A WONG-ZAKAI APPROXIMATION OF STOCHASTIC DIFFERENTIAL EQUATIONS DRIVEN BY A GENERAL SEMIMARTINGALE

XIANMING LIU*
School of Mathematics and Statistics
Huazhong University of Science and Technology
Wuhan, 430074, China

GUANGYUE HAN
Department of Mathematics
The University of Hong Kong
Hong Kong, China

(Communicated by Arnulf Jentzen)

ABSTRACT. We examine a Wong-Zakai type approximation of a family of stochastic differential equations driven by a general càdlàg semimartingale. For such an approximation, compared with the pointwise convergence result by Kurtz, Pardoux and Protter [10, Theorem 6.5], we establish stronger convergence results under the Skorokhod $M_1$-topology, which, among other possible applications, implies the convergence of the first passage time of the solution to the approximating stochastic differential equation.

1. Introduction. Let $L = \{L(t); 0 \leq t < \infty\}$ be a stochastically continuous càdlàg semimartingale [15] defined on a probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. For any $\epsilon > 0$, let $L^{\epsilon}$ be the smooth approximation [6, 10, 23] of $L$ defined by

$$L^{\epsilon}(t) := \frac{1}{\epsilon} \int_{t-\epsilon}^t L(s) ds, \quad 0 \leq t < \infty,$$

and let $X^{\epsilon} = \{X^{\epsilon}(t); 0 \leq t < \infty\}$ be the solution to the following random differential equation

$$dX^{\epsilon}(t) = b(X^{\epsilon}(t))dt + f(X^{\epsilon}(t))dL^{\epsilon}(t), \quad X^{\epsilon}(0) = X_0,$$

where $b, f$ are some functions from $\mathbb{R}$ to $\mathbb{R}$ satisfying certain regularity conditions, and $X_0$ is an $\mathcal{F}_0$-measurable random variable.

In this paper, we will establish some convergence results on $X^{\epsilon}$ as $\epsilon$ tends to 0. More precisely, we will show that, in some sense, $X^{\epsilon}$ converges to $X$, where $X$ is the solution to the following stochastic differential equation:

$$X(t) = X_0 + \int_0^t b(X(s))ds + \int_0^t f(X(s-)) \circ dL(s),$$

2020 Mathematics Subject Classification. Primary: 60H10; Secondary: 34F05.

Key words and phrases. Wong-Zakai approximation, stochastic differential equation, semimartingale, the Skorokhod $M_1$-topology, random time change.

The first author is supported by NSF grants of China No. 11971186.

* Corresponding author: Xianming Liu.
where \( \diamond \) denotes Marcus integral. Here, let us add that (3) is in fact a Marcus canonical equation and can be equivalently rewritten as
\[
X(t) = X_0 + \int_0^t b(X(s)) \, ds + \int_0^t f(X(s)) \, dL^c(s) + \int_0^t f(X(s-)) \, dL^d(s) + \sum_{0 < s \leq t} \left[ \phi(\Delta L(s); X(s-), 1) - X(s-) - f(X(s-)) \Delta L(s) \right],
\]
where \( L^c \) and \( L^d \) are respectively the continuous and discontinuous parts of \( L \), and \( \diamond \) denotes Stratonovich differential, and furthermore \( \phi(\sigma; u, t) \) is the flow generated by a vector field \( \sigma \):
\[
\frac{d\phi(\sigma; u, t)}{dt} = f(\phi(\sigma; u, t)), \quad \phi(\sigma; u, 0) = u. \tag{5}
\]
For more details about Marcus integral and canonical equations, we refer the reader to [1, 3, 7, 8, 10].

Note that the equation (2) is a perturbed version of (3) in the sense of Wong-Zakai [4, 10, 13, 18, 21, 22], the convergence result as above is somewhat expected. The real question, however, is in exactly what sense the result holds. For the special case \( b = 0 \), it has been shown by Kurtz, Pardoux and Protter [10] that for all but countably many \( t \), \( X^\epsilon(t) \) converges in probability to \( X(t) \), as \( \epsilon \) tends to 0. As detailed in the following theorem, we will show that for any \( T > 0 \), \( \{X^\epsilon(t); 0 \leq t \leq T\} \) converges in probability to \( \{X(t); 0 \leq t \leq T\} \) in \( D([0,T], \mathbb{R}) \), the space of all the càdlàg functions over \([0,T]\), under the the Skorokhod \( M_1 \)-topology, or simply put, \( X^\epsilon \) converges in probability to \( X \) under the Skorokhod \( M_1 \)-topology. Here we remark that the same convergence is not possible under the Skorokhod \( J_1 \)-topology due to the simple fact that convergence under the \( J_1 \)-topology retains continuity, whereas \( X^\epsilon \) is continuous for any \( \epsilon > 0 \) and \( X \) may be discontinuous. For the precise definitions of the Skorokhod \( J_1 \) and \( M_1 \)-topologies, see Appendix A.

**Theorem 1.1.** Suppose that the functions \( b, f \) and \( f' \) are bounded and Lipschitz. Then, as \( \epsilon \) tends to 0, \( X^\epsilon \) converges in probability to \( X \) under the Skorokhod \( M_1 \)-topology.

For a real-valued stochastic process \( W = \{W(t); 0 \leq t < \infty\} \) and a positive real number \( a > 0 \), we use \( \tau_a(W) \) to denote the first passage time of \( W \) with respect to \( a \), that is,
\[
\tau_a(W) = \inf\{t \geq 0 : W(t) > a\}.
\]
As an immediate corollary of Theorem 1.1, we have

**Corollary 1.** For any positive real number \( a > 0 \), as \( \epsilon \) tends to 0, \( \tau_a(X^\epsilon) \) converges to \( \tau_a(X) \) in distribution.

**Proof.** The corollary follows from Theorem 1.1, the easily verifiable fact that for any \( t, \epsilon > 0 \),
\[
P(\tau_a(X^\epsilon) \leq t) = P(\sup_{s \in [0,t]} X^\epsilon(s) \geq a), \quad P(\tau_a(X) \leq t) = P(\sup_{s \in [0,t]} X(s) \geq a),
\]
and the fact that the first passage time function and the supremum function as above are continuous under the Skorokhod \( M_1 \)-topology; see e.g., [20, Lemma 1] and [16, Lemma 2.1].

\[\blacksquare\]

\[\text{1Following the usual practice in the theory of stochastic calculus, we will implicitly choose its càdlàg version for a given stochastic process.}\]
The key tool that we used in this work is the so-called method of random time change (see, e.g., [9]), which is a well-known method that has also been used in [10]. On the other hand though, the power of this method somehow has not been fully utilized in [10]: Theorem 1.1 in this work, which is established through a short and simple argument, immediately implies that $X^\epsilon(t)$ converges in probability to $X(t)$ for all but countably many $t$, which further implies Theorem 6.5 in [10]. As a matter of fact, the power of this method can be further showcased in some special setting: For the case that $L$ is a Lévy process, the method of Hintze and Pavlyukevich [5] can be adapted to show that as $\epsilon$ tends to 0, $L^\epsilon$ converges in probability to $L$ under the Skorokhod $M_1$-topology, whereas our proof employing the method of random time change readily yields a stronger result stating that as $\epsilon$ tends to 0, $L^\epsilon$ converges almost surely to $L$ under the Skorokhod $M_1$-topology (see Theorem 2.1 in Section 2). Here we remark that the proof of Theorem 3.1 in [5] is heavily dependent on the structure of the Lévy process and cannot carry over to the case when $L$ is a general semimartingale, in which case our proof however aptly applies.

The remainder of this paper is organized as follows. In Section 2, we use a special case to illustrate the key methodology used in our proof. In Section 3, we prove Theorem 1.1, the main result of this paper. For self-containedness, we recall in Appendix A some basic notions and results on the Skorokhod $J_1$ and $M_1$-topologies.

2. A special case. Note that if we set $b \equiv 0$, $f \equiv 1$ and $X_0 = 0$, then $X^\epsilon$ is nothing but $L^\epsilon$. This section, which is meant to be illustrative, is concerned with this special case, for which we will use the method of random time change to establish the following theorem:

**Theorem 2.1.** As $\epsilon$ tends to 0, $L^\epsilon$ converges almost surely to the semimartingale $L$ under the Skorokhod $M_1$-topology.

**Proof.** Let $[L] = [L,L]$ denote the quadratic variation of $L$, and let $[L]^c$ and $[L]^d$ denote its continuous and purely discontinuous parts, respectively. Define $\gamma^0(t) := [L]^d(t) + t$, and for any $\epsilon > 0$, define

$$\gamma^\epsilon(t) := \frac{1}{\epsilon} \int_{t-\epsilon}^{t} (\gamma^d(s) + s)ds.$$  

It can be shown that for any $t \geq 0$ and any $\epsilon > 0$, $\gamma^\epsilon(t) < \gamma^0(t) < \gamma^0(t + \epsilon)$. For any $\epsilon > 0$, let $\zeta^\epsilon$ be the generalized inverse of $\gamma^\epsilon$, i.e., $\zeta^\epsilon(t) := \inf\{s > 0 : \gamma^\epsilon(s) > t\}$. It can also be shown that for any $t \geq 0$ and any $\epsilon > 0$, $\zeta^\epsilon(t) - \epsilon < \zeta^0(t) < \zeta^\epsilon(t)$, which implies that $\zeta^\epsilon$ converges to $\zeta^0$ uniformly over all $t$ from any bounded time interval. For any $\epsilon > 0$, define $Z^\epsilon(t) = L^\epsilon_{\zeta^\epsilon(t)}$; in other words, the new process $Z^\epsilon$ is the original process $L^\epsilon$ reevaluated with respect to the new time scale $\zeta^\epsilon(\cdot)$. It can be easily verified that $Z^\epsilon(t)$ is continuous in $t$.

The remainder of the proof consists of three steps as follows.

**Step 1:** In this step, we will show that as $\epsilon$ tends to 0, $\{Z^\epsilon(t)\}$ uniformly converges to a continuous process $\{Z(t)\}$.

Defining

$$\eta_-(t) = \sup\{s : \zeta^0(s) < \zeta^0(t)\}, \quad \eta_+(t) = \inf\{s : \zeta^0(s) > \zeta^0(t)\},$$

letting $\{\tau_i, i \in \mathbb{N}\}$ denote the sequence of all the jump times of $L$, we will deal with the following two cases.

**Case 1:** $t \in [0,\gamma^0(\tau_1-))$ or $t \in (\gamma^0(\tau_i),\gamma^0(\tau_{i+1}-))$ for some $i$.  

**Case 2:** $t \in (\gamma^0(\tau_i), \gamma^0(\tau_{i+1}-))$ for some $i$.  

**Case 3:** $t \in (\gamma^0(\tau_i), \gamma^0(\tau_{i+1}))-$.
Moreover, it follows from the verifiable fact

\[ Z(\zeta^0(t)) = L_{\zeta^0(t)} = L_{\zeta^0(t)}, \quad t < 0. \]

**Case 2**: \( t \in [\gamma^0(\tau_i), \gamma^0(\tau_i)]. \)

In this case, \( \zeta^0(t) \equiv \tau_i \) and \( \eta_-(t) \neq \eta^+(t) \), and \( L \) has a discontinuity at \( \zeta^0(t) \). It can be shown that

\[
\lim_{\epsilon \to 0^+} Z'(\gamma^0(\tau_i)) = \lim_{\epsilon \to 0^+} L_{\zeta^0(\gamma^0(\tau_i))} = L_{\tau_i} = L_{\zeta^0(\tau_i)}, \quad (7)
\]

and

\[
\lim_{\epsilon \to 0^+} Z'(\gamma^0(\tau_i)) = \lim_{\epsilon \to 0^+} L_{\zeta^0(\gamma^0(\tau_i))} = L_{\tau_i} = L_{\zeta^0(\tau_i)}. \quad (8)
\]

Moreover, it follows from the verifiable fact

\[
\frac{dZ'_{\zeta^0(t)}}{dt} = \frac{L_{\zeta^0(t)} - L_{\zeta^0(t) - \epsilon}}{[L_{\zeta^0(t)} - [L_{\zeta^0(t)} - \epsilon + \epsilon]}
\]

that for any \( t \in [\gamma^0(\tau_i), \gamma^0(\tau_i)], \)

\[
\lim_{\epsilon \to 0^+} \frac{dZ'_{\zeta^0(t)}}{dt} = \frac{L_{\zeta^0(t)} - L_{\zeta^0(t)} - \epsilon}{[L_{\zeta^0(t)} - [L_{\zeta^0(t)} - \epsilon + \epsilon]}
\]

Consequently, it follows from (6), (7), (8) and (9) that \( \lim_{\epsilon \to 0^+} Z'(t) = Z(t) \) uniformly over all \( t \) from any bounded time interval, where \( Z \) is continuous and admits following expression:

\[ Z(t) = \begin{cases} L_{\zeta^0(t)}, & \text{if } \eta_-(t) = \eta^+(t), \\ \frac{\epsilon - \eta_-(t)}{\eta^+(t) - \eta_-(t)} L_{\zeta^0(t)} + \frac{\eta^+(t) - t}{\eta^+(t) - \eta_-(t)} L_{\zeta^0(t) - \epsilon}, & \text{if } \eta_-(t) \neq \eta^+(t). \end{cases} \]

**Step 2**: This step will lead to the conclusion that as \( \epsilon \) tends to 0, \( \gamma^\epsilon \) converges almost surely to \( \gamma^0 \) under the Skorokhod \( M_1 \)-topology. The proof of this step is postponed to next section (see Lemma 3.2).

**Step 3**: In this step, we will show that as \( \epsilon \) tends to 0, \( L^\epsilon \) converges almost surely to \( L \) under the Skorokhod \( M_1 \)-topology, thereby completing the proof.

It follows from the facts that \( \zeta^0 \circ \gamma^0(t) = t \) and \( \gamma^0(t) \notin [\gamma^0(\tau_i), \gamma^0(\tau_i)] \) for any \( t, \tau_i \) that \( Z_{\zeta^0(t)} \equiv L(t) \). Since \( Z^\epsilon \) uniformly converges to \( Z \), and \( \gamma^\epsilon \) converges almost surely to \( \gamma^0 \) under the Skorokhod \( M_1 \)-topology, we conclude that \( L^\epsilon(\cdot) = Z_{\gamma_{\kappa}(\cdot)}^\epsilon \) converges almost surely to \( Z_{\gamma(\cdot)} = L(\cdot) \) under the Skorokhod \( M_1 \)-topology.

**Remark 1.** Compared to Theorem 1.1, Theorem 2.1 deals with a special setting and yields a stronger result. On the other hand, compared to Theorem 3.1 in [5], as mentioned in Section 1, Theorem 2.1 treats a more general setting yet still produces a stronger result.

3. **Proof of Theorem 1.1.** The proof of Theorem 1.1 roughly follows the framework laid out in the proof of Theorem 2.1 and uses many notations defined therein.

For any \( \epsilon > 0 \), recall that \( Z^\epsilon \) is defined as in the proof of Theorem 2.1, and define \( Y^\epsilon \) as \( Y^\epsilon(t) = X_{\zeta^\epsilon(t)}^\epsilon \) for any \( t \). It can be easily verified that \( \{Z^\epsilon(t)\} \) and \( \{Y^\epsilon(t)\} \) are continuous, and moreover \( \{Y^\epsilon(t)\} \) is the unique solution to the following equation:

\[
Y^\epsilon(t) = X_0 + \int_0^t b(Y^\epsilon(s))d\zeta^\epsilon(s) + \int_0^t f(Y^\epsilon(s))dZ^\epsilon(s), \quad 0 \leq t < \infty. \quad (10)
\]

We will first prove the following lemma.
**Lemma 3.1.** As $\epsilon$ tends to 0, $Y^\epsilon$ converges in probability to a process $Y$ under the compact uniform topology. Moreover, $Y$ is continuous and satisfies

$$Y(t) = X_0 + \sum_i \phi \left( f \Delta L(\tau_i), Y_{\gamma^0(\tau_i)}, \frac{t \wedge \gamma^0(\tau_i) - \gamma^0(\tau_i)}{\| \Delta L(\tau_i) \|^2} \right) - Y^0_{\gamma^0(\tau_i)} - f(Y^0(\tau_i)) \Delta L(\tau_i)$$

$$\times I_{[\gamma^0(\tau_i), +\infty)} + \int_0^t b(Y(s)) \, d\xi^0(s) + \int_0^t f(Y(s)) \, dL^0(s) + \frac{1}{2} \int_0^t f'(Y(s)) \, d[L^0]^\frac{1}{2}(s).$$

**Proof.** The proof largely follows from that of Theorem 6.5 in [10], so we only give a sketch emphasizing the key steps.

As shown in Section 2, $\zeta^\epsilon$ and $Z^\epsilon$ converge to $\zeta^0$ and $Z$, respectively, both uniformly over any bounded time interval, which immediately implies that $U^\epsilon$ converges almost surely to $U$ under the Skorokhod $J_1$-topology, where the processes $U^\epsilon$ and $U$ are defined as

$$U^\epsilon(t) := Z^\epsilon(t) - L_{\zeta^\epsilon(t)}, \quad U(t) := Z(t) - L_{\zeta(t)}.$$

And we note that (10) can be rewritten as

$$Y^\epsilon(t) = X_0 + \int_0^t b(Y^\epsilon(s)) \, d\xi^\epsilon(s) + \int_0^t f(Y^\epsilon(s)) \, dL^\epsilon(s) + \int_0^t f(Y^\epsilon(s)) \, dU^\epsilon(s)$$

$$= X_0 + \int_0^t b(Y^\epsilon(s)) \, d\xi^\epsilon(s) + \int_0^t f(Y^\epsilon(s)) \, dL^\epsilon(s) + f(Y^\epsilon(t)) U^\epsilon(t)$$

$$- \int_0^t f'(Y^\epsilon(s)) U^\epsilon(s) \, dY^\epsilon(s) - [f(Y^\epsilon), U^\epsilon](t)$$

$$= X_0 + \int_0^t b(Y^\epsilon(s)) \, d\xi^\epsilon(s) + \int_0^t f(Y^\epsilon(s)) \, dL^\epsilon(s) + f(Y^\epsilon(t)) U^\epsilon(t)$$

$$- \int_0^t f'(Y^\epsilon(s)) f(Y^\epsilon(s)) U^\epsilon(s) \, dZ^\epsilon(s) - \int_0^t f'(Y^\epsilon(s)) b(Y^\epsilon(s)) U^\epsilon(s) \, d\xi^\epsilon(s),$$

where we have used the fact that $[f(Y^\epsilon), U^\epsilon] \equiv 0$.

By [10, Lemma 6.3], we infer that $\{\int_0^t U^\epsilon(s) \, dZ^\epsilon(s)\}$ and $\{\zeta^\epsilon(\cdot)\}$ are “good” (see Kurtz-Procter [11, 12]), and $\zeta^\epsilon$ converges to $\zeta^0$ uniformly over any bounded time interval, and moreover,

$$\int_0^t U^\epsilon(s) \, dZ^\epsilon(s) \to \frac{U(t)^2 - [L]_{\zeta^\epsilon(t)}}{2} = \frac{(Z(t) - L_{\zeta^\epsilon(t)})^2 - [L]_{\zeta^\epsilon(t)}}{2}$$

(12)

in probability under the Skorokhod $J_1$-topology. Then, parallel to the proof of Lemma 6.4 in [10], using the boundedness and Lipschitzness of $b, f$ and $f'$, we deduce that $f(Y^\epsilon(t)) U^\epsilon(t)$ converges in distribution to $R(t)$ under the Skorokhod $J_1$-topology, where

$$R(t) = \sum_i I_{[\gamma^0(\tau_i), \tau_i)}(t) f \left( \phi \left( f \Delta L(\tau_i), Y_{\gamma^0(\tau_i)}, \frac{t - \gamma^0(\tau_i)}{\| \Delta L(\tau_i) \|^2} \right) \right) U(t);$$

(13)

here, $\{\tau_i, i \in \mathbb{N}\}$, as in the proof of Theorem 2.1, is the sequence of all the jump times of $L$. Moreover, by the definition of $U^\epsilon$ and $\zeta^\epsilon$, we deduce that as $\epsilon$ tends to 0,

$$\int_0^t f'(Y^\epsilon(s)) b(Y^\epsilon(s)) U^\epsilon(s) \, d\xi^\epsilon(s) \to 0.$$  

(14)
Now, combining (11)-(14) as above, we deduce from [12] and [11, Theorem 5.4] that $Y^\epsilon$ converges in distribution to $Y$ under the Skorokhod $J_1$-topology, where
\[
Y(t) = X_0 + \int_0^t b(Y(s))d\zeta^0(s) + \int_0^t f(Y(s))dL\zeta^0(s) + R(t)
- \frac{1}{2} \int_0^t f'(Y(s))f(Y(s))d(U(s)^2 - [L]_{\zeta^0(s)}).
\]
Note that $U(t)$ can be further computed as
\[
U(t) = Z(t) - L\zeta^0(t) = \begin{cases} 0, & \text{if } \eta^+(t) = \eta^-(t), \\
\frac{\eta^+(t) - t}{t - \eta^-(t)}(L\zeta^0(t) - L\zeta^0(t)), & \text{if } \eta^-(t) \neq \eta^+(t). 
\end{cases}
\]
It then follows from the fact for any $t \in [\gamma^0(\tau_i^-), \gamma^0(\tau_i^+))$
\[
\varsigma^0(t) \equiv \tau_i, \quad \eta^+(t) - \eta^-(t) = \Delta[L]_{\tau_i} = |\Delta L(\tau_i)|^2
\]
that
\[
Y(t) = Y_{\gamma^0(\tau_i)} + \int_{\gamma^0(\tau_i^+)}^{\gamma^0(\tau_i^+)} b(Y(s))d\zeta^0(s) + \int_{\gamma^0(\tau_i^-)}^{\gamma^0(\tau_i^-)} f(Y(s))dL\zeta^0(s)
+ f\left( \phi \left( f\Delta L(\tau_i), Y_{\gamma^0(\tau_i^+)} - \gamma^0(\tau_i^-) \right) \right) U(t)
- \int_{\gamma^0(\tau_i^-)}^{\gamma^0(\tau_i)} f'(Y(s))f(Y(s))d\zeta^0(s) - \eta^{+}(s) - \eta^{-}(s)ds
= Y_{\gamma^0(\tau_i)} + f(Y_{\gamma^0(\tau_i)})\Delta L(\tau_i)
+ f\left( \phi \left( f\Delta L(\tau_i), Y_{\gamma^0(\tau_i^+)} - \gamma^0(\tau_i^-) \right) \right) U(t)
- \int_{\gamma^0(\tau_i^-)}^{\gamma^0(\tau_i)} f'(Y(s))f(Y(s))d\zeta^0(s) - \eta^{+}(s) - \eta^{-}(s)ds.
\]
Consequently,
\[
Y(t) = X_0
+ \sum_i \left( \phi \left( f\Delta L(\tau_i), Y_{\gamma^0(\tau_i^+)} - \gamma^0(\tau_i^-) \right) \right) - Y_{\gamma^0(\tau_i)} - f(Y_{\gamma^0(\tau_i)})\Delta L(\tau_i)
\times I_{(\gamma^0(\tau_i^-), +\infty)}(t) + \int_0^t b(Y(s))d\zeta^0(s) + \int_0^t f(Y(s))dL\zeta^0(s) + \frac{1}{2} \int_0^t f'(Y(s))d[U(s)^2] - [L]_{\zeta^0(s)}.
\]
(15)
Since $Y^\epsilon$, $Y$ are continuous, we infer that $Y^\epsilon$ converges in probability to $Y$ under the compact uniform topology. Finally, using a similar argument in [10, Theorem 6.5], we conclude that $Y^\epsilon$ converges in probability to $Y$ under the compact uniform topology, thereby completing the proof. \]

**Remark 2.** With the added assumption that $b^\prime$ is bounded and Lipschitz, the proof of Theorem 6.5 in [10] can be slightly modified to prove that for almost all $t$, $X^\epsilon(t)$ converges in probability to $X(t)$. By comparison, Lemma 3.1 reaches the same conclusion without the added assumption as above.

The following lemma characterizes the convergence behavior of $\gamma^\epsilon$.

**Lemma 3.2.** As $\epsilon$ tends to 0, $\gamma^\epsilon$ converges almost surely to $\gamma^0$ under the Skorokhod $M_1$-topology.
Proof. We first prove that $\gamma^\epsilon$ converges in probability to $\gamma^0$ under the Skorokhod $M_1$-topology. It suffices to verify that $V^\epsilon$ converges in probability to $[L]^d$ under the Skorokhod $M_1$-topology, where

$$V^\epsilon(t) := \frac{1}{\epsilon} \int_{t-\epsilon}^t [L]^d(s)ds.$$ 

To this end, by [15, Theorem 22, Page 66], the quadratic variation process $[L]^d$ of the semimartingale $L^d$ is a càdlàg, increasing and adapted process, which implies that the mapping $t \mapsto \frac{1}{\epsilon} \int_{t-\epsilon}^t [L]^d(s)ds$ is monotone. So, by the definition of $\rho$ (see (A.16)), we have $\rho(V^\epsilon, \delta) = 0$, which implies that for any fixed $\epsilon > 0$, $\lim_{\delta \to 0+} \limsup_\epsilon \mathbb{P}(\rho(V^\epsilon, \delta) > \epsilon) = 0$. Moreover, one verifies that for any $t$, $V^\epsilon(t)$ converges in probability to $[L]^d(t)$. With the preparations as above, we invoke Proposition A.1 to conclude that $V^\epsilon$ converges in probability $[L]^d$ under the Skorokhod $M_1$-topology.

Now we turn to prove that $\gamma^\epsilon$ converges almost surely to $\gamma^0$ under the Skorokhod $M_1$-topology. Using the fact that $[L]^d(t)$ is monotone in $t$ and the definition of $V^\epsilon$, we have that for any $0 = \epsilon_1 < \epsilon_2 < \epsilon_3$,

$$[L]^d(t) = V^0(t) = V^\epsilon(t) \geq V^{\epsilon_2}(t) \geq V^{\epsilon_1}(t).$$

It then follows from the definition of $d_{M_1,T}$ that

$$d_{M_1,T}(V^{\epsilon_2}, [L]^d) \leq d_{M_1,T}(V^{\epsilon_1}, [L]^d),$$

that is to say, for any fixed time $T$, almost all $\omega$ in $\Omega$, $d_{M_1,T}(V^\epsilon, [L]^d)$ is monotonically increasing in $\epsilon$. Now, applying the proven fact $V^\epsilon$ converges in probability to the semimartingale $[L]^d$ under the Skorokhod $M_1$-topology and [2, Lemma 2.5.4], we conclude that $V^\epsilon$ converges almost surely to the semimartingale $[L]^d$ under the Skorokhod $M_1$-topology, which implies that $\gamma^\epsilon$ converges almost surely to $\gamma^0$ under the Skorokhod $M_1$-topology, as desired.

Henceforth, letting $X(t) = Y^\psi, \theta(t)$, we prove the following two lemmas.

**Lemma 3.3.** As $\epsilon$ tends to 0, $X^\epsilon$ converges in probability to $X$ under the Skorokhod $M_1$-topology.

**Proof.** The lemma immediately follows from Lemma 3.1, Lemma 3.2 and [19, Theorem 13.2.3].

**Lemma 3.4.** $X$ is the unique solution to the equation (3), and therefore $X \equiv X$.

**Proof.** Since $X(t) = Y^\psi, \theta(t)$, by the equation (15), we have

$$X(t) = X_0 + \sum_i \left( \phi \left( f \Delta L(\tau_i), Y^0(\tau_i), \frac{\gamma^0(\tau_i)}{|\Delta L(\tau_i)|^2} \right) - Y^\psi(\tau_i) - f(Y^\psi(\tau_i)) \Delta L(\tau_i) \right) \times I_{[\gamma^0(\tau_i), +\infty)}(\gamma^0(\tau_i)) + \int_0^\gamma^0(\tau_i) b(Y(s))ds + \int_0^\gamma^0(\tau_i) f(Y(s))dL^\psi(s) + \frac{1}{2} \int_0^\gamma^0(\tau_i) f'(Y(s))d[L]^\psi(s)$$

$$= X_0 + \sum_i \left( \phi(f \Delta L(\tau_i), Y^0(\tau_i), 1) - Y^\psi(\tau_i) - f(Y^\psi(\tau_i)) \Delta L(\tau_i) \right)$$
\[ X \equiv \hat{X} \]

where in the last equality, we have used the alternative definition of Marcus canonical equation in (4). So, \( \hat{X} \) is the solution to the equation (3), which, together with the uniqueness of the solution to the equation (3), implies that \( \hat{X} \equiv X \). 

With all the lemmas as above, we are finally ready to prove Theorem 1.1.

**Proof of Theorem 1.1.** It follows from Lemma 3.1 that \( Y^ε \) converges in probability to \( Y \) under the compact uniform topology. Moreover, it follows from Lemma 3.3 that \( X^ε \) converges in probability to \( \hat{X} \) under the Skorokhod \( M_1 \)-topology. The theorem then follows from Lemma 3.4, which asserts \( \hat{X} \equiv X \). 

**Appendix.**

**Appendix A. Skorokhod topologies.** Throughout this section, we fix \( T > 0 \).

The following \( J_1 \)-metric has been defined by Skorokhod [17]:

\[
\mathcal{D}_{J_1}(x,y) = \inf_{\lambda \in \Lambda} \left\{ \sup_{0 \leq t \leq T} |x(t) - y(\lambda(t))| + \sup_{s,t \in [0,T], s \neq t} \left| \log \frac{\lambda(s) - \lambda(t)}{s - t} \right| \right\},
\]

where \( x, y \in D([0,T], \mathbb{R}) \), and \( \Lambda \) is the set of all the strictly increasing continuous functions mapping \([0,T]\) onto itself. The topology on \( D([0,T], \mathbb{R}) \) induced by the \( J_1 \)-metric is called the Skorokhod \( J_1 \)-topology.

Skorokhod [17] also defined the \( M_1 \)-metric using the notion of completed graph of a function. More precisely, for any \( x \in D([0,T], \mathbb{R}) \), the *completed graph* of \( x \), denoted by \( \Gamma_x \), is defined as

\[
\Gamma_x := \{(t,z) \in [0,T] \times \mathbb{R} : z \in [x(t-), x(t))\},
\]

where \( x(0-) \) is interpreted as \( x(0) \), \([z_1, z_2]\) is the line segment connecting \( z_1 \) and \( z_2 \), i.e.,

\[
[z_1, z_2] = \{z \in \mathbb{R} : z = az_1 + (1-a)z_2 \text{ for some } a \in [0,1]\}.
\]

Note that \( \Gamma_x \) can be parametrically represented by the following continuous function

\[
(r, u) : [0,1] \to \Gamma_z, \quad (r, u)(0) = (0, z(0)), (r, u)(1) = (T, z(T)),
\]

which is nondecreasing with respect to the following order on \( \Gamma_z \):

\[
(t_1, z_1) \leq (t_2, z_2) \iff t_1 < t_2 \quad \text{or} \quad t_1 = t_2 \text{ and } |x(t_1-)-z_1| \leq |x(t_2-)-z_2|.
\]

Skorokhod [17] defined the \( M_1 \)-metric as follows:

\[
\mathcal{D}_{M_1}(x,y) = \inf_{(r_1, u_1) \in \Pi(x),(r_2, u_2) \in \Pi(y)} \left\{ \sup_{0 \leq t \leq 1} |r_1(t) - r_2(t)| + \sup_{0 \leq t \leq 1} |u_1(t) - u_2(t)| \right\},
\]
where \( x, y \in D([0, T], \mathbb{R}) \), and \( \Pi(\cdot) \) denotes the set of all parametric representations of an element in \( D([0, T], \mathbb{R}) \). The topology on \( D([0, T], \mathbb{R}) \) induced by the \( M_1 \)-metric is called the Skorokhod \( M_1 \)-topology.

Noting that the limit of a sequence of continuous functions under either the uniform or the Skorokhod \( J_1 \)-topology is continuous, we remark that, when approximating a càdlàg function using continuous functions, the Skorokhod \( M_1 \)-topology can be particularly useful. For example, for any \( n \geq 1 \), let

\[
x(t) = I_{[1/2, 1]}(t), \quad x^n(t) = n(t-1/2+1/n)I_{[1/2-1/n, 1/2]}(t) + I_{[1/2, 1]}(t), \quad 0 \leq t \leq 1.
\]

One can verify that, as \( n \) tends to infinity, \( x^n(t) \to x(t) \) in \( D([0, 1], \mathbb{R}) \) under Skorokhod \( M_1 \)-topology but not under the Skorokhod \( J_1 \)-topology.

The following theorem is well known; see, e.g., [14, Theorem 3.2].

**Theorem A.1.** Let \( W = \{W(t); 0 \leq t \leq T\} \) and \( W^n = \{W^n(t); 0 \leq t \leq T\}, n = 1, 2, \ldots, \) be stochastically continuous càdlàg stochastic processes. Then, as \( n \) tends to infinity, \( W^n \) converges in probability to \( W \) under the Skorokhod \( M_1 \)-topology if and only if

1) for any \( t \in [0, T] \), \( W^n(t) \) converges in probability to \( W(t) \);

2) and for any fixed \( \varepsilon > 0 \),

\[
\lim_{\delta \to 0^+} \limsup_{n} P(\rho(W^n, \delta) > \varepsilon) = 0,
\]

where \( \rho : D([0, T], \mathbb{R}) \times \mathbb{R} \to \mathbb{R} \) is defined as

\[
\rho(x, \delta) := \sup_{0 \leq t_1 < t_2 \leq 1} \inf_{a \in [0, 1]} \left| x(t_1) - (ax(t_1) + (1-a)x(t_2)) \right| \quad (A.16)
\]

**Acknowledgments.** The authors acknowledge the support provided by NSFs of China (No. 11971186).

**REFERENCES**

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge Studies in Advanced Mathematics, 116, Cambridge University Press, Cambridge, 2009.

[2] R. B. Ash, *Probability and Measure Theory*, Harcourt/Academic Press, Burlington, MA, 2000.

[3] T. Fujiwara and H. Kunita, Canonical SDE’s based on semimartingales with spatial parameters. I. Stochastic flows of diffeomorphisms, *Kyushu J. Math.*, 53 (1999), 265–300.

[4] M. Hairer and É. Pardoux, A Wong-Zakai theorem for stochastic PDEs, *J. Math. Soc. Japan*, 67 (2015), 1551–1604.

[5] R. Hinze and I. Pavlyukevich, Small noise asymptotics and first passage times of integrated Ornstein-Uhlenbeck processes driven by \( \alpha \)-stable Lévy processes, *Bernoulli*, 20 (2014), 265–281.

[6] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Mathematical Library, 24, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, 1989.

[7] H. Kunita, Stochastic differential equations with jumps and stochastic flows of diffeomorphisms, in *Itô’s Stochastic Calculus and Probability Theory*, Springer, Tokyo, 1996, 197–211.

[8] H. Kunita, Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms, in *Real and Stochastic Analysis*, Trends Math., Birkhäuser, Boston, MA, 2004, 305–373.

[9] T. G. Kurtz, Random time changes and convergence in distribution under the Meyer-Zheng conditions, *Ann. Probab.*, 19 (1991), 1010–1034.

[10] T. G. Kurtz, É. Pardoux and P. Protter, Stratonovich stochastic differential equations driven by general semimartingales, *Ann. Inst. H. Poincaré Probab. Statist.*, 31 (1995), 351–377.

[11] T. G. Kurtz and P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, *Ann. Probab.*, 19 (1991), 1035–1070.
[12] T. G. Kurtz and P. E. Protter, Weak convergence of stochastic integrals and differential equations, in Probabilistic Models for Nonlinear Partial Differential Equations, Lecture Notes in Math., 1627, Fond. CIME/CIME Found. Subser., Springer, Berlin, 1996, 1–41.

[13] S. I. Marcus, Modelling and approximation of stochastic differential equations driven by semimartingales, Stochastics, 4 (1980/81), 223–245.

[14] I. Pavlyukevich and M. Riedle, Non-standard Skorokhod convergence of Lévy-driven convolution integrals in Hilbert spaces, Stoch. Anal. Appl., 33 (2015), 271–305.

[15] P. Protter, Stochastic Integration and Differential Equations, Stochastic Modelling and Applied Probability, 21, Springer-Verlag, Berlin, 2005.

[16] A. A. Puhalskii and W. Whitt, Functional large deviation principles for first-passage-time processes, Ann. Appl. Probab., 7 (1997), 362–381.

[17] A. V. Skorokhod, Limit theorems for stochastic processes, Teor. Veroyatnost. i Primenen., 1 (1956), 289–319.

[18] G. Tessitore and J. Zabczyk, Wong-Zakai approximation of stochastic evolution equations, J. Evol. Equ., 6 (2006), 621–655.

[19] W. Whitt, Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Application to Queues, Springer Series in Operations Research, Springer-Verlag, New York, 2002.

[20] W. Whitt, Weak convergence of first passage time processes, J. Appl. Probability, 8 (1971), 417–422.

[21] E. Wong and M. Zakai, On the relation between ordinary and stochastic differential equations, Internat. J. Engrg. Sci., 3 (1965), 213–229.

[22] E. Wong and M. Zakai, On the convergence of ordinary integrals to stochastic integrals, Ann. Math. Statist., 36 (1965), 1560–1564.

[23] X. Zhang, Derivative formulas and gradient estimates for SDEs driven by $\alpha$-stable processes, Stochastic Process. Appl., 123 (2013), 1213–1228.

Received August 2019; revised March 2020.

E-mail address: xmliu@hust.edu.cn
E-mail address: ghan@hku.hk