,T],$$

so that (3.7) holds for $c := 2c_2\mathrm{e}^{\lambda T}$. □

3.3 Gradient estimate in the singular case

We make the following assumption.

(E) *Assume that* **(C)** *holds uniformly in* $\mu\in\mathscr{C}^w([0,T];\mathscr{P}_D)$, *and there exist* $k_3\in\mathbb{N}$ *and functions* $\{0\le\tilde f_i\in\tilde L_{\tilde q_i}^{\tilde p_i}\}$ *for some* $\{(\tilde p_i,\tilde q_i):\ 1\le i\le k_3\}\subset\mathscr{K}$, *such that*

$$|b_t(x,\mu)-b_t(x,\nu)| \le \hat{\mathbb{W}}_1(\mu,\nu)\sum_{i=1}^{k_3}\tilde f_i(t,x), \quad (t,x,\mu,\nu)\in[0,T]\times\bar D\times\mathscr{P}_D\times\mathscr{P}_D. \tag{3.12}$$

Theorem 3.3. *Assume* **(E)**. *Then there exists a constant* $c>0$ *such that the map* $P_t^{D*}:\mathscr{P}_D\to\mathscr{P}_D$ *associated with* (1.6) *satisfies*

$$\|P_t^{D*}\mu_0 - P_t^{D*}\nu_0\|_{var} \le \frac{c}{\sqrt{t}}\hat{\mathbb{W}}_1^\rho(\mu_0,\nu_0), \quad t\in(0,T],\ \mu_0,\nu_0\in\mathscr{P}_D. \tag{3.13}$$

Proof. For any $\mu_0,\nu_0\in\mathscr{P}_D$, let $\mu_t,\nu_t,P_t^{\mu*}$ and $P_t^{\nu*}$ be in the proof of Theorem 3.2.

By [14, Theorem 3.1], which is proved for $k_1=1$, but the proof also works for $k_1\in\mathbb{N}$ due to [6, Proposition 5.2], the SDE (1.6) is well-posed, and we can find a constant $c_0>0$ such that

$$\hat{\mathbb{W}}_1(\mu_t,\nu_t) \le c_0\hat{\mathbb{W}}_1(\mu_0,\nu_0) \le c_0\hat{\mathbb{W}}_1^\rho(\mu_0,\nu_0), \quad t\in[0,T]. \tag{3.14}$$

By Theorem 2.4 for b^μ in place of b, so that (2.23) holds for $P_t^{D,\mu}$ in place of P_t^D, hence there exists a constant $c_1>0$ uniformly in x such that

$$|\nabla P_t^{D,\mu}f(x)| \le \frac{c_1\|f\|_\infty}{\sqrt{t}(1\wedge\rho_\partial(x))}, \quad t\in(0,T],\ x\in D,\ f\in\mathscr{B}_b(\bar D).$$

By the same reason deducing (1.3) from (1.1), this implies

$$\|P_t^{D,\mu*}\mu_0 - P_t^{D,\mu*}\nu_0\|_{var} \le \frac{c_1}{\sqrt{t}}\hat{\mathbb{W}}_1^\rho(\mu_0,\nu_0), \quad t\in(0,T]. \tag{3.15}$$

Moreover, by (3.12), (3.14) and the Girsanov theorem as in the proof of (3.4), we find constants $c_2, c_3 > 0$ such that

$$\begin{aligned}\|P_t^{D,\mu*}\nu_0 - P_t^{D,\nu*}\nu_0\|_{var} \le & c_2\bigg(\sum_{i=1}^{k_3}\mathbb{E}_{\mathbb{Q}}\int_0^t \tilde{f}_i(s,X_s^\mu)^2\hat{\mathbb{W}}_1(\mu_s,\nu_s)^2\mathrm{d}s\bigg)^{\frac{1}{2}} \\ \le & c_3\hat{\mathbb{W}}^1(\mu_0,\nu_0)\bigg(\sum_{i=1}^{k_3}\mathbb{E}_{\mathbb{Q}}\int_0^t \tilde{f}_i(s,X_s^\mu)^2\mathrm{d}s\bigg)^{\frac{1}{2}}, \quad t\in[0,T],\end{aligned} \tag{3.16}$$

where X_s^μ solves (3.3) for $\mathscr{L}_{X_0^\mu}^D = \nu_0$, and $\mathrm{d}\mathbb{Q} = R_T\mathrm{d}\mathbb{P}$ as defined before (3.4). Noting that the law of X^μ under $\mathbb{Q}$ coincides with that of X^ν under $\mathbb{P}$, where X_s^ν solves (3.3) with $\mathscr{L}_{X_0^\nu}^D = \nu_0$, we have

$$\mathbb{E}_{\mathbb{Q}}\int_0^t \tilde{f}_i(s,X_s^\mu)^2\mathrm{d}s = \mathbb{E}\int_0^t \tilde{f}_i(s,X_s^\nu)^2\mathrm{d}s,$$

so that by Krylov's estimate, see [11, Lemma 1.2.3], we find a constant $c_4 > 0$ such that

$$\sum_{i=1}^{k}\mathbb{E}_{\mathbb{Q}}\int_0^t \tilde{f}_i(s,X_s^\mu)^2\mathrm{d}s \le c_4.$$

Combining this with (3.16), we find a constant $c_5 > 0$ such that

$$\|P_t^{D,\mu*}\nu_0 - P_t^{D,\nu*}\nu_0\|_{var} \le c_5\hat{\mathbb{W}}_1(\mu_0,\nu_0), \quad t\in[0,T].$$

This together with (3.15) and the fact that $\hat{\mathbb{W}}_1 \le \hat{\mathbb{W}}_1^\rho$ proves (3.13) for some constant $c > 0$, since

$$\begin{aligned}&\|P_t^{D*}\mu_0 - P_t^{D*}\nu_0\|_{var} = \|\mu_t-\nu_t\|_{var} = \|P_t^{D,\mu*}\mu_0 - P_t^{D,\nu*}\nu_0\|_{var} \\ &\le \|P_t^{D,\mu*}\mu_0 - P_t^{D,\mu*}\nu_0\|_{var} + \|P_t^{D,\mu*}\nu_0 - P_t^{D,\nu*}\nu_0\|_{var}, \quad t\in[0,T].\end{aligned}$$

□

Acknowledgement. The authors would like to thank the referees for corrections and valuable comments.

Declarations

Funding: This work is supported in part by the National Key R&D Program of China (No. 2022YFA1006000), and State Key Laboratory of Synthetic Biology, Tianjin University.

Conflict of interest: The authors declare that they have no known competing interests.

References

[1] L. Angiuli, L. Lorenzi, *Non autonomous parabolic problems with unbounded coefficients in unbounded domains,* Adv. Differential Equations, 20(2015), 1067–1118.

[2] J. Bao, P. Ren, F.-Y. Wang, *Bismut formula for Lions derivative of distribution-path dependent SDEs,* J. Diff. Equat. 282(2021), 285–329.

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

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

[5] S. Fornaro, G. Metafune, E. Priola, *Gradient estimates for Dirichlet parabolic problems in unbounded domains,* J. Diff. Equat. 205(2)(2004), 329–353.

[6] X. Huang, P. Ren, F.-Y. Wang, *Distribution dependent SDEs with singular interactions: well-posedness and regularity,* arXiv:2505.19787.

[7] X. Huang, Y. Song, F.-Y. Wang, *Bismut formula for intrinsic/Lions derivatives of distribution dependent SDEs with singular coefficients,* Disc. Cont. Dyn. Syst. 42 (2022), 4597–4614.

[8] L. Lorenzi, *Analytical methods for Kolmogorov equations,* Chapman and Hall/CRC, 2016.

[9] P. Ren, *Extrinsic derivative formula for distribution dependent SDEs,* Bernoulli, 31(3) (2025), 2508–2524.

[10] P. Ren, F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications,* J. Diff. Equat. 267(8)(2019), 4745–4777.

[11] P. Ren, F.-Y. Wang, *Distribution Dependent Stochastic Differential Equations,* World Scientific, 2025, Singapore.

[12] A. Thalmaier, *On the differentiation of heat semigroups and Poisson integrals,* Stochastics: An International Journal of Probability and Stochastic Processes, 61(1997), 297–321.

[13] A. Thalmaier, F.-Y. Wang, *Gradient estimates for harmonic functions on regular domains in Riemannian manifolds,* J. Funct. Anal. 155(1998), 109–124.

[14] F.-Y. Wang, *Killed distribution dependent SDE for nonlinear Dirichlet problem,* Disc. Cont. Dyn. Syst. S. 16(2023), 797–818.

[15] F.-Y. Wang, *Derivative formula for singular McKean-Vlasov SDEs,* Commun. Pure Appl. Anal. 22(2023), 1866–1898.

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

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