Convergence of the semi-implicit Euler method for neutral stochastic delay differential equations with phase semi-Markovian switching

Baojian Yin a,⇑, Zhonghua Ma b

a School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China
b College of Basic Sciences, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Article history:
Received 1 August 2009
Received in revised form 26 September 2010
Accepted 12 November 2010
Available online xxxx

Keywords:
Numerical analysis
Semi-implicit Euler–Maruyama method
Neutral stochastic delay differential equations
Phase semi-Markovian switching

ABSTRACT

Recently, in the numerical analysis for stochastic differential equations (SDEs), it is a new topic to study the numerical schemes of neutral stochastic functional differential equations (NSFDEs) (see Wu and Mao [1]). Especially when Markovian switchings are taken into consideration, these problems will be more complicated. Although Zhou and Wu [2] develop a numerical scheme to neutral stochastic delay differential equations with Markovian switching (short for NSDDEwMSSs), their method belongs to explicit Euler–Maruyama methods which are in general much less accurate in approximation than their implicit or semi-implicit counterparts. Therefore, to propose an implicit method becomes imperative to fill the gap. In this paper we will extend Zhou and Wu [2] to the case of the semi-implicit Euler–Maruyama methods and equations with phase semi-Markovian switching rather than Markovian switchings. The employment of phase semi-Markovian chains can avoid the restriction of the negative exponential distribution of the sojourn time at a state. We prove the semi-implicit Euler solution will converge to the exact solution to NSDDEwMSS under local Lipschitz condition. More precise inequalities and new techniques are put forward to overcome the difficulties for the existence of the neutral part.

1. Introduction

Neutral stochastic functional differential equations (NSFDEs) were introduced by Kolmanovskii and Myshkis [3] and had successful applications in chemical and aero-elasticity. They are stochastic functional differential equations (SFDEs) with variable delay argument (called the neutral part) occurring in the state variable. On one hand, most SFDEs, just like the general stochastic differential equations (SDEs), cannot be solved explicitly, so many researchers have turned to the numerical analysis. Here we mention some of them, for example, [4–9] and some references therein. Moreover, by including Markovian switchings, Poisson jumps or both, the numerical solutions of more general SFDEs have been discussed extensively; see [10–20], a few to name. On the other hand, much different from SFDEs, it is the existence of the neutral part in NSFDEs that enables their numerical schemes to be more difficult to be proposed than those of SFDEs, which has been revealed by the recent related literature. So far, only a couple of papers have touched the numerical analysis of NSFDEs. Wu and Mao [1] established the strong mean square convergence theory of the general NSFDEs.

In practical applications, time-delays are frequently encountered, such as engineering, communications and biological systems and may induce instability, oscillation, and poor performance. A special class of NSFDEs are the neutral stochastic delay differential equations (NSDDE). In particular, when Markovian switchings are taken into consideration, the numerical
solutions will be much more complicated. Although Zhou and Wu [2] showed the convergence of numerical solutions to neutral stochastic delay differential equations with Markovian switching, their method belongs to the classic explicit Euler–Maruyama ones and their equations are with the common Markovian switchings. Markovian switching systems have many limitations in applications which originate from some serious weaknesses in Markov processes themselves (for some properties of Markov processes see losifescu [21]).

In general, the switching time of a Markov chain is the exponential distribution whose pervasiveness in stochastic systems is seldom supported by empirical evidence. To loosen this condition will lead to the lack of memory property. Moreover, in a Markov process, the transition from one state to itself is impossible. Therefore, to some extent, existing results obtained on Markovian switching systems are conservative. Fortunately, many researchers turn to the so-called semi-Markovian switching systems, one of which has received great attention is called phase-type semi-Markovian switching systems or PH semi-Markovian switching systems in which exponential distribution has been replaced by PH-distribution. PH-distribution is a generalization of exponential distribution and preserves much of analytic tractability of it. PH-distribution was first introduced by Neuts [23] in 1975. The problem of stochastic stability for linear systems with semi-Markovian switching parameters was first considered by Hou and Luo [24]. Hou et al. [25] further studied Itô differential equations with semi-Markovian switching parameters and established the comparison principle which is very useful in obtaining stochastic stability criteria.

In this paper we will extend Zhou and Wu [2] to the case of the semi-implicit Euler–Maruyama methods and equations with phase semi-Markovian switching. By choosing the proper parameter theta, the semi-implicit methods acquire greater flexibility and hence are more accurate in approximation than their explicit counterparts. In addition, the employment of phase semi-Markovian chains can avoid the restriction of the negative exponential distribution of the sojourn time at a state. Thirdly, more precise inequalities will be applied and new techniques will be developed in dealing with the difficulties for the existence of the neutral part.

The rest of the paper is organized as follows. Section 2 begins with notation and problem. Phase semi-Markov process and its simulation are in Section 3. Section 4 is for the introduction of the semi-implicit Euler–Maruyama method. Section 5 is devoted to many preliminary lemmas. In Section 6 the convergence of numerical solutions will be proved in the sense of mean square.

2. Notation and problem

Throughout the paper, unless otherwise specified, let \( (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}) \) be a complete probability space with a filtration satisfying the usual conditions (i.e., the filtration is continuous on the right and \( \mathcal{F}_t \) contains all \( \mathbb{P} \)-null sets). Let \( w(t) = (w_1(t), \ldots, w_n(t))^T \) be the \( m \)-dimensional Brownian motion defined on the probability space. Let \( C([\tau, 0]; \mathbb{R}^d) \) denote the family of continuous functions \( \phi \) from \( [\tau, 0] \) to \( \mathbb{R}^d \) with the norm definition \( \| \phi \| = \sup_{\tau \leq t \leq 0} \| \phi(t) \| \), where \( \| \cdot \| \) is Euclidean norm in \( \mathbb{R}^d \), i.e., \( |x| = \sqrt{x^T x} \), \( x \in \mathbb{R}^d \). If \( A \) is a matrix, its trace norm is denoted by \( |A| = \sqrt{\text{trace}(A^T A)} \), while its operator norm is denoted by \( \| A \| = \sup \{ |Ax| : |x| = 1 \} \). Moreover, let \( p \geq 2 \), \( t \geq 0 \) and \( L^p_{\mathcal{F}_t}([\tau, 0]; \mathbb{R}^d) \) denote the family of all \( \mathcal{F}_t \)-measurable \( C([\tau, 0]; \mathbb{R}^d) \)-valued random variables \( \xi \) such that \( \mathbb{E}[|\xi|^p] < \infty \). For simplicity, we also denote \( \text{a} \wedge \text{b} = \min\{a,b\} \) and \( \text{a} \vee \text{b} = \max\{a,b\} \).

Consider a d-dimensional Neutral Stochastic Delay Differential Equations with semi-Markovian switching

\[
d[x(t) - \hat{u}(x(t - \tau), \hat{r}(t))] = \hat{f}(x(t), x(t - \tau), \hat{r}(t))dt + \hat{g}(x(t), x(t - \tau), \hat{r}(t))d\hat{w}(t),
\]

on \( t \geq 0 \) with the initial date \( x(0) = \xi \in L^p_{\mathcal{F}_0}([\tau, 0]; \mathbb{R}^d) \) and \( \hat{r}(0) = r_0 \in \hat{S} \). \( \hat{r}(t) \) is so-called phase semi-Markovian chain which takes values in a finite state set \( \hat{S} = \{1, 2, \ldots, m\} \) and is assumed to be independent of the Brownian motion \( w(t) \). The definition and properties of \( \hat{r}(t) \) will be elaborated in next section. Moreover, \( \hat{u} : \mathbb{R}^d \times \hat{S} \to \mathbb{R}^d, \hat{f} : \mathbb{R}^d \times \mathbb{R}^d \times \hat{S} \to \mathbb{R}^d \) and \( \hat{g} : \mathbb{R}^d \times \mathbb{R}^d \times \hat{S} \to \mathbb{R}^{d \times m} \) are all Borel-measurable functions and satisfy the following usual hypotheses.

(H1)(Local Lipschitz Condition) For each integer \( R \geq 1 \), there is a positive constant \( C_R \) such that

\[
|\hat{f}(x_1, y_1, i) - \hat{f}(x_2, y_2, i)|^2 \leq C_R |x_1 - x_2|^2 + |y_1 - y_2|^2
\]

for those \( x_0, y_0 \in \mathbb{R}^d \) and \( i \in \hat{S} \) with \( |x_k| \vee |y_k| \leq R \) and \( k = 1, 2 \).

(H2)(Contractive Mapping Condition) There is a constant \( K \in (0, 1) \) such that for those \( y_k \in \mathbb{R}^d \) and \( i \in \hat{S} \),

\[
|\hat{u}(y_1, i) - \hat{u}(y_2, i)| \leq K |y_1 - y_2|,
\]

and \( \hat{u}(0, i) = 0 \) for \( i \in \hat{S} \).

3. Phase semi-Markov process and its simulation

This section is devoted to the introduction of phase semi-Markovian chain \( \hat{r}(t) \) and its simulation. First, let us recall some definitions and properties of phase type distribution or PH-distribution.
Definition 1. A probability distribution \( F(\cdot) \) on \([0, \infty)\) is a continuous distribution of phase type (PH-distribution), if and only if it is the distribution of the lifetime of a terminating Markov process \( \{r(t)\}_{t \geq 0} \) with finitely many states and time homogeneous transition rates.

Let us turn to details. Firstly, we define a terminating Markov process \( \{r(t)\} \) with state space \( S \) and intensity matrix \( T \). Then we say that \( \{r(t)\} \) is the restriction of a Markov Process \( \{\tilde{r}(t)\}_{t \geq 0} \) on \( S_0 = S \cup \{\delta\} \), where \( \delta \) is absorbing and the state in \( S \) is transient. Write \( m \) for the number of elements of \( S \), then denote an initial distribution by \( x = (a_1, a_2, \ldots, a_m) \) such that \( F \) is the distribution of the time \( \zeta = \inf \{t > 0, \tilde{r}(t) = \delta\} \) to absorption, or \( F(t) = P_x(\zeta \leq t) \). This implies the intensity matrix for \( \{\tilde{r}(t)\} \) can be written in block-partioned form by:

\[
Q = \begin{pmatrix} T & T^0 \\ 0 & 0 \end{pmatrix},
\]

where the matrix \( T = (T_{ij})_{m \times m} \) satisfies \( T_{ii} < 0, T_{ij} \geq 0, i \neq j \), and \( T^0 = (T^0_1, T^0_2, \ldots, T^0_m)^T \) is a nonnegative column vector. The pair \((x, T)\) is called the m-order representation of \( F(\cdot) \). Since \( Q \) is the intensity matrix of a non-terminating Markov process, the rows sum is equal to zero, or we have \( T^0 + Te = 0 \), where \( e \) is the column \( S \)-vector with all components equal to one, so rewrite \( T^0 = -Te \), which interprets \( T^0 \) as existence rate vector, i.e. the \( i \)th component \( T^0_i \) gives the intensity in state \( i \) for leaving \( S \) and going to the absorbing state \( \delta \).

Some basic properties of PH-distribution are gathered as the following both propositions.

Proposition 3.1 ([26]). The cumulative distribution function \( F(\cdot) \) with representation \((x, T)\) is given by:

\[
F(t) = 1 - xe^{Te}.
\]

Proof. Given \( x \) is the initial distribution and \( Q \) is the density matrix, apply the Kolmogorov backward equation, note that from (3.1) the upper left corner of \( Q^r \) is \( T^0_i \) implies the upper left corner of \( e^{Qr} \) is \( e^{th} \), so

\[
F(t) = 1 - P_x(\zeta > t) = 1 - P_x(\tilde{r}(t) \in S) = 1 - xe^{Te}. \quad \square
\]

Proposition 3.3 ([23]). The class of PH-distribution is dense (in the sense of weak convergence) in the set of all probability distributions on \([0, \infty)\).

Sketch of Proof. It is clear that any probability distribution on \([0, \infty)\) may be arbitrary closely and uniformly approximated by a finite mixture of degenerate distributions. Any degenerate distributions at \( x = a > 0 \), is the uniform limit of a sequence of Erlang distributions with mean \( a \) and increasing orders. Any probability distribution \( F(\cdot) \) on \([0, \infty)\), therefore, be obtained as the uniform limit of a sequence of probability distributions, each of which consists of a finite mixture of Erlang distributions and possibly a jump at 0. Such distributions are PH-distribution. \( \square \)

Remark 3.4. By Proposition 3.3, for every probability distribution on \((0, \infty)\), we may choose a PH-distribution to approximate the original distribution in any accuracy.

Due to the density of phase distribution in the families of all probability distribution on \( \mathbb{R}_+ \), we can define the following semi-Markov process.

Definition 2. Let \( \tilde{S} \) be a finite or countable set. A stochastic process \( \{\tilde{r}(t)\} \) on the state space \( \tilde{S} \) is called a phase semi-Markov process or a denumerable semi-Markov process (when \( \tilde{S} \) is finite, \( \tilde{S} \) is also called a finite phase semi-markov process), if the followings hold

1. The sample paths of \( \{\tilde{r}(t); t < +\infty\} \) are right-continuous step functions and have left-handed limits with probability one.
2. Denote the nth jump point of the process \( \tilde{r}(t) \) by \( \tau_n(n = 0, 1, \ldots) \), where \( \tau_0 \equiv 0 < \tau_1 < \cdots < \tau_n \to +\infty \), and \( \tilde{r}(t) \) possesses Markov property at each \( \tau_n(n = 0, 1, \ldots) \).
3. \( F_{ij}(t) = P(\tau_{n+1} - \tau_n \leq t | \tilde{r}(\tau_n) = i, \tilde{r}(\tau_{n+1}) = j) = F_i(t)(i,j \in \tilde{S}, t \geq 0) \) do not depend on \( j \) and \( n \).
4. \( F_i(t)(i \in \tilde{S}) \) is a PH-distribution.

Remark 3.5. By this definition, in the case that $F_i(t) (i \in \tilde{S})$ is a negative exponential distribution, the phase semi-Markovian process becomes a Markov chain.

Next, we will show that the finite phase semi-Markovian process $\tilde{r}(t)$ could be transformed into a finite Markov chain, which is called Markovization.

Theorem 3.6. Given any finite phase semi-Markovian process $\tilde{r}(t)$, there exists an associated finite Markov chain.

Proof. Suppose that $\tilde{S} = \{1, 2, ..., m\}$ is a finite nonempty set and $\{\tilde{r}(t)\}$ is a given finite phase semi-Markov process on the state space $\tilde{S}$. Denote the $n$th jump point of the process $\{\tilde{r}(t)\}$ by $\tau_n (n = 0, 1, ...)$, where $\tau_0 = 0 < \tau_1 < ... < \tau_n + \infty$. Let $F_i(t) (i \in \tilde{S})$ be the distribution of the sojourn time of the process at state $i$. Moreover, for each state $i \in \tilde{S}$, let $\{\tilde{x}(0), \tilde{T}(0)\}$ denote the corresponding $m^{th}$ order representation of $F_i(t)$, and $\tilde{S}^0$ be the transient set, where $m^{th}$ denotes the number of the elements in $\tilde{S}^0$.

$$F_i(t) = P(\tau_{n+1} - \tau_n \leq t | \tilde{r}(\tau_n) = i) (i \in \tilde{S}),$$

$$x^{(i)} = \{a_1^{(i)}, a_2^{(i)}, ..., a_{m^{th}}^{(i)}\},$$

$$T^{(i)} = (T_{0}^{(i)}, \{i, j \in \tilde{S}^0\}).$$

Let $P = \{P_{ij}, i, j \in \tilde{S}\}$, where

$$P_{ij} = P(\tilde{r}(\tau_{n+1}) = j | \tilde{r}(\tau_n) = i)(i, j \in \tilde{S}),$$

$$\{x, T\} = \{\{x^{(i)}, T^{(i)}\}, i \in \tilde{S}\}.$$

It is ready to see that the probability distribution of $\tilde{r}(t)$ can be determined only by $(P, \{x, T\})$.

For any $i \in \tilde{S}$, we define

$$T_j^{(i)} = - \sum_{k=1}^{m^{th}} T_{jk}^{(i)},$$

$$G = \{ (i, k^{(i)}) | i \in \tilde{S}, k^{(i)} = 1, 2, ..., m^{th} \}.$$  

From the above description, it is easy to know that the process $\tilde{r}(t)$ is not a Markov process, unless the sojourn time in each state is exponentially distributed. But we can show that considering the process $\tilde{r}(t)$ only at the jump points $\tau_n$ yields a (discrete-time) Markov process. Moreover, the behavior of $\tilde{r}(t)$ is piecewise deterministic in the intervals between jump points. For the Markovization of $\tilde{r}(t)$, we therefore have to add the information on the neighboring jump points. Hence, to proceed with our study, let $H(t)$ denote the phase of $\tilde{r}(t)$ at time $t$, then we claim that $(\tilde{r}(t), H(t))$ is a Markov process with the state space $G$ and the generator $Q = \{q_{uv}\}, u, v \in G$ is given by:

$$q_{(i,k^{(i)})}(i,k^{(i)}) = T_{k^{(i)}k^{(i)}}^{(i)} - T_{k^{(i)}k^{(i)}}^{(i)};$$

$$q_{(i,k^{(i)})}(i,k^{(i)}) = T_{k^{(i)}k^{(i)}}^{(i)} - T_{k^{(i)}k^{(i)}}^{(i)};$$

$$q_{(i,k^{(i)})}(j,k^{(i)}) = P_{ij}T_{k^{(i)}k^{(i)}}^{(i)};$$

$$i \neq j, (i,k^{(i)}) \in G \text{ and } (j,k^{(i)}) \in G.$$  

In fact, for any $\Delta > 0$,

(1) For any $(i,k^{(i)}) \in G$, we have

$$P\{\tilde{r}(t + \Delta), H(t + \Delta) = (i,k^{(i)}) | \tilde{r}(t), H(t) = (i,k^{(i)})\} = 1 + T_{k^{(i)}k^{(i)}}^{(i)} \Delta + o(\Delta) = 1 + q_{(i,k^{(i)})}(i,k^{(i)}) \Delta + o(\Delta).$$

(2) For any $(i,k^{(i)}) \in G, (i,k^{(i)}) \in G$ and $k^{(i)} \neq k^{(i)}$, we have:

$$P\{\tilde{r}(t + \Delta), H(t + \Delta) = (i,k^{(i)}) | \tilde{r}(t), H(t) = (i,k^{(i)})\} = T_{k^{(i)}k^{(i)}}^{(i)} \Delta + o(\Delta) = q_{(i,k^{(i)})}(i,k^{(i)}) \Delta + o(\Delta).$$

(3) For any $(i,k^{(i)}) \in G, (j,k^{(i)}) \in G$ and $i \neq j$, we have:

$$P\{\tilde{r}(t + \Delta), H(t + \Delta) = (j,k^{(i)}) | \tilde{r}(t), H(t) = (i,k^{(i)})\} = P_{ij}T_{k^{(i)}k^{(i)}}^{(i)} \Delta + o(\Delta) = q_{(i,k^{(i)})}(j,k^{(i)}) \Delta + o(\Delta).$$

Since the set $G$ has $\sum_{i=1}^{m} m^{th}$ elements, the state space of $(\tilde{r}(t), H(t))$ has $s$ states, which are ranked in the alphabetic order:

$$(1, 1), (1, m^{(i)}), (i, 1), (i, m^{(i)}), (i + 1, 1), ..., (1 + 1, m^{(i+1)}), ..., (m, 1), ..., (m, m^{(i)}).$$

So following the above line, we further denote each $(i,k) \in G$ by its ordinal number, or more specifically...
\( \varphi((i,k)) = \sum_{n=1}^{i-1} m^{[n]} + k, \quad i \in \bar{S}, \quad 1 \leq k \leq m^{[0)}. \quad (3.16) \)

Furthermore, let
\[
\begin{align*}
\tau(t) & \doteq \varphi(\bar{r}(t), H(t)) \quad (3.17) \\
\gamma_{\varphi(i,k,i',k')} & \doteq \varphi(i,k,i',k') \quad (3.18)
\end{align*}
\]

Consequently, \( \tau(t) \) is an associated Markov process of the given phase semi-Markovian process \( \bar{r}(t) \) with the state space \( S = \{1, 2, \ldots, s\} \) and the generator \( \Gamma = (\gamma_{ij})_{s \times s} \), so that
\[
P(\tau(t + \Delta) | \tau(t)) = \begin{cases} \gamma_{ij} \Delta + 0(\Delta), & \text{if } i \neq j, \\ 1 + \gamma_{ii} \Delta + 0(\Delta), & \text{if } i = j. \end{cases} \quad (3.19)
\]

where \( \Delta > 0 \). Here \( \gamma_{ij} \geq 0 \) is the transition rate from \( i \) to \( j \) if \( i \neq j \) while \( \gamma_{ii} = -\sum_{j \neq i} \gamma_{ij} \). By the Definition 2, it is known that almost every sample path of \( \tau(t) \) is a right-continuous step function with a finite number of simple jumps in any finite subinterval of \( \mathbb{R}_+ \). \( \Box \)

**Example 3.7.** Now, let us give an example on the phase semi-Markov chain. For simplicity, \( \bar{r}(t) \) has only two states denoted by 1 and 2. The sojourn time in state 1 is a random variable with negative exponential distribution with parameter \( \lambda_1 \), while the sojourn time in state 2 is divided into two parts, and the sojourn times in the two parts are two random variables which are negative exponential distributed with parameters \( \lambda_2 \) and \( \lambda_3 \), respectively. More specially, if the process \( \bar{r}(t) \) enters the state 2, it must first stay at the first part of state 2 for some time and then stay at the second part for some time, at last it will enter the state 1. We assume that \( p_{12} = p_{21} = 1 \). Recalling some notations in (3.5)-(3.8), we have:
\[
P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (3.20)
\]
\[
x^{(1)} = (a_1^{(1)}) = (1), \quad T^{(1)} = (\tau_{11}^{(1)}) = (-\lambda_1), \quad (3.21)
\]
\[
x^{(2)} = (a_1^{(2)}, a_2^{(2)}) = (1, 0), \quad T^{(2)} = \begin{pmatrix} \tau_{11}^{(2)} & \tau_{12}^{(2)} \\ \tau_{21}^{(2)} & \tau_{22}^{(2)} \end{pmatrix} = \begin{pmatrix} -\lambda_2 & -\lambda_3 \\ -\lambda_2 & -\lambda_3 \end{pmatrix}, \quad (3.22)
\]

It is easy to get the state space of the associated Markov chain \( (\bar{r}(t), H(t)) \) can be expressed by \( G = ((1, 1), (2, 1), (2, 2)) \). Define
\[
\begin{align*}
\varphi((1, 1)) & = 1, \\
\varphi((2, 1)) & = 2, \\
\varphi((2, 2)) & = 3. \quad (3.23)
\end{align*}
\]

Therefore, the generator of the associated Markov chain \( (\bar{r}(t), H(t)) \) is
\[
\Gamma = (\gamma_{ij})_{3 \times 3} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 \\ 0 & -\lambda_2 & \lambda_2 \\ \lambda_3 & 0 & -\lambda_3 \end{pmatrix}. \quad (3.24)
\]

Now, to simulate the associated Markov chain \( \tau(t) \), we will need the following lemma (see [27]).

**Lemma 3.8.** Let \( r_k = r(k\Delta) \) for \( \Delta > 0 \) and \( k \geq 0 \), then \( \{ r_k, k = 0, 1, 2, \ldots \} \) is a discrete Markov chain with the one-step transition probability matrix
\[
P(\Delta) = (p_{ij}(\Delta))_{N \times N} = e^{\Delta T}. \quad (3.25)
\]

Since the \( \gamma_{ij} \) are independent of \( x \), the paths of \( r \) can be generated independently of \( x \) and, in fact, before computing \( x \).

Given a stepsize \( \Delta > 0 \), the discrete Markov chain \( \{ r_k, k = 0, 1, 2, \ldots \} \) can be simulated as follows: Compute the one-step transition probability matrix
\[
P(\Delta) = (p_{ij}(\Delta))_{N \times N} = e^{\Delta T}. \quad (3.26)
\]

Let \( r_0 = i_0 \) and generate a random number \( \xi_1 \) which is uniformly distributed in \([0, 1]\). Define
\[
r_1 = \begin{cases} i_1, & \text{if } i_1 \in S \setminus \{N\} \text{ such that } \sum_{j=1}^{i_1-1} P_{b,j}(\Delta) \leq \xi_1 < \sum_{j=1}^{i_1} P_{b,j}(\Delta), \\ N, & \text{if } \sum_{j=1}^{N-1} P_{b,j}(\Delta) \leq \xi_1, \end{cases} \quad (3.27)
\]
where we set $\sum_{j=1}^{n-1} P_{ij}(\Delta) = 0$ as usual. Generate independently a new random number $\xi_2$ which is again uniformly distributed in $[0,1]$ then define

$$r^k_2 = \begin{cases} l_2, & \text{if } l_2 \in S - \{N\} \text{ such that } \sum_{j=1}^{i-1} P_{ij}(\Delta) \leq \xi_1 < \sum_{j=1}^{i} P_{ij}(\Delta), \\ N, & \text{if } \sum_{j=1}^{N-1} P_{ij}(\Delta) \leq \xi_2. \end{cases}$$  

(3.28)

Repeating this procedure can generate a trajectory of $\{r^k_2, k = 0, 1, 2, \ldots\}$. This procedure can be carried out independently to obtain more trajectories.

4. The semi-implicit Euler method

After explaining how to simulate the discrete Markov chain $\{r^k_2, k = 0, 1, 2, \ldots\}$, we can now define the semi-implicit Euler-Maruyama method. Having Theorem 3.6 in mind, the first step is to show how to change the difficult original Eq. (2.1) into a simpler one we are familiar with.

For any $x, y \in \mathbb{R}^d$, $i \in S$ and $1 \leq k \leq m$, we define the function

$$u(y, \varphi((i,k))) = \tilde{u}(y, i), f(x, y, \varphi((i,k))) = \tilde{f}(x, y, i), g(x, y, \varphi((i,k))) = \tilde{g}(x, y, i),$$  

(4.1)

where the function $\varphi$ is defined in (3.16), so it is evident that $u : \mathbb{R}^d \times S \to \mathbb{R}^d$, $f : \mathbb{R}^d \times \mathbb{R}^d \times S \to \mathbb{R}^d$, $g : \mathbb{R}^d \times \mathbb{R}^d \times S \to \mathbb{R}^{d \times m}$ and they all satisfy conditions (H1) and (H2).

In fact, we have the following theorem.

**Theorem 4.1.** System (2.1) is equivalent to the system

$$\begin{cases}
    d\xi(t) = [\varphi(\xi(t), r(t))] dt + g(\xi(t), r(t)) dw(t), \\
    x(0) = x^0 + \int_{0}^{\tau} g(\xi(s), r(s)) dw(s), t \in [\tau, 0], R^d.
\end{cases}$$  

(4.2)

where $r(t)$ is an associated Markov process of the phase semi-Markovian process $\hat{r}(t)$ with the state space $S = \{1, 2, \ldots, s\}$, $s = \sum_{i=1}^{s} m^i$ and $r(0) = \varphi(t_0, k_0)$ = $\sum_{i=1}^{s} m^i + k_0$, $t_0 \in S$, $1 \leq k_0 \leq m^{s_0}$.

From now on, our emphasis will be put on the transformed Eq. (4.2) rather than the original Eq. (2.1).

Given a stepsize $\Delta \in (0, 1)$, which satisfies $\tau = m\Delta$ for a some positive integer $m$, let $t_k = k\Delta$ for $k > 0$. Compute the discrete approximations $y_k = x(t_k)$ by setting $y_0 = x(0)$ and $r^0_0 = l_0$, and performing

$$y_{k+1} = y_k - u(y_k, r^k_1) + \int_{0}^{\Delta} [1 - \varphi(\xi(t), r(t))] dt + g(\xi(t), r(t)) dw(t),$$  

(4.3)

where $0 \leq \varphi \leq 1$, $\Delta t_k = w(t_{k+1}) - w(t_k)$. Let us introduce the following notations $z_i(t) = y_i$, $z_i(t) = y_i$, $z_i(t) = y_i$, $z_i(t) = y_i$, $z_i(t) = y_i$, and then for $i \in \{t_k, t_{k+1}\}$ with the initial value $z_i(t) = \tilde{\xi}(t)$ on $[-\tau, 0]$, $i = 1, 2$.

Then we can extend the discrete numerical solution to the continuous one as follows:

$$y(t) = \begin{cases}
    \tilde{\xi}(t), & t \in [-\tau, 0], \\
    \tilde{u}(z_1(t), r_1(t)) + \tilde{\xi}(z_1(t), r_1(t)) + g(z_1(t), z_1(t), r_1(t)) dz_1 \int_{0}^{t} \tilde{f}(z_1(s), z_1(s), r_1(s)) ds \\
    + \int_{0}^{t} \tilde{f}(z_1(s), z_1(s), r_1(s)) dz_1 \int_{0}^{\Delta} f(z_1(s), z_1(s), r_1(s)) dz_1 \int_{0}^{\Delta} g(z_1(s), z_1(s), r_1(s)) dw(s), t \in [0, T].
\end{cases}$$  

(4.4)

so we have for any $t \geq 0$,

$$y(t) = \tilde{u}(z_1(t), r_1(t)) + \tilde{\xi}(z_1(t), r_1(t)) + \tilde{\xi}(z_1(t), r_1(t)) + \int_{0}^{\Delta} g(z_1(s), z_1(s), r_1(s)) dz_1 \int_{0}^{t} \tilde{f}(z_1(s), z_1(s), r_1(s)) dz_1 \int_{0}^{\Delta} f(z_1(s), z_1(s), r_1(s)) dz_1 \int_{0}^{\Delta} g(z_1(s), z_1(s), r_1(s)) dw(s), t \in [0, T].$$

(4.5)

It is then obvious that $y(t_k) = x(t_k) = z_1(t_k) = y_k$, that is, $y(t)$ and $z_1(t)$ coincide with the discrete approximate solution at the gridpoints.

The primary aim of this paper is to establish the following main result.

**Theorem 4.2.** The approximations $y(t)$ and $z_1(t)$ coincide with the discrete approximate solution at the gridpoints.

Theorem 4.2. Under the local lipschitz condition (H1) and contractive mapping condition (H2), the semi-implicit Euler approximate solution (2.10) converges to the exact solution of the SDDEwMSs (2.2) in the sense
\[
\lim_{\Delta \to 0} E \left( \sup_{0 \leq t \leq T} |x(t) - y(t)|^2 \right) = 0.
\]  
(4.5)
Since the proof of this theorem is rather technical, some preliminary lemmas will be reported in next section, while a complete proof will be presented in Section 6.

5. Preliminary lemmas

From Theorem 4.1, condition (H1) and inequality \(|a+b|^2 \leq 2(|a|^2 + |b|^2)|, it is easy to obtain that for all \( |x_k| \vee |y_k| \leq R (k = 1, 2)\),
\[
|f(x, y, i)|^2 \leq 2|f(x, y, i) - f(0, 0, i)|^2 + 2|f(0, 0, i)|^2 \leq 2C_h(|x|^2 + |y|^2) + 2|f(0, 0, i)|^2 \leq C_k(1 + |x|^2 + |y|^2).
\]  
(5.1)
Likewise, \( |g(x, y)|^2 \leq C_k(1 + |x|^2 + |y|^2)\), where \( C_k = 2(C_h \vee |f(0, 0, i)|^2 \vee |g(0, 0, i)|^2)\). Moreover, by condition (H2), \( |u(y)|^2 \leq K|y|^2\).

Thus, analogous to the proof of Mao [13], we claim: Under conditions (H1) and (H2), system (1) has a unique continuous solution \( x(t) \) on \( t > -\tau \) and the following lemma holds.

Lemma 5.1. Under (H1) and (H2), there exist constants \( p \geq 2 \) and \( C_1 > 0 \) such that
\[
E \left( \sup_{-t \leq s \leq t} |x(t, \xi)|^p \right) \leq C_1.
\]  
(5.2)
To proceed, let us define the following stopping times (As usual we set \( \inf \emptyset = \infty \)):
\[
\sigma_k = \inf \{ t \geq 0 : |y(t)| \geq R \}, \quad \tau_k = \inf \{ t \geq 0 : |x(t)| \geq R \}, \quad \rho_k = \gamma_k \wedge \sigma_k.
\]  
(5.3)

Lemma 5.2. Under the conditions (H1) and (H2),
\[
E \left( \sup_{-t \leq s \leq t} |y(t \wedge \rho_k)|^2 \right) \leq C_2,
\]  
(5.4)
where \( C_2 \) is a positive constant independent of \( \Delta \).

Proof. For \( t \in [t_k, t_{k+1}]\), set \( \bar{y}(t) = y(t) - u(z_1(t - \tau), r_1(t)) \), then for any \( t \in [0, T] \), by the definition of \( y(t) \), we have:
\[
E \left( \sup_{-t \leq s \leq t} |y(t \wedge \rho_k)|^2 \right) \leq E\|\bar{y}\|^2 + E \left( \sup_{-t \leq s \leq t} |y(t \wedge \rho_k)|^2 \right)
\]  
(5.5)
and
\[
\bar{y}(t \wedge \rho_k) = \bar{y}(0) + \int_0^{t \wedge \rho_k} \left[ (1 - \theta) f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s)) \right] ds + \int_0^{t \wedge \rho_k} g(z_1(s), z_1(s - \tau), r_1(s)) dw(s)
\]  
(5.6)
with the initial value \( \bar{y}(0) = \xi(0) - u(\bar{z}(0), r_0) \).

Recall the inequality that for any \( \varepsilon > 0 \),
\[
|x + y|^2 \leq (1 + \varepsilon)|x|^2 + \frac{1 + \varepsilon}{\varepsilon}|y|^2,
\]  
(5.7)
which together with condition (H2) leads to:
\[
|y(t \wedge \rho_k)|^p \leq (1 + \varepsilon)|\bar{y}(t \wedge \rho_k)|^p + \frac{1 + \varepsilon}{\varepsilon}|u(z_1(t \wedge \rho_k - \tau), r_1(t))|^p
\leq (1 + \varepsilon)|\bar{y}(t \wedge \rho_k)|^p + \frac{1 + \varepsilon}{\varepsilon}K^2 \left( \sup_{-t \leq s \leq t} |y(s \wedge \rho_k)|^2 \right).
\]  
(5.8)
Since \( 0 < K \leq 1 \), we can take \( \varepsilon = \frac{K}{1 - K} \), then
\[
|y(t \wedge \rho_k)|^2 \leq \frac{|\bar{y}(t \wedge \rho_k)|^p}{1 - K} + K \left( \sup_{-t \leq s \leq t} |y(s \wedge \rho_k)|^p \right).
\]  
(5.9)
Connecting (5.9) with (5.5) gets
\[
E\left(\sup_{-T \leq t \leq T} |y(t \wedge \rho_k)|^2\right) \leq E\|\xi\|^2 + E\left(\sup_{0 \leq t \leq T} |y(t \wedge \rho_k)|^2\right) + E\left(\sup_{0 \leq t \leq T} |\bar{y}(t \wedge \rho_k)|^2\right) + KE\left(\sup_{-T \leq t \leq T} |y(t \wedge \rho_k)|^2\right).
\]
(5.10)

which can yield
\[
E\left(\sup_{-T \leq t \leq T} |y(t \wedge \rho_k)|^2\right) \leq E\|\xi\|^2 + \frac{E\left(\sup_{0 \leq t \leq T} |\bar{y}(t \wedge \rho_k)|^2\right)}{1 - K}.
\]
(5.11)

By (5.6), applying inequality \(|a + b + c|^2 \leq 3|a|^2 + |b|^2 + |c|^2\) yields
\[
E\left[\sup_{0 \leq t \leq T} |\bar{y}(t \wedge \rho_k)|^2\right] \leq 3E\|\bar{y}\|^2 + 3E \sup_{0 \leq t \leq T} \left|\int_0^{t \wedge \rho_k} \left[(1 - \theta)\bar{f}(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s))\right]ds\right|^2
\]
\[
+ 3E \sup_{0 \leq t \leq T} \left|\int_0^{t \wedge \rho_k} g(z_1(s), z_1(s - \tau), r_1(s))ds\right|^2.
\]
(5.12)

Using condition (H2) gets
\[
E\|\bar{y}\|^2 = E\|\xi\|^2 - u(\xi(-\tau), r_0) \leq E\|\bar{y}\|^2 \leq E\|\xi\|^2 + K\|\xi\|^2 \leq (1 + K)^2 E\|\xi\|^2.
\]
(5.13)

By the Hölder inequality, Fubini's Theorem, condition (5.1) and \(0 \leq \theta < 1\), we have:
\[
E \sup_{0 \leq t \leq T} \left|\int_0^{t \wedge \rho_k} \left[(1 - \theta)\bar{f}(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s))\right]ds\right|^2
\]
\[
\leq TE \sup_{0 \leq t \leq T} \left|\int_0^{t \wedge \rho_k} \left[f(z_1(s), z_1(s - \tau), r_1(s))\right]ds\right|^2
\]
\[
\leq 2TE \sup_{0 \leq t \leq T} \left[\int_0^{t \wedge \rho_k} \left[f(z_1(s), z_1(s - \tau), r_1(s))\right]^2 + \left[f(z_2(s), z_2(s - \tau), r_2(s))\right]^2 \right]ds
\]
\[
\leq 4T^2 C_k + 2TC_k E \sup_{0 \leq t \leq T} \left[\int_0^{t \wedge \rho_k} \left|z_1(s)^2 + |z_1(s - \tau)|^2 + |z_2(s)|^2 + |z_2(s - \tau)|^2 \right]ds
\]
\[
\leq 4T^2 C_k + 8TC_k \int_0^T E \left(\sup_{-T \leq s \leq s \wedge \rho_k} |y(s)|^2\right)ds.
\]
(5.14)

Using the Doob inequality and condition (5.1) implies:
\[
E \sup_{0 \leq t \leq T} \left|\int_0^{t \wedge \rho_k} g(z_1(s), z_1(s - \tau), r_1(s))ds\right|^2 \leq 4E \int_0^{t \wedge \rho_k} |g(z_1(s), z_1(s - \tau), r_1(s))|^2ds
\]
\[
\leq 4C_k E \int_0^{t \wedge \rho_k} (1 + |z_1(s)|^2)^2 + |z_1(s - \tau)|^2)ds
\]
\[
\leq 4C_k + 8C_k \int_0^T E \left(\sup_{-T \leq s \leq s \wedge \rho_k} |y(s)|^2\right)ds.
\]
(5.15)

Inserting (5.11)-(5.15) in (5.10) gives:
\[
E\left[\sup_{-T \leq t \leq T} |y(t \wedge \rho_k)|^2\right] \leq E\|\xi\|^2 + \frac{3}{(1 - K)^2} \left(1 + K\right)^2 E\|\bar{y}\|^2 + 4T^2 C_k
\]
\[
+ 8TC_k \int_0^T E \left(\sup_{-T \leq s \leq s \wedge \rho_k} |y(s)|^2\right)ds + 4C_k \int_0^T E \left(\sup_{-T \leq s \leq s \wedge \rho_k} |y(s)|^2\right)ds + 2C_k T \int_0^T E \left(\sup_{-T \leq s \leq s \wedge \rho_k} |y(s)|^2\right)ds.
\]
(5.16)

where \(\tilde{c}_k = 12C_k(1 + T) > 0\).

Finally, applying the Gronwall inequality completes the proof as follows:
\[
E\left[\sup_{-T \leq t \leq T} |y(t \wedge \rho_k)|^2\right] \leq \frac{1 + K}{1 - K} \left(1 + K\right)^2 E\|\bar{y}\|^2 + \tilde{c}_kT e^{\tilde{c}_kT} = C_2.
\]
From the expressions of $C_2$ and $c_k$ above, we see that they are both positive constants dependent only on $\xi$, $K$, $C_k$ and $T$ but independent of $\Delta$. The assertion must hold. \hfill \Box

**Lemma 5.3.** Under the conditions (H1) and (H2), for any $t \in [0, T]$, there exists a positive constant $C_3$ independent of $\Delta$, such that:

\begin{align}
\int_0^{t_{k+1}} |E[y(s) - z_1(s)]^2| ds &\leq C_3 \Delta. \quad (5.17) \\
\int_0^{t_{k+1}} |E[y(s) - z_2(s)]^2| ds &\leq C_3 \Delta. \quad (5.18)
\end{align}

**Proof.** For any $t \in [t_k, t_{k+1})$, on one hand, by the definition of $y(t)$,

\begin{align}
y(t) &= u(z_1(t - \tau), r_1(t)) + \xi(0) - u(\xi(-\tau), r_0^k) + \int_0^{t} [(1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s))] ds \\
&\quad + \int_0^t g(z_1(s), z_1(s - \tau), r_1(s)) dw(s). \quad (5.19)
\end{align}

On the other hand, noting that $z_1(t) = y_k$, we have:

\begin{align}
z_1(t) &= u(y_k, r_0^k) + \xi(0) - u(\xi(-\tau), r_0^k) + \int_0^{t_{k+1}} [(1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s))] ds \\
&\quad + \int_0^{t_{k+1}} g(z_1(s), z_1(s - \tau), r_1(s)) dw(s). \quad (5.20)
\end{align}

For $t \in [t_k, t_{k+1})$, recalling the following notations: $z_1(t) = y_k$, $z_1(t - \tau) = y_{k-1}$, $z_2(t) = y_{k+1}$, $z_2(t - \tau) = y_{k+1-1}$, $r_1(t) = r_0^k$, and $r_2(t) = r_0^{k+1}$, we have:

\begin{align}
y(t) - z_1(t) &= \int_0^t [(1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s))] ds \\
&\quad + \int_0^t g(z_1(s), z_1(s - \tau), r_1(s)) dw(s) \\
&= [(1 - \theta)f(z_1(t), z_1(t - \tau), r_1(t)) + \theta f(z_2(t), z_2(t - \tau), r_2(t))] (t - t_k) + g(z_1(t), z_1(t - \tau), r_1(t)) (w(t) - w(t_k)). \quad (5.21)
\end{align}

Applying inequality $|a + b|^2 \leq 2(|a|^2 + |b|^2)$ and condition (5.1) yields:

\begin{align}
y(t) - z_1(t) &\leq 4|f(z_1(t), z_1(t - \tau), r_1(t))|^2 + |f(z_2(t), z_2(t - \tau), r_2(t))|^2 (t - t_k)^2 + 2|g(z_1(t), z_1(t - \tau), r_1(t))|^2 (w(t) - w(t_k))^2 \\
&\leq 4[C_k (1 + |z_1(t)|^2 + |z_1(t - \tau)|^2) + C_k (1 + |z_2(t)|^2 + |z_2(t - \tau)|^2)] \Delta^2 + 2C_k (1 + |z_1(t)|^2) \\
&\quad + |z_1(t - \tau)|^2 ((w(t) - w(t_k))^2). \quad (5.22)
\end{align}

By virtue of Lemma 5.2, we know that for any $t \in [-\tau, T \wedge \rho_k]$, $E|z_1(t)|^2 \leq C_2$ and $E|z_2(t)|^2 \leq C_2$, so we have:

\begin{align}
\int_0^{t_{k+1}} E|y(s) - z_1(s)|^2 ds &\leq 4C_k \Delta^2 \int_0^{t_{k+1}} (2 + E|z_1(s)|^2 + E|z_1(s - \tau)|^2 + E|z_2(s)|^2 + E|z_2(s - \tau)|^2) ds \\
&\quad + 2C_k \int_0^{t_{k+1}} (1 + E|z_1(s)|^2 + E|z_1(s - \tau)|^2) (E(w(s) - w(t_k))^2) ds \\
&\leq 4C_k \Delta^2 T (2 + 4C_2) + 2C_k (1 + 2C_2) m \Delta \leq 2C_k T (1 + 2C_2) (4 + m) \Delta \equiv C_3 \Delta, \quad (5.23)
\end{align}

where $C_3 = 2C_k T (1 + 2C_2) (4 + m)$. Analogously, we have $\int_0^{t_{k+1}} E|y(s) - z_2(s)|^2 ds \leq C_3 \Delta$. The proof is then complete. \hfill \Box

**Lemma 5.4.** Under the condition (H2),

\begin{align}
E \left[ \sup_{0 \leq t \leq T} |u(z_1(t - \tau), r(t)) - u(z_1(t - \tau), r_1(t))|^2 \right] &\leq C_4 \Delta + o(\Delta), \quad (5.24)
\end{align}

where $C_4$ is a positive constant independent of $\Delta$. 

---

Proof. Let \( [T/\Delta] \) be the integer of \( T/\Delta \), then

\[
E \left[ \sup_{0 \leq t < T} |u(z_1(t - \tau), r(t)) - u(z_1(t - \tau), r_1(t))|^2 \right] \\
\leq \max_{0 \leq k < [T/\Delta]} E \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} |u(z_1(s - \tau), r(s)) - u(z_1(s - \tau), r_1(s))|^2 \right] \\
\leq 2 \max_{0 \leq k < [T/\Delta]} E \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} |u(z_1(s - \tau), r(s)) - u(z_1(s - \tau), r_1(s))|^2 \right] \\
\leq 4 \max_{0 \leq k < [T/\Delta]} E \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} \left( |u(z_1(s - \tau), r(s))|^2 + |u(z_1(s - \tau), r_1(s))|^2 \right) \right] \\
\leq 8K \max_{0 \leq k < [T/\Delta]} \left( 1 + E \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} |z_1(s - \tau)|^2 \right] \right) E[I_{\{r(t) \neq \tau(t)\}}] \\
\leq 8K \max_{0 \leq k < [T/\Delta]} \left( 1 + E \left[ \sup_{t \in [k\Delta, (k+1)\Delta)} |z_1(t - \tau)|^2 \right] \right) E[I_{\{r(t) \neq \tau(t)\}}] \leq 8K(1 + C_2)E[I_{\{r(t) \neq \tau(t)\}}],
\]

where using the Markov property implies:

\[
E[I_{\{r(t) = \tau(t)\}}] = E[E[I_{\{r(t) = \tau(t)\}} | r(t_k)] \right] = \sum_{i \in S} I_{\{r(t_k) = i\}} P(r(t) = i | r(t_k) = i) = \sum_{i \in S} I_{\{r(t_k) = i\}} \sum_{j \neq i} \gamma_{ij}(t - t_k) + o(t - t_k) \\
\leq \left[ \max_{0 \leq i \leq N} (-\gamma_{ii}) + o(\Delta) \right] \sum_{i \in S} I_{\{r(t_k) = i\}} \leq N \left[ \max_{0 \leq i \leq N} (-\gamma_{ii}) \right] \Delta + o(\Delta),
\]

so

\[
E \left[ \sup_{0 \leq t < T} |u(z_1(t - \tau), r(t)) - u(z_1(t - \tau), r_1(t))|^2 \right] \leq 8K(1 + C_2) \left[ N \max_{0 \leq i \leq N} (-\gamma_{ii}) \Delta + o(\Delta) \right] \equiv C_4 \Delta + o(\Delta),
\]

where \( C_4 = 8NK(1 + C_2) \max_{0 \leq i \leq N} (-\gamma_{ii}). \) \( \Box \)

In the same way of proofs, we can establish the following results.

**Lemma 5.5.** Under the condition (H1), for any \( t \in [0, T \wedge \rho_B] \), there exists a positive constant \( C_5 \) independent of \( \Delta \), such that:

\[
E \int_0^{t \wedge \rho_B} \left| f(z_1(s), z_1(s - \tau), r(s) - f(z_1(s), z_1(s - \tau), r_1(s)) \right|^2 ds \leq C_5 \Delta + o(\Delta), \tag{5.25}
\]

\[
E \int_0^{t \wedge \rho_B} \left| f(z_2(s), z_2(s - \tau), r(s) - f(z_2(s), z_2(s - \tau), r_2(s)) \right|^2 ds \leq C_5 \Delta + o(\Delta), \tag{5.26}
\]

\[
E \int_0^{t \wedge \rho_B} \left| g(z_1(s), z_1(s - \tau), r(s) - g(z_1(s), z_1(s - \tau), r_1(s)) \right|^2 ds \leq C_5 \Delta + o(\Delta). \tag{5.27}
\]

**Lemma 5.6.** Under conditions (H1) and (H2), there exists a positive constant \( C_6 \) independent of \( \Delta \), such that for \( p \geq 2 \),

\[
E \left[ \sup_{-\tau \leq t \leq T} |y(t)|^p \right] \leq C_6. \tag{5.28}
\]

**Proof.** For \( t \in [\tau_k, \tau_{k+1}] \), set \( \tilde{y}(t) = y(t) - u(z_1(t - \tau), r_1(t)) \), then for any \( t \in [0, T] \), by the definition of \( y(t) \), we have:

\[
\tilde{y}(t) = \tilde{y}(0) + \int_0^t \left[ (1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s)) \right] ds + \int_0^t g(z_1(s), z_1(s - \tau), r_1(s)) dw(s)
\]

(5.29)

with the initial value \( \tilde{y}(0) = \xi(0) - u(\xi(-\tau), r_0^1). \)

Recall the inequality that for \( p \geq 1 \) and any \( \varepsilon > 0, \)

\[
|x + y|^p \leq (1 + \varepsilon)^{p-1} |x|^p + \varepsilon \left[ |y(t)|^p \right], \tag{5.30}
\]

which together with condition (H2) implies:

\[
|y(t)|^p \leq (1 + \varepsilon)^{p-1} \left( \tilde{y}(t)^p + \varepsilon \left[ |u(z_1(t - \tau), r_1(t))|^p \right] \right) \leq (1 + \varepsilon)^{p-1} \left[ |y(t)|^p + \varepsilon K^p \left[ \sup_{-\tau \leq t \leq T} |y(t)|^p \right] \right]. \tag{5.31}
\]
Noting that $0 < K < 1$, we choose $\varepsilon = \frac{1}{1-K}$, then
\[ |y(t)|^p \leq (1 - K)^{1-p} |y(t)|^p + KE^\sup_{-\tau \leq s \leq t} |y(s)|^p. \]  
(5.32)

Hence,
\[ E^\sup_{-\tau \leq s \leq t} |y(s)|^p \leq E|\xi|^p + E^\sup_{-\tau \leq s \leq t} |y(s)|^p \leq E|\xi|^p + (1 - K)^{1-p} E^\sup_{-\tau \leq s \leq t} |y(s)|^p + KE^\sup_{-\tau \leq s \leq t} |y(s)|^p. \]
(5.33)

which can yield:
\[ E^\sup_{-\tau \leq s \leq t} |y(s)|^p \leq \frac{E|\xi|^p}{1-K} + \frac{E(\sup_{-\tau \leq s \leq t}|\xi|^p)}{(1-K)^p}. \]
(5.34)

By (5.25) and the inequality $|a + b + c|^p \leq 3^{p-1}(|a|^p + |b|^p + |c|^p)$, for any $t_1 \in [0, T],
\[ E^\sup_{0 \leq s \leq t_1} |y(s)|^p \leq 3^{p-1}E|y(0)|^p + 3^{p-1}E \int_0^{t_1} (1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s)) ds \]
\[ + 3^{p-1}E^\sup_{0 \leq s \leq t_1} \left| \int_0^s g(z_1(s), z_1(s - \tau), r_1(s)) dw(s) \right|^p. \]
(5.35)

Using condition (H2) gets:
\[ E|y(0)|^p = E|\xi(0) - u(\xi(-\tau), r_0)|^p \leq E(|\xi(0)| + K|\xi(-\tau)|)^p \leq (1 + K)^p E|\xi|^p. \]
(5.36)

By the Hölder inequality, the Fubini's theorem, the inequality $|a + b| \leq 2^{p-1}(|a| + |b|)$, condition (5.1) and $0 \leq \theta \leq 1$, we have
\[ E \int_0^{t_1} (1 - \theta)f(z_1(s), z_1(s - \tau), r_1(s)) + \theta f(z_2(s), z_2(s - \tau), r_2(s)) ds \]
\[ \leq (2t_1)^{p-1} E \int_0^{t_1} \left[ f(z_1(s), z_1(s - \tau), r_1(s)) + f(z_2(s), z_2(s - \tau), r_2(s)) \right]^p ds \]
\[ \leq (2t_1)^{p-1} C_K^p \left[ \int_0^{t_1} (1 + E|z_1(s)|^2 + E|z_1(s - \tau)|^2)^\frac{p}{2} ds + \int_0^{t_1} (1 + E|z_2(s)|^2 + E|z_2(s - \tau)|^2)^\frac{p}{2} ds \right] \]
\[ \leq (2t_1)^{p-1} C_K^p \left[ \int_0^{t_1} (1 + 2E \sup_{-\tau \leq s \leq t} |z_1(v)|^2)^\frac{p}{2} ds + \int_0^{t_1} (1 + 2E \sup_{-\tau \leq s \leq t} |z_2(v)|^2)^\frac{p}{2} ds \right] \]
\[ \leq (2T)^{p-1} C_K^p \left[ 2T + \int_0^{t_1} E \sup_{-\tau \leq s \leq t} |z_1(s)|^p ds + \int_0^{t_1} E \sup_{-\tau \leq s \leq t} |z_2(s)|^p ds \right]. \]
(5.37)

Using the Burkholder–Davis–Gundy inequality and condition (5.1) implies:
\[ E^\sup_{0 \leq s \leq t_1} \left| \int_0^s g(z_1(s), z_1(s - \tau), r_1(s)) dw(s) \right|^p \leq C_p^p \left[ \int_0^{t_1} E|g(z_1(s), z_1(s - \tau), r_1(s))^2 ds \right] \]
\[ \leq 2^{p-1} C_p^p C_K^p \left[ T + \int_0^{t_1} E \sup_{-\tau \leq s \leq t} |z_1(v)|^p ds \right], \]
(5.38)

where $C_p^p = \left( \frac{p^{p-1}}{2^{p-1}-1} \right)^\frac{1}{p}$.

Substituting inequalities (5.35)–(5.38) into (5.34) yields:
\[ E^\sup_{-\tau \leq s \leq t} |y(t)|^p \leq \frac{E|\xi|^p}{1-K} + \frac{1}{(1-K)^p} \left[ 3^{p-1} (1 + K)^p E|\xi|^p \right. \]
\[ + (12T)^{p-1} C_K^p \left[ 2T + \int_0^{T} E \sup_{-\tau \leq s \leq t} |z_1(v)|^p ds + \int_0^{T} E \sup_{-\tau \leq s \leq t} |z_2(v)|^p ds \right] \]
\[ + 6^{p-1} C_p^p C_K^p \left[ T + \int_0^{T} E \sup_{-\tau \leq s \leq t} |z_1(v)|^p ds \right] \]
\[ \left. + \frac{1}{(1-K)^p} \left[ 3^{p-1} (1 + K)^p E|\xi|^p + \frac{\tilde{c}_p T}{(1-K)^p} + \frac{\tilde{c}_p}{(1-K)^p} \int_0^{T} E \sup_{-\tau \leq s \leq t} |y(v)|^p ds \right] \right] \]
\[ \leq \frac{1}{(1-K)^p} \left[ 2^{p-1} (1 + K)^p E|\xi|^p + \frac{\tilde{c}_p T}{(1-K)^p} + \frac{\tilde{c}_p}{(1-K)^p} \int_0^{T} E \sup_{-\tau \leq s \leq t} |y(v)|^p ds \right], \]
where $\tilde{c}_p = 6^{p-1} C_K^p \left[ 2^{p-1} T^{1-p} + C_p^p \right] > 0$. 

Finally, applying the Gronwall inequality completes the proof as follows:

$$E \left[ \sup_{-\infty < t < T} |y(t)|^p \right] \leq \left\{ \frac{1}{1 - K} + \frac{4^{p-1}(1 + K)^p}{(1 - K)^p} \right\} E \| \zeta \|^p + \frac{\bar{c}_{p}T}{(1 - K)^p} e^{\bar{c}_{p}T} \equiv C_{6},$$

where $C_{6}$ is a positive constant and independent of $\Lambda$. \qed

6. **Proof of Theorem 4.2**

Let $e(t) \equiv x(t) - y(t)$, then it is evident that:

$$E \left[ \sup_{0 < t < T} |e(t)|^2 \right] = E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k > T\}} + \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right]$$

$$= E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k > T\}} \right] + E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right]$$

$$\leq E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right] + E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right].$$

(6.1)

To obtain the bounds for different parts, we will divide the following proof into three steps:

**Step 1:** Bound for $E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right]$.

For any positive constants: $a, b, \delta, p, q$ and $\frac{1}{p} + \frac{1}{q} = 1$, the Young inequality implies:

$$ab = a^\delta b^{\frac{1}{\delta}} \leq \frac{a^p}{p} + \frac{b^q}{q^\delta}. \quad (6.2)$$

Hence, for any $\delta > 0$,

$$E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right] \leq \frac{2\delta}{p} E \left[ \sup_{0 < t < T} |e(t)|^2 \right] + \frac{1}{\delta} p \left\{ I_{\{\sigma_k < T \text{ or } \gamma_k < T\}} \right\}. \quad (6.3)$$

By Lemma 5.5,

$$P\{\sigma_k < T\} = E \left[ I_{\{\sigma_k < T\}} \frac{|y(\sigma)|^p}{R^p} \right] \leq \frac{1}{R^p} E \left[ \sup_{0 < t < T} |y(t)|^p \right] \leq C_6 R^p. \quad (6.4)$$

Similarly, by Lemma 5.1,

$$P\{\sigma_k < T\} \leq C_1 R^p. \quad (6.5)$$

Then we have

$$P\{\sigma_k < T \text{ or } \gamma_k < T\} \leq P\{\sigma_k < T\} + P\{\gamma_k < T\} \leq \frac{C_1 + C_6}{R^p}. \quad (6.6)$$

So using the inequality $|a + b|^p \leq 2^{p-1}(\|a\|^p + \|b\|^p)$ along with the above bounds leads to:

$$E \left[ \sup_{0 < t < T} |e(t)|^2 I_{\{e_k < T \text{ or } \gamma_k < T\}} \right] \leq \frac{2\delta}{p} E \left[ \sup_{0 < t < T} |x(t)|^p \right] + \frac{1}{\delta} p \left\{ I_{\{\sigma_k < T \text{ or } \gamma_k < T\}} \right\}$$

$$\leq \frac{2\delta(C_1 + C_6)}{p} + \frac{(p - 2)(C_1 + C_6)}{\delta}. \quad (6.7)$$

**Step 2:** Bound for $E \left[ \sup_{0 < t < T} |e(t \wedge \rho_k)|^2 \right]$.

By the definitions of $x(t), y(t)$ and $e(t)$, we have

$$e(t \wedge \rho_k) = u(x(t \wedge \rho_k - \tau), r(t \wedge \rho_k)) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))$$

$$+ \int_{0}^{t \wedge \rho_k} [(1 - \theta)(f(x(s), x(s - \tau), r(s)) - f(z_1(s), z_1(s - \tau), r_1(s))) + \theta f(x(s), x(s - \tau), r(s)) - f(z_2(s), z_2(s - \tau), r_2(s))] ds$$

$$+ \int_{0}^{t \wedge \rho_k} [g(x(s), x(s - \tau), r(s)) - g(z_1(s), z_1(s - \tau), r_1(s))] ds.$$

Recalling the inequality that $|a + b + c|^2 \leq \frac{a^2}{2} + \frac{b^2}{2} + \frac{c^2}{2}$ for any $\varepsilon \in (0, 1)$, so for any $0 < t_1 \leq T$, noting that $0 < \theta \leq 1$, we have:
By the Hölder inequality, condition (H2), Lemmas 5.3 and 5.5, we know:

\[
E \left[ \sup_{0 \leq r \leq t_1} |e(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))|^2 \right] \\
\leq \frac{1}{\varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |u(x(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))|^2 \right]
\]

\[
+ \frac{2}{1 - \varepsilon} E \sup_{0 \leq r \leq t_1} \int_0^{t_1/\rho_k} \left| (1 - \theta)f(x(s), x(s - \tau), r(s) - f(z_1(s), z_1(s - \tau), r_1(s))) \right|^2 \, ds
\]

\[
+ \theta |f(x(s), x(s - \tau), r(s)) - f(z_2(s), z_2(s - \tau), r_2(s))| \, ds
\]

\[
+ \frac{2}{1 - \varepsilon} E \sup_{0 \leq r \leq t_1} \int_0^{t_1/\rho_k} \left| g(x(s), x(s - \tau), r(s) - g(z_1(s), z_1(s - \tau), r_1(s))) \right|^2 \, ds
\]

\[
+ \frac{1}{\varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |u(x(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))|^2 \right]
\]

\[
+ \frac{4}{1 - \varepsilon} E \sup_{0 \leq r \leq t_1} \int_0^{t_1/\rho_k} \left| f(x(s), x(s - \tau), r(s)) - f(z_1(s), z_1(s - \tau), r_1(s)) \right|^2 \, ds
\]

\[
+ \frac{4}{1 - \varepsilon} E \sup_{0 \leq r \leq t_1} \int_0^{t_1/\rho_k} \left| f(x(s), x(s - \tau), r(s)) - f(z_2(s), z_2(s - \tau), r_2(s)) \right|^2 \, ds
\]

\[
+ \frac{2}{1 - \varepsilon} E \sup_{0 \leq r \leq t_1} \int_0^{t_1/\rho_k} \left| g(x(s), x(s - \tau), r(s)) - g(z_1(s), z_1(s - \tau), r_1(s)) \right| \, ds
\]

By the proof process of Lemma 5.3, we know:

\[
E \sup_{0 \leq r \leq t_1} |y(t \wedge \rho_k) - z_1(t \wedge \rho_k)|^2 \leq 2C_\rho(1 + 2C_2)(4\Delta^2 + m\Delta).
\]

(6.7)

\[
E \sup_{0 \leq r \leq t_1} |y(t \wedge \rho_k) - z_2(t \wedge \rho_k)|^2 \leq 2C_\rho(1 + 2C_2)(4\Delta^2 + m\Delta).
\]

(6.8)

then for any \( \varepsilon \in (0, 1) \), by the condition (H2) and Lemma 5.4,

\[
E \left[ \sup_{0 \leq r \leq t_1} |u(x(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))|^2 \right]
\]

\[
\leq \frac{1}{\varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |u(x(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r(t \wedge \rho_k))|^2 \right]
\]

\[
+ \frac{1}{1 - \varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |u(z_1(t \wedge \rho_k - \tau), r(t \wedge \rho_k) - u(z_1(t \wedge \rho_k - \tau), r_1(t \wedge \rho_k))|^2 \right]
\]

\[
\leq \frac{K^2}{\varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |x(t \wedge \rho_k - \tau) - z_1(t \wedge \rho_k - \tau)|^2 \right] + \frac{C_4}{1 - \varepsilon} \Delta + o(\Delta)
\]

\[
\leq \frac{K^2}{\varepsilon} E \left[ \sup_{0 \leq r \leq t_1} |x(t \wedge \rho_k - \tau) - y(t \wedge \rho_k - \tau)|^2 \right] + \frac{K^2}{\varepsilon^2} E \left[ \sup_{0 \leq r \leq t_1} |y(t \wedge \rho_k - \tau) - z_1(t \wedge \rho_k - \tau)|^2 \right] + \frac{C_4}{1 - \varepsilon} \Delta + o(\Delta)
\]

\[
\leq \frac{K^2}{\varepsilon^2} E \left[ \sup_{0 \leq r \leq t_1} |x(t \wedge \rho_k) - y(t \wedge \rho_k)|^2 \right] + \frac{K^2}{\varepsilon^2} E \left[ \sup_{0 \leq r \leq t_1} |y(t \wedge \rho_k) - z_1(t \wedge \rho_k)|^2 \right] + \frac{2C_\rho(1 + 2C_2)(4\Delta^2 + m\Delta)}{1 - \varepsilon} + \frac{C_4}{1 - \varepsilon} \Delta + o(\Delta)
\]

(6.9)

By the Hölder inequality, condition (H2), Lemmas 5.3 and 5.5.

\[
E \sup_{0 \leq t \leq T} \left| \int_{t_\tau}^{t_{\tau'}} g(x(s), x(s - \tau), r(s) - g(z_1(s), z_1(s - \tau), r_1(s)))ds \right|^2 \\
\leq 4 T C_6 \int_{t_\tau}^{t_{\tau'}} \sup_{0 \leq s \leq \tau} |g(x(s), x(s - \tau), r(s) - g(z_1(s), z_1(s - \tau), r_1(s)))|^2 ds \\
\leq 32 C_6 T \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 ds + 8(4C_6C_3 + C_5)\Delta + o(\Delta).
\]

Consequently, substituting all the bounds (6.10), (6.11), (6.12), (6.13) into (6.9) yields
\[
E\left[ \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 \right] \leq \frac{1}{\varepsilon} \left[ \frac{K^2}{\varepsilon^3} E\left[ \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 \right] + \frac{2K^2mC_6(1 + 2C_2) + eC_4}{\varepsilon(1 - \varepsilon)} \Delta + o(\Delta) \right] \\
+ \frac{8}{1 - \varepsilon} \left[ 8T C_6 \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 ds + 2T(4C_6C_3 + C_5)\Delta + o(\Delta) \right] \\
+ \frac{2}{1 - \varepsilon} \left[ 32C_6 T \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 ds + 8(4C_6C_3 + C_5)\Delta + o(\Delta) \right] \\
= \frac{K^2}{\varepsilon^3} E\left[ \sup_{0 \leq t \leq T} |e(t + \rho_\omega)|^2 \right] + \frac{64C_6(1 + T)}{\varepsilon^2(1 - \varepsilon)} \int_{0}^{T} \left[ \sup_{0 \leq s \leq \tau} |e(s + \rho_\omega)|^2 \right] ds \\
+ \frac{2K^2mC_6(1 + 2C_2) + eC_4 + 16(4C_6C_3 + C_5)(1 + T)}{\varepsilon^2(1 - \varepsilon)} \Delta + o(\Delta).
\]
Applying the Gronwall inequality leads to:

\[
\mathbb{E}\left[ \sup_{0 \leq t \leq T} |e(t \land \rho_k)|^2 \right] \leq C_8 e^{\frac{c_2}{p} \delta^2} \Delta + o(\Delta),
\]

where \( C_7 = 64 C_6 (1 + T) \) and \( C_8 = \frac{2 \delta K^2 m C_4 (1 + 2 C_2 + \epsilon^2 K^2 C_4 + 16 (4 C_8 C_3 + C_5) (1 + T))}{(1 - \epsilon^2 K^2 - K^2)} \) are both positive constants independent of \( \Delta \).

Step 3: Bound for \( \mathbb{E}\left[ \sup_{0 \leq t \leq T} |e(t)|^2 \right] \).

Connecting (6.6) and (6.15) with (6.1), we have:

\[
\mathbb{E}\left[ \sup_{0 \leq t \leq T} |e(t)|^2 \right] \leq C_8 e^{\frac{c_2}{p} \delta^2} \Delta + \frac{2 \delta (C_1 + C_6)}{p} + \frac{(p - 2)(C_1 + C_6)}{p \delta R^2} + o(\Delta).
\]

Given any \( \epsilon > 0 \), we can choose \( \delta \) sufficiently small for

\[
\frac{2 \delta (C_1 + C_6)}{p} < \frac{\epsilon}{3},
\]

then choose \( R \) sufficiently large for

\[
\frac{(p - 2)(C_1 + C_6)}{p \delta R^2} < \frac{\epsilon}{3},
\]

and then choose \( \Delta \) sufficiently small for

\[
C_8 e^{\frac{c_2}{p} \delta^2} \Delta + o(\Delta) < \frac{\epsilon}{3},
\]

so that \( \mathbb{E}\left[ \sup_{0 \leq t \leq T} |e(t)|^2 \right] < \epsilon \) as required. The proof is therefore complete. \( \square \)

**Remark 6.1.** Let us make some concluding remarks. It should be pointed out that although our main theorems focus on the convergence in the mean square sense or in \( L^2 \), they can easily be developed to establish the strong convergence in \( L^p \) for any \( p > 2 \).

**Acknowledgements**

This Research was supported by the NNSF of China under Grant No. 10871077 and the fundamental research funds for the central universities under Grant No. 2010MS130.

**References**


[22] A. Jensen, A Distribution Model Applicable to Economics, Munkgaard, Copenhagen, Denmark, 1954.
[23] M.F. Neuts, Probability Distribution of Phase Type, Belgium University of Louvain, Belgium, 1975.