MULTISCALE MODEL. SIMUL.
Vol. 15, No. 4, pp. 1797–1829

ERROR ANALYSIS FOR D-LEAPING SCHEME OF CHEMICAL REACTION SYSTEM WITH DELAY*

CHUCHU CHEN† AND DI LIU‡

Abstract. We perform an error analysis in both strong and weak senses for D-leaping scheme of chemical reactions with delays within the framework of stochastic delay differential equations (SDDEs). In order to establish the convergence orders, we prove an infinite dimensional Itô formula for "tame" functionals acting on the segment process of the solution of SDDEs. It is shown that the mean-square strong convergence is of order 1/2 and the weak convergence is of order 1 for the scheme. Moreover, we propose highly accurate schemes by adding random corrections to the primitive D-leaping scheme in each step. Numerical experiments are provided to illustrate the results.

Key words. stochastic delay differential equation, Poisson random measure, D-leaping, mean-square strong convergence order, weak convergence order, Malliavin calculus

AMS subject classifications. 60H07, 65C20, 65C30

DOI. 10.1137/17M1118464

1. Introduction.

Delay plays a significant role in many chemical dynamics. For example, in genetic regulatory networks, processes such as transcription and translation do not occur instantaneously, and these delays may produce oscillations in the networks [3, 14]. In addition, delayed negative feedback is theorized to govern the dynamics of circadian oscillators [18]. Increasing delay dramatically prolongs the mean residence times near stable states for bistable gene networks, which means that delay stabilizes bistable gene networks [9]. In chemical reactions, noise and delay may interact in subtle and complex ways. For example, in genetic regulatory networks, delay can affect the stochastic properties of gene expression and hence the phenotype of the cell [5]. For bistable gene networks, due to the stability enhanced by the infusion of delay, it may induce an analogue of stochastic resonance [9].

In order to take proper account of these aspects, mathematical modeling, analysis, and simulation of the delayed chemical reactions are necessary. For example, a delay stochastic simulation algorithm (DSSA) was proposed in [5], and three other DSSA-type algorithms were proposed in [3]. Their implementations differ in the ways they handled the waiting time for delayed reactions, as well as in the time steps in the presence of delayed reaction updates and delayed consuming reactions. These algorithms are direct generations of Gillespie's stochastic simulation algorithms (SSAs) to deal with delays. More recently, [7] introduced an exact SSA for chemical reaction systems with delays, which was based on the fundamental premise of stochastic chemical kinetics. Utilizing the fact that the initiation times of the reactions can be represented as the firing times of independent unit rate Poisson processes with internal times given by integrated propensity functions, Anderson [2] derived a modified

*Received by the editors February 27, 2017; accepted for publication (in revised form) September 12, 2017; published electronically December 19, 2017.
http://www.siam.org/journals/mms/15-4/M111846.html
Funding: This work was supported by NSF-DMS 1418959, NSF-DMS 1211292, and a special grant from Michigan Center at Industrial and Applied Mathematics (MCIAM).
†Department of Mathematics, Michigan State University, East Lansing, MI 48824 (chenc118@msu.edu). Current address. LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (chenchuchu@lsec.cc.ac.cn).
‡Department of Mathematics, Michigan State University, East Lansing, MI 48824 (richardl@math.msu.edu).

next reaction method for exactly simulating chemical reaction systems with time- dependent propensities and delays. Thanh, Priami, and Zunino [20] proposed another exact simulation algorithm called rejection-based SSA.

Although SSAs are able to produce the exact time evolution of a chemical reaction system, a great amount of computing time is often required to simulate a desired amount of system time. To this end, an accelerated, approximate algorithm, similar to the τ-leaping method that can produce significant gains in simulation speed with acceptable losses in accuracy, is needed. Bayati, Chatelain, and Koumoutsakos [4] proposed an accelerated algorithm called the D-leaping scheme for the approximate simulation of biochemical systems with delays. Leier, Marques-Lago, and Burrage [12] proposed a generalized binomial τ-leap method to overcome the limit of small step sizes in SSAs.

In this paper, we aim to provide an error analysis for the scheme to approximate the chemical reactions with delays. Mathematically, the chemical reaction process is a pure jump process on a lattice with state-dependent intensity. We may formulate a system of stochastic delay differential equations (SDDEs) via Poisson random measures for jump processes, similarly to [13]. Then we find that the D-leaping scheme is just an explicit Euler-type scheme for this SDDE. Utilizing the Itô formula and the Itô identity for a stochastic integral with Poisson random measure, we prove that

$$\mathbb{E}[|\boldsymbol{X}(t_n) - \boldsymbol{Y}(t_n)|^2] \leq C\delta t \quad \forall\, n = 1, 2, \ldots, N,$$

which means that the mean-square strong convergence for the scheme is of order 1/2. Here $\boldsymbol{X}(t)$ is the exact solution process of the chemical system, $\boldsymbol{Y}(t)$ is the approximated solution generated from the D-leaping scheme, and δt is the maximal time stepsize δt_n, $n = 1, 2, \ldots, N$, with $\sum_{n=1}^{N} \delta t_n = T$. Note that the constant C may depend on the coefficients of the system and the final time.

Using the Markov property of the segment process, we rewrite the expression of weak error as the summation of weak local error. The mathematical analysis of the local error term is technique in two aspects. First, since delays break the Markovian property of the system, by contrast with the nondelay case (stochastic differential equations or SDEs), SDDEs do not correspond to diffusions on Euclidean space. Thus techniques from deterministic PDEs do not apply. Second, techniques used in [6] to derive the weak convergence order of Euler scheme for SDDEs driven by Brownian motions utilize the Fréchet differentiability of the Euler approximation $\boldsymbol{Y}(t_n; t_i, \eta)$ with respect to the initial data η and mean value theorem to show that the local error term is of order $\mathcal{O}(\delta t^2)$. However, since the coefficients in the SDDEs of chemical reactions are not differentiable, the above approach also is not applicable. In order to derive the weak convergence of the scheme, we first establish an Itô formula for tame functionals of segments of the solution process of the SDDEs driven by Poisson random measure. By inserting the functional of the previous step into the weak local error term, we separate the local error term into two parts, and then apply the established tame Itô formula. Moreover, the Malliavin calculus and anticipating stochastic analysis techniques are employed to show that

$$|\mathbb{E}\phi(\boldsymbol{X}(t_n)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n))| \leq C\delta t,$$

which means that the weak convergence of the D-leaping scheme is of order 1.

The construction of high weak order schemes for stochastic systems is a fundamentally interesting topic; see [13, 17] for the case of stochastic differential equations (SDEs) driven by Poisson random measure. In this paper, we also investigate the construction of high weak order schemes for SDDEs driven by Poisson random measure. Due to the difficulties caused by the coupling of delays and noises, we fix the test function $\phi(x)$ first, and then apply the tame Itô formula to obtain the correction term. By adding this random correction term to the primitive D-leaping scheme in every step, we can improve the accuracy of the D-leaping scheme for arbitrary order of moments of the solution.

Finally, we define the following notation in order to describe our setup. $\mathbb{Z}_0^+ = \mathbb{N} \cup \{0\}$ denotes the set of nonnegative integers. Mathematically, a well-stirred chemical reaction system can be accurately described by a discrete state continuous time jump process on the lattice $(\mathbb{Z}_0^+)^N$. Let $\mathbb{R}^n$ be n-dimensional Euclidean space with Euclidean norm $|x| := \sqrt{x_1^2 + \cdots + x_n^2}$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and the inner product in $\mathbb{R}^n$ is denoted by $x \cdot y$, where $x, y \in \mathbb{R}^n$, so that $x \cdot y = \sum_{i=1}^n x_i y_i$. $L([-\tau, 0], \mathbb{R}^n)$ represents the space of all càdlàg paths $[-\tau, 0] \to \mathbb{R}^n$, given the supremum norm $\|\eta\|_\infty = \sup_{-\tau \le s \le 0} |\eta(s)|$ for all $\eta \in L([-\tau, 0], \mathbb{R}^n)$.

The rest of this paper is organized as follows. In section 2, the background of the D-leaping scheme and the SDE formulation of the reacting system is introduced. In section 3, we first give the strong convergence proof for the scheme. Then in order to present the weak convergence, we establish the tame Itô formula and the Malliavin calculus analysis for the solution process. Numerical experiments are performed to support our theoretical results. In section 4, the generation to highly accurate schemes is presented. Conclusions are made in section 5.

2. SDDEs and numerical approximation.

Consider a well-stirred system of N molecular species $\{S_1, S_2, \ldots, S_N\}$ interacting through M chemical reaction channels $\{R_1, R_2, \ldots, R_M\}$. Suppose that some channels involve delays. We let the set I_{nd} consist of all the channels without delay and let the set I_d consist of all the channels with delays, and τ_d is the delay for channel $R_d \in I_d$, i.e., $\{R_1, R_2, \ldots, R_M\} = I_{nd} \cup I_d$. The state of the system is described by the vector

$$\boldsymbol{X}(t) = \Big(X^1(t), X^2(t), \ldots, X^N(t)\Big).$$

Each reaction channel R_j is characterized by its propensity function $a_j(\boldsymbol{x})$ and its state change vector

$$\boldsymbol{\nu}_j = \left(\nu_j^1, \nu_j^2, \ldots, \nu_j^N\right),$$

where $a_j(\boldsymbol{x}) \ge 0$ for physical states. Here $a_j(\boldsymbol{x})\mathrm{d}t$ gives the probability that the system will experience an R_j reaction in the next infinitesimal time $\mathrm{d}t$ when the current state $\boldsymbol{X}(t) = \boldsymbol{x}$. ν_j^i is the change in the number of S_i molecules caused by one R_j reaction.

2.1. Basic model.

The exact DSSA algorithm proposed in [3] is described as follows:

(1) Initialization. Set $t \leftarrow 0$ and the initial number of molecules $\boldsymbol{X}(t) = \boldsymbol{x}$.

(2) Calculate propensity functions $a_m(\boldsymbol{x})$, $m = 1, \ldots, M$. Generate $\tau\prime$ from a standard uniform random variable u_2 as $\tau\prime = -\ln(u_2)/a_0(\boldsymbol{x})$ with $a_0(\boldsymbol{x}) = \sum_{j=1}^M a_j(\boldsymbol{x})$. If there are delayed reaction(s) when finishing in the time interval $[t, t + \tau\prime)$, discard $\tau\prime$, update time $t \leftarrow t_d$, where t_d is the time when the first delayed reaction finishes, update the state vector $\boldsymbol{x}$, and repeat step (2). If there is no delayed reaction when finishing in $[t, t + \tau\prime)$, proceed to step (3).

(3) Generate μ from a standard uniform random variable u_1 by taking μ to be the integer for which $\sum_{j=1}^{\mu-1} a_j(\boldsymbol{x}) < u_1 a_0(\boldsymbol{x}) \le \sum_{j=1}^{\mu} a_j(\boldsymbol{x})$. If $\mu \in I_{nd}$, update the

state vector as $\boldsymbol{X}(t+\tau\prime) = \boldsymbol{X}(t) + \boldsymbol{\nu}_\mu$.

Note that $\boldsymbol{X}(t)$ is actually a compound Poisson process with state-dependent intensity. Given any initial state $\boldsymbol{X}_0 \in (\mathbb{Z}_0^+)^N$, as in [13] the space of the possible physical states generated from $\boldsymbol{X}_0$ is denoted as $\Omega_{\boldsymbol{X}_0}$, which is defined by

$$\Omega_{\boldsymbol{X}_0} = \left\{ \boldsymbol{X} \mid \boldsymbol{X} \in (\mathbb{Z}_0^+)^N,\ \boldsymbol{X} = \boldsymbol{X}_0 + \sum_{j=1}^{M} k_j \boldsymbol{\nu}_j, \quad k_j \in \mathbb{Z}_0^+ \right\},$$

and the space of the possible states generated from $\boldsymbol{X}_0$ is denoted as $\Omega^t_{\boldsymbol{X}_0}$, which may be negative and defined by

$$\Omega^t_{\boldsymbol{X}_0} = \left\{ \boldsymbol{X} \mid \boldsymbol{X} \in \mathbb{Z}^N,\ \boldsymbol{X} = \boldsymbol{X}_0 + \sum_{j=1}^{M} k_j \boldsymbol{\nu}_j, \quad k_j \in \mathbb{Z}_0^+ \right\}.$$

From [13], we notice that the state process $\boldsymbol{X}(t)$ generated by the DSSA algorithm may be formulated as the form of an SDE with delay, also called a stochastic delay differential equation (SDDE). We refer the reader to [8] for approximating the stochastic delay birth-death processes by a SDDE driven by Brownian motion, and to [11] for numerical analysis of SDDEs driven by Brownian motion. In order to unify the equation, we assign the delay $\tau_j = 0$ to a nondelayed channel $R_j \in I_{nd}$. Therefore $\boldsymbol{X}(t)$ is the solution of the following SDDE:

$$\boldsymbol{X}(t) = \begin{cases} \eta(0) + \sum_{j=1}^{M} \int_0^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(s-\tau_j-))\lambda(\mathrm{d}s \times \mathrm{d}a), & t > 0, \\ \eta(t), \quad -\tau \le t \le 0,\ \tau = \max\{\tau_j,\ j \in I_d\}. \end{cases} \tag{1}$$

Here $\lambda(\mathrm{d}t \times \mathrm{d}a)$ is a Poisson random measure with Lebesgue intensity measure $m(\mathrm{d}t \times \mathrm{d}a) = \mathrm{d}t \times \mathrm{d}a$ on the probability space $(\Omega, \mathcal{F}, P)$, and we let $\{\mathcal{F}_t\}_{t\ge 0}$ be the filtration generated by the values of the compensated Poisson random measure $(\lambda - m)(\mathrm{d}t \times \mathrm{d}a)$. The number A denotes the upper bound of total propensity

$$A = \max\{a_0(\boldsymbol{x}),\ \boldsymbol{x} \in \Omega_{\boldsymbol{X}_0}\}.$$

The function $c_j(a;\ \boldsymbol{X}(s-\tau_j-))$ is defined by

$$c_j(a;\ \boldsymbol{X}(s-\tau_j-)) = \begin{cases} 1 & \text{if } a \in (h_{j-1}(\boldsymbol{X}(s-)),\ h_j(\boldsymbol{X}(s-))], \\ 0 & \text{otherwise,} \end{cases}$$

with $h_0 = 0$ and $h_j(\boldsymbol{X}(s-)) = h_{j-1}(\boldsymbol{X}(s-)) + a_j(\boldsymbol{X}(s-\tau_j-))$. Thus intervals $(h_{j-1}(\boldsymbol{X}(s-)),\ h_j(\boldsymbol{X}(s-))]$, $j = 1, 2, \ldots, M$, are disjoint, and the length of the jth interval is $a_j(\boldsymbol{X}(s-\tau_j-))$. We refer the reader to [1] for stochastic integrals with respect to Lévy processes.

We make the following assumptions; see [13] for further information.

ASSUMPTION 2.1 (condition on propensity functions). *The propensity function* $a_j(\boldsymbol{x}) \ge 0$ *for all* $\boldsymbol{x} \in \Omega_{\boldsymbol{X}_0}$*, and* $a_j(\boldsymbol{x}) = 0$ *if* $\boldsymbol{x} \in \Omega_{\boldsymbol{X}_0}$*, but* $\boldsymbol{x} + \boldsymbol{\nu}_j \notin \Omega_{\boldsymbol{X}_0}$.

This assumption is natural. Otherwise the negative states will appear in the physical process.

ASSUMPTION 2.2 (bound on $\boldsymbol{X}(t)$). *The number of elements in* $\Omega_{\boldsymbol{X}_0}$ *is finite; i.e.,* $\boldsymbol{X}(t)$ *is in a bounded lattice.*

This assumption is reasonable because the number of the molecules could not be arbitrarily large in realistic chemical reactions.

In order to perform the analysis, we make the following assumption on $a_j(\boldsymbol{x})$.

ASSUMPTION 2.3 (local Lipschitz condition on $a_j(\boldsymbol{x})$). *The function $a_j(\boldsymbol{x})$ is Lipschitz continuous in a bounded domain. That is, $|a_j(\boldsymbol{x}) - a_j(\boldsymbol{y})| \leq L_j|\boldsymbol{x} - \boldsymbol{y}|$ for any bounded $\boldsymbol{x}$ and $\boldsymbol{y}$, where L_j is a fixed positive real number.*

PROPOSITION 2.4 (redefinition of $a_j(\boldsymbol{x})$). *We define the modification of $a_j(\boldsymbol{x})$ as*

$$\tilde{a}_j(\boldsymbol{x}) = \begin{cases} a_j(\boldsymbol{x}), & \boldsymbol{x} \in (\mathbb{Z}_0^+)^N, \\ 0, & \boldsymbol{x} \in \mathbb{Z}^N/(\mathbb{Z}_0^+)^N. \end{cases}$$

We have

$$|\tilde{a}_j(\boldsymbol{x}) - \tilde{a}_j(\boldsymbol{y})| \leq L_j|\boldsymbol{x} - \boldsymbol{y}| \quad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \Omega^t_{\mathbf{X}_0} \cup \Big(\mathbb{Z}^N/(\mathbb{Z}_0^+)^N\Big).$$

For simplicity we will continue to denote $\tilde{a}_j(\boldsymbol{x})$ as $a_j(\boldsymbol{x})$ in the text.

2.2. Numerical method. As is well known, the DSSA algorithm is exact, but it costs a great amount of time to simulate the system. Therefore, the cost-efficient approximated numerical method should be proposed. When there is no delay involved in the system, the classical approximated method is called the τ-leaping method. The method is established by increasing the leaping time stepsize to allow the fires of a proper number of actions. For the system involving delays, a corresponding scheme (the D-leaping scheme) was proposed in [4]. The pseudocode of the D-leaping scheme reads as follows.

Algorithm 1 D-leaping algorithm.

```
while t < t_final do
    τ′ ~ ξ(Θ)
    X(t + τ′) = X(t)
    for all d such that q_{d,α} ∈ [t, t + τ′]  do
        k̂_d ~ B(k_d, min(t+τ_d−q_{d,α}, span_d) / span_d)
        span_d = span_d − (t + τ′ − q_{d,α})
        k_d = k_d − k̂_d
        q_{d,α} = t + τ′
        X(t + τ′) = X(t + τ′) + Σ_d k̂_d ν_d
        if k_d == 0 then
            Queue.remove([R_d, q_{d,α}, k_d, span_d])
        end if
    end for
    k_{j∪d} ~ Ψ(Θ, τ′)
    for all d such that k_d ≠ 0 do
        Queue.insert([R_d, q_{d,α} = t + τ_d, k_d, span_d = τ′])
    end for
    X(t + τ′) = X(t + τ′) + Σ_j k_j ν_j
    t = t + τ′
end while
```

If we suppose that the above D-leaping scheme is posed in a time interval $[0, T]$ with N_T steps,

$$0 = t_0 < t_1 < \cdots < t_{N_T} = T,$$

then the above algorithm could also be written as

$$\boldsymbol{Z}(t_{n+1}) = \boldsymbol{Z}(t_n) + \sum_{j=1}^{M} k_j \boldsymbol{\nu}_j, \tag{2}$$

where for $j = 1, 2, \ldots, M$, $k_j \sim \mathcal{P}\big(\int_{t_n}^{t_{n+1}} a_j(\boldsymbol{Z} \circ \xi(t - \tau_j))\mathrm{d}t\big)$ with $\xi(t) = t_n$ if $t \in [t_n,\ t_{n+1})$. Note that in the case of $\tau_j = 0$, we have $k_j \in \mathcal{P}\big(a_j(Z(t_n))\delta t_n\big)$ with $\delta t_n := t_{n+1} - t_n$.

Note that the random numbers in (2) are generated from Poisson distribution, while some random numbers in Algorithm 1 are generated from binomial distribution. Because the partial number of executions can be determined by considering a partitioning of the time domain, both binomial distribution and Poisson distribution are suitable to model the number of executions for each molecular species. We know that the processes generated from Algorithm 1 and (2) are equivalent to each other in a distribution sense. In fact, on the one hand, there holds the property of sums of Poisson-distributed random variables; on the other hand, we let the number k be generated from Poisson distribution with parameter $\lambda > 0$, i.e.,

$$k \sim \mathcal{P}(\lambda).$$

If λ is partitioned into two parts denoted by $\lambda_1 > 0$ and $\lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = \lambda$, it follows that the number k_j $(j = 1, 2)$ in each part is

$$k_j \sim \mathcal{P}(\lambda_j) \text{ for } j = 1, 2 \text{ such that } k_1 + k_2 = k.$$

Because of the dependence of λ_1 and λ_2, i.e., for the fixed λ, $\lambda_1 + \lambda_2 = \lambda$, we could express the number in each part by binomial distribution as

$$k_1 \sim \mathcal{B}\left(k, \frac{\lambda_1}{\lambda}\right),$$

where $\mathcal{B}(N, p)$ represents a binomial distribution of N trials with probability p. The condition distribution given k_1 reduces to the number $k_2 = k - k_1$.

Since

$$\begin{aligned}
&\int_{t_n}^{t_{n+1}} \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Z} \circ \xi(s - \tau_j))\lambda(\mathrm{d}s \times \mathrm{d}a) \\
&= \int_{t_n}^{t_{n+1}} \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Z} \circ \xi(s - \tau_j))m(\mathrm{d}s \times \mathrm{d}a) \\
&\quad + \int_{t_n}^{t_{n+1}} \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Z} \circ \xi(s - \tau_j))(\lambda - m)(\mathrm{d}s \times \mathrm{d}a) \\
&= \boldsymbol{\nu}_j \int_{t_n}^{t_{n+1}} a_j(\boldsymbol{Z} \circ \xi(t - \tau_j))\mathrm{d}t \\
&\quad + \left(\boldsymbol{\nu}_j \mathcal{P}\left(\int_{t_n}^{t_{n+1}} a_j(\boldsymbol{Z} \circ \xi(t - \tau_j))\mathrm{d}t\right) - \boldsymbol{\nu}_j \int_{t_n}^{t_{n+1}} a_j(\boldsymbol{Z} \circ \xi(t - \tau_j))\mathrm{d}t\right) \\
&= \boldsymbol{\nu}_j \mathcal{P}\left(\int_{t_n}^{t_{n+1}} a_j(\boldsymbol{Z} \circ \xi(t - \tau_j))\mathrm{d}t\right),
\end{aligned}$$

the scheme (2) may be considered as an explicit Euler-type scheme for SDDE (1), see [19].

In the following analysis we may need the continuous time version of D-leaping scheme (2),

$$\boldsymbol{Y}(t) = \begin{cases} \eta(0) + \sum_{j=1}^{M} \int_0^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y} \circ \xi(s-\tau_j))\lambda(\mathrm{d}s \times \mathrm{d}a), & t > 0, \\ \eta(t), \quad -\tau \le t \le 0,\ \tau = \max\{\tau_j, j \in I_d\}, \end{cases} \tag{3}$$

where $\xi(t) = t_n$ if $t \in [t_n,\ t_{n+1})$. We note that the above process $\boldsymbol{Y}(t)$ coincides with the result obtained from Algorithm 1 in each time-step point, i.e., $\boldsymbol{Y}(t_n) = \boldsymbol{Z}(t_n)$, $n \in \mathbb{Z}$.

3. Convergence order of the D-leaping scheme.

In this section, we will study the convergence order in both mean-square and weak senses of scheme (3) separately. It is shown that the mean-square strong convergence is of order 1/2 and the weak convergence is of order 1.

3.1. Strong convergence order.

First, in this part we prove the strong convergence order of the D-leaping scheme (3). The result is established via the Hölder continuity of the process $\boldsymbol{X}(t)$ (see Proposition 3.3), Itô isometry, and Itô formula. Moreover, the following two properties of the jump operator $\Delta\boldsymbol{X}(t) = \boldsymbol{X}(t) - \boldsymbol{X}(t-)$ are used frequently; see [13] for the proof.

LEMMA 3.1. *For any fixed* $s > 0$, $\Delta\boldsymbol{X}(s) = 0$ *a.s.*

LEMMA 3.2. *For any continuous function* $a(\boldsymbol{x})$ *and two positive reals* $d > c$, *we have* $\int_c^d \Delta a(\boldsymbol{X}(t))\mathrm{d}t = 0$.

In the following analysis, we let C denote the constant depending on the Lipschitz constant $L = \sum_{j=1}^{M} L_j$, the state change vectors $K = \max\{|\boldsymbol{\nu}_j|,\ j = 1, 2, \ldots, M\}$, the number of channels M, and the final time T, but not depending on time step n. Notice that the constant C may be different from line to line.

PROPOSITION 3.3. *Under Assumption* 2.1*, the SDDE driven by Poisson random measure* (1) *is well-posed in the sense that there exists a unique physical solution* $\boldsymbol{X}(t) \in \Omega_{\boldsymbol{X}_0}$ *in* $[0, \infty)$*. Furthermore, we have*

$$\mathbb{E}|\boldsymbol{X}(t) - \boldsymbol{X}(s)|^2 \le C|t - s|.$$

The proof of Proposition 3.3 is postponed to Appendix A. Based on these properties, We obtain the following strong convergence theorem, which shows that the strong convergence of the D-leaping scheme is of order 1/2.

THEOREM 3.4 (mean-square convergence). *Under Assumptions* 2.1–2.3 *and Proposition* 2.4 *we have*

$$\sup_{n \le N_T} \mathbb{E}|\boldsymbol{X}(t_n) - \boldsymbol{Y}(t_n)|^2 \le C\delta t, \tag{4}$$

where $\delta t := \max_n \delta t_n = \max_n(t_{n+1} - t_n)$.

Proof. In order to prove the strong convergence of the D-leaping scheme, we write

(1) from t_n to t_{n+1} as

$$\begin{aligned}\boldsymbol{X}(t_{n+1}) = \boldsymbol{X}(t_n) &+ \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(t-\tau_j-))m(\mathrm{d}t\times\mathrm{d}a)\\ &+ \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(t-\tau_j-))(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\end{aligned} \tag{5}$$

and (3) from t_n to t_{n+1} as

$$\begin{aligned}\boldsymbol{Y}(t_{n+1}) = \boldsymbol{Y}(t_n) &+ \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))m(\mathrm{d}t\times\mathrm{d}a)\\ &+ \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))(\lambda-m)(\mathrm{d}t\times\mathrm{d}a).\end{aligned} \tag{6}$$

Now we subtract (5) and (6) and define the error

$$\boldsymbol{E}(t_n) = \boldsymbol{X}(t_n) - \boldsymbol{Y}(t_n)$$

to get

$$\begin{aligned}&\boldsymbol{E}(t_{n+1}) = \boldsymbol{E}(t_n) + \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\boldsymbol{\nu}_j\Big(a_j(\boldsymbol{X}(t-\tau_j-)) - a_j(\boldsymbol{Y}\circ\xi(t-\tau_j))\Big)\mathrm{d}t\\ &+\sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j\Big(c_j(a;\ \boldsymbol{X}(t-\tau_j-)) - c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))\Big)(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\\ &=: \boldsymbol{E}(t_n) + \mathcal{A}_1 + \mathcal{A}_2,\end{aligned} \tag{7}$$

where we use the identity

$$\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(t-\tau_j-))m(\mathrm{d}t\times\mathrm{d}a) = \int_{t_n}^{t_{n+1}}\boldsymbol{\nu}_j a_j(\boldsymbol{X}(t-\tau_j-))\mathrm{d}t.$$

Squaring both sides of (7) we obtain

$$|\boldsymbol{E}(t_{n+1})|^2 = |\boldsymbol{E}(t_n)|^2 + |\mathcal{A}_1|^2 + |\mathcal{A}_2|^2 + 2\Big(\boldsymbol{E}(t_n)\cdot\mathcal{A}_1 + \boldsymbol{E}(t_n)\cdot\mathcal{A}_2 + \mathcal{A}_1\cdot\mathcal{A}_2\Big).$$

We estimate each term separately for the above equations.

For term $\mathcal{A}_1$, we have

$$\begin{aligned}\mathcal{A}_1 =& \sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\boldsymbol{\nu}_j\Big(a_j(\boldsymbol{X}\circ\xi(t-\tau_j)) - a_j(\boldsymbol{Y}\circ\xi(t-\tau_j))\Big)\mathrm{d}t\\ &+\sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\boldsymbol{\nu}_j\Big(a_j(\boldsymbol{X}(t-\tau_j-)) - a_j(\boldsymbol{X}\circ\xi(t-\tau_j))\Big)\mathrm{d}t\\ =:& \mathcal{A}_1^a + \mathcal{A}_1^b.\end{aligned}$$

By Itô isometry and Proposition 2.4,

$$
\begin{aligned}
&\mathbb{E}\left|\int_{t_n}^{t_{n+1}} \boldsymbol{\nu}_j\Big(a_j(\boldsymbol{X}\circ\xi(t-\tau_j)) - a_j(\boldsymbol{Y}\circ\xi(t-\tau_j))\Big)\mathrm{d}t\right|^2\\
&\le K^2\delta t\mathbb{E}\int_{t_n}^{t_{n+1}} |a_j(\boldsymbol{X}\circ\xi(t-\tau_j)) - a_j(\boldsymbol{Y}\circ\xi(t-\tau_j))|^2\mathrm{d}t\\
&\le K^2L^2\delta t\mathbb{E}\int_{t_n}^{t_{n+1}} |\boldsymbol{E}\circ\xi(t-\tau_j)|^2\mathrm{d}t,
\end{aligned}
$$

and also by Proposition 3.3,

$$
\begin{aligned}
&\mathbb{E}\left|\int_{t_n}^{t_{n+1}} \boldsymbol{\nu}_j\Big(a_j(\boldsymbol{X}(t-\tau_j-)) - a_j(\boldsymbol{X}\circ\xi(t-\tau_j))\Big)\mathrm{d}t\right|^2\\
&\le K^2\delta t\mathbb{E}\int_{t_n}^{t_{n+1}} |a_j(\boldsymbol{X}(t-\tau_j-)) - a_j(\boldsymbol{X}\circ\xi(t-\tau_j))|^2\mathrm{d}t\\
&\le K^2L^2\delta t\mathbb{E}\int_{t_n}^{t_{n+1}} |\boldsymbol{X}(t-\tau_j-) - \boldsymbol{X}\circ\xi(t-\tau_j)|^2\mathrm{d}t\\
&\le C\delta t^3.
\end{aligned}
$$

For term $\mathcal{A}_2$, we have

$$
\begin{aligned}
\mathcal{A}_2 =&\sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j\Big(c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j)) - c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))\Big)(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\\
&+\sum_{j=1}^{M}\int_{t_n}^{t_{n+1}}\int_0^A \boldsymbol{\nu}_j\Big(c_j(a;\ \boldsymbol{X}(t-\tau_j-)) - c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j))\Big)(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\\
=:&\ \mathcal{A}_2^a + \mathcal{A}_2^b.
\end{aligned}
$$

To deal with $\mathcal{A}_2^a$ and $\mathcal{A}_2^b$, we use the Itô formula to get

$$
\begin{aligned}
&\mathbb{E}\left|\int_{t_n}^{t_{n+1}}\int_0^A \Big(c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j)) - c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))\Big)(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\right|^2\\
&=\mathbb{E}\int_{t_n}^{t_{n+1}}\int_0^A |c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j)) - c_j(a;\ \boldsymbol{Y}\circ\xi(t-\tau_j))|^2 m(\mathrm{d}t\times\mathrm{d}a)\\
&\le\int_{t_n}^{t_{n+1}}\mathbb{E}\Big(|h_{j-1}(\boldsymbol{X}\circ\xi(t)) - h_{j-1}(\boldsymbol{Y}\circ\xi(t))| + |h_j(\boldsymbol{X}\circ\xi(t)) - h_j(\boldsymbol{Y}\circ\xi(t))|\Big)\mathrm{d}t\\
&\lesssim \max_{1\le j\le M}\int_{t_n}^{t_{n+1}}\mathbb{E}|\boldsymbol{E}\circ\xi(t-\tau_j)|\mathrm{d}t = \max_{1\le j\le M}\int_{t_n}^{t_{n+1}}\mathbb{E}|\boldsymbol{E}\circ\xi(t-\tau_j)|^2\mathrm{d}t
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left|\int_{t_n}^{t_{n+1}}\int_0^A \Big(c_j(a;\ \boldsymbol{X}(t-\tau_j-)) - c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j))\Big)(\lambda-m)(\mathrm{d}t\times\mathrm{d}a)\right|^2\\
&= \mathbb{E}\int_{t_n}^{t_{n+1}}\int_0^A |c_j(a;\ \boldsymbol{X}(t-\tau_j-)) - c_j(a;\ \boldsymbol{X}\circ\xi(t-\tau_j))|^2 m(\mathrm{d}t\times\mathrm{d}a)\\
&\le \int_{t_n}^{t_{n+1}} \mathbb{E}\Big(|h_{j-1}(\boldsymbol{X}(t-)) - h_{j-1}(\boldsymbol{X}\circ\xi(t))| + |h_j(\boldsymbol{X}(t-)) - h_j(\boldsymbol{X}\circ\xi(t))|\Big)\mathrm{d}t\\
&\lesssim \max_{1\le j\le M}\int_{t_n}^{t_{n+1}} \mathbb{E}|\boldsymbol{X}(t-\tau_j-) - \boldsymbol{X}\circ\xi(t-\tau_j)|\mathrm{d}t\\
&\le \max_{1\le j\le M}\int_{t_n}^{t_{n+1}} \mathbb{E}|\boldsymbol{X}(t-\tau_j) - \boldsymbol{X}\circ\xi(t-\tau_j)|^2\mathrm{d}t \le C\delta t^2.
\end{aligned}
$$

For term $\boldsymbol{E}(t_n)\cdot\mathcal{A}_1^a$, we have

$$
\begin{aligned}
\mathbb{E}\Big(\boldsymbol{E}(t_n)\cdot\mathcal{A}_1^a\Big) &\le C\delta t\mathbb{E}|\boldsymbol{E}(t_n)|^2 + C\frac{1}{\delta t}\mathbb{E}|\mathcal{A}_1^a|^2\\
&\le C\delta t\mathbb{E}|\boldsymbol{E}(t_n)|^2 + C\max_{1\le j\le M}\mathbb{E}\int_{t_n}^{t_{n+1}}|\boldsymbol{E}\circ\xi(t-\tau_j)|^2\mathrm{d}t.
\end{aligned}
$$

By the independence of Poisson random measure, we know that $\mathbb{E}(\boldsymbol{E}(t_n)\cdot\mathcal{A}_2) = 0$. For the other terms, we may use the Hölder inequality $ab \le \frac{1}{2}(a^2+b^2)$.

Finally, we have

$$
\mathbb{E}|\boldsymbol{E}(t_{n+1})|^2 \le C\delta t\mathbb{E}|\boldsymbol{E}(t_n)|^2 + C\max_{1\le j\le M}\mathbb{E}\int_{t_n}^{t_{n+1}}|\boldsymbol{E}\circ\xi(t-\tau_j)|^2\mathrm{d}t + C\delta t^2,
$$

which yields

$$
\mathbb{E}|\boldsymbol{E}(t_n)|^2 \le C\delta t \quad \forall\, n \le N_T.
$$

Thus we finish the proof. ∎

3.2. Weak convergence order. In this part, we will study the weak convergence order of scheme (3). The classical approach to prove the weak convergence order of the SDE is via a Kolmogorov PDE. However, due to the existence of time delay and the nondifferentiability of the coefficients of (1), the technique of a Kolmogorov PDE does not apply. To solve this problem, we rewrite the weak error as

$$
\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}(t_n)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n))\\
&= \sum_{i=1}^n \Big\{\mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))\Big\}\\
&\quad - \Big\{\mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))\Big\},
\end{aligned}
$$

where we utilize the tame property of process $\boldsymbol{Y}(t)$ (see Lemma 3.8 or Proposition 3.9); i.e., there exists a function u such that

$$
u(\Pi(\eta)) = \mathbb{E}\phi(\boldsymbol{Y}(t_n;t_i,\eta)).
$$

Then we estimate that each term in the summation is $\mathcal{O}(\delta t^2)$, where we may need to establish the Itô formula for tame functionals (see Proposition 3.10), and Malliavin calculus for SDDE (1) (see Proposition 3.11).

First, let us introduce some notation. Define the projection $\Pi:\ L([-\tau,0],\mathbb{R}^N)\to\mathbb{R}^{Nk}$ associated with $\mu_1,\dots,\mu_k\in[-\tau,0]$ by

$$\Pi(\eta):=(\eta(\mu_1),\dots,\eta(\mu_k))\in\mathbb{R}^{Nk} \tag{8}$$

for all $\eta\in L([-\tau,0],\mathbb{R}^N)$.

A functional $\Psi:\ [0,T]\times L([-\tau,0],\mathbb{R}^N)\to\mathbb{R}$ is called tame if there exists a function $f:\ [0,T]\times L([-\tau,0],\mathbb{R}^N)\to\mathbb{R}$ and a projection $\Pi:\ L([-\tau,0],\mathbb{R}^N)\to\mathbb{R}^{Nk}$ such that

$$\Psi(t,\eta)=f(t,\Pi(\eta)) \tag{9}$$

for all $t\in[0,T]$ and $\eta\in L([-\tau,0],\mathbb{R}^N)$.

For any continuous N-dimensional process $\boldsymbol{X}:\ [-\tau,T]\times\Omega\to\mathbb{R}^N$, define the segment $\boldsymbol{X}_t:\ [-\tau,0]\to\mathbb{R}^N$, $t\in[0,T]$, by

$$\boldsymbol{X}_t(u)=\boldsymbol{X}(t+u). \tag{10}$$

Denote by D the Malliavin differentiation operator associated with the Poisson random measure. For $F\in\mathbb{D}^{1,2}$, we call $D_{t,z}F$ the Malliavin derivative of F at (t,z). Here $\mathbb{D}^{1,2}$ is a stochastic Sobolev space consisting of all $\mathcal{F}_T$-measurable random variables $F\in L^2(P)$ with chaos expansion $F=\sum_{n=0}^{\infty}I_n(f_n)$ satisfying the convergence criterion $\|F\|^2_{\mathbb{D}^{1,2}}=\sum_{n=1}^{0}nn!\|f_n\|^2_{L^2}<\infty$. The operator D is defined by $D_{t,z}F=\sum_{n=1}^{\infty}nI_{n-1}(f_n(\cdot,t,z))$ for all $F\in\mathbb{D}^{1,2}$. We refer the reader to [16] for more details. In this section, we need some properties of Malliavin derivatives; see [16, Chapter 12].

Proposition 3.5 (chain rule). *Let $F\in\mathbb{D}^{1,2}$ and let ψ be a real continuous function on $\mathbb{R}$. Suppose $\psi(F)\in L^2(P)$ and $\Psi(F+D_{t,z}F)\in L^2(P\times\lambda\times\nu)$. Then $\psi(F)\in\mathbb{D}^{1,2}$ and*

$$D_{t,z}\psi(F)=\Psi(F+D_{t,z}F)-\psi(F). \tag{11}$$

The Skorohod integral can be considered as the adjoint operator to the Malliavin derivative, and it is an extension of the Itô integral. See [16, Definition 11.1] for the definition of Skorohod integral. Below is the relationship between the Malliavin derivative and the Skorohod integral.

Proposition 3.6 (duality formula). *Let $X(t,z)$, $t\in[0,T]$, $z\in[0,A]$, be Skorohod integrable and let $F\in\mathbb{D}^{1,2}$. Then*

$$\mathbb{E}\left[F\int_0^T\int_0^A X(t,z)(\lambda-m)(\mathrm{d}t\times\mathrm{d}z)\right]=\mathbb{E}\left[\int_0^T\int_0^A X(t,z)D_{t,z}Fm(\mathrm{d}t\times\mathrm{d}z)\right]. \tag{12}$$

Proposition 3.7 (fundamental theorem of calculus). *Let $X(s,y)$, $(s,y)\in[0,T]\times[0,A]$ be a stochastic process such that*

$$\mathbb{E}\left[\int_0^T\int_0^A|X(s,y)|^2m(\mathrm{d}s\times\mathrm{d}y)\right]<\infty.$$

Assume that $X(s,y) \in \mathbb{D}^{1,2}$ *for all* $(s,y) \in [0,T]\times[0,A]$ *and that* $D_{t,z}X(\cdot,\cdot)$ *is Skorohod integrable with*

$$\mathbb{E}\left[\int_0^T\int_0^A\Big|\int_0^T\int_0^A D_{t,z}X(s,y)(\lambda-m)(\mathrm{d}s\times\mathrm{d}y)\Big|^2 m(\mathrm{d}t\times\mathrm{d}z)\right]<\infty.$$

Then

$$\int_0^T\int_0^A X(s,y)(\lambda-m)(\mathrm{d}s\times\mathrm{d}y)\in\mathbb{D}^{1,2}$$

and

$$D_{t,z}\int_0^T\int_0^A X(s,y)(\lambda-m)(\mathrm{d}s\times\mathrm{d}y)=X(t,z)+\int_0^T\int_0^A D_{t,z}X(s,y)(\lambda-m)(\mathrm{d}s\times\mathrm{d}y). \tag{13}$$

To simplify notation in the proof of Lemma 3.8, we consider the case of uniform partition, and the delay is a multiple of time stepsize, which means that the D-leaping scheme (3) is equivalent to the following equation:

$$\boldsymbol{Y}(t)=\begin{cases}\eta(0)+\sum_{j=1}^M\int_0^t\int_0^A\boldsymbol{\nu}_jc_j(a;\ \boldsymbol{Y}\circ\xi(s)-\tau_j)\lambda(\mathrm{d}s\times\mathrm{d}a), & t>0,\\ \eta(t), \quad -\tau\le t\le 0.\end{cases} \tag{14}$$

Obviously, the results in what follows still hold for scheme (3).

The first lemma establishes the tame character of the Euler approximation (14). Its proof is given in Appendix B.

LEMMA 3.8. *Fix a partition point* t_i *for some* $i\in\{0,1,\dots,N_T\}$. *Then for a.a.* $\omega\in\Omega$, *the function*

$$[t_i,\ T]\times L([-\tau,\ 0],\mathbb{R})\to\mathbb{R},$$
$$(t,\ \eta)\mapsto\boldsymbol{Y}(t,\omega;t_i,\eta)$$

is a tame functional; i.e., there exists a random function $\boldsymbol{F}$ *such that*

$$\boldsymbol{Y}(t,\omega;t_i,\eta)=\boldsymbol{F}(t,\omega,\Pi(\eta)).$$

PROPOSITION 3.9. *Given any fixed* t, $\mathbb{E}\phi(\boldsymbol{Y}(t;t_i,\eta))$ *is a tame functional, which means there exists a deterministic function* u *such that*

$$\mathbb{E}\phi(\boldsymbol{Y}(t;t_i,\eta))=\mathbb{E}\phi(\boldsymbol{F}(t,\Pi(\eta)))=:u(\Pi(\eta)).$$

In order to derive the weak convergence order of the D-leaping scheme, we shall first establish a tame Itô formula. It describes how the segment process $\boldsymbol{X}_t$ transforms under tame functionals. We state the proof in Appendix C.

PROPOSITION 3.10. *Assume that*

$$X(t)=\begin{cases}\eta(0)+\int_0^t\int_0^A K(s,a)\lambda(\mathrm{d}s\times\mathrm{d}a), & t>0,\\ \eta(t), & -\tau\le t\le 0.\end{cases} \tag{15}$$

Suppose $\phi \in C(\mathbb{R}^k;\mathbb{R})$ and let Π be the tame projection. Then for all $t \in [0,T]$, we have a.s.

$$
\begin{aligned}
&\phi(\Pi(X_t)) - \phi(\Pi(X_0)) = \sum_{i=1}^{k} \int_0^t \int_0^A,\\
&\Big[\phi(X_{s-}(\mu_1),\ldots,X_{s-}(\mu_{i-1}),X_{s-}(\mu_i)+K(s+\mu_i,a),X_s(\mu_{i+1}),\ldots,X_s(\mu_k))\\
&\quad - \phi(X_{s-}(\mu_1),\ldots,X_{s-}(\mu_{i-1}),X_{s-}(\mu_i),X_s(\mu_{i+1}),\ldots,X_s(\mu_k))\Big]\lambda(\mathrm{ds}\times\mathrm{da}).
\end{aligned}
\tag{16}
$$

Let $X(t) := X(t;\sigma,\eta)$, $t \in [\sigma-\tau,T]$ be the solution with initial process η at time σ, i.e.,

$$
X(t) = \begin{cases} \eta(0) + \sum_{j=1}^{M} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau_j-))\lambda(\mathrm{ds}\times\mathrm{d}a), & t > \sigma,\\ \eta(t-\sigma), & \sigma-\tau \le t \le \sigma. \end{cases}
\tag{17}
$$

Moreover, we also need the solution $X(t)$ to be Malliavin differentiable. The proof of the following proposition is stated in Appendix D.

Proposition 3.11. *For any $\eta \in L^2(\Omega, L([-\tau,0],\mathbb{R});\mathcal{F}_\sigma)$ with*

$$
\sup_{\sigma-\tau\le s\le\sigma} \mathbb{E}\int_0^A \|D_{s,z}\eta\|_\infty^2 < \infty,
$$

the solution $X(t)$ of (17) belongs to $\mathbb{D}^{1,2}$ for all $t \in [\sigma-\tau,T]$. Moreover, there exists a positive constant C such that

$$
\sup_{0\le\sigma\le T}\ \sup_{\sigma-\tau\le r,t\le T} \mathbb{E}\int_0^A |D_{r,z}X(t;\sigma,\eta)|^2 \mathrm{d}z \le C\left(1 + \sup_{\sigma-\tau\le s\le\sigma}\mathbb{E}\int_0^A \|D_{s,z}\eta\|_\infty^2\right).
\tag{18}
$$

Finally, we obtain the following weak convergence theorem, which means that the weak convergence order of the D-leaping scheme (3) is 1.

Theorem 3.12 (weak convergence). *There exists a positive constant C such that*

$$
|\mathbb{E}\phi(\boldsymbol{X}(t_n)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n))| \le C\delta t
\tag{19}
$$

for all $n \in \{1,2,\ldots,N_T\}$ and $\phi:\ \mathbb{R}^N \to \mathbb{R}$ of class C_b^2.

Proof. Using the Markov property for the segments $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ (see [15]), we may write the weak error as

$$
\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}(t_n;0,\eta)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n;0,\eta))\\
&= \mathbb{E}\phi(\boldsymbol{Y}(t_n;t_n,\boldsymbol{X}_{t_n}(\cdot;0,\eta))) - \mathbb{E}\phi(\boldsymbol{Y}(t_n;t_0,\boldsymbol{X}_{t_0}(\cdot;0,\eta)))\\
&= \sum_{i=1}^{n}\Big\{\mathbb{E}\phi(\boldsymbol{Y}(t_n;t_i,\boldsymbol{X}_{t_i}(\cdot;0,\eta))) - \mathbb{E}\phi(\boldsymbol{Y}(t_n;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))\Big\}\\
&= \sum_{i=1}^{n}\Big\{\mathbb{E}\phi(\boldsymbol{Y}(t_n;t_i,\boldsymbol{X}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))))\\
&\qquad - \mathbb{E}\phi(\boldsymbol{Y}(t_n;t_i,\boldsymbol{Y}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))))\Big\}.
\end{aligned}
\tag{20}
$$

From Proposition 3.9, we know that there exists a function u such that

$$u(\Pi(\eta)) = \mathbb{E}\phi(\boldsymbol{Y}(t_n; t_i, \eta)).$$

Thus we rewrite (20) as

$$\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}(t_n;0,\eta)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n;0,\eta)) \\
&= \sum_{i=1}^{n} \Big\{ \mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) \Big\} \\
&= \sum_{i=1}^{n} \Big\{ \mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))) \Big\} \\
&\quad - \Big\{ \mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))) \Big\}.
\end{aligned} \tag{21}$$

By the tame Itô formula (Proposition 3.10), we obtain

$$\begin{aligned}
&\mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))) \\
&= \sum_{m=1}^{k} \mathbb{E}\int_{t_{i-1}}^{t_i}\int_0^A \Bigg[u\Bigg(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}_{s-}(\mu_m-\tau_j)), \ldots \Bigg) \\
&\qquad - u(\ldots, \boldsymbol{X}_{s-}(\mu_m), \ldots) \Bigg] \lambda(\mathrm{d}s \times \mathrm{d}a) \\
&= \sum_{m=1}^{k} \mathbb{E}\int_{t_{i-1}}^{t_i} \Bigg\{ \sum_{j=1}^{M} a_j(\boldsymbol{X}_{s-}(\mu_m-\tau_j)) \Big[u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots) \\
&\qquad - u(\ldots, \boldsymbol{X}_{s-}(\mu_m), \ldots) \Big] \Bigg\} \mathrm{d}s
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}(\cdot;t_{i-1},\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta)))) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}(\cdot;0,\eta))) \\
&= \sum_{m=1}^{k} \mathbb{E}\int_{t_{i-1}}^{t_i}\int_0^A \Bigg[u\Bigg(\ldots, \boldsymbol{Y}_{s-}(\mu_m) + \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}_{\xi(s)}(\mu_m-\tau_j)), \ldots \Bigg) \\
&\qquad - u(\ldots, \boldsymbol{Y}_{s-}(\mu_m), \ldots) \Bigg] \lambda(\mathrm{d}s \times \mathrm{d}a) \\
&= \sum_{m=1}^{k} \mathbb{E}\int_{t_{i-1}}^{t_i} \Bigg\{ \sum_{j=1}^{M} a_j(\boldsymbol{Y}(t_{i-1}+\mu_m-\tau_j)) \Big[u(\ldots, \boldsymbol{Y}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots) \\
&\qquad - u(\ldots, \boldsymbol{Y}_{s-}(\mu_m), \ldots) \Big] \Bigg\} \mathrm{d}s.
\end{aligned}$$

We define

$$f_j^m(\Pi(\boldsymbol{X}_s)) = u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots) - u(\ldots, \boldsymbol{X}_{s-}(\mu_m), \ldots).$$

Thus (20) is

$$
\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}(t_n;0,\eta)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n;0,\eta)) \\
&= \sum_{i=1}^{n}\sum_{m=1}^{k}\sum_{j=1}^{M}\mathbb{E}\int_{t_{i-1}}^{t_i}\Big[a_j(\boldsymbol{X}(s+\mu_m-\tau_j-))f_j^m(\Pi(\boldsymbol{X}_s)) \\
&\qquad - a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))f_j^m(\Pi(\boldsymbol{Y}_s))\Big]\mathrm{d}s \\
&=: \sum_{i=1}^{n}\sum_{m=1}^{k}\sum_{j=1}^{M}\mathcal{D}_{m,j}^{i}.
\end{aligned}
\tag{22}
$$

In what follows, we need to show that $\mathcal{D}_{m,j}^i \sim O(\delta t^2)$.

We note that

$$
\begin{aligned}
\mathcal{D}_{m,j}^i &= \int_{t_{i-1}}^{t_i}\mathbb{E}\Big\{[a_j(\boldsymbol{X}(s+\mu_m-\tau_j-)) - a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))]f_j^m(\Pi(\boldsymbol{X}_s))\Big\}\mathrm{d}s \\
&\quad + \int_{t_{i-1}}^{t_i}\mathbb{E}\Big\{a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))[f_j^m(\Pi(\boldsymbol{X}_s)) - f_j^m(\Pi(\boldsymbol{Y}_s))]\Big\}\mathrm{d}s \\
&=: \int_{t_{i-1}}^{t_i}\mathcal{D}_{m,j}^{i,1}(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i}\mathcal{D}_{m,j}^{i,2}(s)\mathrm{d}s.
\end{aligned}
\tag{23}
$$

We claim that for all $s \in [t_{i-1}, t_i]$, we have $\mathcal{D}_{m,j}^{i,1}(s),\ \mathcal{D}_{m,j}^{i,2}(s) \sim O(\delta t)$, which means that $\mathcal{D}_{m,j}^i \sim O(\delta t^2)$.

In fact, by the Itô formula,

$$
\begin{aligned}
\mathcal{D}_{m,j}^{i,1}(s) &= \mathbb{E}\left\{f_j^m(\Pi(\boldsymbol{X}_s))\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A\Big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big]\lambda(\mathrm{d}u\times\mathrm{d}a)\right\} \\
&= \mathbb{E}\left\{f_j^m(\Pi(\boldsymbol{X}_s))\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A\Big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big](\lambda-m)(\mathrm{d}u\times\mathrm{d}a)\right\} \\
&\quad + \mathbb{E}\left\{f_j^m(\Pi(\boldsymbol{X}_s))\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A\Big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big]m(\mathrm{d}u\times\mathrm{d}a)\right\} \\
&= \mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A D_{u,a}f_j^m(\Pi(\boldsymbol{X}_s))\Big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big]m(\mathrm{d}u\times\mathrm{d}a) \\
&\quad + \mathbb{E}\Bigg\{f_j^m(\Pi(\boldsymbol{X}_s))\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\Bigg[\sum_{\ell=1}^{M}a_\ell(\boldsymbol{X}(u-\tau_\ell-)) \\
&\qquad \Big(a_j(\boldsymbol{X}(u-)+\boldsymbol{\nu}_\ell) - a_j(\boldsymbol{X}(u-))\Big)\Bigg]\mathrm{d}u\Bigg\},
\end{aligned}
\tag{24}
$$

where

$$
\tilde{\boldsymbol{X}}(u-) := \boldsymbol{X}(u-) + \sum_{\ell=1}^{M}\boldsymbol{\nu}_\ell c_\ell(a;\ \boldsymbol{X}(u-\tau_\ell-)).
$$

Noting that

$$f_j^m(\Pi(\boldsymbol{X}_s)) = u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots) - u(\ldots, \boldsymbol{X}_{s-}(\mu_m), \ldots),$$

we make the following estimates for functions of f_j^m:

$$\begin{aligned}
\mathbb{E}|f_j^m(\Pi(\boldsymbol{X}_s))|^2 &\le \mathbb{E}\Big|u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots)\Big|^2 + \mathbb{E}\Big|u(\ldots, \boldsymbol{X}_{s-}(\mu_m), \ldots)\Big|^2 \\
&= \mathbb{E}\Big|\mathbb{E}\Big(\phi(\boldsymbol{Y}(t_n; t_i, \eta))\Big|\eta = \widetilde{\boldsymbol{X}}_{s-}\Big)\Big|^2 + \mathbb{E}\Big|\mathbb{E}\Big(\phi(\boldsymbol{Y}(t_n; t_i, \eta))\Big|\eta = \boldsymbol{X}_{s-}\Big)\Big|^2 \\
&\le \mathbb{E}\Big|\phi(\boldsymbol{Y}(t_n; t_i, \widetilde{\boldsymbol{X}}_{s-}))\Big|^2 + \mathbb{E}\Big|\phi(\boldsymbol{Y}(t_n; t_i, \boldsymbol{X}_{s-}))\Big|^2,
\end{aligned}$$

where $\widetilde{\boldsymbol{X}}_{s-} \in L([-\tau, 0], \mathbb{R}^N)$ is defined by

$$\Pi(\widetilde{\boldsymbol{X}}_{s-}) = \Big(\boldsymbol{X}_{s-}(\mu_1) \ldots, \boldsymbol{X}_{s-}(\mu_m) + \boldsymbol{\nu}_j, \ldots, \boldsymbol{X}_{s-}(\mu_k)\Big),$$

and thus

$$\mathbb{E}|f_j^m(\Pi(\boldsymbol{X}_s))|^2 \le C\|\phi\|^2_{C_b^1}(1 + \mathbb{E}\|\boldsymbol{X}(s+\cdot)\|^2_{L^2([-\tau,0])}) \le C.$$

And the estimates for $D_{u,z} f_j^m$ are as follows. Since

$$\begin{aligned}
&\mathbb{E}\int_0^A |D_{u,z} f_j^m(\Pi(\boldsymbol{X}_s))|^2 \mathrm{d}z \\
&= \mathbb{E}\int_0^A \Big|\sum_{\ell=1}^k \Big(f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell) + D_{u,z}\boldsymbol{X}_s(\mu_\ell), \ldots) - f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell), \ldots)\Big)\Big|^2 \mathrm{d}z \\
&\le C\mathbb{E}\int_0^A \sum_{\ell=1}^k \Big|f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell) + D_{u,z}\boldsymbol{X}_s(\mu_\ell), \ldots) - f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell), \ldots)\Big|^2 \mathrm{d}z,
\end{aligned}$$

we have

$$\begin{aligned}
&\mathbb{E}\Big|f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell) + D_{u,z}\boldsymbol{X}_s(\mu_\ell), \ldots) - f_j^m(\ldots, \boldsymbol{X}_s(\mu_\ell), \ldots)\Big|^2 \\
&\le 2\mathbb{E}\Big|u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + D_{u,z}\boldsymbol{X}_{s-}(\mu_m) + \nu_j) - u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + \nu_j)\Big|^2 \\
&\quad + 2\mathbb{E}\Big|u(\ldots, \boldsymbol{X}_{s-}(\mu_m) + D_{u,z}\boldsymbol{X}_{s-}(\mu_m)) - u(\ldots, \boldsymbol{X}_{s-}(\mu_m))\Big|^2 \\
&\le C\|\phi\|^2_{C_b^1}\|D_{u,z}\boldsymbol{X}_s(\cdot)\|^2_{L^2([-\tau,0])},
\end{aligned}$$

where in the last step we use the same technique (conditional expectation) as in the estimate for f_j^m, and the fact that the difference in the solutions $\boldsymbol{Y}$ with different initial data could be controlled by the difference of the initial data; i.e., for any initial data $\eta, \xi \in L([-\tau, 0], \mathbb{R}^N)$, and $t > s$, we get $\mathbb{E}|\boldsymbol{Y}(t, s, \eta) - \boldsymbol{Y}(t, s, \xi)|^2 \le C\mathbb{E}\|\eta - \xi\|^2_{L^2([-\tau,0])}$. Further,

$$\mathbb{E}\int_0^A |D_{u,z} f_j^m(\Pi(\boldsymbol{X}_s))|^2 \mathrm{d}z \le C\|\phi\|^2_{C_b^1}\int_0^A \mathbb{E}\|D_{u,z}\boldsymbol{X}(s+\cdot)\|^2_{L^2([-\tau,0])}\mathrm{d}z \le C.$$

Thus following from Proposition 2.4, we can see that $\mathcal{D}^{i,1}_{m,j}(s) \leq C\delta t$. In fact, the first term on the right-hand side of (24) can be estimated by the Hölder inequality

$$
\begin{aligned}
&\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A D_{u,a}f_j^m(\Pi(\boldsymbol{X}_s))\Big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big]m(\mathrm{d}u\times\mathrm{d}a)\\
&\leq \int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A \Big\{\mathbb{E}|D_{u,a}f_j^m(\Pi(\boldsymbol{X}_s))|^2\\
&\qquad + \mathbb{E}\Big|a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\Big|^2\Big\}m(\mathrm{d}u\times\mathrm{d}a)\\
&\leq C\delta t + L^2\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A |\tilde{\boldsymbol{X}}(u) - \boldsymbol{X}(u)|^2 m(\mathrm{d}u\times\mathrm{d}a)\\
&\leq C\delta t + L^2K^2\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\int_0^A \sum_{\ell=1}^M |c_\ell(a;\ \boldsymbol{X}(u-\tau_\ell-))|^2 m(\mathrm{d}u\times\mathrm{d}a)\\
&= C\delta t + L^2K^2\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\sum_{\ell=1}^M a_\ell(\boldsymbol{X}(u-\tau_\ell-))\mathrm{d}u\\
&\leq C\delta t + CL^3K^2\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}\left(1 + \sum_{\ell=1}^M |\boldsymbol{X}(u-\tau_\ell-)|^2\right)\mathrm{d}u\\
&\leq C\delta t.
\end{aligned}
$$

The second term on the right-hand side of (24) can be estimated similarly as

$$
\begin{aligned}
&\mathbb{E}\left\{f_j^m(\Pi(\boldsymbol{X}_s))\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} g(u)\mathrm{d}u\right\}\\
&\leq \Big(\mathbb{E}|f_j^m(\Pi(\boldsymbol{X}_s))|^2\Big)^{\frac{1}{2}}\left(\mathbb{E}\Big|\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} g(u)\mathrm{d}u\Big|^2\right)^{\frac{1}{2}}\\
&\leq C\delta t^{\frac{1}{2}}\left(\mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j}|g(u)|^2\mathrm{d}u\right)^{\frac{1}{2}} \leq C\delta t,
\end{aligned}
$$

where

$$
g(u) = \sum_{\ell=1}^M a_\ell(\boldsymbol{X}(u-\tau_\ell-))\Big(a_j(\boldsymbol{X}(u-)+\boldsymbol{\nu}_\ell) - a_j(\boldsymbol{X}(u-))\Big)
$$

and

$$
\mathbb{E}|g(u)|^2 \leq L^2K^2\mathbb{E}\left(1 + \sum_{\ell=1}^M |\boldsymbol{X}(u-\tau_\ell-)|^2\right) \leq C.
$$

The estimate of term $\mathcal{D}^{i,2}_{m,j}(s)$ is similar but uses the tame Itô formula.

$$
\begin{aligned}
\mathcal{D}^{i,2}_{m,j}(s) &= \mathbb{E}\Big\{a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\Big[\Big(f_j^m(\Pi(\boldsymbol{X}_s)) - f_j^m(\Pi(\boldsymbol{X}_{t_{i-1}}))\Big)\\
&\quad - \Big(f_j^m(\Pi(\boldsymbol{Y}_s)) - f_j^m(\Pi(\boldsymbol{X}_{t_{i-1}}))\Big)\Big]\Big\}\\
&= \mathbb{E}\int_{t_{i-1}}^{s}\int_0^A D_{u,a}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\sum_{\ell=1}^{k}\\
&\quad \Big\{[f_j^m(\ldots,\tilde{\boldsymbol{X}}_{u-}(\mu_\ell),\ldots) - f_j^m(\ldots,\boldsymbol{X}_{u-}(\mu_\ell),\ldots)]\\
&\quad - [f_j^m(\ldots,\tilde{\boldsymbol{Y}}_{u-}(\mu_\ell),\ldots) - f_j^m(\ldots,\boldsymbol{Y}_{u-}(\mu_\ell),\ldots)]\Big\}m(\mathrm{d}u\times\mathrm{d}a)\\
&\quad + \mathbb{E}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\int_{t_{i-1}}^{s}\sum_{\ell=1}^{k}\sum_{j_1=1}^{M}\Big[a_{j_1}(\boldsymbol{X}(u+\mu_\ell-\tau_{j_1}-))F^\ell_{j_1}(\Pi(\boldsymbol{X}_u))\\
&\quad - a_{j_1}(\boldsymbol{X}(t_{i-1}+\mu_\ell-\tau_{j_1}-))F^\ell_{j_1}(\Pi(\boldsymbol{Y}_u))\Big]\mathrm{d}u,
\end{aligned}
\tag{25}
$$

where

$$
F^\ell_{j_1}(\Pi(\boldsymbol{X}_u)) = f_j^m(\ldots,\boldsymbol{X}_{u-}(\mu_\ell)+\boldsymbol{\nu}_{j_1},\ldots) - f_j^m(\ldots,\boldsymbol{X}_{u-}(\mu_\ell),\ldots).
$$

By Propositions 2.4 and 3.11, we may show that $\mathcal{D}^{i,2}_{m,j}(s)\le C\delta t$. In fact, the first term on the right-hand side of (25) can be estimated by the Hölder inequality,

$$
\begin{aligned}
&\mathbb{E}\int_{t_{i-1}}^{s}\int_0^A D_{u,a}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))g_1(u,a)m(\mathrm{d}u\times\mathrm{d}a)\\
&\le \int_{t_{i-1}}^{s}\int_0^A \Big[\mathbb{E}|D_{u,a}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))|^2 + \mathbb{E}|g_1(u,a)|^2\Big]m(\mathrm{d}u\times\mathrm{d}a)\\
&\le C\delta t + \int_{t_{i-1}}^{s}\int_0^A \mathbb{E}|D_{u,a}\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j)|^2 m(\mathrm{d}u\times\mathrm{d}a)\\
&\le C\delta t,
\end{aligned}
$$

where

$$
\begin{aligned}
g_1(u,a) = \sum_{\ell=1}^{k}&[f_j^m(\ldots,\tilde{\boldsymbol{X}}_{u-}(\mu_\ell),\ldots) - f_j^m(\ldots,\boldsymbol{X}_{u-}(\mu_\ell),\ldots)]\\
&- [f_j^m(\ldots,\tilde{\boldsymbol{Y}}_{u-}(\mu_\ell),\ldots) - f_j^m(\ldots,\boldsymbol{Y}_{u-}(\mu_\ell),\ldots)]
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_0^A \mathbb{E}|g_1(u,a)|^2\mathrm{d}a\\
&\le C\int_0^A \mathbb{E}\|\tilde{\boldsymbol{X}}(u+\cdot)-\boldsymbol{X}(u+\cdot)\|^2_{L^2([-\tau,0])} + \mathbb{E}\|\tilde{\boldsymbol{Y}}(u+\cdot)-\boldsymbol{Y}(u+\cdot)\|^2_{L^2([-\tau,0])}\mathrm{d}a\\
&\le C\int_0^A \sum_{j=1}^{M}\mathbb{E}\|c_j(a;\ \boldsymbol{X}_{u-\tau_j}(\cdot))\|_{L^2([-\tau,0])} + \sum_{j=1}^{M}\mathbb{E}\|c_j(a;\ \boldsymbol{Y}_{u-\tau_j}(\cdot))\|_{L^2([-\tau,0])}\mathrm{d}a\\
&= C\mathbb{E}\left(\sum_{j=1}^{M}\|a_j(\boldsymbol{X}_{u-\tau_j}(\cdot))\|_{L^1([-\tau,0])} + \sum_{j=1}^{M}\|a_j(\boldsymbol{Y}_{u-\tau_j}(\cdot))\|_{L^1([-\tau,0])}\right)\\
&\le C\mathbb{E}(1+\|\boldsymbol{X}_{u-\tau_j}(\cdot)\|^2_{L^2([-\tau,0])} + \|\boldsymbol{Y}_{u-\tau_j}(\cdot)\|^2_{L^2([-\tau,0])}) \le C.
\end{aligned}
$$

The second term on the right-hand side of (25) can be estimated similarly as

$$\begin{aligned}
&\mathbb{E}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\int_{t_{i-1}}^{s} g_2(u)\mathrm{d}u\\
&\le \left(\mathbb{E}|a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))|^2\right)^{\frac{1}{2}}\left(\mathbb{E}|\int_{t_{i-1}}^{s} g_2(u)\mathrm{d}u|^2\right)^{\frac{1}{2}}\\
&\le L\delta t^{\frac{1}{2}}\left(1+\mathbb{E}|\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j)|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\int_{t_{i-1}}^{s}|g_2(u)|^2\mathrm{d}u\right)^{\frac{1}{2}}\\
&\le C\delta t,
\end{aligned}$$

where

$$\begin{aligned}
g_2(u)=\sum_{\ell=1}^{k}\sum_{j_1=1}^{M}\Big[&a_{j_1}(\boldsymbol{X}(u+\mu_\ell-\tau_{j_1}-))F_{j_1}^{\ell}(\Pi(\boldsymbol{X}_u))\\
&-a_{j_1}(\boldsymbol{X}(t_{i-1}+\mu_\ell-\tau_{j_1}-))F_{j_1}^{\ell}(\Pi(\boldsymbol{Y}_u))\Big]
\end{aligned}$$

and

$$\mathbb{E}|g_2(u)|^2\le C\left(1+\sup_{u\in[t_{i-1},s]}\max_{1\le\ell\le k}\max_{1\le j\le M}\mathbb{E}|\boldsymbol{X}(u+\mu_\ell-\tau_j)|^2\right)\le C.$$

Therefore $\mathcal{D}_{m,j}^{i}\le C\delta t^2$. Substituting it into (22), we finish the proof. ∎

3.3. Numerical examples. We apply ghd D-leaping method for two chemical reaction systems. In these systems, exact solutions are obtained from the SSA algorithm. In order to demonstrate the order of accuracy, we follow a procedure that is widely used in the numerical study of SDEs. We simulate $\boldsymbol{X}(t)$ from time $t=0$ to $t=T$, advancing by a fixed time stepsize δt. If the sample size is large enough, the statistical error in the expectation for a function of the solution could be neglected. We double the stepsize to $2\delta t$ to calculate the convergence order.

3.3.1. Example 1. For this system, we consider $S\to\emptyset$ with the propensity function being $a_1(x)=cx$, where the rate constant $c=0.1$. The state change vector is $\nu_1=-1$, the time delay is $\tau=1$, and the initial condition is $X_0=10000$. We simulate the reaction from time 0 to $T=10$ using different stepsizes.

We plot the strong error in Figure 1 and the absolute errors of mean and variance in Figure 2. The sample size is as large as 10^6 so that the magnitude of statistical fluctuation is small. It shows that, for the system, the D-leaping scheme has half order accuracy for the strong convergence and first order accuracy for the weak convergence.

3.3.2. Example 2. This example has two reaction channels:

$$R_1:\ S_1+S_2\to S_3,\quad R_2:\ S_3\to\emptyset.$$

The reaction channel R_2 fires without delay, but the reaction channel R_1 incurs a delay. We assume that R_1 belongs to consuming type, which means that once an R_1 reaction occurs, we immediately have $X_1=X_1-1$ and $X_2=X_2-1$, but we will have $X_2=X_3+1$ after a delay. In our simulations, we chose $c_1=0.001$ and $c_2=0.001$, used $X_1(0)=1000$, $X_2(0)=1000$, and $X_3(0)=0$, and set the delay of R_1 to be $\tau=0.1$.

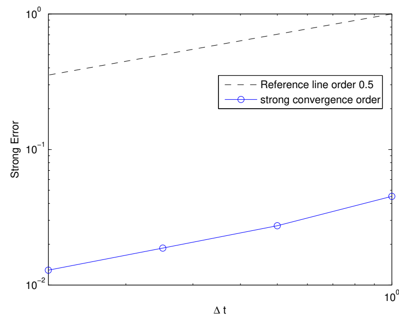


FIG. 1. *Log-log plot of the strong error.*

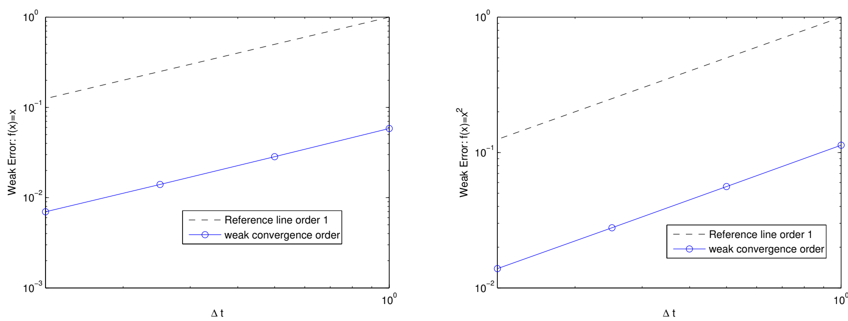


FIG. 2. *Log-log plot of the absolute error with functions $f(x) = x$ and $f(x) = x^2$, respectively.*

We ran simulation 10^5 times, and in each time, simulation starts at $t = 0$ and ends at $T = 1$. Figures 3 and 4 depict the strong and weak convergence behaviors of the D-leaping scheme applied to this system, from which we may observe that the strong convergence of the D-leaping scheme is of half order and the weak convergence is of first order.

4. Generalization for highly accurate methods. The construction of high weak order schemes for SDEs plays an important role in the implications in the efficient simulations of SDEs. Utilizing the Itô formula to expand the solution of SDEs and then truncating the solution series is the common method to construct high weak order schemes; see, for instance, [17]. By adding a random correction to the primitive tau-leaping scheme in each time step, [10] presents a new method which improves the accuracy of the approximations. This gain in accuracy actually comes from the reduction in the local truncation error of the scheme in the order of τ, the marching time stepsize. We introduce the definition of weak consistency (see [10, Definition 1]):

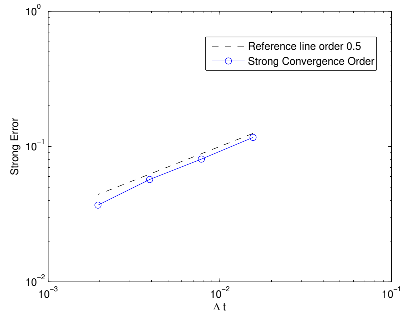


FIG. 3. *Log-log plot of the strong error.*

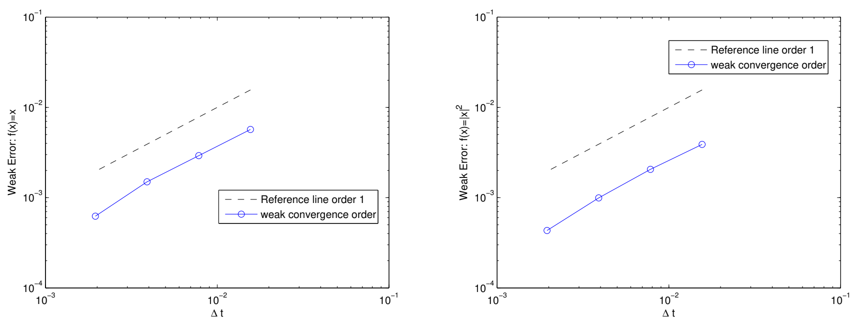


FIG. 4. *Log-log plot of the absolute error with functions $f(x) = x$ and $f(x) = |x|^2$, respectively.*

if there exist $C > 0$ and $\delta > 0$ such that for all $\tau \in [0, \delta]$,

$$\Big|\mathbb{E}_x\big[(X_{n+1} - X_n)^p\big] - \mathbb{E}_x\big[(X(t_n + \tau) - X(t_n))^p\big]\Big| \le C\tau^{q+1},$$

we say that the numerical scheme $\{X_n\}_{n\in\mathbb{N}}$ is weak consistent for the pth moment to qth order. Here $\mathbb{E}_x$ denotes the expectation conditioned on $X(t_n) - X_n = x$. Remark 3 in [10] states that if a numerical scheme is stable and qth order consistent, then it is of qth order accuracy.

In this section, we take Example 1 for demonstration to investigate the method in [10] for the system with delays, i.e., SDDEs driven by the Poisson random measure. Similarly, by adding a random correction to the primitive D-leaping scheme in each step, we are able to greatly improve the accuracy of the D-leaping scheme for the mean. However, delay introduces more technical difficulties for the improvement to higher order moments of the solution. To solve this problem, we first fix the test function (for example, $\phi(x) = |x|^2$ denotes second order moment), and then use the tame Itô formula to obtain the new variable $Z := \phi(X)$, and finally we add the random correction to the augmented variables.

The construction of higher order accuracy methods for mean is similar to that in [10]. From Example 1 we obtain

$$
\begin{aligned}
&\mathbb{E}(X(t_{n+1}) - X(t_n)) \\
&= \mathbb{E}\Big[\int_{t_n}^{t_{n+1}} \int_0^A \nu_1 c_1(a;\ X(s-\tau-))\lambda(\mathrm{d}s \times \mathrm{d}a)\Big] \\
&= \nu_1 \mathbb{E}\int_{t_n}^{t_{n+1}} a_1(X(t-\tau))\mathrm{d}t = \nu_1 \mathbb{E}\int_{t_n-\tau}^{t_{n+1}-\tau} a_1(X(t))\mathrm{d}t \\
&= \nu_1 \mathbb{E}\int_{t_n-\tau}^{t_{n+1}-\tau} \Big[a_1(X(t_n-\tau)) \\
&\quad + \int_{t_n-\tau}^{t}\int_0^A \Big(a_1(X(s-) + \nu_1 c_1(a;\ X(s-\tau-))) - a_1(X(s-))\Big)\lambda(\mathrm{d}s \times \mathrm{d}a)\Big]\mathrm{d}t \\
&= \nu_1 \delta t \mathbb{E} a_1(X(t_n-\tau)) \\
&\quad + \nu_1 \mathbb{E}\int_{t_n-\tau}^{t_{n+1}-\tau}\int_{t_n-\tau}^{t} a_1(X(s-\tau))\Big(a_1(X(s)+\nu_1) - a_1(X(s))\Big)\mathrm{d}s\mathrm{d}t \\
&= \nu_1 \frac{\delta t^2}{2}\mathbb{E} a_1(X(t_n-2\tau))\Big(a_1(X(t_n-\tau)+\nu_1) - a_1(X(t_n-\tau))\Big) \\
&\quad + \nu_1 \delta t \mathbb{E} a_1(X(t_n-\tau)) + \mathcal{O}(\delta t^3).
\end{aligned}
$$

For the numerical scheme, we take the following uniform mesh on $[0,T]$: $0 = t_0 < t_1 < \cdots < t_N = T$, where $t_n = t_0 + n\delta t$, $n = 0, 1, \ldots, N$. In addition, the choice of δt is not arbitrary; it has to be chosen such that $\ell := \tau/\delta t \in \mathbb{N}$. In other words, the delay period τ has to be a multiple of δt. Thus the D-leaping scheme reads

$$Y_{n+1} = Y_n + \nu_1 r_1,$$

with $r_1 = \mathcal{P}(a_1(Y_{n-\ell})\delta t)$.

Consider the D-leaping scheme with a random correction $\tilde{r}_1$:

$$Y_{n+1} = Y_n + \nu_1(r_1 + \tilde{r}_1). \tag{26}$$

We require

$$\mathbb{E}_\eta \mathbb{E}_{r_1}(\tilde{r}_1) = \frac{\delta t^2}{2} a_1(\eta(-2\tau))\xi + \mathcal{O}(\delta t^3)$$

with $\xi = a_1(\eta(-\tau)+\nu_1) - a_1(\eta(-\tau)) = c\nu_1$ to obtain

$$\mathbb{E}\Big(X(t_{n+1}; t_n, \eta) - \eta(0)\Big) - \mathbb{E}\Big(Y(t_{n+1}; t_n, \eta) - \eta(0)\Big) = \mathcal{O}(\delta t^3). \tag{27}$$

In practice, we may take $\tilde{r}_1 = \mathrm{sgn}(\alpha)\mathcal{P}(|\alpha|)$, where

$$\alpha = \frac{\delta t^2}{2} a_1(Y_{n-2\ell})\Big(a_1(Y_{n-\ell}+\nu_1) - a_1(Y_{n-\ell})\Big).$$

Figure 5 plots the absolute error of mean for Example 1. The sample size is taken to be 10^6. We may observe that the scheme (26) has second order accuracy for the mean.

For the generalization to high moment ϕ, as we said in the beginning of this section, we have to take a novel approach, which may require using tame Itô formula.

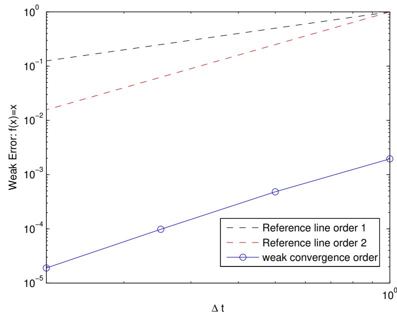


FIG. 5. *Log-log plot of the absolute error with function $f(x) = x$.*

Assume $\phi(\mathbf{x}) = |\mathbf{x}|^2$, and let $Z := X^2$. Following from the Itô formula, we get

(28)

$$Z(t) = Z(0) + \int_0^t \int_0^A \Big[\Big(X(s-) + \nu_1 c_1(a;\ X(s-\tau-))\Big)^2 - \Big(X(s-)\Big)^2\Big]\lambda(\mathrm{d}s \times \mathrm{d}a)$$

$$= Z(0) + \int_0^t \int_0^A \Big[2\nu_1 X(s-) c_1(a;\ X(s-\tau-)) + \nu_1^2 c_1^2(a;\ X(s-\tau-))\Big]\lambda(\mathrm{d}s \times \mathrm{d}a).$$

Consider the equations for X and Z together; then we take the above approach of the generalization to the mean for the augment variable (X, Z) as follows:

$$\begin{aligned}
&\mathbb{E}(Z(t_{n+1}) - Z(t_n)) \\
&= \mathbb{E}\int_{t_n}^{t_{n+1}} \int_0^A \Big[2\nu_1 X(s-) c_1(a;\ X(s-\tau-)) + \nu_1^2 c_1^2(a;\ X(s-\tau-))\Big]\lambda(\mathrm{d}s \times \mathrm{d}a) \\
&= \mathbb{E}\int_{t_n}^{t_{n+1}} \Big[2\nu_1 X(t) a_1(X(t-\tau)) + \nu_1^2 a_1(X(t-\tau))\Big]\mathrm{d}t.
\end{aligned}$$

Applying the tame Itô formula to $X(t)a_1(X(t-\tau))$ and noting that $a_1(x) = cx$ (for the nonlinear case, Taylor expansion may be needed), we get

(29)

$$\begin{aligned}
&X(t)a_1(X(t-\tau)) = X(t_n)a_1(X(t_n-\tau)) \\
&+ \int_{t_n}^t \int_0^A \big(X(s-) + \nu_1 c_1(a; X(s-\tau-))\big) a_1(X(s-\tau)) - X(s-)a_1(X(s-\tau))\lambda(\mathrm{d}s \times \mathrm{d}a) \\
&+ \int_{t_n}^t \int_0^A X(s-)a_1(X(s-\tau) + \nu_1 c_1(a;\ X(s-2\tau-))) - X(s-)a_1(X(s-\tau-))\lambda(\mathrm{d}s \times \mathrm{d}a) \\
&= X(t_n)a_1(X(t_n-\tau)) + \int_{t_n}^t \int_0^A \nu_1 a_1(X(s-\tau)) c_1(a;\ X(s-\tau-))\lambda(\mathrm{d}s \times \mathrm{d}a) \\
&+ \int_{t_n}^t \int_0^A c\nu_1 X(s-) c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s \times \mathrm{d}a)
\end{aligned}$$

and applying the Itô formula to $a_1(X(t-\tau))$ with $a_1(x) = cx$ (for the nonlinear case, the Taylor expansion may be needed), we get

(30)
$$
\begin{aligned}
a_1(X(t-\tau)) &= a_1(X(t_n-\tau)) \\
&+ \int_{t_n}^{t}\int_0^A \Big[a_1(X(s-\tau-)+\nu_1 c_1(a;\ X(s-2\tau-))) - a_1(X(s-\tau-))\Big]\lambda(\mathrm{d}s\times\mathrm{d}a) \\
&= a_1(X(t_n-\tau)) + \int_{t_n}^{t}\int_0^A c\nu_1 c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times\mathrm{d}a).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathbb{E}(Z(t_{n+1})-Z(t_n)) \\
&= 2\nu_1\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[X(t_n)a_1(X(t_n-\tau)) \\
&\quad + \int_{t_n}^{t}\int_0^A \nu_1 a_1(X(s-\tau))c_1(a;\ X(s-\tau-))\lambda(\mathrm{d}s\times\mathrm{d}a) \\
&\quad + \int_{t_n}^{t}\int_0^A c\nu_1 X(s-)c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times\mathrm{d}a)\Big]\mathrm{d}t \\
&+ \nu_1^2\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[a_1(X(t_n-\tau)) + \int_{t_n}^{t}\int_0^A c\nu_1 c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times\mathrm{d}a)\Big]\mathrm{d}t \\
&= 2\nu_1\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[X(t_n)a_1(X(t_n-\tau)) + \int_{t_n}^{t}\nu_1 a_1^2(X(s-\tau))\mathrm{d}s \\
&\quad + \int_{t_n}^{t}\nu_1 X(s)a_1(X(s-2\tau))\mathrm{d}s\Big]\mathrm{d}t \\
&+ \nu_1^2\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[a_1(X(t_n-\tau)) + \int_{t_n}^{t}c\nu_1 a_1(X(s-2\tau))\mathrm{d}s\Big]\mathrm{d}t \\
&= \delta t\mathbb{E}\Big[2\nu_1 X(t_n)a_1(X(t_n-\tau)) + \nu_1^2 a_1(X(t_n-\tau))\Big] \\
&+ \delta t^2\mathbb{E}\Big[\nu_1^2 a_1^2(X(t_n-\tau)) + c\nu_1^2 X(t_n)a_1(X(t_n-2\tau)) + \frac{c\nu_1^3}{2}a_1(X(t_n-2\tau))\Big] + \mathcal{O}(\delta t^3).
\end{aligned}
$$

Consider the D-leaping scheme with a random correction to the equations of X and Z:

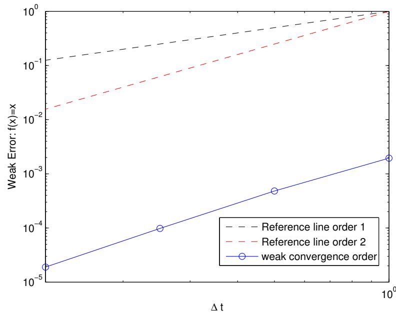
$$
\begin{aligned}
\bar{X}_{n+1} &= \bar{X}_n + \nu_1(r_x+\tilde{r}_x), \\
\bar{Z}_{n+1} &= \bar{Z}_n + \nu_1(r_z+\tilde{r}_z).
\end{aligned} \tag{31}
$$

We require

$$
\begin{aligned}
&\mathbb{E}_\eta\mathbb{E}_{r_x}(\tilde{r}_x) = \frac{\delta t^2}{2}c\nu_1 a_1(\eta(-2\tau)) + \mathcal{O}(\delta t^3), \\
&\mathbb{E}_\eta\mathbb{E}_{r_x}(\tilde{r}_z) = \nu_1\delta t^2 a_1^2(\eta(-\tau)) + c\nu_1\delta t^2\eta(0)a_1(\eta(-2\tau)) + \frac{c\nu_1^2\delta t^2}{2}a_1(\eta(-2\tau)) + \mathcal{O}(\delta t^3)
\end{aligned}
$$

to obtain

$$
\begin{aligned}
&\mathbb{E}\Big(X(t_{n+1};t_n,\eta)-\eta(0)\Big) - \mathbb{E}\Big(\bar{X}(t_{n+1};t_n,\eta)-\eta(0)\Big) = \mathcal{O}(\delta t^3), \\
&\mathbb{E}\Big((X(t_{n+1};t_n,\eta))^2-(\eta(0))^2\Big) - \mathbb{E}\Big(\bar{Z}(t_{n+1};t_n,\eta)-(\eta(0))^2\Big) = \mathcal{O}(\delta t^3).
\end{aligned}
$$



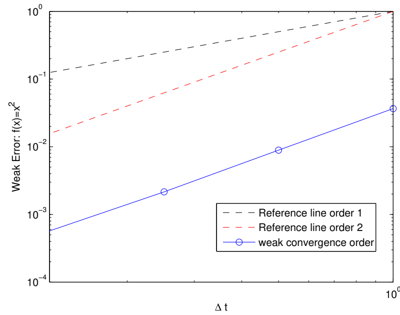


FIG. 6. *Log-log plot of the absolute error with functions $f(x) = x$ and $f(x) = x^2$, respectively.*

In practice, we may take $\tilde{r}_x = \mathrm{sgn}(\alpha)\mathcal{P}(|\alpha|)$, where

$$\alpha = \frac{\delta t^2}{2} c\nu_1 a_1(\bar{X}_{n-2\ell}),$$

and take $\tilde{r}_z = \mathrm{sgn}(\beta)\mathcal{P}(|\beta|)$, where

$$\beta = \nu_1 \delta t^2 a_1^2(\bar{X}_{n-\ell}) + c\nu_1 \delta t^2 \bar{X}_n a_1(\bar{X}_{n-2\ell}) + \frac{c\nu_1^2 \delta t^2}{2} a_1(\bar{X}_{n-2\ell}).$$

Figure 6 plots the absolute errors of first and second moments for Example 1 when applying the generalized scheme (31). The sample size is taken to be 10^6. We may observe that the scheme (31) has second order accuracy for both first and second moments.

Similarly, for the third order moment, we assume $\phi(x) = |x|^3$ and let $Z := X^3$. Following from the Itô formula, we get

$$\begin{aligned} Z(t) &= Z(0) + \int_0^t \int_0^A \Big[\Big(X(s-) + \nu_1 c_1(a;\ X(s-\tau-))\Big)^3 - \Big(X(s-)\Big)^3\Big]\lambda(\mathrm{d}s \times \mathrm{d}a) \\ &= Z(0) + \int_0^t \int_0^A \Big[3\nu_1 X^2(s-)c_1(a;\ X(s-\tau-)) + 3\nu_1^2 X(s-)c_1^2(a;\ X(s-\tau-)) \\ &\quad + \nu_1^3 c_1^3(a;\ X(s-\tau-))\Big]\lambda(\mathrm{d}s \times \mathrm{d}a). \end{aligned}$$

Consider the equations for X and Z together; then we use the approach above to get

$$\begin{aligned} &\mathbb{E}\big(Z(t_{n+1}) - Z(t_n)\big) = \mathbb{E}\int_{t_n}^{t_{n+1}} \int_0^A \Big[3\nu_1 X^2(t-)c_1(a;\ X(t-\tau-)) \\ &\quad + 3\nu_1^2 X(t-)c_1^2(a;\ X(t-\tau-)) + \nu_1^3 c_1^3(a;\ X(t-\tau-))\Big]\lambda(\mathrm{d}t \times \mathrm{d}a) \\ &= \mathbb{E}\int_{t_n}^{t_{n+1}} \Big[3\nu_1 X^2(t)a_1(X(t-\tau)) + 3\nu_1^2 X(t)a_1(X(t-\tau)) + \nu_1^3 a_1(X(t-\tau))\Big]\mathrm{d}t. \end{aligned}$$

Applying the tame Itô formula to $X^2(t)a_1(X(t-\tau))$ and noting that $a_1(x) = cx$ (for the nonlinear case, the Taylor expansion may be needed), we get

$$
\begin{aligned}
&X^2(t)a_1(X(t-\tau)) = X^2(t_n)a_1(X(t_n-\tau))\\
&+\int_{t_n}^{t}\int_0^A \Big[X^2(s)a_1(X(s-\tau-)+\nu_1c_1(a;\ X(s-2\tau-)))\\
&\quad - X^2(s)a_1(X(s-\tau-))\Big]\lambda(\mathrm{d}s\times \mathrm{d}a)\\
&+\int_{t_n}^{t}\int_0^A \Big[\big(X(s-)+\nu_1c_1(a;\ X(s-\tau-))\big)^2a_1(X(s-\tau-))\\
&\quad - X^2(s-)a_1(X(s-\tau-))\Big]\lambda(\mathrm{d}s\times \mathrm{d}a)\\
&= X^2(t_n)a_1(X(t_n-\tau)) + \int_{t_n}^{t}\int_0^A c\nu_1X^2(s)c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times \mathrm{d}a)\\
&+\int_{t_n}^{t}\int_0^A \Big[2\nu_1X(s-)a_1(X(s-\tau-))c_1(a;\ X(s-\tau-))\\
&\quad + \nu_1^2a_1(X(s-\tau-))c_1^2(a;\ X(s-\tau-))\Big]\lambda(\mathrm{d}s\times \mathrm{d}a).
\end{aligned}
$$

Using also (29) and (30), we obtain

$$
\begin{aligned}
&\mathbb{E}(Z(t_{n+1})-Z(t_n))\\
&= 3\nu_1\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[X^2(t_n)a_1(X(t_n-\tau))\\
&+\int_{t_n}^{t}\int_0^A c\nu_1X^2(s)c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times \mathrm{d}a)\\
&+\int_{t_n}^{t}\int_0^A \Big(2\nu_1X(s-)a_1(X(s-\tau-))c_1(a;\ X(s-\tau-))\\
&\quad + \nu_1^2a_1(X(s-\tau-))c_1^2(a;\ X(s-\tau-))\Big)\lambda(\mathrm{d}s\times \mathrm{d}a)\Big]\mathrm{d}t\\
&+3\nu_1^2\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[X(t_n)a_1(X(t_n-\tau))\\
&+\int_{t_n}^{t}\int_0^A \nu_1a_1(X(s-\tau))c_1(a;\ X(s-\tau-))\lambda(\mathrm{d}s\times \mathrm{d}a)\\
&+\int_{t_n}^{t}\int_0^A c\nu_1X(s-)c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times \mathrm{d}a)\Big]\mathrm{d}t\\
&+\nu_1^3\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[a_1(X(t_n-\tau))+\int_{t_n}^{t}\int_0^A c\nu_1c_1(a;\ X(s-2\tau-))\lambda(\mathrm{d}s\times \mathrm{d}a)\Big]\mathrm{d}t\\
&= 2\nu_1\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[X(t_n)a_1(X(t_n-\tau))+\int_{t_n}^{t}\nu_1a_1^2(X(s-\tau))\mathrm{d}s\\
&\quad +\int_{t_n}^{t}\nu_1X(s)a_1(X(s-2\tau))\mathrm{d}s\Big]\mathrm{d}t\\
&+\nu_1^2\mathbb{E}\int_{t_n}^{t_{n+1}}\Big[a_1(X(t_n-\tau))+\int_{t_n}^{t}c\nu_1a_1(X(s-2\tau))\mathrm{d}s\Big]\mathrm{d}t.
\end{aligned}
$$

Simplifying the above equation, we get

$$\begin{aligned}
&\mathbb{E}(Z(t_{n+1}) - Z(t_n))\\
&= \delta t\mathbb{E}\Big[3\nu_1 X^2(t_n)a_1(X(t_n-\tau)) + 3\nu_1^2 X(t_n)a_1(X(t_n-\tau)) + \nu_1^3 a_1(X(t_n-\tau))\Big]\\
&+ \delta t^2\mathbb{E}\Bigg[\frac{3c\nu_1^2}{2}X^2(t_n)a_1(X(t_n-2\tau)) + 3\nu_1^2 X(t_n)a_1^2(X(t_n-\tau)) + \frac{3\nu_1^3}{2}a_1^2(X(t_n-\tau))\\
&+ \frac{3\nu_1^3}{2}a_1^2(X(t_n-\tau)) + \frac{3c\nu_1^3}{2}X(t_n)a_1(X(t_n-2\tau)) + \frac{c\nu_1^4}{2}a_1(X(t_n-2\tau))\Bigg] + \mathcal{O}(\delta t^3).
\end{aligned}$$

Consider the D-leaping scheme with a random correction to the equations of X and Z:

$$\begin{aligned}
\bar{X}_{n+1} &= \bar{X}_n + \nu_1(r_x + \tilde{r}_x),\\
\bar{Z}_{n+1} &= \bar{Z}_n + \nu_1(r_z + \tilde{r}_z).
\end{aligned} \tag{32}$$

We require

$$\begin{aligned}
&\mathbb{E}_\eta\mathbb{E}_{r_x}(\tilde{r}_x) = \frac{\delta t^2}{2}c\nu_1 a_1(\eta(-2\tau)) + \mathcal{O}(\delta t^3),\\
&\mathbb{E}_\eta\mathbb{E}_{r_x}(\tilde{r}_z) = \frac{3c\nu_1\delta t^2}{2}\eta(0)^2 a_1(\eta(-2\tau)) + 3\nu_1\delta t^2\eta(0)a_1^2(\eta(-\tau)) + \frac{3\nu_1^2\delta t^2}{2}a_1^2(\eta(-\tau))\\
&\quad + \frac{3\nu_1^2\delta t^2}{2}a_1^2(\eta(-\tau)) + \frac{3c\nu_1^2\delta t^2}{2}\eta(0)a_1(\eta(-2\tau)) + \frac{c\nu_1^3\delta t^2}{2}a_1(\eta(-2\tau)) + \mathcal{O}(\delta t^3)
\end{aligned}$$

to obtain

$$\begin{aligned}
&\mathbb{E}\Big(X(t_{n+1};t_n,\eta) - \eta(0)\Big) - \mathbb{E}\Big(\bar{X}(t_{n+1};t_n,\eta) - \eta(0)\Big) = \mathcal{O}(\delta t^3),\\
&\mathbb{E}\Big((X(t_{n+1};t_n,\eta))^3 - (\eta(0))^3\Big) - \mathbb{E}\Big(\bar{Z}(t_{n+1};t_n,\eta) - (\eta(0))^3\Big) = \mathcal{O}(\delta t^3).
\end{aligned}$$

In practice, we may take $\tilde{r}_x = \operatorname{sgn}(\alpha)\mathcal{P}(|\alpha|)$, where

$$\alpha = \frac{\delta t^2}{2}c\nu_1 a_1(\bar{X}_{n-2\ell}),$$

and take $\tilde{r}_z = \operatorname{sgn}(\beta)\mathcal{P}(|\beta|)$, where

$$\begin{aligned}
\beta =& \frac{3c\nu_1\delta t^2}{2}\bar{X}_n^2 a_1(\bar{X}_{n-2\ell}) + 3\nu_1\delta t^2\bar{X}_n a_1^2(\bar{X}_{n-\ell}) + \frac{3\nu_1^2\delta t^2}{2}a_1^2(\bar{X}_{n-\ell})\\
&+ \frac{3\nu_1^2\delta t^2}{2}a_1^2(\bar{X}_{n-\ell}) + \frac{3c\nu_1^2\delta t^2}{2}\bar{X}_n a_1(\bar{X}_{n-2\ell}) + \frac{c\nu_1^3\delta t^2}{2}a_1(\bar{X}_{n-2\ell}).
\end{aligned}$$

5. Conclusion. We consider the convergence order in both mean-square strong and weak senses of the D-leaping scheme for chemical reactions within the framework of stochastic delay differential equations (SDDEs) driven by a Poisson random measure. The infinite dimensional tame Itô formula and the Malliavin calculus for the solution process of the SDDEs are established. It is proved that the mean-square convergence order is of 1/2 and the weak convergence order is of 1. Numerical experiments are performed to support the theoretical results. Finally, we propose the construction of highly accurate schemes. We improve the accuracy of the D-leaping scheme for the mean via the addition of a random correction to the primitive D-leaping scheme in each step. For improvement to higher order moments of the solution, and for solving the problems introduced by delay, we first fix the test function and then use the tame Itô formula to obtain the augmented variables. Finally, we add random corrections to the D-leaping scheme in each step. The methodology of the construction of highly accurate schemes for the general system may be a topic of future work. Moreover, the analysis for long time behavior of the chemical reaction system with delay, including the invariant measure and ergodicity, is also an interesting topic.

Appendix A. Proof of Proposition 3.3.

Proof. The idea of the proof is similar to that of Lemma 2.2 in [13] and is obtained directly through the DSSA algorithm. We first define the process

$$N^j(t) = \int_0^t \int_0^A c_j(a;\ \boldsymbol{X}(s-\tau_j-))\lambda(\mathrm{d}s \times \mathrm{d}a), \quad j = 1,2,\dots,M. \tag{33}$$

Define a sequence of processes $\boldsymbol{X}^n(t)$, stopping times T^n, and indices I^n $(n = 0,1,\dots)$ as $\boldsymbol{X}^n(t) = \eta(t)$ for $n \in \mathbb{Z}_0^+$ and $t \in [-\tau, 0]$, $\boldsymbol{X}^0(t) = \eta(0)$ for $t \geq 0$, $T^0 = 0$, $I^0 = 1$, and

$$\begin{aligned}
&\boldsymbol{X}^{n+1}(t) = \eta(0) + \sum_{j=1}^{M} \boldsymbol{\nu}_{I^j} \mathbf{1}_{\{t \geq T^j\}},\\
&T^{n+1} = \min\{\inf\{t :\ N^j(t, \boldsymbol{X}^{n+1}) > N^j(T^n, \boldsymbol{X}^{n+1})\},\ j = 1,2,\dots,M\},\\
&I^{n+1} = \text{Index } j \in \{1,2,\dots,M\} \text{ such that } \Delta N^j(T^{n+1}, \boldsymbol{X}^{n+1}) = 1.
\end{aligned}$$

It is easy to find that under Assumption 2.1, $\boldsymbol{X}^n(t)$ remains in $\Omega_{\boldsymbol{X}_0}$ permanently, $\boldsymbol{X}^n(t) = \boldsymbol{X}^{n-1}(t)$ in $[-\tau, T^{n-1})$, and the stopping time can be extended to ∞. Therefore, we show that the solution of (1) is well-posed.

For the property of Hölder continuous, we may let $t > s \geq 0$ and get

$$\boldsymbol{X}(t) - \boldsymbol{X}(s) = \sum_{j=1}^{M} \int_s^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(u-\tau_j-))\lambda(\mathrm{d}u \times \mathrm{d}a). \tag{34}$$

First, we let $s = 0$ to show that there exists a constant $C := C(\eta, K, L, T, M)$ such that $\mathbb{E}|\boldsymbol{X}(t)|^2 \leq C$. In fact,

$$\begin{aligned}
\mathbb{E}|\boldsymbol{X}(t)|^2 &\lesssim \mathbb{E}|\eta(0)|^2 + \mathbb{E}\Big|\sum_{j=1}^{M} \int_0^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(u-\tau_j-))\lambda(\mathrm{d}u \times \mathrm{d}a)\Big|^2\\
&\lesssim C + \mathbb{E}|\eta(0)|^2 + \sum_{j=1}^{M} \int_0^t \mathbb{E}|\boldsymbol{X}(s-\tau_j)|^2 \mathrm{d}s\\
&\leq C + \mathbb{E}|\eta(0)|^2 + M\|\eta\|^2_{L^2([-\tau,0])} + M\int_0^t \mathbb{E}|\boldsymbol{X}(s)|^2 \mathrm{d}s,
\end{aligned} \tag{35}$$

where we use the following estimation to the stochastic integral with respect to the

Poisson random measure:

$$
\begin{aligned}
&\mathbb{E}\Big|\int_0^t\int_0^A c_j(a;\ \boldsymbol{X}(u-\tau_j-))\lambda(\mathrm{d}u\times\mathrm{d}a)\Big|^2\\
&\lesssim \mathbb{E}\Big|\int_0^t\int_0^A c_j(a;\ \boldsymbol{X}(u-\tau_j-))m(\mathrm{d}u\times\mathrm{d}a)\Big|^2\\
&\qquad+\mathbb{E}\Big|\int_0^t\int_0^A c_j(a;\ \boldsymbol{X}(u-\tau_j-))(\lambda-m)(\mathrm{d}u\times\mathrm{d}a)\Big|^2\\
&=\mathbb{E}\Big|\int_0^t a_j(\boldsymbol{X}(u-\tau_j-))\mathrm{d}u\Big|^2+\mathbb{E}\int_0^t\int_0^A\Big|c_j(a;\ \boldsymbol{X}(u-\tau_j-))\Big|^2 m(\mathrm{d}u\times\mathrm{d}a)\\
&=\mathbb{E}\Big|\int_0^t a_j(\boldsymbol{X}(u-\tau_j-))\mathrm{d}u\Big|^2+\mathbb{E}\int_0^t a_j(\boldsymbol{X}(u-\tau_j-))\mathrm{d}u\\
&\lesssim \mathbb{E}\int_0^t |a_j(\boldsymbol{X}(u-\tau_j))|^2\mathrm{d}u+\mathbb{E}\int_0^t |a_j(\boldsymbol{X}(u-\tau_j))|\mathrm{d}u\\
&\lesssim C+\mathbb{E}\int_0^t|\boldsymbol{X}(u-\tau_j)|^2\mathrm{d}u+\mathbb{E}\int_0^t|\boldsymbol{X}(u-\tau_j)|\mathrm{d}u\le C+\mathbb{E}\int_0^t|\boldsymbol{X}(u-\tau_j)|^2\mathrm{d}u.
\end{aligned}
$$

Here we use the fact that $\boldsymbol{X}(t)\in\Omega_{\boldsymbol{X}_0}$, which means that $|\boldsymbol{X}(t)|\le|\boldsymbol{X}(t)|^2$. The notation $a\lesssim b$ stands for $a\le Cb$, where $C>0$ is a constant.

Applying Gronwall's inequality to (35), we obtain $\mathbb{E}|\boldsymbol{X}(t)|^2\le C$.

Next, we apply $\mathbb{E}|\cdot|^2$ to (34) and estimate similarly:

$$
\begin{aligned}
\mathbb{E}|\boldsymbol{X}(t)-\boldsymbol{X}(s)|^2&\lesssim\mathbb{E}\Big|\sum_{j=1}^M\int_s^t\int_0^A\boldsymbol{\nu}_j c_j(a;\ \boldsymbol{X}(u-\tau_j-))\lambda(\mathrm{d}u\times\mathrm{d}a)\Big|^2\\
&\lesssim(t-s)+\sum_{j=1}^M\mathbb{E}\int_s^t|\boldsymbol{X}(u-\tau_j)|^2\mathrm{d}u\\
&\lesssim(t-s).
\end{aligned}
$$

Thus we finish the proof. □

Appendix B. The proof of Lemma 3.8.

Proof. Suppose $i=0$ and let $t\in[0,T]$. Consider the following case.

Case 1. $0\le t\le t_1$.

$$
\begin{aligned}
\boldsymbol{Y}(t;0,\eta)&=\eta(0)+\sum_{j=1}^M\int_0^t\int_0^A\boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}(\xi(s)-\tau_j;0,\eta))\lambda(\mathrm{dt}\times\mathrm{da})\\
&=\eta(0)+\sum_{j=1}^M\int_0^t\int_0^A\boldsymbol{\nu}_j c_j(a;\ \eta(-\tau_j))\lambda(\mathrm{dt}\times\mathrm{da})\\
&=\boldsymbol{F}_1(t,\omega,\eta(-\tau_j)).
\end{aligned}
$$

Case 2. $t_1\le t\le t_2$.

$$
\begin{aligned}
\boldsymbol{Y}(t;0,\eta)&=\boldsymbol{Y}(t_1)+\sum_{j=1}^M\int_{t_1}^t\int_0^A\boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}(t_1-\tau_j))\lambda(\mathrm{dt}\times\mathrm{da})\\
&=\boldsymbol{F}_2(t,\omega,\eta(-\tau_j),\eta(t_1-\tau_j)).
\end{aligned}
$$

Case 3. $t_2 \le t \le t_3$.

$$\begin{aligned}\boldsymbol{Y}(t;0,\eta) =& \boldsymbol{Y}(t_2) + \sum_{j=1}^{M}\int_{t_1}^{t}\int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}(t_2-\tau_j))\lambda(\mathrm{dt}\times\mathrm{da})\\ =& \boldsymbol{F}_3(t,\omega,\eta(-\tau_j),\eta(t_1-\tau_j),\eta(t_2-\tau_j)).\end{aligned}$$

Case k. $t_{k-1} \le t \le t_k$. By induction, there are fixed number $\mu_1,\dots,\mu_\ell \in [-\tau,0]$ such that

$$\boldsymbol{Y}(t;0,\eta) = \boldsymbol{F}_k(t,\omega,\eta(\mu_1),\dots,\eta(\mu_\ell)),$$

which is a tame function of η.

To complete the proof of the lemma, we take

$$\boldsymbol{F}(t,\omega,\Pi(\eta)) := \sum_{i=1}^{N_T-1} \mathbf{1}_{[t_i,t_{i+1})}(t)\boldsymbol{F}_{i+1}(t,\omega,\Pi(\eta)). \qquad \square$$

Appendix C. Proof of Proposition 3.10.

Proof. We know that

$$\begin{aligned}&\phi(\Pi(X_t)) - \phi(\Pi(X_0))\\ &= \sum_{0\le s\le t}\Big[\phi(\Pi(X_s)) - \phi(\Pi(X_{s-}))\Big]\\ &= \sum_{0\le s\le t}\Big[\phi(X(s+\mu_1),\dots,X(s+\mu_k)) - \phi(X(s+\mu_1-,\dots,X(s+\mu_k-)))\Big]\\ &= \sum_{0\le s\le t}\sum_{i=1}^{k}\Big[\phi(X(s+\mu_1-),\dots,X(s+\mu_{i-1}-),X(s+\mu_i),\dots,X(s+\mu_k))\\ &\quad - \phi(X(s+\mu_1-),\dots,X(s+\mu_{i-1}-),X(s+\mu_i-),\dots,X(s+\mu_k))\Big]\\ &= \sum_{i=1}^{k}\sum_{0\le s\le t}\Big[\phi(X(s+\mu_1-),\dots,X(s+\mu_{i-1}-),X(s+\mu_i),\dots,X(s+\mu_k))\\ &\quad - \phi(X(s+\mu_1-),\dots,X(s+\mu_{i-1}-),X(s+\mu_i-),\dots,X(s+\mu_k))\Big]\\ &=: \sum_{i=1}^{k}\mathcal{B}_i(t).\end{aligned} \tag{36}$$

Let T_n, $n \le 0$ denote the jump time for $\eta(t)$, $-\tau \le t \le 0$. Define $q(t) = \int_0^A a\lambda(\mathrm{d}t\times\mathrm{d}a)$, and let T_n, $n > 0$, denote the jump time for q, which is defined recursively by $T_n =$

$\inf\{t > T_{n-1} :\ \Delta q(t) \in (0, A]\}$. Hence

$$\begin{aligned}\mathcal{B}_i(t) = \sum_{n\geq 1} \Big[&\phi(X((t+\mu_1)\wedge T_n-),\dots,X((t+\mu_{i-1})\wedge T_n-),\\ &X((t+\mu_i)\wedge T_n),\dots,X((t+\mu_k)\wedge T_n))\\ &-\phi(X((t+\mu_1)\wedge T_n-),\dots,X((t+\mu_{i-1})\wedge T_n-),\\ &X((t+\mu_i)\wedge T_n-),\dots,X((t+\mu_k)\wedge T_n))\Big]\\ = \sum_{n\geq 1}\Big[&\phi(\dots,X((t+\mu_i)\wedge T_n-)+K((t+\mu_i)\wedge T_n,\Delta q((t+\mu_i)\wedge T_n)),\dots)\\ &-\phi(\dots,X((t+\mu_i)\wedge T_n-),\dots)\Big]\\ = \int_0^t\int_0^A \Big[&\phi(X_{s-}(\mu_1),\dots,X_{s-}(\mu_{i-1}),X_{s-}(\mu_i)+K(s+\mu_i,a),X_s(\mu_{i+1}),\dots,X_s(\mu_k))\\ &-\phi(X_{s-}(\mu_1),\dots,X_{s-}(\mu_{i-1}),X_{s-}(\mu_i),X_s(\mu_{i+1}),\dots,X_s(\mu_k))\Big]\lambda(\mathrm{ds}\times\mathrm{da}).\end{aligned}$$

Thus we finish the proof. ∎

Appendix D. Proof of Proposition 3.11.

Proof. The proof of $X(t) \in \mathbb{D}^{1,2}$ is similar to that of (18), but the idea is applied to the Picard approximations to (17); see, for instance, [16, Theorem 17.2]. So here we only focus on the proof of (18).

Taking the Malliavin derivative to (17), for $t > \sigma$ we have

$$D_{r,z}X(t) = D_{r,z}\eta(0) + \sum_{j=1}^{M} D_{r,z}\int_\sigma^t\int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau_j-))\lambda(\mathrm{d}s\times\mathrm{d}a), \tag{37}$$

where by Proposition 3.7,

(38)

$$\begin{aligned}&D_{r,z}\int_\sigma^t\int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau_j-))\lambda(\mathrm{d}s\times\mathrm{d}a)\\ &= D_{r,z}\int_\sigma^t \boldsymbol{\nu}_j a_j(X(s-\tau_j-))\mathrm{d}s + D_{r,z}\int_\sigma^t\int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau_j-))(\lambda-m)(\mathrm{d}s\times\mathrm{d}a)\\ &= \boldsymbol{\nu}_j\int_\sigma^t \Big[a_j(X(s-\tau_j-)+D_{r,z}X(s-\tau_j-)) - a_j(X(s-\tau_j-))\Big]\mathrm{d}s\\ &+ \boldsymbol{\nu}_j c_j(z;\ X(r-\tau_j-)) + \boldsymbol{\nu}_j\int_\sigma^t\int_0^A \Big[c_j(a;\ X(s-\tau_j-)+D_{r,z}X(s-\tau_j-))\\ &\qquad - c_j(a;\ X(s-\tau_j-))\Big](\lambda-m)(\mathrm{d}s\times\mathrm{d}a).\end{aligned}$$

For $\sigma-\tau \le t \le \sigma$, we have

$$D_{r,z}X(t) = D_{r,z}\eta(t-\sigma).$$

Take $\mathbb{E}|\cdot|^2$ with respect to (37)–(38) to get

(39)

$$\mathbb{E}|D_{r,z}X(t)|^2 \lesssim \mathbb{E}|D_{r,z}\eta(0)|^2 + \mathbb{E}\Big|\sum_{j=1}^{M}\boldsymbol{\nu}_j c_j(z;\ X(r-\tau_j-))\Big|^2 + \sum_{j=1}^{M}\mathcal{C}_{1,j} + \sum_{j=1}^{M}\mathcal{C}_{2,j},$$

with

$$\mathcal{C}_{1,j} = \mathbb{E}\bigg|\int_\sigma^t \Big[a_j(X(s-\tau_j-)+D_{r,z}X(s-\tau_j-)) - a_j(X(s-\tau_j-))\Big]\mathrm{d}s\bigg|^2,$$

$$\begin{aligned}\mathcal{C}_{2,j} = \mathbb{E}\bigg|\int_\sigma^t\int_0^A &\Big[c_j(a;\ X(s-\tau_j-)+D_{r,z}X(s-\tau_j-)) \\ &- c_j(a;\ X(s-\tau_j-))\Big](\lambda-m)(\mathrm{d}s\times\mathrm{d}a)\bigg|^2.\end{aligned}$$

We estimate each term separately. For term $\mathcal{C}_{1,j}$, we have

$$\begin{aligned}\mathcal{C}_{1,j} &\lesssim (T-\sigma)\mathbb{E}\int_\sigma^t \Big|a_j(X(s-\tau_j-)+D_{r,z}X(s-\tau_j-)) - a_j(X(s-\tau_j-))\Big|^2\mathrm{d}s \\ &\lesssim \mathbb{E}\int_\sigma^t |D_{r,z}X(s-\tau_j-)|^2\mathrm{d}s.\end{aligned}$$

For term $\mathcal{C}_{2,j}$, we have

$$\begin{aligned}&\mathbb{E}\int_\sigma^t\int_0^A |c_j(a;\ X(s-\tau_j-)+D_{r,z}X(s-\tau_j-)) - c_j(a;\ X(s-\tau_j-))|^2 m(\mathrm{d}s\times\mathrm{d}a) \\ &\lesssim \mathbb{E}\int_\sigma^t \Big[|h_{j-1}(X(s-)+D_{r,z}X(s-)) - h_{j-1}(X(s-))| \\ &\quad + |h_j(X(s-)+D_{r,z}X(s-)) - h_j(X(s-))|\Big]\mathrm{d}s \\ &\lesssim \max_{1\le j\le M}\int_\sigma^t \mathbb{E}|D_{r,z}X(s-\tau_j-)|\mathrm{d}s \\ &\lesssim (T-\sigma) + \max_{1\le j\le M}\int_\sigma^t \mathbb{E}|D_{r,z}X(s-\tau_j-)|^2\mathrm{d}s.\end{aligned}$$

Summarizing the above estimates, we obtain

$$\mathbb{E}|D_{r,z}X(t)|^2 \le C(1+\mathbb{E}|D_{r,z}\eta(0)|^2) + C\max_{1\le j\le M}\mathbb{E}\int_\sigma^t |D_{r,z}X(s-\tau_j-)|^2\mathrm{d}s.$$

Integrating with respect to dz, we have

$$\begin{aligned}\mathbb{E}\int_0^A |D_{r,z}X(t)|^2\mathrm{d}z \le& C\bigg(1+\mathbb{E}\int_0^A |D_{r,z}\eta(0)|^2\mathrm{d}z\bigg) \\ &+ C\max_{1\le j\le M}\mathbb{E}\int_\sigma^t\int_0^A |D_{r,z}X(s-\tau_j-)|^2\mathrm{d}s\mathrm{d}z,\end{aligned} \tag{40}$$

which leads to

$$\mathbb{E}\int_0^A |D_{r,z}X(t)|^2\mathrm{d}z \le C\bigg(1+\sup_{\sigma-\tau\le r\le\sigma}\mathbb{E}\int_0^A \|D_{r,z}\eta\|_\infty^2\mathrm{d}z\bigg)e^{C(T-\sigma)}.$$

Thus we finish the proof. ∎

REFERENCES

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge University Press, 2009.

[2] D. Anderson, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), 214107.

[3] M. Barrio, K. Burrage, A. Leier, and T. Tian, *Oscillatory regulation of Hes1: Discrete stochastic delay modelling and simulation*, PLoS Comput. Biol., 2 (2006), e117, https://doi.org/10.1371/journal.pcbi.0020117.

[4] B. Bayati, P. Chatelain, and P. Koumoutsakos, *D-leaping: Accelerating stochastic simulation algorithms for reactions with delays*, J. Comput. Phys., 228 (2009), pp. 5908–5916.

[5] D. Bratsun, D. Volfson, L. Tsimring, and J. Hasty, *Delay-induced stochastic oscillations in gene regulation*, PNAS, 102 (2005), pp. 14593–14598.

[6] E. Buckwar, R. Kuske, S. Mohammed, and T. Shardlow, *Weak convergence of the Euler scheme for stochastic differential delay equations*, LMS J. Comput. Math., 11 (2008), pp. 60–99.

[7] X. Cai, *Exact stochastic simulation of coupled chemical reactions with delays*, J. Chem. Phys., 126 (2007), 124108.

[8] C. Gupta, J. López, R. Azencott, M. Bennett, K. Josić, and W. Ott, *Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations*, J. Chem. Phys., 140 (2014), 204108 https://doi.org/10.1063/1.4878662.

[9] C. Gupta, J. López, W. Ott, K. Josić, and M. Bennett, *Transcriptional delay stabilizes bistable gene networks*, Phys. Rev. Lett., 111 (2013), 058104.

[10] Y. Hu and T. Li, *Highly accurate tau-leaping methods with random corrections*, J. Chem. Phys., 130 (2009), 124109 https://doi.org/10.1063/1.3091269.

[11] Y. Hu, S. Mohammed, and F. Yan, *Discrete-time approximations of stochastic delay equations: Milstein scheme*, Ann. Probab., 32 (2004), pp. 265–314.

[12] A. Leier, T. Marquez-Lago, and K. Burrage, *Generalized binomial τ-leap method for biochemical kinetics incorporating both delay and intrinsic noise*, J. Chem. Phys., 128 (2008), 205107, https://doi.org/10.1063/1.2919124.

[13] T. Li, *Analysis of explicit tau-leaping schemes for simulating chemically reacting systems*, Multiscale Model. Simul., 6 (2007), pp. 417–436, https://doi.org/10.1137/06066792X.

[14] W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, *Delay-induced degrade-and-rire oscillations in small genetic circuits*, Phys. Rev. Lett., 102 (2009), 068105.

[15] S. Mohammed, *Stochastic differential systems with memory: Theory, examples and applications*, in Stochastic Analysis, L. Decreusefond, J. Gjerde, B. Øksendal, and A. Ustunel, eds., Progr. Probab. 42, Birkhäuser, 1998, pp. 1–77.

[16] G. Nunno, B. Øksendal, and F. Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance*, Springer-Verlag, 2009.

[17] E. Platen and N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Springer-Verlag, 2010.

[18] P. Smolen, D. A. Baxter, and J. H. Byrne, *A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator*, Biophys. J., 83 (2002), 2349.

[19] T. Tian, K. Burrage, P. Burrage, and M. Carletti, *Stochastic delay differential equations for genetic regulatory networks*, J. Comput. Appl. Math., 205 (2007), pp. 696–707.

[20] V. Thanh, C. Priami, and R. Zunino, *Efficient rejection-based simulation of biochemical reactions with stochastic noise and delays*, J. Chem. Phys., 141 (2014), 134116, https://doi.org/10.1063/1.4896985.