The Annals of Probability
2014, Vol. 42, No. 3, 994–1019
DOI: 10.1214/13-AOP875

INTEGRATION BY PARTS FORMULA AND SHIFT HARNACK INEQUALITY FOR STOCHASTIC EQUATIONS

BY FENG-YU WANG[1]

Beijing Normal University and Swansea University

A new coupling argument is introduced to establish Driver's integration by parts formula and shift Harnack inequality. Unlike known coupling methods where two marginal processes with different starting points are constructed to move together as soon as possible, for the new-type coupling the two marginal processes start from the same point but their difference is aimed to reach a fixed quantity at a given time. Besides the integration by parts formula, the new coupling method is also efficient to imply the shift Harnack inequality. Differently from known Harnack inequalities where the values of a reference function at different points are compared, in the shift Harnack inequality the reference function, rather than the initial point, is shifted. A number of applications of the integration by parts and shift Harnack inequality are presented. The general results are illustrated by some concrete models including the stochastic Hamiltonian system where the associated diffusion process can be highly degenerate, delayed SDEs and semi-linear SPDEs.

1. Introduction. In stochastic analysis for diffusion processes, the Bismut formula [5] (also known as Bismut–Elworthy–Li formula due to [8]) and the integration by parts formula are two fundamental tools. Let, for instance, $X(t)$ be the (nonexplosive) diffusion process generated by an elliptic differential operator on a Riemannian manifold M, and let P_t be the associated Markov semigroup. For $x \in M$ and $U \in T_x M$, the Bismut formula is of type

$$\nabla_U P_t f(x) = \mathbf{E}\{f(X^x(t))M^x(t)\}, \qquad f \in \mathcal{B}_b(M), t > 0, \tag{1.1}$$

where $X^x(t)$ is the diffusion process starting at point x, $M^x(t)$ is a random variable independent of f and ∇_U is the directional derivative along U. When the curvature of the diffusion operator is bounded below, this formula is available with $M^x(t)$ explicitly given by U and the curvature operator. There exist a number of applications of this formula, in particular, letting $p_t(x, y)$ be the density (or heat kernel) of P_t w.r.t. a nice reference measure μ, we have, formally,

$$\nabla_U \log p_t(\cdot, y)(x) = \mathbf{E}(M^x(t) \mid X^x(t) = y).$$

Received March 2012; revised February 2013.

[1]Supported in part by Lab. Math. Com. Sys., NNSFC (11131003), SRFDP and the Fundamental Research Funds for the Central Universities.

MSC2010 subject classifications. Primary 60K35, 60H10; secondary 47G20.

Key words and phrases. Integration by parts formula, shift Harnack inequality, shift log-Harnack inequality, coupling, Malliavin calculus.

From (1.1) one may also derive gradient-entropy estimates of P_t and thus, the following Harnack inequality introduced in [16] (see [2, 10]):

$$|P_t f|^p(x) \le P_t|f|^p(y)\mathrm{e}^{C_p(t,x,y)}, \qquad t>0, p>1, x,y\in M, f\in B_b(M), \tag{1.2}$$

where $C_p(t,x,y)$ is determined by moments of $M^{\cdot}(t)$ and thus, independent of f. This type of Harnack inequality is a powerful tool in the study of contractivity properties, functional inequalities and heat kernel estimates; see, for example, [19] and references within.

On the other hand, to characterize the derivative of $p_t(x,y)$ in y, which is essentially different from that in x when P_t is not symmetric w.r.t. μ, we need to establish the following integration by parts formula (see [7]):

$$P_t(\nabla_U f)(x) = \mathbf{E}\{f(X^x(t))N^x(t)\}, \qquad f\in C_0^1(M), t>0, x\in M \tag{1.3}$$

for a smooth vector field U and some random variable $N^x(t)$. Combining this formula with (1.1), we are able to estimate the commutator $\nabla P_t - P_t\nabla$ which is important in the study of flow properties; see, for example, [9]. Similar to (1.1), inequality (1.3) can be used to derive a formula for $\nabla_U \log p_t(x,\cdot)(y)$ and the shift Harnack inequality of type

$$|P_t f|^p(x) \le P_t(|f|^p\circ \exp[U])(x)\mathrm{e}^{C_p(t,x,y)}, \qquad t>0, p>1, x,y\in M, f\in B_b(M), \tag{1.4}$$

where $\exp_x : T_xM\to M, x\in M$, is the exponential map on the Riemannian manifold. Differently from usual Harnack inequalities like (1.2), in (1.4) the reference function f, rather than the initial point, is shifted. This inequality will lead to different heat kernel estimates from known ones implied by (1.2).

Before moving on, let us make a brief comment concerning the study of these two formulas. The Bismut formula (1.1) has been widely studied using both Malliavin calculus and coupling argument; cf. [18, 20, 22] and references within. Although (1.3) also has strong potential of applications, it is, however, much less known in the literature due to the lack of efficient tools. To see that (1.3) is harder to derive than (1.1), let us come back to [7] where an explicit version of (1.3) is established for the Brownian motion on a compact Riemannian manifold. Unlike the Bismut formula which only relies on the Ricci curvature, Driver's integration by parts formula involves both the Ricci curvature and its derivatives. Therefore, one can imagine that in general (1.3) is more complicated (and hence harder to derive) than (1.1).

To establish the integration by parts formula and the corresponding shift Harnack inequality in a general framework, in this paper we propose a new coupling argument. In contrast to usual coupling arguments where two marginal processes start from different points and meet at some time (called the coupling time), for the new-type coupling the marginal processes start from the same point, but their difference reaches a fixed quantity at a given time.

In the next section, we will introduce some general results and applications on the integration by parts formula and the shift Harnack inequality using the new coupling method. The general result obtained in Section 2 will be then applied in Section 3 to a class of degenerate diffusion processes, in Section 4 to delayed SDEs and in Section 5 to semi-linear SPDEs.

We remark that the model considered in Section 3 goes back to the stochastic Hamiltonian system, for which the Bismut formula and the Harnack inequalities have been investigated in [10, 20, 22] by using both coupling and Malliavin calculus. As will be shown in Section 2.1 with a simple example of this model, for the study of the integration by parts formula and the shift Harnack inequalities, the Malliavin calculus can be less efficient than the new coupling argument.

2. Some general results.

In Section 2.1 we first recall the argument of coupling by change of measure introduced in [1, 18] for the Harnack inequality and the Bismut formula, and then explain how can we modify the coupling in order to derive the integration by parts formula and the shift Harnack inequality, and introduce the Malliavin calculus for the study of the integration by parts formula. In the second subsection we present some applications of the integration by parts formula and the shift Harnack inequalities to estimates of the heat kernel and its derivatives.

For a measurable space $(E,\mathcal{B})$, let $\mathcal{B}_b(E)$ be the class of all bounded measurable functions on E, and $\mathcal{B}_b^+(E)$ the set of all nonnegative elements in $\mathcal{B}_b(E)$. When E is a topology space, we always take $\mathcal{B}$ to be the Borel σ-field, and let $C_b(E)$ [resp., $C_0(E)$] be the set of all bounded (compactly supported) continuous functions on E. If, moreover, E is equipped with a differential structure, for any $i\ge 1$ let $C_b^i(E)$ be the set of all elements in $C_b(E)$ with bounded continuous derivatives up to order i, and let $C_0^i(E)=C_0(E)\cap C_b^i(E)$. Finally, a contraction linear operator P on $\mathcal{B}_b(E)$ is called a Markov operator if it is positivity-preserving with $P1=1$.

2.1. *Integration by parts formula and shift Harnack inequality.*

DEFINITION 2.1. Let μ and ν be two probability measures on a measurable space $(E,\mathcal{B})$, and let X,Y be two E-valued random variables w.r.t. a probability space $(\Omega,\mathcal{F},\mathbf{P})$.

(i) If the distribution of X is μ, while under another probability measure $\mathbf{Q}$ on $(\Omega,\mathcal{F})$ the distribution of Y is ν, we call (X,Y) a coupling by change of measure for μ and ν with changed probability $\mathbf{Q}$.

(ii) If μ and ν are distributions of two stochastic processes with path space E, a coupling by change of measure for μ and ν is also called a coupling by change of measure for these processes. In this case X and Y are called the marginal processes of the coupling.

Now, for fixed $T > 0$, consider the path space $E^T := E^{[0,T]}$ for some $T > 0$ equipped with the product σ-field $\mathcal{B}^T := \mathcal{B}^{[0,T]}$. Let $\{P^x(A) : x \in E, A \in \mathcal{B}^T\}$ be a transition probability such that $P^x(\{\gamma \in E^T : \gamma(0) = x\}) = 1, x \in E$. For any $t \in [0, T]$, let $P_t(x, \cdot) = P^x(\{\gamma(t) \in \cdot\})$ be the marginal distribution of P^x at time t. Then

$$P_t f(x) := \int_E f(y) P_t(x, \mathrm{d}y), \qquad f \in \mathcal{B}_b(E), x \in E$$

gives rise to a family of Markov operators $(P_t)_{t\in[0,T]}$ on $\mathcal{B}_b(E)$ with $P_0 = I$.

In order to establish the Harnack inequality, for any two different points $x, y \in E$, one constructs a coupling by change of measure (X, Y) for P^x and P^y with changed probability $\mathbf{Q} = R\mathbf{P}$ such that $X(T) = Y(T)$. Then

$$\begin{aligned} |P_T f(y)|^p &= |\mathbf{E}_{\mathbf{Q}} f(Y(T))|^p \\ &= |\mathbf{E}\{R f(X(T))\}|^p \\ &\le (\mathbf{E}|f|^p(X(T)))(\mathbf{E}R^{p/(p-1)})^{p-1} \\ &= (P_T|f|^p(x))(\mathbf{E}R^{p/(p-1)})^{p-1}. \end{aligned}$$

This implies a Harnack inequality of type (1.2) if $\mathbf{E}R^{p/(p-1)} < \infty$.

To establish the Bismut formula, let, for example, E be a Banach space, and $x, e \in E$. One constructs a family of couplings by change of measure (X^ε, X) for $P^{x+\varepsilon e}$ and P^x with changed probability $\mathbf{Q}_\varepsilon := R_\varepsilon \mathbf{P}$ such that $X^\varepsilon(T) = X(T), \varepsilon \in [0, 1]$. Then, if $N^x(T) := \frac{\mathrm{d}}{\mathrm{d}\varepsilon} R_\varepsilon|_{\varepsilon=0}$ exists in $L^1(\mathbf{P})$, for any $f \in \mathcal{B}_b(E)$, we obtain

$$\begin{aligned} \nabla_e P_T f(x) &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbf{E}\{R_\varepsilon f(X^\varepsilon(T))\}\Big|_{\varepsilon=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbf{E}\{R_\varepsilon f(X(T))\}\Big|_{\varepsilon=0} \\ &= \mathbf{E}\{f(X(T)) N^x(T)\}. \end{aligned}$$

Therefore, the Bismut formula (1.1) is derived.

On the other hand, for the integration by parts formula and shift Harnack inequality we need to construct couplings with marginal processes starting from the same point but their "difference" equals to a fixed value at time T. For simplicity, below we only consider E being a Banach space. To extend the result to nonlinear spaces like Riemannian manifolds, one would need to make proper modifications using the geometric structure in place of the linear structure.

THEOREM 2.1. *Let E be a Banach space and $x, e \in E$ and $T > 0$ be fixed.*

(1) *For any coupling by change of measure* (X, Y) *for* P^x *and* P^x *with changed probability* $\mathbf{Q} = R\mathbf{P}$ *such that* $Y(T) = X(T) + e$, *there holds the shift Harnack inequality*

$$\big|P_T f(x)\big|^p \le P_T\big\{|f|^p(e+\cdot)\big\}(x)\big(\mathbf{E}R^{p/(p-1)}\big)^{p-1}, \qquad f \in \mathcal{B}_b(E)$$

and the shift log-Harnack inequality

$$P_T \log f(x) \le \log P_T\big\{f(e+\cdot)\big\}(x) + \mathbf{E}(R\log R), \qquad f \in \mathcal{B}_b(E), f > 0.$$

(2) *Let* $(X, X^\varepsilon), \varepsilon \in [0,1]$, *be a family of couplings by change of measure for* P^x *and* P^x *with changed probability* $\mathbf{Q}_\varepsilon = R_\varepsilon \mathbf{P}$ *such that*

$$X^\varepsilon(T) = X(T) + \varepsilon e, \qquad \varepsilon \in (0,1].$$

If $R_0 = 1$ *and* $N(T) := -\frac{\mathrm{d}}{\mathrm{d}\varepsilon} R_\varepsilon|_{\varepsilon=0}$ *exists in* $L^1(\mathbf{P})$, *then*

$$P_T(\nabla_e f)(x) = \mathbf{E}\big\{f\big(X(T)\big)N(T)\big\}, \qquad f, \nabla_e f \in \mathcal{B}_b(E). \tag{2.1}$$

PROOF. The proof is similar to that introduced above for the Harnack inequality and the Bismut formula.

(1) Note that $P_T f(x) = \mathbf{E}\{Rf(Y(T))\} = \mathbf{E}\{Rf(X(T)+e)\}$. We have

$$\begin{aligned}\big|P_T f(x)\big|^p &\le \big(\mathbf{E}|f|^p\big(X(T)+e\big)\big)\big(\mathbf{E}R^{p/(p-1)}\big)^{p-1}\\ &= P_T\big\{|f|^p(e+\cdot)\big\}(x)\big(\mathbf{E}R^{p/(p-1)}\big)^{p-1}.\end{aligned}$$

Next, by the Young inequality (see [2], Lemma 2.4), for positive f we have

$$\begin{aligned}P_T \log f(x) &= \mathbf{E}\big\{R\log f\big(X(T)+e\big)\big\}\\ &\le \log \mathbf{E} f\big(X(T)+e\big) + \mathbf{E}(R\log R)\\ &= \log P_T\big\{f(e+\cdot)\big\}(x) + \mathbf{E}(R\log R).\end{aligned}$$

(2) Noting that $P_T f(x) = \mathbf{E}\{R_\varepsilon f(X^\varepsilon(T))\} = \mathbf{E}\{R_\varepsilon f(X(T)+\varepsilon e)\}$, we obtain

$$0 = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathbf{E}\big\{R_\varepsilon f\big(X(T)+\varepsilon e\big)\big\}\Big|_{\varepsilon=0} = P_T(\nabla_e f)(x) - \mathbf{E}\big\{f\big(X(T)\big)N(T)\big\},$$

provided $R_0 = 1$ and $N(T) := -\frac{\mathrm{d}}{\mathrm{d}\varepsilon} R_\varepsilon|_{\varepsilon=0}$ exists in $L^1(\mathbf{P})$. $\square$

From Theorem 2.1 and its proof we see that the machinery of the new coupling argument is very clear. So, in applications the key point of the study lies in the construction of new type couplings.

Next, we explain how one can establish the integration by parts formula using Malliavin calculus. Let, for example, $W := (W(t))_{t\ge 0}$ be the cylindrical Brownian motion on an Hilbert space $(H, \langle\cdot,\cdot\rangle, |\cdot|)$ w.r.t. a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with natural filtration $\{\mathcal{F}_t\}_{t\ge 0}$. Let

$$H^1 := \left\{h \in C\big([0,T]; H\big) : \|h\|_{H^1}^2 := \int_0^T \big|h'(s)\big|^2\,\mathrm{d}s < \infty\right\}$$

be the Cameron–Martin space. For a measurable functional of W, denoted by $F(W)$, such that $\mathbf{E}F(W)^2 < \infty$ and

$$H^1 \ni h \mapsto D_h F(W) := \lim_{\varepsilon \downarrow 0} \frac{F(W + \varepsilon h) - F(W)}{\varepsilon}$$

gives rise to a bounded linear operator. Then we write $F(W) \in \mathcal{D}(D)$ and call $DF(W)$ the Malliavin gradient of $F(W)$. It is well known that $(D, \mathcal{D}(D))$ is a densely defined closed operator on $L^2(\Omega, \mathcal{F}_T; \mathbf{P})$; see, for example, [12], Section 1.3. Let $(D^*, \mathcal{D}(D^*))$ be its adjoint operator, which is also called the divergence operator.

THEOREM 2.2. *Let H, W, D and D^* be introduced above. Let $e \in H$ and $X \in \mathcal{D}(D)$. If there exists $h \in \mathcal{D}(D^*)$ such that $D_h X = e$, then*

$$\mathbf{E}(\nabla_e f)(X) = \mathbf{E}\{f(X) D^* h\}, \qquad f \in C_b^1(H).$$

PROOF. Since $D_h X = e$, we have

$$\mathbf{E}(\nabla_e f)(X) = \mathbf{E}(\nabla_{D_h X} f)(X) = \mathbf{E}\{D_h f(X)\} = \mathbf{E}\{f(X) D^* h\}. \qquad \square$$

Finally, as the integration by parts formula (2.1) and by the Young inequality (see [2], Lemma 2.4) imply the derivative-entropy inequality

$$\begin{aligned} |P_T(\nabla_e f)| \le \delta\{P_T(f \log f) - (P_T f)\log P_T f\} \\ + \delta \log \mathbf{E}\left\{\exp\left[\frac{|N(T)|}{\delta}\right]\right\} P_T f, \qquad \delta > 0 \end{aligned}$$

and the L^2-derivative inequality

$$|P_T(\nabla_e f)|^2 \le (\mathbf{E}N(T)^2) P_T f^2,$$

according to the following result it also implies shift Harnack inequalities.

PROPOSITION 2.3. *Let P be a Markov operator on $\mathcal{B}_b(E)$ for some Banach space E. Let $e \in E$.*

(1) *Let $\delta_e \in (0, 1)$ and $\beta_e \in C((\delta_e, \infty) \times E; [0, \infty))$. Then*

$$|P(\nabla_e f)| \le \delta\{P(f \log f) - (Pf)\log Pf\} + \beta_e(\delta, \cdot) Pf, \qquad \delta \ge \delta_e \tag{2.2}$$

holds for any positive $f \in C_b^1(E)$ if and only if

$$\begin{aligned} (Pf)^p &\le (P\{f^p(re + \cdot)\}) \\ &\quad \times \exp\left[\int_0^1 \frac{pr}{1 + (p-1)s} \beta_e\left(\frac{p-1}{r + r(p-1)s}, \cdot + sre\right) \mathrm{d}s\right] \end{aligned} \tag{2.3}$$

holds for any positive $f \in \mathcal{B}_b(E)$, $r \in (0, \frac{1}{\delta_e})$ and $p \ge \frac{1}{1 - r\delta_e}$.

(2) *Let $C \geq 0$ be a constant. Then*

$$|P(\nabla_e f)|^2 \leq CPf^2, \qquad f \in C_b^1(E), f \geq 0 \tag{2.4}$$

is equivalent to

$$Pf \leq P\{f(\alpha e + \cdot)\} + |\alpha|\sqrt{CPf^2}, \qquad \alpha \in \mathcal{R}, f \in \mathcal{B}_b^+(E). \tag{2.5}$$

PROOF. The proof of (1) is similar to that of [10], Proposition 4.1, while (2) is comparable to [17], Proposition 1.3.

(1) Let $\beta(s) = 1 + (p-1)s$, $s \in [0, 1]$. By the monotone class theorem, it suffices to prove for $f \in C_b^1(E)$. Since $\frac{p-1}{r\beta(s)} \geq \delta_e$ for $p \geq \frac{1}{1-r\delta_e}$, it follows from (2.2) that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s} \log\bigl(P\{f^{\beta(s)}(sre + \cdot)\}(x)\bigr)^{p/\beta(s)} \\
&\quad = \frac{1}{\beta(s)^2 P\{f^{\beta(s)}(sre + \cdot)\}(x)} \\
&\qquad \times \bigl(p(p-1)\bigl[P\{(f^{\beta(s)} \log f^{\beta(s)})(sre + \cdot)\} \\
&\qquad\qquad - \bigl(P\{f^{\beta(s)}(sre + \cdot)\}\bigr) \log P\{f^{\beta(s)}(sre + \cdot)\}\bigr] \\
&\qquad\qquad\qquad + prP\{\nabla_e f^{\beta(s)}(sre + \cdot)\}\bigr)(x) \\
&\quad \geq -\frac{rp}{\beta(s)} \beta_e\biggl(\frac{p-1}{r\beta(s)}, x + sre\biggr), \qquad s \in [0, 1].
\end{aligned}$$

Taking the integral over $[0, 1]$ w.r.t. $\mathrm{d}s$ we prove (2.3).

Next, let $z, e \in E$ be fixed, and assume that $P(\nabla_e f)(z) \geq 0$ (otherwise, simply use $-e$ to replace e). Then (2.3) with $p = 1 + \delta_e r$ implies that

$$\begin{aligned}
&\delta\{(Pf) \log Pf\}(z) + |P(\nabla_e f)|(z) \\
&\quad = \limsup_{r \to 0} \frac{(P\{f(re + \cdot)\})^{1+\delta r}(z) - Pf(z)}{r} \\
&\quad \leq \limsup_{r \to 0} \frac{1}{r} \biggl\{(Pf^{1+\delta r})(z) \\
&\qquad\qquad \times \exp\biggl[\int_0^1 \frac{(1+\delta r)r}{1+\delta rs} \beta_e\biggl(\frac{\delta}{1+\delta rs}, \gamma(r)\biggr)\,\mathrm{d}r\biggr] - Pf(z)\biggr\} \\
&\quad = \delta P(f \log f)(z) + \beta_e(\delta) Pf(z).
\end{aligned}$$

Therefore, (2.2) holds.

(2) Let $r > 0$. For nonnegative $f \in C_b^1(E)$, (2.4) implies that

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}s} P\left\{\frac{f}{1+srf}(\alpha(1-s)e+\cdot)\right\} \\
&\quad = -P\left\{\frac{rf^2}{1+srf}(\alpha(1-s)e+\cdot)\right\} - \alpha P\left\{\nabla_e\left(\frac{f}{1+srf}\right)(\alpha(1-s)e+\cdot)\right\} \\
&\quad \le -rP\left\{\frac{f^2}{1+srf}(\alpha(1-s)e+\cdot)\right\} \\
&\qquad + |\alpha|\left(CP\left\{\frac{f^2}{(1+srf)^2}(\alpha(1-s)e+\cdot)\right\}\right)^{1/2} \\
&\quad \le \frac{\alpha^2 C}{4r}.
\end{aligned}
$$

Noting that

$$\frac{f}{1+rf} = f - \frac{rf^2}{1+rf} \ge f - rf^2,$$

we obtain

$$Pf \le P\{f(\alpha e+\cdot)\} + rPf^2 + \frac{\alpha^2 C}{4r}, \qquad r > 0.$$

Minimizing the right-hand side in $r > 0$, we prove (2.5).

On the other hand, let $x \in E$. Without loss of generality we assume that $P(\nabla_e f)(x) \le 0$, otherwise it suffices to replace e by $-e$. Then (2.5) implies that

$$\left|P(\nabla_e f)(x)\right| = \lim_{\alpha\downarrow 0} \frac{Pf(x) - P\{f(\alpha e+\cdot)\}(x)}{\alpha} \le \sqrt{CPf^2(x)}.$$

Therefore, (2.4) holds. $\square$

To conclude this section, we would like to compare the new coupling argument with known coupling arguments and the Malliavin calculus, from which we see that the study of the integration by parts formula and the shift Harnack inequality is, in general, more difficult than that of the Bismut formula and the Harnack inequality.

First, when a strong Markov process is concerned, for a usual coupling $(X(t), Y(t))$ one may ask that the two marginal processes move together after the coupling time, so that to ensure $X(T) = Y(T)$, one only has to confirm that the coupling time is not larger than the given time T. But for the new coupling argument, we have to prove that at time T, the difference of the marginal processes equals to a fixed quantity, which cannot be ensured, even if the difference already reached this quantity at a (random) time before T. From this we see that construction of a new-type coupling is, in general, more difficult than that of a usual coupling.

Second, it is well known that the Malliavin calculus is a very efficient tool to establish Bismut-type formulas. To see the difficulty for deriving the integration by parts formula using Malliavin calculus, we look at a simple example of the model considered in Section 3, that is, $(X(t), Y(t))$ is the solution to the following degenerate stochastic equation on $\mathcal{R}^2$:

$$\begin{cases} \mathrm{d}X(t) = Y(t)\,\mathrm{d}t, \\ \mathrm{d}Y(t) = \mathrm{d}W(t) + Z\big(X(t), Y(t)\big)\,\mathrm{d}t, \end{cases} \tag{2.6}$$

where $W(t)$ is the one-dimensional Brownian motion and $Z \in C_b^1(\mathcal{R}^2)$. For this model the Bismut formula and Harnack inequalities can be easily derived from both the coupling method and Malliavin calculus; see [10, 20, 22]. We now explain how can one establish the integration by parts formula using Malliavin calculus. For fixed $T > 0$ and, for example, $e = (0, 1)$, to derive the integration by parts formula for the derivative along e using Theorem 2.2, one needs to find $h \in \mathcal{D}(D^*)$ such that

$$D_h\big(X(T), Y(T)\big) = e. \tag{2.7}$$

To search for such an element h, we note that (2.6) implies

$$\mathrm{d}\big(D_h X(t), D_h Y(t)\big) = \big(0, h'(t)\big)\,\mathrm{d}t + G(t)\begin{pmatrix} D_h X(t) \\ D_h Y(t) \end{pmatrix}\mathrm{d}t$$

and

$$D_h X(0) = D_h Y(0) = 0,$$

where

$$G(t) := \begin{pmatrix} 0 & 1 \\ Z'\big(\cdot, Y(t)\big)\big(X(t)\big) & Z'\big(X(t), \cdot\big)\big(Y(t)\big) \end{pmatrix}.$$

Then, (2.7) is equivalent to

$$\int_0^T \mathrm{e}^{\int_t^T G(s)\,\mathrm{d}s}\begin{pmatrix} 0 \\ h'(t) \end{pmatrix}\mathrm{d}t = (0, 1).$$

It is, however, very hard to solve h from this equation for general $Z \in C_b^1(\mathcal{R}^2)$. On the other hand, we will see in Section 3 that the coupling argument we proposed above is much more convenient for deriving the integration parts formula for this example.

2.2. *Applications*. We first consider $E = \mathcal{R}^d$ for some $d \geq 1$, and to estimate the density w.r.t. the Lebesgue measure for distributions and Markov operators using integration by parts formulas and shift Harnack inequalities.

THEOREM 2.4. *Let X be a random variable on $\mathcal{R}^d$ such that for some $N \in L^2(\Omega \to \mathcal{R}^d; \mathbf{P})$*

$$\mathbf{E}(\nabla f)(X) = \mathbf{E}\{f(X)N\}, \qquad f \in C_b^1(\mathcal{R}^d). \tag{2.8}$$

(1) *The distribution $\mathbf{P}_X$ of X has a density ρ w.r.t. the Lebesgue measure, which satisfies*

$$\nabla \log \rho(x) = -\mathbf{E}(N \mid X = x), \qquad \mathbf{P}_X\text{-}a.s. \tag{2.9}$$

Consequently, for any $e \in \mathcal{R}^d$ and any convex positive function H,

$$\int_{\mathcal{R}^d} \{H(|\nabla_e \log \rho|)\rho\}(x)\,\mathrm{d}x \leq \mathbf{E}H(|\langle e, N\rangle|).$$

(2) *For any $U \in C_0^1(\mathcal{R}^d; \mathcal{R}^d)$,*

$$\mathbf{E}(\nabla_U f)(X) = \mathbf{E}\{f(X)(\langle U(X), N\rangle - (\operatorname{div} U)(X))\}, \qquad f \in C^1(\mathcal{R}^d).$$

PROOF. (1) We first observe that if $\mathbf{P}_X$ has density ρ, then for any $f \in C_0^1(\mathcal{R}^d)$,

$$\begin{aligned}\int_{\mathcal{R}^d} \{\rho(x)\nabla f(x)\}\,\mathrm{d}x &= \mathbf{E}(\nabla f)(X) \\ &= \mathbf{E}\{f(X)\mathbf{E}(N \mid X)\} \\ &= \int_{\mathcal{R}^d} \{f(x)\mathbf{E}(N \mid X = x)\}\mathbf{P}_X(\mathrm{d}x).\end{aligned}$$

This implies (2.9). To prove the existence of ρ, let ρ_n be the distribution density function of $X_n := X + \frac{\zeta}{n}, n \geq 1$, where ζ is the standard Gaussian random variable on $\mathcal{R}^d$ independent of X and N. It follows from (2.8) that

$$\mathbf{E}(\nabla f)(X_n) = \mathbf{E}\{\nabla f(\zeta/n + \cdot)\}(X) = \mathbf{E}\{f(X_n)N\}.$$

Then

$$4\int_{\mathcal{R}^d} |\nabla\sqrt{\rho_n}|^2(x)\,\mathrm{d}x = \mathbf{E}|\nabla \rho_n|^2(X_n) \leq \mathbf{E}N^2 < \infty.$$

So, the sequence $\{\sqrt{\rho_n}\}_{n\geq 1}$ is bounded in $W^{2,1}(\mathcal{R}^d; \mathrm{d}x)$. Thus, up to a subsequence, $\sqrt{\rho_n} \to \sqrt{\rho}$ in $L^2_{\mathrm{loc}}(\mathrm{d}x)$ for some nonnegative function ρ. On the other hand, we have $\rho_n(x)\,\mathrm{d}x \to \mathbf{P}_X(\mathrm{d}x)$ weakly. Therefore, $\mathbf{P}_X(\mathrm{d}x) = \rho(x)\,\mathrm{d}x$.

(2) As for the second assertion, noting that for $U=\sum_{i=1}^d U_i\partial_i$ one has

$$\nabla_U f=\sum_{i=1}^d \partial_i(U_i f)-f\,\mathrm{div}\,U,$$

it follows from (2.8) that

$$\begin{aligned}\mathbf{E}(\nabla_U f)(X)&=\sum_{i=1}^d \mathbf{E}\{\partial_i(U_i f)(X)\}-\mathbf{E}\{f\,\mathrm{div}\,U\}(X)\\&=\sum_{i=1}^d \mathbf{E}\{(U_i f)(X)N_i\}-\mathbf{E}\{f\,\mathrm{div}\,U\}\\&=\mathbf{E}\{f(X)(\langle U(X),N\rangle-(\mathrm{div}\,U)(X))\}.\end{aligned}$$ □

Next, we consider applications of a general version of the shift Harnack. Let $P(x,\mathrm{d}y)$ be a transition probability on a Banach space E. Let

$$Pf(x)=\int_{\mathcal{R}^d} f(y)P(x,\mathrm{d}y),\qquad f\in\mathcal{B}_b(\mathcal{R}^d)$$

be the associated Markov operator. Let $\Phi:[0,\infty)\to[0,\infty)$ be a strictly increasing and convex continuous function. Consider the shift Harnack inequality

$$\Phi(Pf(x))\le P\{\Phi\circ f(e+\cdot)\}(x)\mathrm{e}^{C_\Phi(x,e)},\qquad f\in\mathcal{B}_b^+(E) \tag{2.10}$$

for some $x,e\in E$ and constant $C_\Phi(x,e)\ge 0$. Obviously, if $\Phi(r)=r^p$ for some $p>1$, then this inequality reduces to the shift Harnack inequality with power p, while when $\Phi(r)=\mathrm{e}^r$, it becomes the log shift Harnack inequality.

THEOREM 2.5. *Let P be given above and satisfy* (2.10) *for all $x,e\in E:=\mathcal{R}^d$ and some nonnegative measurable function C_Φ on $\mathcal{R}^d\times\mathcal{R}^d$. Then*

$$\sup_{f\in\mathcal{B}_b^+(\mathcal{R}^d),\int_{\mathcal{R}^d}\Phi\circ f(x)\,\mathrm{d}x\le 1}\Phi(Pf)(x)\le\frac{1}{\int_{\mathcal{R}^d}\mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e},\qquad x\in\mathcal{R}^d. \tag{2.11}$$

Consequently:

(1) *If $\Phi(0)=0$, then P has a transition density $\varrho(x,y)$ w.r.t. the Lebesgue measure such that*

$$\int_{\mathcal{R}^d}\varrho(x,y)\Phi^{-1}(\varrho(x,y))\,\mathrm{d}y\le\Phi^{-1}\left(\frac{1}{\int_{\mathcal{R}^d}\mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e}\right). \tag{2.12}$$

(2) *If $\Phi(r)=r^p$ for some $p>1$, then*

$$\int_{\mathcal{R}^d}\varrho(x,y)^{p/(p-1)}\,\mathrm{d}y\le\frac{1}{(\int_{\mathcal{R}^d}\mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e)^{1/(p-1)}}. \tag{2.13}$$

PROOF. Let $f \in \mathcal{B}_b^+(\mathcal{R}^d)$ such that $\int_{\mathcal{R}^d} \Phi(f)(x)\,\mathrm{d}x \le 1$. By (2.10) we have

$$\Phi(Pf)(x)\mathrm{e}^{-C_\Phi(x,e)} \le P\{\Phi \circ f(e+\cdot)\}(x) = \int_{\mathcal{R}^d} \Phi \circ f(y+e) P(x,\mathrm{d}y).$$

Integrating both sides w.r.t. $\mathrm{d}e$ and noting that $\int_{\mathcal{R}^d} \Phi \circ f(y+e)\,\mathrm{d}e = \int_{\mathcal{R}^d} \Phi \circ f(e)\,\mathrm{d}e \le 1$, we obtain

$$\Phi(Pf)(x) \int_{\mathcal{R}^d} \mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e \le 1.$$

This implies (2.11). When $\Phi(0) = 0$, (2.11) implies that

$$\sup_{f \in \mathcal{B}_b^+(\mathcal{R}^d), \int_{\mathcal{R}^d} \Phi \circ f(x)\,\mathrm{d}x \le 1} Pf(x) \le \Phi^{-1}\left(\frac{1}{\int_{\mathcal{R}^d} \mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e}\right) < \infty \tag{2.14}$$

since by the strictly increasing and convex properties we have $\Phi(r) \uparrow \infty$ as $r \uparrow \infty$. Now, for any Lebesgue-null set A, taking $f_n = n1_A$ we obtain from $\Phi(0) = 0$ that

$$\int_{\mathcal{R}^d} \Phi \circ f_n(x)\,\mathrm{d}x = 0 \le 1.$$

Therefore, applying (2.14) to $f = f_n$ we obtain

$$P(x, A) = P1_A(x) \le \frac{1}{n} \Phi^{-1}\left(\frac{1}{\int_{\mathcal{R}^d} \mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e}\right),$$

which goes to zero as $n \to \infty$. Thus $P(x,\cdot)$ is absolutely continuous w.r.t. the Lebesgue measure, so that the density function $\varrho(x, y)$ exists, and (2.12) follows from (2.11) by taking $f(y) = \Phi^{-1}(\varrho(x, y))$.

Finally, let $\Phi(r) = r^p$ for some $p > 1$. For fixed x, let

$$f_n(y) = \frac{\{n \wedge \varrho(x,y)\}^{1/(p-1)}}{(\int_{\mathcal{R}^d} \{n \wedge \varrho(x,y)\}^{p/(p-1)}\,\mathrm{d}y)^{1/p}}, \qquad n \ge 1.$$

It is easy to see that $\int_{\mathcal{R}^d} f_n^p(y)\,\mathrm{d}y = 1$. Then it follows from (2.11) with $\Phi(r) = r^p$ that

$$\int_{\mathcal{R}^d} \{n \wedge \varrho(x,y)\}^{p/(p-1)}\,\mathrm{d}y \le \big(Pf_n(x)\big)^{p/(p-1)} \le \frac{1}{(\int_{\mathcal{R}^d} \mathrm{e}^{-C_\Phi(x,e)}\,\mathrm{d}e)^{1/(p-1)}}.$$

Then (2.13) follows by letting $n \to \infty$. □

Finally, we consider applications of the shift Harnack inequality to distribution properties of the underlying transition probability.

THEOREM 2.6. *Let P be given above for some Banach space E, and let (2.10) hold for some $x, e \in E$, finite constant $C_\Phi(x, e)$ and some strictly increasing and convex continuous function Φ.*

(1) *$P(x,\cdot)$ is absolutely continuous w.r.t. $P(x,\cdot - e)$.*

(2) *If $\Phi(r) = r\Psi(r)$ for some strictly increasing positive continuous function Ψ on $(0,\infty)$. Then the density $\varrho(x,e;y) := \frac{P(x,\mathrm{d}y)}{P(x,\mathrm{d}y-e)}$ satisfies*

$$\int_E \Phi\big(\varrho(x,e;y)\big)P(x,\mathrm{d}y-e) \le \Psi^{-1}\big(\mathrm{e}^{C_\Phi(x,e)}\big).$$

PROOF. For $P(x,\cdot - e)$-null set A, let $f = 1_A$. Then (2.10) implies that $\Phi(P(x,A)) \le 0$, hence $P(x,A) = 0$ since $\Phi(r) > 0$ for $r > 0$. Therefore, $P(x,\cdot)$ is absolutely continuous w.r.t. $P(x,\cdot - e)$. Next, let $\Phi(r) = r\Psi(r)$. Applying (2.10) for $f(y) = \Psi(n \wedge \varrho(x,e;y))$ and noting that

$$Pf(x) = \int_E \{\Psi\big(n \wedge \varrho(x,e;y)\big)\}P(x,\mathrm{d}y) \ge \int_E \Phi\big(n \wedge \varrho(x,e;y)\big)P(x,\mathrm{d}y-e),$$

we obtain

$$\int_E \Phi\big(n \wedge \varrho(x,e;y)\big)P(x,\mathrm{d}y-e) \le \Psi^{-1}\big(\mathrm{e}^{C_\Phi(x,e)}\big).$$

Then the proof is complete by letting $n \to \infty$. $\square$

3. Stochastic Hamiltonian system. Consider the following degenerate stochastic differential equation on $\mathcal{R}^{m+d} = \mathcal{R}^m \times \mathcal{R}^d$ $(m \ge 0, d \ge 1)$:

$$\begin{cases} \mathrm{d}X(t) = \big\{AX(t) + BY(t)\big\}\,\mathrm{d}t, \\ \mathrm{d}Y(t) = Z\big(t, X(t), Y(t)\big)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \end{cases} \tag{3.1}$$

where A and B are two matrices of order $m \times m$ and $m \times d$, respectively, $Z : [0,\infty) \times \mathcal{R}^{m+d} \to \mathcal{R}^d$ is measurable with $Z(t,\cdot) \in C^1(\mathcal{R}^{m+d})$ for $t \ge 0$, $\{\sigma(t)\}_{t\ge 0}$ are invertible $d \times d$-matrices measurable in t such that the operator norm $\|\sigma(\cdot)^{-1}\|$ is locally bounded and $W(t)$ is the d-dimensional Brownian motion.

When $m \ge 1$ this equation is degenerate, and when $m = 0$ we set $\mathcal{R}^m = \{0\}$, so that the first equation disappears and thus, the equation reduces to a nondegenerate equation on $\mathcal{R}^d$. To ensure the existence of the transition density (or heat kernel) of the associated semigroup P_t w.r.t. the Lebesgue measure on $\mathcal{R}^{m+d}$, we make use of the following Kalman rank condition (see [11]) which implies that the associated diffusion is subelliptic,

(H) There exists $0 \le k \le m-1$ such that $\mathrm{Rank}\big[B, AB, \ldots, A^kB\big] = m$.

When $m = 0$ this condition is trivial, and for $m = 1$ it means that $\mathrm{Rank}(B) = 1$, that is, $B \ne 0$. For any $m > 1$ and $d \ge 1$, there exist plenty of examples for matrices A and B such that (H) holds; see [11]. Therefore, we allow that m is much larger than d, so that the associated diffusion process is highly degenerate; see Example 3.1 below.

It is easy to see that if $m = d$, $\sigma(t) = I_{d\times d}$, B is symmetric and

$$Z(x,y) = -\big\{\nabla V(x) + A^*y + F(x,y)(Ax+By)\big\}$$

for some smooth functions V and F, then (3.1) reduces to the Hamiltonian system

$$\begin{cases} \mathrm{d}X_t = \nabla H(X_t, \cdot)(Y_t)\,\mathrm{d}t, \\ \mathrm{d}Y_t = -\{\nabla H(\cdot, Y_t)(X_t) + F(X_t, Y_t)\nabla H(X_t, \cdot)(Y_t)\}\,\mathrm{d}t + \mathrm{d}W(t) \end{cases} \tag{3.2}$$

with Hamiltonian function

$$H(x, y) = V(x) + \langle Ax, y\rangle + \tfrac{1}{2}\langle By, y\rangle;$$

see, for example, [14]. If, in particular, $A = 0$, $B = I_{d\times d}$ and $F \equiv c$ for some constant c, the corresponding Fokker–Planck equation is known as the "kinetic Fokker–Planck equation" in PDE (see [15]), and the stochastic equation is called "stochastic damping Hamiltonian system"; see [21].

Let the solution to (3.1) be nonexplosive, and let

$$P_t f = \mathbf{E} f\big(X(t), Y(t)\big), \qquad t \ge 0, f \in \mathcal{B}_b\big(\mathcal{R}^{m+d}\big).$$

To state our main results, let us fix $T > 0$. For nonnegative $\phi \in C([0, T])$ with $\phi > 0$ in $(0, T)$, define

$$Q_\phi = \int_0^T \phi(t)\mathrm{e}^{(T-t)A} B B^* \mathrm{e}^{(T-t)A^*}\,\mathrm{d}t.$$

Then Q_ϕ is invertible; cf. [13]. For any $z \in \mathcal{R}^{m+d}$ and $r > 0$, let $B(z; r)$ be the ball centered at z with radius r.

THEOREM 3.1. *Assume* (H) *and that the solution to* (3.1) *is nonexplosive such that*

$$\sup_{t\in[0,T]} \mathbf{E}\Big\{\sup_{B(X(t),Y(t);r)} \big|\nabla Z(t, \cdot)\big|^2\Big\} < \infty, \qquad r > 0. \tag{3.3}$$

Let $\phi, \psi \in C^1([0, T])$ *such that* $\phi(0) = \phi(T) = 0$, $\phi > 0$ *in* $(0, T)$, *and*

$$\psi(T) = 1, \qquad \psi(0) = 0, \qquad \int_0^T \psi(t)\mathrm{e}^{(T-t)A} B\,\mathrm{d}t = 0. \tag{3.4}$$

Moreover, for $e = (e_1, e_2) \in \mathcal{R}^{m+d}$, *let*

$$h(t) = \phi(t) B^* \mathrm{e}^{(T-t)A^*} Q_\phi^{-1} e_1 + \psi(t) e_2 \in \mathcal{R}^d,$$

$$\Theta(t) = \left(\int_0^t \mathrm{e}^{(t-s)A} B h(s)\,\mathrm{d}s, h(t)\right) \in \mathcal{R}^{m+d}, \qquad t \in [0, T].$$

(1) *For any* $f \in C_b^1(\mathcal{R}^{m+d})$, *there holds*

$$P_T(\nabla_e f) = \mathbf{E}\Big\{ f\big(X(T), Y(T)\big) \times \int_0^T \big\langle \sigma(t)^{-1}\big\{h'(t) - \nabla_{\Theta(t)} Z(t, \cdot)\big(X(t), Y(t)\big)\big\}, \mathrm{d}W(t)\big\rangle\Big\}.$$

(2) *Let* $(X(0), Y(0)) = (x, y)$ *and*

$$R = \exp\biggl[-\int_0^T \langle \sigma(t)^{-1}\xi_1(t), \mathrm{d}W(t)\rangle - \frac{1}{2}\int_0^T \bigl|\sigma(t)^{-1}\xi_1(t)\bigr|^2 \,\mathrm{d}t\biggr],$$

where $\xi_1(t) = h'(t) + Z(t, X(t), Y(t)) - Z(t, X^1(t), Y^1(t))$ *with*

$$X^1(t) = X(t) + \int_0^t \mathrm{e}^{(t-s)A} Bh(s)\,\mathrm{d}s, \qquad Y^1(t) = Y(t) + h(t), \qquad t \ge 0.$$

Then

$$\bigl|P_T f(x, y)\bigr|^p \le P_T\bigl\{|f|^p(e + \cdot)\bigr\}(x, y)\bigl(\mathbf{E}R^{p/(p-1)}\bigr)^{p-1}, \qquad p > 1, f \in \mathcal{B}_b(E),$$

$$P_T \log f(x, y) \le \log P_T\bigl\{f(e + \cdot)\bigr\}(x, y) + \mathbf{E}(R \log R), \qquad 0 < f \in \mathcal{B}_b(E).$$

PROOF. We only prove (1), since (2) follows from Theorem 2.1 with the coupling constructed below for $\varepsilon = 1$. Let $(X^0(t), Y^0(t)) = (X(t), Y(t))$ solve (3.1) with initial data (x, y), and for $\varepsilon \in (0, 1]$ let $(X^\varepsilon(t), Y^\varepsilon(t))$ solve the equation

$$\begin{cases} \mathrm{d}X^\varepsilon(t) = \bigl\{AX^\varepsilon(t) + BY^\varepsilon(t)\bigr\}\,\mathrm{d}t, \\ \qquad X^\varepsilon(0) = x, \\ \mathrm{d}Y^\varepsilon(t) = \sigma(t)\,\mathrm{d}W(t) + \bigl\{Z\bigl(t, X(t), Y(t)\bigr) + \varepsilon h'(t)\bigr\}\,\mathrm{d}t, \\ \qquad Y^\varepsilon(0) = y. \end{cases} \tag{3.5}$$

Then it is easy to see that

$$\begin{cases} Y^\varepsilon(t) = Y(t) + \varepsilon h(t), \\ X^\varepsilon(t) = X(t) + \varepsilon \displaystyle\int_0^t \mathrm{e}^{(t-s)A} Bh(s)\,\mathrm{d}s. \end{cases} \tag{3.6}$$

Combining this with $\phi(0) = \phi(T) = 0$ and (3.4), we see that $h(T) = e_2$ and

$$\begin{aligned} &\int_0^T \mathrm{e}^{(T-t)A} Bh(t)\,\mathrm{d}t \\ &\quad = \int_0^T \phi(t)\mathrm{e}^{(T-t)A} BB^*\mathrm{e}^{(T-t)A^*} Q_\phi^{-1} e_1\,\mathrm{d}t + \int_0^T \psi(t)\mathrm{e}^{(T-t)A} Be_2\,\mathrm{d}t \\ &\quad = e_1. \end{aligned}$$

Therefore,

$$\bigl(X^\varepsilon(T), Y^\varepsilon(T)\bigr) = \bigl(X(T), Y(T)\bigr) + \varepsilon e, \qquad \varepsilon \in [0, 1]. \tag{3.7}$$

Next, to see that $((X(t), Y(t)), (X^\varepsilon(t), Y^\varepsilon(t)))$ is a coupling by change of measure for the solution to (3.1), reformulate (3.5) as

$$\begin{cases} \mathrm{d}X^\varepsilon(t) = \bigl\{AX^\varepsilon(t) + BY^\varepsilon(t)\bigr\}\,\mathrm{d}t, & X^\varepsilon(0) = x, \\ \mathrm{d}Y^\varepsilon(t) = \sigma(t)\,\mathrm{d}W^\varepsilon(t) + Z\bigl(t, X^\varepsilon(t), Y^\varepsilon(t)\bigr)\,\mathrm{d}t, & Y^\varepsilon(0) = y, \end{cases} \tag{3.8}$$

where

$$W^\varepsilon(t) := W(t) + \int_0^t \sigma(s)^{-1}\{\varepsilon h'(s) + Z(s, X(s), Y(s)) - Z(s, X^\varepsilon(s), Y^\varepsilon(s))\}\,\mathrm{d}s, \quad t \in [0, T].$$

Let

$$\xi_\varepsilon(s) = \varepsilon h'(s) + Z(s, X(s), Y(s)) - Z(s, X^\varepsilon(s), Y^\varepsilon(s)) \tag{3.9}$$

and

$$R_\varepsilon = \exp\left[-\int_0^T \langle \sigma(s)^{-1}\xi_\varepsilon(s), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^T |\sigma(s)^{-1}\xi_\varepsilon(s)|^2\,\mathrm{d}s\right].$$

By Lemma 3.2 below and the Girsanov theorem, $W^\varepsilon(t)$ is a d-dimensional Brownian motion under the probability measure $\mathbf{Q}_\varepsilon := R_\varepsilon \mathbf{P}$. Therefore, $((X(t), Y(t)), (X^\varepsilon(t), Y^\varepsilon(t)))$ is a coupling by change of measure with changed probability $\mathbf{Q}_\varepsilon$. Moreover, combining (3.6) with the definition of R_ε, we see from (3.3) that

$$-\frac{\mathrm{d}R_\varepsilon}{\mathrm{d}\varepsilon}\bigg|_{\varepsilon=0} = \int_0^T \langle \sigma_s^{-1}\{h'(s) - \nabla_{\Theta(s)} Z(s, \cdot)(X(s), Y(s))\}, \mathrm{d}W(s)\rangle$$

holds in $L^1(\mathbf{P})$. Then the proof is complete by Theorem 2.1(2). □

LEMMA 3.2. *Let the solution to* (3.1) *be nonexplosive such that* (3.3) *holds, and let* ξ_ε *be in* (3.9). *Then for any* $\varepsilon \in [0, 1]$ *the process*

$$R_\varepsilon(t) = \exp\left[-\int_0^t \langle \sigma(s)^{-1}\xi_\varepsilon(s), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t |\sigma(s)^{-1}\xi_\varepsilon(s)|^2\,\mathrm{d}s\right], \quad t \in [0, T]$$

is a uniformly integrable martingale with $\sup_{t\in[0,T]} \mathbf{E}\{R_\varepsilon(t)\log R_\varepsilon(t)\} < \infty$.

PROOF. Let $\tau_n = \inf\{t \ge 0 : |X(t)| + |Y(t)| \ge n\}, n \ge 1$. Then $\tau_n \uparrow \infty$ as $n \uparrow \infty$. It suffices to show that

$$\sup_{t\in[0,T], n\ge 1} \mathbf{E}\{R_\varepsilon(t \wedge \tau_n)\log R_\varepsilon(t \wedge \tau_n)\} < \infty. \tag{3.10}$$

By (3.6), there exists $r > 0$ such that

$$(X^\varepsilon(t), Y^\varepsilon(t)) \in B(X(t), Y(t); r), \qquad t \in [0, T], \varepsilon \in [0, 1]. \tag{3.11}$$

Let $Q_{\varepsilon,n} = R_\varepsilon(T \wedge \tau_n)\mathbf{P}$. By the Girsanov theorem, $\{W^\varepsilon(t)\}_{t\in[0,T\wedge\tau_n]}$ is the d-dimensional Brownian motion under the changed probability $\mathbf{Q}_{\varepsilon,n}$. Then, due to (3.11),

$$\sup_{t\in[0,T]} \mathbf{E}\{R_\varepsilon(t\wedge\tau_n)\log R_\varepsilon(t\wedge\tau_n)\}$$
$$= \frac{1}{2}\mathbf{E}_{\mathbf{Q}_{\varepsilon,n}}\int_0^{T\wedge\tau_n} |\sigma(s)^{-1}\xi_\varepsilon(s)|^2\,\mathrm{d}s$$
$$\le C + C\mathbf{E}_{\mathbf{Q}_{\varepsilon,n}}\int_0^{T\wedge\tau_n} \sup_{B(X^\varepsilon(t),Y^\varepsilon(t);r)} |\nabla Z(t,\cdot)|^2\,\mathrm{d}t$$

holds for some constant $C>0$ independent of n. Since the law of $(X^\varepsilon(\cdot\wedge\tau_n), Y^\varepsilon(\cdot\wedge\tau_n))$ under $\mathbf{Q}_{\varepsilon,n}$ coincides with that of $(X(\cdot\wedge\tau_n), Y(\cdot\wedge\tau_n))$ under $\mathbf{P}$, combining this with (3.3), we obtain

$$\sup_{t\in[0,T]} \mathbf{E}\{R_\varepsilon(t\wedge\tau_n)\log R_\varepsilon(t\wedge\tau_n)\}$$
$$\le C + C\int_0^T \mathbf{E}\sup_{B(X(t),Y(t);r)} |\nabla Z(t,\cdot)|^2\,\mathrm{d}t < \infty.$$

Therefore, (3.10) holds. $\square$

REMARK 3.1. (a) As shown in [10], Lemma 2.4, condition (3.3) is implied by the Lyapunov condition (A) therein, for which some concrete examples have been presented in [10]. Moreover, as shown in [10], Section 3 (see also Theorem 4.1 in [20]) that under reasonable grown conditions of $\nabla Z(t,\cdot)$ one obtains from Theorem 3.1(1)

$$P_t|\nabla f| \le \delta\{P_t(f\log f) - (P_t f)\log P_t f\} + \frac{W(t,\cdot)}{\delta}P_t f,$$
$$t>0,\ f\in\mathcal{B}_b^+(\mathcal{R}^{m+d}),\ \delta>\delta_0$$

for some constant $\delta_0\ge 0$ and some positive functions $W(t,\cdot)$. According to Theorem 2.2, this inequality implies the shift Harnack inequality.

(b) For any $T_2>T_1$. Applying Theorem 3.1 to $(X(T_1+t), Y(T_1+t))$ in place of $(X(t), Y(t))$, we see that the assertions in Theorem 3.1 hold for

$$P_{T_1,T_2}f(x,y) := \mathbf{E}\big(f(X(T_2),Y(T_2)) \mid (X(T_1),Y(T_1)) = (x,y)\big)$$

in place of $P_T f$ with T and 0 replaced by T_2 and T_1, respectively.

To derive explicit inequalities from Theorem 3.1, we consider below a special case where $\|\nabla Z(t,\cdot)\|_\infty$ is bounded and $A^l=0$ for some natural number $l\ge 1$.

COROLLARY 3.3. *Assume* (H). *If* $\|\nabla Z(t,\cdot)\|_\infty$ *and* $\|\sigma(t)^{-1}\|$ *are bounded in* $t\ge 0$, *and* $A^l=0$ *for some* $l\ge 1$. *Then there exists a constant* $C>0$ *such that for any positive* $f\in\mathcal{B}_b(\mathcal{R}^{m+d})$, $T>0$ *and* $e=(e_1,e_2)\in\mathcal{R}^{m+d}$:

(1) $(P_T f)^p \le P_T\{f^p(e+\cdot)\}\exp[\frac{Cp}{p-1}(\frac{|e_2|^2}{1\wedge T}+\frac{|e_1|^2}{(1\wedge T)^{4k+3}})]$, $p>1$;

(2) $P_T \log f \le \log P_T\{f(e+\cdot)\} + C(\frac{|e_2|^2}{1\wedge T}+\frac{|e_1|^2}{(1\wedge T)^{4k+3}})$;

(3) *for* $f \in C_b^1(\mathcal{R}^{m+d})$, $|P_T\nabla_e f|^2 \le C|P_T f^2|(\frac{|e_2|^2}{1\wedge T}+\frac{|e_1|^2}{(1\wedge T)^{4k+3}})$;

(4) *for strictly positive* $f \in C_b^1(\mathcal{R}^{m+d})$,

$$\begin{aligned}|P_T\nabla_e f|(x,y) &\le \delta\{P_T(f\log f)-(P_T f)\log P_T f\}\\ &\quad+\frac{C}{\delta}\biggl(\frac{|e_2|^2}{1\wedge T}+\frac{|e_1|^2}{(1\wedge T)^{4k+3}}\biggr)P_T f, \qquad \delta>0.\end{aligned}$$

PROOF. According to Remark 3.1(b), $P_T = P_{T-1}P_{T-1,T}$ and the Jensen inequality, we only need to prove for $T\in(0,1]$. Let $\phi(t)=\frac{t(T-t)}{T^2}$. Then $\phi(0)=\phi(T)=0$ and due to [20], Theorem 4.2(1), the rank condition (H) implies that

$$\|Q_\phi^{-1}\| \le cT^{-(2k+1)} \tag{3.12}$$

for some constant $c>0$ independent of $T\in(0,1]$. To fix the other reference function ψ in Theorem 3.1, let $\{c_i\}_{1\le i\le l+1}\in\mathcal{R}$ be such that

$$\begin{cases} 1+\displaystyle\sum_{i=1}^{l+1} c_i = 0,\\ 1+\displaystyle\sum_{i=1}^{l+1}\frac{j+1}{j+1+i}c_i=0, \qquad 0\le j\le l-1.\end{cases}$$

Take

$$\psi(t)=1+\sum_{i=1}^{l+1}c_i\frac{(T-t)^i}{T^i}, \qquad t\in[0,T].$$

Then $\psi(0)=0$, $\psi(T)=1$ and $\int_0^T(T-t)^j\psi(t)\,\mathrm{d}t=0$ for $0\le j\le l-1$. Since $A^l=0$, we conclude that $\int_0^T\psi(t)\mathrm{e}^{(T-t)A}\,\mathrm{d}t=0$. Therefore, (3.4) holds. It is easy to see that

$$|\psi(t)|\le c, \qquad |\psi'(t)|\le cT^{-1}, \qquad t\in[0,T]$$

holds for some constant $c>0$. Combining this with (3.12), (3.6) and the boundedness of $\|\nabla Z\|_\infty$ and $\|\sigma^{-1}\|$, we obtain

$$\begin{aligned}|\xi_1(t)|+|h'(t)| &\le c(T^{-2(k+1)}|e_1|+T^{-1}|e_2|),\\ |\Theta(t)| &\le c(T^{-(2k+1)}|e_1|+|e_2|)\end{aligned} \tag{3.13}$$

for some constant $c>0$. From this and Theorem 3.1, we derive the desired assertions. $\square$

COROLLARY 3.4. *In the situation of Corollary 3.3. Let $\|\cdot\|_{p\to q}$ be the operator norm from L^p to L^q w.r.t. the Lebesgue measure on $\mathcal{R}^{m+d}$. Then there exists a constant $C>0$ such that*

$$\|P_T\|_{p\to\infty}\le C^{1/p}\left(\frac{p}{p-1}\right)^{(m+d)/(2p)}(1\wedge T)^{-(d+(4k+3)m)/(2p)},\quad p>1, T>0. \tag{3.14}$$

Consequently, the transition density $p_T((x,y),(x',y'))$ of P_T w.r.t. the Lebesgue measure on $\mathcal{R}^{m+d}$ satisfies

$$\begin{aligned}&\int_{\mathcal{R}^{m+d}} p_T\big((x,y),(x',y')\big)^{p/(p-1)}\,\mathrm{d}x'\,\mathrm{d}y'\\ &\quad\le C^{1/(p-1)}\left(\frac{p}{p-1}\right)^{(m+d)/(2(p-1))}(1\wedge T)^{-(d+(4k+3)m)/(2(p-1))},\\ &\qquad T>0, (x,y)\in\mathcal{R}^{m+d}, p>1.\end{aligned}\tag{3.15}$$

PROOF. By Corollary 3.3(1), (3.14) follows from (2.11) for $P_T=P$, $\Phi(r)=r^p$ and

$$C_\Phi\big((x,y),(e_1,e_2)\big)=\frac{Cp}{p-1}\left(\frac{|e_2|^2}{1\wedge T}+\frac{|e_1|^2}{(1\wedge T)^{4k+3}}\right).$$

Moreover, (3.15) follows from (2.13). □

EXAMPLE 3.1. A simple example for Theorem 3.3 to hold is that $\sigma(t)=\sigma$ and $Z(t,\cdot)=Z$ are independent of t with $\|\nabla Z\|_\infty<\infty$, $A=0$ and $\mathrm{Rank}(B)=m$. In this case we have $d\ge m$; that is, the dimension of the generate part is controlled by that of the nondegenerate part. In general, our results allow m to be much larger than d. For instance, let $m=ld$ for some $l\ge 2$ and

$$A=\begin{pmatrix}0 & I_{d\times d} & 0 & \cdots & 0 & 0\\ 0 & 0 & I_{d\times d} & \cdots & 0 & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & \cdots & 0 & I_{d\times d}\\ 0 & 0 & 0 & \cdots & 0 & 0\end{pmatrix}_{(ld)\times(ld)},\qquad B=\begin{pmatrix}0\\ \cdot\\ \cdot\\ \cdot\\ 0\\ I_{d\times d}\end{pmatrix}_{(ld)\times d}.$$

Then $A^l=0$ and (H) holds for $k=m-1$. Therefore, assertions in Corollary 3.3 hold for $k=l-1$.

4. Functional stochastic differential equations. The purpose of this section is to establish Driver's integration by parts formula and shift Harnack inequality for delayed stochastic differential equations. In this case the associated segment processes are functional-valued, and thus, infinite-dimensional. As continuation to Section 3, it is natural for us to study the generalized stochastic Hamiltonian system with delay as in [3], where the Bismut formula and the Harnack inequalities are derived using coupling. However, for this model it seems very hard to construct the required new-type couplings. So, we only consider here the nondegenerate setting.

Let $\tau > 0$ be a fixed number, and let $\mathcal{C} = C([-\tau, 0]; \mathcal{R}^d)$ be equipped with uniform norm $\|\cdot\|_\infty$. For simplicity, we will use ∇ to denote the gradient operator both on $\mathcal{R}^d$ and $\mathcal{C}$. For instance, for a differentiable function F on $\mathcal{C}$ and $\xi \in \mathcal{C}$, $\nabla F(\xi)$ is a linear operator from $\mathcal{C}$ to $\mathcal{R}$ with

$$\mathcal{C} \ni \eta \mapsto \nabla_\eta F(\xi) = \lim_{\varepsilon \to 0} \frac{F(\xi + \varepsilon\eta) - F(\xi)}{\varepsilon}.$$

Moreover, let $\|\cdot\|$ be the operator norm for linear operators. Finally, for a function $h \in C([-\tau, \infty); \mathcal{R}^d)$ and $t \ge 0$, let $h_t \in \mathcal{C}$ be such that $h_t(\theta) = h(t + \theta)$, $\theta \in [-\tau, 0]$.

Consider the following stochastic differential equations on $\mathcal{R}^d$:

$$\mathrm{d}X(t) = b(t, X_t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \qquad t \ge 0, \tag{4.1}$$

where $W(t)$ is the Brownian motion on $\mathcal{R}^d$, $b : [0, \infty) \times \mathcal{C} \to \mathcal{R}^d$ is measurable such that $\|\nabla b(t, \cdot)\|_\infty$ is locally bounded in t, and $\sigma : [0, \infty) \to \mathcal{R}^d \otimes \mathcal{R}^d$ is measurable with $\|\sigma(t)^{-1}\|$ locally bounded. We remark that the local boundedness assumption of $\|\nabla b(t, \cdot)\|_\infty$ is made only for simplicity and can be weakened by some growth conditions as in [3].

Now, for any $\xi \in \mathcal{C}$, let $X^\xi(t)$ be the solution to (4.1) for $X_0 = \xi$, and let X_t^ξ be the associated segment process. Let

$$P_t F(\xi) = \mathbf{E} F\big(X_t^\xi\big), \qquad t \ge 0, \xi \in \mathcal{C}, F \in \mathcal{B}_b(\mathcal{C}).$$

We aim to establish the integration by parts formula and shift Harnack inequality for P_T. It turns out that we are only able to make derivatives or shifts along directions in the Cameron–Martin space

$$\mathcal{H} := \left\{ h \in \mathcal{C} : \|h\|_{\mathcal{H}}^2 := \int_{-\tau}^0 |h'(t)|^2\,\mathrm{d}t < \infty \right\}.$$

THEOREM 4.1. *Let $T > \tau$ and $\eta \in \mathcal{H}$ be fixed. For any $\phi \in \mathcal{B}_b([0, T - \tau])$ such that $\int_0^{T-\tau} \phi(t)\,\mathrm{d}t = 1$, let*

$$\Gamma(t) = \begin{cases} \phi(t)\eta(-\tau), & \text{if } t \in [0, T - \tau], \\ \eta'(t - T), & \text{if } t \in (T - \tau, T]. \end{cases}$$

Let $\|\sigma(t)^{-1}\| \le K(T)$, $\|\nabla b(t, \cdot)\|_\infty \le \kappa(T)$ for $t \in [0, T]$.

(1) *For any $F \in C_b^1(\mathcal{C})$,*

$$P_T(\nabla_\eta F) = \mathbf{E}\left(F(X_T) \int_0^T \big\langle \sigma(t)^{-1}\big(\Gamma(t) - \nabla_{\Theta_t} b(t, \cdot)(X_t)\big), \mathrm{d}W(t) \big\rangle \right)$$

holds for

$$\Theta(t)=\int_0^{t\vee 0}\Gamma(s)\,\mathrm{d}s,\qquad t\in[-\tau,T].$$

Consequently, for any $\delta>0$ *and positive* $F\in C_b^1(\mathcal{C})$,

$$\begin{aligned}\big|P_T(\nabla_\eta F)\big| &\le \delta\big\{P_T(F\log F)-(P_TF)\log P_TF\big\}\\ &\quad+\frac{2K(T)^2(1+\kappa(T)^2T^2)}{\delta}\bigg(\|\eta\|_{\mathcal{H}}^2+\frac{|\eta(-\tau)|^2}{T-\tau}\bigg)P_TF.\end{aligned}$$

(2) *For any nonnegative* $F\in\mathcal{B}_b(\mathcal{C})$,

$$\begin{aligned}(P_TF)^p &\le \big(P_T\{F(\eta+\cdot)\}^p\big)\\ &\quad\times\exp\bigg[\frac{2pK(T)^2(1+\kappa(T)^2T^2)}{p-1}\bigg(\|\eta\|_{\mathcal{H}}^2+\frac{|\eta(-\tau)|^2}{T-\tau}\bigg)\bigg].\end{aligned}$$

(3) *For any positive* $F\in\mathcal{B}_b(\mathcal{C})$,

$$P_T\log F\le\log P_T\{F(\eta+\cdot)\}+2K(T)^2\big(1+\kappa(T)^2T^2\big)\bigg(\|\eta\|_{\mathcal{H}}^2+\frac{|\eta(-\tau)|^2}{T-\tau}\bigg).$$

PROOF. For fixed $\xi\in\mathcal{C}$, let $X(t)$ solve (4.1) for $X_0=\xi$. For any $\varepsilon\in[0,1]$, let $X^\varepsilon(t)$ solve the equation

$$\mathrm{d}X^\varepsilon(t)=\big\{b(t,X_t)+\varepsilon\Gamma(t)\big\}\,\mathrm{d}t+\sigma(t)\,\mathrm{d}W(t),\qquad t\ge 0, X_0^\varepsilon=\xi.$$

Then it is easy to see that

$$X_t^\varepsilon=X_t+\varepsilon\Theta_t,\qquad t\in[0,T]. \tag{4.2}$$

In particular, $X_T^\varepsilon=X_T+\varepsilon\eta$. Next, let

$$\begin{aligned}R_\varepsilon=\exp\bigg[&-\int_0^T\big\langle\sigma(t)^{-1}\big\{\varepsilon\Gamma(t)+b(t,X_t)-b\big(t,X_t^\varepsilon\big)\big\},\mathrm{d}W(t)\big\rangle\\ &-\frac12\int_0^T\big|\sigma(t)^{-1}\big\{\varepsilon\Gamma(t)+b(t,X_t)-b\big(t,X_t^\varepsilon\big)\big\}\big|^2\,\mathrm{d}t\bigg].\end{aligned}$$

By the Girsanov theorem, under the changed probability $\mathbf{Q}_\varepsilon:=R_\varepsilon\mathbf{P}$, the process

$$W^\varepsilon(t):=W(t)+\int_0^t\sigma(s)^{-1}\big(\Gamma(s)+b(s,X_s)-b\big(s,X_s^\varepsilon\big)\big)\,\mathrm{d}s,\qquad t\in[0,T]$$

is a d-dimensional Brownian motion. So, (X_t,X_t^ε) is a coupling by change of measure with changed probability $\mathbf{Q}_\varepsilon$. Then the desired integration by parts formula follows from Theorem 2.1 since $R_0=1$ and due to (4.2),

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}R^\varepsilon\bigg|_{\varepsilon=0}=-\int_0^T\big\langle\sigma(t)^{-1}\big(\Gamma(t)-\nabla_{\Theta_t}b(t,\cdot)(X_t)\big),\mathrm{d}W(t)\big\rangle$$

holds in $L^1(\mathbf{P})$. Taking $\phi(t) = \frac{1}{T-\tau}$, we have

$$\int_0^T \left|\Gamma(t)\right|^2 \mathrm{d}t \le \|\eta\|_{\mathcal{H}}^2 + \frac{|\eta(-\tau)|^2}{T-\tau},$$

$$\|\nabla_{\Theta_t} b(t,\cdot)\|_\infty^2 \le \kappa(T)^2\left(\int_0^T \left|\Gamma(t)\right| \mathrm{d}t\right)^2 \le \kappa(T)^2 T \int_0^T \left|\Gamma(t)\right|^2 \mathrm{d}t.$$

Then

$$\int_0^T \left|\Gamma(t) - \nabla_{\theta_t} b(t,\cdot)(X_t)\right|^2 \mathrm{d}t \le 2\bigl(1 + T^2\kappa(T)^2\bigr)\left(\|\eta\|_{\mathcal{H}}^2 + \frac{|\eta(-\tau)|^2}{T-\tau}\right). \tag{4.3}$$

So,

$$\begin{aligned}
&\log \mathbf{E} \exp\left[\frac{1}{\delta}\int_0^T \bigl\langle \sigma(t)^{-1}\bigl\{\Gamma(t) - \nabla_{\Theta_t} b(t,\cdot)(X_t)\bigr\}, \mathrm{d}W(t)\bigr\rangle\right] \\
&\quad \le \frac{1}{2}\log \mathbf{E}\exp\left[\frac{2K(T)^2}{\delta^2}\int_0^T \left|\Gamma(t) - \nabla_{\Theta_t} b(t,\cdot)(X_t)\right|^2 \mathrm{d}t\right] \\
&\quad \le \frac{2K(T)^2(1+T^2\kappa(T)^2)}{\delta^2}\left(\|\eta\|_{\mathcal{H}}^2 + \frac{|\eta(-\tau)|^2}{T-\tau}\right).
\end{aligned}$$

Then the second result in (1) follows from the Young inequality

$$\begin{aligned}
\left|P_T(\nabla_\eta F)\right| &\le \delta\bigl\{P_T(F\log F) - (P_T F)\log P_T F\bigr\} \\
&\qquad + \delta \log \mathbf{E}\exp\left[\frac{1}{\delta}\int_0^T \bigl\langle \sigma(t)^{-1}\bigl\{\Gamma(t) - \nabla_{\Theta_t} b(t,\cdot)(X_t)\bigr\}, \mathrm{d}W(t)\bigr\rangle\right].
\end{aligned}$$

Finally, (2) and (3) can be easily derived by applying Theorem 2.1 for the above constructed coupling with $\varepsilon = 1$, and using (4.2) and (4.3). □

From Theorem 4.1 we may easily derive regularization estimates on $P_T(\xi,\cdot)$, the distribution of X_T^ξ. For instance, Theorem 4.1(1) implies estimates on the derivative of $P_T(\xi, A+\cdot)$ along $\eta \in \mathcal{H}$ for $\xi \in \mathcal{C}$ and measurable $A \subset \mathcal{C}$; and due to Theorems 2.6, 4.1(2) and 4.1(3) imply some integral estimates on the density $p_T(\xi,\eta;\gamma) := \frac{P_T(\xi,\mathrm{d}\gamma)}{P_T(\xi,\mathrm{d}\gamma-\eta)}$ for $\eta \in \mathcal{H}$. Moreover, since $\mathcal{H}$ is dense in $\mathcal{C}$, the shift Harnack inequality in Theorem 4.1(2) implies that $P_T(\xi,\cdot)$ has full support on $\mathcal{C}$ for any $T > \tau$ and $\xi \in \mathcal{C}$.

5. Semi-linear stochastic partial differential equations. The purpose of this section is to establish Driver's integration by parts formula and shift Harnack inequality for semi-linear stochastic partial differential equations. We note that the Bismut formula has been established in [4] for a class of delayed SPDEs, but for technical reasons we only consider here the case without delay.

Let $(H, \langle\cdot,\cdot\rangle, |\cdot|)$ be a real separable Hilbert space, and $(W(t))_{t\ge 0}$ a cylindrical Wiener process on H with respect to a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with the natural filtration $\{\mathcal{F}_t\}_{t\ge 0}$. Let $\mathcal{L}(H)$ and $\mathcal{L}_{\mathrm{HS}}(H)$ be the spaces of all linear bounded operators and Hilbert–Schmidt operators on H, respectively. Denote by $\|\cdot\|$ and $\|\cdot\|_{\mathrm{HS}}$ the operator norm and the Hilbert–Schmidt norm, respectively.

Consider the following semi-linear SPDE:

$$\begin{cases} \mathrm{d}X(t) = \big\{AX(t) + b\big(t, X(t)\big)\big\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \\ X(0) = x \in H, \end{cases} \tag{5.1}$$

where

(A1) $(A, \mathcal{D}(A))$ is a linear operator on H generating a contractive, strongly continuous semigroup $(\mathrm{e}^{tA})_{t\ge 0}$ such that $\int_0^1 \|\mathrm{e}^{sA}\|_{\mathrm{HS}}^2\,\mathrm{d}s < \infty$.

(A2) $b:[0,\infty)\times H \to H$ is measurable, and Fréchet differentiable in the second variable such that $\|\nabla b(t,\cdot)\|_\infty := \sup_{x\in H}\|\nabla b(t,\cdot)(x)\|$ is locally bounded in $t \ge 0$.

(A3) $\sigma:[0,\infty)\to\mathcal{L}(H)$ is measurable and locally bounded, and $\sigma(t)$ is invertible such that $\|\sigma(t)^{-1}\|$ is locally bounded in $t\ge 0$.

Then the equation (5.1) has a unique a mild solution (see [6]), which is an adapt process $(X(t))_{t\ge 0}$ on H such that

$$X(t) = \mathrm{e}^{tA}x + \int_0^t \mathrm{e}^{(t-s)A} b\big(s, X(s)\big)\,\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\sigma(s)\,\mathrm{d}W(s), \qquad t \ge 0.$$

Let

$$P_t f\big(X(0)\big) = \mathbf{E} f\big(X(t)\big), \qquad t\ge 0, X(0)\in H, f \in \mathcal{B}_b(H).$$

Finally, for any $e \in H$, let

$$e(t) = \int_0^t \mathrm{e}^{sA} e\,\mathrm{d}s, \qquad t \ge 0.$$

THEOREM 5.1. *Let $T > 0$ and $e \in \mathcal{D}(A)$ be fixed. Let $\|\sigma(t)^{-1}\| \le K(T)$, $\|\nabla b(t,\cdot)\|_\infty \le \kappa(T)$ for $t \in [0,T]$.*

(1) *For any $f \in C_b^1(H)$,*

$$P_T(\nabla_{e(T)} f) = \mathbf{E}\bigg(f\big(X(T)\big) \int_0^T \big\langle \sigma(t)^{-1}\big(e - \nabla_{e(t)} b(t,\cdot)\big(X(t)\big)\big), \mathrm{d}W(t)\big\rangle\bigg).$$

Consequently, for any $\delta > 0$ and positive $f \in C_b^1(H)$,

$$\begin{aligned} \big|P_T(\nabla_{e(T)} f)\big| &\le \delta\big\{P_T(f\log f) - (P_T f)\log P_T f\big\} \\ &\quad + \frac{K(T)^2|e|^2}{\delta}\bigg(T + T^2\kappa(T) + \frac{T^3\kappa(T)^2}{3}\bigg) P_T f. \end{aligned}$$

(2) *For any nonnegative $F \in \mathcal{B}_b(H)$,*

$$(P_T F)^p \le \big(P_T\big\{F\big(e(T)+\cdot\big)\big\}^p\big)\exp\bigg[\frac{pK(T)^2|e|^2}{p-1}\bigg(T + T^2\kappa(T) + \frac{T^3\kappa(T)^2}{3}\bigg)\bigg].$$

(3) *For any positive* $F \in \mathcal{B}_b(H)$,

$$P_T \log F \le \log P_T\{F(e(T)+\cdot)\} + K(T)^2|e|^2\left(T + T^2\kappa(T) + \frac{T^3\kappa(T)^2}{3}\right).$$

PROOF. For fixed $x \in H$, let $X(t)$ solve (4.1) for $X(0) = x$. For any $\varepsilon \in [0, 1]$, let $X^\varepsilon(t)$ solve the equation

$$\mathrm{d}X^\varepsilon(t) = \{AX^\varepsilon(t) + b(t, X(t)) + \varepsilon e\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W(t), \qquad t \ge 0, X^\varepsilon(0) = x.$$

Then it is easy to see that

$$X^\varepsilon(t) = X(t) + \varepsilon e(t), \qquad t \in [0, T]. \tag{5.2}$$

In particular, $X^\varepsilon(T) = X(T) + \varepsilon e(T)$. Next, let

$$R_\varepsilon = \exp\left[-\int_0^T \langle\sigma(t)^{-1}\{\varepsilon e + b(t, X(t)) - b(t, X^\varepsilon(t))\}, \mathrm{d}W(t)\rangle - \frac{1}{2}\int_0^T |\sigma(t)^{-1}\{\varepsilon e + b(t, X(t)) - b(t, X^\varepsilon(t))\}|^2\,\mathrm{d}t\right].$$

By the Girsanov theorem, under the weighted probability $\mathbf{Q}_\varepsilon := R_\varepsilon\mathbf{P}$, the process

$$W^\varepsilon(t) := W(t) + \int_0^t \sigma(s)^{-1}(\varepsilon e + b(s, X_s) - b(s, X_s^\varepsilon))\,\mathrm{d}s, \qquad t \in [0, T]$$

is a d-dimensional Brownian motion. So, $(X(t), X^\varepsilon(t))$ is a coupling by change of measure with changed probability $\mathbf{Q}_\varepsilon$. Then the desired integration by parts formula follows from Theorem 2.1 since $R_0 = 1$ and due to (5.2),

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}R^\varepsilon\Big|_{\varepsilon=0} = -\int_0^T \langle\sigma(t)^{-1}(e - \nabla_{e(t)}b(t,\cdot)(X(t))), \mathrm{d}W(t)\rangle$$

holds in $L^1(\mathbf{P})$. This formula implies the second inequality in (1) due to the given upper bounds on $\|\sigma(t)^{-1}\|$ and $\|\nabla b(t,\cdot)\|$ and the fact that

$$\begin{aligned}
&|P_T(\nabla_\eta F)| - \delta\{P_T(F\log F) - (P_T F)\log P_T F\} \\
&\quad \le \delta \log \mathbf{E}\exp\left[\frac{1}{\delta}\int_0^T \langle\sigma(t)^{-1}(e - \nabla_{e(t)}b(t,\cdot)(X(t))), \mathrm{d}W(t)\rangle\right]P_T F \\
&\quad \le \frac{\delta}{2}\log \mathbf{E}\exp\left[\frac{2}{\delta^2}\int_0^T |\sigma(t)^{-1}(e - \nabla_{e(t)}b(t,\cdot)(X(t)))|^2\,\mathrm{d}t\right]P_T F.
\end{aligned}$$

Finally, since $|e(t)| \le t|e|$, (2) and (3) can be easily derived by applying Theorem 2.1 for the above constructed coupling with $\varepsilon = 1$. □

Acknowledgment. The author would like to thank the referees for helpful comments and corrections.

REFERENCES

[1] ARNAUDON, M., THALMAIER, A. and WANG, F.-Y. (2006). Harnack inequality and heat kernel estimates on manifolds with curvature unbounded below. *Bull. Sci. Math.* **130** 223–233. MR2215664

[2] ARNAUDON, M., THALMAIER, A. and WANG, F.-Y. (2009). Gradient estimates and Harnack inequalities on non-compact Riemannian manifolds. *Stochastic Process. Appl.* **119** 3653–3670. MR2568290

[3] BAO, J., WANG, F.-Y. and YUAN, C. (2013). Derivative formula and Harnack inequality for degenerate functionals SDEs. *Stoch. Dyn.* **13** 1250013, 22. MR3007251

[4] BAO, J., WANG, F.-Y. and YUAN, C. (2013). Bismut formulae and applications for functional SPDEs. *Bull. Sci. Math.* **137** 509–522. MR3054274

[5] BISMUT, J.-M. (1984). *Large Deviations and the Malliavin Calculus*. Birkhäuser, Boston. MR0755001

[6] DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions*. Cambridge Univ. Press, Cambridge. MR1207136

[7] DRIVER, B. K. (1997). Integration by parts for heat kernel measures revisited. *J. Math. Pures Appl.* (9) **76** 703–737. MR1470881

[8] ELWORTHY, K. D. and LI, X. M. (1994). Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** 252–286. MR1297021

[9] FANG, S., LI, H. and LUO, D. (2011). Heat semi-group and generalized flows on complete Riemannian manifolds. *Bull. Sci. Math.* **135** 565–600. MR2838091

[10] GUILLIN, A. and WANG, F.-Y. (2012). Degenerate Fokker–Planck equations: Bismut formula, gradient estimate and Harnack inequality. *J. Differential Equations* **253** 20–40. MR2917400

[11] KALMAN, R. E., FALB, P. L. and ARBIB, M. A. (1969). *Topics in Mathematical System Theory*. McGraw-Hill, New York. MR0255260

[12] NUALART, D. (1995). *The Malliavin Calculus and Related Topics*. Springer, New York. MR1344217

[13] SEIDMAN, T. I. (1988). How violent are fast controls? *Math. Control Signals Systems* **1** 89–95. MR0923278

[14] SOIZE, C. (1994). *The Fokker–Planck Equation for Stochastic Dynamical Systems and Its Explicit Steady State Solutions*. *Series on Advances in Mathematics for Applied Sciences* **17**. World Scientific, Singapore. MR1287386

[15] VILLANI, C. (2009). Hypocoercivity. *Mem. Amer. Math. Soc.* **202** iv+141. MR2562709

[16] WANG, F.-Y. (1997). On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups. *Probab. Theory Related Fields* **108** 87–101. MR1452551

[17] WANG, F. Y. (2013). Derivative formula and gradient estimates for Gruschin type semigroups. *J. Theor. Probab.* **27** 80–95.

[18] WANG, F.-Y. and XU, L. (2012). Derivative formula and applications for hyperdissipative stochastic Navier-Stokes/Burgers equations. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **15** 1250020, 19. MR2999101

[19] WANG, F.-Y. and YUAN, C. (2011). Harnack inequalities for functional SDEs with multiplicative noise and applications. *Stochastic Process. Appl.* **121** 2692–2710. MR2832420

[20] WANG, F.-Y. and ZHANG, X.-C. (2013). Derivative formula and applications for degenerate diffusion semigroups. *J. Math. Pures Appl.* (9) **99** 726–740. MR3055216

[21] WU, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process*. *Appl*. **91** 205–238. MR1807683

[22] ZHANG, X. (2010). Stochastic flows and Bismut formulas for stochastic Hamiltonian systems. *Stochastic Process*. *Appl*. **120** 1929–1949. MR2673982

SCHOOL OF MATHEMATICAL SCIENCES
BEIJING NORMAL UNIVERSITY
BEIJING 100875
CHINA
E-MAIL: wangfy@bnu.edu.cn
AND
DEPARTMENT OF MATHEMATICS
SWANSEA UNIVERSITY
SINGLETON PARK, SWANSEA SA2 8PP
UNITED KINGDOM
E-MAIL: F.Y.Wang@swansea.ac.uk