GRADIENT ESTIMATES FOR SDES DRIVEN BY MULTIPLICATIVE LÉVY NOISE

FENG-YU WANG, LIHU XU, XICHENG ZHANG

Abstract. Gradient estimates are derived for the semigroup associated to a class of stochastic differential equations driven by multiplicative Lévy noise, where the noise is realized as a subordination of the Brownian motion. In particular, the estimates are sharp for α-stable type noises. To derive these estimates, a new derivative formula is established for the semigroup by using the Malliavin calculus and a finite-jump approximation argument.

1. Introduction

Consider the following stochastic differential equation (abbreviated as SDE):

\[dX_t = b_t(X_t)dt + \sigma_t(X_t)dL_t,\]  

(1.1)

where \(L_t\) is a Lévy process, and

\[b : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d, \quad \sigma : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d\]

are continuous such that

\[
\max \{\langle \nabla v, b_s \rangle, \| \nabla \sigma_s \|_{H.S.}^2 \} \leq K_s |v|^2, \quad s \geq 0, \quad v \in \mathbb{R}^d
\]

(1.2)

holds for some positive function \(K \in C([0, \infty))\). Then for any \(x \in \mathbb{R}^d\), (1.1) has a unique solution \(\{X_t(x)\}_{t \geq 0}\) starting from \(x\). We aim to investigate the gradient estimate of \(P_t\):

\[P_t f(x) := \mathbb{E}f(X_t(x)), \quad t \geq 0, \quad f \in \mathcal{B}_b(\mathbb{R}^d),\]

where \(\mathcal{B}_b(\mathbb{R}^d)\) denotes the space of all bounded Borel measurable functions on \(\mathbb{R}^d\).

When \(b_t(x) = Ax\) for a matrix \(A\) and \(\sigma_t(x) = \text{Id}\), the gradient estimate of type

\[|\nabla P_t f| \leq \|f\|_{\infty} \varphi(t), \quad t > 0, \quad f \in \mathcal{B}_b(\mathbb{R}^d)\]

has been derived in [7] using lower bound condition of the Lévy measure, and in [6] for \(A = 0\) using asymptotic behaviours of the symbol of \(L_t\); see also [8] for a derivative formula by using coupling through the Mecke formula.

Recently, a time-change argument was introduced in [10] to establish the Bismut derivative formula of \(P_t\) for the case that \(L_t\) is the \(\alpha\)-stable process, \(\sigma_t\) is invertible and independent of the space variable (i.e. the noise is additive), and \(\nabla b_t\) is uniformly bounded. In particular, this derivative formula implies that for any \(p > 1\) there exists a constant \(C(p) > 0\) such that (see [10] Theorem 1.1)

\[|\nabla P_t f| \leq \frac{C(p)}{1 + t^\alpha} (P_t |f|^p)^{1/p}, \quad f \in \mathcal{B}_b(\mathbb{R}^d), t > 0.\]  

(1.3)

Using this time-change argument and the coupling method, Harnack inequalities are then established in [9].

Key words and phrases. Gradient estimate, derivative formula, Lévy process, SDEs, time-change.
In this paper, we intend to extend the gradient estimate \((1.3)\) for more general symmetric Lévy process \(L_t\) and space-dependent \(\sigma_t\) (i.e. the noise is multiplicative). Notice that if \(b\) and \(\sigma\) are independent of \(t\), then the generator of \(P_t\) is given by
\[
\mathcal{L}f(x) = b(x) \cdot \nabla f(x) + \text{P.V.} \int_{\mathbb{R}^d} [f(x + \sigma_t(y)) - f(x)]\nu(dy),
\]
where \(\nu(dy)\) is the Lévy measure of \(L_t\), and \text{P.V.} stands for the Cauchy principal value.

From now on, we let \(L_t\) be a symmetric and rotationally invariant Lévy process, which can be formulated as subordination of the Brownian motion. More precisely, let \(\{W_t\}_{t \geq 0}\) be a \(d\)-dimensional Brownian motion and \(\{S_t\}_{t \geq 0}\) an independent subordinator process associated with a Bernstein function \(B\) with \(B(0) = 0\); i.e. \(\{S_t\}_{t \geq 0}\) is an increasing process with stationary independent increments such that
\[
\mathbb{E}e^{-uS_t} = e^{-uB(u)}, \quad u \geq 0, t \geq 0.
\]
Then \(L_t := W_{S_t}\) is a Lévy process with symbol \(\Psi(\xi) := B(|\xi|^2)\) (see e.g. [11]). In particular, if \(B(u) = u^\alpha\) for some constant \(\alpha \in (0, 2)\), then \(L_t\) is called the \(\alpha\)-stable process. So, the equation \((1.1)\) is now reduced to
\[
dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_{S_t}, \quad (1.4)
\]
We assume
\[
(A) \quad \nabla b_t(x), \nabla \sigma_t(x) \text{ are locally (uniformly in } t\text{) Lipschitz continuous in } x, \text{ such that}
\]
\[
\|\nabla b_t\|_t := \sup_{x \in \mathbb{R}^d} \|\nabla b_t(x)\| < +\infty, \quad \|\nabla \sigma_t\|_t := \sup_{x \in \mathbb{R}^d} \|\nabla \sigma_t(x)\| < +\infty, \quad t > 0, \quad (1.5)
\]
where \(\| \cdot \|\) stands for the operator norm, and for some positive increasing function \(c\) and some constant \(m \geq 0\)
\[
\sup_{x \in \mathbb{R}^d} |\sigma_t^{-1}(x)| \leq c_t(1 + |x|^m), \quad t \geq 0, x \in \mathbb{R}^d. \quad (1.6)
\]

Obviously, \((1.5)\) implies \((1.2)\) and hence, the existence and uniqueness of the solution. Before stating our main result, let us briefly recall the main argument introduced in [10] for the study of the additive noise case, and explain why this argument is no-longer valid in the multiplicative case.

In the additive noise case where \(\sigma_t(x) = \sigma_t\) is independent of \(x\), for a fixed path \(\ell\) of \(S\), we may reformulate the equation \((1.4)\) as
\[
dX_t^\ell = b_t(X_t^\ell)dt + \sigma_t dW_{\ell_t},
\]
To establish a derivative formula for \(P_t^\ell f := \mathbb{E} f(X_t^\ell)\), consider the following regularization of \(\ell\) for \(\epsilon > 0\):
\[
\ell^\epsilon_t := \frac{1}{\epsilon} \int_{t-\epsilon}^{t} \ell_s ds + \epsilon t, \quad t \geq 0,
\]
and the associated stochastic differential equation
\[
dX_t^\epsilon = b_t(X_t^\epsilon)dt + \sigma_t dW_{\ell^\epsilon_t}, \quad X_0^\epsilon = X_0^\ell.
\]
Then for any \(t > 0\), as \(\epsilon \downarrow 0\) we have \(\ell_t^\epsilon \downarrow \ell_t\) and \(X_t^\epsilon \rightarrow X_t^\ell, \nabla X_t^\epsilon \rightarrow \nabla X_t^\ell\) in \(L^p(\mathbb{F})\) for any \(p > 1\). Since \(\ell_t^\epsilon\) is absolutely continuous, \(X_t^\epsilon\) is indeed a diffusion process, so that an existing Bismut type derivative formula applies. Therefore, letting \(\epsilon \downarrow 0\) we derive a derivative formula for \(P_t^\ell\).

Now, coming back to the multiplicative noise case, we consider
\[
dX_t^\ell = b_t(X_t^\ell)dt + \sigma_t(X_t^\ell) dW_{\ell_t},
\]
and the corresponding approximation equation
\[ dX^\varepsilon_t = b_t(X^\varepsilon_t)dt + \sigma_t(X^\varepsilon_t)dW^\varepsilon_t, \quad X^\varepsilon_0 = X^\ell_0. \]

Since we only have weak convergence \( d\ell_t \to d\ell \) as Lebesgue-Stieltjes measures, but the function \( s \mapsto \sigma_s(X^\ell_{s-}) \) is however discontinuous, the assertion that \( X^\varepsilon_t \to X^\ell_t \) is no-longer true! A simple counter-example is that \( b_t(x) = 0, \sigma_t(x) = \sqrt{1 + |x|^2} \) and \( \ell = 1_{[1,\infty)} \), since in this case for \( X^\ell_0 = 0 \) we have \( X^\ell_1 = W_1 \) such that \( \mathbb{E}|X^\ell_1|^2 = d \), but
\[ d|X^\ell_1|^2 = (1 + |X^\ell_1|^2)d\ell_t + 2\langle X^\ell_t, \sqrt{1 + |X^\ell_t|^2}dW^\ell_t \rangle, \]
such that
\[ \mathbb{E}|X^\ell_1|^2 = e^{d\ell_t} \int_0^1 d\sqrt{e^{d\ell_t}d\ell_t} = e^{d\ell_t} - 1 \geq d^2 - 1 > d. \]

Therefore, in this case we have to introduce different arguments.

Fortunately, by using a finite-jump approximation (i.e. approximating \( \ell \) by those of finite many jumps in finite intervals) and the Malliavin calculus, we are able to establish a nice derivative formula for \( P^\ell_t \) (see Theorem 3.2 below), which in turn imply the following main result of the paper.

**Theorem 1.1.** Assume (A) and let \( P_t \) be the semigroup associate to equation (1.4).

(i) For fixed \( R > 0 \), let \( \tau := \inf \{ t : S_t > R \} \). If \( \mathbb{E}S^{1/2}_{1\wedge \tau} < \infty \), then for any \( f \in \mathcal{B}_b(\mathbb{R}^d), v \in \mathbb{R}^d \), we have
\[
\nabla_v P_t f = \mathbb{E} \left[ f(X_t) \frac{1}{S_{1\wedge \tau}} \left( \int_0^{1\wedge \tau} \langle \sigma^{-1}_s(X_{s-})\nabla_v X_{s-}, dW_s \rangle + \int_0^{1\wedge \tau} \text{Tr}(\sigma^{-1}_s\nabla_v X_{s-}\sigma_s)(X_{s-})dS_s \right) \right. \\
+ \left. \int_0^{1\wedge \tau} \int_{\mathbb{R}^d} \langle \sigma^{-1}_s(X_{s-})\nabla_v X_{s-}\sigma_s(X_{s-})y, y \rangle N(ds, dy) \right],
\]

where \( N \) is the random measure associated to \( W_{S_t} \), i.e.,
\[
N(t, \Gamma) := \sum_{s \in [0,t]} I_{\{\Delta W_s > 1\}}(\Gamma) = \mathcal{B}_b(\mathbb{R}^d).
\]

(ii) For any \( p > 1 \), there exists a constant \( C > 0 \) such that
\[
|\nabla_v P_t f(x)| \leq C(1 + |x|^m)(P_t |f|^p)^{1/p}(x) \left( \mathbb{E} \left[ \frac{1}{S_{1\wedge \tau}^{p/(2(p-1))}} \wedge 1 \right] \right)^{(p-1)/p}
\]

holds for all \( t > 0, x \in \mathbb{R}^d \) and \( f \in \mathcal{B}(\mathbb{R}^d) \). Consequently, if \( B(u) \geq c u^2 \) holds for \( u \geq u_0 \), where \( \alpha \in (0,2) \) and \( c, u_0 > 0 \) are constants, then for any \( p > 1 \) there exists a constant \( C > 0 \) such that for all \( t > 0, x \in \mathbb{R}^d \) and \( f \in \mathcal{B}(\mathbb{R}^d) \),
\[
|\nabla_v P_t f(x)| \leq \frac{C(1 + |x|^m)}{(t \wedge 1)^{\alpha/2}} (P_t |f|^p)^{1/p}(x).
\]

We remark that according to known derivative estimates of the \( \alpha \)-stable process, the gradient estimate (1.10) is sharp in short-time for \( S_t \), being the \( \alpha \)-stable subordinator (i.e. \( B(u) = u^\alpha \)). Moreover, (1.10) recovers (1.3) in the additive noise case by taking \( m = 0 \).

The remainder of the paper is organized as follows. In Section 2 we briefly recall the integration by parts formula in the Malliavin calculus and present some lemmas on finite-jump approximations. These are then used in Section 3 to establish a derivative formula for \( P^\ell_t \). Finally, in Section 4 we present explicit gradient estimates of \( P^\ell_t \) and a complete proof of the above main result.
2. Preliminaries

Let \((\mathcal{W}, \mathcal{H}, \mu^\mathcal{W})\) be the classical Wiener space, i.e., \(\mathcal{W}\) is the space of all continuous functions \(\omega : \mathbb{R}_+ \to \mathbb{R}^d\) with \(\omega_0 = 0\), \(\mathcal{H}\) is the Cameron-Martin space consisting of all absolutely continuous functions \(h \in \mathcal{W}\) with
\[
\|h\|_\mathcal{H}^2 := \int_0^\infty |\dot{h}_s|^2\,ds < +\infty,
\]
and \(\mu^\mathcal{W}\) is the Wiener measure so that the coordinate process
\[
W_t(\omega) := \omega_t
\]
is a \(d\)-dimensional standard Brownian motion.

Below we recall some basic notions about the Malliavin calculus (cf. [3, 4]). Let \(U\) be a real separable Hilbert space. Let \(C(U)\) be the class of all \(U\)-valued smooth cylindrical functionals on \(\Omega\) with the form:
\[
F = \sum_{i=1}^m f_i(W(h_1), \ldots, W(h_n))u_i,
\]
where \(n \geq 1, f_i \in C_0^\infty(\mathbb{R}^n), u_i \in U, h_1, \ldots, h_n \in \mathcal{H}\) and
\[
W(h) = \int_0^\infty \dot{h}_s dW_s.
\]
The Malliavin derivative of \(F\) is defined by
\[
DF := \sum_{i=1}^m \sum_{j=1}^n (\partial_j f_i)(W(h_1), \ldots, W(h_n))u_i \otimes h_j \in U \otimes \mathcal{H}.
\]
It is well known that the operator \((D, C(U))\) is closable from \(L^2(\mathcal{W}; U)\) to \(L^2(\mathcal{W}; U \otimes \mathcal{H})\) (cf. [4, p.26, Proposition 1.2.1]). The closure is denoted by \((D, D(D))\). The dual operator \(\delta\) of \(D\) (also called divergence operator) is defined by
\[
\mathbb{E}\langle DF, U \rangle_\mathcal{H} = \mathbb{E}(F\delta(U)), \quad F \in D(D), \quad U \in D(\delta).
\]
Notice that the following divergence formula holds: for any \(F \in D(D)\) and \(U \in D(\delta)\),
\[
\delta(FU) = F\delta U + \langle DF, U \rangle_\mathcal{H}.
\]
(2.1)
Let \(\mathcal{F}^\mathcal{W}_t := \sigma\{W_s : s \leq t\}\) be the natural filtration associated to \(W_t\). If \(U_t(\omega) = \int_0^t u_s(\omega)\,ds\), where \(u_s\) is an \(\mathcal{F}^\mathcal{W}_t\)-adapted process on \(\mathbb{R}^d\) with \(\mathbb{E}\int_0^\infty |u_s|^2\,ds < \infty\), then \(u \in D(\delta)\) and
\[
\delta(U) = \int_0^\infty \langle u_s, dW_s \rangle.
\]
(2.3)
Next, for \(\beta : [0, \infty) \to [0, \infty)\) being an absolutely continuous increasing function with \(\beta_0 = 0\) and locally bounded derivative \(\dot{\beta}_t\), set
\[
W_t^\beta := \int_0^\beta \dot{\beta}_s dW_s, \quad t \geq 0.
\]
(2.2)
It is easy to see that \(t \mapsto W_t^\beta\) is a process with independent increments and
\[
\mathcal{L}_t^\beta := \mathbb{E}|W_t^\beta|^2 = \int_0^t |\dot{\beta}_s|^2\,ds.
\]
(2.3)
Let $\mathcal{S}$ be the space of all purely jump càdlàg and increasing functions, i.e. the path space of $(S_t)_{t \geq 0}$. For any $\ell \in \mathcal{S}$ and $s > 0$, we shall denote
\[
\Delta \ell_s := \ell_s - \ell_{s-}.
\]

We recall the following Burkholder’s inequality (cf. [10] Theorem 2.3]).

**Lemma 2.1.** Assume that $\xi_t$ is an $\mathcal{F}_{\ell_t}^{\mathbb{N}}$-adapted càdlàg $\mathbb{R}^d$-valued process and satisfies that for some $p > 0$,
\[
\mathbb{E} \left( \int_0^T |\xi_{s-}|^p d\lambda_{\ell_s} \right) < +\infty, \quad \forall T > 0.
\]

Then there exists a constant $C_p > 0$ such that for all $T > 0$,
\[
\mathbb{E} \left( \sup_{t \in [0, T]} \left| \int_0^t \xi_{s-} dW_{\ell_s}^\beta \right|^p \right) \leq C_p \mathbb{E} \left( \int_0^T |\xi_{s-}|^p d\lambda_{\ell_s} \right).
\]

Finally, for fixed $\ell \in \mathcal{S}$ and $\varepsilon > 0$, define
\[
\ell^\varepsilon_t := \sum_{s \in \ell_t} \Delta \ell_s 1_{[\Delta \ell_s > \varepsilon]} , \quad t \geq 0.
\]

We present below a key lemma about the approximation $\ell^\varepsilon_t$ for later use.

**Lemma 2.2.** Let $\xi_t^\varepsilon$ be an $\mathcal{F}_{\ell_t}^{\mathbb{N}}$-adapted càdlàg $\mathbb{R}^d$-valued process with
\[
\sup_{t \in [0, T]} \mathbb{E} |\xi_t^\varepsilon|^2 < +\infty, \quad T > 0.
\]

Then we have
\[
\lim_{\varepsilon \to 0} \sup_{t \in [0, T]} \mathbb{E} \left| \int_0^t \xi_{s-}^\varepsilon dW_{\ell_s}^\beta - \int_0^t \xi_{s-} dW_{\ell_s}^\beta \right|^2 = 0,
\]
where $W_t^\beta$ is defined by (2.2).

**Proof.** Without loss of generality, we assume $T = 1$. For $t \in [0, 1]$, set
\[
t_n := \lfloor nt \rfloor / n, \quad t^+_n := (\lfloor nt \rfloor + 1) / n.
\]
Fix $\varepsilon \in (0, 1)$. Since $\mathcal{F}_{\ell_t}^{\mathbb{N}} \subset \mathcal{F}_{\ell_t}^{\mathbb{N}}$, by the left-continuity of $s \mapsto \xi_s^\varepsilon$ and (2.5), we have
\[
\mathbb{E} \left| \int_0^t \xi_{s-}^\varepsilon dW_{\ell_s}^\beta - \int_0^t \xi_{s-} dW_{\ell_s}^\beta \right|^2 = \lim_{n \to \infty} \mathbb{E} \left| \int_0^t \xi_{s-}^\varepsilon dW_{\ell_s}^\beta - \int_0^t \xi_{s-}^\varepsilon dW_{\ell_s}^\beta \right|^2
\]
\[
= \lim_{n \to \infty} \mathbb{E} \sum_{k=0}^{\lfloor nt \rfloor} \left( W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta - W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta + W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta \right)^2
\]
\[
= \lim_{n \to \infty} \sum_{k=0}^{\lfloor nt \rfloor} \mathbb{E} \xi_{k/n}^\varepsilon \mathbb{E} \left( W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta - W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta + W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta \right)^2
\]
\[
\leq C \lim_{n \to \infty} \sum_{k=0}^{n-1} \mathbb{E} \left( W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta - W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta + W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta \right)^2,
\]
where we have used the independence of $\xi_{k/n}^\varepsilon$ and $W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta - W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta + W_{\ell^\varepsilon_{\lfloor nt \rfloor}}^\beta$. 

Notice that if \( \ell_{k/n} \geq \ell_{(k+1)/n}^e \), then

\[
\mathbb{E}\left( W_{\ell_{(k+1)/n}^e}^\beta - W_{\ell_{k/n}^e}^\beta - W_{\ell_{(k+1)/n}^e}^\beta + W_{\ell_{k/n}^e}^\beta \right)^2 = \mathbb{E}\left( \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} \beta_s dW_s - \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} \hat{\beta}_s dW_s \right)^2
\]

\[
= \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} |\hat{\beta}_s|^2 ds + \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} |\beta_s|^2 ds = \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta + \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta,
\]

and if \( \ell_{k/n} < \ell_{(k+1)/n}^e \), then

\[
\mathbb{E}\left( W_{\ell_{(k+1)/n}^e}^\beta - W_{\ell_{k/n}^e}^\beta - W_{\ell_{(k+1)/n}^e}^\beta + W_{\ell_{k/n}^e}^\beta \right)^2 = \mathbb{E}\left( \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} \beta_s dW_s - \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} \hat{\beta}_s dW_s \right)^2
\]

\[
= \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} |\hat{\beta}_s|^2 ds + \int_{\ell_{k/n}^e}^{\ell_{(k+1)/n}^e} |\beta_s|^2 ds = \lambda_{\ell_{k/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta + \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta.
\]

Hence,

\[
\sum_{k=0}^{n-1} \mathbb{E}\left( W_{\ell_{(k+1)/n}^e}^\beta - W_{\ell_{k/n}^e}^\beta - W_{\ell_{(k+1)/n}^e}^\beta + W_{\ell_{k/n}^e}^\beta \right)^2 = I_1(n, \varepsilon) + I_2(n, \varepsilon),
\]

where

\[
I_1(n, \varepsilon) := \sum_{k=0}^{n-1} \left( \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta + \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta \right) 1_{\ell_{k/n}^e < \ell_{(k+1)/n}^e}
\]

\[
= \int_{\ell_{n}^e}^{\ell_{n}^e} \left( \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} + \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} \right) 1_{\ell_{n}^e < \ell_{n}^e} d\ell_s,
\]

\[
I_2(n, \varepsilon) := \sum_{k=0}^{n-1} \left( \lambda_{\ell_{k/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta + \lambda_{\ell_{(k+1)/n}^e}^\beta - \lambda_{\ell_{k/n}^e}^\beta \right) 1_{\ell_{k/n}^e < \ell_{(k+1)/n}^e}
\]

\[
= \int_{\ell_{n}^e}^{\ell_{n}^e} \left( \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} + \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} \right) 1_{\ell_{n}^e < \ell_{n}^e} d\ell_s.
\]

Noticing that as \( n \to \infty \),

\[
\ell_{n}^e \uparrow \ell_{s}^e, \quad \ell_{s}^e \downarrow \ell_{s}, \quad \ell_{n}^e \uparrow \ell_{s}^e, \quad \ell_{s}^e \downarrow \ell_{s},
\]

and as \( \varepsilon \downarrow 0 \),

\[
\ell_{s}^e \uparrow \ell_{s}, \quad \ell_{s}^e \downarrow \ell_{s},
\]

by the dominated convergence theorem, we have

\[
\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} I_1(n, \varepsilon) = \lim_{\varepsilon \downarrow 0} \int_{\ell_{n}^e}^{\ell_{n}^e} \left( \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} + \frac{\lambda_{\ell_{n}^e}^\beta - \lambda_{\ell_{n}^e}^\beta}{\ell_{n}^e - \ell_{n}} \right) 1_{\ell_{n}^e < \ell_{n}^e} d\ell_s
\]

\[
= 2 \int_{\ell_{s}^e}^{\ell_{s}^e} \left( \frac{\lambda_{\ell_{s}^e}^\beta - \lambda_{\ell_{s}^e}^\beta}{\ell_{s}^e - \ell_{s}} \right) 1_{\ell_{s}^e < \ell_{s}} d\ell_s = 0,
\]
and
\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} I_2(n, \varepsilon) \leq \lim_{\varepsilon \to 0} \int_0^1 \left( \frac{\lambda_s^b - \lambda_{s-}^b}{\ell_s - \ell_{s-}} + \frac{\lambda_s^p - \lambda_{s-}^p}{\ell_s - \ell_{s-}} \right) 1_{\ell_s < \ell_n} d\ell_s = 0. \quad (2.10)
\]
Combining (2.7)-(2.10), we obtain the desired limit. \qed

3. Derivative formula for \( P_t^\varepsilon \)

In this section, we fix an \( \ell \in \mathbb{S} \) and consider the following SDE:
\[
X_t^\varepsilon(x) = x + \int_0^t b_s(X_s^\varepsilon(x))ds + \int_0^t \sigma_s(X_s^\varepsilon(x))dW_s,
\]
for \( b \) and \( \sigma \) satisfying (A). Under \([1,5]\), it is well-known that \( \{X_t^\varepsilon(), t \geq 0\} \) forms a \( C^1 \)-stochastic flow (cf. \([5]\) p.305, Theorem 39). Let \( \nabla X_t^\varepsilon := (\partial_j(X_t^\varepsilon(x)))_{ij} \) be the derivative matrix. Then
\[
\nabla_s X_t^\varepsilon := (\nabla_s(X_t^\varepsilon(x)))_{1 \leq i \leq d} = (\nabla X_t^\varepsilon)v, \quad v \in \mathbb{R}^d,
\]
and
\[
\nabla_s X_t^\varepsilon = v + \int_0^t \nabla_{s, X_s^\varepsilon} b_s(X_s^\varepsilon)ds + \int_0^t \nabla_{s, X_s^\varepsilon} \sigma_s(X_s^\varepsilon)dW_s.
\]

We first prepare the following lemma.

**Lemma 3.1.** Let \( \alpha_t = \sup_{x \in [0,t]} |b_s(0)| \) and \( \gamma_t = \sup_{x \in [0,t]} \|\sigma_s(0)\| \). Then, for any \( p \geq 2 \) there exists a constant \( c(p) > 0 \) such that for all \( t \geq 0 \) and \( x \in \mathbb{R}^d \),
\[
\mathbb{E} \left( \sup_{x \in [0,t]} |X_t^\varepsilon(x)|^p \right) \leq c(p) \left( |x| + \alpha_t + t + \sqrt{t} \right)^p \exp \left( c(p)(\|\nabla b\|_p t^p + \|\nabla \sigma\|_p \varepsilon) \right), \quad (3.1)
\]
and
\[
\mathbb{E} \left( \sup_{x \in [0,t]} \|\nabla X_t^\varepsilon(x)\|^p \right) \leq c(p) \exp \left( c(p)(\|\nabla b\|_p t^p + \|\nabla \sigma\|_p \varepsilon) \right). \quad (3.2)
\]
Moreover, for any \( T > 0 \) and \( x \in \mathbb{R}^d \),
\[
\lim_{\varepsilon \to 0} \sup_{x \in [0,T]} \mathbb{E}[X_T^\varepsilon(x) - X_T^\varepsilon(x)]^2 = 0, \quad \lim_{\varepsilon \to 0} \sup_{x \in [0,T]} \mathbb{E}[\nabla X_T^\varepsilon(x) - \nabla X_T^\varepsilon(x)]^2 = 0. \quad (3.3)
\]

**Proof.** (1) We simply use \( X_t \) to denote \( X_t^\varepsilon(x) \). Since
\[
|X_t|^p \leq 3^{p-1}|x| + 3^{p-1} \left| \int_0^t b_s(X_s)ds \right|^p + 3^{p-1} \left| \int_0^t \sigma_s(X_s) dW_s \right|^p
\]
and
\[
|b_s(x)| \leq \|\nabla b\|_p |x| + \alpha_s, \quad \|\sigma_s(x)\| \leq \|\nabla \sigma\|_p |x| + \gamma_s,
\]
by Lemma[2.1] and Hölder’s inequality, there exists a constant \( c(p) > 0 \) such that
\[
\mathbb{E} \left( \sup_{x \in [0,t]} |X_t|^p \right) \leq c(p) \left( |x| + t \alpha_t + t \gamma_t \right)^p + c(p) \|\nabla b\|_p t^{p-1} \int_0^t \mathbb{E} \left( \sup_{x \in [0,s]} |X_s|^p \right) ds
\]
\[
+ c(p) \|\nabla \sigma\|_p \varepsilon^{p-1} \int_0^t \mathbb{E} \left( \sup_{x \in [0,s]} |X_s|^p \right) d\ell_s.
\]
By Gronwall’s lemma, this implies (3.1).

Similarly, noticing that
\[
\|\nabla X_t\|_p \leq 3^{p-1} + 3^{p-1} \|\nabla b\|_p \left| \int_0^t \|\nabla X_s\| ds \right|^p + 3^{p-1} \left| \int_0^t \nabla \sigma_s(X_{s-}) \nabla X_s^\varepsilon dW_s \right|^p
\]
by Lemma 2.1 again and Hölder’s inequality, there exists a constant \(c(p) > 0\) such that
\[
\mathbb{E}\left( \sup_{v \in \mathcal{H}} \|\nabla X_t^\ell\|^p \right) \leq c(p) + c(p) \|\nabla b\|^p t^{p-1} \int_0^t \mathbb{E}\left( \sup_{v \in \mathcal{H}} \|\nabla X_s^\ell\|^p \right) ds
+ c(p) \|\nabla \sigma\|^p \ell_t^{p-1} \int_0^t \mathbb{E}\left( \sup_{v \in \mathcal{H}} \|\nabla X_s^\ell\|^p \right) d\ell_s,
\]
which implies (3.2) by Gronwall’s inequality again.

(2) Let \(Y_t^\ell := X_t^\ell - X_t^\ell\). Then for all \(t \leq T\),
\[
\mathbb{E}\|Y_t^\ell\|^2 \leq 3\|\nabla b\|^2 T \int_0^t \mathbb{E}\|Y_s^\ell\|^2 ds + 3\|\nabla \sigma\|^2 \int_0^t \mathbb{E}\|Y_s^\ell\|^2 d\ell_s + 3R_t^\ell,
\]
where
\[
R_t^\ell := \mathbb{E}\left| \int_0^t \sigma_s(S_{x_{s-}})dW_t^\ell - \int_0^t \sigma_s(S_{x_{s-}})dW_{t_s} \right|^2.
\]
By Gronwall’s lemma and Lemma 2.2, we obtain
\[
\lim_{\varepsilon \downarrow 0, \varepsilon \in [0,1]} \mathbb{E}\|Y_t^\ell\|^2 \leq 3e^{3\|\nabla b\|^2 T^2 + 3\|\nabla \sigma\|^2 \ell T} \lim_{\varepsilon \downarrow 0, \varepsilon \in [0,1]} \sup_{\ell \in [0, T]} R_t^\ell = 0.
\]
The proof is finished. □

We now prove the following derivative formula:

**Theorem 3.2.** Assume (1.5) and (1.6). Let \(\beta\) be an increasing \(C^1\)-function with \(\beta_0 = 0\). For any \(v \in \mathbb{R}^d\) and \(f \in \mathcal{B}_b(\mathbb{R}^d)\), we have for all \(t > 0\) with \(\beta_\ell > 0\),
\[
\nabla_v P_t^\ell f = \frac{1}{\beta_\ell} \mathbb{E}\left[ f(X_t^\ell) \left( \int_0^t \langle \sigma_{s-}^{-1}(S_{x_{s-}})\nabla_X X_{s-}, \; dW_t^\ell \rangle + \int_0^t \text{Tr}(\sigma_{s-}^{-1} \nabla_{S_{x_{s-}}} \sigma_s) (X_{s-}) d\beta_{t_s} \right. \\
+ \left. \sum_{s \in [0,t]} \langle \sigma_{s-}^{-1}(X_{s-}) \nabla_{S_{x_{s-}}} \sigma_s(X_{s-}) \Delta W_t^\ell, \Delta W_{t_s} \rangle \right) \right],
\]
where \(W^\ell\) is defined by (2.2).

**Proof.** By Lemma 3.1 (A) and an approximation argument, we may and do assume that \(f \in C^1_b(\mathbb{R}^d)\). (1) Let us first establish formula (3.4) for \(\ell \in \mathcal{S}\) with finite many jumps on any finite time interval. For the simplicity of notations, we shall drop the superscript “\(\ell\)” and variable “\(x\)”. Let \(\mathcal{S}\) solve the following matrix-valued ODE
\[
\mathcal{S}_t = I + \int_0^t \nabla b_s(X_s)\mathcal{S}_s ds.
\]
By the variation of constant formula, it is easy to see that
\[
\nabla_v X_t = \mathcal{S}_t \nabla_v + \mathcal{S}_t \int_0^t \mathcal{S}_s^{-1} \nabla_{S_{x_{s-}}} \sigma_s(X_{s-}) d\beta_{t_s}.
\]
Let \(0 = t_0 < t_1 < t_2 < \cdots < t_n < \cdots\) be the jump times of \(\ell\). Fix \(v \in \mathbb{R}^d\) and set \(h_0 = 0\). Define \(h \in \mathbb{H}\) recursively as follows: if \(s \in (t_{k-1}, t_k)\) for some \(k \geq 1\), then
\[
h_s := h_{t_{k-1}} + (\beta_s - \beta_{t_{k-1}}) \sigma^{-1}_{t_{k-1}}(X_{t_{k-1}}) \left( \nabla_{S_{x_{t_{k-1}}}} + \nabla_{S_{x_{t_{k-1}}}} h_{t_{k-1}} \right) (X_{t_{k-1}}) \Delta W_{t_{k-1}}.
\]
For this \(h\), it is standard to prove that (cf. [4])
\[
D_h X_t = \int_0^t \nabla D_h X_s b_s(X_s) ds + \int_0^t \nabla D_h X_s \sigma_s(X_s) dW_{t_s} + \int_0^t \sigma_s(X_s) d\beta_{t_s}.
\]
As above, one can write
\[ D_{t}X_{t} = \mathcal{J}_{t} \int_{0}^{t} \mathcal{J}_{s}^{-1} \nabla_{D_{t}X_{s}} \sigma_{s}(X_{s}) \, dW_{s} + \mathcal{J}_{t} \int_{0}^{t} \mathcal{J}_{s}^{-1} \sigma_{s}(X_{s}) \, dh_{s}, \tag{3.7} \]
Let us now use induction method to prove
\[ D_{t}X_{t} = \beta_{t} \nabla_{v} X_{t}, \quad t \geq 0. \tag{3.8} \]
First of all, for all \( t \in [0, t_{1}) \), by definition we have
\[ D_{t}X_{t} = 0, \quad \beta_{t} = 0, \]
and so (3.8) holds for all \( t \in (0, t_{1}) \). Suppose that (3.8) holds for all \( t \in [0, t_{m}) \). Then, by (3.6), (3.7) and induction hypothesis, we have for \( t \in (t_{m}, t_{m+1}) \),
\[
D_{t}X_{t} = \mathcal{J}_{t} \int_{0}^{t} \mathcal{J}_{s}^{-1} \nabla_{D_{t}X_{s}} \sigma_{s}(X_{s}) \, dW_{s} + \mathcal{J}_{t} \int_{0}^{t} \mathcal{J}_{s}^{-1} \sigma_{s}(X_{s}) \, dh_{s}
= \mathcal{J}_{t} \sum_{k=1}^{m} \mathcal{J}_{t_{k}}^{-1} \left( \nabla_{D_{t}X_{t_{k}}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}} + \sigma_{t_{k}}(X_{t_{k}}) \Delta h_{t_{k}} \right)
= \mathcal{J}_{t} \sum_{k=1}^{m} \mathcal{J}_{t_{k}}^{-1} \left( \beta_{t_{k}} \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}} + (\beta_{t_{k}} - \beta_{t_{k-1}}) \left( \nabla_{v} X_{t_{k}} + \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}} \right) \right)
= \mathcal{J}_{t} \sum_{k=1}^{m} \left( \beta_{t_{k}} - \beta_{t_{k-1}} \right) \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} + \beta_{t_{k}} \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}} \right) \quad (\because \beta_{t_{k}} = \beta_{t_{k-1}}).
\]
On the other hand, by (3.5) we have
\[
\mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} = \nabla_{v} X_{t_{k}} - \int_{0}^{t_{k}} \mathcal{J}_{s}^{-1} \nabla_{v} X_{s} \sigma_{s}(X_{s}) \, dW_{s} = \nabla_{v} X_{t_{k}} - \int_{0}^{t_{k}} \mathcal{J}_{s}^{-1} \nabla_{v} X_{s} \sigma_{s}(X_{s}) \, dW_{s},
\]
Hence, for all \( t \in (t_{m}, t_{m+1}) \), we further have
\[
D_{t}X_{t} = \mathcal{J}_{t} \sum_{k=1}^{m} \left( \beta_{t_{k}} - \beta_{t_{k-1}} \right) \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} + \beta_{t_{k}} \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}}
= \mathcal{J}_{t} \sum_{k=1}^{m} \left( \beta_{t_{k}} - \beta_{t_{k-1}} \right) \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} + \beta_{t_{k}} \mathcal{J}_{t_{k}}^{-1} \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}}
= \mathcal{J}_{t} \beta_{t_{m}} \nabla_{v} X_{t_{m}} \int_{0}^{t_{m}} \mathcal{J}_{s}^{-1} \nabla_{v} X_{s} \sigma_{s}(X_{s}) \, dW_{s} = \beta_{t} \nabla_{v} X_{t}.
\]
Thus, (3.8) is proven.
Now, if \( \beta_{t} > 0 \),
then by (3.8) and the integration by parts formula in the Malliavin calculus, we have
\[
\nabla_{v} \mathbb{E} f(X_{t}) = \mathbb{E} \left( \nabla f(X_{t}) \nabla v X_{t} \right) = \frac{1}{\beta_{t}} \mathbb{E} \left( \nabla f(X_{t}) D_{t}X_{t} \right) = \frac{1}{\beta_{t}} \mathbb{E} f(X_{t}) \delta(h), \tag{3.9}
\]
provided \( h \in \mathcal{D}(\delta) \). Notice that \( h \) defined in (3.6) is non-adapted because of the term \( \Delta W_{t_{k}} \). Neverethess, \( \delta(h) \) can be explicitly calculated as follows: Define
\[
h_{k}(s) = (\beta_{t \wedge t_{k-1}} - \beta_{t_{k-1}}) \sigma_{t_{k-1}}^{\frac{1}{2}}(X_{t_{k-1}}) \left( \nabla_{v} X_{t_{k-1}} + \nabla_{v} X_{t_{k}} \sigma_{t_{k}}(X_{t_{k}}) \Delta W_{t_{k}} \right), \quad s \geq 0.
\]
Noting that 

\[ h_k = h_k^{(0)} + \sum_{j=1}^{d} h_k^{(j)}(\Delta W_{\ell_k})_{j}, \]

where 

\[ h_k^{(0)}(s) := (\beta_{(s^\delta_{\ell_k-1})_{\Delta \ell_k}} - \beta_{\ell_k})\sigma_{-1}(X_{n-})\nabla_v X_{n-}, \]

\[ h_k^{(j)}(s) := (\beta_{(s^\delta_{\ell_k-1})_{\Delta \ell_k}} - \beta_{\ell_k})\left(\sigma_{-1}(X_{n-})\nabla_v x_{k} - \sigma_{\ell_k}(X_{n-})\right), \]

are adapted, we see from (1.6) and (3.1), (3.2) that \( h \in \mathcal{D}(\delta) \); and by (2.1),

\[ \delta(h) = \sum_{k,\ell_k \leq t} \delta(h_k) + \sum_{k,\ell_k \leq t} \left(\delta(h_k^{(0)}) + \sum_{j=1}^{d} \delta(h_k^{(j)})(\Delta W_{\ell_k})_{j} + D_{h_k^{(j)}}(\Delta W_{\ell_k})_{j}\right) = \sum_{k,\ell_k \leq t} \left(\langle \sigma_{-1}(X_{n-})\nabla_v X_{n-}, \Delta W_{\ell_k}^{\beta}\rangle + \langle \sigma_{-1}(X_{n-})\nabla_v x_{k} - \sigma_{\ell_k}(X_{n-})\Delta W_{\ell_k}^{\beta} - \Delta W_{\ell_k}\rangle\right) \]

\[ + \sum_{k,\ell_k \leq t} (\beta_{\ell_k} - \beta_{\ell_k}) \text{Tr}(\sigma_{-1}\nabla_v x_{k} - \sigma_{\ell_k})(X_{n-}) \]

\[ = \int_{0}^{\tau} \left(\langle \sigma_{-1}(X_{s-})\nabla_v X_{s-}, \Delta W_{\ell_k}^{\beta}\rangle + \int_{0}^{\tau} \text{Tr}(\sigma_{-1}\nabla_v X_{s-}, \sigma_{s})(X_{s-})d\beta_{\ell_k} \right) \]

\[ + \sum_{s \in [0,\tau]} \langle \sigma_{-1}(X_{s-})\nabla_v x_{s} - \sigma_{s}(X_{s-})\Delta W_{\ell_k}^{\beta}, \Delta W_{\ell_k}\rangle. \]

By (3.9), this implies (3.4).

2. For general \( \ell \in \mathbb{S} \). Let \( \ell^e \) be defined by (2.6). It is well known that \( \ell^e \) has finite jumps on any finite interval. By formula (3.4) for \( \ell^e \), we have

\[ \nabla_v \mathbb{E} \bar{f}(X_t^{\ell^e}) = I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon), \]  

where

\[ I_1(\varepsilon) := \frac{1}{\beta_{\ell_1}} \mathbb{E} \left[ f(X_t^{\ell_1}) \int_{0}^{\tau} \sigma_{s}^{-1}(X_{s-}) \left(\nabla_v X_{s-}^{\ell_1}\right) dW_{\ell_1}^{\beta}\right], \]

\[ I_2(\varepsilon) := \frac{1}{\beta_{\ell_2}} \mathbb{E} \left[ f(X_t^{\ell_2}) \int_{0}^{\tau} \text{Tr}(\sigma_{s}^{-1}\nabla_v x_{s} - \sigma_{s})(X_{s-})d\beta_{\ell_2}\right], \]

\[ I_3(\varepsilon) := \frac{1}{\beta_{\ell_3}} \mathbb{E} \left[ f(X_t^{\ell_3}) \sum_{s \in [0,\tau]} \left(\sigma_{s}^{-1}(X_{s-})\nabla_v x_{s} - \sigma_{s}(X_{s-})\Delta W_{\ell_3}^{\beta}, \Delta W_{\ell_3}\right)\right]. \]

Set

\[ \xi_s^e = \sigma_{s}^{-1}(X_{s-})\nabla_v X_{s-}, \quad \xi_s = \sigma_{s}^{-1}(X_{s-})\nabla_v X_{s}. \]

By (1.5), (1.6) and Lemma 3.1 it is easy to see that

\[ \lim_{\varepsilon \downarrow 0} \mathbb{E} |\xi_s^e - \xi_s|^2 = 0. \]

Hence, by Lemma 2.2 and the dominated convergence theorem, we have

\[ \lim_{\varepsilon \downarrow 0} \mathbb{E} \left[ \int_{0}^{\tau} \xi_s^e dW_{\ell_1}^{\beta} - \int_{0}^{\tau} \xi_s dW_{\ell_1}^{\beta}\right]^2 \leq \lim_{\varepsilon \downarrow 0} \mathbb{E} \left[ \int_{0}^{\tau} (\xi_s^e - \xi_s)dW_{\ell_1}^{\beta}\right]^2 = \lim_{\varepsilon \downarrow 0} \int_{0}^{\tau} \mathbb{E} |\xi_s^e - \xi_s|^2 d\lambda_{\ell_1}^{\beta} = 0, \]

and so,

\[ \lim_{\varepsilon \downarrow 0} I_1(\varepsilon) = \frac{1}{\beta_{\ell_1}} \mathbb{E} \left[ f(X_t^{\ell_1}) \int_{0}^{\tau} \sigma_{s}^{-1}(X_{s-}) \left(\nabla_v X_{s-}\right) dW_{\ell_1}^{\beta}\right]. \]  

(3.11)
Similarly, we have
\[
\lim_{\varepsilon \to 0} I_2(\varepsilon) = \frac{1}{\beta_{t_i}} \mathbb{E} \left[ f(X_{t_i}^\varepsilon) \int_0^t \text{Tr}(\sigma_s^{-1} \nabla_{\nabla\lambda X_{t_i}^\varepsilon} \sigma_s) (X_{s-}^\varepsilon) \, \text{d}\beta_s \right].
\]

(3.12)

To treat \( I_3(\varepsilon) \), we set
\[
A_s^\varepsilon := (\sigma_s^{-1} \nabla_{\nabla\lambda X_{t_i}^\varepsilon} \sigma_s) (X_{s-}^\varepsilon), \quad A_s := (\sigma_s^{-1} \nabla_{\nabla\lambda X_{t_i}} \sigma_s) (X_{s-}^\varepsilon).
\]

Then
\[
\eta_t^\varepsilon := \sum_{s \in [0,t]} \langle \sigma_s^{-1} (X_{s-}^\varepsilon) \nabla_{\nabla\lambda X_{t_i}^\varepsilon} \sigma_s (X_{s-}^\varepsilon) \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s}^\varepsilon \rangle = \sum_{s \in [0,t]} \langle A_s^\varepsilon \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s}^\varepsilon \rangle
\]
and
\[
\eta_t := \sum_{s \in [0,t]} \langle \sigma_s^{-1} (X_{s-}^\varepsilon) \nabla_{\nabla\lambda X_{t_i}} \sigma_s (X_{s-}^\varepsilon) \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s} \rangle = \sum_{s \in [0,t]} \langle A_s \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s} \rangle.
\]

Letting \( \| \cdot \|_p \) denote the \( L^p \)-norm with respect to \( \mathbb{P} \), we obtain for \( p \in (1, 2) \) that
\[
\| \eta_t^\varepsilon - \eta_t \|_p \leq J_1(\varepsilon) + J_2(\varepsilon),
\]
where
\[
J_1(\varepsilon) := \sum_{s \in [0,t]} \| \langle A_s \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s} \rangle - \langle A_s \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s} \rangle \|_p,
\]
\[
J_2(\varepsilon) := \sum_{s \in [0,t]} \| \langle (A_s^\varepsilon - A_s) \Delta W_{\ell_s}^\beta, \Delta W_{\ell_s}^\varepsilon \rangle \|_p.
\]

By Hölder’s inequality, \( (1,6) \) and Lemma \( 3.1 \) for \( q = \frac{2p}{2-p} \) and some constants \( C_1, C_2 \) depending on \( p \) and \( t \), we have
\[
J_1(\varepsilon) \leq C_1 \sum_{s \in [0,t]} \| A_s \|_{2q} \| \Delta W_{\ell_s} \|_{2q} \| \Delta W_{\ell_s}^\beta - \Delta W_{\ell_s} \|_2
\]
\[
+ C_1 \sum_{s \in [0,t]} \| A_s \|_{2q} \| \Delta W_{\ell_s}^\beta \|_{2q} \| \Delta W_{\ell_s} - \Delta W_{\ell_s} \|_2
\]
\[
\leq C_2 (1 + |x|^m) \left( \sum_{s \in [0,t]} \| \Delta W_{\ell_s} \|_{2q}^2 \right)^{1/2} \left( \sum_{s \in [0,t]} \| \Delta W_{\ell_s} - \Delta W_{\ell_s} \|_2^2 \right)^{1/2}
\]
\[
+ C_2 (1 + |x|^m) \left( \sum_{s \in [0,t]} \| \Delta W_{\ell_s}^\beta \|_{2q}^2 \right)^{1/2} \left( \sum_{s \in [0,t]} \| \Delta W_{\ell_s} - \Delta W_{\ell_s} \|_2^2 \right)^{1/2},
\]
(3.14)

which converges to zero as \( \varepsilon \downarrow 0 \) due to the argument in the proof of Lemma \( 2.2 \) where \( x \) is the initial point of \( X_{t_i}^\varepsilon \).

Similarly, we have
\[
J_2(\varepsilon) \leq C \sum_{s \in [0,t]} \| A_s^\varepsilon - A_s \|_2 (\Delta X_{t_s}^\varepsilon)^{1/2} (\Delta X_{t_s}^\beta)^{1/2} \leq C \sup_{s \in [0,t]} \| A_s^\varepsilon - A_s \|_2
\]
(3.15)

which goes to zero as \( \varepsilon \downarrow 0 \) by Lemma \( 3.1 \) by the dominated convergence theorem, and combining this with \( (3.14) \) and \( (3.15) \), we obtain
\[
\lim_{\varepsilon \to 0} I_3(\varepsilon) = \frac{1}{\beta_{t_i}} \mathbb{E} \left[ f(X_{t_i}^\varepsilon) \eta_t \right].
\]

Therefore, the proof is finished by taking \( \varepsilon \downarrow 0 \) in \( (3.10) \) and noting that due to \( (3.3) \),
\[
\lim_{\varepsilon \to 0} \nabla_x \mathbb{E} f(X_{t_i}^\varepsilon) = \nabla_x \mathbb{E} f(X_{t_i}^\varepsilon).
\]
4. Derivative formula and gradient estimate for $P_t$: Proof of Theorem 1.1

Let $B_b(\mathbb{R}^d)$ be the set of all bounded measurable functions on $\mathbb{R}^d$. Consider

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

We first prove the following gradient estimate for $P_t^f$.

**Theorem 4.1.** Assume (1.5) and (1.6). For any $p > 1$, there exists a constant $C > 0$ such that for any $f \in B_b(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, $t \in (0, 1]$ with $\ell_t > 0$,

$$|\nabla P_t^f(x)| \leq \frac{C(1 + |x|^m)}{\sqrt{\ell_t} \wedge 1} (P_t^f |f|^p)^{1/p}(x). \quad (4.1)$$

**Proof.** By an approximation argument and the Jensen inequality, we only need to prove (4.1) for $f \in C_b^1(\mathbb{R}^d)$ and $p \in (1, 2)$. In this case, $q := \frac{p}{p-1} > 2$.

Simply denote $X_t^f = X_t^f(x)$. By (1.6), (3.1) and (3.2), there exists a constant $c > 0$ such that for all $t \in (0, 1]$,

$$\mathbb{E}\left(\sup_{s \in [0,t]} |\nabla^s X_s^f|^2\right) \leq c e^{c\ell_t^2}, \quad \mathbb{E}\left(\sup_{s \in [0,t]} |\sigma_s^{-1}(X_s^f)|^2\right) \leq c(1 + |x|^m)^2 e^{c\ell_t^2}. \quad (4.2)$$

Define

$$\tau := \inf\{t : \ell_t \geq 1\}.$$ 

Clearly,

$$\ell_t \geq 1, \quad \ell_{t^-} \leq 1. \quad (4.3)$$

Below we take

$$\beta_t := t \wedge \ell_t, \quad t \geq 0. \quad (4.4)$$

By Theorem 3.2 for any $v \in \mathbb{R}^d$ with $|v| = 1$, we have

$$|\nabla_v P_t^f| \leq I_1 + I_2 + I_3, \quad (4.5)$$

where

$$I_1 := \frac{1}{\beta_t} \mathbb{E}\left|f(X_t^f) \int_0^t \sigma_s^{-1}(X_s^-)\left(\nabla_v X_s^-\right) dW_t^\theta\right|,$$

$$I_2 := \frac{1}{\beta_t} \mathbb{E}\left|f(X_t^f) \int_0^t \text{Tr}(\sigma_s^{-1}\nabla_v X_s^- \sigma_s)(X_s^-) d\beta_t\right|,$$

$$I_3 := \frac{1}{\beta_t} \mathbb{E}\left|f(X_t^f) \sum_{s \in [0,t]} \langle A_s \Delta W_{t^-}^\theta, X_s^- \rangle \right|,$$

where $A_s$ is in (3.13). For $I_1$, by Lemma 2.1 and Hölder’s inequality, and using (4.2) and (2.3), we have

$$I_1 \leq \frac{1}{\beta_t} (P_t^f |f|^p)^{1/p} \left(\mathbb{E}\left(\int_0^t |\sigma_s^{-1}(X_s^-)\left(\nabla_v X_s^-\right)|^q dW_t^\theta\right)^{1/q}\right)^{1/q} \leq \frac{1}{\beta_t} (P_t^f |f|^p)^{1/p} \left(\mathbb{E}\left(\int_0^t |\sigma_s^{-1}(X_s^-)\left(\nabla_v X_s^-\right)|^2 d\lambda_t^\theta\right)^{q/2}\right)^{1/q}$$
Therefore, combining this with (4.6), (4.7) and (4.5), we obtain (4.1).

\[
\begin{align*}
\int_1 I_1 = \int_0^t |\dot{\beta}_1|^2 ds & = \int_0^t 1_{s=t, t = \beta_t} ds \\
\int_0^t e^{c_{\ell t}} d\beta = \int_0^t e^{c_{\ell t}} d\beta & = \sum_{s \in [0,t]} e^{c_{\ell t}} \Delta \beta_t = e^{\beta_t}.
\end{align*}
\]

Therefore,

\[
I_1 \lesssim \frac{C_1(1 + |x|^m) e^{c/q}}{\sqrt{\beta_t}} (P_t[f]_p)^{1/p} \lesssim \frac{C_1(1 + |x|^m) e^{c/q}}{\sqrt{\beta_t} \wedge 1} (P_t[f]_p)^{1/p}.
\] (4.6)

Similarly, there exists a constant \(C_2 > 0\) such that

\[
I_2 \lesssim \frac{1}{\beta_t} (P_t[f]_p)^{1/p} \left( \mathbb{E} \left| \int_0^t \text{Tr}(\sigma_s^{-1} \nabla \psi, \sigma_s^{-1} \nabla \psi)(X_s^\ell) d\beta_t \right|^{q} \right)^{1/q}
\]

\[
\lesssim \frac{1}{\beta_t} (P_t[f]_p)^{1/p} (\beta_t)^{-1/2} \left( \int_0^t \mathbb{E} |\text{Tr}(\sigma_s^{-1} \nabla \psi, \sigma_s^{-1} \nabla \psi)(X_s^\ell)|^q d\beta_t \right)^{1/q}
\]

\[
\lesssim C_2(1 + |x|^m) (P_t[f]_p)^{1/p} (\beta_t)^{-1/2} \left( \int_0^t e^{c_{\ell t}} d\beta_t \right)^{1/q}
\]

\[
\lesssim C_2(1 + |x|^m) e^{c/q} (P_t[f]_p)^{1/p}.
\] (4.7)

Finally, noting that \(\Delta \beta_t = \Delta \ell, 1_{[0,t]}(s)\), we have, for some constant \(C_3 > 0\),

\[
I_3 \lesssim \frac{1}{\beta_t} (P_t[f]_p)^{1/p} \left( \sum_{s \in [0,t]} |A_s \Delta W_{\ell t}^\beta, \Delta W_{\ell t}^\beta| \right)
\]

\[
\lesssim \frac{1}{\beta_t} (P_t[f]_p)^{1/p} \sum_{s \in [0,t]} \|A_s \Delta W_{\ell t}^\beta, \Delta W_{\ell t}^\beta\|_q
\]

\[
\lesssim \frac{1}{\beta_t} (P_t[f]_p)^{1/p} \sum_{s \in [0,t]} \|A_s\|_{2q} \|\Delta W_{\ell t}^\beta\|_{4q} \|\Delta W_{\ell t}\|_{4q}
\]

\[
\lesssim \frac{C_3(1 + |x|^m) (P_t[f]_p)^{1/p}}{\beta_t} \sum_{s \in [0,t]} |\Delta \ell|^{1/2} |\Delta \ell|^{1/2} e^{c/(2q)}
\]

\[
\lesssim \frac{C_3(1 + |x|^m) (P_t[f]_p)^{1/p}}{\beta_t} \sum_{s \in [0,t]} |\Delta \ell|^{1/2} |\Delta \ell|^{1/2} e^{c/(2q)}
\]

\[
= \frac{C_3 e^{c/(2q)} (1 + |x|^m) (P_t[f]_p)^{1/p}}{\beta_t} \sum_{s \in [0,t \wedge \tau]} \Delta \ell
\]

\[
= C_3(1 + |x|^m) e^{c/(2q)} (P_t[f]_p)^{1/p}.
\]

Combining this with (4.6), (4.7) and (4.5), we obtain (4.1). \(\square\)
we have

\[ P_{τ}f(x) := \mathbb{E} f(X(τ)) = \mathbb{E}(\mathbb{E} f(X(τ)|σ_{τ})) = \mathbb{E}(P_{τ}S f(x)). \]

(i) For \( R > 0 \) and \( ℓ ∈ \mathbb{S} \), define

\[ τ(ℓ) := \inf\{t : ℓ_t ≥ R\}. \]

If we choose \( β_t = t ∧ τ(ℓ) \), then by (4.8) and (3.4), we have, for any \( v ∈ \mathbb{R}^d \),

\[ ∇_v P_{τ}f = I_1 + I_2 + I_3, \]

where

\[ I_1 := \mathbb{E}\left( \mathbb{E}\left[ f(X(ℓ)) \frac{1}{β(ℓ)} \int_0^{\tau(ℓ)} \langle \gamma - 1(X(ℓ))∇vX(ℓ), dW(ℓ) \rangle \right]_{ℓ=S} \right), \]

\[ I_2 := \mathbb{E}\left( \mathbb{E}\left[ f(X(ℓ)) \frac{1}{β(ℓ)} \int_0^{\tau(ℓ)} \text{Tr}(\langle \gamma - 1(X(ℓ)), X(ℓ) \rangle dβ(ℓ)) \right]_{ℓ=S} \right), \]

\[ I_3 := \mathbb{E}\left( \mathbb{E}\left[ f(X(ℓ)) \frac{1}{β(ℓ)} \sum_{s \in [0,τ]} \langle \gamma - 1(X(ℓ)), \sigma_s(X(ℓ)) ∆W(ℓ), ∆W(ℓ) \rangle \right]_{ℓ=S} \right). \]

As shown in the proof of Theorem 4.1, it is clear that when \( E S^{-1/2}_{ℓ∧τ} < ∞ \), \( I_1, I_2 \) and \( I_3 \) are well defined. Noticing that

\[ W(ℓ) = W(ℓ∧τ(ℓ)) = W(ℓ∧τ(ℓ)), \]

and

\[ ∆W(ℓ) = ∆W(ℓ∧τ(ℓ)) = W(ℓ1[0,τ])(t), \]

we have

\[ I_1 = \mathbb{E}\left[ f(X(ℓ)) \frac{1}{S_{ℓ∧τ}} \int_0^{ℓ∧τ} \langle \gamma - 1(X(s)), ∇vX(s), dW(s) \rangle \right], \]

\[ I_2 = \mathbb{E}\left[ f(X(ℓ)) \frac{1}{S_{ℓ∧τ}} \int_0^{ℓ∧τ} \text{Tr}(\langle \gamma - 1(X(s), X(s)) dW(s) \rangle \right], \]

and

\[ I_3 = \mathbb{E}\left[ f(X(ℓ)) \frac{1}{S_{ℓ∧τ}} \sum_{s \in [0,τ]} \langle \gamma - 1(X(s)), ∇vX(s), X(s) ∆W(s), ∆W(s) \rangle \right], \]

\[ = \mathbb{E}\left[ f(X(ℓ)) \frac{1}{S_{ℓ∧τ}} \int_0^{ℓ∧τ} \langle \gamma - 1(X(s)), ∇vX(s), X(s), x, x) N(ds, dx) \right], \]

where the random measure \( N \) is defined by (1.8). Thus, the formula (1.7) is proven.

(ii) By the Markov property, it suffices to prove (1.9) for \( t ∈ (0, 1] \). Then the estimate (1.9) follows from Theorem 4.1 and (4.8) by using the Hölder inequality. So, it suffices to prove (1.10) for \( t ∈ (0, 1] \) in the case that \( B(u) \geq c u^α/2, u \geq u_0 \), where \( α ∈ (0, 2) \) and \( u_0 ≥ 0 \). In this case, for any \( γ > 0 \) we may find a constant \( C ≥ 0 \) such that for all \( t > 0 \) (see e.g. [2] page 298) for the formula of \( E S_t^γ \),

\[ \mathbb{E}\left[ S_t^γ \right] ≤ 1 + \mathbb{E}\left[ S_t^γ \right] = 1 + \frac{1}{Γ(γ)} \int_0^∞ u^{γ-1} e^{-tB(u)} du \]

\[ ≤ 1 + \frac{1}{Γ(γ)} \left( ∫_0^∞ u^{γ-1} e^{-cu^{α/2}} du + ∫_0^{u_0} u^{γ-1} du \right) \]
\[ \leq 1 + \frac{\mu_0^\gamma}{\gamma \Gamma(\gamma)} + \frac{1}{\Gamma(\gamma)} \int_0^\infty u^{\gamma-1} e^{-ctu^{\gamma/2}} du \leq \frac{C}{\Gamma} t^{2\gamma/\alpha}. \]

Therefore, (1.10) follows from (1.9).

Acknowledgements:

We would like to deeply thank Professor Zhenqing Chen for his quite useful conversations. Feng-Yu Wang is supported by NNSFC(11131003), SRFDP, the Laboratory of Mathematical and Complex Systems and the Fundamental Research Funds for the Central Universities. Xi-cheng Zhang is supported by NSF of China (No. 11271294) and Program for New Century Excellent Talents in University.

REFERENCES

[1] Jacob N.: Pseudo Differential Operators and Markov Processes (Volume I), Imperial College Press, London, 2001.
[2] Gordiba M., Rockner M., Wang F.-Y.: Dimension-independent Harnack inequality for subordinated semigroups, Potential Analysis 34(2011), 293–307.
[3] Malliavin P.: Stochastic analysis, Springer-Verlag, Berlin, 1997.
[4] Nualart D.: The Malliavin calculus and related topics. Second Edition, Springer-Verlag, Berlin(2005).
[5] Protter P.E.: Stochastic integration and differential equations. Second Edition, Springer-Verlag, Berlin, 2004.
[6] Schilling R. L., Sztonyk P. and Wang J.: Coupling property and gradient estimates of Lévy processes via the symbol. Bernoulli. 18(2012), 1128-1149.
[7] Wang F.-Y.: Gradient estimate for Ornstein-Uhlenbeck jump processes, Stoch. Proc. Appl. 121(2011), 466–478.
[8] Wang F.-Y.: Derivative formula and Harnack inequality for jump processes, \textit{arXiv:1104.5531}.
[9] Wang F. Y., Wang J.: Harnack inequalities for stochastic equations driven by Lévy noise, \textit{arXiv:1212.0405}.
[10] Zhang X.: Derivative formula and gradient estimate for SDEs driven by $\alpha$-stable processes. Stoch. Proc. Appl. Vol. 123, Issue 4, pp.1213-1228(2013).

FENG-YU WANG
School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China; wangfy@bnu.edu.cn

DEPARTMENT OF MATHEMATICS, SWANSEA UNIVERSITY,Singleton Park, SA2 8PP, UK; F.Y.Wang@swansea.ac.uk

LIHU XU
Department of Mathematics, Brunel University, Uxbridge, UB8 3PH, UK; lihu.xu@brunel.ac.uk

CORRESPONDING AUTHOR: XIChENG ZHANG
School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China; xichengzhang@googlemail.com