Ann. Acad. Rom. Sci.
Ser. Math. Appl.
ISSN 2066-6594 Vol. 12, No. 1-2/2020

TWO-TIME-SCALE REGIME-SWITCHING STOCHASTIC KOLMOGOROV SYSTEMS WITH WIDEBAND NOISES*

George Yin† Zhexin Wen‡

Dedicated to Dr. Vasile Drăgan on the occasion of his 70th anniversary

Abstract

In our recent work, in lieu of using white noise, we examined Kolmogorov systems driven by wideband noise. Such systems naturally arise in statistical physics, biological and ecological systems, and many related fields. One of the motivations of our study is to treat more realistic models than the usually assumed stochastic differential equation models. The rationale is that a Brownian motion is an idealization used in a wide range of models, whereas wideband noise processes are much easier to be realized in the actual applications. This paper further investigates the case that in addition to the wideband noise process, there is a singularly perturbed Markov chain. The added Markov chain is used to model discrete events. Although it is a more realistic formulation, because of the non-Markovian formulation due to the wideband noise and the singularly perturbed Markov chain, the analysis is more

*Accepted for publication in revised form on April 8, 2020

†gyin@uconn.edu Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA. Research of this author was supported in part by the National Science Foundation under grant DMS-1710827.

‡zhexin.wen@wayne.edu Department of Mathematics, Wayne State University, Detroit, MI 48202, USA. Research of this author was supported in part by the Wayne State University Graduate Research Assistantship.

difficult. Using weak convergence methods, we obtain a limit result. Then we provide several examples for the utility of our findings.
MSC: 34F05, 60H10, 92D25, 92D40

keywords: Wideband noise, Kolmogorov system, non-Markov model, Markov chain.

1 Introduction

Recently, much effort has been devoted to stochastic Kolmogorov systems because their wide range of applications in statistical physics, ecology, and mathematical biology, among others. For example, there have been resurgent and emerging interests in studying Ginzburg-Landau equations [10] (see also [37]) in statistical physics, Lotka-Volterra models in statistical mechanics of population [21] (see also [9] for such equations with functional responses), the study of ecological models [7, 24], and the work on infectious disease modeling [25] among others. For multi-dimensional Kolmogorov systems, the paper [14] nearly completely classified the "threshold" of coexistence and extinction.

In contrast to the recent development, to the best of our knowledge, most of the stochastic models considered thus far have been concentrated on the "Markovian" formulation. The formulation has been confined to the treatment of Brownian motion and/or jump type noise processes. That is, the resulting systems are Markov processes. Owing to the use of the Markovian formulation, we have good technical machineries to handle the systems. One would naturally ask what if the systems are non-Markovian? Unfortunately, for non-Markovian systems, we lose all the usual analytic tools. There are generally no operators or generators associated to the underlying processes.

In our recent work [38], we examined a class of such non-Markovian systems, from the consideration of approximation and scaling limit. Although the mathematical idealization enables us to take advantage of the Markovian structure, in reality, very often, one does not have true "white" noise, but only has something close to "white" noise. In our paper, we used weak convergence methods to obtain limit systems. The limit systems are driven by a Brownian motion. In this paper, taking the result of [38] as a point of departure, we further examine Kolmogorov systems in which in addition to the wideband noise, the systems are hybrid with a switching component that are subject to both strong and weak interactions. Mathematically, the systems of interest are singularly perturbed systems; see also related work [29, 30] among others. Our objective is to show that such a system is close to a limit problem. The limit is obtained by using methods of weak convergence.

The rest of the paper is arranged as follows. We begin with some preliminary results and formulation of the problem in the next section. Section 3 presents our recent results on the limit of Kolmogorov type systems perturbed by wideband with Markovian switching. Section 4 considers several exams to illustrate what kind of systems can be treated under the framework of this paper. Finally, Section 5 concludes the paper with some further remarks and mention possible future works.

2 Preliminary Results and Formulation

2.1 Kolmogorov Systems

A d-dimensional stochastic Kolmogorov system under white noise is a time-homogenous stochastic differential equation of the form

$$\begin{aligned} &dx_i(t) = x_i(t)b_i(x(t))dt + x_i(t)\widehat{\sigma}_i(x(t))dw_i(t), \\ &x_i(0) = x_{i,0} > 0,\ i = 1, \dots, d, \end{aligned}$$

where $b_i(\cdot)$ and $\widehat{\sigma}_i(\cdot)$ are suitable nonlinear functions, $x = (x_1, \dots, x_d)' \in \mathbb{R}^d$, and $w(\cdot) = (w_1(\cdot), \dots, w_d(\cdot))'$ is a standard d-dimensional Brownian motion. In the above and hereafter, A' denotes the transpose of A, with A being either a vector or a matrix of appropriate dimension. This paper also focuses on Kolmogorov type systems, but we assume the systems are subject to wideband noise perturbations. In addition, we are dealing with a hybrid system in that the system involve both continuous dynamics and discrete events. The discrete event is modeled as a continuous-time Markov chain whose transition rate matrix or the generator includes a fast varying part and a slowly changing part.

An illustration of a wideband noise process can be found in [2, p. 439]. Suppose a stationary wideband noise is given by $G(x, \xi(t))$. One takes the covariance of the noise $R(x, s) = \mathbb{E}G(x, \xi(t))G'(x, \xi(t + s))$. Then one considers its power spectral density (the Fourier transform of the covariance) denoted by $\Gamma(x, \omega) = \int \exp(\imath\omega s)R(x, s)ds$, where $\imath$ is the imaginary number satisfying $\imath^2 = -1$. The premise of the wideband noise is that $\Gamma(x, \omega) = 0$ for sufficiently large ω, or it is band limited in that $\Gamma(x, \omega) = 0$ for $|\omega| > \omega_0 > 0$ and all $x \in \mathbb{R}^d$. So all relevant frequencies are contained in $[-\omega_0, \omega_0]$. In the limit of an appropriate sense, the band width of the power spectral tends to ∞, and the spectral tends to that of the white noise. More precise description and the conditions of the wideband noise will be given in the subsequent section. In [38], we showed how the idealized Brownian model can be approximated by a physically realizable model. The models that we are dealing with is non-Markovian. To obtain the desired limit, we introduce a small parameter $\varepsilon > 0$. Using the methods of weak convergence, we then showed that as $\varepsilon \to 0$, we obtain a limit system, which is the Kolmogorov system driven by a Brownian motion. In this paper, we assume that in addition to the wideband noise, there is a continuous-time Markov chain, which is used to represent a discrete event process. Why should we include both fast and slow motions in the Markov chain? The motivation stems from the consideration of large-scale systems. In the new era, in a wide variety of systems, to take into consideration of different scenarios, the discrete event process has a state space that is very large (i.e., the state space of the Markov chain is large). Naturally, not all the states vary at the same rate. Some of them change rapidly, whereas others are evolving slowly. The chain fits well into the formulation of the so-called nearly completely decomposable models (see [5, 29]) with the different transition rates highlighted by using a small parameter. By aggregating the states in each recurrent class into one state, the computation complexity can be much reduced.

2.2 Formulation

Consider a d-dimensional stochastic Kolmogorov system of the form

$$\begin{aligned} \dot{x}_i^\varepsilon(t) &= x_i^\varepsilon(t) f_i(x^\varepsilon(t), \xi^\varepsilon(t), \alpha^\delta(t)) + \frac{1}{\varepsilon} x_i^\varepsilon(t)\sigma_i(x^\varepsilon(t), \xi^\varepsilon(t), \alpha^\delta(t)), \\ x_i^\varepsilon(0) &= x_{i,0} > 0, \quad \alpha^\delta(0) = \alpha_0, \; i = 1, 2, \dots, d, \end{aligned} \tag{1}$$

where $\varepsilon > 0$ and $\delta > 0$ are small parameters, $\alpha^\delta(t)$ is a continuous-time Markov chain with a finite state space $\mathcal{M}$, $f_i(\cdot,\cdot,\cdot)$ and $\sigma_i(\cdot,\cdot,\cdot)$ are appropriate functions so that for each ξ and each $\alpha \in \mathcal{M}$, $f_i(\cdot, \xi, \alpha)$ and $\sigma_i(\cdot, \xi, \alpha)$ are smooth functions, and $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$, where $\xi(\cdot)$ is a stationary ϕ-mixing process satisfying certain conditions to be specified later. In this paper, we consider the case $\varepsilon = O(\delta)$. That is, ε and δ are varying at the same pace. To further simplify the notation, we simply assume $\varepsilon = \delta$, which is purely for notational simplicity.

Note that (1) is a system of random ordinary differential equations (ODEs); see [13] for a recent study of such random dynamic systems. Note also that (1) is generally non-Markovian. To illustrate, consider a scalar process $y(\cdot)$ (which might be called a "nonlinear" generalization of the Langevin equation) satisfying $\dot{y}(t) = b(y(t)) + c(y(t))\widetilde{\xi}(t)$, where $\widetilde{\xi}(\cdot)$ is a wideband process with a spectral density being roughly equal to a constant v^2 in a large neighborhood of the origin. As illustrated in [18, p.34], if $\widetilde{\xi}(\cdot)$ is not a white Gaussian noise, $y(\cdot)$ is not Markovian, in Chapters 8-10, Kushner provided a number of examples in applications. In [35], van Kampen argued that "non-Markov is the rule, Markov is the exception." He indicated that physical systems are generally non-Markov.

To proceed, denote $x = (x_1, \ldots, x_d)' \in \mathbb{R}^d$,

$$\begin{aligned} f(\cdot,\cdot,\cdot) &= (f_1(\cdot,\cdot,\cdot), \ldots, f_d(\cdot,\cdot,\cdot))' \in \mathbb{R}^d, \\ \sigma(\cdot,\cdot,\cdot) &= (\sigma_1(\cdot,\cdot,\cdot), \ldots, \sigma_d(\cdot,\cdot,\cdot))' \in \mathbb{R}^d. \end{aligned}$$

Note that we use z' to denote the transpose of $z \in \mathbb{R}^{\iota \times \iota_1}$ with ι and $\iota_1 \geq 1$. As usual, denote the corresponding first and second partial derivatives w.r.t. x by f_x, σ_x, and σ_{xx} for f and σ, respectively. Using vector notation, (1) can be written in a more compact form as

$$\begin{aligned} &\dot{x}^\varepsilon(t) = \Lambda(x^\varepsilon(t)) f(x^\varepsilon(t), \xi^\varepsilon(t), \alpha^\varepsilon(t)) + \frac{1}{\varepsilon}\Lambda(x^\varepsilon(t))\sigma(x^\varepsilon(t), \xi^\varepsilon(t), \alpha^\varepsilon(t)), \\ &x^\varepsilon(0) = x_0,\ \alpha^\varepsilon(0) = \alpha_0, \\ &\Lambda(x) = \mathrm{diag}(x_1, \ldots, x_d) \in \mathbb{R}^{d\times d} \quad \text{a } d \times d \text{ diagonal matrix.} \end{aligned} \tag{2}$$

where $x_0 = (x_{0,1}, \ldots, x_{0,d})'$ with $x_{0,i} > 0$, and $\alpha_0 \in \mathcal{M}$. Throughout the paper, we assume that $\alpha^\varepsilon(\cdot)$ and $\xi^\varepsilon(\cdot)$ are independent. Using the notion of p-lim [18], for any $\widehat{h}(\cdot,\cdot) : \mathbb{R}^d \times \mathcal{M}$ satisfying for each $\alpha \in \mathcal{M}$, $\widehat{h}(\cdot, \alpha) \in C_0^2$ (C^2 function with compact support), we can define an operator $\mathcal{L}^\varepsilon$ by

$$\mathcal{L}^\varepsilon \widehat{h}(x,\alpha) = \widehat{h}'_x(x,\alpha)\Big[\Lambda(x) f(x,\xi,\alpha) + \frac{1}{\varepsilon}\Lambda(x)\sigma(x,\xi,\alpha)\Big] + Q^\varepsilon(t)\widehat{h}(x,\cdot)(\alpha), \tag{3}$$

where for each $\alpha \in \mathcal{M}$,

$$Q^\varepsilon(t)\widehat{h}(x,\cdot)(\alpha) = \sum_{\gamma\in\mathcal{M}} q^\varepsilon_{\alpha\gamma}(t)\widehat{h}(x,\gamma),$$

with $Q^\varepsilon(t) = (q^\varepsilon_{\alpha\gamma}(t))$ being the generator of the Markov chain $\alpha^\varepsilon(t)$. In the next section, we provide the specific requirement for the process $\alpha^\varepsilon(\cdot)$.

2.3 Process $\alpha^\varepsilon(\cdot)$

We assume that the process $\alpha^\varepsilon(\cdot)$ has a large state space so that

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cdots \cup \mathcal{M}_l,$$

where for each $1 \leq \ell \leq l$,

$$\mathcal{M}_\ell = \{s_{\ell 1}, \dots, s_{\ell m_\ell}\}.$$

The generator of $\alpha^\varepsilon(\cdot)$ is of the form

$$Q^\varepsilon(t) = \frac{1}{\varepsilon}\widetilde{Q}(t) + \widehat{Q}(t), \tag{4}$$

where $\widetilde{Q}(t) = \mathrm{diag}(\widetilde{Q}^1(t), \dots, \widetilde{Q}^l(t))$, and each $\widetilde{Q}^\ell(t) \in m_\ell \times m_\ell$. Such a model was considered in our work [39]. The rational is that $\mathcal{M}$ is nearly completely decomposable in that the fast varying part is decomposed into l generators, but the actual states are not completely separated as such since the interactions due to $\widehat{Q}(t)$ (not a block diagonal matrix in general) connects the separable parts. Note also that the generator and hence the Markov chain is time dependent. To proceed, we aggregate the states in each $\mathcal{M}_\ell$ into one state to get a new process

$$\overline{\alpha}^\varepsilon(t) = \ell \ \text{ if } \ \alpha^\varepsilon(t) \in \mathcal{M}_\ell. \tag{5}$$

This aggregation process aims at reduction of computation complexity. It views all $s_{\ell j}$ with $j = 1, \dots, m_\ell$ as the same state. Thus all states in $\mathcal{M}_\ell$ can be represented by one state. Although the original state space $\mathcal{M}$ is large. The state space of the aggregated process $\overline{\mathcal{M}} = \{1, \dots, l\}$ is much smaller. The cardinality of $\mathcal{M}$ can be much smaller than that of $\overline{\mathcal{M}}$. If $|\overline{\mathcal{M}}| \ll |\mathcal{M}|$, then significant reduction of complexity is reached. Note that $\overline{\alpha}^\varepsilon(\cdot)$ is not Markovian. However, we can get a limit Markov process by sending $\varepsilon \to 0$. This is stated in the following lemma.

Lemma 2.1 *Assume that for each $\ell = 1, \dots, l$, $\widetilde{Q}^\ell(t)$ is weakly irreducible (see [39]). Then $\overline{\alpha}^\varepsilon(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$, which is a continuous-time Markov chain with generator*

$$\overline{Q}(t) = \mathrm{diag}(\nu^1(t), \dots, \nu^l(t))\widehat{Q}(t)\mathrm{diag}(\mathbb{1}_{m_1}, \dots, \mathbb{1}_{m_l}), \tag{6}$$

where for each $\ell = 1, \dots, l$, $\nu^\ell(t)$ is the quasi-stationary distribution associated with $\widetilde{Q}^\ell(t)$ and $\mathbb{1}_{m_\ell}$ is an m_ℓ-dimensional vector with all entries being 1.

Remark 2.2 For definitions of weak irreducibility and quasi-stationary distribution etc., we refer the reader to [39, Chapter 2]. The proof of the above lemma can be found in [39, Chapter 5].

3 Limit Results

Our objective is to show that the system with wideband noise is close to a limit system. To prove the limit result, we use the methods of weak convergence. The reader can find all related materials in, for example, [19, Chapters 7 and 8].

We have assumed the existence of unique solution of the limit stochastic Kolmogorov system, which is equivalent to that the associated martingale problem with operator $\mathcal{L}$ has a unique solution. Note that we are not working on pathwise solutions. We refer the reader to [11, Chapter 4] for further discussion on martingale problem formulation. Sufficient conditions ensuring the existence of unique solution in the strong sense can be found in [14, Assumption 1.1, p. 1897]. It should be mentioned that we only need the uniqueness to hold in the weak or distributional sense; we do not need the uniqueness in the strong sense. The uniqueness we impose, ensures the convergence to the correct limit. We need the following conditions.

(H1) The wideband noise $\xi^\varepsilon(t)$ is an $\mathbb{R}^d$-dimensional process satisfying $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$, where $\xi(\cdot)$ is a stationary and bounded ϕ-mixing process; $\xi(t)$ has right continuous sample paths and a mixing measure $\widetilde{\phi}(t)$ satisfying

$$\int_0^\infty \widetilde{\phi}^{1/2}(t)dt < \infty;$$

moreover, for each x and each α,

$$\mathbb{E}\sigma(x, \xi(t), \alpha) = 0. \tag{7}$$

(H2) For each $\alpha \in \mathcal{M}$, the functions $f(\cdot,\cdot,\alpha)$, $\sigma(\cdot,\cdot,\alpha)$, and $\sigma_x(\cdot,\cdot,\alpha)$ are continuous in both variables. For each ξ and each α, $f(\cdot,\xi,\alpha)$ and $\sigma(\cdot,\xi,\alpha)$ are locally Lipschitz in that for any x and y, and for each positive integer n, there is a $K_n > 0$ such that

$$|f(x,\xi,\alpha) - f(y,\xi,\alpha)| \vee |\sigma(x,\xi,\alpha) - \sigma(y,\xi,\alpha)| \le K_n|x-y|, \tag{8}$$

where $a \vee b = \max\{a,b\}$ for any two real numbers a and b. The $f_x(\cdot,\xi,\alpha)$ and $\sigma_{xx}(\cdot,\xi,\alpha)$ are continuous in x for each ξ and each α, and bounded on bounded x-set, where $f_x(\cdot,\xi,\alpha)$ and $\sigma_{xx}(\cdot,\xi,\alpha)$ denote the gradient and hessian of f and σ w.r.t. x, respectively.

(H3) Assume that stochastic Kolmogorov system (16) has a unique solution in the sense in distribution on $[0,T]$ for each $0 < T < \infty$ such that $x_i(t) > 0$ for $i = 1,\dots,d$.

Remark 3.1 For each x, each α, each $0 < T < \infty$, and each $t \in [0, T]$,

$$\mathbb{E}f(x, \xi(t), \alpha) = \int f(x, \xi, \alpha)\mu(d\xi), \tag{9}$$

where μ denotes the corresponding stationary measure.

For each x and each $\overline{\alpha} \in \overline{\mathcal{M}}$, denote

$$\begin{aligned} \eta(x, \overline{\alpha}) &= \int_0^\infty \mathbb{E}[\Lambda(x)\sigma(x, \xi(u), \overline{\alpha})]_x \Lambda(x)\sigma(x, \xi(0), \overline{\alpha})du \\ \frac{1}{2}S_0(x, \overline{\alpha}) &= \int_0^\infty \mathbb{E}\sigma(x, \xi(u), \overline{\alpha})\sigma'(x, \xi(0), \overline{\alpha})du. \end{aligned} \tag{10}$$

Each of the improper integrals above is convergent because of the ϕ-mixing property. For subsequent use, denote

$$\begin{aligned} S(x, \overline{\alpha}) &= (S_0(x, \overline{\alpha}) + S_0'(x, \overline{\alpha}))/2, \\ S(x, \overline{\alpha}) &= \overline{\sigma}(x, \overline{\alpha})\overline{\sigma}'(x, \overline{\alpha}). \end{aligned} \tag{11}$$

That is, $\overline{\sigma}(x, \overline{\alpha})$ is the square root of $S(x, \overline{\alpha})$. For each $s_{ij} \in \mathcal{M}$, Define $\overline{f}(x, s_{ij})$ as

$$\overline{f}(x, s_{ij}) = \int_0^\infty \mathbb{E}f(x, \xi(u), s_{ij})\nu_j^i(u). \tag{12}$$

For each $\iota \in \overline{\mathcal{M}}$, define $\overline{F}(x, \iota)$ as

$$\overline{F}(x, \iota) = \overline{f}(x, \iota) + \Lambda^{-1}(x)\eta(x, \iota), \tag{13}$$

We characterize the limit process by the solution of martingale problem with operator $\mathcal{L}$ defined by

$$\begin{aligned} \mathcal{L}h(x, \overline{\alpha}) &= h_x'(x, \overline{\alpha})\Lambda(x)\overline{F}(x, \overline{\alpha}) + \frac{1}{2}\text{tr}[h_{xx}(x, \overline{\alpha})\Lambda(x)S(x, \overline{\alpha})\Lambda(x)] \\ &\quad + \overline{Q}(t)h(x, \cdot)(\overline{\alpha}), \quad \text{for each } \ \overline{\alpha} \in \overline{\mathcal{M}}, \end{aligned} \tag{14}$$

for any $h(\cdot, \overline{\alpha}) \in C_0^2$ (i.e., the class of functions whose partial derivatives up to the second order are continuous with compact support), where $\text{tr}(A)$ denotes the trace of A, $h_x(x, \alpha)$ and $h_{xx}(x, \alpha)$ denote the gradient and hessian of $h(x, \alpha)$ with respect to x, respectively, and $\overline{\sigma}(x, \alpha)$ is defined in (11) below.

Because $\Lambda(x)$ is a diagonal matrix, $\Lambda(x) = \Lambda'(x)$. By $x(t)$ being a solution of a martingale problem, we meant that for any $h(\cdot, \overline{\alpha}) \in C_0^2$ (i.e., C^2 functions with compact support),

$$h(x(t), \overline{\alpha}(t)) - h(x_0, \ell) - \int_0^t \mathcal{L}h(x(s), \overline{\alpha}(s))ds \ \text{ is a martingale,} \tag{15}$$

where $\overline{\alpha}(0) = \ell$; see [11, p. 173], or [18, pp. 15-16], or [39, p. 378] for further details.

We consider the pair of processes $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$. Our effort is devoted to proving that $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(x(\cdot), \overline{\alpha}(\cdot))$ so that the limit is a solution of the stochastic Kolmogorov system

$$\begin{aligned} &dx(t) = \Lambda(x(t))\overline{F}(x(t), \overline{\alpha}(t))dt + \Lambda(x(t))\overline{\sigma}(x(t), \overline{\alpha}(t))dw(t), \\ &x(0) = x_0,\ \overline{\alpha}(0) = \overline{\alpha}, \end{aligned} \tag{16}$$

where $w(\cdot)$ is a d-dimensional standard Brownian motion, and $\eta(x, \iota)$ and $\overline{f}(x, \iota)$ are given by (10) and (12), respectively.

3.1 Truncated Process

Since the solution of (2) is not *a priori* bounded, we use an N-truncation device [18, p.83] or [19, p.284]. We first recall the definition of N-truncation. Let N be a fixed but otherwise arbitrary integer and $B_N = \{x : |x| \le N\}$ be the ball of radius N centered at the origin. Consider a truncated process $x^{\varepsilon,N}(\cdot)$ defined by $x^{\varepsilon,N}(0) = x_0$, and $x^{\varepsilon,N}(t) = x^\varepsilon(t) 1_{\{t \le \tau\}}$, and

$$\lim_{\kappa\to\infty} \limsup_{\varepsilon\to 0} \mathbb{P}(\sup_{t\le T} |x^{\varepsilon,N}(t)| \ge \kappa) = 0, \tag{17}$$

where 1_A is the indicator function, and $\tau = \inf\{s : |x^\varepsilon(s)| > N\}$. That is, τ is the first exit time of $x^\varepsilon(\cdot)$ from the N-ball B_N and $x^{\varepsilon,N}(t)$ is the process that is the same as $x^\varepsilon(\cdot)$ until it exits from the ball B_N and that satisfies (17). To obtain $x^{\varepsilon,N}(\cdot)$, we let a truncation function ρ^N be defined as

$$\rho^N(x) = \begin{cases} 1, \text{ if } x \in B_N, \\ 0, \text{ if } x \in \mathbb{R}^d - B_{N+1}, \\ \text{smooth, otherwise.} \end{cases}$$

Then we can write the corresponding truncated process as a solution of the system

$$\begin{aligned} \dot{x}^{\varepsilon,N}(t) &= \Lambda(x^{\varepsilon,N}(t)) f(x^{\varepsilon,N}(t), \xi^\varepsilon(t), \alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t)) \\ &\quad + \frac{1}{\varepsilon}\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t), \xi^\varepsilon(t), \alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t)). \end{aligned} \tag{18}$$

From the definition of $x^{\varepsilon,N}$, it is readily seen that

$$\mathbb{P}(\sup_{t\le T} |x^{\varepsilon,N}(t)| \ge \kappa) \le \frac{\mathbb{E}\sup_{t\le T} |x^{\varepsilon,N}(t)|}{\kappa} \to 0 \ \text{ as } \ \kappa \to \infty.$$

Thus from the above discussion and the construction, $x^{\varepsilon,N}(\cdot)$ is an N-truncation process. We will first prove that for a fixed N, $\{x^{\varepsilon,N}(\cdot)\}$ is tight. Then we establish its weak convergence. Letting $N \to \infty$, we show that the untruncated process $\{x^{\varepsilon}(\cdot)\}$ also converges. We note that in particular, we are not working on pathwise or strong solution, but rather we are working on convergence of probability measures.

Lemma 3.2 *Under* (H1)-(H3), $\{x^{\varepsilon,N}(\cdot), \overline{\alpha}^{\varepsilon}(\cdot)\}$ *is tight in* $D([0,T] : \mathbb{R}^d \times \overline{\mathcal{M}})$, *where* $D([0,T] : \mathbb{R}^d \times \overline{\mathcal{M}})$ *is the space of functions taking values in* $\mathbb{R}^d \times \overline{\mathcal{M}}$ *satisfying that the functions are defined on* $[0,T]$ *that are right continuous and that have left limit, endowed with the Skorohod topology.*

Proof. We note that we have already established in Lemma 2.1 that $\overline{\alpha}^{\varepsilon}(\cdot)$ converges to $\overline{\alpha}(\cdot)$ hence the tightness of $\{\overline{\alpha}^{\varepsilon}(\cdot)\}$. Because $\overline{\alpha}^{\varepsilon}(\cdot)$ is tight and has a state space $\overline{\mathcal{M}}$, we need only work with the component $x^{\varepsilon,N}(\cdot)$.

We use the perturbed test function methods (see [18, Chapter 4] and [19, p.172]) to prove the desired result. The essence is to define a perturbation that is small in magnitude and that results in the desired cancellation. For any $h(\cdot) \in C_0^2$ (the class of C^2 functions with compact support), using (3), it is easy to see that

$$\begin{aligned}
&\mathcal{L}^{\varepsilon,N} h(x^{\varepsilon,N}(t)) \\
&\quad = h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t)) f(x^{\varepsilon,N}(t), \xi^{\varepsilon}(t), \alpha^{\varepsilon}(t))\rho^N(x^{\varepsilon,N}(t)) \\
&\qquad + \frac{1}{\varepsilon} h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t), \xi^{\varepsilon}(t), \alpha^{\varepsilon}(t))\rho^N(x^{\varepsilon,N}(t)).
\end{aligned} \tag{19}$$

For simplicity, we choose $h(\cdot)$ to be independent of t and α. As a result, the term involving $Q^{\varepsilon}(t)$ in (3) disappears. Moreover, the letter N is used as an index to reflect the dependence on the truncation level N. Comparing to the work [38], the terms involve an $\alpha^{\varepsilon}(t)$ dependence, which makes the analysis more difficult.

To proceed, we aim to average out the $O(\varepsilon^{-1})$ term. Denote by $\mathbb{E}_t^{\varepsilon}$ the conditional expectation on the σ-algebra up to t, i.e., $\mathcal{F}_t^{\varepsilon}$, and define a perturbation by

$$\begin{aligned}
h_1^{\varepsilon}(x^{\varepsilon,N}(t), t) = \frac{1}{\varepsilon}\int_t^T & \mathbb{E}_t^{\varepsilon} h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t), \xi^{\varepsilon}(u), \alpha^{\varepsilon}(u)) \\
& \times \rho^N(x^{\varepsilon,N}(t)) du.
\end{aligned} \tag{20}$$

Because of the independence of $\xi(\cdot)$ and $\alpha^\varepsilon(u)$ and $\mathbb{E}\sigma(x,\xi(u),\iota)=0$ for each $\iota\in\mathcal{M}$,

$$\begin{aligned}\mathbb{E}\sigma(x,\xi(u),\alpha^\varepsilon(u)) &= \sum_{\iota\in\mathcal{M}}\mathbb{E}\sigma(x,\xi(u),\iota)1_{\{\alpha^\varepsilon(u)=\iota\}}\\ &= \sum_{\iota\in\mathcal{M}}\mathbb{E}\sigma(x,\xi(u),\iota)\mathbb{E}1_{\{\alpha^\varepsilon(u)=\iota\}}=0.\end{aligned}$$

Thus,

$$\begin{aligned}&h_1^\varepsilon(x^{\varepsilon,N}(t),t)\\ &=\frac{1}{\varepsilon}\int_t^T \mathbb{E}_t^\varepsilon h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(u),\alpha^\varepsilon(u))\rho^N(x^{\varepsilon,N}(t))du\\ &=\frac{1}{\varepsilon}\int_t^T h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))\{\mathbb{E}_t^\varepsilon\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(u),\alpha^\varepsilon(u))\\ &\qquad\qquad -\mathbb{E}\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(u),\alpha^\varepsilon(u))\}\rho^N(x^{\varepsilon,N}(t))du,\end{aligned}$$

and as a result, making a change of variable $u\to u/\varepsilon^2$,

$$\sup_{t\in[0,T]}|h_1^\varepsilon(x^{\varepsilon,N}(t),t)|\le\varepsilon\sup_{t\in[0,T]}\int_{t/\varepsilon^2}^{T/\varepsilon^2}\widetilde{\phi}(u-(t/\varepsilon^2))du=o(\varepsilon). \tag{21}$$

Define

$$h^\varepsilon(t)=h(x^{\varepsilon,N}(t))+h_1^\varepsilon(x^{\varepsilon,N}(t),t).$$

Using the cancellation of the last term in (19),

$$\begin{aligned}&\mathcal{L}^{\varepsilon,N}h^\varepsilon(t)\\ &=h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))f(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))\\ &+\int_{t/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_t^\varepsilon\Big[h_x'(x^{\varepsilon,N}(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi(u),\alpha^\varepsilon(u))\rho^N(x^{\varepsilon,N}(t))\Big]_x' du\\ &\qquad\times\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))+o(\varepsilon),\end{aligned} \tag{22}$$

where $o(\varepsilon)\to 0$ as $\varepsilon\to 0$. Using boundedness of $x^{\varepsilon,N}(\cdot)$ owing to the truncation, the continuity of $f(\cdot,\cdot,\iota)$ and $\sigma(\cdot,\cdot,\iota)$, the boundedness of $\{\xi(u)\}$, $\xi^\varepsilon(u)=\xi(u/\varepsilon^2)$, and $\{\xi(t)\}$ being stationary mixing, similar to the approach in [38], it can be shown that

$$\begin{aligned}&\{\sup_{|x|\le N}|\psi_1(x,\xi(t),\alpha^\varepsilon(t))|^2\}\ \text{ and}\\ &\sup_{\Delta\le 1}\sup_{|x|\le N}\Big\{\Big|\int_{t+\Delta}^T\psi_2(x,\xi(t),\alpha^\varepsilon(t))du\Big|^2,\ 0<t\le T\Big\}\end{aligned}$$

are uniformly integrable, where $\psi_1(x,\xi,\alpha)$ can be any of the functions $f(x,\xi,\alpha)$, $\sigma(x,\xi,\alpha)$, or $\sigma_x(x,\xi,\alpha)$, and similarly $\psi_2(x,\xi,\alpha)$ can be either $\sigma(x,\xi,\alpha)$ or $\sigma_x(x,\xi,\alpha)$. Therefore, $\{\mathcal{L}^{\varepsilon,N}h^\varepsilon(t)\}$ is uniformly integrable. It then follows from [18, Theorem 3.4, p.48], $\{x^{\varepsilon,N}(\cdot)\}$ is tight. The desired assertion of the lemma therefore follows. □

Because $(x^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ is tight, by Prohorov's theorem, we can extract a weakly convergent subsequence. Choose such a sequence and still use the same index (for simplicity) with limit $(x^N(\cdot),\overline{\alpha}(\cdot))$. By Skorohod representation, with a slight abuse of notation, we may assume $(x^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ converges to $(x^N(\cdot),\overline{\alpha}(\cdot))$ in the sense of w.p.1. We proceed to characterize the limit process. For any $h(\cdot,\cdot):\mathbb{R}^d\times\overline{\mathcal{M}}\mapsto\mathbb{R}$ satisfying for each $\ell\in\overline{\mathcal{M}}$, $h(\cdot,\ell)\in C_0^2$ (C^2 functions with compact support), form a function

$$\overline{h}(x,\alpha)=\sum_{\ell=1}^{l}h(x,\ell)1_{\{\alpha\in\mathcal{M}_\ell\}},\quad (x,\alpha)\in\mathbb{R}^d\times\mathcal{M}. \tag{23}$$

The definition of $\overline{h}(x,\alpha)$ implies that

$$\widetilde{Q}(t)\overline{h}(x,\cdot)(\alpha)=0.$$

Then similar to (19), we have

$$\begin{aligned}
&\mathcal{L}^{\varepsilon,N}\overline{h}(x^{\varepsilon,N}(t),\alpha^\varepsilon(t))\\
&=\overline{h}_x'(x^{\varepsilon,N}(t),\alpha^\varepsilon(t))\Lambda(x^{\varepsilon,N}(t))f(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))\\
&\quad+\frac{1}{\varepsilon}\overline{h}_x'(x^{\varepsilon,N}(t),\alpha^\varepsilon(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))\\
&\quad+\widehat{Q}(t)\overline{h}(x^{\varepsilon,N}(t),\cdot)(\alpha^\varepsilon(t)).
\end{aligned} \tag{24}$$

As (20), we define a perturbation $h_1^\varepsilon(x^{\varepsilon,N}(t),t)$. Then (21) still holds, which indicates that the perturbation is small. Define

$$\overline{h}^\varepsilon(t)=\overline{h}(x,\alpha)+h_1^\varepsilon(x,t). \tag{25}$$

Canceling the $O(\varepsilon^{-1})$ term, detailed calculation reveals that

$$\begin{aligned}
&\mathcal{L}^{\varepsilon,N}\overline{h}^\varepsilon(t)\\
&=\overline{h}_x'(x^{\varepsilon,N}(t),\alpha^\varepsilon(t))\Lambda(x^{\varepsilon,N}(t))f(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))\\
&\quad+\int_{t/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_t^\varepsilon\Big[\overline{h}_x'(x^{\varepsilon,N}(t),\alpha^\varepsilon(t))\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi(u),\alpha^\varepsilon(u))\\
&\qquad\times\rho^N(x^{\varepsilon,N}(t))\Big]_x'du\Lambda(x^{\varepsilon,N}(t))\sigma(x^{\varepsilon,N}(t),\xi^\varepsilon(t),\alpha^\varepsilon(t))\rho^N(x^{\varepsilon,N}(t))\\
&\quad+\widehat{Q}(t)\overline{h}(x^{\varepsilon,N}(t),\cdot)(\alpha^\varepsilon(t))+o(\varepsilon),
\end{aligned} \tag{26}$$

where $o(\varepsilon) \to 0$ as $\varepsilon \to 0$.

We proceed to carry out the averaging procedure. For any $\Delta > 0$, $t, s > 0$, and $s \le \Delta$, using (26), we obtain

$$\begin{aligned}
&\overline{h}^{\varepsilon}(t+s) - \overline{h}^{\varepsilon}(t) \\
&= \int_t^{t+s} \overline{h}'_x(x^{\varepsilon,N}(u), \alpha^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u)) f(x^{\varepsilon,N}(u), \xi^{\varepsilon}(u), \alpha^{\varepsilon}(u)) \\
&\qquad\qquad \times \rho^N(x^{\varepsilon,N}(u))du \\
&\quad + \int_t^{t+s} \int_{u/\varepsilon^2}^{T/\varepsilon^2} \mathbb{E}^{\varepsilon}_u \Big[\overline{h}'_x(x^{\varepsilon,N}(u), \alpha^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u), \xi(v), \alpha^{\varepsilon}(v)) \\
&\qquad \times \rho^N(x^{\varepsilon,N}(u))\Big]'_x dv \Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u), \xi^{\varepsilon}(u), \alpha^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du \\
&\quad + \int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u), \cdot)(\alpha^{\varepsilon}(u))du + o(\varepsilon),
\end{aligned} \tag{27}$$

where $o(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Define

$$\widetilde{c}^{\varepsilon}_{ij}(t) = 1_{\{\alpha^{\varepsilon}(t)=s_{ij}\}} - \nu^i_j(t) 1_{\{\overline{\alpha}^{\varepsilon}(t)=i\}}, \tag{28}$$

where $1_{\{B\}}$ denotes the indicator of the event B. Then

$$\begin{aligned}
&\int_t^{t+s} \overline{h}'_x(x^{\varepsilon,N}(u), \alpha^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u)) f(x^{\varepsilon,N}(u), \xi^{\varepsilon}(u), \alpha^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m_i} \int_t^{t+s} \overline{h}'_x(x^{\varepsilon,N}(u), s_{ij})\Lambda(x^{\varepsilon,N}(u)) f(x^{\varepsilon,N}(u), \xi^{\varepsilon}(u), s_{ij}) \\
&\qquad\qquad \times \rho^N(x^{\varepsilon,N}(u))\nu^i_j(u) 1_{\{\overline{\alpha}^{\varepsilon}(u)=i\}}du \\
&\quad + \sum_{i=1}^{l}\sum_{j=1}^{m_i} \int_t^{t+s} \overline{h}'_x(x^{\varepsilon,N}(u), s_{ij})\Lambda(x^{\varepsilon,N}(u)) f(x^{\varepsilon,N}(u), \xi^{\varepsilon}(u), s_{ij}) \\
&\qquad\qquad \times \rho^N(x^{\varepsilon,N}(u))\widetilde{c}^{\varepsilon}_{ij}(u)du.
\end{aligned} \tag{29}$$

To proceed, we state a lemma.

Lemma 3.3 *Suppose that $\alpha^{\varepsilon}(\cdot)$ satisfies the condition in Lemma* 2.1*, for each $i, j \in \mathcal{M}$, $\theta_{ij}(\cdot)$ is a bounded and continuous process that is independent of $\alpha^{\varepsilon}(\cdot)$. Then for any $0 < t < T < \infty$, we have*

$$\mathbb{E}\left| \int_t^T \theta_{ij}(u)\widetilde{c}^{\varepsilon}_{ij}(u)du \right|^2 = O(\varepsilon). \tag{30}$$

Proof. The proof is similar to [39, Theorem 5.25]. The only modifications needed is to insert a conditional expectation $\mathbb{E}_t$ together with using the independence of $\theta_{ij}(\cdot)$ with $\alpha^{\varepsilon}(\cdot)$. We omit the details here. □

Using Lemma 3.3, the truncation, the continuity of the functions and hence the boundedness, and the Cauchy-Shwartz inequality, we have

$$
\begin{aligned}
&\mathbb{E}\Big|\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s}\overline{h}'_x(x^{\varepsilon,N}(u),s_{ij})\Lambda(x^{\varepsilon,N}(u))f(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),s_{ij})\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\rho^N(x^{\varepsilon,N}(u))\widetilde{c}^{\varepsilon}_{ij}(u)du\Big|\\
&\le K\max_{1\le i\le l,1\le j\le m_i}\mathbb{E}\Big|\int_t^{t+s}\overline{h}'_x(x^{\varepsilon,N}(u),s_{ij})\Lambda(x^{\varepsilon,N}(u))f(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),s_{ij})\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\rho^N(x^{\varepsilon,N}(u))\widetilde{c}^{\varepsilon}_{ij}(u)du\Big|\\
&\le K\max_{1\le i\le l,1\le j\le m_i}\mathbb{E}^{1/2}\Big|\int_t^{t+s}\theta_{ij}(u)\widetilde{c}^{\varepsilon}_{ij}(u)du\Big|^2=O(\sqrt{\varepsilon}),
\end{aligned}
\tag{31}
$$

where K is a generic positive constant whose value may change for different appearances, and $\theta_{ij}(u)$ is chosen as the integrand without the term $\widetilde{c}^{\varepsilon}_{ij}(u)$ in the third line of (31). Denoting

$$
f(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),s_{ij})\nu^i_j(u)=\widetilde{f}(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),s_{ij}),
$$

it is easily seen that

$$
\begin{aligned}
&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s}\overline{h}'_x(x^{\varepsilon,N}(u),s_{ij})\Lambda(x^{\varepsilon,N}(u))f(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),s_{ij})\rho^N(x^{\varepsilon,N}(u))\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\nu^i_j(u)1_{\{\overline{\alpha}^{\varepsilon}(u)=i\}}du\\
&\quad=\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s}h'_x(x^{\varepsilon,N}(u),s_{ij})\Lambda(x^{\varepsilon,N}(u))\widetilde{f}(x^{\varepsilon,N}(u),s_{ij})\rho^N(x^{\varepsilon,N}(u))\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times 1_{\{\overline{\alpha}^{\varepsilon}(u)=i\}}du\\
&\quad=\int_t^{t+s}h'_x(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\widetilde{f}(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du.
\end{aligned}
$$

Moreover, the calculation in (31) leads to

$$
\begin{aligned}
&\mathbb{E}\Big|\int_t^{t+s}\overline{h}'_x(x^{\varepsilon,N}(u),\alpha^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))f(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),\alpha^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\\
&\quad-\int_t^{t+s}h'_x(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\widetilde{f}(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\Big|\\
&\quad\to 0\ \text{ as }\ \varepsilon\to 0.
\end{aligned}
$$

By the weak convergence of $(x^{\varepsilon,N}(\cdot),\overline{\alpha}^{\varepsilon}(\cdot))$ to $(x^N(\cdot),\overline{\alpha}(\cdot))$ and the Skorohod representation, with a slight abusing of notation, we may assume that

$(x^{\varepsilon,N}(\cdot),\overline{\alpha}^\varepsilon(\cdot)) \to (x^N(\cdot),\overline{\alpha}(\cdot))$ w.p.1 as $\varepsilon \to 0$. Partition the interval $[0,s]$ into $t_0 < t_1 < \cdots < t_L < t_{L+1} = s$ so that the subintervals have equal length $\varepsilon^{1-\Delta}$ for some $0 < \Delta < 1$ (i.e., $t_k = k\varepsilon^{1-\Delta}$). Then we have

$$\begin{aligned}
&\int_t^{t+s} h'_x(x^{\varepsilon,N}(u),\overline{\alpha}^\varepsilon(u))\Lambda(x^{\varepsilon,N}(u))\widetilde{f}(x^{\varepsilon,N}(u),\xi^\varepsilon(u),\overline{\alpha}^\varepsilon(u))\rho^N(x^{\varepsilon,N}(u))du \\
&\quad = \sum_{k=0}^{L}\int_{t_k}^{t_{k+1}} h'_x(x^{\varepsilon,N}(u),\overline{\alpha}^\varepsilon(u))\Lambda(x^{\varepsilon,N}(u))\widetilde{f}(x^{\varepsilon,N}(u),\xi^\varepsilon(u),\overline{\alpha}^\varepsilon(u)) \\
&\qquad\qquad \times\rho^N(x^{\varepsilon,N}(u))du \\
&\quad = \sum_{k=0}^{L}\int_{t_k}^{t_{k+1}} h'_x(x^{\varepsilon,N}(t_k),\overline{\alpha}^\varepsilon(t_k^-))\Lambda(x^{\varepsilon,N}(t_k))\widetilde{f}(x^{\varepsilon,N}(t_k),\xi^\varepsilon(u),\overline{\alpha}^\varepsilon(t_k^-)) \\
&\qquad\qquad \rho^N(x^{\varepsilon,N}(t_k))du + o(1),
\end{aligned} \tag{32}$$

where $o(1) \to 0$ in probability uniformly in t. We further have

$$\begin{aligned}
&\sum_{k=0}^{L}\int_{t_k}^{t_{k+1}} h'_x(x^{\varepsilon,N}(t_k),\overline{\alpha}^\varepsilon(t_k^-))\Lambda(x^{\varepsilon,N}(t_k))\widetilde{f}(x^{\varepsilon,N}(t_k),\xi^\varepsilon(u),\overline{\alpha}^\varepsilon(t_k^-)) \\
&\qquad\qquad \times\rho^N(x^{\varepsilon,N}(t_k))du \\
&\quad = \sum_{k=0}^{L}\varepsilon^{1-\Delta}\frac{\varepsilon^2}{t_{k+1}-t_k}\int_{\frac{t_k}{\varepsilon^2}}^{\frac{t_{k+1}}{\varepsilon^2}} h'_x(x^{\varepsilon,N}(t_k),\overline{\alpha}^\varepsilon(t_k^-))\Lambda(x^{\varepsilon,N}(t_k)) \\
&\qquad \times\widetilde{f}(x^{\varepsilon,N}(t_k),\xi(v),\overline{\alpha}^\varepsilon(t_k^-))\rho^N(x^{\varepsilon,N}(t_k))dv.
\end{aligned} \tag{33}$$

In the last line above, we used the change of variable $v = u/\varepsilon^2$. Next using a finite-valued approximation of $x^{\varepsilon,N}(t_k)$; see [19, Chapter 6.1, p.169]. That is, for any $\eta > 0$, let $\{B^\eta_\kappa : \kappa \le \kappa_\eta\}$ be a finite collection of disjoint sets of diameter small than η, then we can write

$$\begin{aligned}
&h'_x(x^{\varepsilon,N}(t_k),\overline{\alpha}^\varepsilon(t_k^-))\Lambda(x^{\varepsilon,N}(t_k))\widetilde{f}(x^{\varepsilon,N}(t_k),\xi(v),\overline{\alpha}^\varepsilon(t_k^-))\rho^N(x^{\varepsilon,N}(t_k)) \\
&\quad = \sum_{i=1}^{l}\sum_{\kappa=1}^{\kappa_\eta} 1_{\{x^{\varepsilon,N}(t_k)\in B^\eta_\kappa\}}1_{\{\overline{\alpha}^\varepsilon(t_k^-)=i\}}h'_x(x^\eta_\kappa,i)\Lambda(x^\eta_\kappa)\widetilde{f}(x^\eta_\kappa,\xi(v),i)\rho^N(x^\eta_\kappa).
\end{aligned}$$

Furthermore, the law of large numbers of mixing process implies that for each κ,

$$\begin{aligned}
&\frac{\varepsilon^2}{t_{k+1}-t_k}\int_{\frac{t_k}{\varepsilon^2}}^{\frac{t_{k+1}}{\varepsilon^2}} h'_x(x^\eta_\kappa,i)\Lambda(x^\eta_\kappa)\widetilde{f}(x^\eta_\kappa,\xi(v),i)\rho^N(x^\eta_\kappa)dv \\
&\qquad \to h'_x(x^\eta_\kappa,i)\Lambda(x^\eta_\kappa)\overline{f}(x^\eta_\kappa,i)\rho^N(x^\eta_\kappa),
\end{aligned} \tag{34}$$

where $\overline{f}$ is an average of $\widetilde{f}$ w.r.t. the stationary measure of ξ. Finally, combining (31)-(34), we arrive at (with the use of Skorohod representation), as $\varepsilon \to 0$,

$$\begin{aligned}&\int_t^{t+s} \overline{h}_x'(x^{\varepsilon,N}(u),\alpha^\varepsilon(u))\Lambda(x^{\varepsilon,N}(u))f(x^{\varepsilon,N}(u),\xi^\varepsilon(u),\alpha^\varepsilon(u))\rho^N(x^{\varepsilon,N}(u))du\\ &\quad \to h_x'(x^N(u),\overline{\alpha}(u))\Lambda(x^N(u))\overline{f}(x^N(u),\overline{\alpha}(u))\rho^N(x^N(u))du \quad \text{w.p.1.}\end{aligned} \tag{35}$$

Next, we have

$$\begin{aligned}&\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(\alpha^\varepsilon(u))du\\ &\quad = \sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(s_{ij})1_{\{\alpha^\varepsilon(u)=s_{ij}\}}du\\ &\quad = \sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(s_{ij})\nu_j^i(u)1_{\{\overline{\alpha}^\varepsilon(u)=i\}}du\\ &\qquad + \sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(s_{ij})\widetilde{c}_{ij}^\varepsilon(u)du.\end{aligned} \tag{36}$$

Similar to the previous estimates, it can be shown that

$$\mathbb{E}\Big|\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(s_{ij})\widetilde{c}_{ij}^\varepsilon(u)du\Big| \to 0 \quad \text{as} \quad \varepsilon \to 0. \tag{37}$$

Note also that

$$\begin{aligned}&\sum_{i=1}^{l}\sum_{j=1}^{m_i}\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(s_{ij})\nu_j^i(u)1_{\{\overline{\alpha}^\varepsilon(u)=i\}}du\\ &\quad = \int_t^{t+s} \operatorname{diag}(\nu^1(u),\ldots,\nu^l(u))\widehat{Q}(u)\operatorname{diag}(\mathbb{1}_{m_1}\ldots,\mathbb{1}_{m_l})\\ &\qquad\qquad \times h(x^{\varepsilon,N}(u),\cdot)(\overline{\alpha}^\varepsilon(u))du\\ &\quad = \int_t^{t+s} \overline{Q}(u)h(x^{\varepsilon,N}(u),\cdot)(\overline{\alpha}^\varepsilon(u))du.\end{aligned} \tag{38}$$

Moreover,

$$\begin{aligned}\mathbb{E}\Big|&\int_t^{t+s} \widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(\alpha^\varepsilon(u))du\\ &- \int_t^{t+s} \overline{Q}(u)h(x^{\varepsilon,N}(u),\cdot)(\overline{\alpha}^\varepsilon(u))du\Big| \to 0.\end{aligned} \tag{39}$$

Thus, the weak convergence of $(x^{\varepsilon,N}(\cdot),\overline{\alpha}^{\varepsilon}(\cdot))$ to $(x^N(\cdot),\overline{\alpha}(\cdot))$, and the Skorohod representation (with a slight abuse of notation) imply that

$$\begin{aligned}&\int_t^{t+s}\widehat{Q}(u)\overline{h}(x^{\varepsilon,N}(u),\cdot)(\alpha^{\varepsilon}(u))du\\ &\quad\to\int_t^{t+s}\overline{Q}(u)h(x^N(u),\cdot)(\overline{\alpha}(u))du\quad\text{w.p.1.}\end{aligned}\tag{40}$$

Likewise, we can examine the term on the third and the fourth lines of (27). Detailed analysis reveals that

$$\begin{aligned}\mathbb{E}\Big|&\int_t^{t+s}\int_{u/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_u^{\varepsilon}\Big[\overline{h}_x'(x^{\varepsilon,N}(u),\alpha^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi(v),\alpha^{\varepsilon}(v))\\ &\quad\times\rho^N(x^{\varepsilon,N}(u))\Big]_x' dv\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),\alpha^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\\ &-\int_t^{t+s}\int_{u/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_u^{\varepsilon}\Big[h_x'(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi(v),\overline{\alpha}^{\varepsilon}(v))\\ &\quad\times\rho^N(x^{\varepsilon,N}(u))\Big]_x' dv\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),\overline{\alpha}^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\Big|\\ &\to 0\quad\text{as}\quad\varepsilon\to 0.\end{aligned}\tag{41}$$

Furthermore, as $\varepsilon\to 0$,

$$\begin{aligned}&\int_t^{t+s}\int_{u/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_u^{\varepsilon}\Big[h_{xx}'(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi(v),\overline{\alpha}^{\varepsilon}(v))\\ &\qquad\times\rho^N(x^{\varepsilon,N}(u))\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),\alpha^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\\ &\to\frac{1}{2}\int_t^{t+s}\text{tr}[h_{xx}(x^N(u),\overline{\alpha}(u))\Lambda(x^N(u))\overline{\sigma}(x^N(u),\overline{\alpha}(u))\overline{\sigma}'(x^N(u),\overline{\alpha}(u))\\ &\qquad\times\Lambda(x^N(u))\rho^N(x^N(u))]du.\end{aligned}\tag{42}$$

Noting (10)-(13), we can show that

$$\begin{aligned}&\int_t^{t+s}\int_{u/\varepsilon^2}^{T/\varepsilon^2}\mathbb{E}_u^{\varepsilon}\Big[h_x'(x^{\varepsilon,N}(u),\overline{\alpha}^{\varepsilon}(u))[\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi(v),\overline{\alpha}^{\varepsilon}(v))\\ &\quad\times\rho^N(x^{\varepsilon,N}(u))]_x\Big]' dv\Lambda(x^{\varepsilon,N}(u))\sigma(x^{\varepsilon,N}(u),\xi^{\varepsilon}(u),\overline{\alpha}^{\varepsilon}(u))\rho^N(x^{\varepsilon,N}(u))du\\ &\to\int_t^{t+s}h_x'(x^N(u),\overline{\alpha}(u))\eta(x^N(u),\overline{\alpha}(u))du.\end{aligned}\tag{43}$$

By putting the estimates obtained thus far, the desired result thus follows.

Combining the results obtained thus, we have shown that $(x^N(\cdot),\overline{\alpha}(\cdot))$ is a solution of the martingale problem with operator $\mathcal{L}^N$, where $\mathcal{L}^N$ is defined as in (14) with $\Lambda(x)\overline{F}(x,\overline{\alpha})$, and $\Lambda(x)S(x,\overline{\alpha})\Lambda(x)$ replaced by $[\Lambda(x)\overline{F}(x,\overline{\alpha})]_N$, and $[\Lambda(x)S(x,\overline{\alpha})\Lambda(x)]_N$, respectively, where $[\Phi(x)]_N = \Phi(x)\rho^N(x)$ is the N-truncation of $\Phi(x)$ with $\Phi(x)$ being $\Lambda(x)\overline{F}(x,\overline{\alpha})$, and $\Lambda(x)S(x,\overline{\alpha})\Lambda(x)$, respectively.

Finally, we show that with the untruncated process $x^\varepsilon(\cdot)$ used, we still have that $(x^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ converges weakly to the desired limit. The last part of the proof is similar to [18, last paragraph on p. 46]. We thus omit the details. We summarize it in the next assertion.

Theorem 3.4 *Assume that conditions* (H1)-(H3) *hold. Then* $(x^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ *converges weakly to* $(x(\cdot),\overline{\alpha}(\cdot))$ *such that* $x(\cdot)$ *is the solution of* (16).

4 Examples

We have obtained the weak convergence of the Kolmogorov systems driven by a wideband noise to the Kolmogorov systems driven by Brownian motions under appropriate conditions. This section provides several examples that illustrate the systems can be treated under our framework.

4.1 SIR Models

Consider the following example

$$\begin{cases} \dot{S}^\varepsilon(t) = \big(\pi(\alpha^\varepsilon(t)) - \beta(\alpha^\varepsilon(t))S^\varepsilon(t)I^\varepsilon(t) - \mu(\alpha^\varepsilon(t))S^\varepsilon(t)\big)dt \\ \qquad\qquad\qquad\qquad\qquad\qquad +\sigma_1(\alpha^\varepsilon(t))S^\varepsilon(t)\xi^\varepsilon(t), \\ \dot{I}^\varepsilon(t) = [\beta(\alpha^\varepsilon(t))S^\varepsilon(t)I^\varepsilon(t) - [\mu(\alpha^\varepsilon(t)) + \rho(\alpha^\varepsilon(t)) + \gamma(\alpha^\varepsilon(t))]I^\varepsilon(t)]dt \\ \qquad\qquad\qquad\qquad\qquad\qquad +\sigma_2(\alpha^\varepsilon(t))I^\varepsilon(t)\xi^\varepsilon(t), \\ \dot{R}^\varepsilon(t) = (\gamma(\alpha^\varepsilon(t))I^\varepsilon(t) - \mu(\alpha^\varepsilon(t))R^\varepsilon(t))dt + \sigma_3(\alpha^\varepsilon(t))R^\varepsilon(t)\xi^\varepsilon(t), \\ S^\varepsilon(0) = s_0,\ I^\varepsilon(0) = i_0,\ R^\varepsilon(0) = r_0, \end{cases} \tag{44}$$

which is a variation of the diffusion model that was studied carefully in [6]. This model is one of the well studied classical epidemic models known as the SIR (Susceptible-Infected-Removed) model with noise added. In the SIR model, a homogeneous host population is subdivided into three epidemiologically distinct types of individuals: (S): The susceptible class, the class of those individuals who are capable of contracting the disease and becoming infective, (I): the infective class, the class of those individuals who are capable of transmitting the disease to others, (R): the removed class, the class of infected individuals who are dead, or have recovered, and are permanently immune, or are isolated. In (44), $S(t)$, $I(t)$, and $R(t)$ are the numbers of individuals at time t in classes (S), (I), and (R), respectively; $W(\cdot)$ is a standard real-valued Brownian motion, σ_1, σ_2, and σ_3 are the intensities of the white noise; π is the per capita birth rate of the population, μ is the per capita disease-free death rate and ρ is the excess per capita death rate of infective class, β is the effective per capita contact rate, and γ is per capita recovery rate of the infective individuals. As it was argued in [6], because the dynamics of the recovered class has no effect on the disease transmission dynamics, we only consider the following system:

$$
\begin{cases}
\dot{S}^\varepsilon(t) = [\pi(\alpha^\varepsilon(t)) - \beta(\alpha^\varepsilon(t))S^\varepsilon(t)I^\varepsilon(t) - \mu(\alpha^\varepsilon(t))S^\varepsilon(t)]dt \\
\qquad\qquad + \dfrac{1}{\varepsilon}\sigma_1(\alpha^\varepsilon(t))S^\varepsilon(t)\xi^\varepsilon(t), \\
\dot{I}^\varepsilon(t) = [\beta(\alpha^\varepsilon(t))S^\varepsilon(t)I^\varepsilon(t) - [\mu(\alpha^\varepsilon(t)) + \rho(\alpha^\varepsilon(t)) + \gamma(\alpha^\varepsilon(t))]I^\varepsilon(t)]dt \\
\qquad\qquad + \dfrac{1}{\varepsilon}\sigma_2(\alpha^\varepsilon(t))I^\varepsilon(t)\xi^\varepsilon(t), \\
S^\varepsilon(0) = s_0,\ I^\varepsilon(0) = i_0.
\end{cases}
\tag{45}
$$

Using Theorem 3.4, we can show that $(S^\varepsilon(\cdot), I^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(S(\cdot), I(\cdot), \overline{\alpha}(\cdot))$, where the limit is the solution of

$$
\begin{cases}
dS(t) = [\overline{\pi}(\overline{\alpha}(t)) - \overline{\beta}(\overline{\alpha}(t))S(t)I(t) - \overline{\mu}(\overline{\alpha}(t))S(t)]dt + \overline{\sigma}_1(\overline{\alpha}(t))S(t)dW(t), \\
dI(t) = [\overline{\beta}(\overline{\alpha}(t))S(t)I(t) - [\overline{\mu}(\overline{\alpha}(t)) + \overline{\rho}(\overline{\alpha}(t)) + \overline{\gamma}(\overline{\alpha}(t))]I(t)]dt \\
\qquad\qquad + \overline{\sigma}_2(\overline{\alpha}(t))I(t)dW(t), \\
S(0) = s_0,\ I(0) = i_0.
\end{cases}
\tag{46}
$$

4.2 Replicator Dynamics

The study of the replicator dynamic systems was initially proposed by Taylor and Jonker in [31]. Since then significant progress has been made in biology (see [15, 26]), economics (see [36]), as well as in optimization and control for a variety of systems (see [3, 27, 34]).

Consider replicator dynamics for a game with d strategies, involving wideband noise. The system is given by

$$\begin{aligned}\dot{x}_i^\varepsilon(t) &= x_i^\varepsilon(t)\Big(f_i(x^\varepsilon(t),\alpha^\varepsilon(t)) - \frac{1}{\mathsf{X}}\sum_{j=1}^d x_j^\varepsilon(t) f_j(x^\varepsilon(t),\alpha^\varepsilon(t))\Big)\\ &\quad + \frac{1}{\varepsilon}x_i^\varepsilon(t)\Big(\sigma_i(\alpha^\varepsilon(t)\xi_i^\varepsilon(t) - \frac{1}{\mathsf{X}}\sum_{j=1}^d \sigma_j(\alpha^\varepsilon(t))x_j^\varepsilon(t)\xi_j^\varepsilon(t)\Big);\ i=1,\dots,d,\end{aligned} \tag{47}$$

where X is the size of the populations; $x_i(t)$ is the portion of population that has selected the i^{th} strategy and the distribution of the whole population among the strategy; $x(t) = (x_1(t),\dots,x_d(t))'$; $\xi^\varepsilon(t)$ is a wideband noise satisfying condition (H1); the fitness functions $f_i(\cdot,\cdot) : \mathbb{R}_+^d \times \mathcal{M} \to \mathbb{R}$, $i = 1,\dots,d$ are the payoffs obtained by the individuals playing the i^{th} strategy; $x(0) = x_0$ is the initial value.

Under the conditions of Theorem 3.4, $(x^\varepsilon(\cdot),\overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(x(\cdot),\overline{\alpha}(\cdot))$ such that $x(\cdot) = (x_1(\cdot),\dots,x_d(\cdot))$ is the solution of

$$\begin{aligned}dx_i(t) &= x_i(t)\Big(\overline{f}_i(x(t),\overline{\alpha}(t)) - \frac{1}{\mathsf{X}}\sum_{j=1}^d x_j(t)\overline{f}_j(x(t),\overline{\alpha}(t))\Big)dt\\ &\quad + x_i(t)\Big(\overline{\sigma}_i(\overline{\alpha}(t))dw_i(t) - \frac{1}{\mathsf{X}}\sum_{j=1}^d \overline{\sigma}_j(\overline{\alpha}(t))x_j(t)dw_j(t)\Big);\ i=1,\dots,d,\end{aligned} \tag{48}$$

and $w(\cdot) = (w_1(\cdot),\dots,w_d(\cdot))'$ is an ℓ-dimensional standard Brownian motion. In recent years, much attention has been devoted to stochastic replicator dynamic systems. One of the most recent one is [23], which considered a system with time delays and rather general dynamics. Our work in wideband noise driven systems enabled the handling of systems that are not necessarily Markovian and the perturbing noise is not necessarily white.

4.3 Lotka-Volterra Model

Consider a stochastic Lotka-Volterra ecosystem in random environments described by the following random differential equation with regime switching in the form

$$dx_i(t) = x_i(t)\left\{\Big[r_i(\alpha^\varepsilon(t)) - \sum_{j=1}^d a_{ij}(\alpha^\varepsilon(t))x_j(t)\Big]dt + \sigma_i(\alpha^\varepsilon(t))\xi_i^\varepsilon(t)\right\},\quad i=1,2,\dots,d, \tag{49}$$

where $\alpha^\varepsilon(\cdot)$ is a continuous-time finite sate Markov chain with state space $\mathcal{M}$ satisfying conditions in Section 2.3. The $\xi^\varepsilon(\cdot)$ is a wideband noise satisfies condition (H1). The processes $\xi^\varepsilon(\cdot)$ and $\alpha^\varepsilon(\cdot)$ are independent. Denote

$$b_i(\alpha) = r_i(\alpha) - \frac{1}{2}\sigma_i^2(\alpha)$$

for each $i = 1, 2, \ldots, d$.

For each $\alpha \in \mathcal{M}$, $b(\alpha) = (b_1(\alpha), \ldots, b_d(\alpha))'$, $A(\alpha) = (a_{ij}(\alpha))$, $\Sigma(\alpha) = \text{diag}(\sigma_1(\alpha), \ldots, \sigma_d(\alpha))$ represent different growth rates, community matrices, and noise intensities in different external environments, respectively; see [40]. We assume $b_i(\alpha) > 0$ for each $\alpha \in \mathcal{M}$ and each $i = 1, \ldots, d$. Denote $x^\varepsilon(t) = (x_1^\varepsilon(t), \ldots, x_d^\varepsilon(t))'$ and $x(t) = (x_1(t), \ldots, x_d(t))'$. Then using the conditions in Theorem 3.4, we can show that $(x^\varepsilon(\cdot), \overline{\alpha}^\varepsilon(\cdot))$ converges weakly to $(x(\cdot), \overline{\alpha}(\cdot))$ such that $x(\cdot)$ is the solution of the stochastic differential equation

$$dx_i(t) = x_i(t)\left\{\left[\overline{r}_i(\overline{\alpha}(t)) - \sum_{j=1}^{d} \overline{a}_{ij}(\overline{\alpha}(t))x_j(t)\right]dt + \overline{\sigma}_i(\overline{\alpha}(t))dw_i(t)\right\}, \quad i = 1, 2, \ldots, d, \tag{50}$$

where $w(\cdot) = (w_1(\cdot), \ldots, w_d(\cdot))'$ is a d-dimensional standard Brownian motion that is independent of the Markov chain $\overline{\alpha}(\cdot)$.

5 Further Remarks

This paper examines stochastic Kolmogorov systems, in which in lieu of Brownian motions, a wideband noise process is serving as the driving noise. In addition, there is a singularly perturbed continuous-time Markov chain that is subject to both strong and weak interactions. In addition to the analysis, several examples are provided.

One of the fundamental issues for systems arising in ecology and population biology is: What are the minimal conditions needed for the population to be permanent (existing for a long time)? Likewise, what are the conditions for the population to reach the extinction. As alluded to in the introduction, usually, this study is based on the Markovian formulation of the underlying systems.

An important question is: Suppose that the Markovian formulation is no longer available and the perturbing noise is not white noise but only approximation of white noise, can we still study the permanence and extinction questions. It is difficult to study Kolmogorov systems in general, but it is possible to study an example where the functions involved and the coefficients are explicitly given. We first recall the following definitions.

We say that the process $X(\cdot)$ is strongly stochastically persistent if for any $\eta > 0$, there exists an $R > 0$ such that for any $x \in \mathbb{R}^d_+$,

$$\liminf_{t\to\infty} \mathbb{P}_x\Big(R^{-1} \leq |X(t)| \leq R\Big) \geq 1 - \eta,$$

where $\mathbb{P}_x$ denotes the probability with initial data $X(0) = x$.

For each $x \in \mathbb{R}^d_+$, we say the population goes extinct with probability p_x if

$$\mathbb{P}_x\Big(\lim_{t\to\infty} X(t) = 0\Big) = p_x,$$

where p_x denotes the dependence on the initial data x. We say that the population $X(t)$ goes extinct if for all $x \in \mathbb{R}^d_+$,

$$\mathbb{P}_x\Big(\lim_{t\to\infty} X(t) = 0\Big) = 1.$$

Consider an SIR model with no switching but with wideband noise perturbations. The results in our work [6] with no switching (i.e., the parameters π, β, ρ, γ, σ_1, σ_2, and σ_3 are all constants) shows that for the limit system driven by a white noise, the Lyapunov exponent associated with the system of interest is given by

$$\lambda = \frac{\pi\beta}{\mu} - \big(\mu + \rho + \gamma + \frac{\sigma_2^2}{2}\big).$$

If $\lambda < 0$, then for any initial value $(S(0), I(0)) = (u, v) \in \mathbb{R}^{2,\circ}_+$ (the interior of the first quadrant), we have

$$\limsup_{t\to\infty} \frac{\ln I_{u,v}(t)}{t} \leq \lambda \ \text{ a.s.}, \tag{51}$$

where $I_{u,v}(t)$ denotes the solution of (46) with initial data $(S(0), I(0)) = (u, v)$. In addition $S_{u,v}(t)$ has an invariant measure. Now, suppose that we are using the setting as in this paper, but with the $Q^\varepsilon(t) = \widetilde{Q}(t)/\varepsilon + \widehat{Q}(t)$, where $\widetilde{Q}(t)$ is weakly irreducible. That is, we only have one irreducible class. In this case, we can then use the methods of perturbed Lyapunov functions [18, 39] to study the asymptotic behavior $(S^\varepsilon(t), I^\varepsilon(t))$.

For more general switching process as given in this paper, an interesting and important problem is to handle the study on permanence and extinction for such systems. Much more thoughts and careful analysis are needed.

Remark

Professor Vasile Dragan has made significant contributions to the fields of differential equations, dynamic systems, and control and optimization. Many of his works, for example [8], have made major impact to the fields. On the occasion of the celebration of his 70th birthday, we send him our best wishes.

References

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

[2] G. Blankenship and G.C. Papanicolaou, Stability and control of stochastic systems with wide-band noise disturbances, I, *SIAM J. Appl. Math.* **34** (1978), 437–476.

[3] I. Bomze, M. Pelillo, and V. Stix, (2000). Approximating the maximum weight clique using replicator dynamics. *IEEE Trans Neural Network.* **11** 1228–1241.

[4] X. Chen, Z.-Q. Chen, K. Tran, and G. Yin, Properties of switching jump diffusions: Maximum principles and Harnack inequalities, *Bernoulli*, **25** (2019), 1045–1075.

[5] P.J. Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, NY, 1977.

[6] N.T. Dieu, D.H. Nguyen, N.H. Du, and G. Yin, Classification of asymptotic behavior in a stochastic SIR model, *SIAM J. Appl. Dynamic Sys.*, **15** (2016), 1062–1084.

[7] N.T. Dieu, N.H. Du, D.H. Nguyen, and G. Yin, Protection zones for survival of species in random environment, *SIAM J. Appl. Math.*, **76** (2016), 1382–1402.

[8] V. Dragan, T. Morozan, A.-M. Stoica, *Mathematical Methods in Robust Control of Linear Stochastic Systems*, Springer, New York, 2006.

[9] N.H. Du, D.H. Nguyen, and G. Yin, Conditions for permanence and ergodicity of certain stochastic predator-prey models, *J. Appl. Probab.*, **53** (2016), 187–202.

[10] N.T. Dung, A stochastic Ginzburg-Landau equation with impulsive effects, Physica A, 392 (9)(2013), 1962-1971.

[11] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, J. Wiley, New York, 1986.

[12] V.L. Ginzburg, L.D. Landau, On the theory of superconductivity, Zh. Eksp. Teor. Fiz. 20 (1950) 1064–1082.

[13] X. Han and P.E. Kloeden, *Random Ordinary Differential Equations and Their Numerical Solution*, Springer, Singapore, 2017.

[14] A. Hening and D. Nguyen, Coexistence and extinction for stochastic Kolmogorov systems, *Ann. Appl. Probab.*, **28** 2018, 1893–1942.

[15] J. Hofbauer and K. Sigmund, (1998). *Evolutionary Games and Population Dynamics.* Cambridge Univ. Press.

[16] J. Hofbauer and L.A. Imhof, (2009). Time averages, recurrence and transience in the stochastic replicator dynamics. *Ann. Appl. Probab.* **19** 1347–1368.

[17] R.Z. Khasminskii and G. Yin, On averaging principles: An asymptotic expansion approach, *SIAM J. Math. Anal.*, **35** (2004), 1534–1560.

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

[19] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed., Springer-Verlag, New York, NY, 2003.

[20] X. Mao, *Stochastic Differential Equations and Applications*, 2nd ed., Horwood, Chichester, UK, 2007.

[21] H. Matsuda, N. Ogita, A. Sasaki, and K, Sato, Statistical mechanics of population: The lattice Lotka-Volterra model, *Progress Theoretical Physics*, **88** (1992), 1035–1049.

[22] D.H. Nguyen and G. Yin, Coexistence and exclusion of stochastic competitive Lotka-Volterra models, *J. Differential Eqs.*, **262** (2017), 1192-1225.

[23] D.H. Nguyen, N.N. Nguyen, and G. Yin, Stochastic functional Kolmogorov equations, preprint, 2019.

[24] D. Nguyen, N. Nguyen, and G. Yin, General nonlinear stochastic systems motivated by chemostat models: Complete characterization of long-time behavior, optimal controls, and applications to wastewater treatment, *Stochastic Process. Appl.*, **130** (2020) 4608–4642.

[25] D. Nguyen, G. Yin, and C. Zhu, Long-term analysis of a stochastic SIRS model with general incidence rates, *SIAM J. Appl. Math.*, **80** (2020), 814–838.

[26] M.A. Nowak and K. Sigmund, Evolutionary dynamics of biological games. *Science*, **303** (2004), 793–799.

[27] G. Obando, A. Pantoja, and N. Quijano, (2014). Building temperature control based on population dynamics. *IEEE Trans. Control Syst. Tech.* **22** 404–412.

[28] S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser, Boston, MA, 1994.

[29] H.A. Simon and A. Ando, Aggregation of variables in dynamic systems, *Econometrica*, **29** (1961), 111–138.

[30] P. Shi and V. Dragan, Asymptotic H^∞ control of singularly perturbed systems with parametric uncertainties, *IEEE Trans. Automat. Control* **44** (1999), 1738–1742.

[31] P.D. Taylor and L.B. Jonker, (1978). Evolutionary stable strategies and game dynamics. *Math Bio.* **40** 145–156.

[32] K. Tran and G. Yin, Numerical methods for optimal harvesting strategies in random environments under partial observations, *Automatica*, **70** (2016), 74–85.

[33] K. Tran and G. Yin, Optimal harvesting strategies for stochastic ecosystems, *IET Control Theory Appl.*, **11** (2017), 2521–2530.

[34] H. Tembine, E. Altman, R. El-Azouzi, and Y. Hayel, (2010). Evolutionary games in wireless networks. *IEEE Trans Syst Man Cybern. Part B.* **40** 634–646.

[35] N.G. van Kampen, Remarks on non-Markov processes, *Brazilian J. Physics*, **28** (1998), 90–96.

[36] J. Weibull, (1997). *Evolutionary Game Theory.* MIT press, Cambridge, MA, USA.

[37] Z. Wu, G. Yin, and D.X. Lei, A class of generalized Ginzburg-Landau equations with random switching, *Physica A*, **506** (2018) 324–336.

[38] G. Yin and Z. Wen, Stochastic Kolmogorov systems driven by wideband noises, *Physica A*, **531** (2019), 121746.

[39] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Two-Time-Scale Approach*, 2nd Ed., Springer, New York, 2013.

[40] C. Zhu and G. Yin, On hybrid competitive Lotka-Volterra ecosystems, *Nonlinear Analysis*, **71** (2009), e1370–e1379.