Acta Mathematica Sinica, English Series
Mar., 2021, Vol. 37, No. 3, pp. 458–470
Published online: November 4, 2020
https://doi.org/10.1007/s10114-020-9365-6
http://www.ActaMath.com

Acta Mathematica Sinica, English Series

Non-uniform Gradient Estimates for SDEs with Local Monotonicity Conditions

Jian WANG[1)]
College of Mathematics and Informatics & Fujian Key Laboratory of Mathematical Analysis and Applications (FJKLMAA) & Center for Applied Mathematics of Fujian Province (FJNU), Fujian Normal University, Fuzhou 350007*, P. R. China*
E-mail: jianwang@fjnu.edu.cn

Bing Yao WU
College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350007*, P. R. China*
E-mail: bingyaowu@163.com

Abstract By using the coupling method and the localization technique, we establish non-uniform gradient estimates for Markov semigroups of diffusions or stochastic differential equations driven by pure jump Lévy noises, where the coefficients only satisfy local monotonicity conditions.

Keywords Gradient estimate, Markov semigroup, monotonicity condition, coupling, stochastic differential equation, Lévy process

MR(2010) Subject Classification 60G51, 60G52, 60J25, 60J75

1 Introduction and Main Results

Let L be the following second order elliptic differential operator on $\mathbb{R}^d$:

$$L = \frac{1}{2} \sum_{1 \le i,j \le d} a_{ij}(x)\partial_{i,j} + \sum_{1 \le i \le d} b_i(x)\partial_i,$$

where $a(x) := (a_{ij}(x))_{1 \le i,j \le d}$ and $b(x) := (b_i(x))_{1 \le i \le d}$ are continuous on $\mathbb{R}^d$, and there exists a constant $\lambda_0 > 0$ such that for any $x, h \in \mathbb{R}^d$,

$$\langle a(x)h, h\rangle \ge \lambda_0 |h|^2. \tag{1.1}$$

Suppose that the martingale problem for the operator $(L, C_b^2(\mathbb{R}^d))$ is well-posed. Equivalently, the following stochastic differential equation (SDE) on $\mathbb{R}^d$:

$$dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt, \quad X_0 = x \in \mathbb{R}^d \tag{1.2}$$

has a unique weak solution. Here, $(B_t)_{t \ge 0}$ is a standard Brownian motion on $\mathbb{R}^d$, and $\sigma(x) := (\sigma_{ij}(x))_{1 \le i,j \le d}$ satisfies $a(x) = \sigma(x)\sigma^T(x)$, where $\sigma^T(x)$ denotes the transpose of the matrix

Received September 4, 2019, revised February 16, 2020, accepted March 26, 2020
Supported by the National Natural Science Foundation of China (Grant No. 11831014), the Program for Probability and Statistics: Theory and Application (Grant No. IRTL1704) and the Program for Innovative Research Team in Science and Technology in Fujian Province University (IRTSTFJ)
1) Corresponding author

$\sigma(x)$. See [7, 11] for more details. Denote by $\mathbb{P}^x$ and $\mathbb{E}^x$ the probability and the expectation of the process $(X_t)_{t\geq 0}$ starting from $x \in \mathbb{R}^d$, respectively. Let $(P_t)_{t\geq 0}$ be the Markov semigroup associated with the process $(X_t)_{t\geq 0}$ (or the operator L). Then,

$$P_t f(x) = \mathbb{E}^x f(X_t), \quad t \geq 0,\ x \in \mathbb{R}^d,\ f \in B_b(\mathbb{R}^d).$$

There are lots of developments on gradient estimates for the diffusion semigroup $(P_t)_{t\geq 0}$, see [1–4, 6, 10, 12] and references therein. In particular, under the assumption that the coefficients $\sigma(x)$ and $b(x)$ fulfill the one-sided Lipschitz condition (monotonicity condition), i.e., there is a constant $C > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$\|\sigma(x) - \sigma(y)\|^2 + 2\langle b(x) - b(y), x - y\rangle \leq C|x - y|^2. \tag{1.3}$$

Priola and Wang in [10, Theorem 3.4] used the probabilistic coupling approach to obtain the following uniform gradient estimates for $(P_t)_{t\geq 0}$:

$$\|\nabla P_t f\|_\infty \leq \frac{C_0}{\sqrt{t \wedge 1}}\|f\|_\infty, \quad t > 0,\ f \in B_b(\mathbb{R}^d), \tag{1.4}$$

where $C_0 > 0$ is a constant independent of $t > 0$ and $f \in B_b(\mathbb{R}^d)$. Here and in what follows,

$$|\nabla P_t f(x)| = \limsup_{y\to x} \frac{|P_t f(y) - P_t f(x)|}{|x - y|}, \quad x \in \mathbb{R}^d.$$

Indeed, the setting of [10, Theorem 3.4] is more general than (1.3), and it was assumed that there is a nonnegative continuous function g on $(0, +\infty)$ with $\int_0^1 g(s)\,ds < \infty$ so that for any $x, y \in \mathbb{R}^d$,

$$\|\sigma(x) - \sigma(y)\|^2 + 2\langle b(x) - b(y), x - y\rangle \leq g(|x - y|)|x - y|. \tag{1.5}$$

In particular, when $g(x) = Cx$ for some constant $C > 0$, (1.5) is reduced into the monotonicity condition (1.3).

Motivated by [10, Theorem 3.4], in this paper we will study non-uniform gradient estimates for the semigroup $(P_t)_{t\geq 0}$ when the coefficients $\sigma(x)$ and $b(x)$ only satisfy local monotonicity conditions. One of our contributions is as follows.

Theorem 1.1 *Suppose that there exist a C^2-function $W : \mathbb{R}^d \to [1, \infty)$ and a nonnegative continuous function g on $(0, +\infty)$ with $\int_0^1 g(s)\,ds < \infty$ such that the following two conditions hold*:

(i) *for any $x, y \in \mathbb{R}^d$,*

$$\|\sigma(x) - \sigma(y)\|^2 + 2\langle b(x) - b(y), x - y\rangle \leq (W(x) + W(y))g(|x - y|)|x - y|; \tag{1.6}$$

(ii) *there exist constants $\lambda > \int_0^1 g(s)\,ds/(2\lambda_0)$ and $C_\lambda > 0$ so that*

$$L\mathrm{e}^{\lambda W} \leq C_\lambda \mathrm{e}^{\lambda W}, \tag{1.7}$$

where the constant λ_0 is given in (1.1).
Then, there is a constant $C > 0$ such that for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 0$,

$$|\nabla P_t f(x)| \leq \frac{C}{\sqrt{t \wedge 1}} \mathrm{e}^{\lambda W(x)} \|f\|_\infty. \tag{1.8}$$

In particular, when $\sigma(x)$ and $b(x)$ satisfy (1.3), we can take $W(x)$ to be the constant function in (1.6). In this case, (1.7) holds trivially. Therefore, Theorem 1.1 extends [10, Theorem 3.4]. On the other hand, when $\lim_{|x|\to\infty} W(x) = \infty$, (1.7) can be regarded as the standard Lyapunov drift condition on the generator L, which guarantees that the process $(X_t)_{t\geq 0}$ is non-explosive. This along with Condition (i) in turn yields that the SDE given by (1.2) indeed has a unique strong solution.

It is natural to ask whether Theorem 1.1 still holds when (1.7) is replaced by a simpler condition like $LW \leq C_0 W$. The latter one is easy to verify and practical in applications. The following statement will address this question in some especial settings.

Theorem 1.2 *Suppose that there exist a C^2-function $W : \mathbb{R}^d \to [1,\infty)$ with $\lim_{|x|\to\infty} W(x) = \infty$ and a nonnegative continuous function g on $(0,+\infty)$ such that $\int_0^1 g(s)\,ds < \infty$. If Condition* (i) *and the following condition hold*:

(ii') *there exist constants C_1, $C_2 > 0$ and $\alpha \in (0,1]$ such that for any $x \in \mathbb{R}^d$,*

$$LW^{1/\alpha}(x) \leq C_1 W^{1/\alpha}(x) \tag{1.9}$$

and

$$|\sigma^T(x)\nabla W(x)|^2 \leq C_2 W(x), \tag{1.10}$$

then, for any $\lambda > \int_0^1 g(s)\,ds/(2\lambda_0)$ with the constant λ_0 in (1.1), *there exists a constant $C > 0$ such that* (1.8) *holds for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 0$.*

To illustrate the power of Theorem 1.2, we consider the following example.

Example 1.3 Suppose that there exist $\alpha \in (0,1]$, $C_0 > 0$ and a continuous nonnegative function g on $(0,+\infty)$ with $\int_0^1 g(s)\,ds < \infty$ so that the following two conditions hold:

(i) for any $x, y \in \mathbb{R}^d$,

$$\|\sigma(x)-\sigma(y)\|^2 + 2\langle b(x)-b(y), x-y\rangle \leq C_0((1+|x|^2)^\alpha + (1+|y|^2)^\alpha)g(|x-y|)|x-y|^2;$$

(ii) for any $x \in \mathbb{R}^d$,

$$\langle x, b(x)\rangle \leq C_0(1+|x|^2), \quad \|\sigma(x)\|^2 \leq C_0(1+|x|^{2(1-\alpha)}).$$

Then there exist constants $C, \lambda > 0$ such that for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 0$,

$$|\nabla P_t f(x)| \leq \frac{C}{\sqrt{t\wedge 1}} \mathrm{e}^{\lambda(1+|x|^2)^\alpha}\|f\|_\infty.$$

Recently, Prato and Priola in [5, Theorem 1.2] also considered gradient estimates for diffusion semigroups under local monotonicity conditions. In details, by checking the proof of [5, Theorem 1.2] (in particular, see the proofs of [5, Proposition 2.1 and Corollary 2.2]), Theorem 1.2 was proved to hold true under the assumptions that both coefficients $\sigma(x)$ and $b(x)$ are C^1, and that there is a nonnegative measurable function $W(x)$ such that

(i*) for any $x, y \in \mathbb{R}^d$,

$$\|\sigma(x)-\sigma(y)\|^2 + 2\langle b(x)-b(y), x-y\rangle \leq (W(x)+W(y))|x-y|^2;$$

(i**) there is a constant $\lambda > 0$ such that for any $x \in \mathbb{R}^d$,

$$LW^4(x) \leq \lambda W^4(x).$$

Since the additional condition (1.10) is not required, [5, Theorem 1.2] is better than Theorem 1.2 when $\alpha = 1/4$. However, the approach of [5, Theorem 1.2] is based on the Bismut–Elworthy–Li formula, and so it is crucial to assume that $\sigma(x)$ and $b(x)$ are C^1 functions; while in the setting of our paper, we only assume that both coefficients are continuous. Moreover, the idea for proofs of Theorems 1.1 and 1.2 is completely different from that of [5, Theorem 1.2]. We mainly apply the coupling technique and the localization argument. Our method is still efficient for gradient estimates of SDEs with Lévy noises under local monotonicity conditions (see Section 3 for more details), for which the approach of [5, Theorem 1.2] seems to be not easy to apply.

The remainder of the paper is arranged as follows. The next section is devoted to proofs of Theorems 1.1 and 1.2 as well as that of Example 1.3. In Section 3, we will consider gradient estimates for SDEs driven by additive pure jump Lévy noises under local monotonicity conditions on the drift term.

2 Proofs of Theorems 1.1 and 1.2

In this section, we will prove Theorems 1.1 and 1.2. We first give the

Proof of Theorem 1.1 The proof is split into two parts.

(1) Let W be the function in Theorem 1.1. For any $k \ge 1$, define

$$\tau_k = \inf\{t > 0 : W(X_t) \ge k\},$$

where we set $\inf \emptyset = \infty$. Then, for any $x \in \mathbb{R}^d$ and $t > 0$,

$$\mathbb{E}^x(\mathrm{e}^{-C_\lambda(t\wedge\tau_k)}\mathrm{e}^{\lambda W(X_{t\wedge\tau_k})}) \ge \mathbb{E}^x(\mathrm{e}^{-C_\lambda(t\wedge\tau_k)}\mathrm{e}^{\lambda W(X_{t\wedge\tau_k})}\mathbb{1}_{\{\tau_k\le t\}}) \ge \mathrm{e}^{-C_\lambda t}\mathrm{e}^{\lambda k}\mathbb{P}^x(\tau_k \le t).$$

Furthermore, according to Condition (ii), we have

$$\begin{aligned}\mathbb{P}^x(\tau_k \le t) &\le \mathrm{e}^{C_\lambda t}\mathrm{e}^{-\lambda k}\mathbb{E}^x(\mathrm{e}^{-C_\lambda(t\wedge\tau_k)}\mathrm{e}^{\lambda W(X_{t\wedge\tau_k})})\\ &= \mathrm{e}^{C_\lambda t}\mathrm{e}^{-\lambda k}\left[\mathrm{e}^{\lambda W(x)} + \mathbb{E}^x\int_0^{t\wedge\tau_k}\mathrm{e}^{-C_\lambda s}(L\mathrm{e}^{\lambda W(X_s)} - C_\lambda\mathrm{e}^{\lambda W(X_s)})\,ds\right]\\ &\le \mathrm{e}^{C_\lambda t}\mathrm{e}^{-\lambda k}\mathrm{e}^{\lambda W(x)}.\end{aligned} \tag{2.1}$$

In particular, by (2.1), the process $(X_t)_{t\ge0}$ is non-explosive.

(2) Let $(X_t, Y_t)_{t\ge0}$ be the Markov coupling constructed in [10, Section 3.1], and

$$T = \inf\{t > 0 : X_t = Y_t\}$$

be the coupling time of the process $(X_t, Y_t)_{t\ge0}$. For any $k \ge 1$, define

$$\begin{aligned}\widetilde{\tau}_k &= \inf\{t > 0 : W(X_t) + W(Y_t) \ge k\},\\ \tau_{k,1} &= \inf\{t > 0 : W(X_t) \ge k\},\\ \tau_{k,2} &= \inf\{t > 0 : W(Y_t) \ge k\}.\end{aligned}$$

Denote by $\mathbb{P}^{(x,y)}$ the probability of the process $(X_t, Y_t)_{t\ge0}$ starting from (x, y). According to (2.1),

$$\begin{aligned}\mathbb{P}^{(x,y)}(\widetilde{\tau}_k \le t \le \widetilde{\tau}_{k+1}) &\le \mathbb{P}^{(x,y)}(\widetilde{\tau}_k \le t)\\ &\le \mathbb{P}^x(\tau_{k/2,1} \le t) + \mathbb{P}^y(\tau_{k/2,2} \le t)\\ &\le \mathrm{e}^{C_\lambda t}\mathrm{e}^{-\lambda k/2}(\mathrm{e}^{\lambda W(x)} + \mathrm{e}^{\lambda W(y)}).\end{aligned} \tag{2.2}$$

Next, we estimate $\mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_{k+1} \geq t \geq \widetilde{\tau}_k)$. Note that, under the event $\{\widetilde{\tau}_{k+1} \geq t \geq \widetilde{\tau}_k\}$, the coupling process $(X_t, Y_t)_{t\geq 0}$ will stay in the region

$$\{(x, y) \in \mathbb{R}^{2d} : W(x) + W(y) \leq k + 1\}$$

before time t. Then, thanks to Condition (i), without loss of generality we may and do assume that the associated coefficients of the coupling process $(X_t, Y_t)_{t\geq 0}$ before time t satisfy the following monotonicity condition

$$\|\sigma(x) - \sigma(y)\|^2 + 2\langle b(x) - b(y), x - y\rangle \leq (k + 1)g(|x - y|)|x - y|, \quad x, y \in \mathbb{R}^d.$$

According to [10, Theorem 3.4 (a)] and its proof (see [10, p. 254–255]), for any $t > 0$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}^d} \limsup_{y\to x} \frac{\mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_k \leq t \leq \widetilde{\tau}_{k+1})}{|x - y|} \\
&\quad\leq \inf_{r>0} \left\{ \frac{\int_0^r \exp[\frac{k+1}{4\lambda_0} \int_0^s g(u)\, du]\, ds}{2\lambda_0 t} + \frac{1}{\int_0^r \exp[-\frac{k+1}{4\lambda_0} \int_0^s g(u)\, du]\, ds} \right\} \\
&\quad\leq \exp\left(\frac{(k + 1)\int_0^1 g(u)\, du}{4\lambda_0}\right)\left(\frac{\sqrt{t \wedge 1}}{2\lambda_0 t} + \frac{1}{\sqrt{t \wedge 1}}\right) \\
&\quad\leq \left(1 + \frac{1}{2\lambda_0}\right) \exp\left(\frac{(k + 1)\int_0^1 g(u)\, du}{4\lambda_0}\right)\frac{1}{\sqrt{t \wedge 1}}.
\end{aligned} \tag{2.3}$$

By (2.2) and (2.3), we find that for any $\lambda > \int_0^1 g(s)\, ds/(2\lambda_0)$, $x \in \mathbb{R}^d$ and $t > 0$,

$$\begin{aligned}
&\limsup_{y\to x} \frac{\mathbb{P}^{(x,y)}(T > t)}{|x - y|} \\
&\quad= \limsup_{y\to x} \frac{\sum_{k=1}^{\infty} \mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_k \leq t \leq \widetilde{\tau}_{k+1})\mathbb{P}^{(x,y)}(\widetilde{\tau}_k \leq t \leq \widetilde{\tau}_{k+1})}{|x - y|} \\
&\quad\leq 2\left(1 + \frac{1}{2\lambda_0}\right) \exp\left(\frac{\int_0^1 g(u)\, du}{4\lambda_0}\right) \frac{\mathrm{e}^{C_\lambda t}}{\sqrt{t \wedge 1}}\mathrm{e}^{\lambda W(x)} \sum_{k=1}^{\infty} \exp\left(\frac{k\int_0^1 g(u)\, du}{4\lambda_0} - \frac{\lambda k}{2}\right) \\
&\quad= \frac{C_0\mathrm{e}^{C_\lambda t}}{\sqrt{t \wedge 1}}\mathrm{e}^{\lambda W(x)},
\end{aligned}$$

where

$$C_0 := 2\left(1 + \frac{1}{2\lambda_0}\right) \exp\left(\frac{\int_0^1 g(u)du}{4\lambda_0}\right) \sum_{k=1}^{\infty} \exp\left[k\left(\frac{\int_0^1 g(u)\, du}{4\lambda_0} - \frac{\lambda}{2}\right)\right] < \infty,$$

thanks to the fact that $\lambda > \int_0^1 g(s)\, ds/(2\lambda_0)$. Therefore, for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t \in (0, 1]$,

$$\begin{aligned}
\limsup_{y\to x} \frac{|P_t f(x) - P_t f(y)|}{|x - y|} &\leq \|f\|_\infty \limsup_{y\to x} \frac{\mathbb{P}^{(x,y)}(T > t)}{|x - y|} \\
&\leq \frac{C_0\mathrm{e}^{C_\lambda t}}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\|f\|_\infty \\
&\leq \frac{C_0\mathrm{e}^{C_\lambda}}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\|f\|_\infty.
\end{aligned}$$

Now, for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 1$, by the property of the semigroup $(P_t)_{t\geq 0}$,

$$\limsup_{y\to x} \frac{|P_t f(x) - P_t f(y)|}{|x - y|} = \limsup_{y\to x} \frac{|P_1 P_{t-1} f(x) - P_1 P_{t-1} f(y)|}{|x - y|}$$

$$\le C_0 e^{C_\lambda} e^{\lambda W(x)} \|P_{t-1} f\|_\infty$$
$$\le C_0 e^{C_\lambda} e^{\lambda W(x)} \|f\|_\infty.$$

Combining with both conclusions above, we prove Theorem 1.1. □

Next, we turn to the

Proof of Theorem 1.2 For simplicity, throughout the proof we will adopt the notations in the proof of Theorem 1.1.

First, we claim that for any $x \in \mathbb{R}^d$, $k \ge 1$ and $t > 0$,

$$\mathbb{P}^x(\tau_k \le t) \le C_0 \exp(-\lambda e^{-\alpha\beta t} k) e^{\lambda W(x)},$$

where

$$\tau_k = \inf\{t > 0 : W(X_t) \ge k\}.$$

Indeed, we set $\overline{W}(x) := W^{1/\alpha}(x)$. By (1.9),

$$L\overline{W}(x) \le C_1 \overline{W}(x).$$

Since the function $W \in C^2(\mathbb{R}^d)$, $\lim_{|x|\to\infty} W(x) = \infty$ and $\alpha \in (0, 1]$, we have $\overline{W} \in C^2(\mathbb{R}^d)$ and $\lim_{|x|\to\infty} \overline{W}(x) = \infty$. Furthermore, due to $|\sigma^T(x)\nabla W(x)|^2 \le C_2 W(x)$, it holds that

$$\frac{1}{\alpha^2} W^{2/\alpha-2}(x) |\sigma^T(x)\nabla W(x)|^2 \le \frac{C_2}{\alpha^2} W^{2/\alpha-1}(x);$$

that is,

$$|\sigma^T(x)\nabla \overline{W}(x)|^2 \le \frac{C_2}{\alpha^2} (\overline{W}(x))^{2-\alpha}.$$

Hence, under Condition (ii'), the function $\overline{W}$ satisfies all the assumptions of [14, Lemma 2.2]. In particular, by [14, Lemma 2.2], for any $\lambda > 0$ and $x \in \mathbb{R}^d$,

$$\mathbb{E}^x \Big[\sup_{t>0} \exp(\lambda e^{-\alpha\beta t} W(X_t)) \Big] \le C_0 e^{\lambda W(x)}, \tag{2.4}$$

where

$$\beta := \beta(\lambda, \alpha, C_1, C_2) = C_1 + C_2 \alpha\lambda/4 + 2/(\alpha\lambda), \quad C_0 = 2(C_2(\alpha\lambda)^2 + 1).$$

Therefore, for any $x \in \mathbb{R}^d$ and $t > 0$,

$$\begin{aligned}\mathbb{E}^x(\exp(\lambda e^{-\alpha\beta(t\wedge\tau_k)} W(X_{t\wedge\tau_k}))) &\ge \mathbb{E}^x(\exp(\lambda e^{-\alpha\beta(t\wedge\tau_k)} W(X_{t\wedge\tau_k})) \mathbb{1}_{\{\tau_k \le t\}}) \\ &\ge \exp(\lambda e^{-\alpha\beta t} k) \mathbb{P}^x(\tau_k \le t),\end{aligned}$$

which along with (2.4) gives us that for any $x \in \mathbb{R}^d$ and $t > 0$,

$$\mathbb{P}^x(\tau_k \le t) \le C_0 \exp(-\lambda e^{-\alpha\beta t} k) e^{\lambda W(x)}. \tag{2.5}$$

In the following, we will make use of the coupling process $(X_t, Y_t)_{t\ge0}$ as that in part (2) of the proof for Theorem 1.1. By (2.5), for any $x, y \in \mathbb{R}^d$ and $t > 0$,

$$\begin{aligned}\mathbb{P}^{(x,y)}(\widetilde{\tau}_k \le t \le \widetilde{\tau}_{k+1}) &\le \mathbb{P}^x(\tau_{k/2} \le t) + \mathbb{P}^y(\tau_{k/2} \le t) \\ &\le C_0 \exp(-\lambda e^{-\alpha\beta t}(k/2))(e^{\lambda W(x)} + e^{\lambda W(y)}).\end{aligned} \tag{2.6}$$

On the other hand, using Condition (i), (2.3) and (2.6), and following the argument in part (2) of the proof for Theorem 1.1, we then can get that for any $\lambda > 0$, $x \in \mathbb{R}^d$ and $t > 0$,

$$\limsup_{y\to x} \frac{\mathbb{P}^{(x,y)}(T > t)}{|x-y|}$$

$$
\begin{aligned}
&= \limsup_{y\to x} \frac{\sum_{k=1}^\infty \mathbb{P}^{(x,y)}(T>t|\widetilde{\tau}_k\le t\le \widetilde{\tau}_{k+1})\mathbb{P}(\widetilde{\tau}_k\le t\le\widetilde{\tau}_{k+1})}{|x-y|}\\
&\le 2\bigg(1+\frac{1}{2\lambda_0}\bigg)\exp\bigg(\frac{\int_0^1 g(u)\,du}{4\lambda_0}\bigg)\frac{C_0}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\sum_{k=1}^\infty \exp\bigg(\frac{k\int_0^1 g(u)\,du}{4\lambda_0}-\frac{k}{2}\lambda \mathrm{e}^{-\alpha\beta t}\bigg).
\end{aligned}
$$

Now, let $\lambda > \int_0^1 g(u)\,du/(2\lambda_0)$. Then, we can choose $t_0>0$ small enough (depending on $\lambda,\alpha,\beta,\lambda_0$ and g) such that

$$
\lambda > \exp(\alpha\beta t_0)\int_0^1 g(u)du/(2\lambda_0). \tag{2.7}
$$

Hence, for any $x\in\mathbb{R}^d$ and $t\in(0,t_0]$,

$$
\begin{aligned}
&\limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T>t)}{|x-y|}\\
&\quad\le 2\bigg(1+\frac{1}{2\lambda_0}\bigg)\exp\bigg(\frac{\int_0^1 g(u)\,du}{4\lambda_0}\bigg)\frac{C_0}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\sum_{k=1}^\infty \exp\bigg(\frac{k\int_0^1 g(u)\,du}{4\lambda_0}-\frac{k}{2}\lambda \mathrm{e}^{-\alpha\beta t_0}\bigg)\\
&\quad=: \frac{C}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\|f\|_\infty,
\end{aligned}
$$

where

$$
C = 2C_0\bigg(1+\frac{1}{2\lambda_0}\bigg)\exp\bigg(\frac{\int_0^1 g(u)du}{4\lambda_0}\bigg)\sum_{k=1}^\infty \exp\bigg[k\bigg(\frac{\int_0^1 g(u)du}{4\lambda_0}-\frac{\lambda}{2}\mathrm{e}^{-\alpha\beta t_0}\bigg)\bigg]<\infty,
$$

thanks to (2.7). Therefore, for any $x\in\mathbb{R}^d$, $f\in B_b(\mathbb{R}^d)$ and $0<t\le t_0$,

$$
\begin{aligned}
\limsup_{y\to x}\frac{|P_tf(x)-P_tf(y)|}{|x-y|} &\le \|f\|_\infty \limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T>t)}{|x-y|}\\
&\le \frac{C}{\sqrt{t}}\mathrm{e}^{\lambda W(x)}\|f\|_\infty.
\end{aligned}
$$

Combining the assertion above with the Markov property of the semigroup $(P_t)_{t\ge 0}$, we obtain that for any $x\in\mathbb{R}^d$, $f\in B_b(\mathbb{R}^d)$ and $t>0$,

$$
|\nabla P_tf(x)| \le \frac{C}{\sqrt{t\wedge 1}}\mathrm{e}^{\lambda W(x)}\|f\|_\infty,
$$

which yields the assertion of Theorem 1.2. □

Proof of Example 1.3 Let $W(x)=(1+|x|^2)^\alpha$. By (ii),

$$
\begin{aligned}
LW^{1/\alpha}(x) = L(1+|x|^2) &= \|\sigma(x)\|^2 + 2\langle x, b(x)\rangle\\
&\le 2C_0(1+|x|^2)+C_0(1+|x|^{2(1-\alpha)})\\
&\le C_1W^{1/\alpha}(x)
\end{aligned}
$$

and

$$
\begin{aligned}
|\sigma^T(x)\nabla W(x)|^2 &= |\sigma^T(x)\nabla(1+|x|^2)^\alpha|^2\\
&= |2\alpha(1+|x|^2)^{\alpha-1}\sigma^T(x)x|^2\\
&\le 4C_0\alpha^2(1+|x|^2)^{2\alpha-2}(1+|x|^{2(1-\alpha)})(1+|x|^2)\\
&\le C_2(1+|x|^2)^\alpha,
\end{aligned}
$$

where $C_1, C_2 > 0$ are independent of $x \in \mathbb{R}^d$. Therefore, the function $W(x)$ satisfies Condition (ii') in Theorem 1.2. On the other hand, by (ii) we see that (i) in Theorem 1.2 holds with the function $W(x)$. Hence, the assertion of Example 1.3 follows from Theorem 1.2. □

3 Gradient Estimates for SDEs with Lévy Noises

In this section, we will consider the following SDE with Lévy noises

$$dX_t = dZ_t + b(X_t)\,dt, \quad X_0 = x \in \mathbb{R}^d, \tag{3.1}$$

where $b(x) : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel measurable function, and $Z := (Z_t)_{t\ge 0}$ is a pure jump Lévy process on $\mathbb{R}^d$. Throughout this section, we always assume that the SDE (3.1) has a unique strong solution, which is denoted by $X := (X_t)_{t\ge 0}$. Then, by the Itô formula, the generator of the process X is given by

$$Lf(x) = \langle b(x), \nabla f(x)\rangle + \int_{\mathbb{R}^d} (f(x+z) - f(x) - \langle \nabla f(x), z\rangle \mathbb{1}_{\{|z|\le 1\}})\,\nu(dz)$$

acting on $f \in C_b^2(\mathbb{R}^d)$, where ν is the Lévy measure of the process Z. We mention that under Conditions (i) and (ii) in the main result Theorem 3.1 below, the SDE (3.1) indeed has a unique strong solution.

There are several recent papers on uniform gradient estimates for the Markov semigroup associated with the SDE (3.1). Zhang in [14, Theorem 1.1] used a time-change argument combined with the Malliavin calculus to obtain a Bismut–Elworthy–Li's derivative formula for the semigroup when Z is a (rotationally invariant) symmetric α-stable process in (3.1). Later, a new derivative formula of Bismut–Elworthy–Li's type was established in [13, Theorem 1.1] for the semigroup, based on the Malliavin calculus and a finite-jump approximation argument, corresponding to a class of SDEs driven by multiplicative Lévy noises. Via the coupling method, gradient estimates for the semigroup of the SDE (3.1) (driven by more general Lévy noises) under monotonicity conditions on the drift term was given in [9, Theorem 5.1]. All gradient estimates of the semigroup given in these quoted papers are uniform with the following form:

$$\|\nabla P_t f\|_\infty \le \psi(t)\|f\|_\infty, \quad t > 0,\ f \in B_b(\mathbb{R}^d),$$

where $\psi(t)$ is a nonnegative function on $(0,\infty)$.

The purpose of this section is to extend Theorem 1.1 to SDEs with Lévy noises, and to study non-uniform gradient estimates for the associated semigroup. The main result of this section is as follows.

Theorem 3.1 *Suppose that there is a C^2-function $W : \mathbb{R}^d \to [1,\infty)$ such that the following three conditions hold*:

(i) *for any $x, y \in \mathbb{R}^d$,*

$$\langle b(x) - b(y), x - y\rangle \le (W(x) + W(y))|x - y|^2;$$

(ii) *there exist constants $\beta, \lambda > 0$ so that for any $x \in \mathbb{R}^d$,*

$$LW^\beta(x) \le \lambda W^\beta(x);$$

(iii) *there are a constant $\varepsilon_0 > 0$ and $\phi \in C[0,\varepsilon_0] \cap C^3(0,\varepsilon_0]$ so that*

(a) *$\phi(0) = 0$, $\phi' \ge 0$, $\phi'' \le 0$ and $\phi''' \ge 0$ on $(0,\varepsilon_0]$;*

(b) *both functions* $\Phi_1(r) := -\frac{1}{4}J(r)r\frac{\phi''(2r)}{\phi'(r)}$ *and* $\Phi_2(r) := -\frac{1}{4}J(r)r^2\phi''(r)$ *are non-increasing on* $(0,\varepsilon_0]$ *with* $\lim_{r\to 0}\Phi_1(r) = +\infty$, *and*

$$\sum_{k=1}^{\infty}\frac{2^{-\beta k}}{\phi(\Phi_1^{-1}(a2^k))} < \infty$$

for some constant $a \geq 1$ *such that* $\Phi_1^{-1}(2a) \leq \varepsilon_0/2$, *where* $J(r)$ *is a nonnegative function on* $(0,\infty)$ *such that*

$$J(r) \leq J_0(r) := \inf_{|x|\leq r}[\nu \wedge (\delta_x * \nu)](\mathbb{R}^d), \quad r \in (0,1]$$

and

$$\Phi_1^{-1}(r) = \inf\{s \in (0,\varepsilon_0] : \Phi_1(s) \leq r\}.$$

Then, there exists a constant $C > 0$ *such that for any* $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ *and* $t > 0$,

$$\limsup_{y\to x}\frac{|P_tf(x) - P_tf(y)|}{\phi(|x-y|)} \leq \frac{C}{\phi(\Phi_3^{-1}(t\wedge 1))}W^\beta(x)\|f\|_\infty,$$

where $\Phi_3(r) = \phi(r)/\Phi_2(r)$.

Proof The idea of the proof is similar to that of proofs for Theorems 1.1 and 1.2, but we require much more delicate estimates due to the Lévy noises. By the property of the semigroup $(P_t)_{t\geq 0}$, we only need to consider the case that $t \in (0,1]$.

Let τ_k be the stopping time given in the proof of Theorem 1.1. Then, for any $x \in \mathbb{R}^d$ and $t > 0$,

$$\mathbb{E}^x(\mathrm{e}^{-\lambda(t\wedge\tau_k)}W^\beta(X_{t\wedge\tau_k})) \geq \mathbb{E}^x(\mathrm{e}^{-\lambda(t\wedge\tau_k)}W^\beta(X_{t\wedge\tau_k})\mathbb{1}_{\{\tau_k\leq t\}}) \geq \mathrm{e}^{-\lambda t}k^\beta\mathbb{P}^x(\tau_k \leq t).$$

This along with Condition (ii) yields that

$$\begin{aligned}\mathbb{P}^x(\tau_k \leq t) &\leq \mathrm{e}^{\lambda t}k^{-\beta}\mathbb{E}^x(\mathrm{e}^{-\lambda(t\wedge\tau_k)}W^\beta(X_{t\wedge\tau_k}))\\ &= \mathrm{e}^{\lambda t}k^{-\beta}\bigg(W^\beta(x) + \mathbb{E}^x\int_0^{t\wedge\tau_k}\mathrm{e}^{-\lambda s}(LW^\beta(X_s) - \lambda W^\beta(X_s))\,ds\bigg)\\ &\leq \mathrm{e}^{\lambda t}k^{-\beta}W^\beta(x).\end{aligned} \tag{3.2}$$

Let $(X_t,Y_t)_{t\geq 0}$ be the coupling process given in the proof of [9, Theorem 5.1]. We still use the notation in part (2) in the proof of Theorem 1.1. According to (3.2), we have

$$\mathbb{P}^{(x,y)}(\widetilde{\tau}_{2^k} \leq t \leq \widetilde{\tau}_{2^{k+1}}) \leq 2^\beta\mathrm{e}^{\lambda t}2^{-\beta k}(W^\beta(x) + W^\beta(y)). \tag{3.3}$$

Under the event $\{\widetilde{\tau}_{2^k} \leq t \leq \widetilde{\tau}_{2^{k+1}}\}$, the coupling process $(X_t,Y_t)_{t\geq 0}$ will stay in the region $\{(x,y) \in \mathbb{R}^{2d} : W(x) + W(y) \leq 2^{k+1}\}$. By (i), we can assume that in this case the coefficients of the process $(X_t,Y_t)_{t\geq 0}$ satisfy that for any $x,y \in \mathbb{R}^d$,

$$\langle b(x) - b(y), x-y\rangle \leq 2^{k+1}|x-y|^2.$$

Then, according to [9, Theorem 5.1] and its proof (see [9, pp. 31–32]) for any $x \in \mathbb{R}^d$, $t > 0$ and $k \geq 1$,

$$\limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_k \leq t \leq \widetilde{\tau}_{k+1})}{\phi(|x-y|)} \leq \inf_{0<\varepsilon\leq\varepsilon_0^*}\bigg\{\frac{1}{\phi(\varepsilon)} + \frac{1}{tA_\varepsilon^k(\phi)}\bigg\}, \tag{3.4}$$

where $\varepsilon_0^* > 0$ is independent of x, t and k, and

$$A_\varepsilon^k(\phi) = \inf_{0<r\le\varepsilon}\left\{\frac{1}{2}J(r)(2\phi(r)-\phi(2r)) - 2^{k+1}\phi'(r)r\right\}, \quad \varepsilon > 0.$$

In the following, for simplicity we will assume that $\varepsilon_0 = \varepsilon_0^*$; otherwise, we will replace them by $\varepsilon_0 \wedge \varepsilon_0^*$. Since $\phi''' \ge 0$ on $(0, \varepsilon_0]$ for some $\varepsilon_0 > 0$,

$$2\phi(r) - \phi(2r) = -\int_0^r \int_s^{r+s} \phi''(u)\,du\,ds \ge -\phi''(2r)r^2, \quad 0 < r \le \varepsilon_0/2.$$

Let $a \ge 1$ large enough such that $\Phi_1^{-1}(2a) \le \varepsilon_0/2$. Thus, for any $\varepsilon \in (0, \Phi_1^{-1}(a2^{k+1})]$,

$$\begin{aligned}
A_\varepsilon^k(\phi) &\ge \inf_{0<r\le\varepsilon}\left\{-\frac{1}{2}J(r)r^2\phi''(2r) - 2^{k+1}\phi'(r)r\right\}\\
&= \inf_{0<r\le\varepsilon}\left\{-\frac{1}{4}J(r)r^2\phi''(2r) - \frac{1}{4}J(r)r^2\phi''(2r) - 2^{k+1}\phi'(r)r\right\}\\
&\ge \inf_{0<r\le\varepsilon}\left\{-\frac{1}{4}J(r)r^2\phi''(2r)\right\} + \inf_{0<r\le\varepsilon}\left\{\phi'(r)r\left(-\frac{1}{4}\frac{J(r)r\phi''(2r)}{\phi'(r)} - 2^{k+1}\right)\right\}\\
&= \inf_{0<r\le\varepsilon}\left\{-\frac{1}{4}J(r)r^2\phi''(2r)\right\} + \inf_{0<r\le\varepsilon}\{\phi'(r)r(\Phi_1(r) - 2^{k+1})\}\\
&\ge \inf_{0<r\le\varepsilon}\left\{-\frac{1}{4}J(r)r^2\phi''(2r)\right\}\\
&= -\frac{1}{4}J(\varepsilon)\varepsilon^2\phi''(2\varepsilon),
\end{aligned}$$

where in the last inequality we used the facts that $\Phi_1(r)$ is non-increasing on $(0, \varepsilon_0]$ and $\phi' \ge 0$ on $(0, \varepsilon_0]$, and the last equality follows from the non-increasing property of the function $\Phi_2(r) = -\frac{1}{4}J(r)r^2\phi''(2r)$ on $(0, \varepsilon_0]$.

Let $\varepsilon_k = \Phi_1^{-1}(a2^{k+1})$ for all $k \ge 1$. By (b) in Condition (ii), $\lim_{k\to\infty}\varepsilon_k = 0$. Then, we can simplify (3.4) as that for any $x \in \mathbb{R}^d$, $t \in (0, 1]$ and $k \ge 1$,

$$\begin{aligned}
\limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_k \le t \le \widetilde{\tau}_{k+1})}{\phi(|x-y|)} &\le \inf_{0<\varepsilon\le\varepsilon_k}\left\{\frac{1}{\phi(\varepsilon)} + \frac{1}{t\Phi_2(\varepsilon)}\right\}\\
&\le \frac{c_0}{\phi(\varepsilon_k \wedge \Phi_3^{-1}(t))}\\
&\le \frac{c_1}{\phi(\varepsilon_k)\phi(\Phi_3^{-1}(t))},
\end{aligned} \tag{3.5}$$

where in the second inequality we used the facts that ϕ is non-decreasing and Φ_2 is non-increasing on $(0, \varepsilon_0]$, and the last inequality follows from $\lim_{k\to\infty}\varepsilon_k = 0$ and $t \in (0, 1]$. Here, c_0, c_1 are independent of x and t.

Combining (3.3) with (3.5), we find that for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t \in (0, 1]$,

$$\begin{aligned}
&\limsup_{y\to x}\frac{|P_tf(x) - P_tf(y)|}{\phi(|x-y|)}\\
&\le 2\|f\|_\infty \limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T > t)}{\phi(|x-y|)}\\
&\le 2\|f\|_\infty \sum_{k=0}^\infty \limsup_{y\to x}\frac{\mathbb{P}^{(x,y)}(T > t|\widetilde{\tau}_{2^{k+1}} \ge t \ge \widetilde{\tau}_{2^k})\mathbb{P}^{(x,y)}(\widetilde{\tau}_{2^{k+1}} \ge t \ge \widetilde{\tau}_{2^k})}{\phi(|x-y|)}
\end{aligned}$$

$$\leq \frac{c_2 W^\beta(x)\|f\|_\infty}{\phi(\Phi_3^{-1}(t))}\bigg[\sum_{k=0}^\infty \frac{2^{-\beta k}}{\phi(\varepsilon_k)}\bigg]$$
$$\leq \frac{c_3 W^\beta(x)\|f\|_\infty}{\phi(\Phi_3^{-1}(t))},$$

where in the last inequality we used (b) in Condition (ii) again. The proof is completed. □

As a consequence of Theorem 3.1, we have the following corollary.

Corollary 3.2 *Under the setting of Theorem* 3.1*, suppose that Conditions* (i) *and* (ii) *hold, and that there are constants* $c_0 > 0$ *and* $\alpha \in (0, 2)$*,*

$$J_0(r) \geq c_0 r^{-\alpha}, \quad r \in (0, 1].$$

Then, the following statements hold.

(1) *If* $\alpha \in (0, 1]$*, and the constant* β *in Condition* (ii) *satisfies that* $\beta > \theta/\alpha$ *with some* $\theta \in (0, \alpha)$*, then there exists a constant* $C > 0$ *such that for any* $x \in \mathbb{R}^d$*,* $f \in B_b(\mathbb{R}^d)$ *and* $t > 0$*,*

$$\limsup_{y\to x} \frac{|P_t f(x) - P_t f(y)|}{|x-y|^\theta} \leq C W^\beta(x)(t\wedge 1)^{-\theta/\alpha}\|f\|_\infty.$$

(2) *If* $\alpha \in (1, 2)$*, and the constant* β *in Condition* (ii) *satisfies that* $\beta > 1/\alpha$*, then for any* $\theta > 0$*, there is a constant* $C > 0$ *such that for all* $x \in \mathbb{R}^d$*,* $f \in B_b(\mathbb{R}^d)$ *and* $t > 0$*,*

$$\limsup_{y\to x} \frac{|P_t f(x) - P_t f(y)|}{|x-y|} \leq C W^\beta(x)\bigg(\frac{\log^{1+\theta}(2/(t\wedge 1))}{t\wedge 1}\bigg)^{1/\alpha}\|f\|_\infty.$$

Proof We set $J(r) = c_0 r^{-\alpha}$.

(1) Let $\phi(r) = r^\theta$ with $\theta \in (0, \alpha)$. It is clear that (a) in Condition (ii) of Theorem 3.1 holds. On the other hand, it is obvious that Φ_1 and Φ_2 satisfy the conditions in (b); moreover, in this case, since $\beta > \theta/\alpha$, for any $a > 1$,

$$\sum_{k=1}^\infty \frac{2^{-\beta k}}{\phi(\Phi_1^{-1}(a2^k))} \leq c_1 \sum_{k=1}^\infty 2^{k(-\beta+\theta/\alpha)} < \infty.$$

Hence, the desired assertion follows from Theorem 3.1.

(2) Let $\phi(r) = r(1 - \log^{-\theta}(1/r))$. By some elementary calculations, we can verify that ϕ satisfies all the conditions in (a). On the other hand, we also can check that $\Phi_1(r)$ and $\Phi_2(r)$ are non-increasing on $(0, r_0]$ for some $r_0 > 0$ small enough. Let $\varepsilon_k = \Phi_1^{-1}(2^{k+1})$. Recall that

$$\Phi_1(r) \simeq r^{-\alpha}\log^{-\theta-1}\bigg(\frac{1}{r}\bigg).$$

Then, there is a constant $c_1 > 0$ such that

$$\varepsilon_k \geq c_1(2^{k+1})^{-1/\alpha}\log^{-(\theta+1)/\alpha}(2^{k+1}).$$

Thus, for $\beta > 1/\alpha$,

$$\sum_{k=1}^\infty \frac{2^{-\beta k}}{\varepsilon_k(1-\log^{-\theta}(1/\varepsilon_k))} \leq c_2 \sum_{k=0}^\infty \frac{2^{-\beta k}}{(2^{k+1})^{-1/\alpha}\log^{-(\theta+1)/\alpha}(2^{k+1})} < \infty.$$

With aid of all the estimates above, we can get the desired assertion by Theorem 3.1. □

We note that it is possible to extend the statements above to SDEs with multiplicative Lévy noises by [8, Theorem 1.2] and its proof. The details are left to readers. To close this section, we present the following example.

Example 3.3 Consider the following SDE

$$dX_t = b(X_t)\,dt + dZ_t,$$

where $(Z_t)_{t\geq 0}$ is a pure jump Lévy process such that its Lévy measure

$$\frac{c_1}{|z|^{d+\alpha}}\mathbb{1}_{\{0<z_1\leq 1\}}(dz) \leq \nu(dz) \leq \frac{c_2}{|z|^{d+\alpha}}\,dz$$

for some $c_1, c_2 > 0$ and $\alpha \in (0,2)$. Suppose that there is a constant $C > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\langle b(x), x\rangle \leq C(1+|x|)^2$$

and

$$\langle b(x)-b(y), x-y\rangle \leq C((1+|x|)^\alpha + (1+|y|)^\alpha)|x-y|^2.$$

Then, we have

(1) for any $\alpha \in (0,1]$ and $0 < \theta_1 < \theta_2 < \alpha$, there exists a constant $C > 0$ such that for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 0$,

$$\limsup_{y\to x}\frac{|P_tf(x)-P_tf(y)|}{|x-y|^{\theta_1}} \leq C(1+|x|)^{\theta_2}(x)(t\wedge 1)^{-\theta_1/\alpha}\|f\|_\infty.$$

(2) for any $\alpha \in (1,2)$, $\theta_1 > 0$ and $\theta_2 > 1$, there exists a constant $C > 0$ such that for any $x \in \mathbb{R}^d$, $f \in B_b(\mathbb{R}^d)$ and $t > 0$,

$$|\nabla P_tf(x)| \leq C(1+|x|)^{\theta_2}\left(\frac{\log^{1+\theta_1}(2/(t\wedge 1))}{t\wedge 1}\right)^{1/\alpha}\|f\|_\infty.$$

Proof With the lower bound of Lévy measure ν, by [9, Example 1.2], we can take $J(r) = c_0r^{-\alpha}$. Let $W(x) = (1+|x|)^\alpha$.

(1) By taking $\theta = \theta_1$ and $\beta = \theta_2/\alpha$, we can get the desired assertion by Corollary 3.2 (1), where we used the upper bound of Lévy measure ν and the assumptions on the drift term $b(x)$.

(2) Similar to (1), the assertion follows from Corollary 3.2 (2) by setting $\theta = \theta_1$ and $\beta = \theta_2/\alpha$. □

Acknowledgements The authors would like to thank the two referees for numerous corrections.

References

[1] Bertoldi, M., Fornaro, S.: Gradient estimates in parabolic problems with unbounded coefficients. *Studia Math.*, **165**, 221–254 (2004)

[2] Bertoldi, M., Lorenzi, L.: Estimates for the derivatives for parabolic operators with unbounded coefficients. *Trans. Amer. Math. Soc.*, **357**, 2627–2664 (2005)

[3] Cerrai, S.: Second Order PDEs in Finite and Infinite Dimension. A Probabilistic Approach, Lecture Notes in Math., vol. 1762, Springer, Berlin, 2001

[4] Cranston, M.: A probabilistic approach to gradient estimates. *Canad. Math. Bull.*, **35**, 46–55 (1992)

[5] Da Prato, G., Priola, E.: Gradient estimates for SDEs without monotonicity type conditions. *J. Differential Equations*, **265**, 1984–2012 (2018)

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

[7] Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, First Ed., North-Holland, Amsterdam, 1981

[8] Liang, M., Wang, J.: Gradient estimates and ergodicity for SDEs driven by multiplicative Lévy noises via coupling. *Stochastic Process. Appl.*, **130**, 3053–3094 (2020)

[9] Luo, D., Wang, J.: Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises. *Stochastic Process. Appl.*, **129**, 3129–3173 (2019)

[10] Priola, E., Wang, F. Y.: Gradient estimates for diffusion semigroups with singular coefficients. *J. Funct. Anal.*, **236**, 244–264 (2006)

[11] Stroock, D. W., Varadhan, S. R. S.: Multidimensional Diffusion Processes, Springer, Berlin, 1979

[12] Wang, F. Y.: Gradient estimates on $\mathbb{R}^d$. *Canad. Math. Bull.*, **37**, 560–570 (1994)

[13] Wang, F. Y., Xu, L., Zhang, X. C.: Gradient estimates for SDEs driven by multiplicative Lévy noises. *J. Funct. Anal.*, **269**, 3135–3219 (2015)

[14] Zhang, X. C.: Stochastic flows and Bismut formulas for stochastic Hamiltonian systems. *Stochastic Process. Appl.*, **120**, 1929–1949 (2010)

[15] Zhang, X. C.: Derivative formula and gradient estimate for SDEs driven by α-stable processes. *Stochastic Process. Appl.*, **123**, 1213–1228 (2013)