SIAM J. NUMER. ANAL.
Vol. 57, No. 1, pp. 70–95

ERGODIC APPROXIMATION TO CHEMICAL REACTION SYSTEM WITH DELAY*

CHUCHU CHEN† AND DI LIU‡

Abstract. In order to inherit numerically the ergodicity of the chemical reaction system with delays, we propose and study an Euler-type numerical method from the point of view of stochastic delay differential equations. We not only prove the unique exponential ergodicity of the numerical solution of the approximation, but also present error estimation on invariant measures, which gives order 1 under certain hypotheses. Numerical experiments are provided to illustrate the results.

Key words. stochastic delay differential equation, invariant measure, ergodicity, weak convergence order, Malliavin calculus, Poisson random measure

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

DOI. 10.1137/17M1149894

1. Introduction.

Many chemical dynamics (processes such as transcription and translation in a genetic regulatory network) are not instantaneous and may have considerable delays associated with them. For example, there is an average delay of 10–20 minutes between the action of a transcription factor on the promoter of a gene and the appearance of the corresponding mature mRNA in the cytoplasm [19]. Taking the delays into account is crucial for the description of transient processes. It is well known that in some delay-sensitive cases, neglecting delays in simulation will lead to erroneous conclusions, since delays can alter qualitatively the dynamical behavior; for example, delay can induce oscillations [2, 17]. 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 and analysis of the delayed chemical reactions is necessary. An effective method is considered via modeling the phenomena by a stochastic dynamical system whose evolution in time is governed by random forces as well as the intrinsic dependence on the state over its history, i.e., stochastic delay differential equations (SDDEs) driven by Poisson random measure [14, 7]. Though the solution of SDDEs is not a Markov process, the segment process of the solution is shown to possess the Markov property [18]. Moreover, under certain conditions, the process is proved to possess the invariant measure [11, 22], and even exponential ergodicity [1, 10, 12, 21].

*Received by the editors September 29, 2017; accepted for publication (in revised form) October 24, 2018; published electronically January 15, 2019.
http://www.siam.org/journals/sinum/57-1/M114989.html
Funding: This work was supported by the NSF-DMS 1418959, NSF-DMS 1211292, NSF-DMS 1720002 and a special grant from the Michigan Center for Industrial and Applied Mathematics (MCIAM).
†Department of Mathematics, Michigan State University, East Lansing, MI 48824, and LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (chenc118@msu.edu, chenchuchu@lsec.cc.ac.cn).
‡Department of Mathematics, Michigan State University, East Lansing, MI 48824 (richardl@math.msu.edu).

In the numerical aspect, we expect to have the accelerated, approximate algorithms not only to compute the solution efficiently and effectively, but also to inherit the properties of the original system. An explicit Euler-type numerical method is considered in this paper, which is called a D-leaping scheme for the approximate simulation of biochemical systems with delays [3, 7]. It is shown that the segments of the solution of the continuous time version of this numerical method not only possess the Markov property, but also exponentially converge to the unique invariant measure, which means that the approximated process is also ergodic. We refer readers to [23, 15] and references therein for the construction and analysis of ergodic numerical methods for ergodic stochastic differential equations without delays.

Since the segment processes of the solutions of SDDEs and approximation possess the unique invariant measures, our other interest is to investigate the error between invariant measures induced by the exact and approximated segment processes, which is obtained via the independence of time for the weak error. In order to estimate the weak error, we take a similar approach as in [7] to utilize the Markov property of the segment processes to rewrite the error as the summation of weak local error. The mathematical analysis of local error term is technical in two ways. First, since delays break the Markov property of the system, by contrast with the nondelay case stochastic ordinary differential equations (SODEs), 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 is also not appliable. In order to derive the time-independent weak convergence order of the scheme, we first establish the boundedness of the segment processes of the exact and approximated solutions, and the Malliavin derivatives such that the bounds are independent of time. And then 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 tame Itô formula. Moreover, the Malliavin calculus and anticipating stochastic analysis techniques are employed to show that

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

where the constant C is independent of time. Here $\boldsymbol{X}(t)$ is the exact solution process of the chemical system and $\boldsymbol{Y}(t)$ is the approximated solution generated from the D-leaping scheme, and δt is the maximal time stepsize. Based on the result of the time-independent weak error analysis and ergodicity of the exact and approximated segment processes, we show that the error between invariant measures is of order 1.

The rest of this paper is organized as follows. In section 2, the main results of this paper are introduced. In sections 3 and 4, we give the proofs for the main results. Section 3 is for the proof of existence and uniqueness of invariant measure, and the exponential ergodicity, while section 4 is for the proof of weak error analysis and the estimate between invariant measures. Numerical experiments are performed to support theoretical results in section 5.

Last, we define the following notations in order to describe our set-up.

1. $\mathbb{Z}_0^+ = \mathbb{N} \cup \{0\}$ 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 $(\mathbb{Z}_0^+)^N$.

2. Let $\mathbb{R}^n$ be n-dimensional Euclidean space with Euclidean norm $|x|$ for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ so that $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$, and the inner product in $\mathbb{R}^n$ is denoted by $\langle x, y\rangle$, where $x, y \in \mathbb{R}^n$, so that $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$.
3. $\mathcal{D} := D([-\tau, 0], \mathbb{R}^n)$ 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 $\eta \in \mathcal{D}$. The space $\mathcal{D}$ is complete but not separable under the metric $\|\cdot\|_\infty$. In order to make $\mathcal{D}$ not only complete but also separable, we introduce the Skorohod metric (see [4, section 12]): let Λ denote the class of strictly increasing continuous mappings, and set
$$\|\lambda\|^o := \sup_{-\tau\le s<t\le 0} \Big| \log \frac{\lambda(t) - \lambda(s)}{t - s} \Big| \qquad \forall \lambda \in \Lambda.$$
For any $x, y \in \mathcal{D}$, the Skorohod metric $d^o(x, y)$ on $\mathcal{D}$ is defined by
$$d^o(x, y) := \inf_{\lambda\in\Lambda} \big\{\|\lambda\|^o \vee \|x - y \circ \lambda\|_\infty\big\},$$
for which we have $d^o(x, y) \le \|x - y\|_\infty$, $x, y \in \mathcal{D}$.
4. The notation $\mathcal{P}(\mathcal{D})$ denotes the collection of all probability measures on $(\mathcal{D},\ \mathcal{B}(\mathcal{D}))$, $\mathcal{B}_b(\mathcal{D})$ means the set of all bounded measurable function $F : \mathcal{D} \to \mathbb{R}$ endowed with the uniform norm $\|F\|_0 := \sup_{\phi\in\mathcal{D}} |F(\phi)|$, and $Lip(\mathcal{D})$ is the family of all Lipschitz continuous $\mathbb{R}$-valued functions defined on $\mathcal{D}$.
5. Let $\boldsymbol{X}^{t_1,\eta}(t_2)$ $(t_1 \le t_2)$ be the solution process, starting with initial data $\eta \in \mathcal{D}$ at time $t = t_1$. If $t_1 = 0$, it is often written as $\boldsymbol{X}^\eta(t_2)$. Sometimes the superscript is omitted, if there's no confusion.
6. Throughout this paper, the notation C denotes the time-independent constant, which may be different from line to line.

2. Main results. Let us consider a well-stirred system of N molecular species $\{S_1, S_2, \ldots, S_N\}$ interacting through M chemical reaction channels $\{R_1, R_2, \ldots, R_M\}$. The state of the system is described by the vector
$$\boldsymbol{X}(t) = \Big(X^1(t), X^2(t), \ldots, X^N(t)\Big),$$
where $X^j(t)$ is the number of S_j molecule at time t. The dynamics of reaction R_j are defined by a state change vector $\boldsymbol{\nu}_j = \big(\nu_j^1, \nu_j^2, \ldots, \nu_j^N\big)$, where ν_j^n gives the changes in the S_n molecular population produced by R_j reaction, and a propensity function $a_j(\boldsymbol{x})$ with $a_j(\mathbf{x}) \ge 0$ for physical states, and $a_j(\boldsymbol{x})\mathrm{d}t$ is the probability that the system will experience an R_j reaction in the next infinitesimal time interval $[t,\ t+\mathrm{d}t)$ given $\boldsymbol{X}(t) = \boldsymbol{x}$. Consider the case that delays are involved; we suppose that a subset of, or all, reaction channels $\{R_1, \ldots, R_M\}$ incur a delay. If we denote this set of channels as I_d, then a reaction $R_j \in I_d$ will finish with a delay of τ_j, after it is initiated. The set I_{nd} consists of all the channels without delay, i.e., $I_{nd} \cap I_d = \emptyset$ and $\{R_1, R_2, \ldots, R_M\} = I_{nd} \cup I_d$.

Since $\boldsymbol{X}(t)$ denotes the numbers of molecules, it should be nonnegative integers; we let $\Omega_{\boldsymbol{X}_0}$ be the set of all the possible physical states generated from some initial state $\boldsymbol{X}_0 \in (\mathbb{Z}_0^+)^N$ at time $t_0 = 0$,
$$\Omega_{\boldsymbol{X}_0} = \left\{ \boldsymbol{X} \mid \boldsymbol{X} \in (\mathbb{Z}_0^+)^N,\ \boldsymbol{X} = \boldsymbol{X}_0 + \sum_{j=1}^M k_j \boldsymbol{\nu}_j, \quad k_j \in \mathbb{Z}_0^+ \right\}.$$

Thus the assumption that the propensity function $a_j(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \Omega_{\boldsymbol{X}_0}$ is natural. Moreover, the number of the molecules could not be arbitrary large in realistic chemical reactions; it is also reasonable to let $\boldsymbol{X}(t)$ be in a bounded lattice. By modifying $a_j(\boldsymbol{x})$ to be zero at negative integers, we can see that a_j is Lipschitz. Denote A the upper bound of total propensity: $A = \max\{a_0(\boldsymbol{x}),\ \boldsymbol{x} \in \Omega_{\boldsymbol{X}_0}\}$, where $a_0(\boldsymbol{x}) = \sum_{j=1}^{M} a_j(\boldsymbol{x})$.

From [14], we notice that the state process $\boldsymbol{X}(t)$ above could be formulated as the form of stochastic differential equation (SDE) with delay, or SDDE driven by Poisson random measure. In order to unify the equation, we set 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 driven by Poisson random measure with initial data $\eta \in \mathcal{D}$:

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

where the characteristic 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 $a_j(\boldsymbol{X}(s-\tau_j-))$. Here $\lambda(\mathrm{d}t \times \mathrm{d}a)$ is a Poisson random measure associated with a Poisson point process $(p(t),\ t \geq 0)$ taking values in $[0, A]$ with Lebesgue intensity measure $m(\mathrm{d}t \times \mathrm{d}a) = \mathrm{d}t \times \mathrm{d}a$ on the probability space $(\Omega, \mathcal{F}, P)$, i.e., $\lambda([0,\ t) \times \mathcal{B}) = \#\{0 \leq s < t; p(s) \in \mathcal{B}\}$ for each $t \geq 0$, Borel set $\mathcal{B}$ in $[0,\ A]$. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be the filtration generated by the values of the compensated Poisson random measure $\widetilde{\lambda}(\mathrm{d}t \times \mathrm{d}a) := (\lambda - m)(\mathrm{d}t \times \mathrm{d}a)$. The mean and variance of Poisson integration for a stochastic $\{\mathcal{F}_t\}_{t\geq 0}$-adapted process $\theta(t, z)$, $t \geq 0$, $z \in [0,\ A]$ are frequently used in the following analysis:

$$\mathbb{E}\int_0^T \int_0^A \theta(s,z)\lambda(\mathrm{d}s \times \mathrm{d}z) = \mathbb{E}\int_0^T \int_0^A \theta(s,z)m(\mathrm{d}s \times \mathrm{d}z),$$
$$\mathbb{E}\left(\int_0^T \int_0^A \theta(s,z)\widetilde{\lambda}(\mathrm{d}s \times \mathrm{d}z)\right)^2 = \mathbb{E}\int_0^T \int_0^A \theta^2(s,z)m(\mathrm{d}s \times \mathrm{d}z).$$

Note that for a Lévy process $X(t)$, its jump $\Delta X(t) = X(t) - X(t-)$ equals to zero a.s. for fixed $t > 0$. Moreover, for any continuous function $b(x)$ and two positive reals $d > c$, it holds that $\int_c^d \Delta a(X(t))\mathrm{d}t = 0$. We refer to [20, Chapter 9] and [14] for the proofs and further properties.

We make the following hypothesis to consider the invariant measure and ergodicity for system (1). Since (1) may be rewritten as

$$d\mathbf{X}(t) = \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\mathbf{X}(t-\tau_j-))\mathrm{d}t + \sum_{j=1}^{M} \int_0^A \boldsymbol{\nu}_j c_j(a;\ \mathbf{X}(s-\tau_j-))\widetilde{\lambda}(\mathrm{d}s \times \mathrm{d}a),$$

the following monotone-type condition is given on the drift and diffusion coefficients.

ASSUMPTION 1. *For any $\phi, \psi \in \mathcal{D}$, there exists $\alpha_1 > \alpha_2 > 0$, such that*

$$2\left\langle \phi(0) - \psi(0),\ \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\phi(-\tau_j)) - \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\psi(-\tau_j)) \right\rangle + \int_0^A \left| \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(a;\ \phi(-\tau_j)) - \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(a;\ \psi(-\tau_j)) \right|^2 \mathrm{d}a \leq -\alpha_1 |\phi(0) - \psi(0)|^2 + \alpha_2 \max_{1\leq j\leq M} |\phi(-\tau_j) - \psi(-\tau_j)|^2.$$

For any N-dimensional stochastic process $\boldsymbol{X} :\ [-\tau, \infty) \times \Omega \to \mathbb{R}^N$, define the segment process $\left\{\boldsymbol{X}_t :\ [-\tau, 0] \times \Omega \to \mathbb{R}^N\right\}_{t\in[0,\infty)}$, by

$$\boldsymbol{X}_t(u) = \boldsymbol{X}(t + u) \quad \forall\ u \in [-\tau, 0], \tag{2}$$

which is also called a past (or memory) of the process $\mathbf{X}$ at the moment t.

Due to the existence of delays, the solution process of (1) is not even a Markov process. However, we know from [1] that the segment process of the solution of (1) possesses the Markov property, and further under the Assumption 1 it has a unique invariant measure $\pi(\cdot) \in \mathcal{P}(\mathcal{D})$, which is exponentially ergodic, i.e.,

$$|P_t F(\xi) - \pi(F)| \leq C(\xi) e^{-\kappa t} \|F\|_{Lip}, \quad t \geq \tau, \xi \in \mathcal{D}, F \in Lip(\mathcal{D}). \tag{3}$$

Here $\pi(F) = \int_{\mathcal{D}} F(\xi)\pi(\mathrm{d}\xi)$, and the Markov transition semigroup P_t for segment process $\boldsymbol{X}_t$ can be given by $P_t F(\xi) := \mathbb{E}F(\boldsymbol{X}_t^{\xi})$, where $\boldsymbol{X}_t^{\xi}$ is the segment process of the solution for (1) with initial data $\xi \in \mathcal{D}$.

For the numerical approximation of system (1), we consider the D-leaping method which is proposed in [3] and is rewritten as the following continuous time version in [7]:

$$\boldsymbol{Y}(t) = \begin{cases} \eta(0) + \sum_{j=1}^{M} \int_0^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y} \circ \zeta(s - \tau_j)) \lambda(\mathrm{d}s \times \mathrm{d}a), & t > 0, \\ \eta(t), \quad -\tau \leq t \leq 0,\ \tau = \max\{\tau_j, j \in I_d\}, \end{cases} \tag{4}$$

where $\zeta(t) = t_n$ if $t \in [t_n,\ t_{n+1})$. Define time stepsize $\delta t_n := t_{n+1} - t_n$ and $\delta t := \max_n\{\delta t_n\}$. We refer readers to [7] for the strong and weak convergence order of scheme (4) in a finite time interval $[-\tau, T]$, i.e.,

$$\left(\mathbb{E}|\boldsymbol{X}(t_n) - \boldsymbol{Y}(t_n)|^2\right)^{1/2} \leq C\delta t^{1/2}, \qquad |\mathbb{E}\phi(\boldsymbol{X}(t_n)) - \mathbb{E}\phi(\boldsymbol{Y}(t_n))| \leq C\delta t.$$

Note that the above constants C depend on the final time T. However, under Assumption 1, we could show that the segment process of scheme (4) has a unique invariant measure $\pi^N(\cdot) \in \mathcal{P}(\mathcal{D})$, which is exponentially ergodic. It could make the constant C get rid of the final time T.

THEOREM 2.1. *Under Assumption* 1, *numerical method* (4) *has a unique invariant measure $\pi^N(\cdot) \in \mathcal{P}(\mathcal{D})$, which is exponentially ergodic, i.e.,*

$$|P_t^N F(\xi) - \pi^N(F)| \leq Ce^{-\kappa t} \quad \forall\ t \geq \tau, \xi \in \mathcal{D}, F \in Lip(\mathcal{D}), \tag{5}$$

where the exponent $\kappa > 0$, and the constant $C := C(\xi, F)$.

The proof of Theorem 2.1 is postponed to section 3.

Moreover, we could show that the error between invariant measure π of (1) and invariant measure π^N of (4) is of order 1.

THEOREM 2.2. *There exists constant C independent of time, such that*

$$|\pi(F) - \pi^N(F)| \le C\delta t \quad \forall\ F \in C_b^2(\mathcal{D}). \tag{6}$$

The proof of Theorem 2.2 is postponed to section 4.

3. Exponential ergodicity of Euler-type scheme. In this part, we will investigate some properties of the solution of scheme (4), such as the Markov property of the segment process, the existence and uniqueness of the invariant measure, and the exponential ergodicity under Assumption 1. We refer to [22] for the Markov and eventually Feller properties for the solution of delay differential equations driven by the Lévy process and to [1] for the ergodic properties for segment processes associated with several classes of retarded SDEs with different types of delays.

The following lemma deals with the Markov property.

LEMMA 3.1. *The segment process $\{\mathbf{Y}_t^\xi(\cdot) : t \ge 0,\ \xi \in \mathcal{D}\}$ describes a Markov process on $\mathcal{D}$ with transition probabilities $p(t_1, \xi; t_2, \cdot)$ given by the following: for any $t_1 \le t_2$,*

$$p(t_1, \xi; t_2, B) = P(\omega \in \Omega :\ \mathbf{Y}_{t_2}^{t_1,\xi}(\omega) \in B). \tag{7}$$

Proof. It is equivalent to prove that the Markov property holds: for all $\xi \in L^2(\Omega, \mathcal{D}; \mathcal{F}_0)$,

$$P(\mathbf{Y}_{t_2}^\xi \in B|\mathcal{F}_{t_1}) = P(\mathbf{Y}_{t_2}^\xi \in B|\mathbf{Y}_{t_1}^\xi).$$

First, we prove

$$P(\mathbf{Y}_{t_2}^\xi \in B|\mathcal{F}_{t_1}) = p(t_1, \mathbf{Y}_{t_1}^\xi(\cdot); t_2, B), \tag{8}$$

which means for a.a. $\omega' \in \Omega$,

$$\Big(P(\omega :\ \mathbf{Y}_{t_2}^\xi(\omega) \in B|\mathcal{F}_{t_1})\Big)(\omega') = p(t_1, \mathbf{Y}_{t_1}^\xi(\omega'); t_2, B). \tag{9}$$

By the definition of conditional probability, we see that (9) is equivalent to

$$\int_D \mathbf{1}_B(\mathbf{Y}_{t_2}^\xi(\omega))dP(\omega) = \int_D \int_\Omega \mathbf{1}_B\{\mathbf{Y}_{t_2}^{t_1,\mathbf{Y}_{t_1}^\xi(\omega')}(\omega)\}dP(\omega)dP(\omega')$$

for all $D \in \mathcal{F}_{t_1}$ and $B \in \mathcal{B}(\mathcal{D})$, and $\mathbf{1}_B$ is the characteristic function of B. Since we could use a sequence of bounded and continuous functions to converge to the characteristic function $\mathbf{1}_B$ for all open sets B in $\mathcal{D}$, we only show the following case:

$$\int_D f(\mathbf{Y}_{t_2}^\xi(\omega))dP(\omega) = \int_D \int_\Omega f\{\mathbf{Y}_{t_2}^{t_1,\mathbf{Y}_{t_1}^\xi(\omega')}(\omega)\}dP(\omega)dP(\omega'), \tag{10}$$

where $D \in \mathcal{F}_{t_1}$ and $f :\ \mathcal{D} \to \mathbb{R}$ is bounded and continuous.

Since $\mathbf{Y}_{t_1}^\xi \in L^2(\Omega, \mathcal{D}; \mathcal{F}_{t_1})$, there is a sequence $\{\psi_j\}_{j=1}^\infty$ of $\mathcal{F}_{t_1}$-simple functions converging to $Y_{t_1}^\xi$ in $L^2(\Omega, \mathcal{D}; \mathcal{F}_{t_1})$, i.e.,

$$\psi_j = \sum_{i=1}^{n_j} \phi_{j,i}\mathbf{1}_{\Omega_{j,i}}, \quad \Omega_{j,i} \in \mathcal{F}_{t_1},\ \phi_{j,i} \in \mathcal{D}.$$

Therefore by $\boldsymbol{Y}_{t_2}^{\xi} = \boldsymbol{Y}_{t_2}^{t_1,\boldsymbol{Y}_{t_1}^{\xi}}$, we have for a.a. $\omega \in \Omega$,

$$\boldsymbol{Y}_{t_2}^{\xi}(\omega) = \lim_{j\to\infty} \boldsymbol{Y}_{t_2}^{t_1,\psi_j}(\omega) = \lim_{j\to\infty} \sum_{i=1}^{n_j} \boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}(\omega)\mathbf{1}_{\Omega_{j,i}}(\omega).$$

Each $\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}} \in L^2(\Omega,\mathcal{D};\mathcal{F}_{t_2}\cap\mathcal{G}^{t_1})$, where $\mathcal{G}^{t_1}$ is the σ-algebra generated by the compensated Poisson random measure $\widetilde{\lambda}$ from the inverse time direction, i.e.,

$$\sigma\{\widetilde{\lambda}(\mathcal{B},(t,u]),\ t \le u,\ \mathcal{B}\in\mathcal{B}((0,A])\}.$$

So we can write $\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}$ as a limit of $\mathcal{G}^{t_1}$-simple functions:

$$\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}} = \lim_{k\to\infty}\sum_{h=1}^{m_k}\theta_{k,h}^{j,i}\mathbf{1}_{\Omega_{k,h}^*}, \quad \theta_{k,h}^{j,i}\in\mathcal{D},\ \Omega_{k,h}^*\in\mathcal{G}^{t_1}.$$

Note that the left-hand side of (10) is equal to

$$\begin{aligned}
&\int_D \lim_{j\to\infty} f\left(\sum_{i=1}^{n_j}\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}(\omega)\mathbf{1}_{\Omega_{j,i}}(\omega)\right)\mathrm{d}P(\omega)\\
&\quad= \lim_{j\to\infty}\int_D\sum_{i=1}^{n_j} f\left(\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}(\omega)\right)\mathbf{1}_{\Omega_{j,i}}(\omega)\mathrm{d}P(\omega)\\
&\quad= \lim_{j\to\infty}\lim_{k\to\infty}\int_D\sum_{i=1}^{n_j}\sum_{h=1}^{m_k} f(\theta_{k,h}^{j,i})\mathbf{1}_{\Omega_{k,h}^*}(\omega)\mathbf{1}_{\Omega_{j,i}}(\omega)\mathrm{d}P(\omega)\\
&\quad= \lim_{j\to\infty}\lim_{k\to\infty}\sum_{i=1}^{n_j}\sum_{h=1}^{m_k} f(\theta_{k,h}^{j,i})P(\Omega_{k,h}^*\cap\Omega_{j,i}\cap D).
\end{aligned}\tag{11}$$

The right-hand side of (10) is equal to

$$\begin{aligned}
&\int_D\int_\Omega \lim_{j\to\infty} f\left(\boldsymbol{Y}_{t_2}^{t_1,\psi_j(\omega')}(\omega)\right)\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\int_D\int_\Omega f\left(\boldsymbol{Y}_{t_2}^{t_1,\psi_j(\omega')}(\omega)\right)\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\int_D\int_\Omega f\left[\left(\sum_{i=1}^{n_j}\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}\mathbf{1}_{\Omega_{j,i}}(\omega')}\right)(\omega)\right]\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\int_D\int_\Omega f\left[\sum_{i=1}^{n_j}\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}(\omega)\mathbf{1}_{\Omega_{j,i}}(\omega')\right]\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\int_D\int_\Omega\sum_{i=1}^{n_j} f\left(\boldsymbol{Y}_{t_2}^{t_1,\phi_{j,i}}(\omega)\right)\mathbf{1}_{\Omega_{j,i}}(\omega')\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\lim_{k\to\infty}\int_D\int_\Omega\sum_{i=1}^{n_j}\sum_{h=1}^{m_k} f(\theta_{k,h}^{j,i})\mathbf{1}_{\Omega_{k,h}^*}(\omega)\mathbf{1}_{\Omega_{j,i}}(\omega')\mathrm{d}P(\omega)\mathrm{d}P(\omega')\\
&\quad= \lim_{j\to\infty}\lim_{k\to\infty}\sum_{i=1}^{n_j}\sum_{h=1}^{m_k} f(\theta_{k,h}^{j,i})P(\Omega_{k,h}^*)P(\Omega_{j,i}\cap D).
\end{aligned}\tag{12}$$

Now $\Omega_{j,i}, D \in \mathcal{F}_{t_1}$, so $\Omega_{j,i} \cap D$ is independent of $\Omega^*_{k,h} \in \mathcal{G}^{t_1}$. Hence

$$P(\Omega^*_{k,h})P(\Omega_{j,i} \cap D) = P(\Omega^*_{k,h} \cap \Omega_{j,i} \cap D).$$

Compare (11) and (12), and it follows that (10) must hold.

Second, we prove

$$p(t_1, \boldsymbol{Y}^\xi_{t_1}(\cdot); t_2, B) = P(\boldsymbol{Y}^\xi_{t_2} \in B | \boldsymbol{Y}^\xi_{t_1}). \tag{13}$$

Due to the measurability of $p(t_1, \boldsymbol{Y}^\xi_{t_1}; t_2, B)$ with respect to the σ-algebra generated by $\boldsymbol{Y}^\xi_{t_1}$, we have

$$p(t_1, \boldsymbol{Y}^\xi_{t_1}; t_2, B) = \mathbb{E}\Big[P(\boldsymbol{Y}^\xi_{t_2} \in B|\mathcal{F}_{t_1})\Big|\boldsymbol{Y}^\xi_{t_1}\Big] = P(\boldsymbol{Y}^\xi_{t_2} \in B|\boldsymbol{Y}^\xi_{t_1}),$$

since the σ-algebra generated by $\boldsymbol{Y}^\xi_{t_1} \subset \mathcal{F}_{t_1}$.

Thus we finish the proof. ∎

The following lemma establishes the time-independent boundedness of the segment process $\boldsymbol{Y}_t$, which gives us the existence of the invariant measure $\pi^N(\cdot)$.

LEMMA 3.2. *There exists a constant C independent of time, such that*

$$\sup_{t\geq\tau} \mathbb{E}\|\boldsymbol{Y}_t\|^2_\infty < C.$$

Proof. By using Itô's formula to $|\boldsymbol{Y}(t)|^2$, for any $t \geq 0$, we have

$$\begin{aligned}
&\mathbb{E}|\boldsymbol{Y}(t)|^2 - \mathbb{E}|\eta(0)|^2 \\
&= \mathbb{E}\int_0^t \int_0^A \left[\left|\boldsymbol{Y}(s-) + \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right|^2 - \left|\boldsymbol{Y}(s-)\right|^2\right]\lambda(\mathrm{d}s\times\mathrm{d}a) \\
&= \mathbb{E}\int_0^t \int_0^A \left[2\left\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right\rangle \right.\\
&\qquad\left. + \left|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right|^2\right] m(\mathrm{d}s\times\mathrm{d}a).
\end{aligned} \tag{14}$$

Under Assumption 1, (14) can be estimated as

$$\begin{aligned}
\mathbb{E}|\boldsymbol{Y}(t)|^2 - \mathbb{E}|\eta(0)|^2 &\leq \mathbb{E}\int_0^t \big[-\alpha_1|\boldsymbol{Y}(s-)|^2 + \alpha_2 \max_j |\boldsymbol{Y}\circ\zeta(s-\tau_j)|^2\big]\mathrm{d}s \\
&\leq -\alpha_1\int_0^t \mathbb{E}|\boldsymbol{Y}(s)|^2\mathrm{d}s + \alpha_2\int_0^t \sup_{s-\tau\leq r\leq s}\mathbb{E}|\boldsymbol{Y}(r)|^2\mathrm{d}s.
\end{aligned}$$

Based on Gronwall's inequality, we know that there exists a constant C independent of time, such that

$$\sup_{t\geq-\tau} \mathbb{E}|\boldsymbol{Y}(t)|^2 \leq C. \tag{15}$$

Since $\|\boldsymbol{Y}_t\|_\infty^2 = \sup_{-\tau\le\theta\le0}|\boldsymbol{Y}(t+\theta)|^2$, we apply the Itô formula to $|\boldsymbol{Y}(t+\theta)|^2$: for any $t \ge \tau$ and $\theta \in [-\tau, 0]$,

$$
\begin{aligned}
&|\boldsymbol{Y}(t+\theta)|^2 - |\boldsymbol{Y}(t-\tau)|^2 \\
&= \int_{t-\tau}^{t+\theta}\int_0^A 2\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle m(\mathrm{d}s\times\mathrm{d}a) \\
&\quad + \int_{t-\tau}^{t+\theta}\int_0^A \left|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right|^2 m(\mathrm{d}s\times\mathrm{d}a) \\
&\quad + \int_{t-\tau}^{t+\theta}\int_0^A 2\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle \widetilde{\lambda}(\mathrm{d}s\times\mathrm{d}a) \\
&\quad + \int_{t-\tau}^{t+\theta}\int_0^A \left|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right|^2 \widetilde{\lambda}(\mathrm{d}s\times\mathrm{d}a).
\end{aligned}
\tag{16}
$$

We apply $\sup_{-\tau\le\theta\le0}$ and expectation to the first and second terms in the right-hand side of (16) and then get

$$
\begin{aligned}
&\mathbb{E}\left\{\sup_{-\tau\le\theta\le0}\int_{t-\tau}^{t+\theta}\int_0^A 2\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle m(\mathrm{d}s\times\mathrm{d}a)\right\} \\
&\quad + \mathbb{E}\left\{\sup_{-\tau\le\theta\le0}\int_{t-\tau}^{t+\theta}\int_0^A \left|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\right|^2 m(\mathrm{d}s\times\mathrm{d}a)\right\} \\
&\le C\mathbb{E}\int_{t-\tau}^{t}\left|\langle \boldsymbol{Y}(s),\ \sum_{j=1}^M \boldsymbol{\nu}_j a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle\right|\mathrm{d}s + \mathbb{E}\int_{t-\tau}^{t}\sum_{j=1}^M |\boldsymbol{\nu}_j|^2 a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\mathrm{d}s \\
&\le C\int_{t-\tau}^{t}\Big(1+\mathbb{E}|\boldsymbol{Y}(s)|^2 + \max_j \mathbb{E}|\boldsymbol{Y}\circ\zeta(s-\tau_j)|^2\Big)\mathrm{d}s \\
&\le C,
\end{aligned}
$$

where we use the fact that $\boldsymbol{Y} \in \Omega_{\boldsymbol{X}_0}$, which means $|\boldsymbol{Y}(t)| \le |\boldsymbol{Y}(t)|^2$. Utilizing the Burkholder–Davis–Gundy inequality to the third and forth terms in the right-hand side of (16), respectively, we get

$$
\begin{aligned}
&\mathbb{E}\left\{\sup_{-\tau\le\theta\le0}\int_{t-\tau}^{t+\theta}\int_0^A 2\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle \widetilde{\lambda}(\mathrm{d}s\times\mathrm{d}a)\right\} \\
&\le C\mathbb{E}\left(\int_{t-\tau}^{t}\int_0^A |\langle \boldsymbol{Y}(s-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\rangle|^2 m(\mathrm{d}s\times\mathrm{d}a)\right)^{\frac{1}{2}} \\
&\le C\max_j \mathbb{E}\left(\int_{t-\tau}^{t}|\boldsymbol{Y}(s)|^2 a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\mathrm{d}s\right)^{\frac{1}{2}} \\
&\le C\max_j \mathbb{E}\left(\int_{t-\tau}^{t}|\boldsymbol{Y}(s)|^2(1+|\boldsymbol{Y}\circ\zeta(s-\tau_j)|)\mathrm{d}s\right)^{\frac{1}{2}}
\end{aligned}
$$

$$
\begin{aligned}
&\le C\mathbb{E}\bigg(\int_{t-\tau}^{t}|\boldsymbol{Y}(s)|^2\mathrm{d}s\bigg)^{\frac{1}{2}} + C\max_j \mathbb{E}\bigg(\int_{t-\tau}^{t}|\boldsymbol{Y}(s)|^2|\boldsymbol{Y}\circ\zeta(s-\tau_j)|\mathrm{d}s\bigg)^{\frac{1}{2}}\\
&\le C + \mathbb{E}\int_{t-\tau}^{t}|\boldsymbol{Y}(s)|^2\mathrm{d}s + \frac{1}{2}\mathbb{E}\|\boldsymbol{Y}_t\|_\infty^2 + C\max_j \mathbb{E}\int_{t-\tau}^{t}|\boldsymbol{Y}\circ\zeta(s-\tau_j)|^2\mathrm{d}s\\
&\le \frac{1}{2}\mathbb{E}\|\boldsymbol{Y}_t\|_\infty^2 + C,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\bigg\{\sup_{-\tau\le\theta\le 0}\int_{t-\tau}^{t+\theta}\int_0^A\Big|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\Big|^2\widetilde{\lambda}(\mathrm{d}s\times\mathrm{d}a)\bigg\}\\
&\le C\mathbb{E}\bigg(\int_{t-\tau}^{t}\int_0^A\bigg|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\bigg|^4 m(\mathrm{d}s\times\mathrm{d}a)\bigg)^{\frac{1}{2}}\\
&\le C\max_j \mathbb{E}\bigg(\int_{t-\tau}^{t} a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\mathrm{d}s\bigg)^{\frac{1}{2}}\\
&\le C\max_j \mathbb{E}\bigg(\int_{t-\tau}^{t}(1+|\boldsymbol{Y}\circ\zeta(s-\tau_j)|)\mathrm{d}s\bigg)^{\frac{1}{2}}\\
&\le C + C\max_j \mathbb{E}\int_{t-\tau}^{t}|\boldsymbol{Y}\circ\zeta(s-\tau_j)|^2\mathrm{d}s \le C,
\end{aligned}
$$

where we use the boundedness of $\boldsymbol{Y}(t)$ (see (15)).

Combining these estimates together, we have

$$
\mathbb{E}\|\boldsymbol{Y}_t\|_\infty^2 \le \mathbb{E}|\boldsymbol{Y}(t-\tau)|^2 + \frac{1}{2}\mathbb{E}\|\boldsymbol{Y}_t\|_\infty^2 + C, \tag{17}
$$

which leads to

$$
\sup_{t\ge\tau}\mathbb{E}\|\boldsymbol{Y}_t\|_\infty^2 \le C.
$$

Thus we complete the proof. ∎

Remark 1. Under the same procedure, we could prove that the moment of $\boldsymbol{X}_t$ is also uniform bounded for all the time, i.e.,

$$
\sup_{t\ge\tau}\mathbb{E}\|\boldsymbol{X}_t\|_\infty^2 < C,
$$

where the constant C is independent of time.

For $\theta \in [-\tau, 0]$ and $\widetilde{\theta} \in [0, \delta]$ with $\delta > 0$ being an arbitrary constant such that $\theta + \delta \in [-\tau, 0]$, by the Itô isometry, for any $t \ge \tau$, we obtain from (4) that

$$
\begin{aligned}
&\mathbb{E}^{t+\theta}\big|\boldsymbol{Y}_{t+\widetilde{\theta}}(\theta) - \boldsymbol{Y}_t(\theta)\big|^2 = \mathbb{E}^{t+\theta}\big|\boldsymbol{Y}(t+\widetilde{\theta}+\theta) - \boldsymbol{Y}(t+\theta)\big|^2\\
&= \mathbb{E}^{t+\theta}\bigg|\int_{t+\theta}^{t+\widetilde{\theta}+\theta}\int_0^A\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}\circ\zeta(s-\tau_j))\lambda(\mathrm{d}s\times\mathrm{d}a)\bigg|^2\\
&\le C\int_{t+\theta}^{t+\theta+\delta}\mathbb{E}^{t+\theta}\bigg\{\bigg|\sum_{j=1}^M \boldsymbol{\nu}_j a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\bigg|^2 + \int_0^A\sum_{j=1}^M |\boldsymbol{\nu}_j|^2 a_j(\boldsymbol{Y}\circ\zeta(s-\tau_j))\bigg\}\mathrm{d}s,
\end{aligned}
$$

where $\mathbb{E}^t(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_t)$. Under the global Lipschitz property of functions a_j and the boundedness of the segment process, there is a random function $\gamma(t,\delta)$ satisfying

$$\mathbb{E}^{t+\theta}|\boldsymbol{Y}_{t+\widetilde{\theta}}(\theta) - \boldsymbol{Y}_t(\theta)|^2 \leq \mathbb{E}^{t+\theta}\gamma(t,\delta) \quad \forall\, \widetilde{\theta} \in [0,\ \delta], \quad \theta \in [-\tau,\ 0],$$

and

$$\lim_{\delta\to 0}\limsup_{t\to\infty}\mathbb{E}\gamma(t,\delta) = 0,$$

It follows from the Kurtz criterion of tightness [13, Theorem 3] that $\boldsymbol{Y}_t$ is tight in $\mathcal{D}$ under the Skorohod metric. Combining the tightness and eventually Feller property (similar to section 3.3 in [22]), we conclude the existence of the invariant measure by the Krylov–Bogoliubov theorem (see [8, Theorem 3.1.1] or [16]).

The following lemma establishes the difference of the segment processes $\boldsymbol{Y}_t^\xi$ and $\boldsymbol{Y}_t^\eta$ with different initial data ξ and η. The difference could be controlled by an exponential decay function, which leads to the uniqueness of the invariant measure, and the exponential ergodicity.

LEMMA 3.3. *Let $\boldsymbol{Y}_t^\xi$ and $\boldsymbol{Y}_t^\eta$, $t \geq 0$, be the segment processes of the solution of* (4) *with initial data ξ and η, respectively. Then there exist a time-independent constant C and a parameter $\kappa > 0$ such that*

$$\sup_{t\geq\tau}\mathbb{E}\|\boldsymbol{Y}_t^\xi - \boldsymbol{Y}_t^\eta\|_\infty^2 \leq Ce^{-2\kappa t}.$$

Proof. By applying Itô formula to $|\boldsymbol{Y}^\xi(t) - \boldsymbol{Y}^\eta(t)|^2$, we could obtain, for $t \geq 0$,

$$\begin{aligned}
&\mathrm{d}\mathbb{E}|\boldsymbol{Y}^\xi(t) - \boldsymbol{Y}^\eta(t)|^2 \\
&= \mathbb{E}\int_0^A 2\langle \boldsymbol{Y}^\xi(t-) - \boldsymbol{Y}^\eta(t-),\ \sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}^\xi\circ\zeta(t-\tau_j)) - \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}^\eta\circ\zeta(t-\tau_j))\rangle \\
&+ \left|\sum_{j=1}^M \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}^\xi\circ\zeta(t-\tau_j)) - \boldsymbol{\nu}_j c_j(a;\ \boldsymbol{Y}^\eta\circ\zeta(t-\tau_j))\right|^2 m(\mathrm{d}t\times\mathrm{d}a).
\end{aligned}$$

Let $\rho(t) = \mathbb{E}|\boldsymbol{Y}^\xi(t) - \boldsymbol{Y}^\eta(t)|^2$, and under Assumption 1,

$$\rho'(t) \leq -\alpha_1\rho(t) + \alpha_2 \sup_{t-\tau\leq s\leq t}\rho(s).$$

By Gronwal's inequality (see Lemma 2.3 in [1]), there exists some $\kappa > 0$ such that

$$\mathbb{E}|\boldsymbol{Y}^\xi(t) - \boldsymbol{Y}^\eta(t)|^2 \leq C\mathbb{E}\|\xi - \eta\|_\infty^2 e^{-2\kappa t} \leq Ce^{-2\kappa t} \quad \forall\, t \geq 0.$$

Using the Burkholder–Davis–Gundy inequality, we estimate as in Lemma 3.2 to show that there is a time-independent constant C such that

$$\mathbb{E}\|\boldsymbol{Y}_t^\xi - \boldsymbol{Y}_t^\eta\|_\infty^2 \leq Ce^{-2\kappa t}, \quad t \geq \tau.$$

Thus we finish the proof. ∎

Remark 2. Under the same procedure, we could prove that the dependence of $\boldsymbol{X}_t$ on initial data, i.e.,

$$\sup_{t\geq\tau}\mathbb{E}\|\boldsymbol{X}_t^\xi - \boldsymbol{X}_t^\eta\|_\infty^2 \leq C\mathbb{E}\|\xi - \eta\|_\infty^2 e^{-2\kappa t}.$$

The uniqueness of invariant measure follows from this exponential decay property. In fact, assume that $\widetilde{\pi}^N(\cdot) \in \mathcal{P}(\mathcal{D})$ is also an invariant measure, and then for all $F \in Lip(\mathcal{D})$,

$$\begin{aligned}
|\pi^N(F) - \widetilde{\pi}^N(F)| &= \left| \int_{\mathcal{D}} F(\xi)\pi^N(\mathrm{d}\xi) - \int_{\mathcal{D}} F(\eta)\widetilde{\pi}^N(\mathrm{d}\eta) \right| \\
&\leq \int_{\mathcal{D}\times\mathcal{D}} \left| P_t^N F(\xi) - P_t^N F(\eta) \right| \pi^N(\mathrm{d}\xi)\widetilde{\pi}^N(\mathrm{d}\eta) \\
&= \int_{\mathcal{D}\times\mathcal{D}} \left| \mathbb{E}F(\boldsymbol{Y}_t^\xi) - \mathbb{E}F(\boldsymbol{Y}_t^\eta) \right| \pi^N(\mathrm{d}\xi)\widetilde{\pi}^N(\mathrm{d}\eta) \\
&\leq C \int_{\mathcal{D}\times\mathcal{D}} \mathbb{E}d^o(\boldsymbol{Y}_t^\xi,\ \boldsymbol{Y}_t^\eta)\pi^N(\mathrm{d}\xi)\widetilde{\pi}^N(\mathrm{d}\eta) \\
&\leq C \int_{\mathcal{D}\times\mathcal{D}} \mathbb{E}\|\boldsymbol{Y}_t^\xi - \boldsymbol{Y}_t^\eta\|_\infty \pi^N(\mathrm{d}\xi)\widetilde{\pi}^N(\mathrm{d}\eta) \to 0, \qquad \text{as } t \to \infty.
\end{aligned}$$

Last, the exponential ergodicity could be shown similarly. By the invariance of the measure $\pi^N \in \mathcal{P}(\mathcal{D})$,

$$\begin{aligned}
|P_t^N F(\xi) - \pi^N(F)| &= \left| P_t^N F(\xi) - \int_{\mathcal{D}} P_t^N F(\eta)\pi^N(\mathrm{d}\eta) \right| \\
&\leq \int_{\mathcal{D}} |P_t^N F(\xi) - P_t^N F(\eta)|\pi^N(\mathrm{d}\eta) \\
&= \int_{\mathcal{D}} |\mathbb{E}F(\boldsymbol{Y}_t^\xi) - \mathbb{E}F(\boldsymbol{Y}_t^\eta)|\pi^N(\mathrm{d}\eta) \\
&\leq C\mathbb{E}d^0(\boldsymbol{Y}_t^\xi, \boldsymbol{Y}_t^\eta) \leq C\mathbb{E}\|\boldsymbol{Y}_t^\xi - \boldsymbol{Y}_t^\eta\|_\infty \\
&\leq C\Big(\mathbb{E}\|\boldsymbol{Y}_t^\xi - \boldsymbol{Y}_t^\eta\|_\infty^2\Big)^{\frac{1}{2}} \leq Ce^{-\kappa t}.
\end{aligned}$$

4. Error estimate on invariant measures. In this section, we will study the error between invariant measure π of (1) and invariant measure π^N of (4).

In fact, based on the ergodicity of processes $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$, we have the following two equations: for any deterministic initial data $\xi \in \mathcal{D}$,

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{E}F(\boldsymbol{X}_t^\xi)\mathrm{d}t &= \int_{\mathcal{D}} F(\eta)\pi(\mathrm{d}\eta) = \pi(F), \\
\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{E}F(\boldsymbol{Y}_t^\xi)\mathrm{d}t &= \int_{\mathcal{D}} F(\eta)\pi^N(\mathrm{d}\eta) = \pi^N(F).
\end{aligned}$$

Supposing we have the time-independent weak convergence order, from

$$\begin{aligned}
\big|\pi(F) - \pi^N(F)\big| &= \left| \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{E}F(\boldsymbol{X}_t^\xi) - \mathbb{E}F(\boldsymbol{Y}_t^\xi)\mathrm{d}t \right| \\
&\leq \lim_{T\to\infty} \frac{1}{T} \int_0^T \big|\mathbb{E}F(\boldsymbol{X}_t^\xi) - \mathbb{E}F(\boldsymbol{Y}_t^\xi)\big|\mathrm{d}t,
\end{aligned} \tag{18}$$

the error between invariant measure π of (1) and invariant measure π^N of (4) has obviously the same order as weak convergence order. Therefore, the key point to prove Theorem 2.2 is to show that the weak error of scheme (4) is independent of time interval.

The classical approach to prove the weak convergence order of SDEs is via a Kolmogorov PDE. At some point the success of this approach is based on the adaptedness and Markov property of the underlying stochastic process. However, due to the existence of time delay, the Markov property is broken, and SDDEs do not correspond to diffusions on Euclidean space. Thus the PDE technique does not apply. Another approach to prove the weak convergence order of SDEs (even the anticipating SDEs) uses the integration by parts of the Malliavin calculus (see [6, 7], for example) instead of using the Markov property and the solution of the PDE. However, the nondifferentiability of the coefficients of (1) adds complications and difficulties to the analysis of weak convergence order. To solve this problem, the tame property of numerical approximation, the Itô formula for tame functionals, and Malliavin calculus are utilized. See [7, section 3.2] for the weak convergence analysis in finite time interval.

The brief outline of the proof of the result of weak convergence order is as follows; see section 4.3.

Step 1. For any test function $\phi:\ \mathbb{R}^N \to \mathbb{R}$, we utilize the Markov property for the segment processes $\mathbf{X}_t$, $\mathbf{Y}_t$ (see Lemma 3.1), and the tame character of $\mathbf{Y}^{s,\eta}(t)$ (see Lemma 4.4 or Proposition 4.5) to rewrite the weak error as

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

Step 2. In this step, we make use of the tame Itô formula to expand each term in the right-hand side of the equation in Step 1. Hence,

$$\mathbb{E}\phi(\boldsymbol{X}_{t_n}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n}) =: \sum_{i=1}^{n}\sum_{m=1}^{k}\sum_{j=1}^{M}\left(\int_{t_{i-1}}^{t_i}\mathcal{D}_{m,j}^{i,1}(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i}\mathcal{D}_{m,j}^{i,2}(s)\mathrm{d}s\right),$$

where

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

Step 3. To estimate the term $\mathcal{D}_{m,j}^{i,1}$ in Step 2, we need the establishment of Itô formula for tame functionals (see Proposition 4.6) and the Malliavin differentiable of SDDE (1) (see Proposition 4.7). It yields

$$\big|\mathcal{D}_{m,j}^{i,1}\big| = O(\delta t).$$

Step 4. The estimate of the term $\mathcal{D}_{m,j}^{i,2}$ in Step 2 is similar as in Step 3, we make use of Itô formula for tame functionals and Malliavin differentiable of SDDE (1) to obtain

$$\big|\mathcal{D}_{m,j}^{i,2}\big| = O(\delta t).$$

4.1. Preliminaries. In this part, we will introduce some notation and propositions about Malliavin calculus and the tame Itô formula.

First, let us start with a brief introduction of Malliavin calculus for Poisson random measure, and note that all the definitions and properties are from [20, Chapter 12], which is by means of the chaos expansion in terms of iterated integrals with respect to compensated Poisson random measure. Denote D the Malliavin differentiation operator associated with 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 $\|F\|^2_{\mathbb{D}^{1,2}} = \sum_{n=1}^{0} nn!\|f_n\|^2_{L^2} < \infty$ (see [20, Definition 12.1]). The operator D is defined by $D_{t,z}F = \sum_{n=1}^{\infty} nI_{n-1}(f_n(\cdot,t,z))$ for all $F \in \mathbb{D}^{1,2}$ (see [20, Definition 12.2]). To perform the weak convergence analysis, we also need some properties of Malliavin derivatives. First, we present the chain rule for Malliavin derivative; for the proof we refer to [20, Theorem 12.8].

PROPOSITION 4.1 (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*

$$D_{t,z}\psi(F) = \Psi(F + D_{t,z}F) - \psi(F). \tag{19}$$

The Skorohod integral can be considered as an adjoint operator to the Malliavin derivative, and it is an extension of the Itô integral. See [20, Definition 11.1] for the definition of the Skorohod integral. Below is the relationship between the Malliavin derivative and the Skorohod integral; for the proof we refer to [20, Theorem 12.10].

PROPOSITION 4.2 (duality formula). *Let $X(t,z)$, $t \in [0,T]$, $z \in [0,A]$, be the Skorohod integrable and $F \in \mathbb{D}^{1,2}$. Then*

$$\mathbb{E}\left[F \int_0^T \int_0^A X(t,z)\widetilde{\lambda}(\mathrm{d}t \times \mathrm{d}z)\right] = \mathbb{E}\left[\int_0^T \int_0^A X(t,z) D_{t,z}F m(\mathrm{d}t \times \mathrm{d}z)\right]. \tag{20}$$

The following result is the fundamental theorem of calculus for Poisson random measure; for the proof we refer to [20, Theorem 12.15].

PROPOSITION 4.3 (fundamental theorem of calculus). *Let $X(s,y)$, $(s,y) \in [0,T]\times [0,A]$, be a stochastic process such that*

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

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

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

Then

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

and

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

Next, let us introduce some notation about the tame functional and the Itô formula for tame functionals; see [6, 7], for instance. Define the tame projection $\Pi:\ \mathcal{D} \to \mathbb{R}^{Nk}$ associated with $\mu_1, \dots, \mu_k \in [-\tau, 0]$ by

$$\Pi(\eta) := (\eta(\mu_1), \dots, \eta(\mu_k)) \in \mathbb{R}^{Nk} \tag{21}$$

for all $\eta \in \mathcal{D}$.

A functional $\Psi:\ [0,T] \times \mathcal{D} \to \mathbb{R}$ is called tame if there exits a functional $f:\ [0,T] \times \mathcal{D} \to \mathbb{R}$ and a tame projection $\Pi:\ \mathcal{D} \to \mathbb{R}^{Nk}$ such that

$$\Psi(t, \eta) = f(t, \Pi(\eta)) \tag{22}$$

for all $t \in [0,T]$ and $\eta \in \mathcal{D}$.

The following lemma gives the tame character of the solution $\mathbf{Y}^{s,\eta}(t)$ of the numerical approximation (4).

LEMMA 4.4 (see [7, Lemma 3.8]). *Let t_i be a fixed partition point for some $i \in \{0, 1, \dots, N_T\}$ Then for a.a. $\omega \in \Omega$, the function*

$$\begin{aligned} [t_i,\ T] \times \mathcal{D} &\to \mathbb{R}^N \\ (t,\ \eta) &\mapsto \mathbf{Y}^{t_i,\eta}(t, \omega) \end{aligned}$$

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

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

Considering the segment process $\boldsymbol{Y}_t^{t_i,\eta}$, from the above result in Lemma 4.4, there exists a random function $\boldsymbol{F}$ such that, for any $\theta \in [-\tau, 0]$,

$$\mathbf{Y}_t^{t_i,\eta}(\theta) = \boldsymbol{Y}^{t_i,\eta}(t+\theta) = \boldsymbol{F}(t+\theta, \Pi(\eta)) =: \boldsymbol{F}_t(\theta, \Pi(\eta)).$$

Therefore, we have the following result.

PROPOSITION 4.5. *Given any fixed t and a function $\phi:\ \mathcal{D} \to \mathbb{R}$, $\mathbb{E}\phi(\mathbf{Y}_t^{t_i,\eta})$ is a tame functional, which means there exists a deterministic function u such that*

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

The following proposition presents the Itô formula for tame functionals, which describes how the segment process $\boldsymbol{X}_t$ transforms under tame functionals.

PROPOSITION 4.6 (see [7, Proposition 3.10]). *Assume that*

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

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.

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

4.2. Properties. Now we are in the position to show that the solution $X(t)$ of (23) is Malliavin differentiable, and the bound is independent of time interval.

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

$$X(t) = \begin{cases} \eta(0) + \sum_{j=1}^{M} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau_j-))\lambda(\mathrm{d}s \times \mathrm{d}a), & t > \sigma, \\ \eta(t-\sigma), & \sigma - \tau \le t \le \sigma. \end{cases} \tag{23}$$

The Malliavin differentiability of the solution $X(t)$ of (23) is stated below.

PROPOSITION 4.7. *For any $\eta \in L^2(\Omega, \mathcal{D}; \mathcal{F}_\sigma)$ 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 (23) belongs to $\mathbb{D}^{1,2}$ for all $t \in [\sigma - \tau, \infty)$. Moreover, there exists a positive constant C independent of time such that*

$$\sup_{\sigma\ge 0} \sup_{r,t\ge\sigma-\tau} \mathbb{E} \int_0^A |D_{r,z}X(t;\sigma,\eta)|^2 \mathrm{d}z \le \left(1 + \sup_{\sigma-\tau\le s\le\sigma} \mathbb{E} \int_0^A \|D_{s,z}\eta\|_\infty^2\right). \tag{24}$$

Proof. We assume $\tau_j \equiv \tau$ in order to simplify notation. If $t \in [\sigma - \tau, \sigma]$, then $X(t) = \eta(t - \sigma)$, so it is obvious that (24) holds. We consider the case of $t > \sigma$: for given $r \le t$, we have

$$\begin{aligned} D_{r,z}X(t) &= D_{r,z}\eta(0) + \sum_{j=1}^{M} D_{r,z} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau-))\lambda(\mathrm{d}s \times \mathrm{d}a) \\ &= D_{r,z}\eta(0) + \sum_{j=1}^{M} D_{r,z} \int_\sigma^t \boldsymbol{\nu}_j a_j(X(s-\tau-))\mathrm{d}s \\ &\quad + \sum_{j=1}^{M} D_{r,z} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j c_j(a;\ X(s-\tau-))\widetilde{\lambda}(\mathrm{d}s \times \mathrm{d}a). \end{aligned} \tag{25}$$

Propositions 4.3 and 4.1 give us

$$\begin{aligned} D_{r,z}X(t) &= D_{r,z}\eta(0) + \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(z;\ X(r-\tau-)) \\ &\quad + \sum_{j=1}^{M} \int_\sigma^t \boldsymbol{\nu}_j \big[a_j(X(s-\tau-) + D_{r,z}X(s-\tau-)) - a_j(X(s-\tau-))\big]\mathrm{d}s \\ &\quad + \sum_{j=1}^{M} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j \big[c_j(a;\ X(s-\tau-) + D_{r,z}X(s-\tau-)) - c_j(a;\ X(s-\tau-))\big]\widetilde{\lambda}(\mathrm{d}s \times \mathrm{d}a) \\ &= D_{r,z}\eta(0) + \sum_{j=1}^{M} \boldsymbol{\nu}_j c_j(z;\ X(r-\tau-)) \\ &\quad + \sum_{j=1}^{M} \int_\sigma^t \int_0^A \boldsymbol{\nu}_j \big[c_j(a;\ X(s-\tau-) + D_{r,z}X(s-\tau-)) - c_j(a;\ X(s-\tau-))\big]\lambda(\mathrm{d}s \times \mathrm{d}a). \end{aligned}$$

Since $D_{r,z}X(t) = 0$ for $r > t$, we get

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

Letting $H(t,z) := D_{r,z}X(t)$, the above equation is

$$H(t,z) = H(r+\tau,z) + \sum_{j=1}^{M}\int_{r+\tau}^{t}\int_0^A \boldsymbol{\nu}_j\big[c_j(a;\ X(s-\tau-) + H(s-\tau-,z)) - c_j(a;\ X(s-\tau-))\big]\lambda(\mathrm{d}s\times\mathrm{d}a),$$

where $H(r+\tau,z) = D_{r,z}\eta(0) + \sum_{j=1}^{M}\boldsymbol{\nu}_j c_j(z;\ X(r-\tau-))$. Applying the Itô formula to $|H(t,z)|^2$, we have

$$\begin{aligned}|H(t,z)|^2 &= |H(r+\tau,z)|^2 \\ &\quad + \int_{r+\tau}^{t}\int_0^A \big[|H(s-,z) + K(s,a)|^2 - |H(s-,z)|^2\big]\lambda(\mathrm{d}s\times\mathrm{d}a) \\ &= |H(r+\tau,z)|^2 \\ &\quad + \int_{r+\tau}^{t}\int_0^A \big[2\langle H(s-,z),\ K(s,a)\rangle + |K(s,a)|^2\big]\lambda(\mathrm{d}s\times\mathrm{d}a),\end{aligned}$$

where

$$K(s,a) = \sum_{j=1}^{M}\boldsymbol{\nu}_j\big[c_j(a;\ X(s-\tau-) + H(s-\tau-,z)) - c_j(a;\ X(s-\tau-))\big].$$

Under Assumption 1,

$$\mathbb{E}|H(t,z)|^2 \le \mathbb{E}|H(r+\tau,z)|^2 + \int_{r+\tau}^{t}\big[-\alpha_1\mathbb{E}|H(s,z)|^2 + \alpha_2\mathbb{E}|H(s-\tau,z)|^2\big]\mathrm{d}s,$$

and then integrate with respect to z,

$$\begin{aligned}\mathbb{E}\int_0^A |H(t,z)|^2\mathrm{d}z \le \mathbb{E}\int_0^A |H(r+\tau,z)|^2\mathrm{d}z - \alpha_1\int_{r+\tau}^{t}\int_0^A \mathbb{E}|H(s,z)|^2\mathrm{d}z\mathrm{d}s \\ + \alpha_2\int_{r+\tau}^{t}\int_0^A \mathbb{E}|H(s-\tau,z)|^2\mathrm{d}z\mathrm{d}s.\end{aligned}$$

Hence, by Gronwall's inequality, there exists a constant C independent of time, such that

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

Thus we finish the proof. □

4.3. Time-independent weak convergence order. Recalling the result in (18),

$$\left|\pi(F) - \pi^N(F)\right| \le \lim_{T\to\infty} \frac{1}{T}\int_0^T \left|\mathbb{E}F(\boldsymbol{X}_t^\xi) - \mathbb{E}F(\boldsymbol{Y}_t^\xi)\right| \mathrm{d}t \quad \forall\, \xi \in \mathcal{D},$$

all we need is to prove the following weak convergence theorem, which means that the time-independent weak convergence order of the D-leaping scheme (4) is 1. Hence, from (18), the convergence order of invariant measures is also 1, which completes the proof of Theorem 2.2.

THEOREM 4.8 (weak convergence). *There exists a positive constant C independent of time such that*

$$|\mathbb{E}\phi(\boldsymbol{X}_{t_n}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n})| \le C\delta t \tag{26}$$

for all $n \in \{1, 2, \dots\}$ and $\phi:\ \mathcal{D} \to \mathbb{R}$ of class C_b^2.

Proof. Step 1. Using the Markov property for the segments $\boldsymbol{X}_t$ (see [18]) and $\boldsymbol{Y}_t$ (see Proposition 3.1), we may rewrite

$$\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}_{t_n}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n})\\
&\quad = \mathbb{E}\phi(\boldsymbol{Y}_{t_n}^{t_n,\boldsymbol{X}_{t_n}} - \mathbb{E}\phi(\boldsymbol{Y}_{t_n}^{0,\boldsymbol{X}_0})\\
&\quad = \sum_{i=1}^n \left\{\mathbb{E}\phi(\boldsymbol{Y}_{t_n}^{t_i,\boldsymbol{X}_{t_i}}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}})\right\}\\
&\quad = \sum_{i=1}^n \left\{\mathbb{E}\phi\left(\boldsymbol{Y}_{t_n}^{t_i,\boldsymbol{X}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}}}\right) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n}^{t_i,\boldsymbol{Y}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}}})\right\}.
\end{aligned} \tag{27}$$

From Proposition 4.5, 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 (27) into

$$\begin{aligned}
&\mathbb{E}\phi(\boldsymbol{X}_{t_n}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n})\\
&\quad = \sum_{i=1}^n \left\{\mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}})) - \mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}}))\right\}\\
&\quad = \sum_{i=1}^n \left\{\mathbb{E}u(\Pi(\boldsymbol{X}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}})) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}))\right\}\\
&\qquad - \left\{\mathbb{E}u(\Pi(\boldsymbol{Y}_{t_i}^{t_{i-1},\boldsymbol{X}_{t_{i-1}}})) - \mathbb{E}u(\Pi(\boldsymbol{X}_{t_{i-1}}))\right\}.
\end{aligned} \tag{28}$$

Step 2. By the tame Itô formula (Proposition 4.6), we obtain

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

and

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

We define

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

Thus (27) is

$$\begin{aligned}
\mathbb{E}\phi(\boldsymbol{X}_{t_n}) - \mathbb{E}\phi(\boldsymbol{Y}_{t_n}) &= \sum_{i=1}^{n} \sum_{m=1}^{k} \sum_{j=1}^{M} \mathbb{E} \int_{t_{i-1}}^{t_i} \Big[a_j(\boldsymbol{X}(s + \mu_m - \tau_j -)) f_j^m(\Pi(\boldsymbol{X}_s))\\
&\qquad - a_j(\boldsymbol{X}(t_{i-1} + \mu_m - \tau_j)) f_j^m(\Pi(\boldsymbol{Y}_s)) \Big] \mathrm{d}s \qquad (29)\\
&=: \sum_{i=1}^{n} \sum_{m=1}^{k} \sum_{j=1}^{M} \mathcal{D}_{m,j}^i.
\end{aligned}$$

In the following part, we need to show that there exists a time-independent constant $C > 0$ such that $\mathcal{D}_{m,j}^i \le C\delta t^2$.

We note that

$$\begin{aligned}\mathcal{D}^i_{m,j} &= \int_{t_{i-1}}^{t_i} \mathbb{E}\Big\{[a_j(\boldsymbol{X}(s+\mu_m-\tau_j-)) - a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))] f^m_j(\Pi(\boldsymbol{X}_s))\Big\}\mathrm{d}s \\ &\quad + \int_{t_{i-1}}^{t_i} \mathbb{E}\Big\{a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))[f^m_j(\Pi(\boldsymbol{X}_s)) - f^m_j(\Pi(\boldsymbol{Y}_s))]\Big\}\mathrm{d}s \qquad (30)\\ &=: \int_{t_{i-1}}^{t_i} \mathcal{D}^{i,1}_{m,j}(s)\mathrm{d}s + \int_{t_{i-1}}^{t_i} \mathcal{D}^{i,2}_{m,j}(s)\mathrm{d}s.\end{aligned}$$

We claim that for all $s \in [t_{i-1}, t_i]$, $\mathcal{D}^{i,1}_{m,j}(s)$, $\mathcal{D}^{i,2}_{m,j}(s) \le \delta t$ with $C > 0$ being independent of time, which means that $\mathcal{D}^i_{m,j} \le C\delta t^2$.

Step 3. Estimate of the term $\mathcal{D}^{i,1}_{m,j}(s)$. In fact, by the Itô formula

$$\begin{aligned}&\mathcal{D}^{i,1}_{m,j}(s) \\ &\quad = \mathbb{E}\Big\{f^m_j(\Pi(\boldsymbol{X}_s)) \int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} \int_0^A \big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\big]\lambda(\mathrm{d}u\times\mathrm{d}a)\Big\} \\ &\quad = \mathbb{E}\Big\{f^m_j(\Pi(\boldsymbol{X}_s)) \int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} \int_0^A \big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\big]\widetilde{\lambda}(\mathrm{d}u\times\mathrm{d}a)\Big\} \\ &\quad + \mathbb{E}\Big\{f^m_j(\Pi(\boldsymbol{X}_s)) \int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} \int_0^A \big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\big]m(\mathrm{d}u\times\mathrm{d}a)\Big\} \qquad (31)\\ &\quad = \mathbb{E}\int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} \int_0^A D_{u,a} f^m_j(\Pi(\boldsymbol{X}_s))\big[a_j(\tilde{\boldsymbol{X}}(u-)) - a_j(\boldsymbol{X}(u-))\big]m(\mathrm{d}u\times\mathrm{d}a) \\ &\quad + \mathbb{E}\Big\{f^m_j(\Pi(\boldsymbol{X}_s)) \int_{t_{i-1}+\mu_m-\tau_j}^{s+\mu_m-\tau_j} \Big[\sum_{\ell=1}^M a_\ell(\boldsymbol{X}(u-\tau_\ell-)) \\ &\qquad\qquad \Big(a_j(\boldsymbol{X}(u-)+\boldsymbol{\nu}_\ell) - a_j(\boldsymbol{X}(u-))\Big)\Big]\mathrm{d}u\Big\},\end{aligned}$$

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

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

we make the following estimates for functions of f^m_j:

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

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

$$\mathbb{E}|f_j^m(\Pi(\boldsymbol{X}_s))|^2 \le C\|\phi\|^2_{C_b^1}(1+\mathbb{E}\|\boldsymbol{X}_s\|^2_\infty) \le C.$$

And the estimates for $D_{u,z}f_j^m$ follow: Since

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

we have

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

where in the last step we use the result in Lemma 3.3. Further via Proposition 4.7,

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

Thus we can get that $\mathcal{D}^{i,1}_{m,j}(s) \le C\delta t$, where C does not depend on time. In fact, the first term on the right-hand side of (31) can be estimated by the Hölder inequality:

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

Note that the constants C in the above estimate are independent of time. The second term on the right-hand side of (31) can be estimated similarly,

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

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

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

Step 4. Estimate of the term $\mathcal{D}_{m,j}^{i,2}(s)$. The estimate of term $\mathcal{D}_{m,j}^{i,2}(s)$ is similar, but by using the tame Itô formula,

$$\begin{aligned}&\mathcal{D}_{m,j}^{i,2}(s)\\&=\ \mathbb{E}\Big\{a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\Big[\Big(f_j^m(\Pi(\boldsymbol{X}_s))-f_j^m(\Pi(\boldsymbol{X}_{t_{i-1}}))\Big)\\&\quad-\Big(f_j^m(\Pi(\boldsymbol{Y}_s))-f_j^m(\Pi(\boldsymbol{X}_{t_{i-1}}))\Big)\Big]\Big\}\\&=\ \mathbb{E}\int_{t_{i-1}}^{s}\int_0^A D_{u,a}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\sum_{\ell=1}^{k}\\&\qquad\{[f_j^m(\ldots,\tilde{\boldsymbol{X}}_{u-}(\mu_\ell),\ldots)-f_j^m(\ldots,\boldsymbol{X}_{u-}(\mu_\ell),\ldots)]\\&\qquad\quad-[f_j^m(\ldots,\tilde{\boldsymbol{Y}}_{u-}(\mu_\ell),\ldots)-f_j^m(\ldots,\boldsymbol{Y}_{u-}(\mu_\ell),\ldots)]\}m(\mathrm{d}u\times\mathrm{d}a)\\&\quad+\mathbb{E}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\int_{t_{i-1}}^{s}\sum_{\ell=1}^{k}\sum_{j_1=1}^{M}\Big[a_{j_1}(\boldsymbol{X}(u+\mu_\ell-\tau_{j_1}-))F_{j_1}^{\ell}(\Pi(\boldsymbol{X}_u))\\&\quad-a_{j_1}(\boldsymbol{X}(t_{i-1}+\mu_\ell-\tau_{j_1}-))F_{j_1}^{\ell}(\Pi(\boldsymbol{Y}_u))\Big]\mathrm{d}u,\end{aligned}\tag{32}$$

where

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

By Proposition 4.7, we may show that $\mathcal{D}_{m,j}^{i,2}(s)\le C\delta t$ with C being independent of time. In fact, the first term on the right-hand side of (32) 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

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

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 (32) can be estimated similarly,

$$\begin{aligned}
&\mathbb{E}a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))\int_{t_{i-1}}^{s}g_2(u)\mathrm{d}u\\
&\quad\le \Big(\mathbb{E}|a_j(\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j))|^2\Big)^{\frac12}\left(\mathbb{E}|\int_{t_{i-1}}^{s}g_2(u)\mathrm{d}u|^2\right)^{\frac12}\\
&\quad\le L\delta t^{\frac12}\Big(1+\mathbb{E}|\boldsymbol{X}(t_{i-1}+\mu_m-\tau_j)|^2\Big)^{\frac12}\left(\mathbb{E}\int_{t_{i-1}}^{s}|g_2(u)|^2\mathrm{d}u\right)^{\frac12}\\
&\quad\le C\delta t,
\end{aligned}$$

where

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

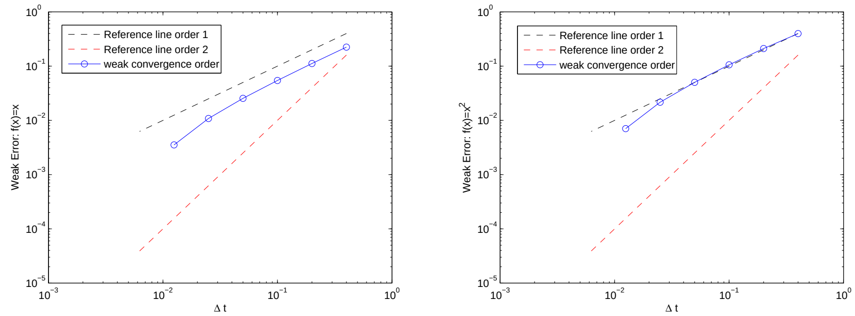

FIG. 1. *Log-log plot of the absolute error with functions* $f(x) = x$ *and* $f(x) = x^2$*, respectively.*

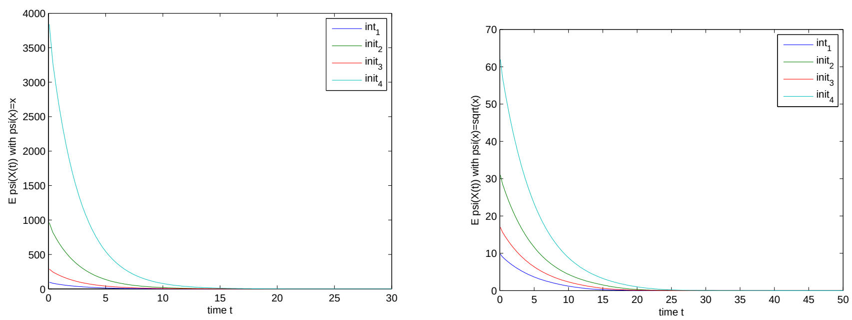

FIG. 2. *Long time behaviors of the solutions with different initial data for functions* $f(x) = x$ *and* $f(x) = \sqrt{x}$*, respectively.*

and

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

Therefore $\mathcal{D}^i_{m,j} \leq C\delta t^2$ with C being independent of time. Replacing it into (29), we finish the proof. ∎

5. Examples.

In this section, two examples are presented to support our theoretical analysis.

5.1. Example 1: Linear case.

For this system, we consider $\emptyset \to S$ with the propensity function being $a_1(x) = \alpha x$, where the rate constant $\alpha = 0.1$, the state-change vector is $\nu_1 = 1$, and the time delay is $\tau = 0.4$. And $S \to \emptyset$ with the propensity function being $a_2(x) = \beta x$, where the rate constant $\beta = 0.5$.

We plot the absolute errors of mean and variance in Figure 1. We simulate the reaction from time 0 to $T = 8$ using different stepsizes. 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 scheme has first order accuracy for the weak convergence.

We plot the behaviors of the solution in Figure 2 using different initial data, and observe that though the solution starts at different value, the averages go to the same state. It shows that the approximation possesses a unique invariant measure.

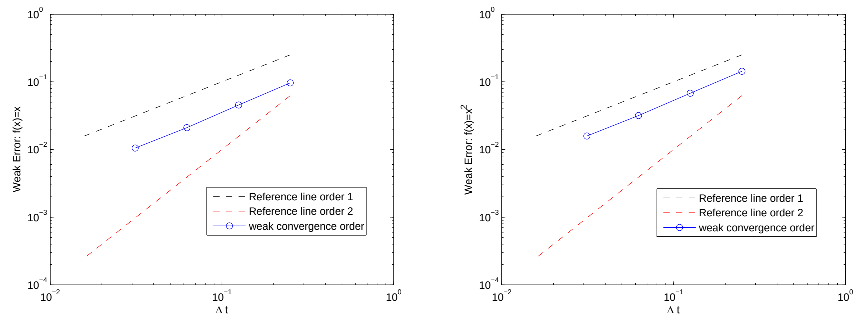

FIG. 3. *Log-log plot of the absolute error with functions $f(x) = x$ and $f(x) = x^2$, respectively.*

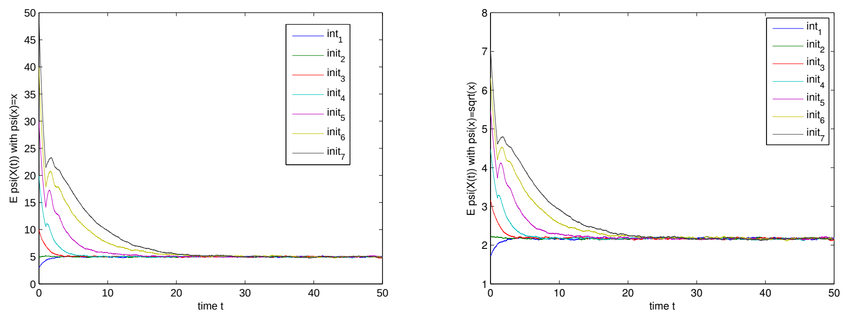

FIG. 4. *Long time behaviors of the solutions with different initial data for functions $f(x) = x$ and $f(x) = \sqrt{x}$, respectively.*

5.2. Example 2: Nonlinear case. For this system, we consider $\emptyset \to S$ with the propensity function being $a_1(x) = \alpha + \beta \frac{x(t-\tau)^b}{c^b + x(t-\tau)^b}$, where the constants are $\alpha = 5$, $\beta = 20$, $b = 10$, and $c = 19$, the state-change vector is $\nu_1 = 1$, and the time delay is $\tau = 1$. And $S \to \emptyset$ with the propensity function being $a_2(x) = \gamma x$, where the rate constant $\gamma = 1$.

We plot the absolute errors of mean and variance in Figure 3. We simulate the reaction from time 0 to $T = 10$ using different stepsizes. 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 scheme has first order accuracy for the weak convergence.

We plot the behaviors of the solution in Figure 4 using different initial data, and observe that though the solution starts at a different value, the averages go to the same state. It shows that the approximation possesses a unique invariant measure.

REFERENCES

[1] J. Bao, G. Yin, C. Yuan, and L. Wang, *Exponential ergodicity for retarded stochastic differential equations*, Appl. Anal., 93 (2014), pp. 2330–2349.

[2] M. Barrio, K. Burrage, A. Leier, and T. Tian, *Oscillatory regulation of Hes*1: *Discrete stochastic delay modelling and simulation*, PLoS Comput. Biol., 2 (2006), e117, https://doi.org/10.1371/journal.pcbi.0020117.

[3] 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.

[4] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[5] D. Bratsun, D. Volfson, L. Tsimring, and J. Hasty, *Delay-induced stochastic oscillations in gene regulation*, Proc. Natl. Acad. Sci. USA, 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., (2008), pp. 1461–1570.

[7] C. Chen and D. Liu, *Error analysis for D-leaping scheme of chemical reaction system with delay*, Multiscale Model. Simul., 15 (2017), pp. 1797–1829.

[8] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite-Dimensional Systems*, London Math. Soc. Lecture Note Ser. 229, Cambridge University Press, Cambridge, 1996.

[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] M. Hairer and A. Ohashi, *Ergodic thoery for SDEs with extrinsic memory*, Ann. Probab., 35 (2007), pp. 1950–1977.

[11] A. Ivanov, Y. Kazmerchuk, and A. Swishchuk, *Theory, stochastic stability and applications of stochastic delay differential equations: A survey of resent results*, Differ. Equ. Dyn. Syst., 11 (2003), pp. 55–115.

[12] A. Kulik, *Exponential ergodicity of the solutions to SDE's with a jump noise*, Stochastic Process Appl., 119 (2009), pp. 602–632.

[13] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Application to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.

[14] T. Li, *Analysis of explicit Tau-leaping schems for simulating chemically reacting systems*, Multiscale Model. Simul., 6 (2007), pp. 417–436.

[15] J. Mattingly, A. Stuart, and D. Higham, *Ergodicity for SDEs and approximations: A locally Lipschitz vector fields and degenerate noise*, Stoch. Process. Appl., 101 (2002), pp. 185–232.

[16] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, Cambridge University Press, Cambridge, 2009.

[17] W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, *Delay-Induced Degrade-and-Fire Oscillations in Small Genetic Circuits*, Phys. Rev. Lett., 102 (2009), 068105.

[18] 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, Basel, Switzerland, 1998, pp. 1–77.

[19] N. A. Monk, *Oscillatory expression of Hes*1*, p*53*, and NF-*κB *driven by transcriptional time delays*, Curr. Biol., 13 (2003), pp. 1409–1413

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

[21] E. Priola, A. Shirikyan, L. Xu, and J. Zabczyk, *Exponential ergodicity and regularity for equations with Lèvy noise*, Stochastoc Process Appl. 1 (2012), pp. 106–133.

[22] M. Reiss, M. Riedle, and O. van Gaans, *Delay differential equations driven by Lèvy processes: Stationarity and Feller properties*, Stochastic Process Appl., 116 (2006), pp. 1409–1432.

[23] D. Talay, *Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Process. Related Fields, 8 (2002), pp. 1–36.