Spectral solution method for distributed delay stochastic differential equations

Alexandre René

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the M.Sc. degree in Physics

Ottawa-Carleton Institute of Physics
Department of Physics
Faculty of Science
University of Ottawa
Ottawa, Canada

Abstract

Stochastic delay differential equations naturally arise in models of complex natural phenomena, yet continue to resist efforts to find analytical solutions to them: general solutions are limited to linear systems with additive noise and a single delayed term. In this work we solve the case of distributed delays in linear systems with additive noise. Key to our solution is the development of a consistent interpretation for integrals over stochastic variables, obtained by means of a virtual discretization procedure. This procedure makes no assumption on the form of noise, and would likely be useful for a wider variety of cases than those we have considered. We show how it can be used to map the distributed delay equation to a known multivariate system, and obtain expressions for the system's time-dependent mean and autocovariance. These are in the form of series over the system's natural modes and completely define the solution.

An interpretation of the system as an amplitude process is explored. We show that for a wide range of realistic parameters, dynamics are dominated by only a few modes, implying that most of the observed behaviour of stochastic delayed equations is constrained to a low-dimensional subspace.

The expression for the autocovariance is given particular attention. A recurring problem for stochastic delay equations is the description of their temporal structure. We show that the series expression for the autocovariance does converge over a meaningful range of time lags, and therefore provides a means of describing this temporal structure.

Résumé

Alors que les modèles de phénomènes naturels complexes sont souvent exprimés en terme d'équations à délais stochastiques, ces dernières continuent de résister aux efforts de leur trouver des solutions analytiques. Les solutions générales sont limitées aux systèmes linéaires avec bruit additif et un seul terme de délai. Dans ce travail, nous résolvons le cas d'un système également linéaire avec bruit additif, mais où le délai est distribué. La clé de notre solution est le développement d'une interprétation consistante pour les intégrales de variables aléatoires, obtenue par le moyen d'une procédure de discrétization virtuelle. Cette procédure ne faisant aucune supposition par rapport à la forme du bruit, il est probable qu'elle puisse être appliquée à une plus grande variété de cas que ceux que nous avons considérés. Nous démontrons comment elle peut être utilisée pour associer une équation à délais distribués à un système multivarié connu, et ainsi obtenir des expressions pour la moyenne et l'autocovariance dépendantes du temps du système. Ces dernières sont sous la forme de séries où la somme porte sur les modes naturels du système; elles définissent complètement la solution.

Une interprétation du système en tant que processus d'amplitudes est explorée. Nous démontrons que pour un large étendu de paramètres réalistes, la dynamique est dominée par seulement quelques modes; cela suggère que la plupart du comportement observé pour les équations stochastiques à délais est limité à un sous-espace de faible dimension.

Une attention particulière est donnée à l'expression pour l'autocovariance. Un problème fréquent pour les équations stochastiques à délais est la description de leur structure temporelle. Nous démontrons que l'expression en série de l'autocovariance converge effectivement sur un interval suffisamment grand, et donc fournit un moyen de décrire cette structure temporelle.

Acknowledgements

I first would like to thank my advisor, Prof. André Longtin, for his essential advice, direction and support over the course of this project. I further thank Profs. Ivan L'heureux and Victor Leblanc for their help in ensuring accuracy throughout this text, as well as Prof. Rowan Thomson for her helpful comments.

Throughout this project I have relied on the experience and expertise of my colleagues of the neurophysics group at the University of Ottawa. It has been a pleasure to work with them.

I am also grateful to my family and friends for their moral support throughout this journey.

Finally, I thank the FRQNT for financially supporting this work.

Table of contents

List of Figures

List of Tables

Definition of symbols

Symbol	Definition
$\mathcal{S}$	state space for a delayed differential equation, i.e. $\mathcal{S} = C([-\tau_{max}, 0], \mathbb{R})$
x, y	function from $[0, \infty)$ to $\mathcal{S}$
$x(t)$, $y(t)$	state in $\mathcal{S}$; $t \in [0, \infty)$
τ, ν	continuous look-back variable for a functional state
x_τ	$x_\tau(t) = x_0(t - \tau)$
a, b, c, $x_0(t)$	scalar variable or parameter in $\mathbb{R}$ or $\mathbb{C}$
A, B, C, $X_0(t)$	random scalar variable or parameter of type $\mathbb{R}$ or $\mathbb{C}$
$\boldsymbol{X(t)}$, $\boldsymbol{Y(t)}$, $\boldsymbol{C}$	multivariate random variables of type $\mathbb{R}$ or $\mathbb{C}$
$X(t)$, $Y(t)$	functional random variables of type $\mathcal{S}$
$\langle A \rangle$, $\mathrm{E}[A]$	expectation value of A over the ensemble of noise realizations
$\mathrm{Cov}(A, B)$	covariance of A and B
$\Sigma_{ij}(t)$	$\mathrm{Cov}(X_i(t), X_j(t))$, where $\boldsymbol{X}(t)$ is a multivariate random variable provided be the context
$\Sigma_{\tau\nu}(t)$	$\mathrm{Cov}(X_\tau(t), X_\nu(t))$, where $X(t)$ is a functional random variable provided by the context
λ_n	eigenvalue associated to the index n
u_n	eigenfunction associated to the eigenvalue λ_n
$u_n^\dagger$	dual to the eigenfunction u_n, a.k.a. eigenfunctional
$u_{n,\tau}$	eigenfunction u_n, evaluated at the lag time τ
$\boldsymbol{u}_n$, $\boldsymbol{u}_n^\dagger$	right and left eigenvectors
$\boldsymbol{L}$	differential operator
$\boldsymbol{L}^{(N)}$	$N \times N$ matrix approximating $\boldsymbol{L}$
$\boldsymbol{u}_n^{(N)}$, $\boldsymbol{u}_n^{(N)\dagger}$	right and left eigenvectors of $\boldsymbol{L}^{(N)}$
$u_{n,i}^{(N)}$	i^{th} component of the eigenvector $\boldsymbol{u}_n^{(N)}$

Chapter 1

Introduction

As physicists, we are trained to think of the world in terms of differential equations. And given their versatility and power, it is not surprising that they have been applied, by scientists of all backgrounds, to model phenomena going well beyond traditional physics. However, there remain important distinctions between a physical theory and a mathematical model. The former is developed following a "bottom-up" approach: starting from the fundamental interaction of particles and fields, we build an understanding of an entire system. These fundamental interactions are local in both space and time: they only occur between immediate neighbours. And above the scale of quantum mechanics, they are also perfectly predictable.

For very complex systems, however, this ideal of a fully grounded theory is no longer effective. Weather models are a good example: while all the individual processes are purely physical phenomena that can be precisely described, their combined dynamic cannot (and never will be, due to its chaotic nature). That does not mean that as physicists, we should give up on describing or predicting phenomena of complex systems. But it *does* mean that we need to relax our strict bottom-up approach and make use of the heuristic and phenomenological reasoning power permitted by models. Applied judiciously, it can allow us to make approximations that beautifully isolate the dominant dynamics of a system. Ultimately the goal is always the same: to explain, and possibly predict, the phenomena we observe.

As always though, there is no free lunch: if our reasoning is less constrained, so too must be our tools. To reach that elusive goal of the eloquent mathematical model, it is often necessary to extend ordinary differential equations (ODEs) in such ways that violate some previously conceived truths. And as a consequence, this makes the resulting mathematics much more intricate.[1] The subject of this thesis will be the study of models that introduce two such violations: randomness and temporal nonlocality.

These violations appear when we remove elements of a model in order to simplify it. For instance, in a population model, rather than trying to pin down exactly when each individual dies, we might consider that the death rate is subject to random fluctuations, or *noise*. Differential equations that include the effect of noise are qualified as *stochastic*.

The main difficulty with a stochastic system is that due to the randomness, every time we simulate it we get a different result. Thus a "solution" is in fact a probability density,

[1] But also more interesting !

obtained only after thousands of simulations. Thankfully, in most cases,[2] we can convert the stochastic differential equation (SDE) into a *Fokker-Planck equation*, which, while more difficult, allows us to obtain the probability density directly – this is why the Fokker-Planck equation is one of the cornerstones in the study of stochastic differential equations.

In the aforementioned population model, we would also expect the best predictor for the number of births today to be the size of the population at some time in the past – which leads us to formulate a *delayed* differential equation (DDE) to reproduce the effect of the gestation period. Unfortunately, it turns out that delayed equations are incompatible with the Fokker-Planck formalism – due to their inherently infinite dimensional nature – which makes the study of delayed stochastic differential equations (SDDEs) particularly hard. So hard in fact, that despite serious attempts over at least the last two decades [1–4], no general solution to even the simplest linear SDDE has been published. Some progress has been made in adapting the Fokker-Planck formalism to delay equations equations [2, 3]. With the exemption of approximation results published in the last few months [5], the focus has remained on special cases.[3]

In a number of situations we can avoid the difficulty of dealing with SDDEs by neglecting either the noise or the delay, and then modelling the system as an SDE or a DDE. In many other situations however it is simply not justifiable to treat the system as anything less than a full SDDE.

Our work of course builds upon that of others. Of particular note are the results of Amann et. al. [4], who studied an SDDE of the form (the notation for stochastic and delay differential equations will be explained in chapter 2)

$$dX_0(t) = \Big(aX(t) + bX_0(t \underbrace{-\tau_1}_{\text{delay}}) \Big) dt \underbrace{+ q\, dW}_{\substack{\text{additive Gaussian} \\ \text{white noise}}};$$

they solve this equation by decomposing solutions into eigenfunctions. While by itself this idea is not new [6], they are able to find the linear forms associated to each eigenfunction, making decomposition much more direct than with the traditional approach. However, these linear forms are not derived but posited as an ansatz.

Our principal result will be the development of a procedure to systematically derive this dual set, borrowing some ideas from Hutt and Lefebvre. [7], most notably the use of constructs appearing in the study of functional differential equations (FDEs). Moreover, we are able to derive this dual set for DDEs with distributed delays of the form[4]

$$dX_0(t) = \left(\int_0^{\tau_{max}} \kappa(\tau) X_0(t-\tau)\, d\tau \right) dt \; + \; q\, dW\,. \tag{1.1}$$

Real systems are often better described by distributed delays. For example, a single fixed delay may not accurately reproduce the effect of gestation periods in a population model,

[2] In particular, this applies to any SDE with Gaussian noise.
[3] Specifically, efforts have focused on stationary solutions.
[4] The first form is recovered by setting $\kappa(\tau) = a\delta(\tau) + b\delta(\tau - \tau_1)$.

considering the wide variety these can have. In contrast, eq. (1.1) essentially considers that every possible delay between 0 and τ_{max} might influence the dynamics, each with its own weight $\kappa(\tau)$. Such expressions can thus reproduce the distributed delays often found in real-word phenomena. We will assume that both X and κ are real valued, as that is most usually the case for models of real phenomena. Equation (1.1), under these assumptions, forms the core object of study of this thesis.

The primary motivating examples that lead us to eq. (1.1) were models of neuronal networks. Such networks are subject to very large amounts of noise, as well as important transmission delays between neurons, and thus fall squarely in the category of models for which a full SDDE description is necessary. Due to the large number of connections between neurons, each with its own propagation speed, delays also tend to be distributed. That being said, the interest in solving eq. (1.1) is not limited to neuronal models, as such SDDEs are found in a much wider variety contexts [8–12]. Consequently, while this work was performed with neuronal models in mind, they will be little more than mentioned throughout the text.

The organization of this thesis is as follows. Chapter 2 reviews the notions and results forming the basis of later chapters. This is done in two parts, with delayed and stochastic equations covered separately. Chapter 3 then briefly summarizes the various recent attempts to tackle equations that are simultaneously delayed and stochastic. This sets the context for our own solution to the linear SDDE, which is presented in chapter 4. Chapter 5 finally concludes this thesis with a look at possible future research directions.

Some complementary material is included in appendices. Appendices A and B collect technical proofs that would otherwise break the text's flow, while appendix C shares some insights on appropriate numerical approaches. Implementation details for the presented numerical calculations are also included in the latter.

The work presented here falls within the larger project of developing a practical solution method for delayed Fokker-Planck equations. Such a solution would not be limited to linear equations such as eq. (1.1), but be applicable to a wide range of nonlinear systems.[5] A discussion of how such a solution might be built upon the tools and ideas developed in this thesis is also included in chapter 5.

[5] This is of particular interest to us, since while linear neuronal models do exist [7], most are strongly nonlinear.

Chapter 2

Building the foundation

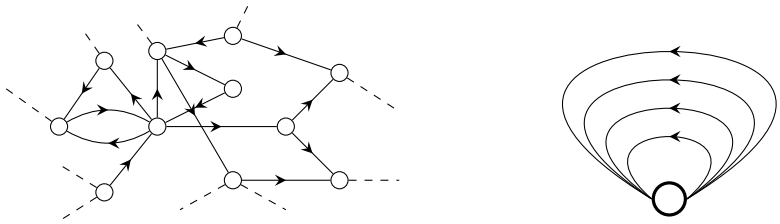
The construction of SDDEs relies heavily on the constructs of delay and stochastic differential equations (DDEs and SDEs): to a large extent, SDDEs are obtained by simply joining them all into one equation. Making sense of the result, of course, presents its own host of challenges. In this chapter we will look at DDEs and SDEs separately, and worry about how to put them together in chapter 3. Since stochastic equations and delay equations are seldom taught as part of an undergraduate degree in physics, we've made this chapter as self-standing as possible, assuming little prior knowledge beyond ordinary differential equations.

2.1 Delay differential equations

2.1.1 Introduction

Delayed dynamical systems go hand in hand with delayed feedback systems, and consequently are in fact quite common. A number of examples where they occur are detailed in chapter 1 of [13], among them: population dynamics, car following models, optical ring cavities, economics and mechanical control systems. However, since the motivation of this work stems from neural network models, we will illustrate the concepts in that context.

Brains are extremely complex systems, and have been the focus of a wide range of models. At the smallest scale are attempts to explain the complicated structures on a



(a) **(b)**

Figure 2.1 – **(a)** A possible schematic diagram of a neuronal network. **(b)** A toy model whose dynamics might be described by Eq. (1.1). The restriction to a single node makes the equation scalar rather than vectorial.

neuron's membrane that enable it to function. At the largest scales, neurons, or even entire populations of them, are reduced to single nodes. Neurons interact by sending and receiving spikes; thus we can assign to each node a time-dependent spike rate, or *activity*. This is a node's state, which we denote by $x(t)$. The spikes propagate along neural pathways, which, in a diagram such as fig. 2.1a, are represented by edges. Since the propagation speed is finite, a mathematical representation of such a network might be:[1]

$$\begin{aligned}\frac{dx_0^{(1)}(t)}{dt} &= f_1(x_0^{(1)}(t), x_0^{(2)}(t-\tau_{12}), \dots,) \\ \frac{dx_0^{(2)}(t)}{dt} &= f_2(x_0^{(1)}(t-\tau_{21}), x_0^{(2)}(t), \dots,) \\ &\vdots\end{aligned}$$

The reason for the subscript on x_0 will become clear in section 2.1.2.

Now for pedagogical reasons and to lighten the notation, we will limit our considerations to scalar systems. However, we still want to allow for the possibility of multiple delays; these might be conceived as in fig. 2.1b, where the effect of the relaying neurons along the multiple feedback pathways has been absorbed into the function f. Thus we can drop the indices and our dynamical equation reduces to

$$\frac{dx_0(t)}{dt} = f(x_0(t), x_0(t-\tau_1), x_0(t-\tau_2), \dots)\,, \tag{2.1}$$

where f is an arbitrary function. As mentioned in the introduction (chapter 1), we want to focus our attention on linear DDEs. Consequently, we can rewrite this as

$$\frac{dx_0(t)}{dt} = a_0 x_0(t) + a_1 x_0(t-\tau_1) + a_2 x_0(t-\tau_2), \dots\,,$$

where $a_0, a_1, \dots$ are constant parameters. Finally, to accommodate continuously distributed delays, we collect the parameters into a kernel κ and write the DDE in the same form as (1.1):[2]

$$\frac{dx_0(t)}{dt} = \int_0^{\tau_{max}} \kappa(\tau) x_0(t-\tau)\, d\tau\,. \tag{2.2}$$

Since the initial time is arbitrary anyway, we will always assume that systems start at $t = 0$.

Before diving into the details of delayed equations, let's make a few quick general remarks regarding DDEs.

[1] We neglect here the effect of noise, which is the subject of section 2.3.

[2] The distinction between lower and upper case variables, which distinguishes between normal and random variables, will be made in section 2.3.

Distinguishing delay differential, differential-difference and integro-differential equations. One finds that much of the basic results on delay differential equations were developed in a more general context, that of *differential-difference* equations.[3] These encompass all differential equations involving a finite number of delays, with *delayed* equations[4] such as eq. (2.1) corresponding to those where delays appear only in the non-differentiated terms. Among differential-difference equations, DDEs are the most tractable, are the ones most commonly found in models and have received the most research attention.

Equation (2.2), however, is an *integro-differential*, rather than a delay-difference, equation. In the current literature the former is often synonymous with *distributed-delay* equations, a term we will prefer as it better expresses the close similarity with delay equations. In fact, in this work we consider discrete-delay equations such as eq. (2.1) to be simply particular cases of distributed delay equations, and only rarely distinguish the two. Thus eq. (2.2) is referred to as a DDE (and eq. (1.1) as an SDDE), with the distributed nature of the delay implicit.

A solution to a DDE is almost as simple to generate as for an ODE. By "generating a solution", we refer to numerically solving a DDE for a given initial condition. An equation such as eq. (2.1) is an unambiguous update rule, and the same techniques as for ODEs can be applied with minimal additional overhead. Some extra care does have to be taken when evaluating the delayed terms $x(t-\tau_i)$, but these are technical considerations. We cover them briefly in appendix C.2.

Linear delayed equations have richer behaviour than linear ODEs. A nontrivial scalar linear ODE is a very constrained object: it 1) has exactly one fixed point at $x=0$ and 2) will either increase or decrease exponentially for all t (or stay at the fixed point). In particular, it is impossible for it to oscillate.

The first property is a consequence of its linearity and still applies to an equation such as (2.2). The second property, however, no longer applies, since it is a consequence of a scalar ODE having only one dimension. In particular, even for scalar DDEs oscillations are the norm rather than the exception; an example is given in fig. 2.2.

The initial condition of a DDE is a function. Looking at the form of eq. (2.2), it is clear that in order to evaluate the right-hand side at $t=0$ we need to know $x(t)$ over the interval $[-\tau_{max}, 0]$. So an initial condition for a DDE is in fact a function, which is usually denoted ϕ. Thus a well-posed DDE might be written:

$$\begin{aligned} \frac{dx_0(t)}{dt} &= \int_0^{\tau_{max}} \kappa(\tau) x_0(t-\tau)\, d\tau && t \geq 0 \\ x_0(t) &= \phi(t) && -\tau_{max} \leq t \leq 0\,. \end{aligned}$$

[3] Many of these results are presented in the classic text by Bellman and Cooke [6].
[4] Also referred to as *retarded* equations.

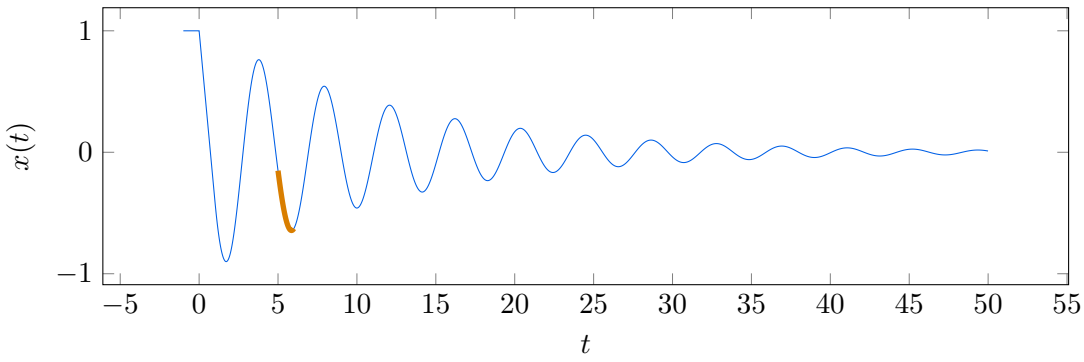


Figure 2.2 – A solution to the system $dx_0(t)/dt = -1.4\,x_0(t-1)$, where we set $\tau_{max} = 1$. The initial condition is $\phi(t) = 1, -1 \leq t \leq 0$, as plotted. In order to continue the propagation from any point t, we need the value of $x_0(t')$ for $t' \in [t-1, t]$; thus the proper description of a state $x(t)$ is a function over that interval. The state $x(6)$ is indicated in orange.

2.1.2 Describing the state of a DDE

The immediate corollary of the previous remark is that the proper way to express the state of a delayed system at some time t is also by a *continuous*[5] *function* from $[-\tau_{max}, 0]$ to $\mathbb{R}$, rather than by a scalar as is the case for ODEs. An example is given in fig. 2.2, where the state $x(6)$ is highlighted.

For a DDE given by eq. (2.2), the set of all possible states corresponds to the set of continuous functions from $[-\tau_{max}, 0]$ to $\mathbb{R}$. Hence, if we define $\mathcal{S}$ to be the state space of the given DDE, then

$$\mathcal{S} = C([-\tau_{max}, 0], \mathbb{R})\,. \tag{2.3}$$

Since $\mathcal{S}$ is a space of functions, the state space of a delay differential equation is necessarily infinite dimensional.

Remark 2.1

An important idea will be that we can discretize a state function[6] $x(t)$ as a vector $\boldsymbol{x}^{(N)}(t)$ – the use of a bold character indicates that this is a vector quantity, while the superscript $^{(N)}$ indicates that it is the discretization of a continuous one. To obtain $\boldsymbol{x}^{(N)}(t)$, we first choose a discretization level $N \in \mathbb{N}$. Then we discretize $x(t)$ as an $N+1$ dimensional vector, associating to each component $x_k^{(N)}(t)$ the value of $x_0(t - \frac{k}{N}\tau_{max})$, as illustrated in fig. 2.3.

We note that this mapping implies

$$x_{n-1}^{(N)}\left(t - \tfrac{\tau_{max}}{N}\right) = x_0^{(N)}\left(t - \tfrac{n\,\tau_{max}}{N}\right) = x_n^{(N)}(t)\,,$$

[5] In the context of neural models, it is usually safe to assume that solutions are continuous, even if they may not be differentiable. At the very least they should be piecewise continuous.

[6] We employ "state" and "state function" interchangeably, using the latter when we want to stress the functional nature of a state.

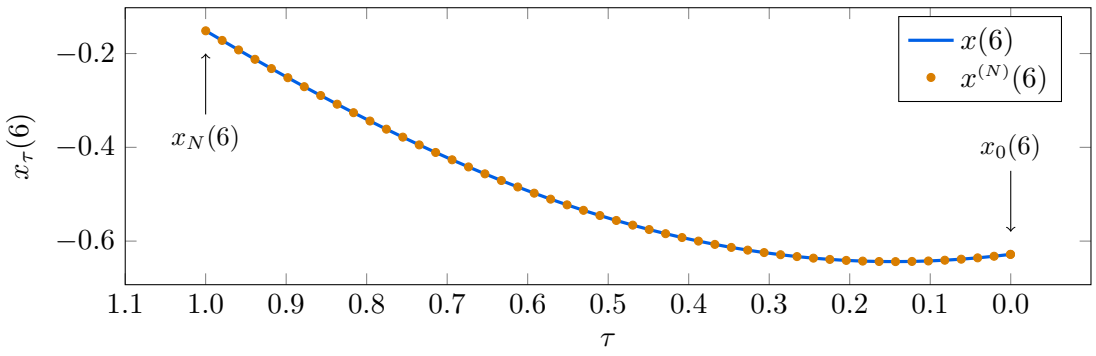


Figure 2.3 – Example of a discretized state function with $N = 50$, corresponding to $x(6)$ of fig. 2.2. The solid line corresponds to the state $x(6)$ while the dots indicate the value of each component $x_k^{(N)}(6)$, $0 \leq k \leq N$. We note the inversion of the abscissa: since $x_\tau(t) = x_0(t-\tau)$, the lag variable τ in fact takes positive values between 0 and τ_{max} (we have $\tau_{max} = 1$ in this case).

and therefore that we have the recursion relation

$$x_n^{(N)}(t) = x_{n-1}^{(N)}\left(t - \tfrac{\tau_{max}}{N}\right) \tag{2.4}$$

for $1 \leq n \leq N$.

In an expression such as $x_0(t-\tau)$, the *look-back* variable[7] τ plays essentially the same role for the state function $x(t)$ as the index k does for the state vector $\boldsymbol{x}^{(N)}(t)$. Because of this, it makes sense to define

$$x_\tau(t) \equiv x_0(t-\tau)\,. \tag{2.5}$$

This allows us to associate values of a discretized state $x^{(N)}(t)$ to those of a continuous state $x(t)$, as illustrated in fig. 2.3. Explicitly, this association is be expressed as

$$x_\tau(t) = x_k^{(N)}(t)\,,$$

where

$$k = \frac{\tau}{\tau_{max}} N\,. \tag{2.6}$$

We can now rewrite eq. (2.2) using the notation that will be standard throughout this document:

$$\dot{x}_0(t) = \int_0^{\tau_{max}} \kappa(\tau) x_\tau(t) d\tau\,, \tag{2.7}$$

where we've also used the conventional dot to indicate derivative with respect to time. One advantage of this notation is that the dependence on t can be made implicit, as is often done for ODEs.

In order to prevent confusion, we maintain strict adherence to our index notation throughout this text, such that $x(t)$ always refers to a member of the function space $\mathcal{S}$ and $x_0(t)$ to the value of the solution at time t.

[7] We will alternatively use the term *time lag* as a synonym for "look-back variable".

We want to stress the fact that $x(t)$ is *itself* a function. Indeed, while x is a function of the time t,

$$x\colon \begin{array}{ccc} [0,\infty) & \to & \mathcal{S} \\ t & \mapsto & x(t) \end{array},$$

$x(t)$ is a function of the look-back variable τ,

$$x(t)\colon \begin{array}{ccc} [0,\tau_{max}] & \to & \mathbb{R} \\ \tau & \mapsto & x_\tau(t) = x_0(t-\tau) \end{array}.$$

Furthermore, one notices that while the time variable t runs forward, the look-back variable τ runs *backwards*: $x_\tau(t)$ *precedes* $x_\nu(t)$, when $\tau > \nu$.

2.1.3 Standard solution methods for DDEs

2.1.3.1 The method of steps

Suppose we have a DDE

$$\dot{x}_0(t) = f(t, x_0(t), x_{\tau_1}(t))\,, \qquad x(0) = \phi\,,$$

with a single delay τ_1 and an initial function ϕ. Then over the interval $[0, \tau_1]$, it is an ODE,

$$\dot{x}_0(t) = f(t, x_0(t), \phi(t-\tau_1))\,,$$

where the initial condition can be considered a time dependent parameter of f. This can be solved using standard ODE methods. The solution over $[0, \tau_1]$ is then used as an initial condition for $t \in [\tau_1, 2\tau_1]$. This process can be repeated ad infinitum to generate the entire solution.

This way of computing a solution is called the *method of steps*, which is detailed e.g. in [14]. It's extremely convenient for basic proofs such as the existence and uniqueness of solutions, since it allows one to leverage well-established results for ODEs. It also matches how DDEs are solved numerically, although algorithms do not really need to artificially partition time into intervals of length τ_1.

On the other hand, the analytical solutions obtained by the method of steps are clunky, and become unwieldy very quickly as t increases. Thus the study of long term or general behaviour using this approach, while possible in at least some occasions (see e.g. [1]), is difficult. The method is also inapplicable to distributed delay systems where the kernel $\kappa(\tau)$ is non-zero over some interval $\tau \in [0, \epsilon]$, $\epsilon > 0$.

2.1.3.2 Applying the Laplace transform

For a suitably non-pathological function[8] f, the Laplace transform $\mathcal{L}(f)$ is defined as

$$\mathcal{L}(f)(s) = \int_0^\infty e^{-st} f(t)\, dt\,.$$

[8] Basically, it must not grow faster than e^{Mt} for some M.

This operation maps a function to what is sometimes termed "*Laplace space*", where the s are complex frequencies of sorts. This is a common solution technique for ODEs, and turns out to be particularly useful for DDEs because the integral over t is a natural way of dealing with the dependence on history.

Applying the Laplace transform to a differential equation transforms it into an *algebraic* equation, which is much easier to solve – see [15, chap. 5] or [6, chap. 1]. We can then apply the inverse transform

$$\mathcal{L}^{-1}(F)(t) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F(s)e^{st}\,ds\,, \tag{2.8}$$

where $c \in \mathbb{R}$ must be chosen sufficiently large, to recover the solution as a function of t.

While the solution method we present in chapter 4 does not make use of this technique, it employs the same idea of applying a transform to a frequency space, and then the inverse transform to recover the solution.

2.1.3.3 Decomposing into eigenfunctions

The idea of constructing the solution to a differential equation from eigenfunctions – whose dynamics are typically simpler – should be familiar to most physicists. This works equally well with delay equations, and has traditionally been achieved by a combination of Laplace transforms and careful use of the residue theorem (see for example [6]). However Amann et al. [4] recently demonstrated an approach that is perhaps more amenable to generalization. Since we perform such a generalization in chapter 4, we present their approach here in some detail.[9]

To demonstrate the idea, let's consider the homogeneous equation

$$\dot{x}_0(t) = ax_0(t) + bx_{\tau_1}(t)\,, \qquad x(0) = \phi\,. \tag{2.9}$$

Since the state of a DDE is a function rather than a scalar (see section 2.1.2), we want to rewrite this as a functional equation in the form

$$\dot{x}(t) = \boldsymbol{L}x(t)\,, \tag{2.10}$$

where the dot indicates derivative with respect to time and $\boldsymbol{L}$ is an endomorphism of $\mathcal{S}$.

We first need to determine the form of the functional differential operator $\boldsymbol{L}$. Given eq. (2.5), it is clear that we must have

$$\frac{dx_\tau(t)}{dt} = -\frac{dx_\tau(t)}{d\tau}\,,$$

which, combined with eq. (2.9), gives us

$$[\boldsymbol{L}x(t)]_\tau = \begin{cases} ax_0(t) + bx_{\tau_1}(t) & \tau = 0 \\ -dx_\tau(t)/d\tau & 0 < \tau \le \tau_1\,. \end{cases} \tag{2.11}$$

[9] The presentation is somewhat modified from [4] in order to maintain better coherence with subsequent chapters.

A delay differential operator in this form appears in [7, eq. 10].

Next we need the form of the eigenfunctions u_n, which should verify

$$\boldsymbol{L}u_n = \lambda_n u_n$$

for some $\lambda_n \in \mathbb{C}$. Comparing with eq. (2.11), we see that

$$u_{n,\tau} = e^{-\lambda_n \tau}$$

would satisfy this condition for $\tau \in (0, \tau_1]$. This leaves only $u_{n,0}$ unspecified. But we expect modes to be continuous in τ, which will only be true if $u_{n,0} = 1$. Thus we are lead to define the eigenfunctions as

$$u_{n,\tau} \equiv e^{-\lambda_n \tau}\,. \tag{2.12}$$

> The second index following an eigenfunction indicates evaluation, consistent with the notation in section 2.1.2. We'll maintain the convention throughout this thesis of indicating eigenfunction indices with latin letters, and look-back times with greek letters.

Assuming the eigenfunctions form a basis of the state space, we can write

$$x(t) = \sum_n c_n(t) u_n\,;$$

where the coefficients c_n must satisfy the initial condition

$$\phi = \sum_n c_n(0) u_n\,, \tag{2.13}$$

as well as the differential equations

$$\dot{c}_n(t) = \lambda_n c_n(t)\,. \tag{2.14}$$

(It's important to keep in mind when reading these equations that $x(t)$ and u_n are not simply scalars, but functions from $[0, \tau_1]$ to $\mathbb{R}$.) By virtue of the linearity of $\boldsymbol{L}$, the general solution can then be written

$$\begin{aligned} x(t) &= \sum_n c_n(t) u_n \\ &= \sum_n c_n(0) e^{\lambda_n t} u_n\,. \end{aligned} \tag{2.15}$$

In this form, it becomes clear that the single fixed point at $x = 0$ is stable if and only if the real part of every eigenvalue is negative, as we would expect.

Of course, in obtaining eq. (2.15) we've glossed over two key points:

- how we might compute the eigenvalues λ_n,
- how we determine the coefficients $c_n(0)$,

which we will now address.

Remark 2.2 Validity of the eigenfunction decomposition

One might wonder whether the set of eigenfunctions $\{u_n\}$ even forms a *complete basis*[10] of $\mathcal{S}$; in fact, in most cases they likely don't – in fig. 4.3 (page 51) for instance, we are unable to reproduce the constant function as a sum over eigenfunctions. That being said, Bellman and Cooke *do* show that as long as the initial condition ϕ is differentiable over $[-\tau_{max}, 0]$, a solution in the form of eq. (2.15) exists for $t > 0$ [6, thm 4.2].[11]

In fact, the behaviour observed in fig. 4.3 is analogous to the Gibb's phenomenon, a well-known effect for Fourier decompositions. For an initial condition that is only continuous, their proof is somewhat weaker, ensuring the existence of the series solution only for $t > \tau_{max}$ [6, thm 4.1c].

We interpret this to mean that the span of eigenfunctions u_n is *dense*[12] in $\mathcal{S}$: even if an initial state $\phi \in \mathcal{S}$ cannot be decomposed as in eq. (2.13), there should be a function $\tilde{\phi} \in \mathrm{Span}(\{u_n\})$ so similar that both ϕ and $\tilde{\phi}$ produce exactly the same dvnamics for $t > 0$.

Strictly speaking, Bellman and Cooke show their result for linear equations with a single discrete delay and constant coefficients. However, given that all linear equations show a similar distribution of eigenvalues (section 4.4.1), we expect it to remain true for distributed delays. (We needn't worry about nonlinear and equations with time-dependent coefficients since we excluded them from this study.)

Computing the eigenvalues

The characteristic equation is obtained simply by substituting $x(t) = c_n(t)u_n$ into the first row of eq. (2.11), yielding

$$a + b\, e^{-\lambda_n \tau_1} = \lambda_n \,. \tag{2.16}$$

Rearranging terms and multiplying by $\tau_1 e^{-a\tau_1}$, we get

$$(\lambda_n - a)\, \tau_1\, e^{(\lambda_n - a)\tau_1} = b\, \tau_1\, e^{-a\tau_1} \,.$$

Comparing this with the definition of the Lambert W function (eq. (2.27)),

$$W_n(\alpha) e^{W_n(\alpha)} = \alpha \,,$$

we see that the two equations are equivalent when we make the substitutions $\alpha = b\, \tau_1\, e^{-a\tau_1}$ and $W_n(\alpha) = (\lambda_n - a)\tau_1$. Combining these two equalities yields an expression for the

[10] The basis $\{u_n\}$ is *complete* if every element of $\mathcal{S}$ can be decomposed as a linear combination of $\{u_n\}$.

[11] If some of the eigenvalues have multiplicity greater than one, the form changes slightly.

[12] For a set U to be *dense* in another set V essentially means that there are elements of U arbitrarily close to any element of V. Here we use the term somewhat loosely since we don't formally check that the property is verified. Moreover, in order to define "closeness", we would in turn be required to endow $\mathcal{S}$ with a norm – most likely one would want to use either the L^1 or L^2 norms.

eigenvalues:

$$W_n(b\tau_1 e^{-a\tau_1}) = (\lambda_n - a)\,\tau_1$$
$$\lambda_n = a + \frac{W_n(b\,\tau_1\,e^{-a\tau_1})}{\tau_1}\,. \tag{2.17}$$

The Lambert function has multiple branches, which we indicate by the index n, with each branch yielding a different eigenvalue. As a special function, it has already been characterized by a good body of research (see esp. [16]) and is implemented in most computational platforms. Thus, while it must still be evaluated numerically, it's nonetheless much more convenient to obtain the eigenvalues this way than by solving the characteristic equation (2.16) directly. Some properties of the Lambert function that we will need in later chapters are presented in section 2.2.

Of course, for more complex equations such as those with distributed delays, a solution to the characteristic equations (2.16) in terms of special functions such as W might not be known. In those cases one has to resort to solving the characteristic equation numerically. We present thoughts on this problem in appendix C.1.

Determining the coefficients $c_n(0)$

There are at least three known approaches for obtaining the decomposition coefficients $c_n(0)$. One is to deform the integration contour of eq. (2.8) such as to close it and then apply the residue theorem. Although somewhat involved, it is entirely based on well-established results of complex analysis; details can be found in [6].

Another approach is to tackle the problem entirely numerically. One constructs a matrix where each column corresponds to a discretized eigenfunction. Inverting this matrix and applying it to the initial condition then yields a good estimate of the coefficients $c_n(0)$. This is the approach presented in [17], where it is also extended to vector delay differential equations. A similar method, combined with transforms to Laplace space, is used in [18] and [19].

Finally, the approach Amann et al. [4] take is to define the dual set $\{u_n^\dagger\}$ to the eigenfunctions.[13]

Dual spaces, linear functionals, biorthogonal sets

> A dual space $\mathcal{S}^*$ consists of linear functionals mapping functions in $\mathcal{S}$ to scalars in $\mathbb{R}$. In other words, given a function $y \in \mathcal{S}$ and a linear functional $\omega \in \mathcal{S}^*$, applying ω to y yields a real number $\omega(y)$. And since ω is linear, its action can be defined by an integral over the domain of functions in $\mathcal{S}$.[14] (It's sometimes enlightening to think in terms of finite dimensional spaces, in which case elements of the space and its dual can be respectively expressed as column and row vectors.)

[13] While one might argue that u_n^* would be more consistent with standard notations for dual bases, for vectors we prefer to use † in order to avoid confusion with complex conjugation.

[14] Assuming of course that $\mathcal{S}$ admits the definition of an integral. In our case $\mathcal{S} = C([0, \tau_{max}])$, so that's a safe assumption.

The *dual set* $\{u_n^\dagger\}$ to the basis $\{u_n\}$ is the set of functionals satisfying

$$u_n^\dagger(u_m) = \delta_{n,m}\,, \tag{2.18}$$

where $\delta_{n,m}$ is the Kronecker delta. Together, a basis and its dual form a *biorthogonal set* [20, § 6.2].[15] Since we assume that the basis is for all practical purposes complete (remark 2.2), generally we also assume that the elements of the biorthogonal set fulfill the *completeness relation* [20, eq. 6.28]

$$\sum_n u_n u_n^\dagger(x) = x, \quad \forall x \in \mathcal{S}\,, \tag{2.19}$$

unless we have reason to suspect otherwise. We refer to the linear functionals $u_n^\dagger$, complements to the eigenfunctions u_n, as *eigenfunctionals* in this thesis.

We claim[16] that if the characteristic equation is given by eq. (2.16), then the duals $u_n^\dagger$ to the eigenfunctions, which we've termed eigenfunctionals, are given by[17]

$$\begin{array}{rccc} u_n^\dagger\colon & \mathcal{S} & \to & \mathbb{R} \\ & y & \mapsto & \mathcal{N}_n \cdot \left(y_0 + b\int_0^{\tau_1} d\tau\, e^{\lambda_n(\tau-\tau_1)}\ y_\tau\right), \end{array} \tag{2.20}$$

where

$$\mathcal{N}_n = \left(1 + b\tau_1 e^{-\lambda_n\tau_1}\right)^{-1} = (1 - a\tau_1 + \lambda_n\tau_1)^{-1} = \left(1 + W_n(b\tau_1 e^{-a\tau_1})\right)^{-1} \tag{2.21}$$

is a normalization constant. Indeed, substituting (2.12) for y, one can readily verify that $u_n^\dagger(u_n) = 1$, while for $n \neq m$,

$$\begin{aligned} u_n^\dagger(u_m) &= \mathcal{N}_n \left[1 + b\, e^{-\lambda_n\tau_1} \int_0^{\tau_1} e^{\lambda_n\tau} e^{-\lambda_m\tau} d\tau\right] \\ &= \mathcal{N}_n \left[1 + b\, e^{-\lambda_n\tau_1} \frac{e^{\lambda_n\tau_1}\, e^{-\lambda_m\tau_1} - 1}{\lambda_n - \lambda_m}\right] \\ &= \mathcal{N}_n \left[1 + \frac{b\, e^{-\lambda_m\tau_1} - b\, e^{-\lambda_n\tau_1}}{\lambda_n - \lambda_m}\right] \\ &= \mathcal{N}_n \left[1 + \frac{\lambda_m - \lambda_n}{\lambda_n - \lambda_m}\right] \\ &= 0\,, \end{aligned}$$

[15] Strictly speaking, a basis is complete *by definition*, but since we won't worry too much about this anyway, we'll use "basis" to mean a set that we *think* spans $\mathcal{S}$ (or $\mathcal{S}^*$).

[16] Amann et al. similarly justify this claim after the fact. We will give a systemic way of obtaining the $u_n^\dagger$ in chapter 4.

[17] Since Amann et al. define their dual basis functions $u_n^\dagger$ as members of C rather than the dual space C^*, they are forced to define a non-standard delta function $\delta(\tau + 0)$ in order to obtain the left term in the parenthesis of eq. (2.20). Without this there is a risk of ambiguity for functions that jump at $\tau = 0$.

where we've used the fact that the characteristic equation (2.16) implies

$$be^{-\lambda_m \tau_1} - be^{-\lambda_n \tau_1} = \lambda_m - \lambda_n \,.$$

Once we have the dual basis $\{u_n^\dagger\}$, computing the coefficients c_n is as simple as applying the $u_n^\dagger$ to the initial condition:

$$\begin{aligned} u_n^\dagger(\phi) &= u_n^\dagger \left(\sum_m c_m(0)\, u_m \right) = \sum_m c_m(0)\, u_n^\dagger(u_m) \\ &= c_n(0)\,. \end{aligned}$$

Assuming the expression is convergent, we can define the system's propagator (a mapping from $\mathcal{S}$ to $\mathcal{S}$) in terms of the $u_n^\dagger$:

$$G(t,t')(\cdot) = \sum_n e^{\lambda_n (t-t')}\, u_n\, u_n^\dagger(\cdot)\,. \tag{2.22}$$

(Here the dot represents an arbitrary element of $\mathcal{S}$.) Thus the general scalar solution can be written,

$$\begin{aligned} x_0(t) &= \sum_n e^{\lambda_n t} c_n(0) \\ &= \sum_n e^{\lambda_n t} u_n^\dagger(\phi) \\ &= \sum_n e^{\lambda_n t} \cdot \mathcal{N}_n \cdot \left(\phi_0 + b \int_0^{\tau_1} d\tau\, e^{\lambda_n (\tau - \tau_1)}\ \phi_\tau \right) . \end{aligned} \tag{2.23}$$

To get the functional solution, we multiply each term by u_n,

$$\begin{aligned} x(t) &= \sum_n e^{\lambda_n t} u_n u_n^\dagger(\phi) \\ &= G(t,0)\,\phi\,. \end{aligned} \tag{2.24}$$

Remark 2.3

We note that while eq. (2.23) is exactly equivalent to the expression obtained by Amann et al. [4, eq. 14], they derive a *scalar*, rather than functional, propagator. Their form is

$$G_{scal}(t,t') = \begin{cases} \sum_n \mathcal{N}_n\, e^{\lambda_n (t-t')} & t > t' \\ 0 & t < t'\,. \end{cases} \tag{2.25}$$

This has a couple of consequences. The most obvious one is that eq. (2.24) must now be written in scalar form [4, eq. (14)]

$$x_0(t) = G_{scal}(t,0)\phi_0 + \int_0^{\tau_{max}} G_{scal}(t, \tau_{max} - \tau)\, b\, \phi_\tau\, d\tau \tag{2.26}$$

which is certainly less compact than eq. (2.24).[18] The more serious consequence

however is that $G_{scal}(t,t')$ remains undefined for $t = t'$: at that point, the series becomes

$$\sum_n \frac{1}{1 + W_n(b\,\tau_1\, e^{-a\tau_1})}\,,$$

which is only *conditionally* convergent.[19] Consequently, following Riemann's summation theorem, the series' value depends on the order in which we take the sum. Since $G(t,t)$ *should* be equal to 1, we can conceive using this constraint to establish a "sum rule". However, this process is far from trivial and unlikely to be generalizable.

In contrast, eq. (2.22) unambiguously tends towards the unity map as t approaches t', since in that case we recover the completeness relation (2.19). To our knowledge, this is the first instance of such an expression in the context of DDEs.

2.2 The Lambert W function

The Lambert W function is defined by the property

$$W_n(\alpha)e^{W_n(\alpha)} = \alpha\,. \tag{2.27}$$

It is defined over the entire complex plane, however as long as we consider DDEs with real state variables and coefficients, the equations it appears in will always have α real. Thus for the rest of this section we implicitly assume $\alpha \in \mathbb{R}$.

The canonical reference for the Lambert function is the comprehensive study by Corless et al. [16]. For a discussion more focused on the distribution of the roots of characteristic equations such as eq. (2.16) – which is what we are interested in here – one can refer to chapters 4 and 12 of [6].

It is easy to verify that if $W_n(\alpha)$ satisfies eq. (2.27), then its complex conjugate $W_n(\alpha)^*$ must do the same. Thus complex solutions to eq. (2.27) come in conjugate pairs, as evidenced by fig. 2.4. For this reason, it is really quite convenient to label the branches of W such that $W_{-n}(\alpha) = W_n(\alpha)^*$ if $W_n(\alpha)$ is complex. To achieve this, we index them from $-\infty$ to ∞, with smaller values of $|n|$ corresponding to smaller values of $|\mathrm{Re}\, W_n(\alpha)|$. Crucially, *we skip zero when the number of real solutions is even.*[20] We find that the

[18] One might argue that eq. (2.23) simply conceals the integral in eq. (2.24). However, note that only in eq. (2.26) is the integral time-dependent. In eq. (2.23) we've essentially rewritten G we only have one linear operation (the application of G) rather than two nested ones (the *integral* of the application of G) to perform. It also simplifies intuitive interpretation, since more often then not we think of ϕ in terms of the mode amplitudes $c_n(0)$.

[19] A series is conditionally convergent if $\sum_{n=1}^{\infty} |a_n|$ diverges while $\sum_{n=1}^{\infty} a_n$ converges. In this case, using the fact that for large n, $W_n(\alpha) \sim -\log|n| + i2\pi n$ (see eq. (2.28) below), we get $\sum_{n=-\infty}^{\infty} \left|\frac{1}{1+W_n(b\tau_1 e^{-a\tau_1})}\right| \sim \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n} = \infty$, while $\sum_{n=-\infty}^{\infty} \frac{1}{1+W_n(b\tau_1 e^{-a\tau_1})} \sim \sum_{n=1}^{\infty} \frac{1}{W_n(b\tau_1 e^{-a\tau_1})} + \frac{1}{W_{-n}(b\tau_1 e^{-a\tau_1})} \sim -\frac{1}{2\pi^2}\sum_{n=1}^{\infty}\frac{\log n}{n^2} < -\frac{1}{2\pi^2}\zeta\left(\frac{3}{2}\right) < \infty$.

[20] We make this skipping of 0 implicit, so that we can give a ranges such as $-10 \le n \le 10$.

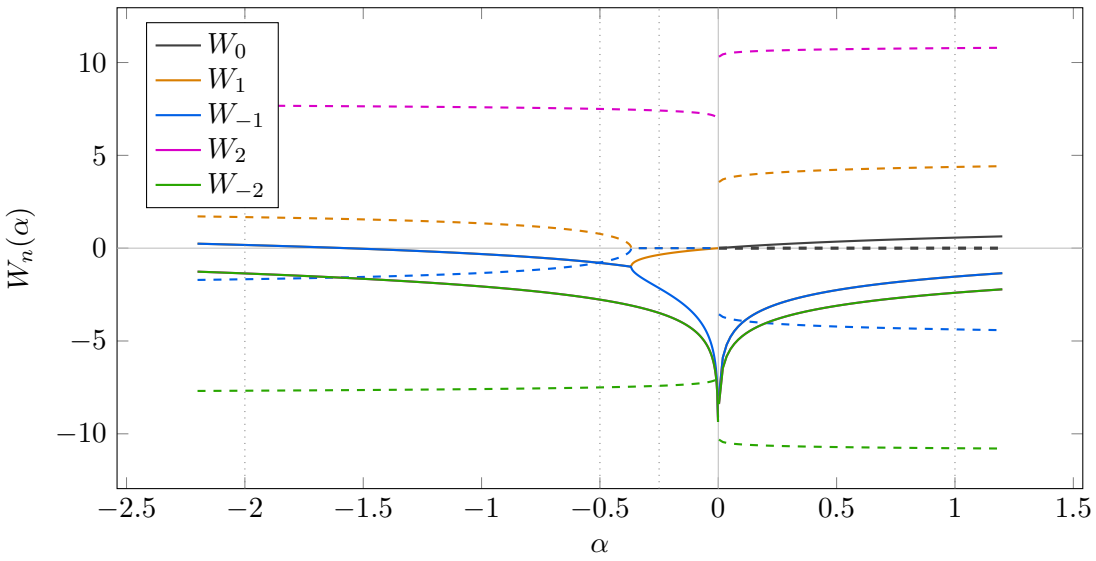


Figure 2.4 – The first few branches of the Lambert function. Solid lines correspond to $\mathrm{Re}(W_n(\alpha))$; dashed lines to $\mathrm{Im}(W_n(\alpha))$. The W_0 branch is defined only when there are an odd number of real eigenvalues, which is the case when $\alpha > 0$. Complex branches come in conjugate pairs, so their real parts superimpose. The vertical dotted lines correspond to the values of α appearing in fig. 2.5.

symmetry our expressions gain in this way is well worth the trouble of introducing a non-standard indexing scheme.

Depending on the value of α, eq. (2.27) may have anywhere between zero and two real solutions. Only when the number of real solutions is exactly one do we define $W_0(\alpha)$. In the case of two real solutions, these are identified as $W_1(\alpha)$ and $W_{-1}(\alpha)$ – in this case the symmetry of solutions is broken for $n = 1$.

While the Lambert function must be computed numerically, for intuition and asymptotic analysis it's quite useful to have an approximate analytical form of $W_n(\alpha)$ for $n \gg 1$ and $\alpha \in \mathbb{R} \setminus 0$ fixed. From examining fig. 2.5, we might try the ansatz

$$W_n(\alpha) = -\log c(\alpha)(|n| - 1) + i(2\pi(n - \mathrm{sgn}(n)) + \theta(\alpha)) \qquad |n| > 1$$

where the constants $c(\alpha) \in \mathbb{R}^+$ and $\theta(\alpha) \in [0, 2\pi]$ are constants to be determined. (The sign function $\mathrm{sgn}(\alpha)$ is -1 if $\alpha < 0$, 0 if $\alpha = 0$ and 1 if $\alpha > 0$.) Substituting this expression into eq. (2.27) and taking the limit as n goes to infinity, we get

$$i2\pi \cos\theta(\alpha) - 2\pi \sin\theta(\alpha) = \alpha c\,.$$

The fact that c and α are real requires $\theta = \pm\frac{\pi}{2}$. And since c must be positive, we have in fact $\theta = -\mathrm{sgn}(\alpha)\frac{\pi}{2}$ and $c = \left|\frac{2\pi}{\alpha}\right|$. Hence we obtain an asymptotic expression for the

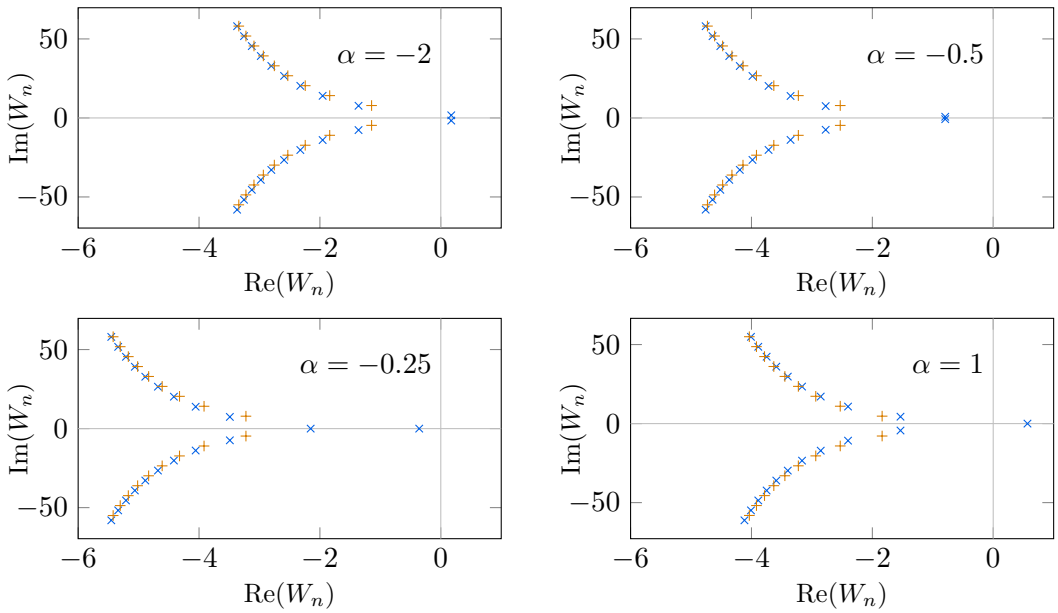


Figure 2.5 – Value of $W_n(\alpha)$ for branches $-10 \leq n \leq 10$; these plots can be seen as "vertical cuts" of fig. 2.4. True values and approximate ones (given by eq. (2.28)) are indicated by 'x' and '+' respectively. For $n = \pm 1$ eq. (2.28) is undefined, so only the exact value is given.

Lambert function,

$$W_n(\alpha) \approx -\log \frac{2\pi(|n| - 1)}{|\alpha|} + i\left(2\pi(n - \mathrm{sgn}(n)) - \mathrm{sgn}(\alpha)\frac{\pi}{2}\right). \tag{2.28}$$

This expression is compared to the exact function in fig. 2.5; it shows surprisingly good agreement even for small n (except for $n = 1$, for which it is undefined). While more general asymptotic expressions exist, for instance [16, eq. 4.20], they consider $\alpha \in \mathbb{C}$ – we are not aware of a similar derivation or expression exploiting a restriction to $\alpha \in \mathbb{R}$.

2.3 Stochastic differential equations

At this point we switch gears and delve into the wonderful[21] world of stochastic differential equations (SDEs). But before we go into the details, it's a good idea to have in mind why a model might have randomness (or noise) in the first place.

Perhaps the most obvious source of random fluctuations between a deterministic model and an experimental measurement is *measurement error*. This kind of fluctuation is "added on top": *after* the system has run its course, the measurement itself introduces variations.

[21] and/or terrifying

This is "noise" in the purest sense, but since it doesn't influence the evolution of the system, it is not an SDE.

In order to have a system described by an SDE, the source of noise needs to "kick" it in some way: predators capturing prey sporadically; wind gusts warping a flag; quantum fluctuations on an Ising lattice; unpredictable input from the 100 billion or so other neurons in a human brain; spontaneous opening and closing of ion channels on a cell membrane, etc. Sometimes the process underlying the noise is truly random, while other times we simply do not wish to (or can not) provide a detailed (e.g. microscopic) model for it. In all cases, the nature of the process matters only insofar as determining the noise's statistical properties (as well, of course, as how exactly it appears in the model).

At first glance, it might seem that generating a realization of an SDE isn't fundamentally more challenging than for an ODE: At each time step, pick a random number, stick that into the noise and compute the next step using the SDE. Unfortunately it is not quite that simple for two reasons. The first is that determining the statistical distribution from which to pick the random number is far from trivial. The second is that there are many mathematically valid ways of computing the next step once we've chosen a random number for the noise.

We will illustrate these ideas using the most classical example of an SDE: Brownian motion. Our presentation will broadly follow the much more complete treatment by Gillespie [21, chap. 8,9].

2.3.1 Brownian motion, the Langevin equation and the Wiener process

Brownian motion takes its name from Robert Brown, who observed the phenomenon in 1827. Although he never knew this, what he was observing were packets of starch and fat, which while tiny, are nonetheless much larger than the nanoscopic water molecules surrounding them [21, p. v]. This is the core assumption of Brownian diffusion:[22] that we have a relatively small number of *solute* particles in a bath of much smaller *solvent* particles. As we know, liquid water molecules are not stationary (otherwise they would form ice), and so as they move around they bump into the larger solute particles, kicking them to and fro. Now, there are *a lot* of water molecules, so it makes more sense to describe their statistics than the individual movement of each one.

The mathematical transcription of this idea is the *Langevin equation*. To derive it, we first invoke the superposition principle[23] in order to consider only the x coordinate of a solute particle. We will represent the kicks for the solvent molecules by a force $m \cdot g\xi(t)$, where $\xi(t)$ is a standardized random variable (which we define later), g is a constant which determines the strength of the kicks and m is the mass of the solute particle. This drives the random motion which we call *diffusion*. In addition to the kicks, we also expect there to be the usual force of drag for a particle in a fluid; at low speeds, this is proportional to

[22] Brownian *diffusion* is what Robert Brown observed. Brownian *motion* is more generic and includes behaviour described by the same mathematics but not necessarily relating to solutes in a solvent.

[23] Or in mathematical terms, the fact that Newtonian forces in a system combine linearly.

the particle's speed V_x. This force is deterministic, and drives the *drift*. Thus, applying "$F = ma$" we get

$$\dot{V}_x(t) = \tfrac{1}{m}(F_{\text{drift}} + F_{\text{diffusion}}) = f\, V_x(t) + g\, \xi(t)\,, \tag{2.29}$$

where f and g are constants representing the strengths of the drift and diffusion forces respectively.

By now it has likely become clear that $V_x(t)$, just like $\xi(t)$, is a *random variable*. Indeed, it makes little sense to ask about the value of V_x at time t: the series of kicks will be different every time we repeat the experiment, leading to a different time series, or *realization*. What we need to ask instead is: "What is the probability that $V_x(t)$ be between v_x and $v_x + dv_x$?" In other words, solutions to stochastic equations take the form of time-dependent probability densities.

Throughout this thesis, we will apply the usual convention that random variables be denoted by capital letters.[24] This helps distinguish them from their realizations, which are conventionally denoted by lowercase letters. For instance, we write

$$P_{X(t)}(X(t) < x)$$

for the probability that $X(t)$ be smaller than x. (For continuous variables of course, the probability that $X(t)$ *equals* x is zero.) If there is a function p such that

$$\int_{-\infty}^{x} p_{X(t)}(x)\, dx = P_{X(t)}(X(t) < x)\,,$$

then we call this the *probability density function* (p.d.f.). It is easy to see that this definition is equivalent to the derivative of $P_{X(t)}(X(t) < x)$, when the latter is differentiable. For continuous variables the probability density is often more convenient. In particular, the expectation value of a random variable across realizations is often written

$$\langle X(t) \rangle = \int_{-\infty}^{\infty} x\, p_{X(t)}(x)\, dx\,. \tag{2.30}$$

One shorthand we will allow ourselves is to write

$$p_{X(t),Y(t),\dots}(x, y, \dots) \equiv p_{X(t),Y(t),\dots}(X(t) = x, Y(t) = y, \dots)\,,$$

since here the order of the indices clearly indicates which random variable is compared to which value. For one particularly recurrent case (see page 26), we will even drop the indices altogether.

Now we need to determine the statistical properties of the noise $\xi(t)$. Since each kick is due to a different water molecule, and each molecule is independent, then so too should

[24] For whatever reason this convention isn't as well established in the physics as it is in the mathematical literature. It will however be rigorously applied in this text.

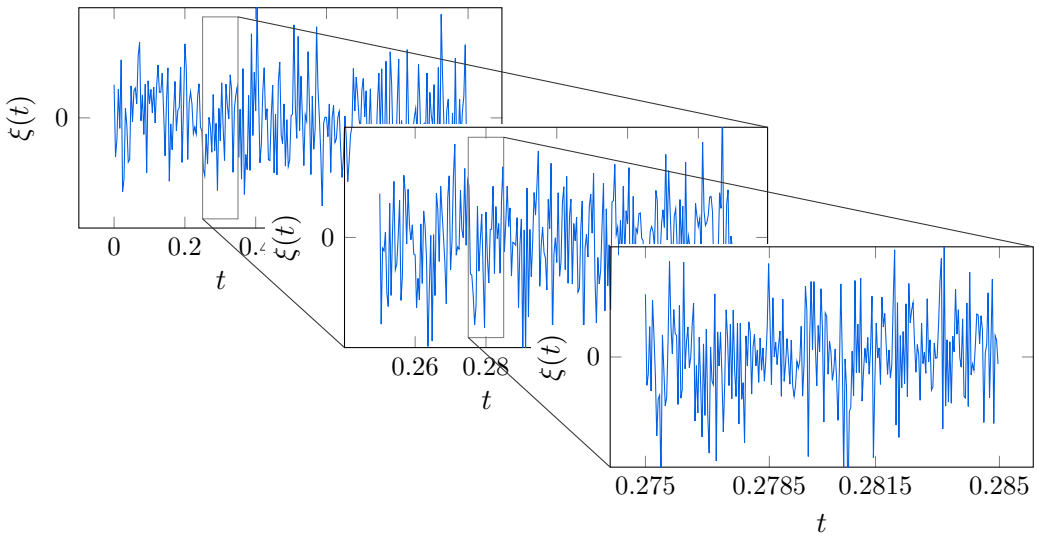


Figure 2.6 – Delta-correlated noise is equally irregular at all time-scales. In other words, it is never smooth, even in the tiniest of intervals, and is thus nowhere differentiable. (Note that the *value* of $\xi(t)$ is meaningless since it is infinite almost everywhere, as shown by eq. (2.35). This is why no values are given on the ordinate.)

the value of $\xi(t)$ be independent from the value of $\xi(t')$ for $t \neq t'$.[25] In addition, since the molecules have no preferable direction, the mean of $\xi(t)$ should be zero. We can express these constraints mathematically as

$$\begin{aligned} \langle \xi(t) \rangle &= 0 \\ \langle \xi(t)\, \xi(t') \rangle &= \delta(t - t')\,; \end{aligned} \tag{2.31}$$

where the brackets $\langle \cdot \rangle$ denote the expectation value across realizations. Such noise is often termed "zero mean, delta correlated", or again "white noise".[26] It is also possible to define a noise where $\langle \xi(t)\, \xi(t') \rangle \neq \delta(t - t')$, in which case it is termed "coloured".

The conditions (2.31) don't yet fully determine the statistics of ξ, but they already introduce a serious problem: they imply that ξ is (almost)-nowhere differentiable. This is easy to see when we view differentiability as the property of a function to become smoother as we reduce the time scale: No matter how close we may choose t and t', since $\xi(t)$ and $\xi(t')$ are independent they can never be correlated, and so the function never smooths out. In fact, a delta-correlated random function displays fractal-like self-similarity throughout time scales, as fig. 2.6 illustrates.

[25] This stops being true at very short time scale, for instance if $t - t'$ is less than the time of an individual collision.

[26] "White" refers to the fact that the frequency spectrum of $\xi(t)$, as would be obtained by Fourier transform, is flat. This is in analogy to white light, which also has a flat spectrum.

To deal with this, we rewrite eq. (2.29) as an "update formula" involving differentials rather than derivatives,

$$V_x(t+dt) = V_x(t) + f\,V_x(t)dt + g\,dW(t)\,, \tag{2.32}$$

where we've replaced $\xi(t)dt$ by the infinitesimal random variable $dW(t)$.[27] To be consistent, the update formula has to yield the same result when we "refine" the time scale:

$$\begin{aligned} V_x(t+dt) &= V_x(t)(1+f\,dt) + g\,dW_{dt}(t) \\ &= \Big[V_x(t)(1+f\,\tfrac{dt}{2}) + g\,dW_{dt/2}(t)\Big](1+f\,\tfrac{dt}{2}) + g\,dW_{dt/2}(t+\tfrac{dt}{2}) + \mathcal{O}(dt^{3/2})\,. \end{aligned} \tag{2.33}$$

In order to satisfy this self-consistency requirement, $dW_{dt}(t)$ *must* be a Gaussian random variable whose variance is proportional to dt [22] – and since we have already included a scaling factor g, we set this proportionality constant to 1. Hence we can define the *Wiener increment*[28] to be

$$dW(t) \sim \mathcal{N}(0, dt)\,. \tag{2.34}$$

(We've dropped the subscript dt since it almost always clear from the context.) This justifies the presence of $\mathcal{O}(dt^{3/2})$ rather than $\mathcal{O}(dt^2)$ in eq. (2.33) : it is the standard deviation of dW that has the same units as dV_X, and this scales as $dt^{1/2}$.

Formally substituting this expression into eq. (2.32), we get

$$\begin{aligned} \frac{V_x(t+dt) - V_x(t)}{dt} &= f\,V_x(t) + \frac{g}{dt}\,\mathcal{N}(0, dt) \\ \frac{dV_x(t)}{dt} &= f\,V_x(t) + g\,\lim_{dt\to 0}(dt)^{-\frac{1}{2}}\,\mathcal{N}(0,1)\,. \end{aligned}$$

Comparing with eq. (2.29), we see that $\xi(t)$ should satisfy

$$\xi(t) \equiv \lim_{dt\to 0}\mathcal{N}(0, 1/dt)\,. \tag{2.35}$$

The mathematical validity of this expression is of course questionable, but it is a useful conceptual tool. Assuming we could somehow rescale this function by its standard deviation,[29] it would look something like fig. 2.6.

In terms of Wiener increments, the white noise properties of eq. (2.31) become

$$\begin{aligned} \langle dW(t)\rangle &= 0 \\ \langle dW(t)\,dW(t')\rangle &= \begin{cases} dt & \text{if } t = t' \\ 0 & \text{otherwise}\,. \end{cases} \end{aligned} \tag{2.36}$$

[27] At least in terms of developing an intuition it is really best here to take the physicist's interpretation of a differential, namely to view it as the limit of a small but finite increment.

[28] Named for Norbert Wiener, who was one of the first to establish the robust mathematical bases for equations such as Langevin's.

[29] We can only do this formally of course, since the standard deviation is infinite.

General form of the Langevin equation

By substituting X for V_x, eq. (2.32) may be slightly reworked and generalized into the standard form

$$dX(t) = f(t, X(t))\, dt + g(t, X(t))\, dW(t)\,. \tag{2.37}$$

This is the most general form of the scalar Langevin equation, which is said to have *multiplicative noise* because the function g depends on X. When g is independent of X, as in

$$dX(t) = f(t, X(t))\, dt + g(t)\, dW(t)\,,$$

the equation is said to have *additive noise*.

2.3.2 Integrating a stochastic differential equation

Let's now turn to the problem of computing a realization of eq. (2.37). This gets a little bit tricky since $dW(t)$ is nowhere continuous: classical Riemann integration, for example, requires continuity over most of the interval (specifically, its set of points of discontinuity must have measure zero). But it is possible to extend the definition of the Riemann-Stieltjes integral to accommodate stochastic increments.

Riemann-Stieltjes integral

The Riemann-Stieltjes integral defines a value for integrals of the form

$$\int_a^b g(t)\, dh(t)\,,$$

where g is continuous and h is of *bounded variation*.[30] The idea is to choose t''s that split the interval $[a, b]$ into a partition : $a = t_N < t_{N-1} < \cdots < t_1 < t_0 = b$. (We maintain the convention that t_0 corresponds to "present" time.) Then the Riemann-Stieltjes integral is schematically defined as

$$\int_a^b g(t)dh(t) \equiv \lim_{N\to\infty} \sum_{n=N}^{1} g(t'_n)(h(t_{n-1}) - h(t_n))\,, \qquad t'_n \in [t_n, t_{n-1}]\,. \tag{2.38}$$

There are of course a number of sometimes grudging details involved in making sure this limit actually makes sense, but those are not our focus; they are the same as for the Riemann integral. One important fact is that the integral *does not depend on the choice of* t'_n. This is a little less surprising when we note that if the derivative of h' exists and is bounded, we can rewrite the integral as a Riemann integral,

$$\int_a^b g(t)dh(t) = \int_a^b g(t)h'(t)dt\,,$$

where there is no such choice to make.

[30] Basically, being of bounded variation means that h may have discontinuous jumps under the conditions that a) there be finitely many jumps between a and b and b) that every jump be between two finite values.

Applying eq. (2.38) to a stochastic equation, we get[31]

$$X(t_0) = \int_0^{t_0} g(t, X(t))dW(t) \tag{2.39}$$

$$= \lim_{N\to\infty} \sum_{n=N}^{1} g(t'_n, X(t'_n))(W(t_{n-1}) - W(t_n)) \,, \quad t'_n \in [t_n, t_{n-1}] \,. \tag{2.40}$$

The catch is that now W is no longer of bounded variation. This means that in general, the integral is no longer independent of the choice of t'_n.

When we have additive noise, g only depends on t'_n. Noting that $g(t'_n)$ is continuous in t'_n, and that the difference between t_n and t_{n-1} is infinitesimally small, we see that in this case the choice of t'_n remains inconsequential.

For multiplicative noise the choice of t'_n can strongly influence the integral, and to evaluate eq. (2.39) it is essential to establish an "integration rule"; the two most common are the Itô and Stratonovich conventions. Determining the appropriate rule is often not easy [24], and at least historically has been the topic of heated debates [25].

Given that for this particular study we focus exclusively on Langevin equations with additive noise, we will not distinguish between Itô or Stratonovich conventions.

Wiener process

The *Wiener process* is defined by the integral

$$W(t) = \int_0^t dW \,.$$

Since every dW is Gaussian, so too is $W(t)$; in fact we have [21, eq. 7.36]

$$W(t) = \mathcal{N}(0, t) \,. \tag{2.41}$$

A realization of this process is a continuous, nowhere-differentiable function of t; we show many such realizations in fig. 2.7a. As we accumulate realizations, we begin reconstructing the p.d.f. numerically: at each time point, the value of $W(t)$ is normally distributed (figs. 2.7b and 2.7c).

2.3.3 The Fokker-Planck equation

It is unfortunately extremely rare to be able to evaluate a closed-form solution such as eq. (2.41) for an SDE. In general, the only method to obtain a probability density function (p.d.f.) from a Langevin equation is to simulate it multiple times and then draw some statistics from the aggregate results. This kind of Monte Carlo approach is relatively

[31] The proper mathematical definition of the stochastic integral involves the *mean square limit* [23]. For building an intuitive picture the looser formalism we use here is fine – as long we keep in mind that it isn't completely bullet-proof.

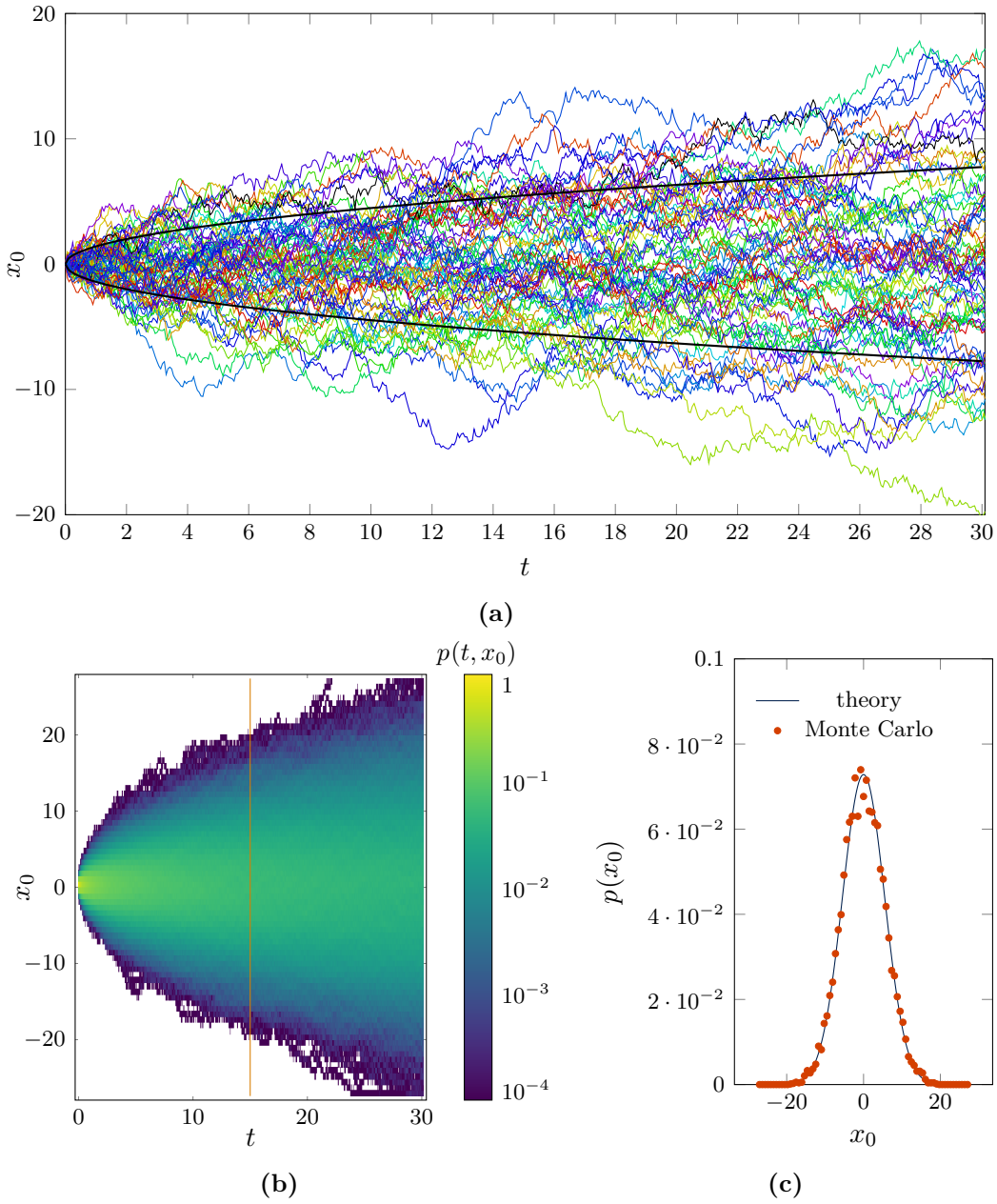


Figure 2.7 – **(a)** Overlay of 75 realizations of the Wiener process $dX(t) = \sqrt{2}\, dW(t)$. The standard deviation, given by $\pm\sqrt{2t}$, is drawn in black. **(b)** Approximate p.d.f. obtained from 10 000 realizations over the same time interval. The vertical orange bar indicates the time slice which is illustrated in (c). **(c)** Slice of the numerical p.d.f. at $t = 15$. The theoretical result given by eq. (2.41) is drawn for reference.

straightforward to do once we've established the form of noise and how we integrate it, but it's still annoying for two reasons:

- It takes a long time. Even for the Wiener process, about 10 000 runs are required to produce a reasonably smooth distribution such as the one shown in fig. 2.7, with running times measured in hours. A sweep of the initial conditions might therefore very well take days. And of course, the more peaks and trough in the p.d.f., the more runs we need to adequately sample it.
- It might get things wrong. The more complicated the p.d.f., the more careful we have to be in generating the random numbers. If for example the system has an unlikely but extremely stable state, properly sampling that state becomes difficult since it is rarely visited by the Monte Carlo process. As the complexity of the system grows, just sampling it with Monte Carlo becomes a research problem in itself.

One solution to these problems is to recast the systems dynamics completely, in the form of a partial differential equation (PDE) rather than an SDE:

$$\frac{\partial}{\partial t}p(t,x_0|x^{init}) = -\frac{\partial}{\partial x_0}\left[A(t,x_0)p(t,x_0|x^{init})\right] + \frac{1}{2}\frac{\partial^2}{\partial x_0^2}\left[D(t,x_0)p(t,x_0|x^{init})\right]. \tag{2.42}$$

Here $A(t,x_0)$ and $D(t,x_0)$ are functions related to the system's drift and diffusion, and they define what is known as the *Fokker-Planck equation* (see § 3 of [21] for a derivation. More extensive treatments can be found in [23] and [20]).

> Since we very frequently write expressions involving the p.d.f. of the current state $X(t)$ conditioned on the initial value $X(0)$, for this particular case we drop the indices and make the random variables implicit:
>
> $$p(t,x|x^{init}) \equiv p_{X(t)|X(0)}(x|x^{init}),$$
>
> a notation we use in eq. (2.42). If the initial condition is clear from the context, especially if it unnecessarily clutters the notation, we may drop the condition altogether:
>
> $$p(t,x) \equiv p(t,x|x^{init}).$$

Compared to solving the SDE directly, the Fokker-Planck equation is advantageous in a number of ways:

- The p.d.f. is computed directly, circumventing the need to generate thousands of realizations. The computation is also *deterministic*.
- Since the p.d.f. is smooth, there is no need to specify an integration rule: all SDEs describing the same process lead to the same Fokker-Planck equation.[32]
- By virtue of the p.d.f.'s smoothness, it is also much easier to evaluate the magnitude of errors introduced by numerical schemes – so the "unlikely but stable" states mentioned above can be dealt with simply by setting a sufficiently small mesh.

[32] In general, two equivalent SDEs using different integration rules will have different expressions for the drift and diffusion coefficients.

The Fokker-Planck approach also has its drawbacks of course: Partial differential equations are notoriously more difficult to solve than ODEs, and can quickly choke numerical solvers as dimensions are added.[33] But as these are in many cases easier to manage than the difficulties that arise when trying to solve SDEs directly, the Fokker-Planck formalism is often the preferred method for describing stochastic processes.

Remark 2.4 Markov assumption

At the heart of the derivation of eq. (2.42) lies a critical assumption: that $X(t)$ is a *Markov process*. For such processes, if we know the state at time t_1 and want to predict $X(t)$, where $t > t_1$, then also knowing the process' state at time $t_2 < t_1$ provides *no additional information*. Consequently, for a Markov process $X(t)$,

$$p_{X(t)|X(t_1),X(t_2)} = p_{X(t)|X(t_1)}\,, \tag{2.43}$$

if $t > t_1 > t_2$. Thus Markov processes can be thought of as "memoryless", and are sometimes referred to as such.

For delayed equations this assumption becomes a real problem because such equations are intrinsically *non*-Markovian. For example, for the delayed process defined by

$$dX(t) = a\,X(t) + b\,X(t-1) + c\,dW\,,$$

knowing $X(t-1)$ still provides additional information on $dX(t)$, even when $X(t)$ is given. In this case we would need the *joint* probability density $p_{X(t),X(t-1)|X(0)}(x_0, x_1|x^{init})$, rather than the simpler probability density $p_{X(t)|X(0)}(x|x^{init})$. In fact, in general non-Markovian processes are characterized by an *infinite* hierarchy of densities

$$p_{X(t),X(t-\tau_1)}(x_0, x_{\tau_1}),\ p_{X(t),X(t-\tau_1),X(t-\tau_2)}(x_0, x_{\tau_1}, x_{\tau_2})\ldots$$

It is this impossibility to collapse p.d.f.'s as in eq. (2.43) that has stymied efforts to formulate a Fokker-Planck-type description for SDDEs.

2.3.4 Multivariate SDEs

Although we've focused on scalar SDEs until now, there is no fundamental restriction to extending the formalism to multivariate equations.[34] Analogously to multivariate ODEs, the general Langevin equation becomes

$$dX_i(t) = f_i(t, \boldsymbol{X}(t))\,dt + \sum_{j=1}^{M} g_{ij}(t, \boldsymbol{X}(t))\,dW_j \qquad i = 0, 1, \ldots, N\,, \tag{2.44}$$

[33] Solutions to multivariate PDEs with more than 3 or 4 dimensions are rarely attempted without special computing resources.

[34] A nice succinct exposition of this extension can be found in [26].

where we view $\boldsymbol{X}(t) = (X_1(t), X_2(t), \ldots, X_N(t))^\intercal$ as an N-dimensional column vector of random variables. In the fully general case, there can be up to N independent noise sources, hence $d\boldsymbol{W}$ itself can be a vector.

2.3.5 The Ornstein-Uhlenbeck process

One of the most ubiquitous classes of the Langevin equation is the *Ornstein-Uhlenbeck process*. Defined by having linear drift and additive noise, it is the natural extension of linear ODEs to stochastics and one of the few cases of SDEs for which complete analytical solutions exist.

For constant parameters a_{ij} and Q_{ij}, a general multivariate Ornstein-Uhlenbeck process takes the form

$$dX_i(t) = \sum_{j=1}^{N} a_{ij} X_j(t) dt + \sum_{k=1}^{M} Q_{ik} dW_k \,. \tag{2.45}$$

Both terms on the right can evidently be expressed as matrix products, which means we can write this more compactly as

$$d\boldsymbol{X}(t) = \boldsymbol{L}\boldsymbol{X}(t)dt + \boldsymbol{Q}d\boldsymbol{W} \,, \tag{2.46}$$

where the components of $\boldsymbol{L}$ and $\boldsymbol{Q}$ are the parameters a_{ij} and Q_{ij} respectively.

2.3.5.1 Integral solution to the Ornstein-Uhlenbeck process

Equation (2.46) can be viewed as a standard inhomogeneous ODE, with the time-dependent driving function given by $\boldsymbol{Q}d\boldsymbol{W}$. The usual approach for such ODE systems is to first solve the homogeneous equation, and then find the general solution by using the method of variation of parameters to obtain an inhomogeneous solution; the general solution will be the sum of the two. Here the homogeneous equation

$$d\boldsymbol{X}(t) = \boldsymbol{L}\boldsymbol{X}(t)dt \tag{2.47}$$

is linear, so as long as $\boldsymbol{L}$ is diagonalizable, eq. (2.47) is easily solved by finding the eigenvectors

$$\boldsymbol{L}\boldsymbol{u}_n = \lambda_n \boldsymbol{u}_n \,,$$

and adding them to form the general solution:

$$\boldsymbol{X}^{\mathrm{h}}(t) = \sum_{n=1}^{N} C_n \boldsymbol{u}_n e^{\lambda_n t} \,. \qquad (\boldsymbol{X}^{init} = \textstyle\sum_n C_n \boldsymbol{u}_n)$$

(We consider the coefficients C_n to be random variables, to allow for the fact that the initial condition may itself be probabilistic.)

We can determine the coefficients c_n by constructing the dual set $\{\boldsymbol{u}^\dagger\}$ to the basis $\{\boldsymbol{u}_n\}$, just as we did on page 13. The situation here is in fact easier: since we are dealing with finite vectors, the linear forms (i.e. row vectors) $\boldsymbol{u}^\dagger$ are simply the left eigenvectors of $\boldsymbol{L}$,

$$\boldsymbol{u}_n^\dagger \boldsymbol{L} = \lambda_n \boldsymbol{u}_n^\dagger \,,$$

which we can normalize such that they are orthonormal with the u_n,

$$\boldsymbol{u}_n^\dagger \boldsymbol{u}_m = \delta_{nm} \,. \tag{2.48}$$

Then the coefficients are given by $c_n = \boldsymbol{u}_n^\dagger(\boldsymbol{X}^{init})$ and we can write the solution

$$\boldsymbol{X}^{\mathrm{h}}(t) = \sum_{n=1}^{N} e^{\lambda_n t} \boldsymbol{u}_n \boldsymbol{u}_n^\dagger(\boldsymbol{X}^{init}) \,.$$

We can diagonalize $\boldsymbol{L}$ in the basis of its eigenvectors,

$$\boldsymbol{L} = \sum_n \lambda_n \underbrace{\boldsymbol{u}_n \boldsymbol{u}_n^\dagger}_{N\times N \text{ matrix}} , \tag{2.49}$$

and therefore we can define[35]

$$\boldsymbol{G}^{\mathrm{h}}(t) = e^{\boldsymbol{L}t} = \sum_{n=1}^{N} e^{\lambda_n t} \boldsymbol{u}_n \boldsymbol{u}_n^\dagger \,, \tag{2.50}$$

such that we have

$$\boldsymbol{X}^{\mathrm{h}}(t) = \boldsymbol{G}^{\mathrm{h}}(t) \boldsymbol{X}^{init} \,. \tag{2.51}$$

Substituting this back into the eq. (2.47), we also see that

$$d\boldsymbol{G}^{\mathrm{h}}(t) = L\boldsymbol{G}^{\mathrm{h}}(t) dt \,.$$

Propagation operator

The operator $\boldsymbol{G}^{\mathrm{h}}(t)$ is called the *Green's function* or *propagator* of the homogeneous system. It's not required that the $\{u_n, u_n^\dagger\}$ form a biorthogonal set for the Green's function to exist, but if they do, then it can be expressed as a matrix, as shown in eq. (2.50). Although they act on different spaces, the similarity of this equation with eq. (2.22) is not a coincidence: delayed equations can be recast as multivariate equations, as we show in section 4.1.

Propagators in general must satisfy

$$\boldsymbol{G}(t)\boldsymbol{G}(t') = \boldsymbol{G}(t+t') \,,$$

[35] The rightmost equality is easy to see once we note that $\boldsymbol{L}^\alpha = \sum_n \lambda_n^\alpha \boldsymbol{u}_n \boldsymbol{u}_n^\dagger$ and $e^{\boldsymbol{L}t} = 1 + \boldsymbol{L}\,t + \frac{1}{2}\boldsymbol{L}^2\,t^2 + \cdots$.

and as a consequence,

$$\boldsymbol{G}^{-1}(t) = \boldsymbol{G}(-t)\,.$$

Following the method of variation of parameters, we now make the ansatz

$$\boldsymbol{X}^{\mathrm{inh}}(t) = \boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{c}(t)\,,$$

by replacing $\boldsymbol{X}^{init}$ by $\boldsymbol{c}(t)$ in eq. (2.51). Substituting the ansatz into eq. (2.46) leads to

$$\begin{aligned} d\left(\boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{c}(t)\right) &= \boldsymbol{L}\boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{c}(t)\,dt + \boldsymbol{Q}\,dW \\ \underbrace{\left[d\boldsymbol{G}^{\mathrm{h}}(t)\right]}_{=L\boldsymbol{G}^{\mathrm{h}}(t)dt}\boldsymbol{c}(t) + \boldsymbol{G}^{\mathrm{h}}(t)\,d\boldsymbol{c}(t) &= \boldsymbol{L}\boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{c}(t)\,dt + \boldsymbol{Q}\,dW \\ \boldsymbol{G}^{\mathrm{h}}(t)d\boldsymbol{c}(t) &= \boldsymbol{Q}\,d\boldsymbol{W}(t) \\ d\boldsymbol{c}(t) &= \boldsymbol{G}^{\mathrm{h}}(-t)\,\boldsymbol{Q}\,d\boldsymbol{W}(t)\,. \end{aligned}$$

Setting the initial value $\boldsymbol{X}^{\mathrm{inh}}(0) = 0$, this gives us

$$\boldsymbol{X}^{\mathrm{inh}}(t) = \boldsymbol{G}^{\mathrm{h}}(t)\int_0^t \boldsymbol{G}^{\mathrm{h}}(-t')\,\boldsymbol{Q}\,d\boldsymbol{W}(t')\,.$$

Applying the property $\boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{G}^{\mathrm{h}}(t') = \boldsymbol{G}^{\mathrm{h}}(t+t')$, this becomes

$$\boldsymbol{X}^{\mathrm{inh}}(t) = \int_0^t \boldsymbol{G}^{\mathrm{h}}(t-t')\,\boldsymbol{Q}\,d\boldsymbol{W}(t')\,,$$

which we rewrite as

$$\boldsymbol{X}^{\mathrm{inh}}(t) = \int_0^t \boldsymbol{G}^{\mathrm{h}}(t')\boldsymbol{Q}d\boldsymbol{W}(t-t')$$

using the change of variables $t - t' \to t'$. Hence the general solution takes the form

$$\boldsymbol{X}(t) = \boldsymbol{X}^{\mathrm{h}}(t) + \boldsymbol{X}^{\mathrm{inh}}(t) = \boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{X}^{init} + \int_0^t \boldsymbol{G}^{\mathrm{h}}(t')\,\boldsymbol{Q}\,d\boldsymbol{W}(t-t')\,. \tag{2.52}$$

2.3.5.2 Evaluating the Ornstein-Uhlenbeck stochastic integral

We can evaluate the stochastic integral in eq. (2.52) by treating the infinitesimal increments as though they were well-defined Gaussian random variables (eq. (2.41)),

$$\boldsymbol{G}^{\mathrm{h}}(t')\,\boldsymbol{Q}\,d\boldsymbol{W}(t-t') \sim \mathcal{N}(\boldsymbol{\mu}^{\mathrm{incr}}(t')\,,\boldsymbol{\Sigma}^{\mathrm{incr}}(t'))\,,$$

for which we need to determine $\boldsymbol{\mu}^{\mathrm{incr}}(t')$ and $\boldsymbol{\Sigma}^{\mathrm{incr}}(t')$.

> The notation $X \sim \mathcal{D}(\theta)$ is standard for indicating that a random variable X follows a distribution $\mathcal{D}$ parameterized by θ. This of course conflates with the use of the symbol $\sim$ to indicate asymptotic proportionality, but in practice context is usually sufficient to avoid ambiguity.

Since $\langle dW(t) \rangle$ vanishes independently of t (see eq. (2.36)), we likewise have

$$\boldsymbol{\mu}^{\text{incr}}(t') = \langle \boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \rangle = 0$$

(as before, the brakets $\langle \cdot \rangle$ denote an expectation value over the ensemble of possible noise realizations). Therefore the components of the increment's covariance matrix become

$$\begin{aligned}
\Sigma^{\text{incr}}_{ij}(t') &\equiv \text{Cov}\left(\left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_i , \left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_j \right) = \Sigma^{\text{incr}}_{ji}(t') \\
&= \text{E}\left[\left(\left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_i - \underbrace{\mu^{\text{incr}}_i(t')}_{=0} \right) \left(\left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_j - \underbrace{\mu^{\text{incr}}_j(t')}_{=0} \right) \right] \\
&= \left\langle \left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_i \left[\boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\, d\boldsymbol{W}(t-t') \right]_j \right\rangle \\
&= \left\langle \left(\sum_{k,l} G^{\text{h}}_{ik}(t')\, \boldsymbol{Q}_{kl}\, dW_l(t-t') \right) \left(\sum_{k',l'} G^{\text{h}}_{jk'}(t')\, \boldsymbol{Q}_{k'l'}\, dW_{l'}(t-t') \right) \right\rangle \\
&= \sum_{k,k',l,l'} G^{\text{h}}_{ik}(t')\, \boldsymbol{Q}_{kl}\, G^{\text{h}}_{jk'}(t')\, \boldsymbol{Q}_{k'l'}\, \underbrace{\langle dW_l(t-t')\, dW_{l'}(t-t') \rangle}_{\delta_{ll'}\, dt'} \\
&= \sum_{k,k',l,l'} G^{\text{h}}_{ik}(t')\, \boldsymbol{Q}_{kl}\, G^{\text{h}}_{jk'}(t')\, \boldsymbol{Q}_{k'l'}\, \delta_{ll'}\, dt' \\
\boldsymbol{\Sigma}^{\text{incr}}(t') &= \boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\boldsymbol{Q}^{\intercal} (\boldsymbol{G}^{\text{h}}(t'))^{\intercal} dt' \,.
\end{aligned}$$

The increments being independent multivariate Gaussian random variables, their sum (i.e. their integral) is also Gaussian.[36] Moreover, the mean and covariance of their sum are respectively the sums of their means and of their covariances. Thus the result of the stochastic integral is also a Gaussian random variable, with mean 0 and covariance $\int_0^t \boldsymbol{\Sigma}^{\text{incr}}(t') dt'$. The general solution then becomes becomes

$$\boldsymbol{X}(t) = \boldsymbol{G}^{\text{h}}(t) \boldsymbol{X}^{init} + \boldsymbol{Y}(t) \,,$$

where $\boldsymbol{Y}(t)$ is a multivariate Gaussian random variable following

$$\boldsymbol{Y}(t) \sim \mathcal{N}\left(0, \int_0^t \boldsymbol{G}^{\text{h}}(t')\, \boldsymbol{Q}\boldsymbol{Q}^{\intercal} (\boldsymbol{G}^{\text{h}}(t'))^{\intercal} \right) .$$

[36] It is not true in general that the sum of multivariate Gaussians is itself a multivariate Gaussian, but it is when they are independent. Therefore this reasoning would not be valid for coloured noise. However, coloured noise is itself most often generated from an Ornstein-Uhlenbeck process (which involves white noise by definition) – e.g. [27]. We suspect that at least in such cases, we could simply append the noise generating process to the system: if the original system with coloured noise is N-dimensional, and the noise generating Ornstein-Uhlenbeck process is M-dimensional, than we can construct an $N+M$-dimensional white noise process whose N first dimensions match the process of interest.

If the initial condition is known with certainty, then this can be formally expressed as a Gaussian variable with a vanishing covariance matrix,

$$\boldsymbol{X}^{init} \sim \mathcal{N}(\boldsymbol{X}^{init}, 0)\,.$$

In this case we get

$$\boldsymbol{X}(t) \sim \mathcal{N}\left(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t)\right)\,, \tag{2.53}$$

where

$$\boldsymbol{\mu}(t) = \boldsymbol{G}^{\mathrm{h}}(t)\boldsymbol{X}^{init} \tag{2.54}$$

$$\boldsymbol{\Sigma}(t) = \int_0^t \boldsymbol{G}^{\mathrm{h}}(t')\,\boldsymbol{Q}\boldsymbol{Q}^{\intercal}(\boldsymbol{G}^{\mathrm{h}}(t'))^{\intercal}dt'\,. \tag{2.55}$$

Equations (2.52), (2.54) and (2.55) are given in [20, pp. 38-40], with some differences in the derivation of the covariance.

2.3.5.3 Evaluating the covariance an Ornstein-Uhlenbeck process

When a biorthogonal set exists, we can do even better than eq. (2.55) and actually evaluate the integral. In a similar vein as [20, p. 155], we substitute the matrix form (2.50) of $\boldsymbol{G}^{\mathrm{h}}(t)$ into eq. (2.55), getting

$$\begin{aligned}
\boldsymbol{\Sigma}(t) &= \int_0^t \sum_{n,m} e^{\lambda_n t'}\boldsymbol{u}_n\boldsymbol{u}_n^{\dagger}\boldsymbol{Q}\boldsymbol{Q}^{\intercal}e^{\lambda_m t'}\left(\boldsymbol{u}_m^{\dagger}\right)^{\intercal}\left(\boldsymbol{u}_m\right)^{\intercal}dt' \\
&= \sum_{n,m}\boldsymbol{u}_n\underbrace{\left(\boldsymbol{u}_n^{\dagger}\boldsymbol{Q}\boldsymbol{Q}^{\intercal}\left(\boldsymbol{u}_m^{\dagger}\right)^{\intercal}\right)}_{\in\mathbb{C}}\left(\boldsymbol{u}_m\right)^{\intercal}\int_0^t e^{(\lambda_n+\lambda_m)t'}\,dt' \\
&= \sum_{n,m}\boldsymbol{u}_n^{\dagger}\boldsymbol{Q}\boldsymbol{Q}^{\intercal}\left(\boldsymbol{u}_m^{\dagger}\right)^{\intercal}\frac{e^{(\lambda_n+\lambda_m)t}-1}{\lambda_n+\lambda_m}\boldsymbol{u}_n\left(\boldsymbol{u}_m\right)^{\intercal}\,.
\end{aligned} \tag{2.56}$$

The terms in this sum should be viewed as composed of three parts:

- $\boldsymbol{u}_n^{\dagger}\boldsymbol{Q}\boldsymbol{Q}^{\intercal}\left(\boldsymbol{u}_m^{\dagger}\right)^{\intercal}$: a constant scalar proportional to the noise (or to the square of $\boldsymbol{Q}$);
- $\frac{e^{(\lambda_n+\lambda_m)t}-1}{\lambda_n+\lambda_m}$: a time-dependent scalar;
- $\boldsymbol{u}_n(\boldsymbol{u}_m)^{\intercal}$: a matrix related to the subspaces of the eigenvectors u_n and u_m.

Equation (2.56) is used in section 4.2, and forms the basis for our time-dependent auto-covariance function.

2.3.6 Stationary solutions

Whereas fixed points are a staple in the study of ODEs, they cannot occur in SDEs due to the noise. However, we can have stationarity in the p.d.f.: a probability density that satisfies

$$\frac{\partial p^{\mathrm{s}}(t,\boldsymbol{x})}{\partial t} = 0$$

is called a *stationary solution*.[37] In many ways these solutions play a role analogous to fixed points. For instance, similarly to stable fixed points, a stationary solution is the p.d.f. to which a system converges when we let time go to infinity

$$p^{\mathrm{s}}(\boldsymbol{x}) = \lim_{t\to\infty} p(t, \boldsymbol{x})\,. \tag{2.57}$$

Stationary solutions and fixed points are not, however, complete analogues. Consider for instance the p.d.f. of a process described by the Langevin equation. Since Gaussian noise can take any value between $-\infty$ and ∞, the process can jump to *any* state, at *any* point in time. Consequently, there can only be one stationary p.d.f. satisfying eq. (2.57).[38] For linear systems, each stable (resp. unstable) fixed point corresponds to a maximum (resp. minimum) of the p.d.f., although for nonlinear systems this is not always the case [28]. The notion of an "unstable" stationary solution would also be bizarre, since p.d.f.'s are aggregates of many realizations.

Mathematically, studying stationary solutions has many advantages. They allow us to assign a unique solution to an SDE, independent of the initial condition. The independence of the process' statistics on time and the large t limit also make it easier to simplify the solution. And while we lose all the transient dynamics, in many situations they are not our main interest anyway. For all these reasons, stationary solutions have received a lot more attention that general solutions.

On the other hand, we can also have systems which are constantly kicked away from their stationary solution[39] due to time-dependent inputs, in which case we really need the general solution to be able to study transient behaviour. In particular, this is often what happens in neuronal systems, which originally motivated this study. Still, stationary solutions remain the most natural baseline to which we can compare the large t limit of a general solution.

[37] Sometimes "steady state" is also used.

[38] This is easy to see by contradiction. Suppose there are two stationary p.d.f.'s. Each p.d.f. has its own "attractor": a set of initial conditions for which it is the stationary solution. But because of the noise, we can always jump from one attractor to the other. Ergo, the two p.d.f.s must be the same.

[39] It's usually safe to assume a stationary solution exist for systems modelling realistic processes.

Chapter 3

Stochastic delay differential equations today

As we've seen, possibly the most powerful assumption we can make about a stochastic process is that it's Markovian. It is this assumption that breaks the "infinite hierarchy" mentioned in remark 2.4 and obtain a differential equation for the process' p.d.f.. And quite often it's very reasonable: as long as the interactions are local and instantaneous, a process will always fill the "memoryless" requirement to be Markovian. Of course, with delay equations all this falls apart since they are anything *but* memoryless. This doesn't mean things are hopeless for SDDEs, and in fact a number of efforts over the last two decades have considerably advanced our understanding – most of them focusing on stationary solutions. In this short chapter, we present an overview of these efforts and identify the challenges we later address in chapter 4.

Delayed Ornstein-Uhlenbeck process

A commonly studied equation is the delayed Ornstein-Uhlenbeck process

$$dX_0(t) = [aX_0(t) + bX_0(t-\tau_1)]dt + q\,dW(t)\,, \tag{3.1}$$

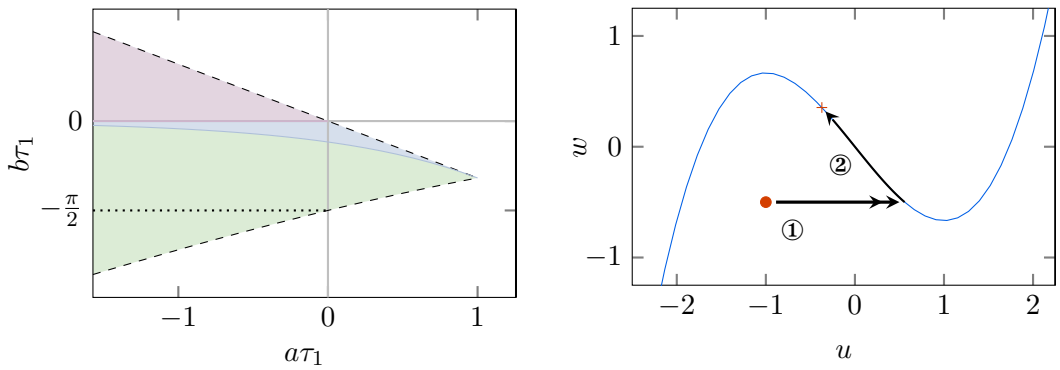
in large part because it's one of the simplest possible SDDEs. It is in fact the particular case of eq. (1.1) where $\kappa(\tau) = a\delta(\tau)+b\delta(\tau-\tau_1)$. And while it looks deceptively simple, we remain unaware of a full analytical characterization of this equation.

We do however have complete characterization of its stationary solution. This is largely helped by the fact that since the process is Gaussian, it is completely described by its mean and *auto*covariance[1]– the covariance of the process with itself at different

$$\Sigma_{\tau\nu}(t) \equiv \mathrm{Cov}\left(X_0(t-\tau), X_0(t-\nu)\right)\,. \tag{3.2}$$

(This notation is slightly redundant in that we could fix $\nu = 0$, but has the advantage of matching the earlier notation for the covariance matrix.)

[1]Strictly speaking this is only the case when the joint distribution of any pair of normal variables $(X(t_1), X(t_2))$ is itself normal, which is not true in general. However we know that for two *independent* normal random variables the joint distribution is also normal, and since the increments dW are in fact independent normal random variables, we are assured that this condition is satisfied for Ornstein-Uhlenbeck processes. For other processes we may not have this property, and then higher order moments would be needed.



(a) **(b)**

Figure 3.1 – **(a)** Stability diagram for the delayed Ornstein-Uhlenbeck process given by eq. (3.1). The shaded area delimited by dashed lines corresponds to the parameters satisfying the inequalities in eq. (3.3). The purple, blue and green shaded regions further indicate for which parameters we respectively find 1, 2 and 0 real eigenvalues. **(b)** A visualization of a two-variable fast-slow system where $\dot{u} \gg \dot{w}$. ① The fast subsystem quickly brings the system onto one of the states on the *critical manifold*. ② On the critical manifold the fast derivative is zero, and the "slow" derivative brings the system to the equilibrium. This specific example is a FitzHugh-Nagumo system, as given by eq. 1.13 of [30] with $q_1 = 0$, $q_2 = 15$, $q_3 = -1.2$, $q_4 = 0.8$.

3.1 Stationary solution of the Ornstein-Uhlenbeck process

It is possible to use the method of steps to extend SDE methods to SDDEs. Küchler and Mensch used this approach to provide existence and uniqueness proofs for the delayed Ornstein-Uhlenbeck process of eq. (3.1) [1]. They then determine that necessary and sufficient conditions for the existence of the stationary solution [1, p. 30] are

$$a + b < 0 \qquad \text{and} \qquad \tau_1 < \frac{\cos^{-1}(-a/b)}{\sqrt{b^2 - a^2}}\,. \tag{3.3}$$

Figure 3.1a illustrates the set of parameters satisfying these conditions, which interestingly are independent of q. This is a consequence of the noise being additive: in a similar study by Mackey and Nechaeva [27,29], where results are extended to coloured and multiplicative noise, examples are given where the existence conditions depend on q.

Küchler and Mensch also obtain a closed-form expression for the autocovariance $\Sigma_{\tau 0}(\infty)$ of the stationary solution [1, eqs. 2.27,2.28], valid on the interval $\tau \in [0, \tau_1]$.

3.2 Multiscale decompositions

Multiscale systems exhibit dynamics on two or more (usually time) scales. A classic example is the FitzHugh-Nagumo model, developed to reproduce the spiking behaviour of

neurons, which can be separated into fast and slow subsystems [30, sec. 1.4]. Using multiscale techiques, we can separate such systems into "fast" and "slow" subsystems, effectively reducing the dimensionality of the problem (fig. 3.1b). Such approaches have been applied to delay [31] and stochastic [32] systems, so there's no reason not to apply them to SDDEs – indeed, they already have [33, 34].

Multiscale techniques can typically be used wherever we have a clear separation of time scales. However, their main interest is in reducing a system's dimensionality: we still need to solve the reduced system using other methods. Nevertheless, since DDEs are never really scalar, there may be ways in which we can apply ideas from multiscale approaches to general cases such as eq. (1.1) – in fact, the reduction to dominant amplitudes we present in section 4.3.2 can be seen as just that. But they won't constitute the thrust of our argument.

3.3 Decomposition into eigenstates

In section 2.1.3.3 we presented a method for solving DDEs by decomposing the states into eigenfunctions, which we illustrated by solving eq. (3.1) when $q = 0$. We showed that each eigenfunction led to an independent non-delayed equation for its amplitude, as given by eq. (2.14). Amann et al. [4] show that in general, the eigenvalue decomposition of eq. (3.1) yields

$$dc_n = \lambda_n c_n dt + q\mathcal{N}_n dW\,. \tag{3.4}$$

A priori this doesn't get us very far since a solution is composed of an infinite number of these amplitudes. However, we might get somewhere if we can show that these amplitudes can be ordered such that truncating after more and more terms gives an increasingly accurate estimate of the solution (in, say, L^2); to our knowledge this has only been done under rather strong assumptions – e.g. in [4, sec. 5] it is assumed that a few eigenvalues are positive. In section 4.3 we relax these conditions substantially.

Using this approach, it's not too hard to compute the expectation value $\langle X_0(t)\rangle$. Indeed, it is given by the solution to the deterministic portion of eq. (3.1):

$$\begin{aligned}
d\langle X_0(t)\rangle = \langle dX_0(t)\rangle &= \langle aX_0(t) + bX_0(t-\tau_1)\rangle\, dt + q\underbrace{\langle dW(t)\rangle}_{=0} \\
\frac{d\langle X_0(t)\rangle}{dt} &= a\langle X_0(t)\rangle + b\langle X_0(t-\tau_1)\rangle \\
\langle X_0(t)\rangle &= \sum_n c_n(t)\underbrace{u_{n,0}}_{=1} \qquad (3.5)\\
&= \sum_n c_n(t)\,. \qquad (3.6)
\end{aligned}$$

Amann et al. are also able to compute the autocovariance of the stationary solution[2] of

[2]They do not give an expression for a time-dependent autocovariance, although it is obtainable using a

eq. (3.1) to be[3]

$$\Sigma^{\mathrm{s}}_{0,-\tau}(\infty) = -q^2 \sum_{kl} \mathcal{N}_k \mathcal{N}_l \frac{e^{\lambda_k \tau}}{\lambda_k + \lambda_l}\,, \quad \tau \geq 0\,. \tag{3.7}$$

A time dependent version of this expression also appears in [35, eq. 8]. A generalization of this expression is one of the key results of chapter 4 (compare especially with eq. (4.55)), where we also extensively study the convergence of the series.

We use a negative time lag here to indicate a comparison between $X_0(t)$ and a later variable $X_0(t+\tau)$. This is not just notational preference, as the sum on the right-hand-side of eq. (3.7) does not always correspond to the autocovariance; exactly where they correspond is investigated in section 4.4, and especially fig. 4.8.

Being able to compute the mean and autocovariance, we are already very close to a full solution to eq. (3.1): all we really need to do is extend the latter to all times t. An eigenfunction decomposition such as this has the advantage of being familiar to physicists and of allowing a more intuitive understanding of the process in terms of modes. This is why, despite forcing us to work with infinite sums, we feel that they remain our best bet for developing a generalizable solution method for SDDEs.

3.4 Reduction to the centre manifold

Linear equations such as eq. (3.1) generally occur after linearizing a nonlinear differential equation around an equilibrium point: in the vicinity of the equilibrium, the linearized system can usually be used to approximate the nonlinear one – unless one or more eigenvalues are zero. States associated to these zero eigenvalues tend to decay more slowly and form a lower dimensional subspace called the centre manifold. In many cases one can argue that the dominant behaviour occurs in this subspace, and therefore that the system can be reduced to the centre manifold. Techniques can then be used to introduce correction terms, thereby extending the neighbourhood in which the reduced system is valid.

Within the last decade, a number of studies have extended or applied centre manifold theory to SDEs with either delays or forcing [36, 37]. In particular, Hutt and Lefebvre [7] consider a centre manifold reduction for an SDDE with distributed delays. In order to do this, they use results developed for the centre manifold analysis of stochastic *functional* differential equations, for which SDDEs are a special case. This is a powerful but rather specialized language, which we prefer to avoid; the bulk of their work also deals with the nonlinearities of their system, which are not our focus. At the core of their centre manifold reduction, however, is a linear system which they treat in a manner very similar to the way in which we treat our own results: a functional differential operator analogous to eq. (2.11) is defined, and the solution developed in terms of eigenfunctions and associated

very similar procedure.

[3] This is a reformulated form of their eq. 27, where we have redefined variables to match their definitions in chapter 4.

eigenfunctionals. The manner in which the latter are obtained is however different than the approach we use ourselves in chapter 4.

3.5 The conditional average drift and perturbative methods

None of the methods discussed above attempt to derive an equivalent to the Fokker-Planck equation, which as we've seen can be a powerful complementary approach to SDEs. Part of the reason is the key part the Markov assumption plays in its derivation. Nevertheless, we can at least write a "Fokker-Planck-like" equation for SDDEs; for instance, Guillouzic et al. [38] show that for an equation of the form

$$dX_0 = f(X_0, X_{\tau_1})dt + g(X_0)dW\,,$$

the p.d.f. satisfies

$$\frac{\partial}{\partial t}p_{X_0(t)}(x_0,t) = -\frac{\partial}{\partial x_0}\left\{p_{X_0(t)}(t,x_0)\bar{f}(t,x_0)\right\} + \frac{1}{2}\frac{\partial^2}{\partial x_0^2}\left\{p_{X_0(t)}(t,x_0)g^2(x_0)\right\}\,, \tag{3.8}$$

where $\bar{f}$ is the *conditional average drift* (CAD), defined as

$$\bar{f}(x_0,t) \equiv \int f(x_0,x_{\tau_1})p_{X_{\tau_1}(t)|X_0(t)}(t,x_{\tau_1}|x_0)\,dx_{\tau_1}\,. \tag{3.9}$$

Remark 3.1

A state x of an SDDE is a function in $\mathcal{S} = C([-\tau_{max},0],\mathbb{R})$, and therefore a true Fokker-Planck equation would be over that space. But we don't really want a probability density over $\mathcal{S}$: with the latter being uncountably infinite dimensional, it's not even clear how we'd define such a p.d.f.. Ergo the interest in a Fokker-Planck-*like* equation for p_{X_0}, which is over $\mathbb{R}$ instead. Indeed, this makes a lot of sense, since we can always recover $p_{X_\tau(t)}$ as $p_{X_0(t-\tau)}$.

What we lose by doing this is information on the joint probability densities such as p_{X_0,X_τ}, which are required to evaluate the CAD. Consequently, in contrast to a Fokker-Planck equation, knowing the solution to an equation like eq. (3.8) up to some time t is *not* sufficient to step the solution forward to $t+dt$. This issue remains unresolved and is part of the reason efforts have focused on stationary solutions; we give some ideas as to how it could be addressed in section 5.3.

If we could evaluate the CAD, we'd be well on our way towards a usable Fokker-Planck equation for SDDEs. The problem of course is that it involves a p.d.f. for X_{τ_1} which is conditioned on the *future* state X_0.[4] A number of efforts have tackled this problem [3,38,39] by assuming the delay – here τ_1 – is small, but what is missing is a solution that works for any delay size.

[4] Incidentally, this condition on a future state is where we see the non-Markovian nature of the process.

We should mention here the recent work by Bhat and Madushani [5]. While their approach is purely numerical and makes a rather strong assumption on the joint probabilities, they do succeed in relaxing the assumption on the delay.

Finding a general scheme to evaluate the CAD is the overarching goal that has guided the work presented in this thesis. This essentially boils down to determining the joint probability densities, which in turn can be described by the set of their moments. Thus ultimately we want to develop a formalism capable of extracting these moments from the density $p_{X_0}(t)$. We outline how we envision our results might serve to reach this goal in chapter 5.

3.6 Approximation by a random walk

A *random walk* is a stochastic process with discrete random steps taken at regular intervals. Each step is always of the same size, so the stochasticity is reduced to a simple binary choice between stepping right or left. Despite this apparent limitation, it's possible to construct a random walk that is statistically equivalent to the continuous process defined by eq. (3.1) with $a = 0$. We can then work out the delayed Fokker-Planck equation as a limit when the interval between steps goes to zero [40, eq. 11.68] – the result is the same expression as eq. (3.8). We can also obtain expressions for the covariance using this approach.

Viewing a stochastic process as a random walk is enticing: it drastically simplifies the mathematics, since it means that we deal with easy-to-define probabilities over a discrete space. It is also much easier to visualize a process composed of regular, discrete jumps. However, we still have the same issue of needing to evaluate the CAD. In addition, for every new stochastic process we would need to derive the appropriate mapping to a random walk – we are not aware of systematic ways of doing this, or indeed whether it is always possible.

3.7 Path integral methods

Path integral methods are relatively well established for SDEs [41, 42], but very few efforts have applied them to SDDEs, one of the exceptions being Navarra et. al. [43]. These methods are in a similar situation as the functional analysis techniques mentioned in section 3.5: they have the potential to tackle very complex problems, but suffer from a more limited target audience due to their sophistication. A similar statement can be said for the application of variational to SDDEs, of which there are still few examples [44, 45].

We don't need the sophistication of these approaches for our own efforts, and don't look into them further.

3.8 Closing remarks

Having completed our survey of currently employed tools to tackle SDDEs, we see that current methods are still a bit haphazard. The core issue we are still learning to cope

with is the need for joint probability densities and the prohibitively high-dimensional state space required to track them. This is especially difficult in the case of distributed delays, as we have in eq. (1.1).

In the next chapter, we revisit the eigenstate approach suggested by Amann et al., fixing the issues raised in remark 2.3, extending the approach to distributed delays in the form of eq. (1.1) and obtaining a full time-dependent solution.

While these results are limited to linear SDDEs, they fall within a wider project to develop a means to compute the CAD. The scheme, which would be similar in spirit to the one by Bhat and Madushani, should allow for distributed delays and almost any form of drift. We sketch out how this might be achieved in chapter 5

Three main factors made us pursue the eigenstate approach rather than the other ones presented here:

- it seems to have the robustness needed to be generalizable to a wide class of problems;
- it paints an intuitive picture of the stochastic process in terms of simpler modes;
- it builds on concepts with which many researchers, especially physicists, are familiar.

It is certainly possible that the choice of some other approach may be better suited, especially under different selection criteria. However, given the natural way in which we are able to achieve our results, it is difficult to dispute that it satisfies our requirements.

Chapter 4

Solving the distributed Ornstein-Uhlenbeck process

We are now in a position to solve the distributed-delay equation (1.1), which we recall was

$$dX_0(t) = \left(\int_0^{\tau_{max}} \kappa(\tau) X_\tau(t)\, d\tau \right) dt + q\, dW\,, \tag{4.1}$$

with the deterministic initial condition

$$X(0) = \phi\,. \tag{4.2}$$

Recall that, following the notation of section 2.1.2, ϕ is a function in $\mathcal{S} = C([-\tau_{max}, 0], \mathbb{R})$. The variable $X(t)$ is a random variable on $\mathcal{S}$, the fact that it is a random variable indicated by the use of an upper case character. Both ϕ and $X(t)$ are functions of the time lag, usually denoted by τ, such that random variable $X_\tau(t)$ is the evaluation of $X(t)$ at a particular lag τ. The variable $X_0(t)$ corresponds to the "current" state of the system, at some time $t \geq 0$.

Since eq. (4.1) an Ornstein-Uhlenbeck process with the addition of a distributed delay, we refer to it as the "distributed Ornstein-Uhlenbeck process".

A solution to an SDDE can take many forms; at a minimum it should characterize the random variable $X_0(t)$ for all times $t > 0$. However, the function[1] X_0 – which we will refer to as the *pointwise solution* – says nothing of the joint p.d.f.'s of the process. (See our discussion of a "Fokker-Planck-like" equation in section 3.5, especially remark 3.1.) We often *would* need more information on the temporal structure, but as we noted in remark 3.1, looking for a full joint p.d.f. is unlikely to be successful. A more tractable approach is to consider the various joint moments, starting with the autocovariance.

In sections 4.1 and 4.2 we develop the tools to express the mean and autocovariance of eq. (4.1) as series over the process' modes. We then describe the dynamics of the modes themselves in section 4.3 and look into how reliable the series expressions would be in practice in section 4.4. Section 4.5 simply highlights the peculiar mathematical properties of a result obtained in the previous section and section 4.6 concludes this chapter with an example; those looking for a flavour of how our solution method works in practice may want to skip ahead to this last section.

[1] X_0 is a function from $[0, \infty)$ to a space of random variables.

——————— *** ———————

As we've seen, it's often best to treat states of a DDE as functions (section 2.1.2). This leads us to rewrite eq. (4.1) as a functional equation, as we did in eq. (2.10):

$$dX(t) = \boldsymbol{L}X(t)dt + \boldsymbol{Q}dW\,, \tag{4.3}$$

where

$$[\boldsymbol{L}X(t)]_\tau = \begin{cases} \int_0^{\tau_{max}} \kappa(\nu)X_\nu(t)\,d\nu & \text{if } \tau = 0 \\ -dx_\tau(t)/d\tau & \text{if } 0 < \tau \leq \tau_{max} \end{cases}$$

$$\boldsymbol{Q} = \begin{cases} q & \text{if } \tau = 0 \\ 0 & \text{if } 0 < \tau \leq \tau_{max}\,. \end{cases}$$

Note that the expression for $[\boldsymbol{L}X(t)]_\tau$ is little more than a transcription of eq. (2.11), with the first line given by the DDE or SDDE, and the second by $-dx_\tau(t)/d\tau$.

For the same reasons as in section 2.1.3.3, we know that the eigenfunctions of $\boldsymbol{L}$ must be

$$u_{n,\tau} = e^{-\lambda_n \tau}\,, \tag{4.4}$$

where the eigenvalues λ_n satisfy

$$\int_0^{\tau_{max}} \kappa(\tau)e^{-\lambda_n\tau}\,d\tau = \lambda_n\,, \tag{4.5}$$

a generalization of eq. (2.16). Equations (4.4) and (4.5) can be considered known quantities, derived from established theory which we presented section 2.1.3.3. We will use them to perform a sanity test on our results.

While we have expressions for the eigenvalues and eigenfunctions, we are still missing an expression for the dual set of eigenfunctionals $\{\boldsymbol{u}_n^{(N)\dagger}\}$. In section 2.1.3.3 we just posited a form and then verified its correctness, but that's hardly satisfying (or generalizable); now we develop a systematic procedure to derive the $\boldsymbol{u}_n^{(N)\dagger}$.

The core of our idea is to discretize the operators $\boldsymbol{L}$ and $\boldsymbol{Q}$ as $\boldsymbol{L}^{(N)}$ and $\boldsymbol{Q}^{(N)}$ respectively. Since the right and left eigenvectors of a finite dimensional matrix form a biorthogonal set, if the former converge to eq. (4.4) in the continuous limit (i.e. $N \to \infty$), then we can be fairly confident that the latter converge to proper expressions for the $u_n^\dagger$.

With expressions for the u_n and $u_n^\dagger$ in hand, we can switch freely between function and multivariate representations of an SDDE. Problems can be solved using either representation, and confidently translated to the other. This idea is illustrated in fig. 4.1 to serve as a visual reference.

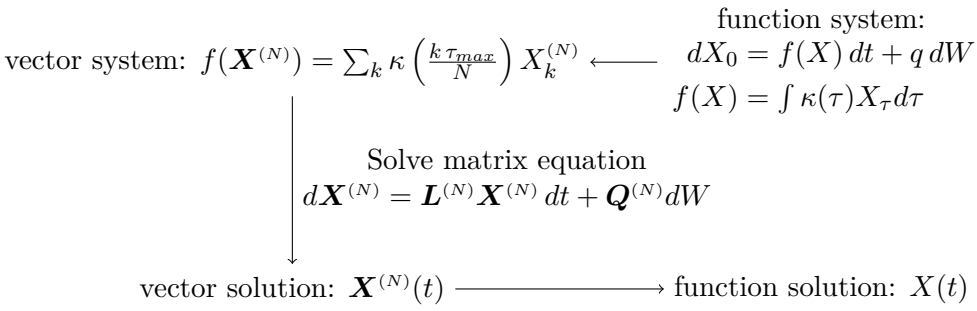


Figure 4.1 – "Big picture" illustration of our solution method. By building a robust correspondence between functional and matrix operators, we are able to use finite-dimensional linear algebra to solve problems in a function space.

4.1 The discretized integro-differential operator

We begin by discretizing the functional state X as the vector $\boldsymbol{X}$ indexed from 0 to N, exactly as we did in section 2.1.2.[2] (The value of N is arbitrary but remains fixed throughout our argument.) This allows us to approximate the integral in eq. (4.1) as a Riemann sum:[3]

$$\int_0^{\tau_{max}} \kappa(\tau) X_\tau(t)\,d\tau \approx \sum_{k=0}^{N-1} \kappa\left(\frac{k\,\tau_{max}}{N}\right) X_k^{(N)}(t)\frac{\tau_{max}}{N}\,, \tag{4.6}$$

where $X_k^{(N)}(t) = X_0^{(N)}\left(t - \frac{k\,\tau_{max}}{N}\right)$. The components satisfy the recursion relation (2.4), which we recall was

$$X_n^{(N)}(t) = X_{n-1}^{(N)}\left(t - \tfrac{\tau_{max}}{N}\right).$$

We can approximate $dX_0^{(N)}(t)$ by substituting eq. (4.6) into eq. (4.1):

$$dX_0^{(N)}(t) \approx \left(\sum_{k=0}^{N-1} \kappa\left(\frac{k\,\tau_{max}}{N}\right) X_k^{(N)}(t)\cdot\frac{\tau_{max}}{N}\right) dt + q\,dW\,. \tag{4.7}$$

To approximate $dX_n^{(N)}(t)$, we write

$$\frac{dX_n^{(N)}(t)}{dt} \approx \frac{X_n^{(N)}\left(t + \frac{\tau_{max}}{N}\right) - X_n^{(N)}(t)}{\tau_{max}/N}\,;$$

[2] We should mention here that there is in fact a very minor difference between the discretizations: here we exclude the point X_N, such that $X^{(N)}$ is N dimensional. We have to do this because the integral in eq. (4.6) is written as a sum of rectangles, and for a discretization with $N+1$ points there are N rectangles.

[3] This expression is somewhat symbolic in that it assumes that κ is continuous everywhere on $[0, \tau_{max}]$. But beyond simplifying the notation, this assumption doesn't have any real consequence. In particular, the weight of a Dirac delta at $\kappa(\tau_{max})$ may be shifted to $\kappa\left(\frac{(N-1)}{N}\tau_{max}\right)$ to ensure it is included in the sum. (A Dirac delta at $\kappa(\tau_{max})$ is in fact quite common, since it occurs any time we have discrete delays.)

substituting the recursion relation into this expression and multiplying by dt, we find

$$dX_n^{(N)}(t) \approx \left(\frac{X_{n-1}^{(N)}(t) - X_n^{(N)}(t)}{\tau_{max}/N} \right) dt \tag{4.8}$$

for $1 \leq n \leq N$. These approximations allow us to express the system in matrix form (from here onwards we drop the superscript on $\boldsymbol{X}^{(N)}$ to lighten the notation):

$$\begin{pmatrix} dX_0 \\ dX_1 \\ \vdots \\ dX_{N-1} \end{pmatrix} \approx \underbrace{\begin{pmatrix} \kappa(0)\frac{\tau_{max}}{N} & \kappa(\frac{\tau_{max}}{N})\frac{\tau_{max}}{N} & & \kappa\left(\frac{(N-1)}{N}\tau_{max}\right)\frac{\tau_{max}}{N} \\ N/\tau_{max} & -N/\tau_{max} & & \\ 0 & N/\tau_{max} & -N/\tau_{max} & \\ & & \ddots & \\ & & N/\tau_{max} & -N/\tau_{max} \end{pmatrix}}_{\equiv \boldsymbol{L}^{(N)}} \begin{pmatrix} X_0 \\ X_1 \\ \vdots \\ X_{N-1} \end{pmatrix} dt + \underbrace{\begin{pmatrix} q \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\equiv \boldsymbol{Q}^{(N)}} dW. \tag{4.9}$$

Defining $\boldsymbol{L}^{(N)}$ and $\boldsymbol{Q}^{(N)}$ by analogy with their continuous forms, we obtain the discretized version of eq. (4.3)

$$d\boldsymbol{X} = \boldsymbol{L}^{(N)}\boldsymbol{X}dt + \boldsymbol{Q}^{(N)}dW\,. \tag{4.10}$$

Using standard methods from linear algebra, we can compute the eigenvalues and eigenvectors of $\boldsymbol{L}^{(N)}$; we find these to be (see eqs. (A.3) to (A.5) of appendix A)

$$\lambda_n = \sum_{k=0}^{N-1} \kappa\left(\frac{k\,\tau_{max}}{N}\right) \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{-k} \left(\frac{\tau_{max}}{N}\right), \tag{4.11}$$

$$u_{n,i}^{(N)} = \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{-i}, \qquad (0 \leq i < N) \tag{4.12}$$

$$u_{n,i}^{(N)\dagger} = \begin{cases} \mathcal{N}_n^{(N)}\,, & (i = 0) \\ \mathcal{N}_n^{(N)} \sum_{k=i}^{N-1} \kappa\left(\frac{k\,\tau_{max}}{N}\right) \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{i-k-1} \left(\frac{\tau_{max}}{N}\right)^2, & (1 \leq i < N) \end{cases} \tag{4.13}$$

where the normalization constant $\mathcal{N}_n^{(N)}$ is given by (eq. (A.6))

$$\mathcal{N}_n^{(N)} = \left[1 + \sum_{k=1}^{N-1} k \cdot \kappa\left(\frac{k\,\tau_{max}}{N}\right) \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{-k-1} \left(\frac{\tau_{max}}{N}\right)^2\right]^{-1}. \tag{4.14}$$

The constant $\mathcal{N}_n^{(N)}$ is chosen such that

$$\boldsymbol{u}_n^{(N)\dagger}(\boldsymbol{u}_m^{(N)}) = \delta_{n,m} \equiv \begin{cases} 1\,, & \text{if } n = m, \\ 0\,, & \text{if } n \neq m. \end{cases} \tag{4.15}$$

Since the matrix $\boldsymbol{L}^{(N)}$ isn't singular[4], we can reasonably expect its eigenvectors to span the state space $\mathbb{R}^N$. Nevertheless, there are very particular forms of κ for which this is not true – see section 5.2 for an example.

[4]Unless κ is identically zero or $\tau_{max} = 0$, but then there'd be no DDE to begin with. We can ignore the case where we are "unlucky" and choose a κ which is zero at every point $\kappa(\frac{k\,\tau_{max}}{N})$, since in such cases we can just choose a different N.

4.1.1 Taking the continuous limit

To find the continuous forms of eqs. (4.11) to (4.13), we recall that each component k of a state vector $\boldsymbol{u}^{(N)}$ corresponds to a delay time $\tau = k\,\tau_{max}/N$. Correspondingly, when we substitute $k = N\tau/\tau_{max}$ into eq. (4.11) and take the limit,[5] we find that the characteristic equation now becomes

$$\begin{aligned}\lambda_n &= \lim_{N\to\infty} \sum_{\{\tau\}} \kappa(\tau) \left[\left(1 + \tfrac{\lambda_n \tau_{max}}{N}\right)^N\right]^{-\tau/\tau_{max}} \left(\tfrac{\tau_{max}}{N}\right) \\ &= \lim_{N\to\infty} \sum_{\{\tau\}} \kappa(\tau) \left[e^{\lambda_n \tau_{max}}\right]^{-\tau/\tau_{max}} \left(\tfrac{\tau_{max}}{N}\right) \\ &= \int_0^{\tau_{max}} \kappa(\tau) e^{-\lambda_n \tau} d\tau\,,\end{aligned}$$

which is in correspondence with eq. (4.5). Similarly, the continuous form of eq. (4.12),

$$\begin{aligned}u_{n,\tau} &= \lim_{N\to\infty} \left[\left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^N\right]^{-\tau/\tau_{max}} \\ u_{n,\tau} &= e^{-\lambda_n \tau}\,,\end{aligned}$$

matches that of eq. (4.4).

So we're in good shape: our discretization is consistent with expected results. What about the elements of the dual then? Well, following the same procedure, we are led to define the normalization constant as

$$\begin{aligned}\mathcal{N}_n &= \lim_{N\to\infty} \left[1 + \sum_{\{\tau\}\setminus 0} \left(\frac{N\tau}{\tau_{max}}\right) \cdot \kappa(\tau) \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{-\frac{N\tau}{\tau_{max}} - 1} \left(\frac{\tau_{max}}{N}\right)^2\right]^{-1} \\ &\equiv 1 \Big/ 1 + \int_{0^+}^{\tau_{max}} \kappa(\tau) \tau e^{-\lambda_n \tau} d\tau\,. \end{aligned} \tag{4.16}$$

Remark 4.1

The lower bound of 0^+ in the integral is a consequence of the sum in eq. (4.14) starting with $k = 1$. Excluding zero becomes important when κ includes a term proportional to $\delta(0)$, something that in fact happens rather frequently in models. Failure to do would otherwise lead to incorrect results.

[5] "Taking the limit", here and subsequently, is shorthand for assigning the value to which the Riemann sums converge as the partition fineness increases with N. This will involve a somewhat cavalier treatment of limits as it confounds partition fineness and proper limits, and we apologize to those this may offend.

Then for the eigenfunctionals we have

$$u^\dagger_{n,\tau} = \begin{cases} \mathcal{N}_n\,, & (\tau = 0) \\ \mathcal{N}_n \lim_{N\to\infty} \sum_{\{\nu\}_{\nu>\tau}} \kappa(\nu) \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{\frac{N(\tau-\nu)}{\tau_{max}} - 1} \left(\frac{\tau_{max}}{N}\right)^2, & (0 < \tau \le \tau_{max}) \end{cases}$$
$$= \begin{cases} \mathcal{N}_n\,, & (\tau = 0) \\ \mathcal{N}_n\, d\tau \int_\tau^{\tau_{max}} d\nu\, \kappa(\nu) e^{\lambda_n(\tau-\nu)}\,, & (0 < \tau \le \tau_{max}) \end{cases}$$

which suggests that we should define their action on an arbitrary function $x \in \mathcal{S}$ as

$$\begin{aligned} u^\dagger_n(x) &= \lim_{N\to\infty} \boldsymbol{u}^{(N)}_n(\boldsymbol{x}) \\ &\equiv \mathcal{N}_n x_0 + \mathcal{N}_n \int_{0^+}^{\tau_{max}} d\tau \int_\tau^{\tau_{max}} d\nu\, \kappa(\nu) e^{\lambda_n(\tau-\nu)} x_\tau \\ &= \mathcal{N}_n x_0 + \mathcal{N}_n \int_{0^+}^{\tau_{max}} d\tau\, e^{\lambda_n \tau} x_\tau \int_\tau^{\tau_{max}} d\nu\, \kappa(\nu) e^{-\lambda_n \nu}\,. \end{aligned} \tag{4.17}$$

To complete our construction of a biorthogonal basis for $\mathcal{S}$, we need to ensure that

$$u^\dagger_n(u_m) = \delta_{n,m}\,. \tag{4.18}$$

For this it suffices to note that eq. (4.15) is independent of N, and therefore remains true when we take the limit $N \to \infty$.[6]

——————— *** ———————

The vector $\boldsymbol{Q}^{(N)}$ is zero in every component except the first. For the first component, we have $Q^{(N)}_0 = q$. As a consequence, the only reasonable limit for $\boldsymbol{Q}^{(N)}$ when $N \to \infty$ is

$$Q_\tau \equiv q\, \delta_{\tau 0} = \begin{cases} q & \text{if } \tau = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

Note that $\delta_{\tau 0}$ is not a Dirac but a *Kronecker* delta, even though τ is a continuous index. This is a consequence of q being independent of N.

——————— *** ———————

[6] We haven't proved that the set $\{u_n\}$ forms a complete basis for $\mathcal{S}$, which strictly speaking is required to call the set $\{u_n, u^\dagger_n\}$ a biorthogonal basis. However the limiting procedure *does* show that this set forms a complete basis for discretized functions of arbitrary fineness. Since data and simulations are invariably discretized that's really all we care about, and any function of $\mathcal{S}$ not in $\lim_{N\to\infty} \mathrm{Span}\{u_n\}$ can be safely ignored.

We can check our expressions with the example of section 2.1.3.3, the one case where we *do* know the eigenfunctionals. In that example, we considered the equation

$$dx_0(t) = \big(ax_0(t) + bx_\tau(t)\big)\,dt\,, \qquad x(0) = \phi\,. \tag{eq. (2.11))}$$

We now put it in the form of eq. (4.1) by defining

$$\kappa(\tau) \equiv a\delta(\tau) + b\delta(\tau - \tau_1)\,. \tag{4.20}$$

Then the characteristic equation is

$$\lambda_n = \int_0^{\tau_1} [a\delta(\tau) + b\delta(\tau - \tau_1)]e^{-\lambda_n \tau} d\tau \tag{4.21}$$

$$= a + b\,e^{-\lambda_n \tau_1}\,, \tag{4.22}$$

which is the same expression we had in eq. (2.16). Similarly, the expressions we get when we substitute eq. (4.20) into eqs. (4.16) and (4.17),

$$\mathcal{N}_n = \frac{1}{1 + b\tau_1 e^{-\lambda_n \tau_1}} \tag{4.23}$$

$$u_n^\dagger(x) = \mathcal{N}_n x_0 + \mathcal{N}_n \int_{0^+}^{\tau_1} d\tau\, b e^{\lambda_n(\tau - \tau_1)} x_\tau\,, \tag{4.24}$$

match those of eq. (2.20).[7]

Remark 4.2

One might wonder why this process of transforming to and from a discrete representation should work with stochastic equations. After all, the procedure must assume that realizations "aren't too pathological", an often dangerous assumption to make when dealing with stochastic equations. The key is that realizations only need to be continuous, and that is the one nice property that stochastic realizations *do* have.

Another way to look at this is that this process of treating a differential equation as a set of discrete steps, and then obtaining the solution as the limit when the step size goes to zero, is very similar to how we defined stochastic integrals in the first place (section 2.3.2).

4.2 The general solution to the delayed Ornstein-Uhlenbeck process

Now that we are set up, solving eq. (4.3) is just a matter of applying known results to the discretized process and then passing to the continuous limit.[8]

[7] In eq. (2.20), the lower bound of the integral was 0 rather than 0^+ in order to be equivalent with the form in [4]. As long as x_0 remains finite (and it really should !), both integrals are equivalent.

[8] Implicit in this process is a change in the characteristic equation, so the eigenvalues are different (and in different numbers) for the discrete and continuous system, even though we label them the same way.

Noting that eq. (4.10) describes a multivariate Ornstein-Uhlenbeck process of the same form as eq. (2.46), we know that its solution is also given by the same Gaussian (eq. (2.53)),

$$\boldsymbol{X}(t) \sim \mathcal{N}(\boldsymbol{\mu}^{(N)}(t), \boldsymbol{\Sigma}^{(N)}(t)),$$

and therefore that the first component of the discretized system follows

$$X_0(t) \sim \mathcal{N}(\mu_0^{(N)}(t), \Sigma_{00}^{(N)}(t))\,. \tag{4.25}$$

So solving eq. (4.3) boils down to determining the continuous mean $\mu_0(t)$ and variance $\Sigma_{00}(t)$ as limits of $\mu_0^{(N)}(t)$ and $\Sigma_{00}^{(N)}(t)$. We will need the discretized initial condition $\phi^{(N)}$, which is obtained in the same way as $X^{(N)}$: $\phi_k^{(N)} \equiv \phi_\tau$, where $\tau = k\,\tau_{max}/N$.

Equation (2.54), after substitution of eq. (2.50) and making the correspondence $\boldsymbol{X}^{init} \to \phi^{(N)}$, gives the mean vector

$$\boldsymbol{\mu}^{(N)}(t) = \sum_{n=0}^{N-1} e^{\lambda_n t} \boldsymbol{u}_n^{(N)} \underbrace{\boldsymbol{u}_n^{(N)\dagger}(\phi^{(N)})}_{\in\mathbb{C}}\,.$$

(Recall that $\boldsymbol{u}_n^{(N)\dagger}$ is a row vector, and therefore its application to ϕ returns a scalar.) To get $\mu_0^{(N)}(t)$, all we need to do is evaluate each eigenvector $\boldsymbol{u}_n^{(N)}$ at time lag 0. Since $u_{n,0}^{(N)} = 1$ for all n, we find

$$\mu_0^{(N)}(t) = \sum_{n=0}^{N-1} e^{\lambda_n t} \boldsymbol{u}_n^{(N)\dagger}(\phi^{(N)})\,. \tag{4.26}$$

Taking the continuous limit now is simply a matter of replacing $\boldsymbol{u}_n^{(N)\dagger}$ by $u_n^\dagger$ and $\phi^{(N)}$ by ϕ:

$$\mu_0(t) = \sum_{n=0}^{N-1} e^{\lambda_n t} u_n^\dagger(\phi)\,. \tag{4.27}$$

We can get the variance from the covariance matrix given in section 2.3.5. There we considered an arbitrary noise coupling matrix $\boldsymbol{Q}$, but here $\boldsymbol{Q}^{(N)}$ is a single column whose only non-zero entry is the first row. Thus we substitute $Q_i = Q_i^{(N)} = q\,\delta_{i,0}$ into eq. (2.56), allowing us to simplify it to

$$\begin{aligned}
\boldsymbol{\Sigma}^{(N)}(t) &= q^2 \sum_{n,m} \left(\sum_{k,l=0}^{N-1} u_{n,k}^{(N)\dagger} \delta_{k,0} \delta_{l,0} u_{m,l}^{(N)\dagger} \right) \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} \boldsymbol{u}_n \left(\boldsymbol{u}_m\right)^\intercal \\
&= q^2 \sum_{n,m} u_{n,0}^{(N)\dagger} u_{m,0}^{(N)\dagger} \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} \boldsymbol{u}_n \left(\boldsymbol{u}_m\right)^\intercal\,; \\
\Sigma_{ij}^{(N)}(t) &= q^2 \sum_{n,m} \mathcal{N}_n^{(N)} \mathcal{N}_m^{(N)} \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} u_{n,i}^{(N)} u_{m,j}^{(N)}\,.
\end{aligned}$$

Moving to the continuous limit is again simply a question of substituting the eigenfunctions:

$$\Sigma_{\tau\nu}(t) = q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} e^{-\lambda_n \tau} e^{-\lambda_m \nu} \tag{4.28}$$

$$\mathrm{Var}(X_0(t)) = \Sigma_{00}(t) = q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} \,. \tag{4.29}$$

Thus the pointwise solution to eq. (4.3), under deterministic initial conditions, is

$$X_0(t) \sim \mathcal{N}\left(\mu_0(t), \Sigma_{00}(t)\right) \,, \qquad (t > 0) \tag{4.30}$$

where the mean and variance are given by eqs. (4.27) and (4.29) respectively. Equation (4.28), for its part, gives us the covariance between two processes $X_\tau(t)$ and $X_\nu(t)$. Since these are really the same process $X_0(t)$ at two different time lags, this gives us a way to compute its autocovariance.

Virtual discretization Since the discretization step does not appear in the final eqs. (4.27) to (4.29), our discretization is effectively *virtual*: a very useful calculation tool that only appears in the intermediate steps.

Time-dependent autocovariance function The "indices" τ and ν of the *autocovariance function* $\mathbf{\Sigma}$ – or of the eigenfunctions u_n – range from 0 to τ_{max}. Accordingly, we can use eq. (4.28) to compute autocovariances with lags within this range. Attempting to use the same equation to compute the autocovariance with time lags *outside* this range is akin to trying to extract the fourth component of a three component vector, and may lead to similarly nonsensical results (see fig. 4.7 on page 62). While the range

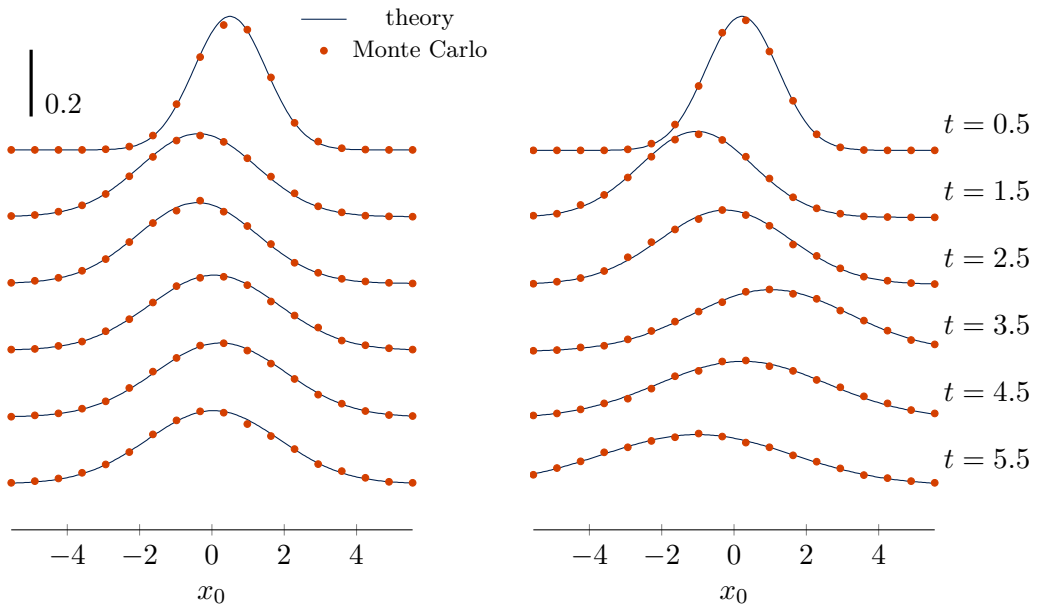
$$I_{\text{lag}} \equiv [0, \tau_{max}] \,. \tag{4.31}$$

is rather tight, recall that to compute the conditional average drift – and eventually derive a Fokker-Planck-like equation for SDDEs – we only need the process' temporal structure for lags up to τ_{max}.

Autocovariance functions are frequently sought after in the study of stochastic processes because they provide a lot of information on the temporal structure of a system's behaviour. Consequently, eq. (4.28) is one the key results of this work, and the subject of further study in section 4.4.2. Time lags for an autocovariance function, usually denoted s or t', can range from $-\infty$ to ∞. However in our case we denote them with τ or ν to remind the reader that in this case time lags are limited to I_{lag}.

Although it normally makes little sense for an autocovariance function to be "time-dependent" since it is defined for all times *by definition*, in the case here we can truly say that of $\mathbf{\Sigma}$. Indeed, if it weren't, the limitation of time lags to I_{lag} would be disastrous.

A separate question to the range of valid time lags is whether or not the double series in eq. (4.28) converges. This is something we look into in section 4.4



(a) $\beta = -1$ **(b)** $\beta = -1.56$

Figure 4.2 – Numerical verification of the theoretical expression eq. (4.30) for processes given by eq. (4.32). Monte Carlo estimates are obtained as averages over 10 000 simulations; computing these takes a few hours on a personal computer, while the theoretical expressions are computed in seconds. Simulations are done with initial condition $\phi_\tau = 1$, $\tau \in [0, \tau_{max}]$ and noise amplitude $q = \sqrt{2}$. The theoretical variance is calculated using the truncation scheme (I) (described on page 58) with $n_{max} = 40$.

4.2.1 Numerical verification

It's relatively straightforward to check the validity of eq. (4.30) numerically by integrating eq. (4.1) using standard algorithms.[9] Aggregating the runs, we can construct an approximate p.d.f. for each time step – for these Gaussian processes, 10 000 runs is approximately what is needed to get reasonably smooth probability densities. This approach is laborious of course, which is one reason to prefer the theoretical expression (4.30).

Figure 4.2 shows such comparison tests for processes

$$dX_0(t) = \beta X_1(t) + q\, dW\,; \tag{4.32}$$

these correspond to those given by eq. (4.42) when we set $\alpha = 0$. Referring to fig. 3.1a, we see that this system is stable when $-\pi/2 < \beta < 0$. We can limit ourselves to values of β within this range since otherwise the system diverges exponentially, which is rarely something we want in a model.

[9] See appendix C.

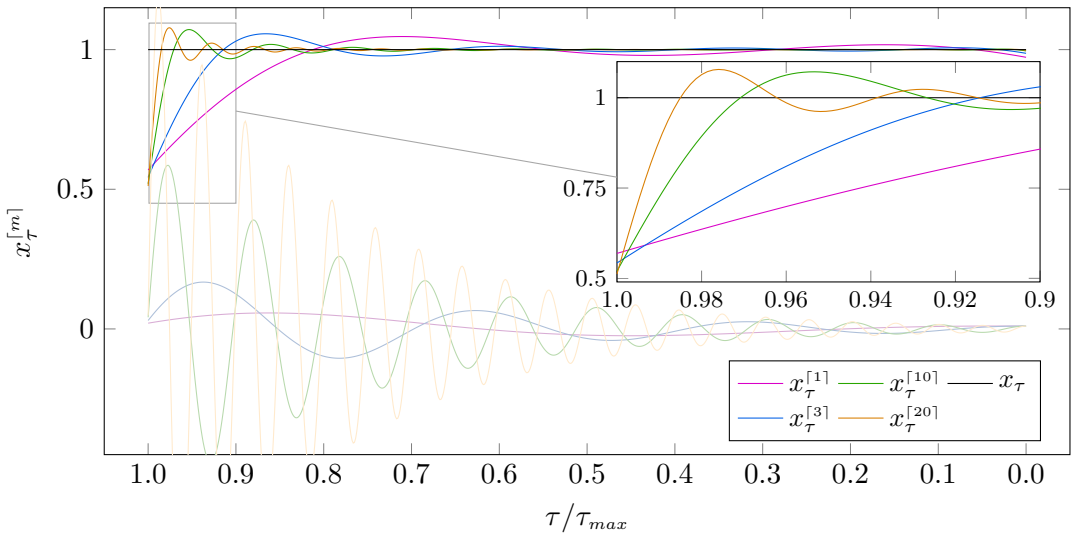


Figure 4.3 – Reconstruction of the constant state $x_\tau = 1$, $0 \leq \tau \leq \tau_{max}$ using the modes of eq. (4.35). As more modes are included, $x_\tau^{\lceil m \rceil}$ converges to x_τ everywhere except at $\tau = \tau_{max}$. The abscissa is inverted to reflect the fact that τ is a look-back variable. The 1st, 3rd, 10th and 20th modes (scaled down by a factor 1000) are drawn as fainter lines centred on zero. Modes correspond to the kernel $\kappa(\tau) = -\delta(\tau - \tau_{max})$.

4.3 SDDEs as amplitude processes

4.3.1 Deriving dynamical equations for the amplitudes

At any particular point in time, a state $x(t)$ can be decomposed into eigenstates u_n:

$$x(t) = \sum_n c_n(t) u_n \,, \tag{4.33}$$

where the coefficients c_n are given by

$$c_n(t) = u_n^\dagger(x(t)) \,. \tag{4.34}$$

Since the eigenfunctions u_n correspond to damped oscillations, we interpret them as modes with amplitudes c_n, and although it may be a slight abuse of the term, we refer to eq. (4.33) as the *spectral decomposition* of a state. In practice we need to truncate the sum, which we indicate by

$$x^{\lceil m \rceil} \equiv \sum_{n=-m}^{m} c_n u_n \,. \tag{4.35}$$

Figure 4.3 shows how a state is increasingly better approximated as we increase the number of modes, as we would expect. The exception is the point $x_{\tau_{max}}^{\lceil m \rceil}$ at the very end, which in this particular example begins around 0.67 and *decreases* monotonically with

m towards $\frac{1}{2}$ (while $x_{\tau_{max}} = 1$). Thus when using a spectral decomposition (4.33), we shouldn't put too much trust in the computed value of $x_{\tau_{max}}$, nor in expressions in which it appears. Any other point however, e.g. $x_{\tau_{max}-\epsilon}$, should be fine as long as it's computed with enough modes. Since the convergence is not uniform, for a given accuracy, increasingly more modes will be required as ϵ is reduced.

Fourier series exhibit a similar effect, called the "Gibb's phenomenon", where the series does not converge to the expected value around discontinuous jumps. While it does introduce some artifacts, it does not invalidate the Fourier series, and we expect the same to be true in our case. In particular, the functions $x^{\lceil m \rceil}$ should converge to x in L^2.

If the states $x(t)$ are realizations of a random variable $X(t)$, then the amplitudes $c_n(t)$ are realizations of random variables $C_n(t)$:

$$X(t) = \sum_n C_n(t) u_n \,. \tag{4.36}$$

In section 4.2 we essentially performed a spectral decomposition of the states $X(t)$, computed the mean and covariance of the amplitudes, and from those, we derived the mean and covariance of the original states $X(t)$. This conversion to and from an eigenspace is all packaged up into eqs. (4.27) to (4.29), but it's sometimes useful to think in terms of the amplitudes themselves. For this we need dynamical equations for them, which can be formally obtained by differentiating eq. (4.34),[10]

$$\begin{aligned} dC_n = d\left(u_n^\dagger(X(t))\right) &= u_n^\dagger(dX(t)) \\ &= u_n^\dagger\left(\boldsymbol{L}X(t)\,dt \,+\, \boldsymbol{Q}dW\right) \\ &= u_n^\dagger\left(\sum \lambda_n C_n u_n\right) dt \,+\, \mathcal{N}_n\Big(\underbrace{Q_0}_{=q} + \int_{0^+}^{\tau_{max}} d\tau \int_\tau^{\tau_{max}} d\nu\, \kappa(\nu) e^{\lambda_n(\tau-\nu)} \underbrace{Q_\tau}_{=0}\Big) dW \\ &= \lambda_n C_n\, dt + q\,\mathcal{N}_n\, dW \,. \end{aligned} \tag{4.37}$$

> Although eq. (4.37) shows that amplitude dynamics can be solved individually, they remained coupled (and therefore correlated) through their common scalar noise dW. Moreover, eq. (4.37) is an SDE rather than an SDDE.
>
> **Remark 4.3**

Since eq. (4.37) is Markovian, we can use a standard technique to obtain from it the mean amplitude,

$$\begin{aligned} d\left\langle C_n\right\rangle = \left\langle dC_n\right\rangle &= \left\langle \lambda_n\, C_n\, dt + q\,\mathcal{N}_n\, dW\right\rangle = \lambda_n \left\langle C_n\right\rangle dt \\ \left\langle C_n(t)\right\rangle &= e^{\lambda_n t}\left\langle C_n(0)\right\rangle \,, \end{aligned} \tag{4.38}$$

[10] Strictly speaking, the quantity $u_n^\dagger(X(t))$ would involve integrating a random variable, and it is not the goal of this project to explore how such a quantity might be defined. One can, however, sidestep this issue by working out the discretized expressions and translating them back to continuous space as in section 4.1.1; the result is the same as eq. (4.37).

and mean power,

$$\begin{aligned}
C_n^*(t+dt)\,C_n(t+dt) &= \Big[C_n^*(t) + \lambda_n^*\,C_n^*(t)dt + q\,\mathcal{N}_n^*\,dW\Big]\Big[C_n(t) + \lambda_n\,C_n(t)dt + q\,\mathcal{N}_n\,dW\Big] \\
&= C_n^*(t)\,C_n(t) + (\lambda_n + \lambda_n^*)\,C_n^*(t)\,C_n(t)\,dt + q^2\,\mathcal{N}_n^*\mathcal{N}_n\,dW^2 \\
&\qquad + (C_n^*\,\mathcal{N}_n + C_n\,\mathcal{N}_n^*)\,dW + \mathcal{O}(dt\,dW) \\
d\,|C_n(t)|^2 &= C_n^*(t+dt)\,C_n^*(t+dt) - C_n^*(t)\,C_n^*(t) \\
\frac{d}{dt}\left\langle |C_n(t)|^2 \right\rangle &= 2\,\mathrm{Re}(\lambda_n)\left\langle |C_n(t)|^2 \right\rangle + q^2\;|\mathcal{N}_n|^2 \\
\left\langle |C_n(t)|^2 \right\rangle &= e^{2\,\mathrm{Re}(\lambda_n)t}\left(\left\langle |C_n(0)|^2 \right\rangle + \frac{q^2\;|\mathcal{N}_n|^2}{2\,\mathrm{Re}(\lambda_n)}\right) - \frac{q^2\;|\mathcal{N}_n|^2}{2\,\mathrm{Re}(\lambda_n)}\,,
\end{aligned} \tag{4.39}$$

of each mode. In particular, in the stationary state, the mean power of each mode is

$$\left\langle |C_n(\infty)|^2 \right\rangle = -\frac{q^2\;|\mathcal{N}_n|^2}{2\,\mathrm{Re}(\lambda_n)}\,. \tag{4.40}$$

4.3.2 Dimensionality reduction due to stochasticity

We can get an idea of the importance of each mode by computing its relative stationary power,[11]

$$\text{rel. stat. power(mode } n) = \frac{\left\langle |C_n(\infty)|^2 \right\rangle}{\sum_k \left\langle |C_k(\infty)|^2 \right\rangle}\,; \tag{4.41}$$

table 4.1 shows how this quantity varies with feedback strength for the SDDE[12]

$$dX_0(t) = [\alpha\,X_0(t) + \beta\,X_1(t)]\,dt + dW \tag{4.42}$$

for the three most important modes. Every mode corresponds to a unique eigenvalue; most of these are complex, and so come in conjugate pairs. This lead to two modes having the same relative power. For a given α and β however, between zero and two eigenvalues are real (and when two are real, they are necessarily distinct). These can be identified in table 4.1 by the fact that their relative power is different from that of all other modes – examples would be the parameter choices $\alpha = -1$, $\beta = 0.5$ (one real eigenvalue) and $\alpha = 0$, $\beta = -0.3$ (two real eigenvalues – modes 3 and 4 in this case have the same power). Along the stability boundaries, one or two of the eigenvalues have zero real part; looking at eq. (4.40), we see that this implies that their power is infinite. Thus the *relative* power of all other modes, as given in the table, is zero.

[11] Equations (2.17) and (4.16) can be used to calculate the eigenvalues and normalization constants required to compute eq. (4.41).

[12] Note that eq. (4.42) is simply a non-dimensionalized form of eq. (3.1), where we've made the changes of variables $t/\tau_1 \to t$, $X(t)/q \to X(t)$, $a\tau_1 \to \alpha$ and $b\tau_1 \to \beta$.

Further inspection of table 4.1 reveals that for feedback strengths of order unity or less, the overwhelming majority of the power is concentrated in the first mode. This brings us to the interesting realization that stochasticity in some way *reduces* the dimensionality of SDDEs. For example, for the single delay system with $\alpha = 0$, $\beta = -1$, around 99.2% of the power is concentrated in the first two complex conjugate modes.[13] What happens is that the drift component of eq. (4.37), by driving the process towards equilibrium, erases memory of the initial condition. At the same time, the diffusion component $q\mathcal{N}_n\, dW$ adds power to each mode following a ratio fixed by the normalization constants $\mathcal{N}_n$. The net effect is a gradual rebalancing of the power distribution until it matches the equilibrium values shown in table 4.1.

If this effect were primarily driven by the drift, we would expect it to happen on the timescale of $\tau_{\text{decay}} = -\frac{1}{\text{Re}(\lambda_2)}$ (the time for the second slowest mode to decay). However in fig. 4.4 we see that this collapse onto the dominant modes happens almost instantaneously, whereas $\tau_{\text{decay}} \sim 0.5$. This suggests that the observed collapse to dominant modes is indeed primarily a stochastic effect.

4.4 Convergence of the covariance series

In order to evaluate eq. (4.28), we need to know how well the series converges. We can do this analytically, which provides some measure of intellectual satisfaction, but doesn't get us much further than determining the domain of *absolute* convergence.[14] It does however allow us to make a reasonable conjecture for conditional convergence, which we then verify empirically. Empirical tests can also give us a good idea of the convergence speed, whether absolute or conditional.[15]

Our analytical treatment requires a sufficiently precise understanding of the distribution of eigenvalues λ_n, so we start with that.

4.4.1 Distribution of the eigenvalues

Recall that the eigenvalues of the operator $\boldsymbol{L}$ are given by eq. (4.5),

$$\int_0^{\tau_{max}} \kappa(\tau) e^{-\lambda_n \tau}\, d\tau = \lambda_n\,. \tag{4.43}$$

For the kinds of models of interest to us, X is always real; likewise κ is a real function. We also assume the system to be stable – or equivalently, that κ is such that all eigenvalues of $\boldsymbol{L}$ have negative real part.

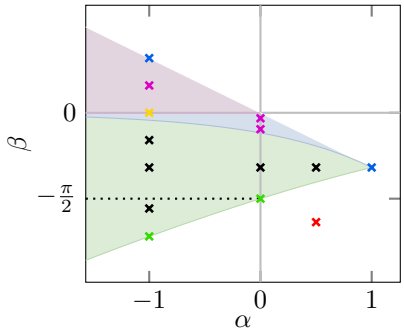
[13] When two modes are complex conjugate, they are best seen as a single mode along with its phase, rather than as completely distinct.

[14] We are not claiming here that it is impossible to pursue further analytical treatment; this is simply left for future work.

[15] For numerical applications, convergence speed is the most important criterion. Conditional convergence is not necessarily a problem, as long as the summation order is clear. In remark 2.3 we had a non-trivial "sum rule", which is why conditional convergence was a problem in that case.

Table 4.1 – Relative mode powers for the process (4.42) as computed by eq. (4.41). Figure 3.1a is reproduced with added marks for reference. Each mode corresponds to an eigenvalue; conjugate pairs lead to two modes of equal power. Unstable systems (red) have nonsensical values for relative power since they don't have a stationary state. Systems on the stability boundary (blue and green marks) have eigenvalues with zero real part which completely dominate. Systems with no delay at all (orange) have only a single mode.

		Relative stationary power		
α	β	mode 1	mode 2	mode 3
1	-1	1	0	0
0.5	-1	0.499	0.499	0.001
0.5	-2	0.508	0.508	-0.006
0	-0.1	0.995	0.004	0.0003
0	-0.3	0.893	0.105	0.0006
0	-1	0.496	0.496	0.003
0	$-\pi/2$	0.5	0.5	0
-1	1	1	0	0
-1	0.5	0.952	0.019	0.019
-1	0	1	–	–
-1	-0.5	0.484	0.484	0.010
-1	-1	0.481	0.481	0.012
-1	-1.75	0.489	0.489	0.007
-1	-2.261	0.5	0.5	0
-5	-1	0.360	0.360	0.081
-10	-1	0.231	0.231	0.12
-100	99	0.152	0.127	0.127
-100	-1	0.023	0.023	0.023
-100	-100	0.410	0.410	0.045



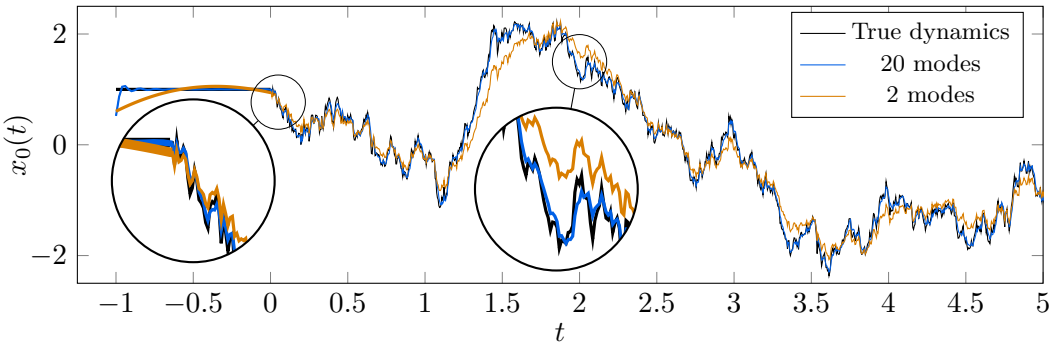


Figure 4.4 – Approximating the realization of an SDDE process by its amplitude dynamics (eq. (4.37)). The process has a single delay feedback kernel $\kappa(\tau) = \delta(\tau - 1)$, noise amplitude $q = \sqrt{2}$ and initial condition $\phi \equiv 1$. All three processes are generated with the same sequence of stochastic increments so that they correspond to the same realization. While restricting to only the first two (complex conjugate) modes makes for a very poor approximation of the initial segment where $-1 \leq t \leq 0$, the large-scale dynamics are nevertheless captured for $t > 0$. The numerical integration step size is quite small, $\Delta t = 0.0003$, in order to correctly capture the dynamics of the higher modes.

A first thing to note is that complex eigenvalues come in conjugate pairs. Indeed, since we impose that κ be a real function, if λ satisfies eq. (4.43), then so does its complex conjugate λ^*:

$$\begin{aligned}\int_0^{\tau_{max}} \kappa(\tau) e^{-\lambda_n^* \tau}\, d\tau &= \int_0^{\tau_{max}} \kappa(\tau) \left(e^{-\lambda_n \tau}\right)^*\, d\tau \\ &= \left(\int_0^{\tau_{max}} \kappa(\tau) e^{-\lambda_n \tau}\right)^*\, d\tau \\ &= \lambda_n^* \,.\end{aligned}$$

We know there are an infinite number of eigenvalues, and we've assumed that they all have negative real part. To reflect their organization into conjugate pairs, we use the same ordering scheme as in section 2.2:

- $|n| < |m|$ if $\lambda_n \in \mathbb{R}$ and $\lambda_m \in \mathbb{C}$
- $|n| < |m|$ if $\mathrm{Re}(\lambda_n) > \mathrm{Re}(\lambda_m)$
- $\lambda_n \in \mathbb{R}$ iff $\lambda_{-n} \in \mathbb{R}$
- $\lambda_n > \lambda_{-n} > \lambda_{n+1}$ if $\lambda_n, \lambda_{n+1} \in \mathbb{R}$
- if real eigenvalues are even-numbered, then λ_0 is left undefined
- $\lambda_{-n} = \lambda_n^*$, $\mathrm{Im}(\lambda_n) > 0$ if $\lambda_n \in \mathbb{C}$.

For the values of the Lambert function illustrated in fig. 2.5, this scheme corresponds to assigning the same index in absolute value to conjugate pairs, and ordering them right-to-left.

The distribution of eigenvalues for the single delay case (see eq. (2.17)) suggests that the real part of eigenvalues goes to $-\infty$ as we increase n. If this is true – and we assume it is – then *the characteristic equation* (4.43) *is determined, for large n, only by* $\kappa(\tau_{max})$;[16] we show this in appendix B.1.[17] For questions of convergence, we can therefore approximate the characteristic equation (4.43) by

$$b e^{-\lambda_n \tau_{max}} \sim \lambda_n \,,$$

where b is some expression dependent on $\kappa(\tau_{max})$. This implies that the eigenvalues are asymptotically given by

$$\lambda_n \sim \frac{W_n(b\tau_{max})}{\tau_{max}} \,, \tag{4.44}$$

and the normalization constants by

$$\mathcal{N}_n \sim \frac{1}{1 + \lambda_n \tau_{max}} \,.$$

[16] Unless of course $\kappa(\tau_{max}) = 0$, which should only happen when κ tails off to infinity. In that case the distribution of eigenvalues may be substantially different. We give an example in section 5.2 of a system with a *finite* number of eigenvalues.

[17] Note that appendix B.1 only shows the result for a kernel with a Dirac delta at $\tau = \tau_{max}$. Although we expect the property to generalize to the continuous case, this remains to be formally shown.

(We can get the last two expressions by substituting $a = 0$ into eqs. (2.17) and (2.21).) Substituting the asymptotic expression (2.28) for the Lambert function in eq. (4.44), we finally obtain an expression suitable for our convergence analysis,

$$\lambda_n \sim -\frac{1}{\tau_{max}} \log \left| \frac{2\pi(n-1)}{b\,\tau_{max}} \right| + \frac{i}{\tau_{max}} \left(2\pi(n-1) - \operatorname{sgn}(b) \frac{\pi}{2} \right) . \tag{4.45}$$

Remark 4.4 The fact that eigenvalues are asymptotically given by the Lambert function implies that only a finite number of them can be real. (The Lambert function itself has at most two real eigenvalues.)

4.4.2 Analytical convergence results

Now that we have a handle on the distribution of eigenvalues, we can proceed with an analytical characterization of the convergence of eq. (4.28). This will serve as a reference for interpreting the empirical results of the next section.

For convenience, let's define the terms

$$\Phi_{\tau,\nu,t}(n,m) \equiv q^2 \mathcal{N}_n \mathcal{N}_m \frac{e^{(\lambda_n+\lambda_m)t} - 1}{\lambda_n + \lambda_m} e^{-\lambda_n \tau} e^{-\lambda_m \nu} \tag{4.46}$$

$$\sim q^2 \frac{e^{(\lambda_n+\lambda_m)t} - 1}{(1+\lambda_n\tau_{max})(1+\lambda_m\tau_{max})(\lambda_n+\lambda_m)} e^{-\lambda_n \tau} e^{-\lambda_m \nu} , \tag{4.47}$$

where we have substituted the asymptotic expressions for $\mathcal{N}_n$ and $\mathcal{N}_m$. The covariance series of eq. (4.28) can now be written as

$$\Sigma_{\tau\nu}(t) = \sum_{n,m} \Phi_{\tau,\nu,t}(n,m) . \tag{4.48}$$

Remark 4.5 The series on the right-hand side of eq. (4.48) is always real, as we show in appendix B.2.

Since the imaginary parts of $\Phi_{\tau,\nu,t}(n,m)$ cancel out, we are only really interested in their real parts. But $\mathrm{Re}(\Phi_{\tau,\nu,t}(n,m))$ varies in an extremely non-uniform fashion and it's very difficult to get an intuitive sense of how it behaves by looking only at the analytical forms. So instead we view the terms as distributed on an infinite two-dimensional grid, indexed by n and m. We represent the value of each term by a colour, and repeat for different time lags.

The result is a set of figures like those in fig. 4.5, which shows the distribution of the stationary terms

$$\Phi_{\tau,\tau,\infty}(n,m) = -q^2 \mathcal{N}_n \mathcal{N}_m \frac{e^{-(\lambda_n+\lambda_m)\tau}}{\lambda_n + \lambda_m}, \tag{4.49}$$

for a single delay kernel and $q = 1$. The symmetry of these figures is of course a consequence of having set $\tau = \nu$.

When viewing these figures, one wants to keep in mind that we are looking for two things. First we want to identify which terms *dominate* the sum, i.e. which terms have the largest magnitude compared to the others. These are the terms in *bright* colours, either bright blue or bright red. Thus when a figure is brighter than another, it means that more of its terms are non-negligible – in practice, longer computation times, since we need to keep more terms in the sum. Conversely, very dark areas of the figure, whether dark blue or dark gold, don't contribute much to the sum.

The other thing we are looking for is the *sign* of the terms, indicated by their colour: blue for positive values and gold or red for negative values.

One prominent feature is that the dominant terms in eq. (4.48) are those for which n or m are close to zero: they occupy the centre of each figure – the two bright blue squares corresponding to $n = \pm 1$, $m = \mp 1$ particularly stand out. This is what we would expect, since the central terms correspond to eigenvalues with the least negative real part.

In order to actually evaluate eq. (4.48) will need to choose a truncation scheme; since the dominant terms are in the centre, it makes sense to sum them by "radiating outwards" from there. The two most obvious schemes that do this are

$$\text{(I)} \quad \sum_{n=1}^{n_{max}} \sum_{m=1}^{n_{max}} \Phi_{\tau,\nu,t}(n,m) + \Phi_{\tau,\nu,t}(-n,m) + \Phi_{\tau,\nu,t}(n,-m) + \Phi_{\tau,\nu,t}(-n,-m), \text{ and}$$

$$\text{(II)} \quad \sum_{k=2}^{k_{max}} \sum_{n=1}^{k-1} \Phi_{\tau,\nu,t}(n,k-n) + \Phi_{\tau,\nu,t}(-n,k-n) + \Phi_{\tau,\nu,t}(n,-k+n) + \Phi_{\tau,\nu,t}(-n,-k+n),$$

with the exact sum being recovered by setting either n_{max} or k_{max} to infinity. On the two-dimensional grid, these both correspond to squares, with one rotated 45° with respect to the other.

Another feature highlighted by fig. 4.5 is that while the central terms don't change much with τ, those around them increase by several orders of magnitude when τ is increased. This means that *for larger time lags, more and more terms become non-negligible*, and therefore that the sum likely requires more terms to be computed accurately.

——————— *** ———————

We want to start by determining the domain within which eq. (4.48) converges absolutely, i.e. the values of τ and ν which ensure that

$$\sum_{n,m} |\Phi_{\tau,\nu,t}(n,m)| < \infty\,, \qquad \forall t\,. \tag{4.50}$$

For absolute convergence the specific summation order doesn't really matter; we choose the one described by (II) since it makes the analysis simplest. (This choice is not, however, a particularly good one to evaluate the series, as fig. 4.7 (page 62) illustrates.)

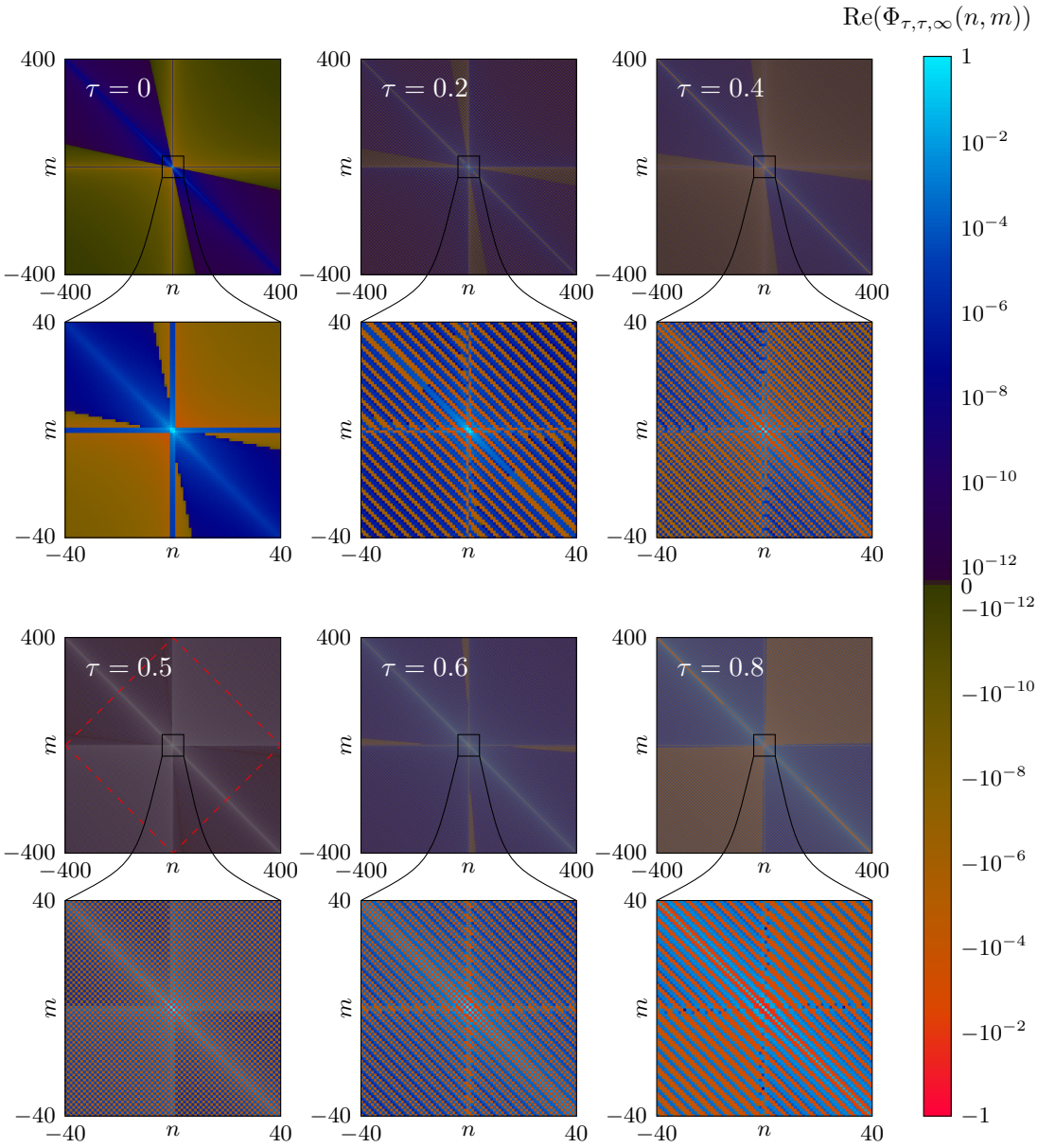


Figure 4.5 – Real part of the stationary state terms $\Phi_{\tau,\tau,\infty}(n,m)$, as given by eq. (4.49), for various time lags τ and system parameters $q = 1$, $\kappa(\tau) = -\delta(\tau - 1)$. Figures are inherently pixelated since we associate to each term a 1x1 uniformly coloured square.
Terms become negligible as $|n|$ and $|m|$ increase, with associated squares correspondingly becoming darker. Plots for small values of τ show a lot of dark blue and dark gold, indicating that most values are near zero and therefore that good estimations of the covariance can be obtained with very few terms. (Overall lightness can be visually estimated by noting the contrast of the plot with the black lines of the blowup square.)
The terms appearing in these plots are exactly those included in a sum following summation convention (I) with $n_{max} = 400$. The dashed red line indicates the terms that would be included with convention (II) and $k_{max} = 400$.

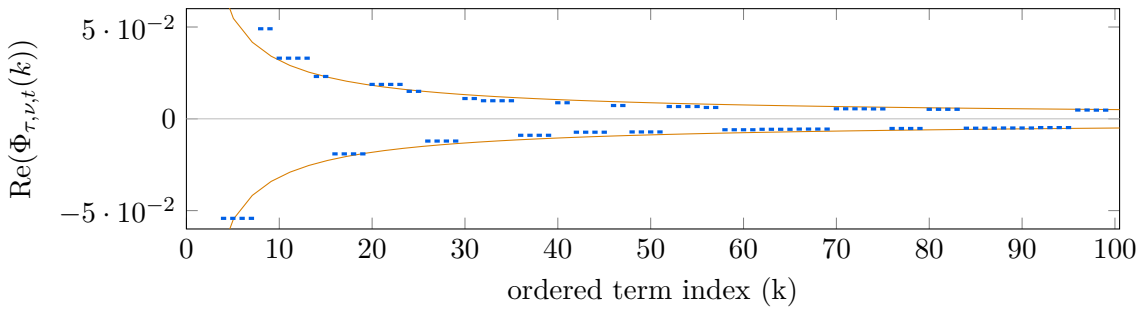


Figure 4.6 – Rearranged in decreasing order of $|\mathrm{Re}(\Phi_{\tau,\nu,t})|$, the terms of eq. (4.48) form a sequence roughly similar to an alternated p-series. Shown here is such a reordering for the terms with parameters $\tau = \nu = 0.6$ and $t = \infty$ where we've kept the 5^{th} to 100^{th} terms. (The first four terms fall beyond the figure's boundaries.) The solid lines are guides to the eye: they correspond to the functions $\pm 0.5\, k^{-0.8}$.

With some effort (see appendix B.3), we can show that asymptotically the absolute series follows

$$\sum_{n,m} |\Phi_{\tau,\nu,t}(n,m)| \sim \sum_{|n|,|m|<n'} |\Phi_{\tau,\nu,t}(n,m)| + C_{\tau,\nu} \sum_{k=n'}^{\infty} \frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,, \tag{4.51}$$

where $C_{\tau,\nu}$ is a real constants and n' is some finite positive integer. Thus it has the same convergence domain as the *p-series* $\sum k^{-p} = \sum k^{-2+(\tau+\nu)/\tau_{max}}$, which converges for $\tau + \nu < \tau_{max}$.[18] We note that the convergence domain has no dependence on the current time t as long as it's non-zero, although t *does* influence the number of terms we need to compute to reach a certain degree of accuracy (smaller values of t require more terms).

In itself, eq. (4.51) isn't terribly useful: what we really want to evaluate is eq. (4.48), in which terms *don't* appear in absolute value. However, it does strongly suggest that eq. (4.48) should be comparable to a *p*-series. And since about half of the terms $\Phi_{\tau,\nu,t}$ are negative (we see this in the particular case of fig. 4.5, and we strongly suspect it to be true in general), it seems reasonable to suspect that the eq. (4.48) is comparable to an *alternating* p-series where $p = 2 - (\tau + \nu)/\tau_{max}$. Indeed, when we reorder the terms in eq. (4.48) according to the magnitude of their real part, as in fig. 4.6, that seems to be what we observe.[19]

Alternating series will converge as long as the terms tend towards zero, thus within the

[18] For any $\beta > 0$, we can choose constants k_0 and c such that $\log k < ck^{\beta}$, $\forall k \geq k_0$. So the logarithm can be viewed as proportional to an arbitrarily small power as we increase k, and therefore $\sum \frac{\log k}{k^{\beta}}$ has the same convergence domain as $\sum \frac{1}{k^{\beta}}$.

[19] Visual inspection of power laws is notoriously error-prone, so the reader shouldn't place too much faith in the fact that we are able to match the terms with the "predicted" power of $p = 2-(\tau+\nu)/\tau_{max} = 0.8$ in fig. 4.6.

square domain defined by $0 \leq \tau \leq \tau_{max}$, $0 \leq \nu \leq \tau_{max}$ we expect that

$$\sum_{n,m} \Phi_{\tau,\nu,t}(n,m) \begin{cases} \text{converges absolutely if} & \tau + \nu < \tau_{max}\,; \\ \text{converges conditionally if} & \tau_{max} \leq \tau + \nu < 2\,\tau_{max}\,; \\ \text{and diverges if} & \tau = \nu = \tau_{max}\,. \end{cases} \tag{4.52}$$

In the next section we add empirical support to eq. (4.52) and give an idea of how many terms might actually be needed to compute the covariance.

4.4.3 Empirical convergence analysis

Here we test the convergence of eq. (4.48) numerically, comparing the performance of schemes (I) and (II). To avoid having to explore the three parameters τ, ν and t, we define a one-dimensional test function.

Given the interpretation of eq. (4.28) as an autocovariance, the "translated variance"

$$\Sigma_{\tau\tau}(t+\tau) = q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{(\lambda_n+\lambda_m)t} - e^{-(\lambda_n+\lambda_m)\tau}}{\lambda_n + \lambda_m} \tag{4.53}$$

should be equal to $\mathrm{Var}(X_0(t))$ for all τ in $I_{\mathrm{lag}} = [0, \tau]$. The fact that eq. (4.53) is constant as a function of τ makes it a convenient test equation for the convergence of eq. (4.48).[20] Of course, that also means that it doesn't tell us anything about the *co*variance, but we'll get to that afterwards.

We established in the previous section that the primary factor determining convergence was the time lag τ. It is therefore convenient to remove the current time t from our test equation by considering its limit as $t \to \infty$, which gives us the stationary translated variance

$$\Sigma^{\mathrm{s}}_{\tau\tau} \overset{\mathrm{def}}{=} \lim_{t\to\infty} \Sigma_{\tau\tau}(t+\tau) = -q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{-(\lambda_n+\lambda_m)\tau}}{\lambda_n + \lambda_m}\,. \tag{4.54}$$

(This is also the expression we studied in fig. 4.5.)

Figure 4.7b shows eq. (4.54) computed using truncation scheme (II). It is interesting in that it clearly exhibits a change in the convergence behaviour when the series goes from being absolutely to conditionally convergent. Still, the overall convergence is quite poor, especially when compared to the result of using scheme (I), shown in fig. 4.7a. The reason for this seems to be that the bounds of the summation domain of scheme (II), delimited by the dashed red line on fig. 4.5, align perfectly with the striped pattern formed by the term values seen on the same figure. Because of this alignment, the sum may include an entire positive (blue) stripe in the first or third quadrant without including a balancing negative

[20] If we were to consider, for example, the function $\Sigma_{\tau 0}(t)$, we would have to distinguish effects due to a change in the nature of convergence from those due to the autocovariance simply having a different value for different lags. Using eq. (4.53) helps isolate the former.

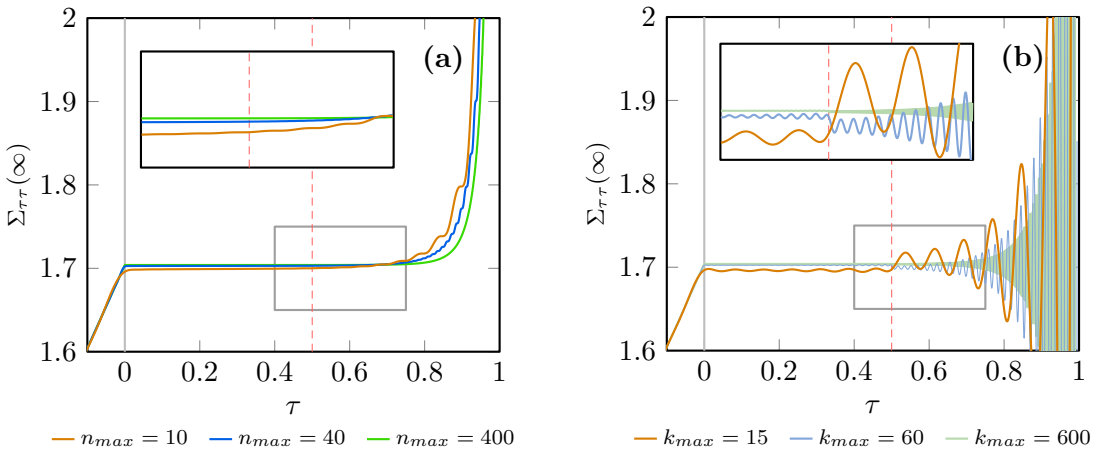


Figure 4.7 – Evaluation of the test series (4.54) at various time lags, using either truncation scheme (I) (fig. 4.7a) or (II) (fig. 4.7b). The series converges absolutely for τ between 0 and 0.5, the latter bound indicated by the dashed red line. The values of n_{max} and k_{max} were chosen to ensure curves of the same colour are computed using a comparable number of terms, with curves in (b) including a few more. Scheme (I) is seen to converge much more quickly and reliably, even when the series is only conditionally convergent. The value of the series veers away from $\mathrm{Var}(X_0(\infty))$ when $\tau < 0$ as it is no longer interpretable as a variance. Series was computed for kernel $\kappa(\tau) = -\delta(\tau - 1)$ and noise amplitude $q = 1$.

(red-orange) stripe. Scheme (I), by forming a 45° angle with the striped pattern in each quadrant, does a much better job of balancing positive and negative contributions.

It's important to note that the 45° stripe pattern observed in fig. 4.5 is entirely a consequence of us having computed $\Sigma_{\tau\nu}$ with $\tau = \nu$. In general we expect there to be some form of stripe pattern, but the angle it forms with the axes depends on the difference $\tau - \nu$. Consequently, the difference between both summation schemes won't always be so marked, and there may be as many "optimal" schemes as there are differences $\tau - \nu$.

> For a given number of terms, around 83% of those that would be included following scheme (I) would *also* be included following scheme (II). So relatively few terms induce the differences observed in fig. 4.7.

Remark 4.6

The takeaway from all this is that

- the choice of summation scheme strongly affects speed of convergence;
- graphically representing term values on a grid, such as we did in fig. 4.5, may help identify efficient summation schemes;
- based on fig. 4.7, for a close to optimal scheme, around $40^2 = 1600$ terms should give reasonable estimates.[21]

(The last statement is of course at best a rule of thumb, and will heavily depend on exactly how precise we need the estimates to be.)

[21] For *very* short times t, more terms might be needed.

In addition to providing a test of our convergence result, fig. 4.7 also drives home a fact stated on page 49, namely that the autocovariance series is strictly limited to lags between 0 and τ_{max}. That's why we see it veer away from $\mathrm{Var}(X_0(\infty))$ for negative τ.

——————— *** ———————

This is all well and good, but ultimately we want to compute autocovariances $\Sigma_{\tau\nu}$ with $\tau \neq \nu$. We know both lags must be within $I_{\mathrm{lag}} = [0, \tau_{max}]$. Also, following eq. (4.52), minimizing the sum $\tau + \nu$ should improve convergence. As a consequence, for practical calculations we really see no point in setting ν to any value other than zero. Considering again the stationary state, our equation of interest is therefore now the stationary autocovariance

$$\Sigma^{\mathrm{s}}_{\tau 0} = -q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{-\lambda_n \tau}}{\lambda_n + \lambda_m} \,. \tag{4.55}$$

Figure 4.8 compares eq. (4.55) to a Monte Carlo estimate of the autocovariance. We use scheme (I) to evaluate the series; this scheme isn't optimal over the entire range $[0, \tau_{max}]$, but we can easily compensate by including more terms. At precisely $\tau = \tau_{max}$ however, no amount of terms will do, as the series simply does not converge to the expected covariance. This isn't entirely unexpected, as we saw in fig. 4.3.

However, since we have good convergence for $\tau \in [0, \tau_{max})$, this is more of a technicality than a real problem. For example, we could simply approximate $\Sigma_{\tau_{max},0}$ by $\Sigma_{\tau_{max}-\epsilon,0}$ for some small ϵ, or compute a second point (perhaps $\Sigma_{\tau_{max}-2\epsilon,0}$) and linearly extrapolate if we need more precision. In brief, despite the fact that eq. (4.55) does not give the correct value for $\Sigma_{\tau_{max},0}$, with a bit of care it should be possible to evaluate it to any required degree of accuracy.

Ultimately, a better approach may actually be to use negative time lags. Probabilistically they are equivalent, since the autocovariance is symmetric ($\mathrm{Cov}(X_\tau, X_\nu) = \mathrm{Cov}(X_\nu, X_\tau)$), and fig. 4.8 clearly suggests that eq. (4.55) continues to be correct for $\tau < 0$. At the moment however, our formalism provides no grounds to conclude anything for negative lags (see remarks on page 49), and we've already seen in fig. 4.7 that eq. (4.28) *cannot* always be extended to negative lags. Nevertheless, if we could determine when this extension can be made, we would be in a position to break the restriction that time lags be limited to I_{lag} for the autocovariance.

——————— *** ———————

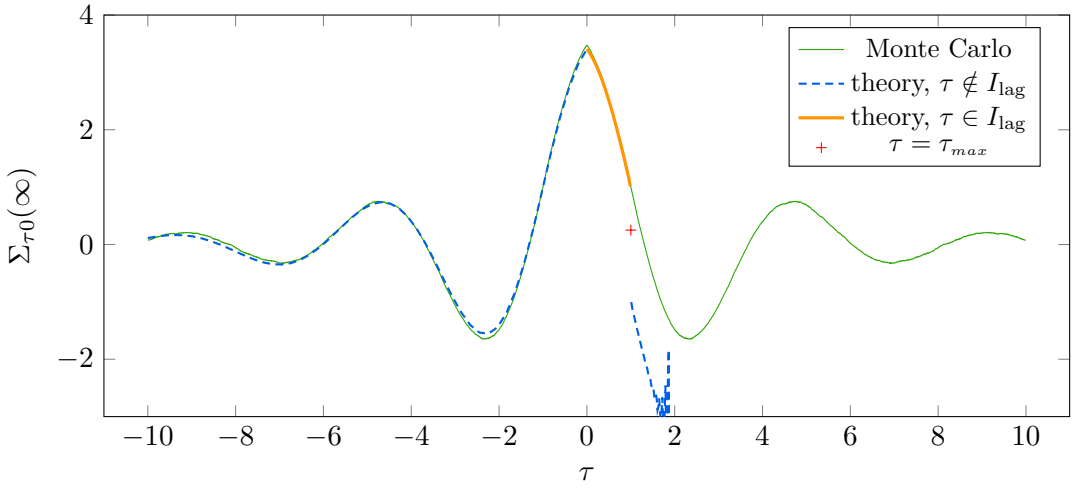


Figure 4.8 – Comparison of eq. (4.55) to a Monte Carlo estimate of the autocovariance, for the single delay process with $\kappa(\tau) = -\sqrt{2}\delta(\tau - 1)$. The latter is the averaged result of 10,000 runs, with values for $\tau > 0$ obtained by exploiting the autocovariance's symmetry. Equation (4.55) is expected to yield correct values for all lags τ within $I_{\text{lag}} \setminus \tau_{max}$. It indeed starkly deviates from the estimate for $\tau \geq 1$, but surprisingly both curves remain in agreement for all negative time lags.

We've shown that eq. (4.28) is an efficient method for computing the autocovariance of a delayed process, applicable to most[22] linear distributed-delay equations. We've identified elements for which care must be taken when evaluating the series in eq. (4.28) to ensure rapid convergence. Further work would almost certainly bring improvements in this regard; of particular interest would be extending the evaluation to arbitrarily large negative time lags. As it stands now however, this approach is already much more efficient than obtaining the same values using Monte Carlo, the only other method we know to be applicable to this class of SDDEs. In addition, it also provides some analytical intuition, something Monte Carlo estimates will never do.

4.5 Curious property of the Lambert function

In section 4.4.3, we justified eq. (4.53) by its interpretation as a variance. But as a mathematical object, it demonstrates a very curious property of the Lambert W function which we find deserves to be highlighted. To illustrate it, we consider a single delay kernel $\kappa(\tau) = -\delta(\tau - 1)$, for which the normalization constant is given by eq. (4.23), and look at

[22] In special cases feedback kernels may have degenerate eigenvalues, or only a finite number of them. We haven't addressed these possibilities at this time.

the stationary state. For $\tau = 0$ and $q = 1$, eq. (4.54) reduces to

$$\sum_{n,m} \frac{1}{(\lambda_n + \lambda_m)(1 + \lambda_n)(1 + \lambda_m)} \,. \tag{4.56}$$

Now, eq. (4.54) corresponds to the *same* physical quantity $\mathrm{Var}(X_0(\infty))$, independently of the chosen value of τ. Equalling eqs. (4.54) and (4.56), we conclude that

$$\sum_{n,m} \frac{e^{-(\lambda_n+\lambda_m)\tau}}{(\lambda_n + \lambda_m)(1 + \lambda_n)(1 + \lambda_m)} = \sum_{n,m} \frac{1}{(\lambda_n + \lambda_m)(1 + \lambda_n)(1 + \lambda_m)} \tag{4.57}$$

should hold for all $\tau \in \left[0, \frac{\tau_{max}}{2}\right]$, where $\lambda_n = W_n(-1)$. Based on fig. 4.7, this property might even extend to $\tau \in [0, \tau_{max})$.

A proper proof of eq. (4.57) which doesn't make the "physical" appeal to its interpretation as a variance is actually far from trivial. Therefore, in a strict mathematical sense, eq. (4.57) (and by extension eq. (4.53)) remains a conjecture.

4.6 An example: the uniform feedback kernel

In previous sections we always used single delay kernels to test our results. This of course limits the number of parameters and allows us to focus on essential features, but it may come off as disingenuous after claiming early on in this thesis that our results hold for most distributed delays. The analysis of our solution method also somewhat obscures its efficiency in actual applications. To amend these issues, we conclude this chapter by solving a true integro-differential equation.

We consider the SDDE process given by

$$\begin{aligned} dX_0 &= \beta \left(\int_0^1 X_\tau \, d\tau \right) dt \, + \, q \, dW \\ X(0) &= \phi \,, \end{aligned} \tag{4.58}$$

with initial condition

$$\phi_\tau = 1 \,, \qquad 0 \leq \tau \leq 1 \,.$$

The eigenvalues associated to eq. (4.58) must satisfy eq. (4.5), which in this case reduces to

$$\begin{aligned} \lambda_n &= \int_0^1 \beta \, e^{-\lambda_n \tau} \, d\tau \\ &\begin{cases} \lambda_n = \beta & \text{if } \lambda_n = 0 \\ \lambda_n^2 = \beta \left(1 - e^{-\lambda_n}\right) & \text{otherwise} \,. \end{cases} \end{aligned} \tag{4.59}$$

This allows us to conclude that no eigenvalue is zero since $\beta \neq 0$, but otherwise the lower equation needs to be solved numerically. This isn't entirely trivial, but it is also not that difficult to construct a reliable solver; see appendix C.1 for a brief discussion.

Once we have the eigenvalues, eigenvectors are immediately obtained as $u_{n,\tau} = e^{-\lambda_n \tau}$ (eq. (4.4)). The actions of the eigenfunctionals can be computed with eq. (4.17), and since none of the eigenvalues are zero, we find

$$\begin{aligned} u_n^\dagger(x) &= \mathcal{N}_n x_0 + \mathcal{N}_n \int_{0^+}^1 d\tau\, e^{\lambda_n \tau} x_\tau \int_\tau^1 d\nu\, \beta e^{-\lambda_n \nu} \\ &= \mathcal{N}_n x_0 - \mathcal{N}_n \frac{\beta}{\lambda_n} \int_{0^+}^1 d\tau\, e^{\lambda_n \tau} x_\tau \left(e^{-\lambda_n} - e^{-\lambda_n \tau} \right) \\ &= \mathcal{N}_n x_0 - \mathcal{N}_n \frac{\beta}{\lambda_n} \left(e^{-\lambda_n} \int_{0^+}^1 d\tau\, e^{\lambda_n \tau} x_\tau - \int_{0^+}^1 d\tau\, x_\tau \right) , \end{aligned} \tag{4.60}$$

where the normalization constant is given by eq. (4.16)

$$\begin{aligned} \mathcal{N}_n &= \left(1 + \beta \int_0^1 \tau e^{-\lambda_n \tau}\, d\tau \right)^{-1} \\ &= \left(1 + \beta \left[\frac{\tau e^{-\lambda_n \tau}}{-\lambda_n} - \frac{e^{-\lambda_n \tau}}{\lambda_n^2} \right]_0^1 \right)^{-1} \\ &= \left(1 - \frac{\beta}{\lambda_n^2} \left[(\lambda_n + 1) e^{-\lambda_n} - 1 \right] \right)^{-1} . \end{aligned}$$

Applying the eigenfunctionals to the initial condition, we get

$$\begin{aligned} c_n = u_n^\dagger(\phi) &= \mathcal{N}_n - \mathcal{N}_n \frac{\beta}{\lambda_n} \left(e^{-\lambda_n} \frac{e^{\lambda_n} - 1}{\lambda_n} - 1 \right) \\ &= \mathcal{N}_n \left(1 - \frac{\beta}{\lambda_n^2} (1 - e^{-\lambda_n} - \lambda_n) \right) , \end{aligned}$$

where

$$\phi = \sum_n c_n\, u_n \, .$$

Using eq. (4.27), the pointwise solution's mean is given by

$$\langle X_0(t) \rangle = \sum_{n=0}^{N-1} e^{\lambda_n t} c_n \, ,$$

while its variance is given by

$$\mathrm{Var}(X_0(t)) = q^2 \sum_{n,m} \mathcal{N}_n \mathcal{N}_m \frac{e^{(\lambda_n + \lambda_m)t} - 1}{\lambda_n + \lambda_m} \, ,$$

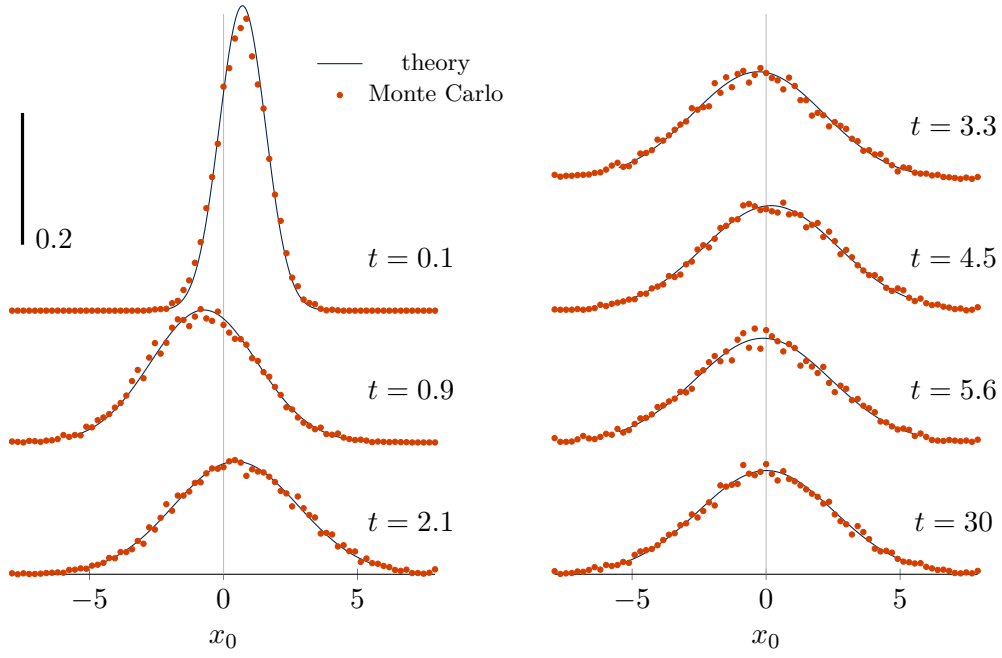


Figure 4.9 – Numerical verification of eq. (4.61) for the process described by eq. (4.58) with constant initial condition $\phi(\tau) = 1$. Parameters are $\beta = -3$, $\tau_{max} = 1$ and $q = 3$. The Monte Carlo estimates are aggregates of 10,000 runs.

thanks to eq. (4.29). The solution is then a time-dependent Gaussian,

$$X_0(t) \sim \mathcal{N}\left(\langle X_0(t)\rangle, \mathrm{Var}(X_0(t))\right), \qquad (t > 0) \tag{4.61}$$

which we show on fig. 4.9.

Chapter 5

Going forward

The results we obtained in this work are specific to the linear case eq. (4.1). This gives them somewhat limited appeal, as most models of real systems are *non*linear. In this final chapter, we discuss how one might leverage the ideas we've developed to explore novel solution and analysis methods applicable to a much wider class of equations. The proposed research directions are those we feel demand the most immediate attention in order to tackle the problems of interest to us, namely models of neurological networks. There are of course many more.

5.1 Multidimensional equations

Perhaps the most natural and obvious extension of the formalism of chapter 4 would be to consider multidimensional SDDEs. Since the equations remain linear, such an extension should pose no fundamental hurdle. By the same token, it also still wouldn't be applicable to nonlinear models, which in many applications are the norm rather than the exception. Nevertheless, most models *are* multivariate, so in most cases such an extension will be necessary, even if not sufficient.

There is no doubt that the notation would become heavier: for instance, while in the scalar case we obtained the autocovariance function as the limit of a matrix, in the multidimensional case the matrix would probably need to be replaced by a rank 4 tensor. Thus keeping the formalism manageable may not be trivial, but neither should it be impossible.

5.2 Kernels with special structure

In order to construct the biorthogonal set $\{u_n, u_n^\dagger\}$, we assumed that all eigenvalues λ_n were distinct (appendix A, page 76). We also require the eigenfunctions to be dense in $\mathcal{S}$, so that we can approximate arbitrary functions (see remark 2.2, page 12); in most cases, having an infinite number of eigenfunctions defined over a finite interval should ensure this. (An illustration of the reconstruction of an arbitrary state by eigenfunctions is given in fig. 4.3.) These are properties we expect to be true for a kernel chosen "at random", but they aren't true in general.

One case where these assumptions fail spectacularly is the gamma kernel, which is defined as [14, eq. 7.2]

$$g_a^p(\tau) \equiv \frac{a^p\,\tau^{p-1}\,e^{-a\tau}}{(p-1)!}\,, \qquad 0 \leq \tau < \infty\,,$$

where $a > 0$ and $p = 1, 2, \ldots$ are parameters. A DDE of the form of eq. (1.1) (with $q = 0$) is then defined with a kernel $\kappa = \alpha g_a^p$, with $\alpha \in \mathbb{R}$ a simple parameter giving the strength of the feedback. In this case not only do we have an infinite interval, but we also have only a *finite* number of eigenvalues. Indeed, the characteristic equation (4.5) in this case becomes (see [14, eq. 7.10]

$$\lambda_n\,(a + \lambda_n)^p - \alpha\,a^p = 0\,,$$

which has only $p + 1$ solutions. As a consequence, we can't decompose an arbitrary state into eigenfunctions because there is no way they can be dense in $\mathcal{S}$, and the formalism of chapter 4 falls apart.

It's not uncommon for properties such as degenerate eigenvalues to be connected to interesting physical phenomena,[1] so it may well be worth developing special techniques for treating these "pathological" kernels. Gamma kernels in particular are heavily used in models with distributed delays.[2]

5.3 A Fokker-Planck equation for SDDEs

Our primary motivation for this project was to develop a formalism allowing one to compute the p.d.f. for an SDDE directly, without resorting to Monte Carlo methods – the same way we can do so for an SDE with the Fokker-Planck equation. Thus we concludes this thesis by describing how the results of chapter 4 fit into this effort.

As was said in section 3.5, the remaining challenge in the development of a Fokker-Planck-like formalism for SDDEs is devising a method for evaluating the conditional average drift (eq. (3.9)),

$$\bar{f}(x_0, t) \equiv \int f(x_0, x_{\tau_1}) p_{X_{\tau_1}(t)|X_0(t)}(t, x_{\tau_1}|x_0)\,dx_{\tau_1}\,,$$

using as data only the collection of pointwise p.d.f.'s $\mathcal{P}_t^{\text{pt}} \equiv \{p_{X_\tau(t)} \mid 0 \leq \tau \leq \tau_{max}\}$. This would effectively "close the loop" in eq. (3.8) and allow one to compute p.d.f.'s for SDDEs directly and in deterministic time.

We envision that this could be done in two steps: First one would decompose the drift function f as a series over the moments of the random variable X. Then each of these moments could be determined from $\mathcal{P}^{\text{pt}}$, using the same kind of formalism as the one we developed in chapter 4.

[1] One might cite as examples the Berry phase [46] and certain phase transitions [47].

[2] This isn't a coincidence: their special eigenvalue structures is what permits the "linear chain trick"; see for example [14, section 7.1.2].

To see how our results fit into the second step, recall that the random variable X, whose realizations are elements of $\mathcal{S} = C([-\tau_{max}, 0], \mathbb{R})$, can be written as a sum of random amplitudes C_n over basis functions u_n (eq. (4.36)):

$$X = \sum_k C_k \, u_k \, .$$

The basis functions u_k don't necessarily need to be the eigenfunctions we used in section 4.3, although in many cases those are likely to be a good choice.

The expectation value $\langle X \rangle$ is a quantity that can be computed from $\mathcal{P}^{\mathrm{pt}}$. Furthermore, it is a function on $[-\tau_{max}, 0]$, i.e. an element of $\mathcal{S}$; therefore we can apply the eigenfunctionals $\langle X \rangle$ to get the mean amplitudes:

$$\begin{aligned} \langle X \rangle &= \sum_k \langle C_k \rangle \, u_k \\ u_n^\dagger(\langle X \rangle) &= \sum_k \langle C_k \rangle \, u_n^\dagger(u_k) \\ &= \langle C_n \rangle \, . \end{aligned}$$

The idea would be to define for each moment $\langle X^n \rangle$ an appropriate biorthogonal set. Since higher moments like $\langle X^2 \rangle$ contains all the joint moments $\langle C_n C_m \rangle$, we can extract enough information to determine the joint moments of X that we need to compute the CAD. There are still of course a number of challenges to overcome to get this to work, but exploratory work suggests that it is possible.

Chapter 6

Conclusion

Although differential equations involving either noise or delays have been studied for over half a century [6, 48], equations combining *both* have been the focus of surprisingly little efforts. In fact, it is the personal experience of the author that many researchers studying delayed equations balk at the prospect of introducing stochastics into their work, feeling the mathematical challenges are too great. That is not to say that there haven't been any efforts in this direction – we listed many in chapter 3 – but they are limited by rather stringent conditions on the original equations, such as considering only small delays or only very specific forms of SDDEs. The most potent approaches (sections 3.4 and 3.7), while being less restrictive, also require a much higher level of mathematical sophistication than standard solution methods for either DDEs or SDEs.

When we started this journey a little more than two years ago, we felt that there had to be another way, a way that would be powerful enough to handle general cases, but also remain palatable to the variety of researchers studying delayed or stochastic systems today. As a case study, we chose to consider a general SDE with distributed delays (eq. (1.1)), which to our knowledge was previously unsolved, and succeeded in reducing it to an eigenvalue problem. Such problems are among the most well known among scientists and engineers of all stripes.

With the eigenvalues in hand, one can immediately construct the solution from a set of eigenfunctions, following a common pattern for solving ODEs.[1] The principal challenge lies in assigning the correct weight to each eigenfunction. To do this we need the set of complementary eigenfunctionals, which in textbook cases is relatively easy to obtain thanks to the orthogonality of the eigenfunctions [15, § 8]. In the case of eq. (1.1), the eigenfunctions are *not* orthogonal, which forced us to come up with another way to find the eigenfunctionals. Amann et. al. proposed correct expressions for the specific case of a single delay DDE in 2007 (section 3.3), but simply posited their form with little rationale. For a general solution method this simply doesn't do: we need a systematic method that is not limited to specific forms of DDEs.

We solved this by constructing a robust bidirectional mapping between continuous linear DDEs and discretized approximations in the form of multivariate ODEs (see fig. 4.1). The bidirectionality was key: while others have used similar mappings to obtain *discretized*

[1] For example, any 2^{nd} order, nonlinear ODE that can be written as a Sturm-Liouville problem can be solved in this way [15, § 8].

eigenfunctions and -functionals [17, 19], our discretization is virtual: it does not appear in the final expressions, resulting in clean, continuous expressions. A bidirectional mapping between continuous and discretized forms is a very powerful tool: it allows for a natural transition from analytics to numerical simulations and back.

The same mapping can be applied to linear SDDEs, yielding approximate multivariate linear SDEs. Since complete time-dependent solutions to the latter are already known, we were able to solve the problem of our study case simply by mapping the multivariate solutions back to the original problem space (section 4.2). These solutions take the form of a Gaussian random variable, with independent series expressions for the mean and time-dependent autocovariance. Here again, while the discretization is an integral part of the derivation, it disappears completely in the final expressions.

Thus we devised a solution method that can be simply understood as adding the contributions of a dynamical system's various modes, which we feel should satisfy the requirement of palatability we initially set forth as a goal. However, while the solution *method* may be relatively simple to apply, ensuring its *validity* was far from trivial. In particular, in order to convince ourselves that our series expression for the time-dependent autocovariance function behaved as expected, we performed an extensive convergence analysis combining analytical and numerical results (section 4.4). This analysis also highlighted the corner cases where extra care would be warranted, and left a number of open avenues for future intrepid analysts to pursue, such as the relation proposed in section 4.5.

The work performed for this thesis was in some ways a proof of concept. We demonstrated that linear stochastic differential equations with distributed delays *can* be solved in a practical manner, and that the eigenfunction method we propose can make the high-dimensionality of the state space manageable (section 4.3). Ultimately of course, the aim is to go beyond linear equations. Of particular interest is the long-standing problem of developing an analogue to the Fokker-Planck equation, applicable to SDDEs. Given our results, we now have strong reason to believe that this is indeed possible (chapter 5).

We build mathematical models in order to explain and predict the behaviour of observable systems. It is therefore imperative that the models be determined by the systems themselves. However, despite the ubiquity of noise and delays in natural systems, relatively few mathematical models are based on SDDEs due to their intractability. By presenting a practical solution method for stochastic delayed equations, our work reduces this hurdle – allowing us physicists to view even more of the world as differential equations.

Appendix A

Diagonalizing the $\boldsymbol{L}^{(N)}$ operator

We wish to determine the left and right eigenvectors of

$$\boldsymbol{L}^{(N)} = \begin{pmatrix} \kappa(0)\frac{\tau_{max}}{N} & \kappa(\frac{\tau_{max}}{N})\frac{\tau_{max}}{N} & \kappa(\frac{2\tau_{max}}{N})\frac{\tau_{max}}{N} & \dots & \kappa(\frac{(N-2)\tau_{max}}{N})\frac{\tau_{max}}{N} & \kappa(\frac{(N-1)\tau_{max}}{N})\frac{\tau_{max}}{N} \\ N/\tau_{max} & -N/\tau_{max} & 0 & & & 0 \\ 0 & N/\tau_{max} & -N/\tau_{max} & & & \vdots \\ & & & \ddots & 0 & 0 \\ & & & N/\tau_{max} & -N/\tau_{max} & 0 \\ & & & & N/\tau_{max} & -N/\tau_{max} \end{pmatrix}, \tag{A.1}$$

which we do by solving $\lambda_n I - \boldsymbol{L}^{(N)} = 0$. Using the standard transformation rules between similar matrices, and assuming that $\lambda_n \neq -N/T$ (since we want to be able to vary the coarseness of the discretization, any issue arising only for particular values of N is irrelevant), we can put $\lambda_n I - \boldsymbol{L}^{(N)}$ in echelon form (to make them more legible, we define $T \equiv \tau_{max}$ for subsequent matrices):

$$\lambda_n I - \boldsymbol{L}^{(N)} = \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -\kappa(\frac{T}{N})\frac{T}{N} & -\kappa(\frac{2T}{N})\frac{T}{N} & \dots & -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & -\kappa(\frac{(N-1)T}{N})\frac{T}{N} \\ -N/T & \lambda_n + N/T & 0 & & & 0 \\ 0 & -N/T & \lambda_n + N/T & & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & & -N/T & \lambda_n + N/T & 0 \\ 0 & 0 & \dots & 0 & -N/T & \lambda_n + N/T \end{pmatrix}$$

$$\sim \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -\kappa(\frac{T}{N})\frac{T}{N} & -\kappa(\frac{2T}{N})\frac{T}{N} & \dots & -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & -\kappa(\frac{(N-1)T}{N})\frac{T}{N} \\ -1 & \frac{\lambda_n T}{N} + 1 & 0 & & & 0 \\ 0 & -1 & \frac{\lambda_n T}{N} + 1 & & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & & -1 & \frac{\lambda_n T}{N} + 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & \frac{\lambda_n T}{N} + 1 \end{pmatrix}.$$

Multiplying the second row by $(\frac{\lambda_n T}{N}+1)^{-1}$ and adding to the third, we get

$$\sim \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -\kappa(\frac{T}{N})\frac{T}{N} & -\kappa(\frac{2T}{N})\frac{T}{N} & \dots & -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & -\kappa(\frac{(N-1)T}{N})\frac{T}{N} \\ -1 & \frac{\lambda_n T}{N}+1 & 0 & & & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-1} & 0 & \frac{\lambda_n T}{N}+1 & & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ 0 & 0 & & -1 & \frac{\lambda_n T}{N}+1 & 0 \\ 0 & 0 & \dots & 0 & -1 & \frac{\lambda_n T}{N}+1 \end{pmatrix}.$$

Repeating this process all the way down ultimately yields

$$\sim \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -\kappa(\frac{T}{N})\frac{T}{N} & -\kappa(\frac{2T}{N})\frac{T}{N} & \dots & -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & -\kappa(\frac{(N-1)T}{N})\frac{T}{N} \\ -1 & \frac{\lambda_n T}{N}+1 & 0 & & & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-1} & 0 & \frac{\lambda_n T}{N}+1 & & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-(N-3)} & 0 & & 0 & \frac{\lambda_n T}{N}+1 & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-(N-2)} & 0 & \dots & 0 & 0 & \frac{\lambda_n T}{N}+1 \end{pmatrix}$$

$$\sim \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -\kappa(\frac{T}{N})\frac{T}{N} & -\kappa(\frac{2T}{N})\frac{T}{N} & \dots & -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & -\kappa(\frac{(N-1)T}{N})\frac{T}{N} \\ -(\frac{\lambda_n T}{N}+1)^{-1} & 1 & 0 & & & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-2} & 0 & 1 & & & \vdots \\ \vdots & & & \ddots & 0 & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-(N-2)} & 0 & & 0 & 1 & 0 \\ -(\frac{\lambda_n T}{N}+1)^{-(N-1)} & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Finally, we get

$$(\lambda_n I - \boldsymbol{L}^{(N)})\boldsymbol{u}_n^{(N)} \sim \begin{pmatrix} \lambda_n - \sum_{k=0}^{N-1}\kappa(\frac{kT}{N})(\frac{\lambda_n T}{N}+1)^{-k}\frac{T}{N} & 0 & 0 & \dots & \\ -(\frac{\lambda_n T}{N}+1)^{-1} & 1 & & & \\ \vdots & & & \ddots & \\ -(\frac{\lambda_n T}{N}+1)^{-(N-2)} & & & 1 & \\ -(\frac{\lambda_n T}{N}+1)^{-(N-1)} & \dots & & & 1 \end{pmatrix} \begin{pmatrix} u_{n,0}^{(N)} \\ u_{n,1}^{(N)} \\ \vdots \\ u_{n,N-2}^{(N)} \\ u_{n,N-1}^{(N)} \end{pmatrix} = 0\,. \tag{A.2}$$

In order for λ_n to be an eigenvalue, the rows of the matrix in eq. (A.2) must be linearly dependent. The only way to satisfy this condition is for the first row to contain only zeroes, which means that we write the characteristic equation of $\boldsymbol{L}^{(N)}$ as

$$\sum_{k=0}^{N-1}\kappa(\tfrac{k\,\tau_{max}}{N})(1+\tfrac{\lambda_n \tau_{max}}{N})^{-k}\left(\frac{T}{N}\right) = \lambda_n\,. \tag{A.3}$$

Once the first row is set to zero, the right eigenvectors can be read off eq. (A.2); they are given by

$$u_{n,i}^{(N)} = \left(1 + \frac{\lambda_n \tau_{max}}{N}\right)^{-i} . \qquad (0 \leq i \leq N-1) \quad \text{(A.4)}$$

To find the left eigenvectors, we need to solve

$$\boldsymbol{u}^{(N)\dagger}(\lambda_n I - \boldsymbol{L}^{(N)}) = (\lambda_n I - \boldsymbol{L}^{(N)})^\intercal (\boldsymbol{u}^{(N)\dagger})^\intercal = 0\,.$$

Proceeding in the same way as before, we have

$$(\lambda_n I - \boldsymbol{L}^{(N)})^\intercal$$

$$= \begin{pmatrix} \lambda_n - \kappa(0)\frac{T}{N} & -N/T & 0 & \dots & 0 & 0 \\ -\kappa(\frac{T}{N})\frac{T}{N} & \lambda_n + N/T & -N/T & & & \\ -\kappa(\frac{2T}{N})\frac{T}{N} & 0 & \lambda_n + N/T & & & \\ \vdots & \vdots & & \ddots & -N/T & \\ -\kappa(\frac{(N-2)T}{N})\frac{T}{N} & 0 & & & \lambda_n + N/T & -N/T \\ -\kappa(\frac{(N-1)T}{N})\frac{T}{N} & 0 & & & & \lambda_n + N/T \end{pmatrix}$$

$$\sim \begin{pmatrix} \frac{\lambda_n T}{N} - \kappa(0)\left(\frac{T}{N}\right)^2 & -1 & 0 & \dots & 0 & 0 \\ -\kappa(\frac{T}{N})\left(\frac{T}{N}\right)^2 & \frac{\lambda_n T}{N} + 1 & -1 & & & \\ -\kappa(\frac{2T}{N})\left(\frac{T}{N}\right)^2 & 0 & \frac{\lambda_n T}{N} + 1 & & & \\ \vdots & \vdots & & \ddots & -1 & \\ -\kappa(\frac{(N-2)T}{N})\left(\frac{T}{N}\right)^2 & 0 & & & \frac{\lambda_n T}{N} + 1 & -1 \\ -\kappa(\frac{(N-1)T}{N})\left(\frac{T}{N}\right)^2 & 0 & & & & \frac{\lambda_n T}{N} + 1 \end{pmatrix}$$

$$\sim \begin{pmatrix} \frac{\lambda_n T}{N} - \sum_{k=0}^{N-1} \kappa(\frac{kT}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-k}\left(\frac{T}{N}\right)^2 & 0 & \dots & 0 & 0 \\ -\sum_{k=1}^{N-1} \kappa(\frac{kT}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-k}\left(\frac{T}{N}\right)^2 & 1 & & & \\ -\sum_{k=2}^{N-1} \kappa(\frac{kT}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-k+1}\left(\frac{T}{N}\right)^2 & 0 & 1 & & \\ \vdots & \vdots & & \ddots & \\ -\kappa(\frac{(N-1)T}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-2}\left(\frac{T}{N}\right)^2 - \kappa(\frac{(N-2)T}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-1}\left(\frac{T}{N}\right)^2 & 0 & & 1 & 0 \\ -\kappa(\frac{(N-1)T}{N})\left(\frac{\lambda_n T}{N} + 1\right)^{-1}\left(\frac{T}{N}\right)^2 & 0 & & & 1 \end{pmatrix} .$$

We recognize in the first row the same characteristic equation as in eq. (A.3), as we would expect. Again, after we've set this row to zero, the eigenvectors can be read off. In this case we find

$$u_{n,i}^{(N)\dagger} = \begin{cases} \mathcal{N}_n^{(N)} & (i=0) \\ \mathcal{N}_n^{(N)} \sum_{k=i}^{N-1} \kappa\left(\frac{kT}{N}\right)\left(1+\frac{\lambda_n T}{N}\right)^{i-k-1}\left(\frac{T}{N}\right)^2, & (1\le i<N) \end{cases} \tag{A.5}$$

where $\mathcal{N}_n^{(N)}$ is an arbitrary constant.

As long as we don't have multiplicities in the eigenvalues, right and left eigenvectors of a matrix always form a biorthogonal set. Consequently, we must have

$$\boldsymbol{u}_n^{(N)\dagger}(\boldsymbol{u}_m^{(N)}) = 0$$

when $n \neq m$. This is arduous to verify by direct computation, but must be true since it's the only way to avoid a contradiction with

$$\lambda_n\, \boldsymbol{u}_n^{(N)\dagger}\, \boldsymbol{u}_m^{(N)} = (\boldsymbol{u}_n^{(N)\dagger}\, \boldsymbol{L}^{(N)})\, \boldsymbol{u}_m^{(N)} = \boldsymbol{u}_n^{(N)\dagger}\, (\boldsymbol{L}^{(N)}\, \boldsymbol{u}_m^{(N)}) = \lambda_m\, \boldsymbol{u}_n^{(N)\dagger}\, \boldsymbol{u}_m^{(N)}\,.$$

(Assuming none of the eigenvalues have multiplicity higher than 1.)

On the other hand, $\boldsymbol{u}_n^{(N)\dagger}(\boldsymbol{u}_n^{(N)})$ is of course not zero. In this case we get

$$\begin{aligned} \boldsymbol{u}_n^{(N)\dagger}(\boldsymbol{u}_n^{(N)}) &= \mathcal{N}_n^{(N)}\left[1+\sum_{i=1}^{N-1}\sum_{k=i}^{N-1}\kappa\left(\frac{kT}{N}\right)\left(1+\frac{\lambda_n T}{N}\right)^{-k-1}\left(\frac{T}{N}\right)^2\right] \\ &= \mathcal{N}_n^{(N)}\left[1+\sum_{k=i}^{N-1} k\cdot\kappa\left(\frac{kT}{N}\right)\left(1+\frac{\lambda_n T}{N}\right)^{-k-1}\left(\frac{T}{N}\right)^2\right]. \end{aligned}$$

Since it's a lot more convenient to work with a biortho*normal* system, we choose to set

$$\mathcal{N}_n^{(N)} \equiv \left[1+\sum_{k=i}^{N-1} k\cdot\kappa\left(\frac{kT}{N}\right)\left(1+\frac{\lambda_n T}{N}\right)^{-k-1}\left(\frac{T}{N}\right)^2\right]^{-1}. \tag{A.6}$$

This ensures that

$$\boldsymbol{u}_n^{(N)\dagger}(\boldsymbol{u}_n^{(N)}) = 1\,.$$

Appendix B

Covariance series

B.1 Asymptotic equivalence to the single delay memory kernel

What we want to show here is that the characteristic equation

$$\int_0^{\tau_{max}} \kappa(\tau) e^{-\lambda_n \tau}\, d\tau = \lambda_n \,, \tag{B.1}$$

for sufficiently large n and $\tau_{max} < \infty$, can be reasonably well approximated by the much simpler equation

$$b e^{-\lambda_n \tau_{max}} = \lambda_n \,. \tag{B.2}$$

Eigenvalues are ordered in the same way as in section 4.4.1, so large values of n correspond to eigenvalues with large negative real part.

Assumption that negative eigenvalues are arbitrarily large

> We know that if the kernel has only a single discrete delay, i.e. if $\kappa(\tau) = a\,\delta(\tau) + b\,\delta(\tau - \tau_{max})$, then the eigenvalues are distributed as in fig. 2.4: at most finitely many have positive real part, and negative real parts become arbitrarily large. Since delayed equations almost always have an infinite number of eigenvalues,[1] we make this assumption here. Kernels for which this is not the case require very particular structure; we choose to ignore these special cases.

Any reasonable kernel should be decomposable into a sum of a "comb-like" and a "continuous" kernel. The former is composed only of Dirac deltas,

$$\kappa_{\text{comb}}(\tau) = \sum_{i=0}^{M} b_i \delta(\tau - \tau_i) \,, \tag{B.3}$$

while the latter is a continuous bounded function on $[0, \tau_{max}]$. In this appendix we show the property (B.2) for a comb-like kernel. We suspect that it still holds for continuous kernels, if only by the fact that we can take a comb with a very large M, but leave a formal proof for a later time. (Note that the proof below is easily extended to mixed cases, as long as there is a Dirac delta at $\tau = \tau_{max}$.)

[1] An example of an exception would be gamma kernels, as seen in section 5.2.

Comb-like kernel For a comb-like kernel, eq. (B.1) becomes

$$\sum_{i=0}^{M} b_i \, e^{-\lambda_n \tau_i} = \lambda_n \, . \tag{B.4}$$

Let $\epsilon > 0$ be given and set $\tau_M = \tau_{max}$. Choose $i \in \mathbb{N}$, where $0 \leq i < M$. Since the real parts of λ_n take arbitrarily large negative values, and since $\tau_M - \tau_i$ is finite, there must exist an $n^* \in \mathbb{N}$ such that

$$\left| b_i e^{-\lambda_n \tau_i} \right| < \frac{\epsilon}{M} \left| b_M e^{-\lambda_n \tau_M} \right| , \quad \forall n \geq n^* .$$

Repeating this process for each $i < M$, we get

$$\sum_{i=0}^{M-1} \left| b_i \, e^{-\lambda_n \tau_i} \right| < \epsilon \left| b_M \, e^{-\lambda_n \tau_M} \right| , \quad \forall n \geq n^* \, .$$

Since ϵ can be chosen arbitrarily small, we see that

$$\sum_{i=0}^{M-1} \left| b_i \, e^{-\lambda_n \tau_i} \right| \approx \sum_{i=0}^{M-1} b_i \, e^{-\lambda_n \tau_i} \approx 0 \, , \quad \forall n \geq n^* \, ,$$

and therefore that

$$\begin{aligned} b_M \, e^{-\lambda_n \tau_M} &\approx b_M \, e^{-\lambda_n \tau_M} + \sum_{i=0}^{M-1} b_i \, e^{-\lambda_n \tau_i} \, , && \forall n \geq n^* \\ b_M \, e^{-\lambda_n \tau_M} &\approx \sum_{i=0}^{M} b_i \, e^{-\lambda_n \tau_i} \, , && \forall n \geq n^* \\ b_M \, e^{-\lambda_n \tau_{max}} &\approx \lambda_n \, , && \forall n \geq n^* \, , \end{aligned}$$

where the last expression is obtained by substituting eq. (B.4).

Hence eq. (B.2) is satisfied for large enough n.

——————— *** ———————

The purpose of this appendix was to justify our use of a single-delay kernel (B.2) for the asymptotic analysis in appendix B.3. We've made a number of assumptions on the kernel κ, both explicit and implicit, which – apart from the assumption that τ_{max} be finite – are generally consistent with the degree to which we would expect κ to be well-behaved. Nevertheless, we cannot completely exclude the possibility that in some very rare cases one would find a kernel κ that invalidates our assertion that eq. (B.1) is asymptotically equivalent to eq. (B.2).

B.2 Covariance series is always real

We certainly expect the covariance of a real process to be real, so the same should be true of its series expansion for given τ, ν and t. To see this, we set n' as the lowest positive index for which λ_n is complex. Then the series in eq. (4.48) becomes

$$\begin{aligned}\Sigma_{\tau\nu}(t) = &\sum_{|n|,|m|<n'} \Phi_{\tau,\nu,t}(n,m) + 2\sum_{\substack{|n|<n'\\ m\geq n'}} [\Phi_{\tau,\nu,t}(n,m) + \Phi_{\tau,\nu,t}(n,-m)] \\ &+ \sum_{n,m\geq n'} [\Phi_{\tau,\nu,t}(n,m) + \Phi_{\tau,\nu,t}(-n,-m) + \Phi_{\tau,\nu,t}(n,-m) + \Phi_{\tau,\nu,t}(-n,m)] \ . \quad \text{(B.5)}\end{aligned}$$

The 2 in front of the second sum appears because we make use of the reflexivity property

$$\Phi_{\tau,\nu,t}(n,m) = \Phi_{\tau,\nu,t}(m,n)\,.$$

The terms of the first sum of eq. (B.5) are real, so we don't need to worry about those. In the second sum, we have

$$\begin{aligned}\Phi_{\tau,\nu,t}(n,-m) &= q^2\mathcal{N}_n\mathcal{N}_{-m}\frac{e^{(\lambda_n+\lambda_{-m})t}-1}{\lambda_n+\lambda_{-m}}e^{-\lambda_n\tau}e^{-\lambda_{-m}\nu} \\ &= q^2\mathcal{N}_n\mathcal{N}_m^*\frac{e^{(\lambda_n+\lambda_m^*)t}-1}{\lambda_n+\lambda_m^*}e^{-\lambda_n\tau}e^{-\lambda_m^*\nu} \\ &= q^2[\mathcal{N}_n^*\mathcal{N}_m]^*\frac{\left[e^{(\lambda_n^*+\lambda_m)t}-1\right]^*}{[\lambda_n^*+\lambda_m]^*}\left[e^{-\lambda_n^*\tau}e^{-\lambda_m\nu}\right]^* \\ &= \Phi_{\tau,\nu,t}(-n,m)^*\,.\end{aligned}$$

We've used here the fact that for $n \geq n'$,

$$\mathcal{N}_{-n} = 1\Big/1 + \int_{0^+}^{\tau_{max}} \kappa(\tau)\tau e^{-\lambda_n^*\tau}d\tau = \mathcal{N}_n^*\,.$$

Thus

$$\Phi_{\tau,\nu,t}(n,m) + \Phi_{\tau,\nu,t}(n,-m) = 2\,\mathrm{Re}\left(\Phi_{\tau,\nu,t}(n,m)\right)\,, \qquad (|n|<n', m>n')$$

and that sum is also real.

For the third sum we can similarly show that the imaginary parts cancel since

$$\Phi_{\tau,\nu,t}(-n,-m) = \Phi_{\tau,\nu,t}(n,m)^* \qquad \text{and} \qquad \Phi_{\tau,\nu,t}(n,-m) = \Phi_{\tau,\nu,t}(-n,m)^*\,, \qquad \text{(B.6)}$$

making the entire sum real.

B.3 Asymptotic behaviour of the covariance series

In section 4.4.2, we claim that (eq. (4.51))

$$\sum_{n,m} |\Phi_{\tau,\nu,t}(n,m)| \sim \sum_{|n|,|m|<n'} |\Phi_{\tau,\nu,t}(n,m)| + C_{\tau,\nu} \sum_{k=n'}^{\infty} \frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,, \tag{B.7}$$

where $C_{\tau,\nu}$ is a real constant to be determined and n' a finite positive integer. To show this, we will make heavy use of the large n asymptotic expression for (4.45) for the eigenvalues,

$$\lambda_n \sim -\frac{1}{\tau_{max}} \log\left|\frac{2\pi(n-1)}{b\,\tau_{max}}\right| + \frac{i}{\tau_{max}}\left(2\pi(n-1) - \operatorname{sgn}(b)\frac{\pi}{2}\right)\,, \tag{B.8}$$

which provides a reasonable approximation for $n \geq 2$. The terms of the series are given by eq. (4.47), so in absolute value they are

$$|\Phi_{\tau,\nu,t}(n,m)| \sim q^2 \frac{\left|e^{(\lambda_n+\lambda_m)t} - 1\right|}{|\lambda_n+\lambda_m|\,|1+\lambda_n\tau_{max}|\,|1+\lambda_m\tau_{max}|} \left|e^{-\lambda_n\tau}\right| \left|e^{-\lambda_m\nu}\right|\,. \tag{B.9}$$

We start by separating the terms where either one or both eigenvalues are real. We define a cutoff $n' \geq 2$ such that any λ_n is complex if $n \geq n'$. (This is always possible since there can be at most a finite number of real eigenvalues, as per remark 4.4. In the case of eigenvalues given by the Lambert W function, there are in fact at most two real eigenvalues.) The condition that n' be at least 2 is to exclude the cases $n = \pm 1$, where eq. (B.8) is undefined. Taking the magnitudes of the terms in eq. (B.5), we can split the series as follows

$$\begin{aligned}
\sum_{n,m} |\Phi_{\tau,\nu,t}(n,m)| = & \sum_{|n|,|m|<n'} |\Phi_{\tau,\nu,t}(n,m)| \\
& + 2 \sum_{|n|<n'} \sum_{|m|\geq n'} |\Phi_{\tau,\nu,t}(n,m)| \\
& + \sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(n,m)| \quad + \sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(-n,-m)| \\
& + \sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(n,-m)| \quad + \sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(-n,m)|\,;
\end{aligned} \tag{B.10}$$

the factor 2 on the second line again arises due to the symmetry of n and m we'll work through this expression line by line.

The sum on the first line of eq. (B.10) gives us the first sum in eq. (B.7) and requires no further analysis.

On the second line, only the inner sum has an infinite number of terms. This means that λ_n can simply be treated as a parameter. Making use of eq. (B.8), and of the fact that

$$m \sim m - 1$$

for large m, we see that the terms of eq. (B.9) become

$$|1 + \lambda_m \tau_{max}| \sim \left| 1 - \log \frac{2\pi\,|m|}{b\,\tau_{max}} + i\left(2\pi(m-1) - \mathrm{sgn}(b)\frac{\pi}{2}\right)\right| \sim 2\pi\,|m| \tag{B.11a}$$

$$|\lambda_n + \lambda_m| \sim |\lambda_m| \sim \frac{2\pi\,|m|}{\tau_{max}} \tag{B.11b}$$

$$\left|e^{-\lambda_m \nu}\right| \sim \exp\left[\frac{\nu}{\tau_{max}} \log\left|\frac{2\pi(m-1)}{b\,\tau_{max}}\right|\right] \sim \left[\frac{2\pi\,|m|}{b\,\tau_{max}}\right]^{\nu/\tau_{max}} \tag{B.11c}$$

$$\left|e^{(\lambda_n+\lambda_m)t} - 1\right| \approx 1\,. \tag{B.11d}$$

Putting these together, we get

$$\begin{aligned}\sum_{|m|\geq n'} |\Phi_{\tau,\nu,t}(n,m)| &\sim \sum_{|m|\geq n'} \left|\frac{e^{-\lambda_n \tau}}{1+\lambda_n\tau_{max}}\right| q^2\, b^{-\nu/\tau_{max}} \tau_{max}^{1-\nu/\tau_{max}} (2\pi)^{-2+\nu/\tau_{max}}\ \frac{1}{|m|^{2-\nu/\tau_{max}}} \\ &\sim C^{\mathrm{I}}_{\tau,\nu,n} \sum_{m\geq n'} \frac{1}{m^{2-\nu/\tau_{max}}}\,,\end{aligned} \tag{B.12}$$

where $C^{\mathrm{I}}_{\tau,\nu,n}$ is some constant. This well-known series – corresponding to the Riemann function $\zeta(s)$ with real s – converges for $\nu < \tau_{max}$.

——————— *** ———————

To treat the double sums of the third and fourth lines of eq. (B.10), we need to set a summation order. We choose to collect the terms for which $n + m = k$ for some k, following scheme (II) of section 4.4.2. Since the first line of eq. (B.10) sums all terms where $n, m < n'$, the smallest possible value for k is $n' + n' = 2n'$. Consequently the third line can be written

$$\sum_{k=2n'}^{\infty} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(n, k-n)| + \sum_{k=2n'}^{\infty} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(-n, -k+n)|\,.$$

We can find an asymptotic expression for the entire inner sum, and treat that as a single term in the series to determine its convergence domain. Both double sums are analogous, so we only need to treat one; we'll do the one on the left.

We want to determine the terms of eq. (B.9) for $n' \ll k$ and $n' \leq n \leq k - n'$. Since our choice for n' was arbitrary (it only needs to be finite), we'll additionally make the assumption that $1 \ll n'$. This leads us to

$$\begin{aligned} |\lambda_n + \lambda_{k-n}| &\sim \left| -\frac{1}{\tau_{max}} \log\left[\left(\frac{2\pi\, n}{b\,\tau_{max}}\right)\left(\frac{2\pi\,(k-n)}{b\,\tau_{max}}\right)\right] + \frac{i}{\tau_{max}}\left(2\pi\, n + 2\pi\,(k-n)\right)\right| \\ &\sim \frac{2\pi\, k}{\tau_{max}} \end{aligned} \tag{B.13a}$$

$$|1 + \lambda_n \tau_{max}| \sim 2\pi n \tag{B.13b}$$

$$|1 + \lambda_{k-n} \tau_{max}| \sim 2\pi(k-n) \tag{B.13c}$$

$$\left|e^{-\lambda_n \tau}\right| \sim \left[\frac{2\pi(n-1)}{b\,\tau_{max}}\right]^{\tau/\tau_{max}} \sim \left[\frac{2\pi\, n}{b\,\tau_{max}}\right]^{\tau/\tau_{max}} \tag{B.13d}$$

$$\left|e^{-\lambda_{k-n} \nu}\right| \sim \left[\frac{2\pi(k-n-1)}{b\,\tau_{max}}\right]^{\nu/\tau_{max}} \sim \left[\frac{2\pi(k-n)}{b\,\tau_{max}}\right]^{\nu/\tau_{max}} \tag{B.13e}$$

$$\left|e^{(\lambda_n + \lambda_{k-n})t} - 1\right| \approx 1\,. \tag{B.13f}$$

Substituting these back into eq. (B.9), we get

$$\begin{aligned} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(n,m)| &\sim \sum_{n=n'}^{k-n'} \frac{\tau_{max}}{(2\pi)^3}\left(\frac{2\pi}{b\tau_{max}}\right)^{\tau+\nu/\tau_{max}} k^{-1}\, n^{-1+\tau/\tau_{max}}\,(k-n)^{-1+\nu/\tau_{max}} \\ &\sim C^{\mathrm{II}}_{\tau,\nu} \sum_{n=n'}^{k-n'} n^{-1+\tau/\tau_{max}}\,(k-n)^{-1+\nu/\tau_{max}} \frac{1}{k}\,. \end{aligned}$$

If we define $x \equiv n/k$, then we can write this as

$$\sim C^{\mathrm{II}}_{\tau,\nu} k^{-2+(\tau+\nu)/\tau_{max}} \sum_{n=n'}^{k-n'} x^{-1+\tau/\tau_{max}}\,(1-x)^{-1+\nu/\tau_{max}} \frac{1}{k}\,,$$

where we recognize here a Riemann sum. In the limit of large k we can replace it by an integral, giving us

$$\sim C^{\mathrm{II}}_{\tau,\nu} k^{-2+(\tau+\nu)/\tau_{max}} \int_{\frac{n'}{k}}^{1-\frac{n'}{k}} x^{-1+\tau/\tau_{max}}\,(1-x)^{-1+\nu/\tau_{max}}\, dx \tag{B.14}$$

$$\sim C^{\mathrm{II}}_{\tau,\nu} k^{-2+(\tau+\nu)/\tau_{max}} \int_{0}^{1} x^{-1+\tau/\tau_{max}}\,(1-x)^{-1+\nu/\tau_{max}}\, dx\,. \tag{B.15}$$

Here we recognize the integral as the Beta function $B(\frac{\tau}{\tau_{max}}, \frac{\nu}{\tau_{max}})$, which is finite for all strictly positive τ and ν. (The case where either τ or ν is zero still converges, but needs

to be treated separately.) The result is that the sums on the third line of eq. (B.10) can be written as

$$\sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(n,m)| \sim C^{\text{II}}_{\tau,\nu} B\left(\tfrac{\tau}{\tau_{max}}, \tfrac{\nu}{\tau_{max}}\right) \sum_{k=n'}^{\infty} \frac{1}{k^{2-(\tau+\nu)/\tau_{max}}}\,, \tag{B.16a}$$

and therefore will converge for $\tau + \nu < \tau_{max}$.

When either τ or ν is zero, the beta function diverges because of the poles at $x = 0$ and $x = 1$ of the integrand. But since we always keep a distance $\frac{n'}{k}$ from each pole, we can avoid this inconvenient blow up – we just have to work a little harder. For the sake of argument, let's assume that $\tau = 0$.

Let $0 < \epsilon \ll 1$. For $k > \frac{n'}{\epsilon}$, we split the integral in eq. (B.14) as

$$\int_{\frac{n'}{k}}^{1-\frac{n'}{k}} x^{-1}\,(1-x)^{-1+\nu/\tau_{max}}\,dx = \int_{\frac{n'}{k}}^{\epsilon} x^{-1}\,(1-x)^{-1+\nu/\tau_{max}}\,dx + \int_{\epsilon}^{1-\frac{n'}{k}} x^{-1}\,(1-x)^{-1+\nu/\tau_{max}}\,dx\,.$$

The integral on the left is dominated by the x^{-1} term. The one on the right, which we'll label $\mathcal{B}_\nu$, is of a finite integrand over a finite domain, so whatever it is equal to, it must be finite. Thus

$$\begin{aligned}
\int_{\frac{n'}{k}}^{1-\frac{n'}{k}} x^{-1}\,(1-x)^{-1+\nu/\tau_{max}}\,dx &\approx \int_{\frac{n'}{k}}^{\epsilon} x^{-1}dx + \mathcal{B}_\nu \\
&\approx \log x\big|_{\frac{n'}{k}}^{\epsilon} + \mathcal{B}_\nu \\
&\approx \log k + \log\frac{\epsilon}{n'} + \mathcal{B}_\nu \\
&\sim \log k\,,
\end{aligned}$$

and eq. (B.16a) becomes

$$\sum_{n\geq n',m\geq n'} |\Phi_{\tau,\nu,t}(n,m)| \sim C^{\text{II}}_{\tau,\nu}\mathcal{B}_\nu \sum_{k=n'}^{\infty} \frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,. \tag{B.16b}$$

If it is ν that is equal to zero, we instead cut off the upper end of the integral by an amount ϵ, and if both τ and ν are zero, than we just cut off both ends. In all these cases, we get a series that is asymptotically proportional to

$$\sum_{k=n'}^{\infty} \frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,,$$

which again converges for $\tau + \nu < \tau_{max}$. (See the footnote on page 60.)

——————— *** ———————

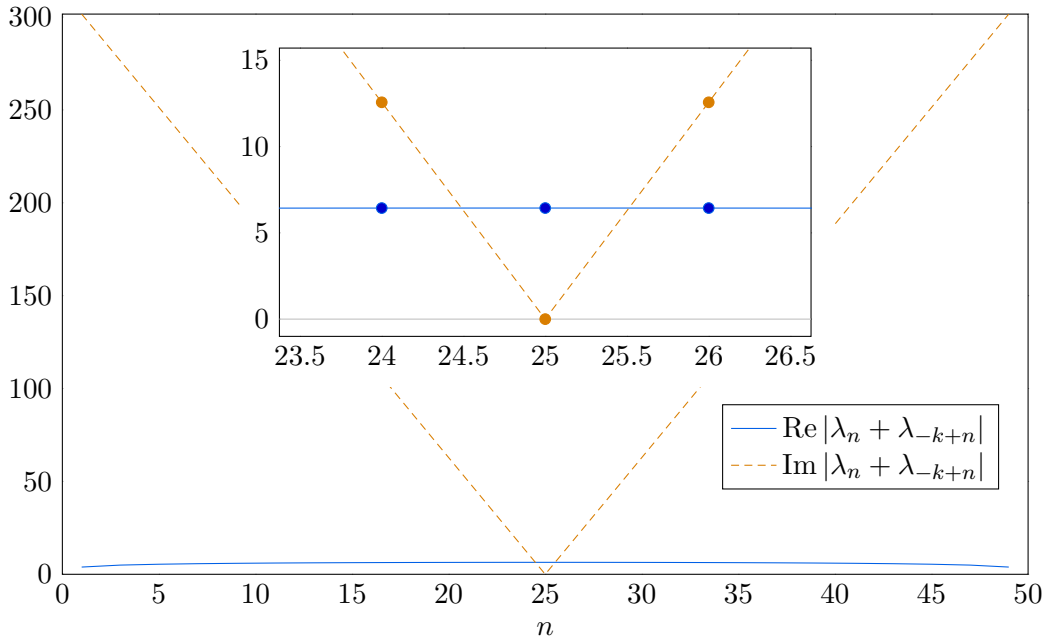


Figure B.1 – Plot of the real and imaginary parts of eq. (B.17) when $k = 50$. The inset is a blow-up of the region around $n = 25$. In this case n_{b-} and n_{b+} are 24 and 26 respectively; plots with very large values of k would show a logarithmic increase in the distance of the intersection of both curves.

The fourth and final line of eq. (B.10) is treated in a similar fashion as the third. We again define $k = n + m$ and collect the terms with same k such that the line is rewritten

$$\sum_{k=2n'}^{\infty} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(n, -k+n)| + \sum_{k=2n'}^{\infty} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(-n, k-n)| \ .$$

Both double sums are again analogous, and we will consider only the one on the left. Most terms appearing in $\Phi_{\tau,\nu,t}(n, -k+n)$ have the same asymptotic expressions, namely eqs. (B.13b) to (B.13f). The main difference is $|\lambda_n + \lambda_{-k+n}|$ – the term previously described by eq. (B.13a) – whose asymptotic limit is a little more subtle.

Indeed, now we have

$$\begin{aligned} |\lambda_n + \lambda_{-k+n}| &\sim \left| -\frac{1}{\tau_{max}} \log \left[\left(\frac{2\pi\, n}{b\, \tau_{max}} \right) \left(\frac{2\pi\, (k-n)}{b\, \tau_{max}} \right) \right] + \frac{i}{\tau_{max}} \left(2\pi\, n + 2\pi\, (-k+n) \right) \right| \ . \\ &\sim \left| -\frac{1}{\tau_{max}} \log \left[n(k-n) \right] + \frac{2\pi\, i}{\tau_{max}} (2n-k) \right| \ , \end{aligned} \tag{B.17}$$

and while previously the imaginary part was always $\frac{2\pi\, k}{\tau_{max}}$ and could easily dominate the real part, in this case there is a range of values for n – when n is close to its central value

of $\frac{k}{2}$ – where the imaginary part is almost zero (see fig. B.1). In this central range the real part dominates. To identify its bounds, we solve

$$\begin{aligned} \operatorname{Re}|\lambda_{n_b} + \lambda_{-k+n_b}| &= \operatorname{Im}|\lambda_{n_b} + \lambda_{-k+n_b}| \\ \log[n_b(k-n_b)] &\approx 2\pi\ |2n_b - k|\ . \end{aligned} \tag{B.18}$$

Within our central region of interest, $\log[n_b(k-n_b)]$ is practically flat (and ever more so as we increase k), so we can approximate it by its central value:

$$\begin{aligned} \log[n_b(k-n_b)] &\approx \log\left[\frac{k}{2}\left(k-\frac{k}{2}\right)\right] = \log\left(\frac{k^2}{4}\right) \\ &\sim 2\log\frac{k}{2}\,. \end{aligned}$$

Inserting this back into eq. (B.18), this gives us the bounds of the central, real-part dominated range as

$$n_{b-} \equiv \left\lfloor \frac{k}{2} - \frac{\log k/2}{2\pi} \right\rfloor\,, \qquad\qquad n_{b+} \equiv \left\lceil \frac{k}{2} + \frac{\log k/2}{2\pi} \right\rceil\,.$$

(The floor $\lfloor\cdot\rfloor$ and ceiling $\lceil\cdot\rceil$ functions are just to ensure that n_{b-} and n_{b+} are whole numbers at equal distance from $\frac{k}{2}$.)

Since we only need a first order approximation to get the asymptotic behaviour, we can make the crude estimate

$$|\lambda_n + \lambda_{-k+n}| \sim \begin{cases} 2\pi(k-2n)/\tau_{max} & \text{if } n \le n_{b-} \\ 2\log(\frac{k}{2})/\tau_{max} & \text{if } n_{b-} \le n \le n_{b+} \\ 2\pi(2n-k)/\tau_{max} & \text{if } n_{b+} \le n, \end{cases} \tag{B.19}$$

and split the inner sum into three parts:

$$\begin{aligned} \sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(n,-k+n)| = \sum_{n=n'}^{n_{b-}} |\Phi_{\tau,\nu,t}(n,-k+n)| + \sum_{n=n_{b-}}^{n_{b+}} |\Phi_{\tau,\nu,t}(n,-k+n)| \\ + \sum_{n=n_{b+}}^{k-n'} |\Phi_{\tau,\nu,t}(n,-k+n)| \end{aligned} \tag{B.20}$$

which we can evaluate separately.

Substituting eqs. (B.13b) to (B.13f) and (B.19) into eq. (B.9), we find for the first part

$$\sum_{n=n'}^{n_{b-}} |\Phi_{\tau,\nu,t}(n,-k+n)| \sim \frac{\tau_{max}}{(2\pi)^3}\left(\frac{2\pi}{b\tau_{max}}\right)^{\tau+\nu/\tau_{max}} \sum_{n=n'}^{n_{b-}} \frac{1}{(k-2n)\,n^{1-\tau/\tau_{max}}\,(k-n)^{1-\nu/\tau_{max}}}\,.$$

Substituting again $x \equiv n/k$ to pass to an integral form, this becomes

$$\begin{aligned}
&\sim C^{\mathrm{III}}_{\tau,\nu} \frac{1}{k^{2-(\tau+\nu)/\tau_{max}}} \sum_{n=n'}^{n_{b-}} \frac{1}{(1-2x)\, x^{1-\tau/\tau_{max}}\, (1-x)^{1-\nu/\tau_{max}}} \frac{1}{k} \\
&\sim C^{\mathrm{III}}_{\tau,\nu} \frac{1}{k^{2-(\tau+\nu)/\tau_{max}}} \int_{\frac{n'}{k}}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} \underbrace{\frac{1}{(1-2x)\, x^{1-\tau/\tau_{max}}\, (1-x)^{1-\nu/\tau_{max}}}}_{\equiv h_{\tau,\nu}(x)} \, dx \\
&\sim C^{\mathrm{III}}_{\tau,\nu} \frac{1}{k^{2-(\tau+\nu)/\tau_{max}}} \int_{\frac{n'}{k}}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} h_{\tau,\nu}(x)\, dx \,,
\end{aligned}$$

where we've defined $h_{\tau,\nu}(x)$ to be the function appearing in the integrand.

Using the same cutoff trick as before, we let $\frac{n'}{k} < \frac{\log k/2}{2\pi k} < \epsilon \ll \frac{1}{2}$ and split this integral into

$$\int_{\frac{n'}{k}}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} h_{\tau,\nu}(x) \sim \int_{\frac{n'}{k}}^{\epsilon} h_{\tau,\nu}(x)\, dx + \frac{1}{k^2} \underbrace{\int_{\epsilon}^{\frac{1}{2}-\epsilon} h_{\tau,\nu}(x)}_{\text{finite constant}} \, dx + \int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} h_{\tau,\nu}(x)\, dx \,. \tag{B.21}$$

Similarly to how we obtained eq. (B.16), the first integral either tends towards a finite limit or grows as $\log k$:

$$\begin{cases}
\int_{\frac{n'}{k}}^{\epsilon} h_{\tau,\nu}(x)\, dx \sim \log k & \text{if } \tau = 0 \\
\int_{\frac{n'}{k}}^{\epsilon} h_{\tau,\nu}(x)\, dx < \frac{\tau_{max}}{\tau} \epsilon^{\tau/\tau_{max}} \approx 0 & \text{if } \tau > 0
\end{cases}$$

The second integral is constant (for some set ϵ) and can be absorbed into $C^{\mathrm{III}}_{\tau,\nu}$.

Within the range of the third integral ($\frac{1}{2} - \epsilon \le x \le \frac{1}{2} - \frac{\log k/2}{2\pi k}$), $h_{\tau,\nu}$ is dominated by $(1-2x)^{-1}$; its other terms can be approximated by

$$x^{-1+\tau/\tau_{max}}\, (1-x)^{-1+\nu/\tau_{max}} \approx \left(\frac{1}{2}\right)^{-1+\tau/\tau_{max}} \left(1-\frac{1}{2}\right)^{-1+\nu/\tau_{max}} = 2^{2-(\tau+\nu)/\tau_{max}} \,.$$

As a consequence, we have

$$\begin{aligned}
\int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} h_{\tau,\nu}(x) &\sim 2^{2-(\tau+\nu)/\tau_{max}} \int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}-\frac{\log k/2}{2\pi k}} \frac{dx}{1-2x} \\
&\sim 2^{1-(\tau+\nu)/\tau_{max}} \log k
\end{aligned}$$

Reassembling the terms in eq. (B.21) and aggregating all the constants, we finally have

$$\sum_{n=n'}^{n_{b-}} |\Phi_{\tau,\nu,t}(n, -k+n)| \sim C^{\mathrm{IV}}_{\tau,\nu} \frac{\log k}{k^{2+(\tau+\nu)/\tau_{max}}} \tag{B.22}$$

That takes care of the first term in eq. (B.20). The third term is exactly analogous, so all we have left to consider is the second term in that equation. In this case we have

$$\sum_{n=n_{b-}}^{n_{b+}} |\Phi_{\tau,\nu,t}(n,-k+n)| \sim \frac{\tau_{max}}{8\pi^2}\left(\frac{2\pi}{b\tau_{max}}\right)^{\tau+\nu/\tau_{max}} \frac{1}{\log k/2} \sum_{n=n_{b-}}^{n_{b+}} \frac{1}{n^{1-\tau/\tau_{max}}\,(k-n)^{1-\nu/\tau_{max}}}$$
$$\sim C^{\mathrm{V}}_{\tau,\nu}\,\frac{k^{-1+(\tau+\nu)/\tau_{max}}}{\log k} \int_{\frac{1}{2}-\frac{\log k/2}{2\pi k}}^{\frac{1}{2}+\frac{\log k/2}{2\pi k}} x^{-1+\tau/\tau_{max}}\,(1-x)^{-1+\nu/\tau_{max}}\,dx\,.$$

Over the range of the integral, the integrand is practically constant and can be approximated by $2^{2-(\tau+\nu)/\tau_{max}}$; pulling it out of the integral, we get

$$\sum_{n=n_{b-}}^{n_{b+}} |\Phi_{\tau,\nu,t}(n,-k+n)| \sim C^{\mathrm{V}}_{\tau,\nu}\,\frac{2^{2-(\tau+\nu)/\tau_{max}}}{\pi} k^{-2+(\tau+\nu)/\tau_{max}}\,. \tag{B.23}$$

Inserting eqs. (B.22) and (B.23) into eq. (B.20), we keep only the terms proportional to $\log k/k^{2-(\tau+\nu)/\tau_{max}}$ and get the asymptotic expression for the last line of eq. (B.10),

$$\sum_{k=2n'}^{\infty}\sum_{n=n'}^{k-n'} |\Phi_{\tau,\nu,t}(n,-k+n)| \sim 2C^{\mathrm{VI}}_{\tau,\nu}\,\frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,, \tag{B.24}$$

which, again, converges when $\tau+\nu<\tau_{max}$.

——————— *** ———————

Reassembling eq. (B.10) from eqs. (B.12), (B.16) and (B.24) and combining the constants into $C_{\tau,\nu}$, we finally get the result of eq. (B.7),

$$\sum_{n,m} |\Phi_{\tau,\nu,t}(n,m)| \sim K + C_{\tau,\nu} \sum_{k=n'}^{\infty} \frac{\log k}{k^{2-(\tau+\nu)/\tau_{max}}}\,,$$

which demonstrates convergence of the series for $\tau+\nu<\tau_{max}$.

Appendix C

Details on numerical methods

This appendix collects details on numerical methods that may be of use to those looking to reproduce or extend our results. A few additional thoughts on appropriate solution methods are also included. All simulations were run on C++ code, and subsequent analysis performed using Python with the SciPy library.

C.1 Numerically solving integral characteristic equations

Solving a characteristic equation such as eq. (4.5) or eq. (4.59) numerically can be a little tricky for two principal reasons:

- the solutions (λ_n) are complex,
- and it is essential for us not to miss any solution.

The fact that solutions are complex just means that we essentially have a 2D problem, and therefore must use a multivariate solver. There are many such routines available, and while they require more care than single variables solvers, they can certainly get the job done. Some implementations of common general purpose algorithms are listed in [49].

The second point is in fact the most critical: because we are going to be summing over all the eigenvalues, missing even just one[1] of them can throw the entire sum off. This can be a challenge because numerical solvers are only good at finding "a" solution close to some initial guess or within given bounds. Here there is an infinity of solutions, which can be quite close (see fig. 2.5), so we need very good guesses to initiate the solver.

The best way to obtain such guesses is probably to use the discretized operator $\boldsymbol{L}^{(N)}$ (eq. (4.9)). Since it's expressed as a matrix, we can find its eigenvalues using linear algebra methods – a task for which very robust and efficient algorithms are already commonplace.[2] If we choose N large enough, the eigenvalues of $\boldsymbol{L}^{(N)}$ should be extremely close to the solutions of the characteristic equation (4.5). For many applications in fact, it will likely

[1] Obviously we can't really find *all* the eigenvalues since there are infinitely many. What's important is that none are missed within a given search domain, chosen to include the eigenvalues corresponding to dominant terms in the sum.

[2] For many years the standard numerical routines for these kinds of operations have been packaged as LAPACK [50]. They've been integrated into linear algebra packages for almost every modern programming language; the go-to Python implementation is the one included in the SciPy library [51].

be more efficient to choose a large N and simply use the eigenvalues of $\boldsymbol{L}^{(N)}$, without going through the trouble of refining the estimate with a 2D solver.

Remark C.1

Since in most of our examples the feedback kernel had only a single delay, we could compute the eigenvalues directly using eq. (2.17). All that was required was an implementation of the Lambert W function, which is included in most computational packages. The exception was section 4.6, where we considered a uniform kernel over $[0, 1]$. For the figure in that section, we calculated the first 100 eigenvalues for the discretized operator with $N = 400$. We estimated the differences of the dominant eigenvalues to be within a few percent of their exact values, which was precise enough for our purposes.

Other applications may require more precise estimates of the eigenvalues, in a further refinement of the values with a numerical solver, as we described above, will likely become necessary

Remark C.2

When using eigenvalues estimated from the discretized operator $\boldsymbol{L}^{(N)}$, we found it best to also use the matrix's right and left eigenvectors (which can be obtained along with the eigenvalues) rather than the expressions of eqs. (4.4) and (4.17). Similarly, the normalization constant should be calculated using eq. (4.14) rather than eq. (4.16).

C.2 Integrating delay differential equations

Any numerical integrator produces a discrete time series; if a constant step size of Δt is used, then the solution is only computed at times $t = \Delta t, 2\Delta t, 3\Delta t$ and so on. Therefore, to ensure we can compute the right-hand side of a DDE given by,

$$dx_0(t) = f(x_0(t), x_0(t - \tau_1))\, dt\,,$$

it's often easiest to chose Δt such that it is a divisor of τ_1. For a single delay this constraint is not much more than an annoyance, but with multiple delays it can be harder to respect. The greatest common divisor of multiple delays may also be unnecessarily small, thereby increasing the computation time.

In order to be able to deal with multiple or distributed delays, it can therefore be easier to compute $x_0(t - \tau_1)$ using interpolation between the two closest steps. One must only take care that the order of the interpolation is at least as high as the integration order (for example, with a 4^{th}-order Runge-Kutta integrator, interpolation should be done with a polynomial of degree at least 4). Special care must also be taken for those points where the function isn't sufficiently smooth. For our simulations, we implemented the integration algorithm described by Oberle and Pesh [52], which is specifically optimized for delayed equations.

On the other hand, for stochastic integrators we are pretty much forced to fall back on first-order interpolation since integrators are usually either first- or half-order [53, § 6] and the solutions are nowhere smooth. This makes most of the considerations in [52] somewhat moot.

C.3 Integrating stochastic delay differential equations

There are at least a few published algorithms for numerically integrating stochastic differential equations, including some second order schemes [54]. However, with stochastic solutions being non-differentiable, higher order methods are significantly more sophisticated than their deterministic counterparts. And with many of them being less than a decade old, at present they still lack the time-tested "stock implementations" we have for example for deterministic Runge-Kutta algorithms.

For these reasons, it is quite common to sacrifice computational efficiency for simplicity and stick with the Euler-Maruyama algorithm [53, § 6.1]. It is the natural extension of Euler's method for deterministic equations, and while of lowest possible order (i.e. $\frac{1}{2}$), it is also easiest to understand and implement. This is the method we used for all our stochastic computations.

C.4 Obtaining the source code

Most of the code written for numerical simulations is generic: very often, the difference between one simulation and another is just a few lines of code. To avoid re-introducing bugs by rewriting, our integration was packaged code into a reusable library which can be found on the author's personal repository: `https://bitbucket.org/acrene/frantic`. This library is in fact designed to be adaptable to a wide array of initial value problems, and may be used as an extensible C++ integrator.

The code written specifically for the computations presented in this thesis is available upon request to the author.

Bibliography

[1] Küchler, U. and Mensch, B. *Langevins stochastic differential equation extended by a time-delayed term.* Stochastics and Stochastic Reports, **40**(1-2):23–42 (1992). ISSN 1045-1129. http://dx.doi.org/10.1080/17442509208833780. (Cited on pages 2, 9, and 35.)

[2] Guillouzic, S. *Fokker-Planck approach to stochastic delay differential equations.* Ph.D., University of Ottawa, Ottawa, Canada (2001). http://www.ruor.uottawa.ca/en/handle/10393/244. (Cited on page 2.)

[3] Frank, T.D. *et al. Fokker-Planck perspective on stochastic delay systems: Exact solutions and data analysis of biological systems.* Physical Review E, **68**(2):021912 (2003). http://dx.doi.org/10.1103/PhysRevE.68.021912. (Cited on pages 2 and 38.)

[4] Amann, A. *et al. Some basic remarks on eigenmode expansions of time-delay dynamics.* Physica A: Statistical Mechanics and its Applications, **373**:191–202 (2007). ISSN 0378-4371. http://dx.doi.org/10.1016/j.physa.2005.12.073. (Cited on pages 2, 10, 13, 15, 36, and 47.)

[5] Bhat, H.S. and Madushani, W.M.A. *Computing the Density Function for a Nonlinear Stochastic Delay System.* In *Proceedings of the 12th IFAC Workshop on Time Delay Systems.* University of Michigan, Ann Arbor, MI, USA (2015). In press. (Cited on pages 2 and 39.)

[6] Bellman, R. *Differential-difference equations*, volume 6 of *Mathematics in science and engineering.* Academic Press, New York (1963). (Cited on pages 2, 6, 10, 12, 13, 16, and 71.)

[7] Hutt, A. and Lefebvre, J. *Stochastic center manifold analysis in scalar nonlinear systems involving distributed delays and additive noise.* Markov Processes and Related Fields (2014). https://hal.inria.fr/hal-01006388/document. (Cited on pages 2, 3, 11, and 37.)

[8] Bahar, A. and Mao, X. *Stochastic delay Lotka–Volterra model.* Journal of Mathematical Analysis and Applications, **292**(2):364–380 (2004). ISSN 0022-247X. http://dx.doi.org/10.1016/j.jmaa.2003.12.004. (Cited on page 3.)

[9] Lei, J. and Mackey, M. *Stochastic Differential Delay Equation, Moment Stability, and Application to Hematopoietic Stem Cell Regulation System.* SIAM Journal on Applied

Mathematics, **67**(2):387–407 (2007). ISSN 0036-1399. http://dx.doi.org/10.1137/060650234. (Cited on page 3.)

[10] Jaurigue, L. *et al. Timing jitter of passively-mode-locked semiconductor lasers subject to optical feedback: A semi-analytic approach.* Physical Review A, **92**(5):053807 (2015). http://dx.doi.org/10.1103/PhysRevA.92.053807. (Cited on page 3.)

[11] Burgers, G. *The El Niño stochastic oscillator.* Climate Dynamics, **15**(7):521–531 (1999). ISSN 0930-7575, 1432-0894. http://dx.doi.org/10.1007/s003820050297. (Cited on page 3.)

[12] Kashima, K. and Yamamoto, N. *Control of Quantum Systems Despite Feedback Delay.* IEEE Transactions on Automatic Control, **54**(4):876–881 (2009). ISSN 0018-9286. http://dx.doi.org/10.1109/TAC.2008.2010969. (Cited on page 3.)

[13] Erneux, T. *Applied delay differential equations.* Number v. 3 in Surveys and tutorials in the applied mathematical sciences. Springer, New York (2009). ISBN 978-0-387-74371-4. (Cited on page 4.)

[14] Smith, H. *An Introduction to Delay Differential Equations with Applications to the Life Sciences.* Number 57 in Texts in Applied Mathematics. Springer-Verlag New York, 1 edition (2007). ISBN 978-1-4419-7645-1. http://www.springer.com/mathematics/dynamical+systems/book/978-1-4419-7645-1. (Cited on pages 9 and 69.)

[15] Lea, S. *Mathematics for physicists.* Brooks/Cole-Thomson Learning, Belmont CA (2004). ISBN 978-0-534-37997-1. (Cited on pages 10 and 71.)

[16] Corless, R.M. *et al. On the LambertW function.* Advances in Computational Mathematics, **5**(1):329–359 (1996). ISSN 1019-7168, 1572-9044. http://dx.doi.org/10.1007/BF02124750. (Cited on pages 13, 16, and 18.)

[17] Asl, F.M. and Ulsoy, A.G. *Analysis of a System of Linear Delay Differential Equations.* Journal of Dynamic Systems, Measurement, and Control, **125**(2):215–223 (2003). ISSN 0022-0434. http://dx.doi.org/10.1115/1.1568121. (Cited on pages 13 and 72.)

[18] Szalai, R. and Orosz, G. *Decomposing the dynamics of heterogeneous delayed networks with applications to connected vehicle systems.* Physical Review E, **88**(4) (2013). ISSN 1539-3755, 1550-2376. http://dx.doi.org/10.1103/PhysRevE.88.040902. (Cited on page 13.)

[19] Yi, S. *et al. Solution of Systems of Linear Delay Differential Equations via Laplace Transformation.* In *2006 45th IEEE Conference on Decision and Control*, pages 2535–2540 (2006). http://dx.doi.org/10.1109/CDC.2006.377712. (Cited on pages 13 and 72.)

[20] Risken, H. *The Fokker-Planck equation: methods of solution and applications*. Number v. 18 in Springer series in synergetics. Springer-Verlag, New York, 2nd edition (1996). ISBN 3-540-61530-X. (Cited on pages 14, 26, and 32.)

[21] Gillespie, D.T. *Simple brownian diffusion: an introduction to the standard theoretical models*. Oxford University Press, New York, NY (2012). ISBN 978-0-19-966450-4. (Cited on pages 19, 24, and 26.)

[22] Gillespie, D.T. *The mathematics of Brownian motion and Johnson noise*. American Journal of Physics, **64**(3):225–240 (1996). ISSN 0002-9505, 1943-2909. http://dx.doi.org/10.1119/1.18210. (Cited on page 22.)

[23] Gardiner, C.W. *Handbook of stochastic methods: for physics, chemistry and the natural sciences ; with 29 figures*. Number 13 in Springer series in synergetics. Springer, Berlin (1983). ISBN 978-3-540-11357-7 978-0-387-11357-9. (Cited on pages 24 and 26.)

[24] Lau, A.W.C. and Lubensky, T.C. *State-dependent diffusion: Thermodynamic consistency and its path integral formulation*. Physical Review E, **76**(1):011123 (2007). http://dx.doi.org/10.1103/PhysRevE.76.011123. (Cited on page 24.)

[25] Kampen, N.G.v. *Itô versus Stratonovich*. Journal of Statistical Physics, **24**(1):175–187 (1981). ISSN 0022-4715, 1572-9613. http://dx.doi.org/10.1007/BF01007642. (Cited on page 24.)

[26] Gillespie, D.T. *The multivariate Langevin and Fokker–Planck equations*. American Journal of Physics, **64**(10):1246–1257 (1996). ISSN 0002-9505, 1943-2909. http://dx.doi.org/10.1119/1.18387. (Cited on page 27.)

[27] Mackey, M.C. and Nechaeva, I.G. *Solution moment stability in stochastic differential delay equations*. Physical Review E, **52**(4):3366 (1995). http://journals.aps.org/pre/abstract/10.1103/PhysRevE.52.3366. (Cited on pages 31 and 35.)

[28] Horsthemke, W. *Noise-induced transitions theory and applications in physics, chemistry, and biology*. Springer-Verlag, Berlin (2006). ISBN 978-3-540-36852-6. http://dx.doi.org/10.1007/3-540-36852-3. (Cited on page 33.)

[29] Mackey, M.C. and Nechaeva, I.G. *Noise and stability in differential delay equations*. Journal of Dynamics and Differential Equations, **6**(3):395–426 (1994). http://link.springer.com/article/10.1007/BF02218856. (Cited on page 35.)

[30] Kuehn, C. *Multiple time scale dynamics*. Number 191 in Applied mathematical sciences. Springer, Cham (2015). ISBN 978-3-319-12316-5 978-3-319-12315-8. (Cited on pages 35 and 36.)

[31] Carr, T. *et al. Delayed-Mutual Coupling Dynamics of Lasers: Scaling Laws and Resonances*. SIAM Journal on Applied Dynamical Systems, **5**(4):699–725 (2006). http://dx.doi.org/10.1137/050647918. (Cited on page 36.)

[32] Yu, N. *et al.* *Stochastic phase dynamics: Multiscale behavior and coherence measures.* PHYSICAL REVIEW-SERIES E-, **73**(5):056205 (2006). http://www.researchgate.net/profile/Yue-Xian_Li/publication/6981735_Stochastic_phase_dynamics_multiscale_behavior_and_coherence_measures/links/0912f51301eb0ce48f000000.pdf. (Cited on page 36.)

[33] Klosek, M.M. and Kuske, R. *Multiscale Analysis of Stochastic Delay Differential Equations.* Multiscale Modeling & Simulation, **3**(3):706–729 (2005). ISSN 1540-3459, 1540-3467. http://dx.doi.org/10.1137/030601375. (Cited on page 36.)

[34] Gaudreault, M. *et al.* *Analytical determination of the bifurcation thresholds in stochastic differential equations with delayed feedback.* Physical Review E, **82**(5):051124 (2010). http://dx.doi.org/10.1103/PhysRevE.82.051124. (Cited on page 36.)

[35] Hunt, D. *et al.* *Network Synchronization in a Noisy Environment with Time Delays: Fundamental Limits and Trade-Offs.* Physical Review Letters, **105**(6):068701 (2010). http://dx.doi.org/10.1103/PhysRevLett.105.068701. (Cited on page 37.)

[36] Hutt, A. *et al.* *Additive Global Noise Delays Turing Bifurcations.* Physical Review Letters, **98**(23):230601 (2007). http://dx.doi.org/10.1103/PhysRevLett.98.230601. (Cited on page 37.)

[37] Lefebvre, J. *et al.* *Reduced dynamics for delayed systems with harmonic or stochastic forcing.* Chaos: An Interdisciplinary Journal of Nonlinear Science, **22**(4):043121 (2012). ISSN 1054-1500, 1089-7682. http://dx.doi.org/10.1063/1.4760250. (Cited on page 37.)

[38] Guillouzic, S. *et al.* *Small delay approximation of stochastic delay differential equations.* Physical Review E, **59**(4):3970–3982 (1999). http://dx.doi.org/10.1103/PhysRevE.59.3970. (Cited on page 38.)

[39] Gaudreault, M. *et al.* *Correlation times in stochastic equations with delayed feedback and multiplicative noise.* Physical Review E, **83**(1):011903 (2011). http://dx.doi.org/10.1103/PhysRevE.83.011903. (Cited on page 38.)

[40] Milton, J. and Ohira, T. *Delayed random walks: Investigating the interplay between delay and noise.* In B. Balachandran, T. Kalmár-Nagy and D. Gilsinn, editors, *Delay differential equations: recent advances and new directions*, pages 305–332. Springer, New York (2009). ISBN 978-0-387-85595-0. (Cited on page 39.)

[41] Carson C. Chow and Buice, M.A. *Path Integral Methods for Stochastic Differential Equations* (2012). http://arxiv.org/pdf/1009.5966v2.pdf. (Cited on page 39.)

[42] Dickman, R. and Vidigal, R. *Path integrals and perturbation theory for stochastic processes.* Brazilian Journal of Physics, **33**(1):73–93 (2003). ISSN 0103-9733. http://dx.doi.org/10.1590/S0103-97332003000100005. (Cited on page 39.)

[43] Navarra, A. *et al. The Path Integral Formulation of Climate Dynamics*. PLoS ONE, **8**(6):e67022 (2013). http://dx.doi.org/10.1371/journal.pone.0067022. (Cited on page 39.)

[44] Schwartz, I.B. *et al. Noise-induced switching and extinction in systems with delay*. Physical Review E, **91**(1):012139 (2015). http://dx.doi.org/10.1103/PhysRevE.91.012139. (Cited on page 39.)

[45] Carr, T.W. *Stochastic Extinction in the Presence of Delayed Feedback* (2011). http://faculty.smu.edu/tcarr/publications/escape.pdf. (Cited on page 39.)

[46] Berry, M.V. *Quantal Phase Factors Accompanying Adiabatic Changes*. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, **392**(1802):45–57 (1984). http://dx.doi.org/10.1098/rspa.1984.0023. (Cited on page 69.)

[47] Stanley, H.E. *et al. Eigenvalue Degeneracy as a Possible "Mathematical Mechanism" for Phase Transitions*. Journal of Applied Physics, **41**(3):1278–1280 (1970). ISSN 0021-8979, 1089-7550. http://dx.doi.org/10.1063/1.1658912. (Cited on page 69.)

[48] Langevin, P. *Sur la théorie du mouvement brownien*. Comptes Rendus de l'Académie des Sciences (Paris), **146**:530–533 (1908). (Cited on page 71.)

[49] Scipy community. *Optimization and root finding (scipy.optimize) — SciPy v0.14.0 Reference Guide* (2014). http://docs.scipy.org/doc/scipy-0.14.0/reference/optimize.html. (Cited on page 88.)

[50] *LAPACK — Linear Algebra PACKage* (2015). http://www.netlib.org/lapack/. (Cited on page 88.)

[51] Scipy community. *Linear algebra (scipy.linalg) — SciPy v0.14.0 Reference Guide* (2014). http://docs.scipy.org/doc/scipy-0.14.0/reference/linalg.html. (Cited on page 88.)

[52] Oberle, H.J. and Pesch, H.J. *Numerical treatment of delay differential equations by Hermite Interpolation*. Numerische Mathematik, **37**(2):235–255 (1981). ISSN 0029-599X, 0945-3245. http://dx.doi.org/10.1007/BF01398255. (Cited on page 89.)

[53] Jacobs, K. *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press, Cambridge. UK; New York (2010). ISBN 978-0-521-76542-8 0-521-76542-0. (Cited on pages 89 and 90.)

[54] Rößler, A. *Second order Runge–Kutta methods for Stratonovich stochastic differential equations*. BIT Numerical Mathematics, **47**(3):657–680 (2007). ISSN 0006-3835, 1572-9125. http://dx.doi.org/10.1007/s10543-007-0130-3. (Cited on page 90.)