Perturbations of linear delay differential equations at the verge of instability

N. Lingala* and N. Sri Namachchivaya
Department of Aerospace Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, USA

(Received 27 October 2015; revised manuscript received 3 May 2016; published 2 June 2016)

The characteristic equation for a linear delay differential equation (DDE) has countably infinite roots on the complex plane. This paper considers linear DDEs that are on the verge of instability, i.e., a pair of roots of the characteristic equation lies on the imaginary axis of the complex plane and all other roots have negative real parts. It is shown that when small noise perturbations are present, the probability distribution of the dynamics can be approximated by the probability distribution of a certain one-dimensional stochastic differential equation (SDE) without delay. This is advantageous because equations without delay are easier to simulate and one-dimensional SDEs are analytically tractable. When the perturbations are also linear, it is shown that the stability depends on a specific complex number. The theory is applied to study oscillators with delayed feedback. Some errors in other articles that use multiscale approach are pointed out.

DOI: 10.1103/PhysRevE.93.062104

I. INTRODUCTION

Delay differential equations (DDEs) arise when the evolution of a variable at any time depends on the history of the variable. The evolution of many physical systems depends on their history owing to finite conduction velocities. Naturally, these systems are modeled by DDEs. Delay differential equations arise in many areas: biological systems, population dynamics, machining processes, viscoelasticity, laser optics, etc. See [1] for a description of some examples. Many models of physiological systems, disease models, and population dynamics involve DDE; see the Mackey-Glass equation [2] for example.

The subject of this paper is linear DDEs at the verge of instability. For example, consider the equation

$$\dot{x}(t) = \kappa x(t-1). \tag{1}$$

Seeking a solution of the form $x(t) = e^{t\lambda}$, we find that λ must satisfy the characteristic equation $\lambda - \kappa e^{-\lambda} = 0$. When $\kappa \in (-\frac{\pi}{2},0)$, all roots of the characteristic equation have negative real parts (see Corollary 3.3 in [3]). When $\kappa = -\frac{\pi}{2}$ a pair of roots $\pm i\frac{\pi}{2}$ is on the imaginary axis and all others have negative real parts. When $\kappa < -\frac{\pi}{2}$ some of the roots have a positive real part. Hence, the system (1) is on the verge of instability at $\kappa = -\frac{\pi}{2}$. We study the effect of perturbations on such systems, for example,

$$\dot{x}(t) = \left(-\frac{\pi}{2} + \varepsilon\xi(t)\right)x(t-1),$$

where ξ is a noise and $\varepsilon \ll 1$ is the strength of the perturbation.

Such instability situations arise, for example, in machining processes. An oscillator of the form

$$\ddot{q}(t) + 2\zeta\dot{q}(t) + p^2 q(t) = -\kappa p^2[q(t) - q(t-r)] \tag{2}$$

is used to describe a phenomenon called regenerative chatter in machining processes [4]. The model is as follows. A cutting tool is placed on a workpiece that is attached to a shaft rotating with time period r. The tool vibrates as it cuts the material from the workpiece. Let $q(t)$ describe the position of a point on the machine tool. The force acting on the tool is proportional to the depth of the chip being cut and the depth is approximated as the difference between the present position $q(t)$ of the tool and its position one revolution earlier $q(t-r)$. The coefficient κ is a force coefficient that depends on, among other factors, the width of cut. It is known that, for a fixed r, there exists a critical κ_c such that the amplitude q of the oscillator decreases exponentially if $\kappa < \kappa_c$ and increases exponentially if $\kappa > \kappa_c$. When $\kappa = \kappa_c$ oscillations of constant amplitude persist. This oscillatory behavior is called chatter. In machining, the goal is to have a large rate of cut. The greater the rate, the larger κ is, and chatter occurs when κ is larger than a critical value resulting in poor surface finish. Researchers explored the possibility of achieving chatter suppression by varying structural parameters of the tool such as damping and stiffness (see [5,6]). Suppose there are small random perturbations in the natural frequency p in (2) such that $p = p_0[1 + \varepsilon\sigma(\xi(t))]$, where σ is a mean-zero function of the noise ξ and $\varepsilon \ll 1$ is the strength of the perturbation; then on expanding in powers of ε and discarding terms of higher order, we have

$$\begin{aligned}\ddot{q}(t) &+ 2\zeta\dot{q}(t) + p_0^2 q(t)\\ &= -\kappa p_0^2[q(t) - q(t-r)] + \varepsilon\sigma(\xi(t))\big[-2(1+\kappa)p_0^2 q(t)\big]\\ &\quad + \varepsilon\sigma(\xi(t))\big[2\kappa p_0^2 q(t-r)\big],\end{aligned} \tag{3}$$

which can be studied as a perturbation of (2). Also, small random perturbations in the properties of the material being cut could affect the tool dynamics (see [7]).

Delay equations on the verge of instability arise also, for example, in the study of eye pupil [8] and act of human balancing [9]. In [10] Gaudreault *et al.* make a case for studying the effect of noise on oscillators with delayed feedback. As a prototypical oscillator they consider the van der Pol model

$$\begin{aligned}\ddot{q}(t) &+ \omega_0^2 q(t) + \eta q(t-r)\\ &= \beta\dot{q}(t) + \kappa\dot{q}(t-r) - bq^2(t)\dot{q}(t) + q(t)\xi(t),\end{aligned} \tag{4}$$

with ξ a Gaussian white noise with zero mean and variance $\langle\xi(t)\xi(t')\rangle = 2D\delta(t-t')$.

Deterministic and stochastic DDEs have been well studied in the literature; see, for example, Refs. [11] (deterministic) and [12] (stochastic). Deterministic DDEs at the verge of

*lingala1@illinois.edu

instability are also well studied; see [13] for the averaging approach and [14,15] for the multiscale approach. Stochastic DDEs at the verge of instability, with the noise being white, have been studied by employing the multiscale approach in [10,16–18]; the averaging approach in [19–21]; and the center-manifold approach in [22].

However, [10,16–18] have committed serious errors in the analysis. These are pointed out in Appendix A. Appendixes A 1 (errors of [16,17]) and A 2 (errors of [10,18]) can be read without further preparation. References [19–21] restrict their analysis to noise being white and do not consider stronger deterministic perturbations as considered here in Sec. V. Reference [23] considers a different kind of instability (one root of the characteristic equation is zero and all other roots have negative real parts), which is reviewed in Sec. VII.

This article deals with systems that can be studied as perturbations of linear DDEs at the verge of instability. In recent articles [24,25] we have shown rigorously that, under certain conditions, the dynamics of such systems forced by white noise can be approximated (in a distributional sense) by the dynamics of a one-dimensional stochastic differential equation (SDE) without delay. The purpose of this article is threefold.

The first is to exploit the results of [24,25] to show how the analysis of systems at the verge of instability can be simplified. The advantage arises because equations without delay are easier to simulate and one-dimensional SDEs are analytically tractable. References [24,25] deal rigorously with scalar systems forced by white noise. In this article we give (without proofs) explicit formulas for the approximating dynamics of vector-valued systems forced by white noise [equations of the form (7) and (43)].

The approach taken in this article is similar to those in [19–21], in the sense that all use the spectral theory for DDEs and averaging. However, [19–21] consider specific applications of the equations of the form (7), but do not consider the stronger perturbations as in Eq. (43). When dealing with Eq. (43), the averaging approach that we take does not assume the existence of a center manifold (rigorous results about a center manifold for stochastic DDEs are not known[1]). Further, the formulas (55) and (56) presented here, regarding the stronger perturbations G_q in (43), are of independent interest. When applied in the deterministic DDE setting, they provide an alternate way to compute the effect of center-manifold terms on the amplitude of the critical mode (more details are provided in Sec. V).

The second purpose is to point out the errors in existing approaches that deal with the white noise case.

The third is to study systems forced by other general kind of noises (for example, a continuous-time two-state Markov chain). Theoretical results for this case [equations of the form (8)] are dealt with in Sec. VI. A sketch of the proof of the main result (Theorem 3) is provided in Appendix D.

These claims would become more clear after the next two sections where the mathematical framework is explained. Also, in the case where the perturbations are also linear, a complex number is identified that alone dictates the stability of the system.

The organization of the rest of the paper is given at the end of the next section, after presenting the preliminaries.

II. MATHEMATICAL SETUP OF DDES

A. Notation

We use the following notation throughout: $\mathbb{R}$ denotes the set of real numbers and $\mathbb{C}$ denotes the set of complex numbers; $e^{\lambda\bullet}$ means a function whose evaluation at $\theta \in \mathbb{R}$ is $e^{\lambda\theta}$; an asterisk as a superscript indicates a transpose; $\bar{z}$ is the complex conjugate of z; $\underline{v} \in \mathbb{R}^n$ means that $\underline{v}$ is an $n \times 1$ matrix with each entry in $\mathbb{R}$ and $\underline{v} \in \mathbb{R}^{n*}$ means that $\underline{v}$ is a $1 \times n$ matrix with each entry in $\mathbb{R}$. The line underneath serves as a reminder that the quantity is multidimensional. The notation is similar for $\mathbb{C}^n$ and $\mathbb{C}^{n*}$.

B. Equations considered in the article

Let $x(t)$ be an $\mathbb{R}^n$-valued process governed by a DDE with maximum delay r. The evolution of x at each time t requires the history of the process in the time interval $[t-r,t]$. So the state space can be taken as $\mathcal{C} := C([-r,0];\mathbb{R}^n)$, the space[2] of continuous functions on the interval $[-r,0]$ with values in $\mathbb{R}^n$. At each time t, denote the $[t-r,t]$ segment of x by $\Pi_t x$, i.e., $\Pi_t x \in \mathcal{C}$ and

$$\Pi_t x(\theta) = x(t+\theta) \quad \text{for } \theta \in [-r,0].$$

Now a linear DDE can be represented in the form

$$\begin{aligned} \dot{x}(t) &= L_0(\Pi_t x), \quad t \geqslant 0 \\ \Pi_0 x &= \varphi \in \mathcal{C}, \end{aligned} \tag{5}$$

where $L_0 : \mathcal{C} \to \mathbb{R}^n$ is a continuous linear mapping on $\mathcal{C}$ and φ is the initial history required. For example, $\dot{x}(t) = -\frac{\pi}{2}x(t-1)$ can be represented using the linear operator given by $L_0(\eta) = -\frac{\pi}{2}\eta(-1)$ for $\eta \in \mathcal{C}$.

We assume that there exists a bounded matrix-valued function $\mu : [-r,0] \to \mathbb{R}^{n\times n}$, continuous from the left on the interval $(-r,0)$ and normalized with $\mu(0) = 0_{n\times n}$, such that

$$L_0\eta = \int_{[-r,0]} d\mu(\theta)\eta(\theta) \quad \text{for } \eta \in \mathcal{C}. \tag{6}$$

This is not a restriction: Every continuous linear operator L_0 has such a representation. For example, $\dot{x} = -\frac{\pi}{2}x(t-1)$ can be represented by

$$\mu(\theta) = \begin{cases} \frac{\pi}{2}, & \theta = -r \\ 0, & \theta > -r. \end{cases}$$

[1]However, see [26] for related results. One of the special cases of Theorem 4.1 of [26] is the following: In the case that zero is a fixed point of a stochastic DDE and the stochastic system linearized about zero does not have zero as a Lyapunov exponent, local stable and unstable manifolds exist. These manifolds are the set of initial conditions that converge to or diverge from zero at an exponential rate.

[2]The space $\mathcal{C}$ is Banach space when equipped with a sup norm: $\|\eta\| := \sup_{\theta\in[-r,0]} |\eta(\theta)|$ for $\eta \in \mathcal{C}$.

This article deals with perturbations of linear DDEs, i.e., equations of the form

$$dx(t) = L_0(\Pi_t x)dt + \varepsilon^2 G(\Pi_t x)dt + \varepsilon F(\Pi_t x)dW(t),\quad \Pi_0 x = \varphi \in \mathcal{C}, \tag{7}$$

where $F,G : \mathcal{C} \to \mathbb{R}^n$ are possibly nonlinear, W is a $\mathbb{R}$-valued Wiener process, and $\varepsilon \ll 1$ is a small number signifying perturbation. The following equations are also considered:

$$dx(t) = L_0(\Pi_t x)dt + \varepsilon^2 G(\Pi_t x)dt + \varepsilon\sigma(\xi(t))F(\Pi_t x)dt,\quad \Pi_0 x = \varphi \in \mathcal{C}, \tag{8}$$

where $F,G : \mathcal{C} \to \mathbb{R}^n$ are possibly nonlinear, ξ is a noise process (satisfying some assumptions), and σ is a mean-zero function of the noise ξ. For example, one can have ξ as a finite-state Markov chain.

As an example, consider $\dot{\tilde{x}} = \kappa\tilde{x}(t-1) - \tilde{x}^3(t)$, where κ has small perturbations about $-\frac{\pi}{2}$ according to $\kappa = -\frac{\pi}{2} + \varepsilon\sigma(\xi(t)) + \varepsilon^2$, where ξ is a noise. Then $x(t) = \varepsilon^{-1}\tilde{x}(t)$ can be put in the form (8) with $L_0(\eta) = -\frac{\pi}{2}\eta(-1)$, $F(\eta) = \eta(-1)$, and $G(\eta) = -\eta^3(0) + \eta(-1)$.

The operator L_0 is assumed to be such that the unperturbed system (5) is on the verge of instability, i.e., L_0 satisfies the following assumption.

Assumption 1. Define

$$\Delta(\lambda) = \lambda I_{n\times n} - \int_{[-r,0]} d\mu(\theta)e^{\lambda\theta},$$

where I is the identity matrix. The characteristic equation

$$\det[\Delta(\lambda)] = 0, \quad \lambda \in \mathbb{C} \tag{9}$$

has a pair of purely imaginary solutions $\pm i\omega_c$ and all other solutions[3] have negative real parts.

Since (7) and (8) would be studied as perturbations of the linear DDE (5), a brief overview of the unperturbed system (5) is given now.

C. Unperturbed system (5)

The content in this section can be found in [11] (see Chap. 7) and [27] (see Chap. 4).

1. Projection onto eigenspaces

The space $\mathcal{C}$ can be split as $\mathcal{C} = P \oplus Q$, where P is the eigenspace of the critical eigenvalues $\pm i\omega_c$. Since P corresponds to the critical eigenvalues $\pm i\omega_c$, the projection of the dynamics of the unperturbed system onto P is purely oscillatory with frequency ω_c. Since Q corresponds to the eigenvalues with a negative real part, the projection of the dynamics of the unperturbed system onto Q decays exponentially fast.

Here we show, given an $\eta \in \mathcal{C}$, how to find the projection onto the space P. For details, see [11] (Chap. 7) and [27] (Chap. 4).

Any $\eta \in \mathcal{C}$ can be written as $\eta = \pi\eta + (I-\pi)\eta$, where $\pi\eta \in P$ and $(I-\pi)\eta \in Q$. Here π is the projection operator $\pi : \mathcal{C} \to P$ and I is the identity operator. The projection π can be constructed as follows. Let

$$\Phi = [\Phi_1,\Phi_2], \quad \Phi_1(\bullet) = \underline{d}e^{i\omega_c\bullet}, \quad \Phi_2(\bullet) = \bar{\underline{d}}e^{-i\omega_c\bullet}, \tag{10}$$

where $\underline{d} \in \mathbb{C}^n$ is chosen such that

$$\Delta(i\omega_c)\underline{d} = 0_{n\times 1}. \tag{11}$$

Note that each Φ_i belongs to $C([-r,0];\mathbb{C}^n)$. Define the bilinear form $\langle\cdot,\cdot\rangle : C([0,r];\mathbb{C}^{n*}) \times C([-r,0];\mathbb{C}^n) \to \mathbb{C}$, given by

$$\langle\psi,\eta\rangle := \psi(0)\eta(0) - \int_{-r}^{0}\int_{0}^{\theta}\psi(s-\theta)d\mu(\theta)\eta(s)ds. \tag{12}$$

Let

$$\Psi = \begin{bmatrix}\Psi_1\\ \Psi_2\end{bmatrix}, \quad \Psi_1(\bullet) = c\underline{d_2}e^{-i\omega_c\bullet}, \quad \Psi_2(\bullet) = \bar{c}\bar{\underline{d_2}}e^{i\omega_c\bullet}, \tag{13}$$

where $\underline{d_2} \in \mathbb{C}^{n*}$ is chosen such that

$$\underline{d_2}\Delta(i\omega_c) = 0_{1\times n} \tag{14}$$

and the constant c is chosen such that

$$\langle\Psi_i,\Phi_j\rangle = \delta_{ij}. \tag{15}$$

(Here $\delta_{ij} = 1$ if $i = j$ and zero if $i \neq j$.)

Writing $\langle\Psi,\eta\rangle = [{}^{\langle\Psi_1,\eta\rangle}_{\langle\Psi_2,\eta\rangle}]$, we obtain for the projection $\pi : \mathcal{C} \to P$,

$$\pi(\eta) = \Phi\langle\Psi,\eta\rangle = \Phi_1\langle\Psi_1,\eta\rangle + \Phi_2\langle\Psi_2,\eta\rangle. \tag{16}$$

Note that $\langle\Psi_1,\eta\rangle$ and $\langle\Psi_2,\eta\rangle$ are complex conjugates and so are Φ_1 and Φ_2.

2. Behavior of the solution on the eigenspaces

The solution to the unperturbed system (5) can be written as

$$\Pi_t x = \pi\Pi_t x + (I-\pi)\Pi_t x = \Phi z(t) + y_t,$$

where $z(t) = \langle\Psi,\Pi_t x\rangle$ and $y_t = \Pi_t x - \Phi z(t)$. Note that $z \in \mathbb{C}^2$ is a two-component vector with $z_2 = \bar{z_1}$, and $\Phi z(t) \in P$ and $y_t \in Q$. It can be shown that

$$\dot{z}(t) = Bz(t), \quad B = \begin{bmatrix}i\omega_c & 0\\ 0 & -i\omega_c\end{bmatrix}, \tag{17}$$

i.e., z oscillates with constant amplitude and frequency ω_c. So $2z_1z_2$ is a constant in time. Further, it can be shown that $\|y_t\|$ decreases[4] to zero exponentially fast (because the dynamics on Q is governed by eigenvalues with negative real parts).

[3] Typically there are countably infinite other roots.

[4] This is the sup norm on $\mathcal{C}$.

D. Perturbed systems (7) and (8)

Define the function $\mathfrak{h} : \mathcal{C} \to \mathbb{R}$ by

$$\mathfrak{h}(\eta) := 2\langle \Psi_1,\eta\rangle\langle \Psi_2,\eta\rangle, \quad \eta \in \mathcal{C}. \tag{18}$$

As noted above,

$$2z_1(t)z_2(t) = 2\langle \Psi_1,\Pi_t x\rangle\langle \Psi_2,\Pi_t x\rangle = \mathfrak{h}(\Pi_t x)$$

is a constant for the unperturbed system (5). When we deal with the perturbed system (7) or (8), the quantity $\mathcal{H}(t) := \mathfrak{h}(\Pi_t x)$ evolves much more slowly than x and z_i. In (7), because a Wiener process has the property that the rescaled process $t \mapsto \varepsilon W(t/\varepsilon^2)$ has the same probability distribution as that of a Wiener process, the noise perturbations take $O(1/\varepsilon^2)$ time to significantly affect the $\mathcal{H}$ dynamics. Also, the perturbation G is of strength ε^2. Hence, significant changes in $\mathcal{H}$ occur only in times of order $1/\varepsilon^2$. In (8), even though the strength of the noise perturbation is ε, because σ is a mean-zero function of the noise, significant changes in $\mathcal{H}$ occur only in times of order $1/\varepsilon^2$.

Our claim is that, under certain conditions on the coefficients F and G, the probability distribution of the process $\mathcal{H}(t/\varepsilon^2)$ converges to the probability distribution of a SDE without delay. Because of the nature of decay on Q, $\|y_t\|$ decays to small values exponentially fast and so studying $\mathcal{H}$ is enough to obtain a good approximation to the behavior of x in (7) and (8). How to obtain the SDE is shown in later sections.

Remark 1. The reason why studying $\mathcal{H}$ would be useful is the following. For the moment assume that the part of the solution in the stable eigenspace Q is zero, i.e., $\Pi_t x = \Phi z(t)$ and $(I-\pi)\Pi_t x = 0$. Then, for the jth component of x we have $x_j(t) = [\Pi_t x(0)]_j = (\underline{d})_j z_1(t) + (\bar{\underline{d}})_j z_2(t)$, where $\underline{d}$ is chosen in (10). Noting that $z_2 = \bar{z}_1$ and that the dynamics of z_i is predominantly oscillatory with frequency ω_c, we find that the dynamics of x_j is predominantly oscillatory with amplitude $2|(\underline{d})_j z_1|$ or, equivalently, $\sqrt{4(\underline{d})_j(\bar{\underline{d}})_j z_1 z_2} = |(\underline{d})_j|\sqrt{4z_1z_2} = |(\underline{d})_j|\sqrt{2\mathcal{H}}$. Hence the magnitude of $\mathcal{H}$ indicates the amplitude of oscillation of x [usually the amplitude might differ from $|(\underline{d})_j|\sqrt{2\mathcal{H}}$ by a slight amount because the part of the solution in Q, i.e., $(1-\pi)\Pi_t x$, is not exactly zero].

The rest of the paper is organized as follows. Equations of the form (7) are considered in Sec. III and convergence of the probability distribution of the $\mathcal{H}$ process for such equations is stated in Theorem 1. Examples illustrating the usefulness of Theorem 1 are given in Sec. IV. Equations similar to (7) but with stronger perturbations [Eq. (43)] are considered in Sec. V and convergence of the probability distribution of the $\mathcal{H}$ process for such equations is stated in Theorem 2. The physical arguments leading to Theorem 2 are explained in Sec. V. However, the application-oriented reader can utilize Remark 4 to immediately apply Theorem 2 (notation is available in Sec. V A). Analogous results for equations of the form (8) are in Sec. VI.

A crucial role is played by the vector $\Psi(0)$. So the symbol $\hat{\Psi}$ is reserved for $\Psi(0)$:

$$\hat{\Psi} \stackrel{\text{def}}{=} \Psi(0).$$

III. PERTURBED SYSTEM (7)

As noted above, $\mathfrak{h}(\Pi_t x)$ for the perturbed system (7) varies slowly compared to x. Changes in $\mathfrak{h}(\Pi_t x)$ are significant only for times of order $1/\varepsilon^2$. Hence, we rescale time and write $X^\varepsilon(t) = x(t/\varepsilon^2)$, where x is governed by (7).

Under the above time scaling, the x time series would be compressed by a factor of ε^2. So, in order to be able to write the evolution equation for X^ε, we need to define a new segment extractor Π_t^ε as follows: For an $\mathbb{R}^n$-valued function f defined on $[-\varepsilon^2 r,\infty)$, the $[t-\varepsilon^2 r,t]$ segment is given by

$$\left(\Pi_t^\varepsilon f\right)(\theta) = f(t+\varepsilon^2\theta), \quad -r \leqslant \theta \leqslant 0. \tag{19}$$

Now the process X^ε has the same probability law as that of a process satisfying

$$dX^\varepsilon(t) = \frac{1}{\varepsilon^2}L_0(\Pi_t^\varepsilon X^\varepsilon)dt + G(\Pi_t^\varepsilon X^\varepsilon)dt + F(\Pi_t^\varepsilon X^\varepsilon)dW(t), \quad t \geqslant 0 \tag{20}$$

$$\Pi_0^\varepsilon X^\varepsilon = \varphi \in \mathcal{C},$$

where W is $\mathbb{R}$-valued Wiener process.[5] Write $\mathcal{H}^\varepsilon(t) := \mathfrak{h}(\Pi_t^\varepsilon X^\varepsilon)$ with $\mathfrak{h}$ defined in (18). Using Itô's formula, it can be shown that $\mathcal{H}^\varepsilon(t)$ satisfies

$$d\mathcal{H}^\varepsilon(t) = b\left(\Pi_t^\varepsilon X^\varepsilon\right)dt + \sigma\left(\Pi_t^\varepsilon X^\varepsilon\right)dW, \quad \mathcal{H}^\varepsilon(0) = \mathfrak{h}(\varphi), \tag{21}$$

where

$$b(\eta) = E(\eta)G(\eta) + \tfrac{1}{2}4[\hat{\Psi}_1 F(\eta)][\hat{\Psi}_2 F(\eta)], \tag{22}$$

$$\sigma(\eta) = E(\eta)F(\eta), \tag{23}$$

$$E(\eta) = 2(\langle \Psi_1,\eta\rangle\hat{\Psi}_2 + \langle \Psi_2,\eta\rangle\hat{\Psi}_1). \tag{24}$$

Recall that we can write the solution as $\Pi_t^\varepsilon X^\varepsilon = \Phi z(t) + (I-\pi)\Pi_t^\varepsilon X^\varepsilon$, where $z(t) := \langle \Psi,\Pi_t^\varepsilon X^\varepsilon\rangle$. Note that the evolution of $z_i(t) = \langle \Psi_i,\Pi_t^\varepsilon X^\varepsilon\rangle$ is fast compared to the evolution of $\mathcal{H}^\varepsilon$ and is predominantly oscillatory. Heuristically, the z_i oscillate fast along trajectories of constant $\mathfrak{h}$ (the effect of $\frac{1}{\varepsilon^2}L_0$) while at the same time diffusing slowly across the constant $\mathfrak{h}$ trajectories (the effect of perturbations G,F). Hence, the z_i in the above coefficients b and σ can be averaged.

Theorem 1. In the case when (i) F is constant and G is cubic and has a stabilizing effect or (ii) F is either linear or constant and G is Lipschitz continuous, the probability distribution of $\mathcal{H}^\varepsilon$ from (21) until any finite time $T > 0$ converges, as $\varepsilon \to 0$, to the probability distribution of a process $\check{h}$, which is the solution of the SDE

$$d\check{h}(t) = b_H(\check{h}(t))dt + \sigma_H(\check{h}(t))dW(t), \quad \check{h}(0) = \mathfrak{h}(\varphi),$$

where b_H and σ_H are obtained by averaging the functions in (22) and (23) as described below in Sec. III A. The perturbation G is said to have a stabilizing effect if the deterministic system $\dot{\hbar} = b_H(\hbar)$ is stable.

Note that $\mathcal{H}$ encodes information only about the critical component $\pi\Pi^\varepsilon X^\varepsilon$ of the solution. The above results should

[5] We have used the fact that for a Wiener process W, $\varepsilon W(t/\varepsilon^2)$ has the same probability law as a Wiener process.

be augmented with the result that the stable component $(I-\pi)\Pi^{\varepsilon}X^{\varepsilon}$ is small. Proof of Theorem 1 and a result to the effect that the stable component of the solution is small are presented in [25] (also see [24] for the case when G is Lipschitz continuous and F is constant).

A. Evaluation of b_H and σ_H

To evaluate b_H and σ_H at a specific value $\hbar \in \mathbb{R}$, we consider a solution $\Pi_t x$ of the unperturbed system (5) that remains in the space P for all time and such that $\mathfrak{h}(\Pi_t x) = \hbar$. For this purpose define

$$\eta_t^{\hbar} \overset{\text{def}}{=} \frac{1}{2}\sqrt{2\hbar}\Phi\begin{bmatrix} e^{i\omega_c t} \\ e^{-i\omega_c t}\end{bmatrix}. \tag{25}$$

Note that $\eta_t^{\hbar} \in P$ for all time and the z coordinates of $\eta_t^{\hbar}$ given by $\frac{1}{2}\sqrt{2\hbar}[\substack{e^{i\omega_c t}\\ e^{-i\omega_c t}}]$ evolve according to (17). Hence $\eta_t^{\hbar}$ is the solution of the unperturbed system with the initial condition $\eta_0^{\hbar}$. Further, $\mathfrak{h}(\eta_t^{\hbar}) = 2(\frac{1}{2}\sqrt{2\hbar}e^{i\omega_c t})(\frac{1}{2}\sqrt{2\hbar}e^{-i\omega_c t}) = \hbar$.

Now the averaged coefficients b_H and σ_H are given by

$$b_H(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} b(\eta_t^{\hbar})dt, \tag{26}$$

$$\sigma_H^2(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} \sigma^2(\eta_t^{\hbar})dt. \tag{27}$$

The following fact would be useful in the evaluation of the above averages: For $\eta_t^{\hbar}$, E defined in (24) becomes [on using (15)]

$$E(\eta_t^{\hbar}) = \sqrt{2\hbar}(\hat{\Psi}_1 e^{-i\omega_c t} + \hat{\Psi}_2 e^{i\omega_c t}).$$

IV. EXAMPLES

In this section we show three examples. The first is a simple scalar system: We study the perturbations of $\dot{x}(t) = -\frac{\pi}{2}x(t-1)$. In Sec. IV A, while studying cubic nonlinear perturbations and additive white noise perturbations, we illustrate the results of the previous section and show how the averaged process can yield information about the x process. This example is a running one in the sense that we revisit it when studying stronger deterministic perturbations in Sec. V and different kinds of noise in Sec. VI.

The purpose of the second example is to propose a conjecture. When perturbations are linear as well, we identify a complex number and claim that it alone dictates the stability of the system. We provide support to our conjecture using numerical simulations on $\dot{x}(t) = -\frac{\pi}{2}x(t-1)$.

The third is the van der Pol oscillator (4). Here we illustrate the stabilizing and destabilizing effects of noise and show how the averaging results obtained in the previous section give a good enough description of the effects of noise and allow us to compute how much bifurcation thresholds are displaced in the presence of noise when compared to the deterministic case.

A. Scalar equation

Consider the following equation:

$$dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon^2 x^3(t-1)dt + \varepsilon\sigma dW. \tag{28}$$

In this case $L_0\eta = -\frac{\pi}{2}\eta(-1)$, $G(\eta) = \eta^3(-1)$, and $F(\eta) = \sigma$. The characteristic equation $\lambda + \frac{\pi}{2}e^{-\lambda} = 0$ has countably infinite roots on the complex plane. The roots with the largest real part are $\pm i\omega_c = \pm i\frac{\pi}{2}$. Let $\Phi(\theta) = [e^{i(\pi/2)\theta}\ e^{-i(\pi/2)\theta}]$. Now Ψ can be evaluated [using (12)–(15)] to be

$$\Psi(\bullet) = \begin{bmatrix} \left(1 + i\frac{\pi}{2}\right)^{-1} e^{-i(\pi/2)\bullet} \\ \left(1 - i\frac{\pi}{2}\right)^{-1} e^{i(\pi/2)\bullet}\end{bmatrix}.$$

The averaged drift and diffusion can be calculated using (22)–(27) as

$$b_H(\hbar) = 2\hat{\Psi}_1\hat{\Psi}_2\sigma^2 - \tfrac{3}{2}i(\hat{\Psi}_1 - \hat{\Psi}_2)\hbar^2, \tag{29}$$

$$\sigma_H^2(\hbar) = 4\hat{\Psi}_1\hat{\Psi}_2\sigma^2\hbar. \tag{30}$$

In Sec. V D we illustrate how the averaged equation $d\hbar = b_H(\hbar)dt + \sigma_H(\hbar)dW$ can be used to gain information about (28) (recall Remark 1). Section V D can be read now, setting $\gamma_q = 0$ in (59).

B. Linear perturbations

In this section we consider the case where perturbations are also linear and identify a complex number that alone dictates the stability of the system. Note that we restrict the discussion to systems satisfying Assumption 1. Reference [28] discusses methods to obtain bounds on the maximal exponential growth rates of a more general class of delay equations. However, the bounds given in [28] are not optimal for systems satisfying Assumption 1.

Consider

$$dx(t) = L_0(\Pi_t x)dt + \varepsilon L_1(\Pi_t x)dW(t), \tag{31}$$

where L_i are linear operators, with L_0 satisfying Assumption 1. The averaged equation corresponding to (31) is

$$d\hbar(t) = b_H(\hbar)dt + \sigma_H(\hbar)dW(t), \tag{32}$$

where b_H and σ_H can be evaluated using (22)–(27) as

$$\begin{aligned} b_H(\hbar) &= C_b\hbar, \quad \sigma_H^2(\hbar) = C_\sigma\hbar^2, \\ C_b &= (\hat{\Psi}_1 L_1\Phi_1)(\hat{\Psi}_2 L_1\Phi_2) + (\hat{\Psi}_1 L_1\Phi_2)(\hat{\Psi}_2 L_1\Phi_1), \\ C_\sigma &= (\hat{\Psi}_1 L_1\Phi_1 + \hat{\Psi}_2 L_1\Phi_2)^2 + 2(\hat{\Psi}_1 L_1\Phi_2)(\hat{\Psi}_2 L_1\Phi_1). \end{aligned}$$

The solution to (32) is given by

$$\hbar(t) = \hbar(0)\exp\left[\left(C_b - \tfrac{1}{2}C_\sigma\right)t + \sqrt{C_\sigma}W(t)\right]. \tag{33}$$

The Lyapunov exponent for the averaged equation (32) can be calculated to be

$$\begin{aligned} \lambda_{\text{avg}} &= \lim_{t\to\infty}\frac{1}{t}\ln\hbar(t) \\ &= \lim_{t\to\infty}\frac{1}{t}\ln\hbar(0) + \left(C_b - \frac{1}{2}C_\sigma\right) + \sqrt{C_\sigma}\lim_{t\to\infty}\frac{W(t)}{t} \\ &= C_b - \frac{1}{2}C_\sigma = -\frac{1}{2}[(\hat{\Psi}_1 L_1\Phi_1)^2 + (\hat{\Psi}_2 L_1\Phi_2)^2]. \end{aligned}$$

Define $\lambda_j^{\varepsilon}(t) := \frac{1}{t}\ln\sup_{s\in[t-mr,t]}|x_j(s)|$ with $m \in \mathbb{N}$ such that $mr > \frac{2\pi}{\omega_c}$ (here m is chosen so as to avoid oscillations in the modulus of x). We conjecture that for large t, $\lambda^{\varepsilon}(t)$ is close to $\varepsilon^2\frac{1}{2}\lambda_{\text{avg}}$. The $\frac{1}{2}$ arises from the fact that $\hbar$ is quadratic in x.

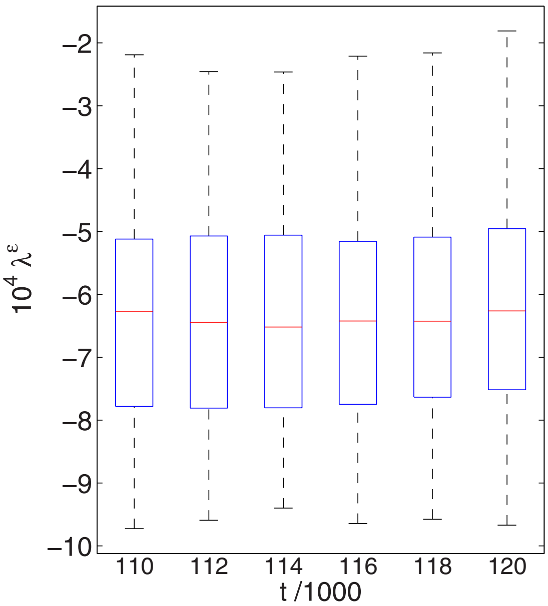

FIG. 1. Box plot of $\lambda^\varepsilon(t) = \frac{1}{t}\ln\sup_{s\in[t-5r,t]}|x(s)|$ for t between 110 000 and 120 000 in steps of 2000. The x process is simulated using (34) with $\varepsilon = 0.1$. The line inside the box (red) is the mean of 80 realizations. The lower end of the box (blue) is the 25th percentile and the upper end of the box (blue) is the 75th percentile.

We verify the above conjecture using the system

$$dx = -\frac{\pi}{2}x(t-1)dt + \varepsilon x(t-1)dW, \tag{34}$$

i.e., $L_0\eta = -\frac{\pi}{2}\eta(-1)$ and $L_1\eta = \eta(-1)$. The Lyapunov exponent for (32) can be calculated to be $\lambda_{\text{avg}} \approx -0.122$ (the matrices $\hat{\Psi}$ and Φ are already calculated in Sec. IV A). Eighty realizations of trajectories of (34) are simulated with $\varepsilon = 0.1$ and the initial condition $(\Pi_0 x)(\theta) = \cos(\omega_c\theta)$ for $\theta \in [-r,0]$. In Fig. 1 we show a box plot for $\lambda^\varepsilon(t) := \frac{1}{t}\ln\sup_{s\in[t-5,t]}|x(s)|$. For t large, the mean of $\lambda^\varepsilon(t)$ is close to -0.0006 and we have $\varepsilon^2\frac{1}{2}\lambda_{\text{avg}} \approx -0.0006$. For details of the numerical scheme see Appendix E.

Recalling that $\hat{\Psi}_2$ and $L_1\Phi_2$ are the complex conjugates of $\hat{\Psi}_1$ and $L_1\Phi_1$, respectively, we find that

$$\lambda_{\text{avg}} = -\text{Re}[(\hat{\Psi}_1 L_1 \Phi_1)^2] = -|\hat{\Psi}_1 L_1 \Phi_1|^2\cos(2\theta_*),$$

where θ_* is the angle of the complex number $\hat{\Psi}_1 L_1\Phi_1$. The stability condition $\lambda_{\text{avg}} < 0$ translates to $\cos(2\theta_*) > 0$. If the conjecture that for large t, $\lambda^\varepsilon(t)$ is close to $\varepsilon^2\frac{1}{2}\lambda_{\text{avg}}$ is true, then the complex number $\hat{\Psi}_1 L_1\Phi_1$ alone dictates the stability of (31).

C. van der Pol oscillator

In this section we consider the oscillator modeled by Eq. (4), which was considered in [10]. In studying (4) our intentions are threefold: (i) to point out[6] the errors in the analysis of [10], (ii) to illustrate the stabilizing and destabilizing effects of noise, and (iii) to show that the averaging results obtained in the previous section give a good enough description of the effects of noise.

The oscillator (4) has natural frequency ω_0, which would be altered by the delayed feedbacks $\eta q(t-r)$ and $\kappa\dot{q}(t-r)$. The negative of β indicates the strength of linear damping in the oscillator. The coefficient b, if positive, is the strength of nonlinear damping in the oscillator.

Since we intend to study the effect of small noise perturbations, we scale $D = \varepsilon^2\tilde{D}$ with $\varepsilon \ll 1$. Since we study the dynamics close to the zero fixed point, we zoom-in and write $x_1(t) = \frac{1}{\varepsilon}q(t)$ and $x_2(t) = \frac{1}{\varepsilon}\dot{q}(t)$. Then the oscillator (4) can be put in the form (using Itô's interpretation)

$$dx(t) = L_0(\Pi_t x)dt + \varepsilon^2\begin{pmatrix}0\\ -bx_1^2(t)x_2(t)\end{pmatrix}dt + \varepsilon\sqrt{2\tilde{D}}\begin{pmatrix}0\\ x_1(t)\end{pmatrix}dW(t), \tag{35}$$

where W is a Wiener process and $L_0\phi = \int_{-r}^0 d\mu(\theta)\phi(\theta)$ with

$$d\mu(\theta) = \begin{pmatrix}0 & 1\\ -\omega_0^2 & \beta\end{pmatrix}\delta_0(\theta) + \begin{pmatrix}0 & 0\\ -\eta & \kappa\end{pmatrix}\delta_{-r}(\theta),$$

where δ_0 and δ_{-r} are delta functions, i.e., $\int\delta_0\phi = \phi(0)$ and $\int\delta_{-r}\phi = \phi(-r)$ for $\phi \in \mathcal{C}$.

The characteristic equation becomes

$$-\lambda\beta + \lambda^2 + (\eta - \kappa\lambda)e^{-\lambda r} + \omega_0^2 = 0. \tag{36}$$

Since our intention is to study the effect of small noise perturbations on the oscillator when it is at the verge of instability, we assume that the parameters of the problem are such that the characteristic equation has two roots $\pm i\omega_c$ on the imaginary axis and all other roots have negative real parts. With this assumption the unperturbed system $\dot{x}(t) = L_0(\Pi_t x)$ is on the verge of instability. Figure 2 shows the stability boundary.

The matrices Φ and Ψ can be evaluated [using (10)–(15)] as

$$\Phi(\bullet) = \begin{pmatrix} e^{i\omega_c\bullet} & e^{-i\omega_c\bullet}\\ i\omega_c e^{i\omega_c\bullet} & -i\omega_c e^{-i\omega_c\bullet}\end{pmatrix} = (\Phi_1\Phi_2),$$

$$\Psi(\bullet) = \begin{pmatrix} c(\omega_0^2 + \eta e^{-i\omega_c r})e^{-i\omega_c\bullet} & c(-i\omega_c)e^{-i\omega_c\bullet}\\ \bar{c}(\omega_0^2 + \eta e^{i\omega_c r})e^{i\omega_c\bullet} & \bar{c}(i\omega_c)e^{i\omega_c\bullet}\end{pmatrix} = \begin{pmatrix}\Psi_1\\ \Psi_2\end{pmatrix},$$

where

$$c = \left[\omega_c^2 + e^{-i\omega_c r}\left(\eta + i\eta r\omega_c + \kappa r\omega_c^2\right) + \omega_0^2\right]^{-1}. \tag{37}$$

Remark 2. The process $\mathfrak{h}(\Pi_t x)$ with $\mathfrak{h}$ defined in (18) has additional significance for this problem. If $\Pi_t x$ is such that the stable part $(I-\pi)\Pi_t x$ is zero, then $\Pi_t x = \pi\Pi_t x = \Phi z(t)$, which gives

$$x(t) = \Pi_t x(0) = \Phi_1(0)z_1(t) + \Phi_2(0)z_2(t) = \begin{bmatrix} z_1(t) + z_2(t)\\ i\omega_c[z_1(t) - z_2(t)]\end{bmatrix},$$

from which we get $\mathfrak{h}(\Pi_t x) \overset{\text{by def}}{=} 2z_1(t)z_2(t) = \frac{1}{2}\{[x_1(t)]^2 + [x_2(t)/\omega_c]^2\}$, which represents some kind of energy in the oscillator (note that x_1 is the position and x_2 is the velocity). Usually $\|(I-\pi)\Pi_t x\|$ decays to very small values

[6]This is done in Appendix A.

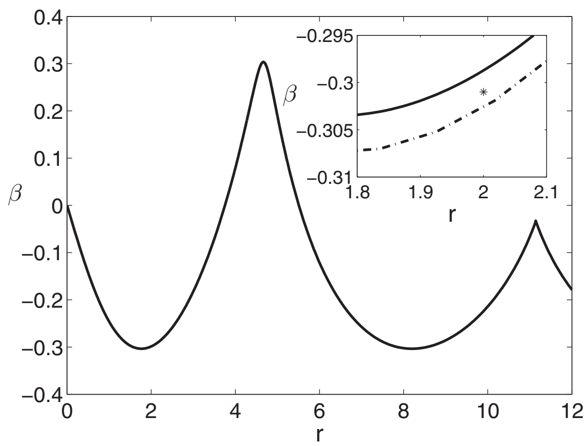

FIG. 2. Boundary of stability for the fixed point ($x_1 = 0, x_2 = 0$) of the system (35) with $\varepsilon = 0$, $\omega_0 = 1$, $\kappa = 0$, and $\eta = 0.3$. For each delay r there exists a critical value β_c such that for $\beta < \beta_c$ the fixed point is stable and for $\beta > \beta_c$ the fixed point is unstable. In the inset, the (theoretically predicted) stability boundary in the presence of noise is shown with a dashed line [obtained using (41)]. For this, $\varepsilon = 0.1$, $\tilde{D} = 1$, and $b = 1$. For β in the region below the dashed line, theoretical results predict that the (0,0) fixed point is stable in the presence of noise. Above the dashed line the fixed point loses stability; nevertheless, invariant density exists. So, theoretical results predict that the noise has destabilized the region between solid and dashed lines. The point marked by an asterisk in the inset is ($r = 2, \beta = -0.301$). For this point we show in Fig. 3 the invariant density obtained by numerical simulations. The theoretically obtained invariant density [obtained in (42)] is in very good agreement with the actual density obtained from numerical simulations.

exponentially fast and hence $\mathfrak{h}(\Pi_t x)$ differs from the energy $\frac{1}{2}\{[x_1(t)]^2 + [x_2(t)/\omega_c]^2\}$ by a little amount.

Using (22)–(27) we have

$$b_H(\hbar) = (2\tilde{D})2|c|^2\omega_c^2\hbar - b\omega_c^2\tfrac{1}{2}(c+\bar{c})\hbar^2,$$
$$\sigma_H^2(\hbar) = (2\tilde{D})\{2|c|^2\omega_c^2 + [i\omega_c(\bar{c}-c)]^2\}\hbar^2.$$

To understand whether noise has a stabilizing or destabilizing effect, let us consider the damping β as a bifurcation parameter. Write $\beta = \beta_c + \varepsilon^2\tilde{\beta}$ and assume that at $\varepsilon = 0$, β satisfies the characteristic equation (36). Then the effect of $\tilde{\beta}$ is to add another term $\tilde{\beta}(c+\bar{c})\omega_c^2\hbar$ to b_H. Then we can write the averaged equation as

$$d\hbar = b_H(\hbar)dt + \sigma_H(\hbar)dW, \tag{38}$$

where

$$b_H(\hbar) = C_b\hbar + C_b^{(2)}\hbar^2, \quad \sigma_H^2(\hbar) = C_\sigma\hbar^2,$$
$$C_b = (2\tilde{D})2|c|^2\omega_c^2\left(1 + \frac{\tilde{\beta}}{2\tilde{D}}\frac{(c+\bar{c})/2}{|c|^2}\right),$$
$$C_b^{(2)} = -b\omega_c^2\frac{1}{2}(c+\bar{c}),$$
$$C_\sigma = (2\tilde{D})2|c|^2\omega_c^2\left(1 + \frac{2[(\bar{c}-c)/2i]^2}{|c|^2}\right).$$

To focus on the effect of noise, for the moment we ignore the nonlinearities by setting $b = 0$ in (35). The corresponding averaged system then becomes

$$d\hbar = C_b\hbar + \sqrt{C_\sigma}\hbar dW. \tag{39}$$

The above system is unstable when[7] $C_b - \frac{1}{2}C_\sigma > 0$, i.e., when

$$\frac{\tilde{\beta}}{2\tilde{D}|c|}\frac{(c+\bar{c})/2}{|c|} > \frac{[(\bar{c}-c)/2i]^2}{|c|^2} - \frac{1}{2}. \tag{40}$$

Let $\varsigma_1 = \frac{(c+\bar{c})/2}{|c|}$ and $\varsigma_2 = (\frac{[(\bar{c}-c)/2i]^2}{|c|^2} - \frac{1}{2})$. It can be shown[8] that if $\beta_c < 0$, then $\varsigma_1 > 0$.

Assume $\beta_c < 0$. Then (40) holds when

$$\frac{\tilde{\beta}}{2\tilde{D}|c|} > \frac{\varsigma_2}{\varsigma_1}. \tag{41}$$

If noise were not present, i.e., $\tilde{D} = 0$ in (35), then the fixed point ($x_1 = 0, x_2 = 0$) of (39) would have been unstable for any $\tilde{\beta} > 0$ (this is because $-\tilde{\beta}$ specifies how much additional damping is present in the system). If noise is present and $\varsigma_2 > 0$, then the ($x_1 = 0, x_2 = 0$) fixed point of (39) is stable even for $0 < \tilde{\beta} < 2\tilde{D}|c|\varsigma_2/\varsigma_1$ So, noise has a stabilizing effect if $\varsigma_2 > 0$.

Similar reasoning shows that the noise has a destabilizing effect if $\varsigma_2 < 0$. If the noise were not present, then the fixed point ($x_1 = 0, x_2 = 0$) of (39) would have been stable for any $\tilde{\beta} < 0$. If noise is present and $\varsigma_2 < 0$, then (39) is unstable even for $2\tilde{D}|c|\varsigma_2/\varsigma_1 < \tilde{\beta} < 0$. So noise has a destabilizing effect if $\varsigma_2 < 0$. This is the scenario presented in the inset of Fig. 2.

The stability of (35) when $b \neq 0$ depends on the stability of averaged nonlinear system (38). However, Theorem 1 deals with only weak convergence of probability distributions and hence is not adequate to transfer the stability properties from the averaged system to the original system (35). Nevertheless, we give an account of the stability of the averaged system (38). When the nonlinearity is destabilizing, i.e., $C_b^{(2)} > 0$, the system (35) cannot be stable. When $C_b^{(2)} < 0$ and $C_b - \frac{1}{2}C_\sigma < 0$ the trivial solution $\hbar = 0$ is the only equilibrium point of (38) and is stable. When $C_b^{(2)} < 0$ and $C_b - \frac{1}{2}C_\sigma > 0$ the trivial solution of (38) becomes unstable; nevertheless, an invariant density exists. It is given by (obtained by solving steady-sate

[7]Note that the solution is similar to (33).

[8]Note that $\mathrm{sgn}(\varsigma_1) = \mathrm{sgn}(\frac{c+\bar{c}}{c\bar{c}}) = \mathrm{sgn}(\frac{1}{c} + \frac{1}{\bar{c}})$. Using (37) we have

$$c^{-1} + (\bar{c})^{-1} = 2(\omega_c^2 + \omega_0^2) + \eta(e^{i\omega_c r} + e^{-i\omega_c r}) + ir\omega_c e^{-i\omega_c r}(\eta - i\omega_c k) - ir\omega_c e^{i\omega_c r}(\eta + i\omega_c k).$$

Employing $\lambda = \pm i\omega_c$ in the characteristic equation (36) we get

$$ir\omega_c e^{-i\omega_c r}(\eta - i\omega_c k) - ir\omega_c e^{i\omega_c r}(\eta + i\omega_c k) = -2\beta_c r\omega_c^2,$$
$$2\eta(e^{i\omega_c r} + e^{-i\omega_c r}) = (\omega_c^2 - \omega_0^2)(e^{i\omega_c r} + e^{-i\omega_c r})^2 + \beta_c i\omega_c(e^{2i\omega_c r} - e^{-2i\omega_c r}).$$

Hence $c^{-1} + (\bar{c})^{-1} = 2(\omega_c^2 + \omega_0^2) + \frac{1}{2}(\omega_c^2 - \omega_0^2)(e^{i\omega_c r} + e^{-i\omega_c r})^2 + \frac{1}{2}\beta_c i\omega_c(e^{2i\omega_c r} - e^{-2i\omega_c r}) - 2\beta_c r\omega_c^2$, which can be simplified as $c^{-1} + (\bar{c})^{-1} = 2\omega_c^2(1 + \cos^2\omega_c r) + 2\omega_0^2(1 - \cos^2\omega_c r) - \beta_c\omega_c(2r\omega_c + \sin 2\omega_c r)$, which is positive if $\beta_c < 0$.

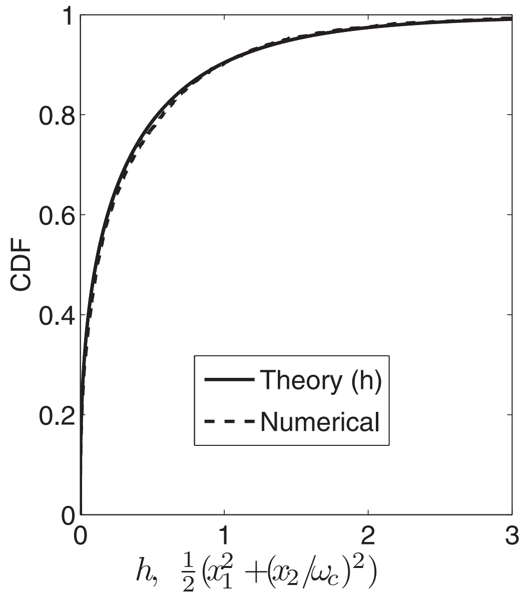

FIG. 3. Cumulative distribution function of the invariant density of $\frac{1}{2}[x_1^2+(x_2/\omega)^2]$ obtained from numerical simulation of (35) with parameters specified by the point marked by an asterisk in the inset of Fig. 2 ($\omega_0=1$, $\kappa=0$, $\eta=0.3$, $\varepsilon=0.1$, $\tilde{D}=1$, $b=1$, $r=2$, and $\beta=-0.301$). This agrees with the CDF of the density given in (42). For this case, the deterministic bifurcation threshold is $\beta_c=-0.2987$ and the predicted threshold in the presence of noise is $\beta_c+\varepsilon^2 2\tilde{D}|c|\varsigma_2/\varsigma_1=-0.3027$.

Fokker-Planck equation)

$$p(\hbar)=\frac{\chi^{2C_b/C_\sigma-1}}{\Gamma\left(\frac{2C_b}{C_\sigma}-1\right)}\hbar^{2(C_b/C_\sigma-1)}e^{-\hbar\chi},\quad \chi=2\left(-C_b^{(2)}\right)/C_\sigma, \tag{42}$$

where Γ is the Gamma function.

The averaging results for (35) hold for times of order $1/\varepsilon^2$, whereas stability concerns times $t\to\infty$. Nevertheless, we expect that, for small ε, (i) the invariant density from (42) is a good approximation to the steady-state density of $\frac{1}{2}[x_1^2+(x_2/\omega_c)^2]$ from (35) and (ii) the bifurcation threshold as predicted by averaging would be a good approximation to the actual bifurcation threshold of (35). The usefulness of the averaging results is shown in Fig. 3. Let the parameters be specified by the point marked by the asterisk in the inset of Fig. 2. When $\varepsilon=0$, the fixed point ($x_1=0,x_2=0$) of the oscillator (35) would be stable because the asterisk lies below the stability boundary (solid line in Fig. 2). However, in the presence of noise the stability boundary is shifted by $\varepsilon^2 2\tilde{D}|c|\varsigma_2/\varsigma_1$ (dashed line in Fig. 2). Now the fixed point loses stability; nevertheless, invariant density exists. Numerical simulation is done with 3200 samples and the cumulative distribution function (CDF) of the steady-state density of $\frac{1}{2}(x_1^2+(x_2/\omega_c)^2)$ is plotted in Fig. 3. Also shown is the CDF arising from the averaging result (42). Figure 3 indeed shows that the density from (42) is a good approximation to the steady-state density of $\frac{1}{2}(x_1^2+(x_2/\omega_c)^2)$ from (35).

Numerical simulations in the case $\varsigma_2<0$ with $\varepsilon=0.1$ show very good agreement with theoretical averaging results for β in the range $\beta_c+0.9\varepsilon^2(2\tilde{D}|c|\varsigma_2/\varsigma_1)<\beta<\beta_c$. Very close to the theoretically predicted bifurcation threshold in the presence of noise, i.e., $\beta\approx\beta_c+\varepsilon^2(2\tilde{D}|c|\varsigma_2/\varsigma_1)$, the agreement is not very good. The actual bifurcation threshold in the presence of noise (denoted by $\beta_{c,\mathrm{noi}}$), obtained from numerical simulations of (35), is within 10% of the theoretically predicted value,[9] i.e., $\beta_c+1.1\varepsilon^2(2\tilde{D}|c|\varsigma_2/\varsigma_1)<\beta_{c,\mathrm{noi}}<\beta_c+\varepsilon^2(2\tilde{D}|c|\varsigma_2/\varsigma_1)$. For details of the numerical scheme see Appendix E. For the numerical simulations verifying this claim, see the Supplemental Material [29].

V. STRONGER DETERMINISTIC PERTURBATIONS

Here we consider systems with slightly stronger deterministic perturbations

$$dx(t)=L_0(\Pi_t x)dt+\varepsilon G_q(\Pi_t x)dt+\varepsilon^2 G(\Pi_t x)dt + \varepsilon F(\Pi_t x)dW(t), \tag{43}$$

where W is an $\mathbb{R}$-valued Wiener process. As an example, consider the noisy perturbation $d\tilde{x}=-\frac{\pi}{2}\tilde{x}(t-1)dt+\tilde{x}^2(t)dt+\varepsilon^2\sigma dW$ of the DDE $\dot{\tilde{x}}(t)=-\frac{\pi}{2}\tilde{x}(t-1)+\tilde{x}^2(t)$. Then $x(t)=\varepsilon^{-1}\tilde{x}(t)$ can be put in the form (43) with $L_0(\eta)=-\frac{\pi}{2}\eta(-1)$, $F(\eta)=\sigma$, $G(\eta)=0$, and $G_q(\eta)=\eta^2(0)$.

The effect of G_q in (43) is significant in just times of order $1/\varepsilon$, whereas the effects of G and F are significant in times of order $1/\varepsilon^2$. So we consider only those G_q that are such that a certain kind of time-averaged effect of G_q is zero:

$$\frac{1}{2\pi/\omega}\int_0^{2\pi/\omega} e^{-i\omega_c t}\hat{\Psi}_1 G_q\left(\eta_t^{\hbar}\right)dt=0, \tag{44}$$

where $\eta_t^{\hbar}$ is defined in (25). The assumption (44) is a natural one: For example, G_q that are homogenously quadratic in η (say $G_q(\eta)=[\eta(0)]^2$) satisfy the property (44).

Writing $X^\varepsilon(t)=x(t/\varepsilon^2)$, the equation analogous to (20) becomes

$$dX^\varepsilon(t)=\frac{1}{\varepsilon^2}L_0\left(\Pi_t^\varepsilon X^\varepsilon\right)dt+\frac{1}{\varepsilon}G_q\left(\Pi_t^\varepsilon X^\varepsilon\right)dt+G\left(\Pi_t^\varepsilon X^\varepsilon\right)dt + F\left(\Pi_t^\varepsilon X^\varepsilon\right)dW(t),\quad t\geqslant 0$$
$$\Pi_0^\varepsilon X^\varepsilon=\varphi\in\mathcal{C}. \tag{45}$$

Using Itô's formula, $\mathcal{H}^\varepsilon(t):=\mathfrak{h}(\Pi_t^\varepsilon X^\varepsilon)$ satisfies

$$d\mathcal{H}^\varepsilon(t)=\frac{1}{\varepsilon}\left[b^{q,(1)}\left(\Pi_t^\varepsilon X^\varepsilon\right)+b^{q,(2)}\left(\Pi_t^\varepsilon X^\varepsilon\right)\right]dt + b\left(\Pi_t^\varepsilon X^\varepsilon\right)dt+\sigma\left(\Pi_t^\varepsilon X^\varepsilon\right)dW,\quad \mathcal{H}^\varepsilon(0)=\mathfrak{h}(\varphi), \tag{46}$$

where b, σ, and E are the same as in (22), (23), and (24) respectively, and

$$b^{q,(1)}(\eta)=E(\eta)G_q(\pi\eta), \tag{47}$$
$$b^{q,(2)}(\eta)=E(\eta)[G_q(\eta)-G_q(\pi\eta)]. \tag{48}$$

Recall that we can write the solution as $\Pi_t^\varepsilon X^\varepsilon=\Phi z(t)+(I-\pi)\Pi_t^\varepsilon X^\varepsilon$, where $z(t):=\langle\Psi,\Pi_t^\varepsilon X^\varepsilon\rangle$. Note that the evolution of $z_i(t)=\langle\Psi_i,\Pi_t^\varepsilon X^\varepsilon\rangle$ is fast compared to the evolution

[9]Note that the theoretical averaging results concern the limit $\varepsilon\to 0$, but here we take $\varepsilon=0.1$.

of $\mathcal{H}^\varepsilon$ and is predominantly oscillatory. Heuristically, the z_i oscillate fast *along* trajectories of constant $\mathfrak{h}$ (the effect of $\frac{1}{\varepsilon^2}L_0$) while at the same time diffusing slowly *across* the constant $\mathfrak{h}$ trajectories (the effect of perturbations G,G_q,F). Hence, the effect of z_i in the above coefficients b and σ can be averaged out. Our goal is to obtain an averaging result akin to Theorem 1. However, the terms arising from G_q should be dealt with carefully. The assumption (44) would entail that $\frac{1}{2\pi/\omega}\int_0^{2\pi/\omega} E(\eta_t^\hbar)G_q(\eta_t^\hbar)dt$ equals zero as well.[10] Hence, when the oscillations are averaged, the leading-order contribution of $b^{q,(1)}$ is zero. However, because of the $\frac{1}{\varepsilon}$ multiplying $b^{q,(1)}$, higher-order effects must be taken into account.

We give explicit formulas for the contributions from $b^{q,(1)}$ and $b^{q,(2)}$, using solutions of the unperturbed system with n specific initial conditions. At least when G_q is purely quadratic, the averaged terms arising from $b^{q,(k)}$ would be the same as what one gets from a formal center manifold and normal-form calculation. However, we do not assume the existence of a center manifold. The following method, however, has an advantage in that numerical integration can be used to find the answers. To provide an illustration of how the method works, a simple example without delay is worked out in Appendix B. To state the formulas, we need to set up some notation.

A. Notation

For $\varphi \in \mathcal{C}$, let $\hat{T}(t)\varphi$ denote the solution at time t of the unperturbed linear system (5) with initial condition $\Pi_0 x = \varphi$, i.e., $\hat{T}(t)\varphi = \Pi_t x$, where x is governed by (5). Let $\mathbf{1}_{\{0\}} : [-r,0] \to \mathbb{R}^{n\times n}$ denote the matrix-valued function

$$\mathbf{1}_{\{0\}}(\theta) = \begin{cases} I_{n\times n}, & \theta = 0 \\ 0_{n\times n}, & \theta \neq 0, \end{cases} \tag{49}$$

where I is the identity matrix. For a constant $n \times 1$ vector $\underline{v}$, one can solve the unperturbed linear system (5) with $\Pi_0 x = \mathbf{1}_{\{0\}}\underline{v}$. The solution is indicated by $\hat{T}(t)\mathbf{1}_{\{0\}}\underline{v}$.

Recall that π is the projection operator onto the critical eigenspace and is given by (16). Even though $\mathbf{1}_{\{0\}}\underline{v}$ does not belong to $\mathcal{C}$ (because it is not continuous), the definition $\pi(\mathbf{1}_{\{0\}}\underline{v}) := \Phi\langle\Psi,\mathbf{1}_{\{0\}}\underline{v}\rangle$ still makes sense[11] using the bilinear form (12). On evaluation of the bilinear form we find that

$$\pi(\mathbf{1}_{\{0\}}\underline{v}) = \Phi\hat{\Psi}\underline{v}. \tag{50}$$

The meaning of $\hat{T}(t)\pi\mathbf{1}_{\{0\}}\underline{v}$ and $\hat{T}(t)(I-\pi)\mathbf{1}_{\{0\}}\underline{v}$ should now be clear.

Suppose that $G : \mathcal{C} \to \mathbb{R}^k$ and let $\eta,\xi \in \mathcal{C}$. Then $(\xi\cdot\nabla)G(\eta)$ denotes the Fréchet differential of G evaluated at η in the direction of ξ, i.e.,

$$(\xi\cdot\nabla)G(\eta) = \lim_{\delta\to 0}\frac{G(\eta+\delta\xi)-G(\eta)}{\delta}.$$

Later we will see the motivation for defining the following:

$$\rho(\eta) := \inf\left\{t > 0 : \langle\Psi,\hat{T}(t)\pi\eta\rangle = \frac{1}{2}\sqrt{2\mathfrak{h}(\eta)}\begin{bmatrix}1\\1\end{bmatrix}\right\}, \tag{51}$$

$$a_q^{(1)}(\eta) = \int_0^{\rho(\eta)}((\hat{T}(s)\pi\mathbf{1}_{\{0\}}G_q(\eta))\cdot\nabla)b^{q,(1)}(\hat{T}(s)\pi\eta)ds, \tag{52}$$

$$a_q^{(2)}(\eta) = \int_0^{\infty}((\hat{T}(s)\mathbf{1}_{\{0\}}G_q(\eta))\cdot\nabla)b^{q,(2)}(\hat{T}(s)\pi\eta)ds. \tag{53}$$

B. Averaging

Theorem 2. In the case when F is constant, G and G_q are Lipschitz continuous, and G_q satisfies (44), the probability distribution of $\mathcal{H}^\varepsilon$ until any finite time $T > 0$ converges, as $\varepsilon \to 0$, to the probability distribution of a process $\check{h}$, which is the solution of the SDE

$$d\check{h}(t) = \left(b_H + b_H^{q,(1)} + b_H^{q,(2)}\right)(\check{h}(t))dt + \sigma_H(\check{h}(t))dW(t),$$
$$\check{h}(0) = \mathfrak{h}(\varphi),$$

where b_H and σ_H are the same as in (26) and (27) and $b_H^{q,(k)}$ for $k = 1,2$ are given by

$$b_H^{q,(k)}(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} a_q^{(k)}(\eta_t^\hbar)dt, \tag{54}$$

with $\eta_t^\hbar$ defined in (25). The coefficients $b_H^{q,(k)}$ are written more explicitly in (55) and (56).

The proof of the above result can be found in [24]. The key idea in obtaining the averaged effect of G_q is this: Let $c^{q,(1)}$ be the function whose differential along the trajectory of the unperturbed system equals $b^{q,(1)}$ defined in (47). Then the average effect of $b^{q,(1)}$ is the negative of the average of the differential of $c^{q,(1)}$ along the direction of the perturbations. In symbols, the function $c^{q,(1)}(\eta) = -\int_0^{\rho(\eta)} b^{q,(1)}(\hat{T}(s)\eta)ds$ is such that $\frac{d}{dt}|_{t=0}c^{q,(1)}(\hat{T}(t)\eta) = b^{q,(1)}(\eta)$. The differential of $c^{q,(1)}$ along the direction of the perturbations is $(\mathbf{1}_{\{0\}}G_q(\eta)\cdot\nabla)c^{q,(1)}(\eta)$, which evaluates to $-a^{q,(1)}(\eta)$ [plus an additional term whose average turns out to be zero due to the assumption (44)]. The average effect of $b^{q,(1)}$ is the average of $a^{q,(1)}$. The reasoning is similar for $b^{q,(2)}$. For details see[12] Sec. 9 of [24]. To illustrate the above idea, a simple example without delay is worked out in Appendix B. We urge the reader to study Appendix B to gain intuition about the process of obtaining the drift coefficients $b_H^{q,(i)}$.

---

[10] This follows from the fact that $E(\eta_t^\hbar) = \sqrt{2\hbar}(e^{-i\omega_c t}\hat{\Psi}_1 + e^{i\omega_c t}\hat{\Psi}_2)$ and $\hat{\Psi}_2$ is the conjugate of $\hat{\Psi}_1$.

[11] A rigorous way to extend the space $\mathcal{C}$ to include the discontinuities and the decomposition of the extended space as $P \oplus \hat{Q}$ is discussed in [11].

[12] Reference [24] deals with scalar systems and does not employ polar coordinates. Hence the form of expression differs from here. However, Ref. [24] evaluates to the same numbers as here. The key difference is that in [24] an element $\eta \in P$ is written as $z_1\cos(\omega_c\cdot) + z_2\sin(\omega_c\cdot)$ with $z_i \in \mathbb{R}$. Here we write $z_1 e^{i\omega_c\cdot} + z_2 e^{-i\omega_c\cdot}$ with $z_i \in \mathbb{C}$ and $z_2 = \bar{z}_1$.

The term $b_H^{q,(1)}$ is solely due to the critical eigenspace and the term $b_H^{q,(2)}$ arises from the interaction between the stable eigenspace and critical eigenspace. When G_q is purely quadratic, these are the same terms that arise from a formal center-manifold calculation.

Note that $\mathcal{H}$ encodes information only about the critical component of the solution $\pi\Pi^\varepsilon X^\varepsilon$. The above results should be augmented with a result that the stable component $(I-\pi)\Pi^\varepsilon X^\varepsilon$ is small. Proof of Theorem 2 and a result to the effect that the stable component of the solution is small are presented in [24].

Remark 3. It is clear from (48) that, if we had totally ignored the stable component, i.e., if we had set $(I-\pi)\Pi_t^\varepsilon X^\varepsilon = 0$ at the very beginning of the analysis, we would have missed the term $b_H^{q,(2)}$.

Remark 4. The coefficients $b_H^{q,(k)}$ can be written more explicitly as

$$b_H^{q,(1)}(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt \int_0^{(2\pi/\omega_c)-t} ds \Big(2\big(\hat{\Psi}G_q(\eta_t^\hbar)\big)^* \begin{bmatrix} 0 & e^{i\omega_c s} \\ e^{-i\omega_c s} & 0 \end{bmatrix} \hat{\Psi}G_q(\eta_{t+s}^\hbar)\Big)$$
$$+ \frac{\sqrt{2\hbar}}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt \int_0^{(2\pi/\omega_c)-t} ds \big(\big(\Phi e^{sB}\hat{\Psi}G_q(\eta_t^\hbar)\big)\cdot\nabla\big)\big(\mathcal{E}_{t+s}G_q(\eta_{t+s}^\hbar)\big), \tag{55}$$

$$b_H^{q,(2)}(\hbar) = \frac{\sqrt{2\hbar}}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt \int_0^\infty ds \sum_{j=1}^n \big(G_q(\eta_t^\hbar)\big)_j\big((\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}\underline{e_j})\cdot\nabla\big)\big(\mathcal{E}_{t+s}G_q(\eta_{t+s}^\hbar)\big), \tag{56}$$

where $\eta_t^\hbar$ is defined in (25),

$$\mathcal{E}_t := e^{-i\omega_c t}\hat{\Psi}_1 + e^{i\omega_c t}\hat{\Psi}_2, \tag{57}$$

and $\underline{e_j}$ denotes the unit vector in the jth direction of $\mathbb{R}^n$. To check how these explicit forms follow from (51)–(54) refer to Appendix C. If G_q is a polynomial, the terms in (55) can be put in *Mathematica* to get an explicit functional dependence on $\hbar$; otherwise numerical integration can be done at specific $\hbar$ values. For the term in (56) the integral $\int_0^{2\pi/\omega_c}$ can be evaluated first using *Mathematica* and then $\int_0^\infty$ can be done using numerical integration. All that we would need are the solutions of the unperturbed system with n different initial conditions $(I-\pi)\mathbf{1}_{\{0\}}\underline{e_j}$ for $j=1,\dots,n$. Since the initial condition $(I-\pi)\mathbf{1}_{\{0\}}\underline{e_j}$ belong to the stable space Q, the solution $\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}\underline{e_j}$ decays exponentially fast to zero and hence integral $\int_0^\infty$ need not be evaluated until infinity; a reasonable large upper limit would be enough to get a good enough approximation. An example is done in the next section to illustrate the above computations. Note that, when applied in a deterministic DDE setting, the above formulas provide an alternate way to compute the effect of center-manifold terms on the amplitude of critical mode.

C. Example

Consider Eq. (28) with added quadratic nonlinearity $G_q(\eta) = [\eta(-1)]^2$:

$$dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon^2 x^3(t-1)dt + \varepsilon\sigma dW$$
$$+ \varepsilon x^2(t-1)dt. \tag{58}$$

We apply Theorem 2. Note that b_H and σ_H have already been evaluated [see Eqs. (29) and (30)]. We continue using the Φ and Ψ from Sec. IV A.

Now we evaluate $b_H^{q,(1)}$ and $b_H^{q,(2)}$ using (54). In Sec. V D we show by numerical simulations how the averaged dynamics would be useful to gain information about (58).

Note that $(\xi\cdot\nabla)G_q(\eta) = 2\eta(-1)\xi(-1)$. We also write it as $2\eta|_{-1}\xi|_{-1}$ to avoid writing too many brackets. Using the formula (55), we have

$$b_H^{q,(1)}(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c}\Big(\int_0^{(2\pi/\omega_c)-t}\mathscr{G}(t,s)ds\Big)dt,$$

where

$$\mathscr{G}(t,s) = 2\hat{\Psi}_1\hat{\Psi}_2(e^{i\omega_c s}+e^{-i\omega_c s})\big(\eta_t^\hbar|_{-1}\big)^2\big(\eta_{t+s}^\hbar|_{-1}\big)^2 + \sqrt{2\hbar}\mathcal{E}_{t+s}2\big(\eta_{t+s}^\hbar|_{-1}\big)(\Phi|_{-1}e^{sB}\hat{\Psi})\big(\eta_t^\hbar|_{-1}\big)^2,$$

where $\eta_t^\hbar$ is defined in (25). Using *Mathematica* we get $b_H^{q,(1)}(\hbar) = -64\hbar^2/(4+\pi^2)^2 \approx -0.3327\hbar^2$.

To evaluate $b_H^{q,(2)}(\hbar)$ using (56), we first evaluate the $\int_0^{2\pi/\omega_c}$ integral. We have

$$b_H^{q,(2)}(\hbar) = \int_0^\infty\Big(\frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c}\sqrt{2\hbar}\mathcal{E}_{t+s}2\big(\eta_{t+s}^\hbar|_{-1}\big)[\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}|_{-1}]\big(\eta_t^\hbar|_{-1}\big)^2 dt\Big)ds$$
$$= -\frac{4\hbar^2}{4+\pi^2}\int_0^\infty[2\pi+\pi\cos(\pi s)+2\sin(\pi s)][\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}|_{-1}]ds.$$

The $\int_0^\infty$ integral can be evaluated numerically by simulating the unperturbed system with the initial condition $(I-\pi)\mathbf{1}_{\{0\}}$, i.e., $\mathbf{1}_{\{0\}} - \Phi\hat{\Psi}$. We get $b_H^{q,(2)}(\hbar) \approx -0.7893\hbar^2$.

D. Verification by numerical simulations

This section illustrates the results of Theorems 1 and 2 using numerical simulations and also shows how the averaged $\hbar$ process can be used to gain information about the original x dynamics (recall Remark 1). For details of the numerical scheme see Appendix E.

Consider

$$dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon^2\gamma_c x^3(t-1)dt + \varepsilon\sigma dW + \varepsilon\gamma_q x^2(t-1)dt. \tag{59}$$

Draw a random sample of size N_{samp} with $\hbar$ values $\{\hbar_i^0\}_{i=1}^{N_{\mathrm{samp}}}$. Simulate them according to

$$d\hbar(t) = \left(b_H + b_H^{q,(1)} + b_H^{q,(2)}\right)(\hbar(t))dt + \sigma_H(\hbar(t))dW \tag{60}$$

for $0 \leqslant t \leqslant T_{\mathrm{end}}$, where b_H and σ_H are obtained from (29) and (30) and $b_H^{q,(i)}$ are obtained in Sec. V C:

$$\begin{aligned}\left(b_H + b_H^{q,(1)} + b_H^{q,(2)}\right)(\hbar) &= 2\hat{\Psi}_1\hat{\Psi}_2\sigma^2 - \gamma_c\tfrac{3}{2}[i(\hat{\Psi}_1 - \hat{\Psi}_2)]\hbar^2 \\ &\quad - \gamma_q^2(0.3327 + 0.7893)\hbar^2, \\ \sigma_H^2(\hbar) &= 4\hat{\Psi}_1\hat{\Psi}_2\sigma^2\hbar.\end{aligned} \tag{61}$$

Fix ε. Simulate (59) for $0 \leqslant t \leqslant T_{\mathrm{end}}/\varepsilon^2$ using the initial history $\{\sqrt{2\hbar_i^0}\cos(\omega_c\bullet)\}_{i=1}^{N_{\mathrm{samp}}}$.

Fix a number H^* and let τ^ε be the first time $|x(t)|$ exceeds $\sqrt{2H^*}$ and $\tau^\hbar$ be the first time $\hbar(t)$ exceeds H^*, i.e.,

$$\tau^\varepsilon := \inf\{t \geqslant 0 : |x(t)| \geqslant \sqrt{2H^*}\},$$
$$\tau^\hbar := \inf\{t \geqslant 0 : \hbar(t) \geqslant H^*\}.$$

We can check whether the following pairs are close: (i) the distribution of $\mathfrak{h}(\Pi_{T_{\mathrm{end}}/\varepsilon^2}x)$ from (59) [where $\mathfrak{h}$ is defined in (18)] and the distribution of $\hbar(T_{\mathrm{end}})$ from (60) and (ii) the distribution of $\varepsilon^2\tau^\varepsilon$ and the distribution of $\tau^\hbar$. We take $\varepsilon = 0.025$, $H^* = 1.5$, $T_{\mathrm{end}} = 2$, $N_{\mathrm{samp}} = 4000$, and $\sqrt{2\{\hbar_i^0\}_{i=1}^{N_{\mathrm{samp}}}} = 1.2$. Figures 4 and 5 answer the above questions. Three cases are considered with $\sigma = 1$ fixed: ($\gamma_q = 0, \gamma_c = 0$), ($\gamma_q = 0, \gamma_c = 1$), and ($\gamma_q = 1/\sqrt{3}, \gamma_c = 0$).

From the figures we can see that it is enough to study the averaged equations for $\mathfrak{h}(\Pi_t x)$ to get a good approximation of the behavior of x. The distribution of $\mathfrak{h}(\Pi_t x)$ (note that $\sqrt{2\mathfrak{h}}$ gives the amplitude of oscillations) is well predicted by the distribution of the averaged system $\hbar$ and the distribution of time taken by x to exceed a threshold $\sqrt{2H^*}$ is well predicted by the time taken by the averaged process $\hbar$ to exceed H^*. Because the averaged equations do not contain any delay, they are easier to analyze and simulate numerically.

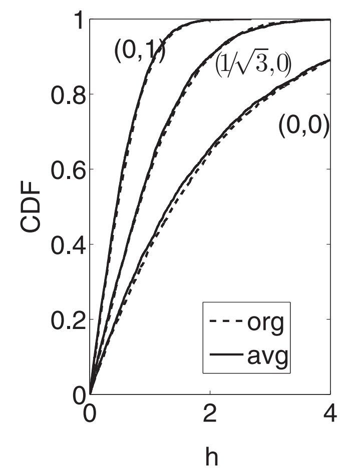

FIG. 4. Cumulative distribution function of $\mathfrak{h}(\Pi_{2/\varepsilon^2}x)$ (org) and $\hbar(2)$ (avg). The numbers in parentheses are (γ_q,γ_c) values.

VI. OTHER KINDS OF NOISE

Here we consider equations of the form

$$\begin{aligned}dx(t) &= L_0(\Pi_t x)dt + \varepsilon\sigma(\xi_t)F(\Pi_t x)dt, \quad t \geqslant 0 \\ \Pi_0 x &= \varphi \in \mathcal{C},\end{aligned} \tag{62}$$

where $F : \mathcal{C} \to \mathbb{R}^n$ is Lipschitz continuous, with at most linear growth and three bounded derivatives; ξ is a noise process whose state space is denoted by $\mathbf{M}$; and $\sigma : \mathbf{M} \to \mathbb{R}$.

We make the following assumptions about the noise ξ.

Assumption 2. The noise ξ is an $\mathbf{M}$-valued time-homogeneous Markov process with transition probability function ν given by

$$\nu(t,\xi,B) = \mathbb{P}\{\xi_t \in B | \xi_0 = \xi\}$$

for B a Borel subset of $\mathbf{M}$. There exist a unique invariant probability measure $\bar{\nu}$ and positive constants c_1 and c_2 such

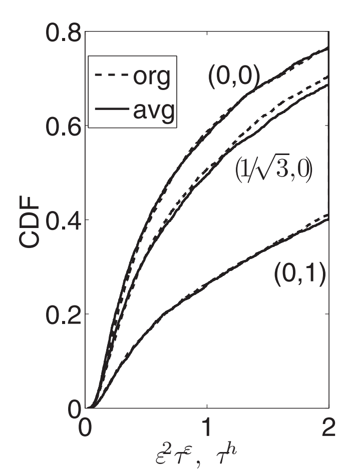

FIG. 5. Cumulative distribution function of $\varepsilon^2\tau^\varepsilon$ (org) and CDF of $\tau^\hbar$ (avg). The numbers in parentheses are (γ_q,γ_c) values. The CDF value at $\varepsilon^2\tau^\varepsilon = 2$ indicates the fraction of the sample whose modulus exceeded $\sqrt{2H^*}$ before the time $2/\varepsilon^2$.

that for all $t \geqslant 0$,

$$\sup_{\xi\in\mathbf{M}} \int_{\mathbf{M}} |\nu(t,\xi,d\zeta) - \bar{\nu}(d\zeta)| \leqslant c_1 e^{-c_2 t},$$

i.e., the transition probability density converges to a stationary density exponentially fast. The function σ is bounded, and is such that $\int_{\mathbf{M}} \sigma(\xi)\bar{\nu}(d\xi) = 0$.

Other requirements are that $\mathbf{M}$ is a locally compact separable metric space and the transition semigroup is a Feller semigroup with $\sigma(\cdot)$ in the domain of the infinitesimal generator. For example, a finite-state continuous-time Markov chain satisfies the above requirements.

The autocorrelation of the noise process ξ is denoted by R:

$$R(s) = \int_{\mathbf{M}} \sigma(\xi)\left(\int_{\mathbf{M}} \sigma(\zeta)\nu(s,\xi,d\zeta)\right)\bar{\nu}(d\xi). \tag{63}$$

For the perturbed system (62), $\mathfrak{h}(\Pi_t x)$ varies slowly compared to x. Changes in $\mathfrak{h}(\Pi_t x)$ are significant only for times of order $1/\varepsilon^2$. Hence, we rescale time and write $X^\varepsilon(t) = x(t/\varepsilon^2)$, where x is governed by (62). Also, we write $\xi_t^\varepsilon = \xi(t/\varepsilon^2)$.

Using the segment extractor Π_t^ε defined in (19), X^ε satisfies

$$dX^\varepsilon(t) = \frac{1}{\varepsilon^2}L_0(\Pi_t^\varepsilon X^\varepsilon)dt + \frac{1}{\varepsilon}\sigma(\xi_t^\varepsilon)F(\Pi_t^\varepsilon X^\varepsilon)dt, \quad t \geqslant 0$$
$$\Pi_0^\varepsilon X^\varepsilon = \varphi \in \mathcal{C}. \tag{64}$$

Write $\mathcal{H}^\varepsilon(t) := \mathfrak{h}(\Pi_t^\varepsilon X^\varepsilon)$. Then $\mathcal{H}^\varepsilon(t)$ satisfies

$$d\mathcal{H}^\varepsilon(t) = \frac{1}{\varepsilon}\sigma(\xi_t^\varepsilon)b(\Pi_t^\varepsilon X^\varepsilon)dt, \quad \mathcal{H}^\varepsilon(0) = \mathfrak{h}(\varphi), \tag{65}$$

where

$$b(\eta) = E(\eta)F(\eta), \tag{66}$$

with E defined in (24).

Using the technique of martingale problem, we can prove[13] the following result (a sketch of proof is given in Appendix D).

Theorem 3. Under the conditions on F and noise ξ listed before, the probability distribution of $\mathcal{H}^\varepsilon$ converges, as $\varepsilon \to 0$, to the distribution of the process $\check{h}$, which is the solution of the SDE

$$d\check{h}(t) = b_H(\check{h}(t))dt + \sigma_H(\check{h}(t))dW(t), \quad \check{h}(0) = \mathfrak{h}(\varphi),$$

with coefficients b_H and σ_H given by

$$\sigma_H^2(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} 2b(\eta_t^\hbar)\left(\int_0^\infty R(s)b(\eta_{t+s}^\hbar)ds\right)dt,$$
$$b_H(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c}\left(\int_0^\infty R(s)\big(\hat{T}(s)\mathbf{1}_{\{0\}}F(\eta_t^\hbar)\cdot\nabla\big)b(\eta_{t+s}^\hbar)ds\right)dt,$$

where $\eta_t^\hbar$ is defined in (25).

We urge the reader to study Appendix D to gain intuition about the process of obtaining the coefficients b_H and σ_H. Akin to the formulas (55) and (56), the coefficient b_H can be written more explicitly as

$$b_H(\hbar) = \frac{1}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt\int_0^\infty ds\left(2R(s)\big(\hat{\Psi}F(\eta_t^\hbar)\big)^*\begin{bmatrix}0 & e^{i\omega_c s}\\ e^{-i\omega_c s} & 0\end{bmatrix}\hat{\Psi}F(\eta_{t+s}^\hbar)\right)$$
$$+\frac{\sqrt{2\hbar}}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt\int_0^\infty ds\,R(s)\sum_{j=1}^n \big(F(\eta_t^\hbar)\big)_j\big((\hat{T}(s)\mathbf{1}_{\{0\}}\underline{e_j})\cdot\nabla\big)\big(\mathcal{E}_{t+s}F(\eta_{t+s}^\hbar)\big),$$

with $\eta_t^\hbar$ defined in (25), $\mathcal{E}$ defined in (57), and $\underline{e_j}$ the unit vector in the jth direction of $\mathbb{R}^n$. Similarly,

$$\sigma_H^2(\hbar) = \frac{4\hbar}{2\pi/\omega_c}\int_0^{2\pi/\omega_c} dt\int_0^\infty ds\big(\mathcal{E}_t F(\eta_t^\hbar)\big)R(s)\big(\mathcal{E}_{t+s}F(\eta_{t+s}^\hbar)\big).$$

It would be easier to do the $\int_0^{2\pi/\omega_c}$ integral before the $\int_0^\infty$ integral.

Analogous results for systems without delay are found in Sec. 4 of [30]. Even systems with delay can be put in the framework of [30]. Equations of the form (62) with $F(0) = 0$ and $\int_{\mathbf{M}} \sigma(\xi)\bar{\nu}(d\xi) \neq 0$ (i.e., noise is not mean zero) are studied in [31].

Remark 5. In Eq. (62) we could have included the deterministic perturbations G and G_q as in Eq. (43), but the averaged drift terms arising from these would be the same as in the previous sections.

[13]Proof of Theorem 3 and a result to the effect that the stable component of the solution is small are not included herein.

A. Linear perturbations

When $F(\eta) = L_1\eta$, where $L_1 : \mathcal{C} \to \mathbb{R}^n$ is a linear operator, the expressions for b_H and σ_H can be more explicitly evaluated using the autocorrelation function as follows. Let Υ be the 2×2 matrix $\Upsilon_{ij} = \hat{\Psi}_i L_1 \Phi_j$. Let

$$R_0 = \int_0^\infty R(s)ds,$$

$$R_{2c} = \int_0^\infty R(s)\cos(2\omega_c s)ds,$$

$$\hat{R}_1 = \int_0^\infty R(s)e^{-i\omega_c s}\hat{\Psi}_1 L_1(\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}L_1\Phi_1)ds,$$

$$\hat{R}_2 = \int_0^\infty R(s)e^{i\omega_c s}\hat{\Psi}_2 L_1(\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}L_1\Phi_2)ds.$$

Then

$$b_H(\hbar) = C_b\hbar, \quad \sigma_H^2(\hbar) = C_\sigma \hbar^2,$$

where

$$C_b = (\Upsilon_{11} + \Upsilon_{22})^2 R_0 + 4\Upsilon_{12}\Upsilon_{21} R_{2c} + \hat{R}_1 + \hat{R}_2,$$
$$C_\sigma = 2[(\Upsilon_{11} + \Upsilon_{22})^2 R_0 + 2\Upsilon_{12}\Upsilon_{21} R_{2c}].$$

Remark 6. Note that if we had totally ignored the stable modes, i.e., if we had set $(I-\pi)\Pi_t^\varepsilon X^\varepsilon = 0$ at the very beginning of the analysis, we would not have the terms $\hat{R}_1$ and $\hat{R}_2$.

The Lyapunov exponent for the averaged equation

$$d\hbar(t) = b_H(\hbar)dt + \sigma_H(\hbar)dW \tag{67}$$

can be calculated to be

$$\lambda_{\text{avg}} = C_b - \tfrac{1}{2}C_\sigma = 2\Upsilon_{12}\Upsilon_{21}R_{2c} + \hat{R}_1 + \hat{R}_2. \tag{68}$$

Using singular perturbation methods and the Furstenberg-Khasminskii formula, the following theorem for scalar processes is proved in [32,33].

Theorem 4. Consider (62) with $F(\eta) = L_1(\eta)$, where $L_1 : \mathcal{C} \to \mathbb{R}$ is linear. Let the top Lyapunov exponent of the process x be defined by

$$\lambda^\varepsilon := \limsup_{t\to\infty} \frac{1}{t} \ln \sup_{s\in[t-r,t]} |x(s)|. \tag{69}$$

Then $\lambda^\varepsilon = \varepsilon^2 \frac{1}{2}\lambda_{avg} + O(\varepsilon^3)$.

The same can be said about vector-valued processes.

B. Verification by numerical simulation

Consider the system

$$dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon\sigma(\xi_t)x(t-1)dt. \tag{70}$$

Let ξ be a two-state symmetric Markov chain with switching rate $g/2$, i.e.,

$$\lim_{t\downarrow 0} \frac{1}{t}P_{1\to 2}(t) = g/2 = \lim_{t\downarrow 0}\frac{1}{t}P_{2\to 1}(t), \tag{71}$$

where $P_{i\to j}(t)$ is the probability of transition from state i to state j in time t. Let $\sigma(\xi = 1) = -\sigma(\xi = 2) = \sigma_0$. We then have the autocorrelation as $R(s) = \sigma_0^2 e^{-gs}$.

We consider two cases $g = 2$ and $g = 6$ with $\sigma_0 = 1$. The averaged equations are

$$d\hbar(t) = 0.3734\hbar dt + \sqrt{0.9873}\hbar dW \quad \text{for } g = 2,$$
$$d\hbar(t) = 0.1715\hbar dt + \sqrt{0.4245}\hbar dW \quad \text{for } g = 6.$$

Using the same notation as in Sec. V D, we fix $\varepsilon = 0.025$, $T_{\text{end}} = 1$, $H^* = 1$, $N_{\text{samp}} = 4000$, and $\sqrt{2\{\hbar_i^0\}_{i=1}^{N_{\text{samp}}}} = 1$. Equation (70) is simulated for time $T_{\text{end}}/\varepsilon^2$ with initial history $\{\sqrt{2\hbar_i^0}\cos(\omega_c\bullet)\}_{i=1}^{N_{\text{samp}}}$. We obtain Figs. 6 and 7, which show that the averaged system gives a good approximation of the original system. For details of the numerical scheme see Appendix E.

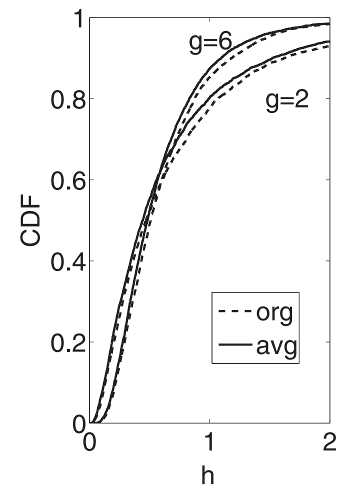

FIG. 6. Cumulative distribution function of $\mathfrak{h}(\Pi_{1/\varepsilon^2}x)$ (org) and $\hbar(1)$ (avg).

VII. DISCUSSION

Delay equations with noise perturbations as considered in Sec. VI display interesting similarities to nondelay systems. For example, Ref. [34] considers coupled oscillators with one of the oscillators stable, in the following form. Let J be the symplectic matrix $\left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$, I be the 2×2 identity matrix, and O be the 2×2 zero matrix. Let $x \in \mathbb{R}^4$ be governed by

$$\dot{x}(t) = \begin{pmatrix}\omega_1 J & O\\ O & -\delta I + \omega_2 J\end{pmatrix}x(t) + \varepsilon\sigma(\xi(t))\begin{pmatrix}K & M\\ N & L\end{pmatrix}x(t), \tag{72}$$

where K,L,M,N are 2×2 matrices. The oscillator with frequency ω_1 is coupled to the stable oscillator of frequency ω_2. Reference [34] shows that the Lyapunov exponent of the above system can be written in terms of quantities analogous to $R_0, R_{2c}, \hat{R}_i$ defined in Sec. VI A. Further, they show that both

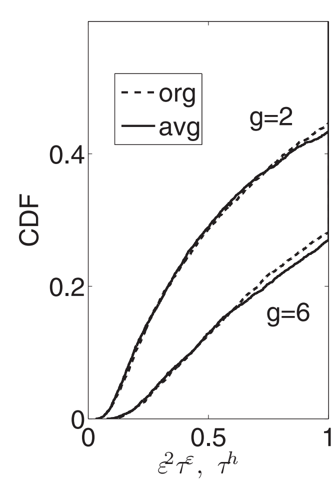

FIG. 7. Cumulative distribution function of $\varepsilon^2\tau^\varepsilon$ (org) and CDF of $\tau^\hbar$ (avg). The CDF value at $\varepsilon^2\tau^\varepsilon = 1$ indicates the fraction of particles whose modulus exceeded $\sqrt{2H^*}$ before the time $1/\varepsilon^2$.

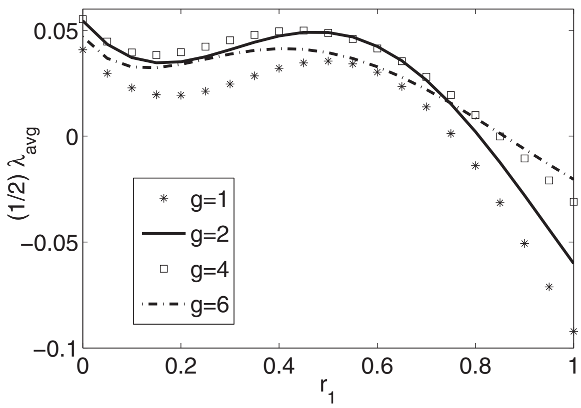

FIG. 8. Lyapunov exponent $\frac{1}{2}\lambda_{\text{avg}}$ as a function of the delay in the perturbation r_1 and the rate of switching of the noise g for Eq. (73). The top Lyapunov exponent λ^ε is close to $\varepsilon^2\frac{1}{2}\lambda_{\text{avg}}$ by Theorem 4. Note that both $\lambda_{\text{avg}} < 0$ (stabilization) and $\lambda_{\text{avg}} > 0$ (destabilization) are possible.

stabilization and destabilization are possible depending on the matrix coefficients K, M, and N.

The delay system that we considered under Assumption 1 can be thought of as a coupled oscillator system with one critical mode and infinitely many stable modes (the characteristic equation has a pair of roots $\pm i\omega_c$ and all other roots have a negative real part). The Lyapunov exponent obtained in (68) suggests that both stabilization and destabilization are possible. To illustrate this, consider

$$dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon\sigma(\xi_t)x(t-r_1)dt, \qquad (73)$$

with ξ a two-state symmetric Markov chain with states $\sigma(\xi) \in \{+1,-1\}$ and rate of switching $g/2$ [defined in (71)]. Theorem 4 says that the Lyapunov exponent λ^ε [defined in (69)] is close to $\varepsilon^2\frac{1}{2}\lambda_{\text{avg}}$, where λ_{avg} is evaluated in (68). Figure 8 shows how $\frac{1}{2}\lambda_{\text{avg}}$ varies with the delay in the perturbation r_1 and the rate of switching g of the two-state Markov chain. Note that both $\lambda_{\text{avg}} < 0$ (stabilization) and $\lambda_{\text{avg}} > 0$ (destabilization) are possible.

Even the white noise allows for both possibilities. As mentioned in Sec. IV B, the Lyapunov exponent λ_{avg} corresponding to (31) equals $-\text{Re}[(\hat{\Psi}_1 L_1 \Phi_1)^2]$. Applying to $dx(t) = -\frac{\pi}{2}x(t-1)dt + \varepsilon x(t-r_1)dW$, we find that $\lambda_{\text{avg}} < 0$ for $r_1 < 0.8609$ and $\lambda_{\text{avg}} > 0$ for $0.8609 < r_1 \leqslant 1$.

The above examples raise the question whether stabilization or destabilization is possible when the noise is additive, i.e., the coefficient F is a constant independent of the state x. To answer this question, as an example consider (59). The corresponding averaged equation is (60), with the averaged drift and diffusion coefficients given by (61). Note that the diffusion σ_H^2 is zero only if $\hbar = 0$ and when $\hbar = 0$, the drift is $2\hat{\Psi}_1\hat{\Psi}_2\sigma^2 = 2|\hat{\Psi}_1|^2\sigma^2 > 0$. Thus additive noise destroys the fixed points and hence stabilization is not possible.

The averaging results presented in this article allow us to simplify the analysis of delay systems at the verge of instability. The averaged dynamics does not involve any delay and hence is easier to analyze. Using numerical simulations, we have amply demonstrated the usefulness of the theoretical results in approximating the probability distribution of the time-delay system with that of the averaged system. In Sec. IV C we have shown how these results would be useful in computing an approximation to the shift of bifurcation thresholds in the presence of noise.

Note that the $\mathcal{H}$ process only deals with the amplitude of oscillations and does not concern the phase. In applications where phase is also important, the following methods might be useful.

One method is the study of the individual projections $z_i(t) = \langle \Psi_i, \Pi_t x\rangle$ without averaging. However, to study the behavior of x for times of order $1/\varepsilon^2$, the z_i processes should also be studied for times of order $1/\varepsilon^2$. However, the averaged $\mathcal{H}$ equations need to be simulated only for times of order 1 to study the amplitude of oscillations of x for times of order $1/\varepsilon^2$.

Another method is the study of the slowly varying process $\alpha(t) := \frac{z_1(t)}{z_2(t)}e^{-2i\omega_c t}$. As an example consider the scalar process $dx = L_0(\Pi_t x)dt + \varepsilon\sigma dW$, with L_0 satisfying Assumption 1. Let $z_i(t) = \langle \Psi_i, \Pi_t x\rangle$ and let δ_t be the angle of the complex number $z_1(t)$. Since the dynamics of z_i is predominantly an oscillation with frequency ω_c, the quantity $\alpha(t) = \cos[2(\delta_t - \omega_c t)] + i\sin[2(\delta_t - \omega_c t)]$ is slowly varying. Writing $\alpha^\varepsilon(t) = \alpha(t/\varepsilon^2)$ and $z_i^\varepsilon(t) = z(t/\varepsilon^2)$ and applying It's formula, we find that $\mathcal{H}^\varepsilon$ and α^ε have the same distribution as the equations

$$d\mathcal{H}^\varepsilon = 2\sigma^2\hat{\Psi}_1\hat{\Psi}_2 dt + 2\sigma\left(\hat{\Psi}_1 z_2^\varepsilon + \hat{\Psi}_2 z_1^\varepsilon\right)dW,$$

$$d\alpha^\varepsilon = \sigma^2\alpha^\varepsilon\left(\frac{\hat{\Psi}_2^2}{\left(z_2^\varepsilon\right)^2} - \frac{2\hat{\Psi}_1\hat{\Psi}_2}{\mathcal{H}^\varepsilon}\right)dt + \sigma\alpha^\varepsilon\left(\frac{\hat{\Psi}_1}{z_1^\varepsilon} - \frac{\hat{\Psi}_2}{z_2^\varepsilon}\right)dW,$$

where z^ε evolve according to $dz^\varepsilon = \frac{1}{\varepsilon^2}Bz^\varepsilon dt + \sigma\hat{\Psi}dW$. Heuristically, on averaging the fast oscillations of z_i^ε we get[14] that the distribution of $(\mathcal{H}^\varepsilon, \alpha^\varepsilon)$ converges as $\varepsilon \to 0$ to the distribution of $(\hbar, \alpha^\circ)$,

$$d\hbar = 2\sigma^2\hat{\Psi}_1\hat{\Psi}_2 dt + 2\sigma\sqrt{\hat{\Psi}_1\hat{\Psi}_2\hbar}\,dW_1,$$
$$d\alpha^\circ = -(2\sigma^2\hat{\Psi}_1\hat{\Psi}_2/\hbar)\alpha^\circ dt + i\sigma\alpha^\circ\sqrt{4\hat{\Psi}_1\hat{\Psi}_2/\hbar}\,dW_2, \qquad (74)$$

where W_1 and W_2 are independent $\mathbb{R}$-valued Wiener processes. The phase of the oscillation of $x^\varepsilon(t) = x(t/\varepsilon^2)$ is $\delta_{t/\varepsilon^2} = (\omega_c t/\varepsilon^2) + \frac{1}{2}\arg(\alpha^\varepsilon(t))$, the distribution of which can be approximated by the distribution of $(\omega_c t/\varepsilon^2) + \frac{1}{2}\arg(\alpha^\circ(t))$ where α° is the process from (74). Writing $\beta^\circ := \frac{1}{2}\arg(\alpha^\circ(t)) = \frac{1}{2i}\ln\alpha^\circ(t)$ and applying It's formula, we find that β° evolves according to $d\beta^\circ = \sigma\sqrt{\hat{\Psi}_1\hat{\Psi}_2/\hbar}\,dW_2$.

We conclude this article with one other related work in this context. The instability in Assumption 1 is not the only kind of instability possible. For example, one can have the following.

Assumption 3. The characteristic equation (9) has zero as a simple root and all other roots have negative real parts.

[14]Note that z_i^ε are fast oscillating and hence $(z_i^\varepsilon)^2$ are also fast oscillating, but $2z_1^\varepsilon z_2^\varepsilon = \mathcal{H}^\varepsilon$ is slow varying. Denoting the average by $\mathbb{A}$, we have $\mathbb{A}[(\hat{\Psi}_1 z_2^\varepsilon + \hat{\Psi}_2 z_1^\varepsilon)^2] = \hat{\Psi}_1^2\mathbb{A}[(z_2^\varepsilon)^2] + \hat{\Psi}_2^2\mathbb{A}[(z_1^\varepsilon)^2] + \hat{\Psi}_1\hat{\Psi}_2\mathcal{H}^\varepsilon = \hat{\Psi}_1\hat{\Psi}_2\mathcal{H}^\varepsilon$. Similarly, $\mathbb{A}[(\frac{\hat{\Psi}_1}{z_1^\varepsilon} - \frac{\hat{\Psi}_2}{z_2^\varepsilon})^2] = -4\hat{\Psi}_1\hat{\Psi}_2/\mathcal{H}^\varepsilon$ and $\mathbb{A}[(\hat{\Psi}_1 z_2^\varepsilon + \hat{\Psi}_2 z_1^\varepsilon)(\frac{\hat{\Psi}_1}{z_1^\varepsilon} - \frac{\hat{\Psi}_2}{z_2^\varepsilon})] = 0$.

The analysis under Assumption 3 is similar to the analysis in previous sections. Choose $\underline{d}$ such that $\Delta(0)\underline{d} = 0_{n\times 1}$ and $\underline{d_2}$ such that $\underline{d_2}\Delta(0) = 0_{1\times n}$. Define Φ by the constant $\Phi(\bullet) = \underline{d}$ and Ψ by $\Psi(\bullet) = c\underline{d_2}$, where the constant c is chosen so that $\langle\Psi,\Phi\rangle = 1$ for the bilinear form in (12). The space $\mathcal{C}$ can be split as $\mathcal{C} = P \oplus Q$, where P is the space spanned by the constant function Φ. The projection operator is $\pi : \mathcal{C} \to P$ given by $\pi(\eta) = \Phi\langle\Psi,\eta\rangle$. Define $\hat{\Psi} \stackrel{\text{def}}{=} \Psi(0)$. Let $\hat{T}$ and $\mathbf{1}_{\{0\}}$ be as defined in Sec. V A. For the unperturbed system (5), writing $\Pi_t x = \pi\Pi_t x + (1-\pi)\Pi_t x = \Phi z(t) + (I-\pi)\Pi_t x$, we find that $\dot{z} = 0$ and $\|(I-\pi)\Pi_t x\|$ decays exponentially fast. So, defining $\mathfrak{h}(\eta) = \langle\Psi,\eta\rangle$, we find that $\mathcal{H}(t) = \mathfrak{h}(\Pi_t x)$ is a constant for the unperturbed system (note that $\mathcal{H}$ is the same as z). Now consider equations of the form (43). Akin to the condition (44), we need to impose that

$$\hat{\Psi} G_q(\Phi h) = 0 \quad \text{for } h \in \mathbb{R}. \tag{75}$$

[If the above is not imposed, then the dynamics of $\mathcal{H}$ for times of order $1/\varepsilon$ converges to that of a deterministic process[15] given by $\dot{\mathcal{H}} = \hat{\Psi} G_q(\Phi\mathcal{H})$.] When (75) is imposed, significant changes in $\mathcal{H}$ occur only for times of order $1/\varepsilon^2$. So writing $X^\varepsilon(t) = x(t/\varepsilon^2)$, we find that X^ε has the same probability distribution as the process satisfying (45). Defining $\mathcal{H}^\varepsilon(t) := \mathfrak{h}(\Pi_t^\varepsilon X^\varepsilon)$ and using Itô's formula, we get that $\mathcal{H}$ satisfies (46) with $b^{q,(1)}(\eta) = \hat{\Psi} G_q(\pi\eta) = 0$, $b^{q,(2)}(\eta) = \hat{\Psi}(G_q(\eta) - G_q(\pi\eta))$, $b(\eta) = \hat{\Psi} G(\eta)$, and $\sigma(\eta) = \hat{\Psi} F(\eta)$. It can be shown that a result analogous to Theorem 2 holds with the averaged drift and diffusion coefficients given by $b_H(\hbar) = \hat{\Psi} G(\Phi h)$, $\sigma_H^2(\hbar) = [\hat{\Psi} F(\Phi\hbar)]^2$, $b_H^{q,(1)} = 0$, and

$$b_H^{q,(2)}(\hbar) = \int_0^\infty ((\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}} G_q(\Phi\hbar)) \cdot \nabla)\hat{\Psi} G_q(\Phi\hbar) ds. \tag{76}$$

For scalar systems the condition (75) would necessarily mean that $G_q(\Phi\hbar) = 0$, which would result in $\mathbf{1}_{\{0\}} G_q(\Phi\hbar) = 0$ and hence $b_H^{q,(2)} = 0$. This means that, when (43) is scalar valued, G_q terms would have a negligible effect on the dynamics on P subspace for times of order $1/\varepsilon^2$.

Reference [23] considers scalar systems satisfying Assumption 3, but do not impose (75). Reference [23] gives a method to construct higher-order corrections to the center manifold in the presence of periodic forcing and white noise. They show that having higher-order corrections in the center manifold would improve accuracy of reconstructing the trajectories (Figs. 2 and 6 in [23]). However, these corrections should be evaluated through numerical simulations of a delay equation; for example, the correction to the center manifold in Eq. (52) of [23] should be numerically simulated. In scalar equations this task can be circumvented by employing series solutions as in Eq. (53) of [23]. However, for a multidimensional system this involves evaluating a reasonable number of eigenvalues and eigenvectors of the linear delay system. Further, the computations require memory for storing the history of Brownian motion for computing the convolutions [Eq. (55) in [23]]. The extra effort required from the methods in [23] allows one to reconstruct trajectories. The averaging methods presented in our article deal with distributions alone in the limit of small ε and cannot reconstruct trajectories.

Finally, for completeness, we consider equations of the form (62) with Assumption 3. In this case it can be shown that Theorem 3 holds with

$$b_H(\hbar) = \left(\int_0^\infty R(s)ds\right)(\mathbf{1}_{\{0\}} F(\Phi\hbar)\cdot\nabla)\hat{\Psi} F(\Phi h),$$
$$\sigma_H^2(\hbar) = 2\left(\int_0^\infty R(s)ds\right)(\hat{\Psi} F(\Phi h))^2. \tag{77}$$

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the suggestions of Professor Volker Wihstutz. The work was financially supported by the National Science Foundation under Grants No. CMMI 1000906 and No. 1030144. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

APPENDIX A: ERRORS IN [16,17] AND [10,18]

1. Errors in [16,17]

One of the equations considered in [17] is

$$dX^\varepsilon(t) = \frac{1}{\varepsilon^2}(-\alpha X^\varepsilon(t) + \beta X^\varepsilon(t-\varepsilon^2\tau))dt + X^\varepsilon(t)dW(t), \tag{A1}$$

where W is a Wiener process.[16] The above system is studied as a perturbation of the linear system

$$\dot{x}(t) = \frac{1}{\varepsilon^2}(-\alpha x(t) + \beta x(t-\varepsilon^2\tau)). \tag{A2}$$

Seeking a solution of the form $e^{\lambda t/\varepsilon^2}$, the characteristic equation is found to be $\lambda = -\alpha + \beta e^{-\lambda\tau}$. Let the parameters $\alpha,\beta,\tau = \tau_c + \varepsilon^2\tau_2$ be such that when $\tau_2 = 0$, a pair of roots $\pm i\omega$ is on the imaginary axis and all other roots are with a negative real part. In this scenario we have $i\omega = -\alpha + \beta e^{-i\omega\tau_c}$, which on solving gives[17]

$$\omega = \sqrt{\beta^2 - \alpha^2}, \quad \beta\cos(\omega\tau_c) = \alpha, \quad \beta\sin(\omega\tau_c) = -\omega. \tag{A3}$$

Reference [17] employs multiscale analysis and for that purpose is written[18]

$$dW(t) = \mathcal{K}_0 dW_0(t) + \mathcal{K}_{2,1}\cos\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,1}(t)$$
$$+ \mathcal{K}_{2,2}\sin\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,2}(t), \tag{A4}$$

[15] A stochastic limit can be obtained by strengthening the noise.

[16] This is a time-rescaled version of Eq. (1.1) in [17]. The analysis below appears in Sec. 2 of [17].

[17] This is Eq. (2.1) in [17].

[18] This is Eq. (2.11) in [17].

where W_i are independent Brownian motions. Reference [17] assumes that the solution X^ε is of the form[19]

$$X^\varepsilon(t) = A(t)\cos(\omega t/\varepsilon^2) + B(t)\sin(\omega t/\varepsilon^2). \tag{A5}$$

Here A,B vary on a different scale (in the spirit of multiscale analysis) than cosine and sine.

According to [17], on one the hand, applying Itô's formula we have[20]

$$dX^\varepsilon = \frac{1}{\varepsilon^2}(-\omega\mathfrak{s}A + \omega\mathfrak{c}B)dt + \mathfrak{c}dA + \mathfrak{s}dB, \tag{A6}$$

where $\mathfrak{c} = \cos(\omega t/\varepsilon^2)$ and $\mathfrak{s} = \sin(\omega t/\varepsilon^2)$. On the other hand, since X^ε must satisfy (A1) we must have[21]

$$\begin{aligned} dX^\varepsilon = \frac{1}{\varepsilon^2}\bigg\{ -\alpha(\mathfrak{c}A + \mathfrak{s}B) + \beta\bigg[A_\tau \cos\left(\frac{\omega(t-\varepsilon^2\tau)}{\varepsilon^2}\right) \\ + B_\tau \sin\left(\frac{\omega(t-\varepsilon^2\tau)}{\varepsilon^2}\right)\bigg]\bigg\}dt \\ + (\mathfrak{c}A + \mathfrak{s}B)\bigg[\mathcal{K}_0 dW_0(t) + \mathcal{K}_{2,1}\cos\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,1}(t) \\ + \mathcal{K}_{2,2}\sin\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,2}(t)\bigg], \end{aligned} \tag{A7}$$

where A_τ means $A(t-\varepsilon^2\tau)$.

Using $\tau = \tau_c + \varepsilon^2\tau_2$ and (A3) we have

$$\beta\cos\left(\frac{\omega(t-\varepsilon^2\tau)}{\varepsilon^2}\right) = (\alpha\mathfrak{c} - \omega\mathfrak{s}) + \varepsilon^2\omega\tau_2(\omega\mathfrak{c} + \alpha\mathfrak{s}), \tag{A8}$$

$$\beta\sin\left(\frac{\omega(t-\varepsilon^2\tau)}{\varepsilon^2}\right) = (\omega\mathfrak{c} + \alpha\mathfrak{s}) + \varepsilon^2\omega\tau_2(-\alpha\mathfrak{c} + \omega\mathfrak{s}). \tag{A9}$$

Using the above in (A7) and comparing the resulting equation with (A6) we have

$$\begin{aligned} &\frac{1}{\varepsilon^2}[-\alpha(\mathfrak{c}A + \mathfrak{s}B) + A_\tau(\alpha\mathfrak{c} - \omega\mathfrak{s}) + B_\tau(\alpha\mathfrak{s} + \omega\mathfrak{c})]dt + \omega\tau_2[\omega(\mathfrak{c}A_\tau + \mathfrak{s}B_\tau) + \alpha(\mathfrak{s}A_\tau - \mathfrak{c}B_\tau)]dt \\ &\quad + (\mathfrak{c}A + \mathfrak{s}B)\bigg[\mathcal{K}_0 dW_0(t) + \mathcal{K}_{2,1}\cos\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,1}(t) + \mathcal{K}_{2,2}\sin\left(\frac{2\omega t}{\varepsilon^2}\right)dW_{2,2}(t)\bigg] \\ &\quad - \frac{1}{\varepsilon^2}(-\omega\mathfrak{s}A + \omega\mathfrak{c}B)dt - \mathfrak{c}dA - \mathfrak{s}dB = 0. \end{aligned} \tag{A10}$$

Reference [17] then multiplies the above with $\mathfrak{c}$ or $\mathfrak{s}$ and integrates over a time period, while treating A and B as constants, to get the following equations:

$$\begin{aligned} dA &= -\alpha\hat{d}A - \omega\hat{d}B + \omega\tau_2(\omega A_\tau - \alpha B_\tau)dt + A\mathcal{K}_0 dW_0 + \tfrac{1}{2}A\mathcal{K}_{2,1}dW_{2,1} + \tfrac{1}{2}B\mathcal{K}_{2,2}dW_{2,2}, \\ dB &= \omega\hat{d}A - \alpha\hat{d}B + \omega\tau_2(\alpha A_\tau + \omega B_\tau)dt + B\mathcal{K}_0 dW_0 - \tfrac{1}{2}B\mathcal{K}_{2,1}dW_{2,1} + \tfrac{1}{2}A\mathcal{K}_{2,2}dW_{2,2}, \end{aligned} \tag{A11}$$

where $\hat{d}A$ means $\frac{A(t)-A(t-\varepsilon^2\tau)}{\varepsilon^2}dt$.

In (A11) the constants $\mathcal{K}$ are not yet determined. Reference [17] determines them in the following way: Reference [17] compares the diffusive part of the generator for X^ε and for (A,B). The diffusive part of the generator for (A,B) is

$$\begin{aligned} &(A^2\partial_A\partial_A + B^2\partial_B\partial_B + 2AB\partial_A\partial_B)\mathcal{K}_0^2 \\ &\quad + \tfrac{1}{4}(A^2\partial_A\partial_A + B^2\partial_B\partial_B - 2AB\partial_A\partial_B)\mathcal{K}_{2,1}^2 \\ &\quad + \tfrac{1}{4}(B^2\partial_A\partial_A + A^2\partial_B\partial_B + 2AB\partial_A\partial_B)\mathcal{K}_{2,2}^2. \end{aligned} \tag{A12}$$

The diffusive part of the generator for x is

$$x^2\partial_x\partial_x = (\mathfrak{c}A + \mathfrak{s}B)^2(\mathfrak{c}\partial_A + \mathfrak{s}\partial_B)^2. \tag{A13}$$

Averaging (A13) over one time period, Ref. [17] obtains[22]

$$\frac{3A^2 + B^2}{8}\partial_A\partial_A + \frac{3B^2 + A^2}{8}\partial_B\partial_B + \frac{1}{2}AB\partial_A\partial_B. \tag{A14}$$

Reference [17] equates (A14) and (A12) to find that

$$\mathcal{K}_0 = \frac{1}{2}, \quad \mathcal{K}_{2,1} = \mathcal{K}_{2,2} = \frac{1}{\sqrt{2}}. \tag{A15}$$

Then [17] presents a figure showing that the density of $A(T)\cos(\omega T/\varepsilon^2) + B(T)\sin(\omega T/\varepsilon^2)$, with A,B simulated from (A11), gives a good approximation to the density of $X^\varepsilon(T)$.

The above procedure is not convincing due to the following reasons.

(i) It is not clear whether the error in transferring from (A10) to (A11) would go to zero in some sense as $\varepsilon \to 0$.

(ii) Note that (A11) is still a delay equation and hence there would not be much advantage in simulating A,B compared to simulating X^ε. The delay itself is small $O(\varepsilon^2)$, but the difference $A(t) - A(t-\varepsilon^2\tau)$ is magnified by ε^{-2}.

(iii) Note that, heuristically, the left-hand side of (A4) is a normal random variable with variance dt and hence, for consistency, we must have

$$\mathcal{K}_0^2 + \mathcal{K}_{2,1}^2\cos^2\left(\frac{2\omega t}{\varepsilon^2}\right) + \mathcal{K}_{2,2}^2\sin^2\left(\frac{2\omega t}{\varepsilon^2}\right) = 1. \tag{A16}$$

The above is possible only if we take $|\mathcal{K}_{2,1}| = |\mathcal{K}_{2,2}|$ and set

$$\mathcal{K}_0^2 + \mathcal{K}_{2,1}^2 = 1. \tag{A17}$$

[19] This is Eq. (2.2) in [17].

[20] This is Eq. (2.4) in [17].

[21] This is Eq. (2.5) in [17].

[22] This is Eq. (2.16) in [17].

However, note that (A15) contradicts the consistency equation (A17). We have from (A15) that $\mathcal{K}_0^2 + \mathcal{K}_{2,1}^2 = \frac{3}{4} \neq 1$.

We show by means of numerical simulation that the above procedure is indeed wrong. In (A1) set $\alpha = 0$, $\beta = -\frac{\pi}{2}$, $\tau_c = 1$, and $\tau_2 = 0$. Then $\omega = \frac{\pi}{2}$ and this system satisfies Assumption 1. Equations (A11) in this case become

$$\begin{aligned}\begin{pmatrix} dA \\ dB \end{pmatrix} &= \frac{1}{\varepsilon^2}\begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix}\begin{pmatrix} A(t) - A(t-\varepsilon^2) \\ B(t) - B(t-\varepsilon^2) \end{pmatrix} dt \\ &\quad + \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} dW_{2,0} \\ &\quad + \frac{1}{2\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} dW_{2,1} \\ &\quad + \frac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} dW_{2,2}. \end{aligned} \tag{A18}$$

Numerical simulations show that splitting W into harmonics as in (A4) is unnecessary. For this purpose consider

$$\begin{aligned}\begin{pmatrix} dA \\ dB \end{pmatrix} &= \frac{1}{\varepsilon^2}\begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix}\begin{pmatrix} A(t) - A(t-\varepsilon^2) \\ B(t) - B(t-\varepsilon^2) \end{pmatrix} dt \\ &\quad + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} dW_{2,0}, \end{aligned} \tag{A19}$$

i.e., $\mathcal{K}_0 = 1$ and $\mathcal{K}_{2,1} = 0 = \mathcal{K}_{2,2}$.

We set $\varepsilon = 0.05$ and $T = 1$. The initial condition is $X^\varepsilon(t) = \cos(\omega t/\varepsilon^2)$ for $t \in [-\varepsilon^2, 0]$, i.e., $\Pi_0^\varepsilon X^\varepsilon(\theta) = \cos(\omega\theta)$ for $\theta \in [-1, 0]$, i.e., $A(t) = 1$ for $t \leqslant 0$ and $B(t) = 0$ for $t \leqslant 0$. The cumulative distribution in Fig. 9 is obtained with 2400 realizations.

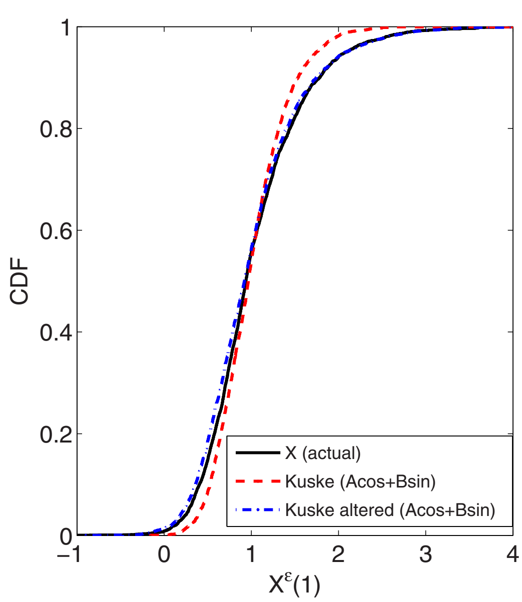

FIG. 9. Cumulative distribution function of $X^\varepsilon(T)$ with $\varepsilon = 0.05$ and $T = 1$: X (actual) is obtained from simulating the original dynamics (A1), Kuske ($A\cos + B\sin$) is $A(T)\cos(\omega T/\varepsilon^2) + B(T)\sin(\omega T/\varepsilon^2)$ obtained from simulating (A18), and Kuske altered ($A\cos + B\sin$) is $A(T)\cos(\omega T/\varepsilon^2) + B(T)\sin(\omega T/\varepsilon^2)$ obtained from simulating (A19).

Figure 9 shows that (A19) better matches the actual dynamics (A1) than (A18). However, note that (A19) is still a delay equation and there is no advantage in simulating (A, B) compared to simulating X.

2. Errors in [10,18]

There are two errors in the analysis of [10,18], one of which is similar in nature to the previous section. We illustrate the errors using a special case of the equation considered in [10].

Reference [10] considers

$$\ddot{x}(t) + x(t) + \eta x(t-1) - \beta\dot{x}(t) = \sqrt{2D}x(t)\xi(t), \tag{A20}$$

where ξ is a white noise process with correlation $\mathbb{E}[\xi(t)\xi(t')] = \delta(t-t')$. For now, let us set $D = 0$. The characteristic equation is $\lambda^2 + 1 + \eta e^{-\lambda} - \beta\lambda = 0$. Given η, solve $\eta\cos\omega = \omega^2 - 1$ for ω and get $\beta_c = -\eta\sin\omega/\omega$. With $\beta = \beta_c$ the system (A20) (with $D = 0$) satisfies Assumption 1 with critical roots of the characteristic equation being $\pm i\omega$. We assume $\beta = \beta_c$.

Reference [10] assumes that the solution is of the form

$$x(t, T) = \varepsilon A(T)\cos\omega t - \varepsilon B(T)\sin\omega t, \tag{A21}$$

where $T = \varepsilon^2 t$ is the slow time scale. Then

$$\begin{aligned} x(t-1, T-\varepsilon^2) &= x(t,T)\cos\omega - (\sin\omega/\omega)\partial_t x(t,T) \\ &\quad - \varepsilon^2\varepsilon\frac{A(T) - A(T-\varepsilon^2)}{\varepsilon^2}\cos[\omega(t-1)] \\ &\quad + \varepsilon^2\varepsilon\frac{B(T) - B(T-\varepsilon^2)}{\varepsilon^2}\sin[\omega(t-1)]. \end{aligned} \tag{A22}$$

However, Ref. [10] sets the last two terms on the right-hand side to zero claiming $A(T) \approx A(T-\varepsilon^2)$ and $B(T) \approx B(T-\varepsilon^2)$. However, as $\varepsilon \to 0$ it is easy to see that (if derivatives of A and B exist) these terms go to $\partial_T A$ and $\partial_T B$, respectively. At which ε should we ignore these and which ε should we consider it as a derivative?

Differentiating, we get

$$\begin{aligned} \dot{x}(t) &= (\varepsilon^2\partial_T + \partial_t)x(t,T) \\ &= \varepsilon^2(\varepsilon\partial_T A\cos\omega t - \varepsilon\partial_T B\sin\omega t) + \partial_t x(t,T), \end{aligned} \tag{A23}$$

$$\begin{aligned} \ddot{x}(t) &= (\varepsilon^2\partial_T + \partial_t)^2 x(t,T) = \varepsilon^4\left(\varepsilon\partial_T^2 A\cos\omega t - \varepsilon\partial_T^2 B\sin\omega t\right) \\ &\quad - \varepsilon^2 2\omega(\varepsilon\partial_T A\sin\omega t + \varepsilon\partial_T B\cos\omega t) - \omega^2 x(t,T). \end{aligned} \tag{A24}$$

Putting (A22)–(A24) together in (A20), using $\eta\cos\omega = \omega^2 - 1$ and $\beta_c = -\eta\sin\omega/\omega$, and ignoring terms of order more than ε^3 we get that

$$\begin{aligned} &-2\omega\varepsilon^3(\partial_T A\sin\omega t + \partial_T B\cos\omega t) \\ &\quad - \varepsilon^3\eta\{\Delta A(T)\cos[\omega(t-1)] - \Delta B(T)\sin[\omega(t-1)]\} \\ &\quad - \varepsilon^3\beta_c(\partial_T A\cos\omega t - \partial_T B\sin\omega t) \\ &= \sqrt{2D}\varepsilon[A(T)\cos\omega t - B(T)\sin\omega t]\xi(t), \end{aligned} \tag{A25}$$

where $\Delta A(T)$ means $\frac{A(T)-A(T-\varepsilon^2)}{\varepsilon^2}$, etc. The corresponding equation that [10] arrives at[23] is

$$-\omega\varepsilon^3(\partial_T A\sin\omega t+\partial_T B\cos\omega t)$$
$$=\sqrt{2D}\varepsilon[A(T)\cos\omega t-B(T)\sin\omega t]\xi(t), \quad \text{(A26)}$$

Equation (A26) does not match with (A25) when ΔA and ΔB are set to zero or when they are set as actual derivatives $\partial_T A$ and $\partial_T B$. Reference [10] proceeds with (A26), multiplies by $\sin\omega t$, and averages over a time period to arrive at

$$\begin{aligned}-\omega\varepsilon^3\tfrac{1}{2}\partial_T A &= \sqrt{2D}\varepsilon\{A(T)[\![\cos\omega t\sin\omega t\xi(t)]\!]\\&\quad -B(T)[\![\sin^2\omega t\xi(t)]\!]\}\\&=\sqrt{2D}\varepsilon\tfrac{1}{2}\{A(T)[\![\sin 2\omega t\xi(t)]\!]-B(T)[\![\xi(t)]\!]\\&\quad +B(T)[\![\cos 2\omega t\xi(t)]\!]\},\end{aligned} \quad \text{(A27)}$$

where $[\![\ ]\!]$ is used for time averaging.

The intermediate steps in [10] are not clear, but the end result of [10] is that D is scaled as $D=\varepsilon^2\tilde{D}$ and three new Gaussian processes ξ_0,ξ_1,ξ_2 are defined on a slow time scale and the following are used:

$$[\![\xi(t)]\!]=\varepsilon\xi_0,\quad [\![\cos 2\omega t\xi(t)]\!]=\frac{\varepsilon}{\sqrt{2}}\xi_1,$$
$$[\![\sin 2\omega t\xi(t)]\!]=\frac{\varepsilon}{\sqrt{2}}\xi_2. \quad \text{(A28)}$$

Employing this in (A27), the following is arrived at:

$$-\frac{\omega}{\sqrt{2\tilde{D}}}\partial_T A=-B\xi_0+\frac{1}{\sqrt{2}}B\xi_1+\frac{1}{\sqrt{2}}A\xi_2. \quad \text{(A29)}$$

Similarly, [10] multiplies (A26) by $\cos\omega t$, averages over a time period, and employs (A28) to arrive at

$$-\frac{\omega}{\sqrt{2\tilde{D}}}\partial_T B=A\xi_0+\frac{1}{\sqrt{2}}A\xi_1-\frac{1}{\sqrt{2}}B\xi_2. \quad \text{(A30)}$$

Equations (A29) and (A30) are, respectively, (16) and (17) in [10].

Now we show that the above method is not consistent with itself. From (A29) and (A30) we get

$$-\frac{\omega}{\sqrt{2\tilde{D}}}(\partial_T A\sin\omega t+\partial_T B\cos\omega t)$$
$$=(-B\mathfrak{s}+A\mathfrak{c})\xi_0+\frac{1}{\sqrt{2}}(B\mathfrak{s}+A\mathfrak{c})\xi_1+\frac{1}{\sqrt{2}}(A\mathfrak{s}-B\mathfrak{c})\xi_2 \quad \text{(A31)}$$
$$=:\mathfrak{F}(T), \quad \text{(A32)}$$

where $\mathfrak{s}=\sin\omega t$ and $\mathfrak{c}=\cos\omega t$. Now $\mathbb{E}[\mathfrak{F}(T)\mathfrak{F}(T)]$ equals

$$(-B\mathfrak{s}+A\mathfrak{c})^2+\tfrac{1}{2}(B\mathfrak{s}+A\mathfrak{c})^2+\tfrac{1}{2}(A\mathfrak{s}-B\mathfrak{c})^2$$
$$=(A\mathfrak{c}-B\mathfrak{s})^2+\tfrac{1}{2}(A^2+B^2). \quad \text{(A33)}$$

However, from (A26),

$$-\frac{\omega}{\sqrt{2\tilde{D}}}(\partial_T A\sin\omega t+\partial_T B\cos\omega t)$$
$$=\varepsilon(A\mathfrak{c}-B\mathfrak{s})\xi(t)=:\varepsilon\mathfrak{F}(T). \quad \text{(A34)}$$

Now $\mathbb{E}[\mathfrak{F}(T)\mathfrak{F}(T)]$ equals $(A\mathfrak{c}-B\mathfrak{s})^2$. So the system (A29) and (A30) has an extra variance of $\frac{1}{2}(A^2+B^2)$ [see (A33)] than what is required.

APPENDIX B: EXAMPLE ILLUSTRATING THE APPROACH FOR CALCULATION OF $b_H^{q,(i)}$ IN THEOREM 2

Consider the system without delay given by $\ddot{x}+x=\varepsilon\dot{x}y$ and $\dot{y}=-y+\varepsilon\dot{x}^2$. Here x is oscillatory and y is stable. The quantity $\mathcal{H}=\frac{1}{2}(x^2+\dot{x}^2)$ evolves slowly compared to x and y. Writing in state-space form $z_1=x$ and $z_2=\dot{x}$, we have

$$\begin{pmatrix}\dot{z}_1\\ \dot{z}_2\\ \dot{y}\end{pmatrix}=\begin{pmatrix}z_2\\ -z_1\\ -y\end{pmatrix}+\varepsilon\begin{pmatrix}0\\ z_2 y\\ z_2^2\end{pmatrix} \quad \text{(B1)}$$

and $\dot{\mathcal{H}}=\varepsilon b^{(q)}(z,y)$, where $b^{(q)}(z,y)=z_2^2 y$.

The unperturbed system is obtained by setting $\varepsilon=0$ in (B1). The differential of any function f along a trajectory of the unperturbed system is given by $\mathcal{L}_0 f$, where $\mathcal{L}_0=z_2\frac{\partial}{\partial z_1}-z_1\frac{\partial}{\partial z_2}-y\frac{\partial}{\partial y}$. The differential along the perturbations is given by $\mathcal{L}_1 f$, where $\mathcal{L}_1=z_2 y\frac{\partial}{\partial z_2}+z_2^2\frac{\partial}{\partial y}$. Note that $\dot{f}(z_t,y_t)=[(\mathcal{L}_0+\varepsilon\mathcal{L}_1)f](z_t,y_t)$.

Now let

$$H(z,y)=\mathcal{H}(z)-\varepsilon c(z,y)+\varepsilon^2 g_1(z,y)+\varepsilon^2 g_2(z), \quad \text{(B2)}$$

where c,g are yet to be determined. On differentiating we get (until order ε^2)

$$\begin{aligned}\dot{H}(z_t,y_t)&=\varepsilon[b^{(q)}(z_t,y_t)-(\mathcal{L}_0c)(z_t,y_t)]-\varepsilon^2(\mathcal{L}_1c)(z_t,y_t)\\&\quad+\varepsilon^2(\mathcal{L}_0g_1)(z_t,y_t)+\varepsilon^2(\mathcal{L}_0g_2)(z_t,y_t)+O(\varepsilon^3).\end{aligned} \quad \text{(B3)}$$

Now choose c such that $\mathcal{L}_0c=b^{(q)}$. Choose g_1 such that $(\mathcal{L}_0g_1)(z,y)=(\mathcal{L}_1c)(z,y)-(\mathcal{L}_1c)(z,0)$. Such a choice of g_1 is possible because, according to the unperturbed dynamics, y decays to zero exponentially fast. Note that $(\mathcal{L}_1c)(z,0)$ is a function of z alone and the unperturbed z dynamics is an oscillation with constant amplitude $\sqrt{2\mathcal{H}}$. Now let the average of $(\mathcal{L}_1c)(z,0)$ along an orbit of constant $\mathcal{H}$ be denoted by $\{\mathcal{L}_1c\}$. This $\{\mathcal{L}_1c\}$ would be a function only of $\frac{1}{2}(z_1^2+z_2^2)$ or, equivalently, $\mathcal{H}$. Choose $g_2(z)$ such that $(\mathcal{L}_0g_2)(z,0)=(\mathcal{L}_1c)(z,0)-\{\mathcal{L}_1c\}|_{\frac{1}{2}(z_1^2+z_2^2)}$. Plugging the above choices of functions in (B3) we get

$$\dot{H}(z_t,y_t)=-\varepsilon^2\{\mathcal{L}_1c\}|_{\mathcal{H}}+O(\varepsilon^3). \quad \text{(B4)}$$

Hence, for times of $O(1/\varepsilon^2)$ we have $H(z_t,y_t)=H(z_0,y_0)+\varepsilon^2\int_0^t\{\mathcal{L}_1c\}|_{\mathcal{H}_s}ds+O(\varepsilon)$. Since H differs from $\mathcal{H}$ only by $O(\varepsilon)$ [see (B2)] we can write $\mathcal{H}_t=\mathcal{H}_0+\varepsilon^2\int_0^t\{\mathcal{L}_1c\}|_{\mathcal{H}_s}ds+O(\varepsilon)$. So, for times of order $O(1/\varepsilon^2)$, if we use

$$\dot{\mathcal{H}}=-\varepsilon^2\{\mathcal{L}_1c\}|_{\mathcal{H}}, \quad \text{(B5)}$$

[23] This is Eq. (9) in [10]. The quantity μ defined under Eq. (7) of [10] is zero for the special case that we consider.

then the error resulting in $\mathcal{H}$ would be only of $O(\varepsilon)$. Such a method is shown in [13]; we have adapted it to stochastic delay equations in [24].

To see why the above method is useful, note that c in $\mathcal{L}_0 c = b^{(q)}$ can be immediately solved using the method of characteristics. Since the solution to the unperturbed system is $z_1(t) = z_1(0)\cos t + z_2(0)\sin t$, $z_2(t) = -z_1(0)\sin t + z_2(0)\cos t$, $y(t) = y(0)e^{-t}$, and $b^{(q)}(z,y) = z_2^2 y$, we get $c(z,y) = -\int_0^\infty (-z_1\sin t + z_2\cos t)^2 y e^{-t} dt$. Now $(\mathcal{L}_1 c)(z,0) = -\int_0^\infty z_2^2(-z_1\sin t + z_2\cos t)^2 e^{-t} dt$. Hence $\{\mathcal{L}_1 c\}|_{\mathcal{H}}$ is

$$\frac{1}{2\pi}\int_0^{2\pi}\left(-\int_0^\infty z_2^2(-z_1\sin t + z_2\cos t)^2 e^{-t}dt\right)\bigg|_{(z_1,z_2)=\sqrt{2\mathcal{H}}(\sin s,\cos s)} ds = -\mathcal{H}^2\int_0^\infty \frac{1}{2}(2+\cos 2t)e^{-t}dt = -\frac{11}{10}\mathcal{H}^2.$$

So we have $\dot{\mathcal{H}} = \varepsilon^2\frac{11}{10}\mathcal{H}^2 + O(\varepsilon^3)$. The reader can check using conventional center-manifold calculations that the same answer would be obtained. However, the method presented here would easily adapt to multidimensional delay equations as shown in Sec. V.

APPENDIX C: EXPLICIT EVALUATION OF $b_H^{q,(k)}$ USING (51)–(54)

In this section we show how the explicit formulas (55) and (56) can be derived from (51)–(54). First we give a few preliminaries.

Recall that, for $\varphi \in \mathcal{C}$, $\hat{T}(t)\varphi$ denotes the solution at time t of the unperturbed linear system (5) with initial condition $\Pi_0 x = \varphi$. Recall that $\mathcal{C} = P \oplus Q$, where P is the space corresponding to the critical eigenvalues $\pm i\omega_c$. Recalling the evolution on P defined by (17), we have that for $\underline{u} \in \mathbb{C}^2$ with $u_2 = \bar{u_1}$,

$$\hat{T}(t)\Phi\underline{u} = \Phi e^{Bt}\underline{u}. \tag{C1}$$

Using (50) and (C1) we have for the $n\times 1$ vector $\underline{v}$,

$$\hat{T}(t)\pi\mathbf{1}_{\{0\}}\underline{v} = \Phi e^{Bt}\hat{\Psi}\underline{v}. \tag{C2}$$

For $\eta_t^\hbar$ defined in (25), we have $\hat{T}(s)\eta_t^\hbar = \eta_{t+s}^\hbar$. The z coordinates $\langle\Psi,\hat{T}(s)\eta_t^\hbar\rangle$ are given by $\frac{1}{2}\sqrt{2\hbar}\left[\begin{smallmatrix} e^{i\omega_c(t+s)} \\ e^{-i\omega_c(t+s)}\end{smallmatrix}\right]$ and hence for ρ defined in (51), we can take $\rho(\eta_t^\hbar) = \frac{2\pi}{\omega_c} - t$.

Using the product rule for differentiation on $b^{q,(1)}$ [defined in (47)] and linearity of the function E, we have, for $\xi,\eta \in \mathcal{C}$,

$$(\xi\cdot\nabla)b^{q,(1)}(\eta) = E(\xi)G_q(\pi\eta) + E(\eta)(\pi\xi\cdot\nabla)G_q(\pi\eta).$$

Using the product rule for differentiation on $b^{q,(2)}$ [defined in (48)] we have, for $\xi,\eta \in \mathcal{C}$,

$$(\xi\cdot\nabla)b^{q,(2)}(\eta) = ((\xi\cdot\nabla)E(\eta))(G_q(\eta) - G_q(\pi\eta)) + E(\eta)(\xi\cdot\nabla)G_q(\eta) - E(\eta)(\pi\xi\cdot\nabla)G_q(\pi\eta).$$

Since $\eta_t^\hbar$ [used in (54)] belongs to P, i.e., $\eta_t^\hbar = \pi\eta_t^\hbar$, the first term vanishes. Using linearity of differentials we have that

$$(\xi\cdot\nabla)b^{q,(2)}(\eta_t^\hbar) = E(\eta_t^\hbar)((I-\pi)\xi\cdot\nabla)G_q(\eta_t^\hbar) \quad \forall\xi \in \mathcal{C}. \tag{C3}$$

Now we show how (56) can be derived. Using (53) in (54) we encounter the task of evaluating the differential $(\xi\cdot\nabla)b^{q,(2)}(\hat{T}(s)\eta_t^\hbar)$ with $\xi = \hat{T}(s)\mathbf{1}_{\{0\}}G_q(\eta_t^\hbar)$. Using $\hat{T}(s)\eta_t^\hbar = \eta_{t+s}^\hbar$ and (C3) we get the differential as $E(\eta_{t+s}^\hbar)((I-\pi)\hat{T}(s)\mathbf{1}_{\{0\}}G_q(\eta_t^\hbar)\cdot\nabla)G_q(\eta_{t+s}^\hbar)$. It is a property of the unperturbed system that $\hat{T}$ commutes with $(I-\pi)$. Defining $\mathcal{E}_t = e^{-i\omega_c t}\hat{\Psi}_1 + e^{i\omega_c t}\hat{\Psi}_2$, we can write $E(\eta_t^\hbar) = \sqrt{2\hbar}\mathcal{E}_t$. So we can rewrite the differential as $\sqrt{2\hbar}(\hat{T}(s)(I-\pi)\mathbf{1}_{\{0\}}G_q(\eta_t^\hbar)\cdot\nabla)(\mathcal{E}_{t+s}G_q(\eta_{t+s}^\hbar))$. Writing $G_q(\eta_t^\hbar) = \sum_{j=1}^n (G_q(\eta_t^\hbar))_j\underline{e_j}$ and using linearity of differentials, we get the desired form in (56). Equation (55) can be similarly derived.

APPENDIX D: SKETCH OF PROOF OF THEOREM 3

One way to characterize the probability distribution of a stochastic process Y is by an operator called the infinitesimal generator $\mathcal{L}$ defined as follows: For any nice real-valued function f of the process Y,

$$(\mathcal{L}f)(y) \stackrel{\text{def}}{=} \lim_{t\to 0}\frac{1}{t}\{\mathbb{E}[f(Y_t)|Y_0 = y] - f(y)\}. \tag{D1}$$

Here the $\mathbb{E}$ term means the average of $f(Y_t)$ given that the initial condition Y_0 equals y. For example, the process whose infinitesimal generator is defined by $\mathcal{L}f = \frac{1}{2}f''$ has the same probability distribution as the standard Brownian motion. The process whose infinitesimal generator is defined by $(\mathcal{L}f)(y) = b(y)f'(y) + \frac{1}{2}\sigma^2(y)f''(y)$ has the same probability distribution as the process governed by the SDE $dY = b(Y)dt + \sigma(Y)dW$ with W a Wiener process. The process whose infinitesimal generator is $(\mathcal{L}f)(y) = b(y)f'(y)$ is the ordinary differential equation $\dot{Y} = b(Y)$. The infinitesimal generator characterizes the probability distribution of a process.

We consider the system (64) and (65) and try to find the infinitesimal generator $\mathcal{L}_H$ of the process $\lim_{\varepsilon\to 0}\mathcal{H}^\varepsilon$. For this purpose consider the triplet process $(\Pi^\varepsilon X,\xi^\varepsilon,\mathcal{H}^\varepsilon)$. It has the infinitesimal generator $\mathcal{L}^\varepsilon = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1$, where for function f of (η,ξ,h),

$$(\mathcal{L}_0 f)(\eta,\xi,h) = (\mathfrak{G}f)(\eta,\xi,h) + \frac{d}{dt}\bigg|_{t=0} f(\hat{T}(t)\eta,\xi,h), \tag{D2}$$

$$(\mathcal{L}_1 f)(\eta,\xi,h) = \sigma(\xi)(\mathbf{1}_{\{0\}}F(\eta)\cdot\nabla)f(\eta,\xi,h) + \sigma(\xi)b(\eta)\frac{\partial f}{\partial h}(\eta,\xi,h). \tag{D3}$$

Here $\mathfrak{G}$ is the infinitesimal generator of the noise process ξ. Recall that $\hat{T}(t)\eta$ is the solution at time t of the unperturbed system (5) with initial condition η and $\mathbf{1}_{\{0\}}$ is the matrix-valued function defined in (49).

The following comments help in gaining insight into the structure of $\mathcal{L}^\varepsilon$. Consider (64) and (65). If there were no noise perturbations at all, then $\mathcal{H}^\varepsilon$ would have remained a constant and $\Pi^\varepsilon X^\varepsilon$ would have evolved according to the unperturbed system whose solution at time t with initial condition η is given by $\hat{T}(t)\eta$. Applying the definition (D1) for this case, we get the $\frac{d}{dt}|_{t=0}$ term in (D2). If there were noise alone we would get the $\mathfrak{G}$ term in (D2). The rate of change of $\mathcal{H}^\varepsilon$ in (65) is σb, which explains the $\sigma b \frac{\partial f}{\partial h}$ term in (D3). The other term in (D3) is due to the perturbation coefficient σF in (64).

The problem of finding the infinitesimal generator $\mathcal{L}_H$ of the process $\lim_{\varepsilon\to 0}\mathcal{H}^\varepsilon$ can be simplified as follows (for details see the technique of the martingale problem in Chap. 5 of [35]): Find an operator $\mathcal{L}_H$ such that given any nice function f_H of h alone, there exists a function f^ε of (η,ξ,h) such that $|f_H(h) - f^\varepsilon(\eta,\xi,h)|$ and $|(\mathcal{L}_H f_H)(h) - (\mathcal{L}^\varepsilon f^\varepsilon)(\eta,\xi,h)|$ are of order ε.

Now we show how to find $\mathcal{L}_H$. Formally, consider $f^\varepsilon(\eta,\xi,h) \stackrel{\text{def}}{=} f_H(h) + \varepsilon f_1(\eta,\xi,h) + \varepsilon^2 f_2(\eta,\xi,h)$ with f_1 and f_2 yet to be determined. Computing $\mathcal{L}^\varepsilon f^\varepsilon$ we find

$$\mathcal{L}^\varepsilon f^\varepsilon = \frac{1}{\varepsilon^2}\mathcal{L}_0 f_H + \frac{1}{\varepsilon}(\mathcal{L}_0 f_1 + \mathcal{L}_1 f_0) + (\mathcal{L}_0 f_2 + \mathcal{L}_1 f_1) + O(\varepsilon). \tag{D4}$$

Note that $\mathcal{L}_0 f_H = 0$ because $\mathcal{L}_0$ involves differentials with respect to (η,ξ), whereas f_H is a constant as a function of (η,ξ) (it is function only of h). Now f_1 can be chosen so that $\mathcal{L}_0 f_1 + \mathcal{L}_1 f_0 = 0$. It can be verified that f_1 is

$$f_1(\eta,\xi,h) = \int_0^\infty ds \left(\int_{\mathbf{M}} [\nu(s,\xi,d\zeta) - \bar{\nu}(d\zeta)]\sigma(\zeta) \right) \times b(\hat{T}(s)\eta)\frac{\partial f_H(h)}{\partial h}.$$

We would not be able to select f_2 such that $\mathcal{L}_0 f_2 + \mathcal{L}_1 f_1 = 0$. However, $\mathcal{L}_0 f_2 + (\mathcal{L}_1 f_1 - \{\mathcal{L}_1 f_1\}) = 0$ can be solved, where $\{\mathcal{L}_1 f_1\}$ is certain kind of average. With this choice of f_2, now (D4) gives $|\mathcal{L}^\varepsilon f^\varepsilon - \{\mathcal{L}_1 f_1\}| \sim O(\varepsilon)$. Inspecting $\{\mathcal{L}_1 f_1\}$ gives $\mathcal{L}_H$. Note that $\mathcal{L}_1 f_1$ equals

$$\int_0^\infty ds \left(\sigma(\xi) \int_{\mathbf{M}} (\nu(s,\xi,d\zeta) - \bar{\nu}(d\zeta))\sigma(\zeta) \right) \times \left((\mathbf{1}_{\{0\}} F(\eta)\cdot\nabla) b^{(s)}(\eta) \frac{\partial f_H(h)}{\partial h} + b(\eta) b^{(s)}(\eta) \frac{\partial^2 f_H(h)}{\partial h^2} \right),$$

where $b^{(s)}(\eta) \stackrel{\text{def}}{=} b(\hat{T}(s)\eta)$. In the above expression (i) averaging the noise ξ with respect to its invariant measure $\bar{\nu}$ and recalling the definition of autocorrelation in (63), (ii) realizing that $(\mathbf{1}_{\{0\}} F(\eta)\cdot\nabla) b^{(s)}(\eta) = (\hat{T}(s)\mathbf{1}_{\{0\}} F(\eta)\cdot\nabla) b(\hat{T}(s)\eta)$, and (iii) averaging the η on trajectories of constant h, we get $\{\mathcal{L}_1 f_1\}$ as $b_H(h)\frac{\partial f_H(h)}{\partial h} + \frac{1}{2}\sigma_H^2(h)\frac{\partial^2 f_H(h)}{\partial h^2}$, where b_H and σ_H are as stated in Theorem 3.

APPENDIX E: NUMERICAL SCHEME FOR SIMULATIONS

All simulations in this paper are done with the Euler-Maruyama scheme. For example, (59) with $\gamma_c = 0$ is simulated as follows. Select a time step Δ. Let $N = r/\Delta$, where r is the delay in the system. Specify initial conditions at the time points of the form $j\Delta$ for $j = -N, -N+1, \ldots, -2, -1, 0$. Then, for $j \geqslant 0$,

$$x|_{(j+1)\Delta} = x|_{j\Delta} + \Delta\left(-\frac{\pi}{2}x + \varepsilon\gamma_q x^2 \right)\Big|_{(j-N)\Delta} + \varepsilon\sigma\sqrt{\Delta}\mathcal{N}_j,$$

where $\mathcal{N}_j$ is a standard normal random variable.

For (70) we first simulate the two-state Markov chain and then use

$$x|_{(j+1)\Delta} = x|_{j\Delta} + \Delta\left(-\frac{\pi}{2} + \varepsilon\sigma(\xi|_{j\Delta}) \right) x\Big|_{(j-N)\Delta}.$$

The following values of Δ are used: For Sec. VI B, $\Delta = 5\times 10^{-5}$; for Sec. V D, $\Delta = 2\times 10^{-5}$; for Sec. IV B, $\Delta = 10^{-5}$; and for the stationary density in Fig. 3, $\Delta = 5\times 10^{-6}$. Further evidence for the usefulness of averaging results of Sec. IV C is provided in the Supplemental Material [29].

[1] V. Kolmonovskii and A. Myshkis, *Applied Theory of Functional Differential Equations* (Springer, Berlin, 1992).
[2] L. Glass and M. Mackey, Scholarpedia **5**, 6908 (2010).
[3] G. Stepan, *Retarded Dynamical Systems: Stability and Characteristic Functions* (Longman Scientific & Technical, Harlow, 1989).
[4] G. Stepan, Philos. Trans. R. Soc. London A **359**, 739 (2001).
[5] D. Mei, Z. Yao, T. Kong, and Z. Chen, Int. J. Adv. Mfg. Technol. **46**, 1071 (2010).
[6] G. Quintana and J. Ciurana, Int. J. Mach. Tools Manu. **51**, 363 (2011).
[7] E. Buckwar, R. Kuske, B. L'Esperance, and T. Soo, Int. J. Bifurcat. Chaos **16**, 2407 (2006).
[8] A. Longtin, J. G. Milton, J. E. Bos, and M. C. Mackey, Phys. Rev. A **41**, 6992 (1990).
[9] W. Yao, P. Yu, and C. Essex, Phys. Rev. E **63**, 021902 (2001).
[10] M. Gaudreault, F. Drolet, and J. Vinals, Phys. Rev. E **85**, 056214 (2012).
[11] J. Hale and S. Verduyn Lunel, *Introduction to Functional Differential Equations* (Springer, Berlin, 1993).
[12] S. Mohammed, *Stochastic Functional Differential Equations* (Pitman, London, 1984).
[13] S. N. Chow and J. Mallet-Paret, J. Diff. Equ. **26**, 112 (1977).
[14] S. Das and A. Chatterjee, Nonlinear Dynam. **30**, 323 (2002).
[15] A. H. Nayfeh, Nonlinear Dynam. **51**, 483 (2008).
[16] M. Klosek and R. Kuske, Multiscale Model. Simul. **3**, 706 (2005).
[17] R. Kuske, Stoch. Dynam. **05**, 233 (2005).
[18] M. Gaudreault, F. Drolet, and J. Vinals, Phys. Rev. E **82**, 051124 (2010).
[19] M. S. Fofana, Mach. Sci. Technol. **6**, 129 (2002).
[20] M. Fofana, Probabilistic Eng. Mech. **17**, 385 (2002).
[21] C. Wang and J. Xu, Math. Problems Eng. **2010**, 829484 (2010).
[22] J. Lefebvre and A. Hutt, Europhys. Lett. **102**, 60003 (2013).
[23] J. Lefebvre, A. Hutt, V. G. LeBlanc, and A. Longtin, Chaos **22**, 043121 (2012).

[24] N. Lingala and N. S. Namachchivaya, Stoch. Dynam. **16**, 1650013 (2016).
[25] N. Lingala, arXiv:1510.06310.
[26] S.-E. A. Mohammed and M. K. Scheutzow, J. Funct. Anal. **206**, 253 (2004).
[27] O. Diekmann, S. van Gils, S. Verduyn Lunel, and H. Walther, *Delay Equations* (Springer, Berlin, 1995).
[28] S. E. A. Mohammed and M. K. R. Scheutzow, Ann. Probab. **25**, 1210 (1997).
[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.93.062104 for more evidence of the accuracy of the averaging results in Sec. IV C.
[30] G. C. Papanicolaou and W. Kohler, Commun. Math. Phys. **45**, 217 (1975).
[31] L. Katafygiotis and Y. Tsarkov, J. Appl. Math. Stoch. Anal. **12**, 1 (1999).
[32] N. Sri Namachchivaya and V. Wihstutz, Stoch. Dynam. **12**, 1150010 (2012).
[33] N. Sri Namachchivaya and V. Wihstutz (unpublished).
[34] N. S. Namachchivaya and L. Vedula, Dynam. Stability Syst. **15**, 185 (2000).
[35] S. Ethier and T. Kurtz, *Markov Processes. Characterization and Convergence* (Wiley, New York, 2005).