1

Introduction

The topic of this book is stochastic differential equations (SDEs). As their name suggests, they really are differential equations that produce a different "answer" or solution trajectory each time they are solved. This peculiar behaviour gives them properties that are useful in modeling of uncertainties in a wide range of applications, but at the same time it complicates the rigorous mathematical treatment of SDEs.

The emphasis of the book is on applied rather than theoretical aspects of SDEs and, therefore, we have chosen to structure the book in a way that we believe supports learning SDEs from an applied point of view. In the following, we briefly outline the purposes of each of the remaining chapters and explain how the chapters are connected to each other. In the chapters, we have attempted to provide a wide selection of examples of the practical application of theoretical and methodological results. Each chapter (except for the Introduction and Epilogue) also contains a representative set of analytic and hands-on exercises that can be used for testing and deepening understanding of the topics.

Chapter 2 is a brief outline of concepts and solutions methods for deterministic ordinary differential equations (ODEs). We especially emphasize solution methods for linear ODEs, because the methods translate quite easily to SDEs. We also examine commonly used numerical methods such as the Euler method and Runge–Kutta methods, which we extend to SDEs in the later chapters.

Chapter 3 starts with a number of motivating examples of SDEs found in physics, engineering, finance, and other applications. It turns out that in a modeling sense, SDEs can be regarded as noise-driven ODEs, but this notion should not be taken too far. The aim of the rest of the chapter is to show where things start to go wrong. Roughly speaking, with linear SDEs we are quite safe with this kind of thinking, but anything beyond them will not work.

In **Chapter 4**, we reformulate SDEs properly as stochastic integral equations where one of the terms contains a new kind of integral called the Itô integral. We then derive the change of variable formula, that is, the Itô formula for the integral, and use it to find complete solutions to linear SDEs. We also discuss some methods to solve nonlinear SDEs and look briefly at Stratonovich integrals.

The aim of **Chapter 5** is to analyze the statistics of SDEs as stochastic processes. We discuss and derive their generators, the Fokker–Planck–Kolmogorov equations, as well as Markov properties and transition densities of SDEs. We also derive the formal equations of the moments, such as the mean and covariance, for the SDE solutions. It turns out, however, that these equations cannot easily be solved for other than linear SDEs. This challenge will be tackled later in the numerical methods chapters.

As linear SDEs are very important in applications, we have dedicated **Chapter 6** to solution methods for their statistics. Although explicit solutions to linear SDEs and general moment equations for SDEs were already given in Chapters 4 and 5, here we also discuss and derive explicit mean and covariance equations, transition densities, and matrix fraction methods for the numerical treatment of linear SDEs. We also discuss steady-state solutions and Fourier analysis of linear time-invariant (LTI) SDEs as well as temporal covariance functions of general linear SDEs.

In **Chapter 7**, we discuss some useful theorems, formulas, and results that are typically required in more advanced analysis of SDEs as well as in their numerical methods. In addition to the Lamperti transform, Girsanov theorem, and Doob's h-transform, we also show how to find solutions to partial differential equations with Feynman–Kac formulas and discuss some connections to path integrals in physics. This chapter is not strictly necessary for understanding the rest of the chapters and can be skipped during a first reading.

Although the Itô stochastic calculus that is derivable from the Itô formula is theoretically enough for defining SDEs, it does not help much in practical solution of nonlinear SDEs. In **Chapter 8**, we present numerical simulation-based solution methods for SDEs. The methods are based primarily on Itô–Taylor series and stochastic Runge–Kutta methods, but we also discuss the Verlet and exact algorithm methods.

In many applications we are interested in the statistics of SDEs rather than their trajectories per se. In **Chapter 9**, we develop methods for approximate computation of statistics such as means and covariances or probability densities of SDEs – however, many of the methods are suitable for

numerical simulation of SDEs as well. We start with classical and modern Gaussian "assumed density" approximations and then proceed to other linearization methods. We also discuss Taylor and Hermite series approximations of transition densities and their moments, numerical solutions of Fokker–Planck–Kolmogorov equations, simulation-based approximations, and finally pathwise Wong–Zakai approximations of SDEs.

An important and historically one of the first applications of SDEs is the filtering and smoothing theory. In **Chapter 10**, we describe the basic ideas of filtering and smoothing and then proceed to the classical Kushner–Stratonovich and Zakai equations. We also present the linear and nonlinear Kalman–Bucy and Kalman filters and discuss their modern variants. Finally, we present formal equations and approximation methods for the corresponding smoothing problems.

The aim of **Chapter 11** is to give an overview of parameter estimation methods for SDEs. The emphasis is on statistical likelihood-based methods that aim at computing maximum likelihood (ML) or maximum a posteriori (MAP) estimates or are targeted to full Bayesian inference on the parameters. We start with brief descriptions of the ideas of ML and MAP estimates as well as Markov chain Monte Carlo (MCMC) methods. Parameter estimation in linear SDEs is then discussed, and finally we give approximate likelihood methods for parameter estimation in nonlinear SDEs. We also discuss some parameter estimation methods for indirectly observed SDEs.

Chapter 12 addresses the somewhat less traditional topic of connections between machine learning and SDEs. The aim is to discuss links between Gaussian process regression, Kalman filtering, and SDEs, along with applications of the methods across the fields of signal processing and machine learning.

Finally, **Chapter 13** concludes the book with an overview and gives some hints where to go next. We also discuss additional topics such as fractional Brownian motions, Lévy process driven SDEs, and stochastic control problems.

2

Some Background on Ordinary Differential Equations

The chapter provides background on deterministic (nonstochastic) ordinary differential equations (ODEs) from points of view especially suited to the context of stochastic differential equations (SDEs). As SDEs are inherently inhomogeneous differential equations (i.e., they have an input), we will concentrate on solution methods suitable for them. Furthermore, as linear and especially linear time-invariant (LTI) ODE systems are important in applications, we review the matrix exponential– and transition matrix–based methods of solution. We also discuss Fourier– and Laplace transform–based solution methods for LTI ODEs and for computing matrix exponentials. For more details on ODE methods and theory, the reader is referred to the books of Kreyszig (1993), Tenenbaum and Pollard (1985), and Hairer et al. (2008), although the same information can be found in many other books as well.

2.1 What Is an Ordinary Differential Equation?

An ODE is an equation in which the unknown quantity is a function, and the equation involves derivatives of the unknown function. For example, the second-order differential equation for a forced spring–mass system (or, e.g., a resonator circuit in telecommunications) can be generally expressed as

$$\frac{\mathrm{d}^2 x(t)}{\mathrm{d}t^2} + \gamma \frac{\mathrm{d}x(t)}{\mathrm{d}t} + \nu^2 x(t) = w(t), \tag{2.1}$$

where ν and γ are constants that determine the resonant angular velocity and damping of the spring. The *force* $w(t)$ is some given function that may or may not depend on time. In this equation, the position variable x is called the *dependent variable* and time t is the *independent variable*. The equation is of *second order*, because it contains the second derivative and no higher-order terms are present. It is *linear*, because $x(t)$ appears linearly

in the equation. The equation is *inhomogeneous*, because it contains the *forcing* term $w(t)$. This inhomogeneous term will become essential in later chapters, because replacing it with a random process leads to a stochastic differential equation.

Here a *solution* to the differential equation is defined as a *particular solution*, a function that satisfies the equation and does not contain any arbitrary constants. A *general solution* on the other hand contains every particular solution of the equation parameterized by some free constants. To actually solve the differential equation, it is necessary to tie down the general solution by some initial conditions. In the preceding case, this means that we need to know the spring–mass position $x(t)$ and velocity $\mathrm{d}x(t)/\mathrm{d}t$ at some fixed initial time $t = t_0$. Given these initial values, there is a unique solution to the equation (provided that $w(t)$ is continuous). Instead of initial conditions, we could also fix some other (boundary) conditions of the differential equation to get a unique solution, but here we only consider differential equations with given initial conditions.

Note that it is common not to write the dependencies of x and w on t explicitly, and write the equation as

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \gamma \frac{\mathrm{d}x}{\mathrm{d}t} + \nu^2 x = w. \tag{2.2}$$

Although it sometimes is misleading, this "ink saving" notation is very commonly used, and we will also employ it here whenever there is no risk of confusion. Furthermore, because in this section and in this whole book we mainly consider ordinary differential equations, we often drop the word "ordinary" and just talk about differential equations.

Time derivatives are also sometimes denoted with dots over the variable, such as $\dot{x} = \mathrm{d}x/\mathrm{d}t$, $\ddot{x} = \mathrm{d}^2x/\mathrm{d}t^2$ and so on. In this *Newtonian notation*, the previous differential equation would be written as

$$\ddot{x} + \gamma \dot{x} + \nu^2 x = w. \tag{2.3}$$

Differential equations of an arbitrary order n can (almost) always be converted into vector differential equations of order one. For example, in the preceding spring model, if we define a *state variable* $\mathbf{x}(t) = (x_1(t), x_2(t)) = (x(t), \mathrm{d}x(t)/\mathrm{d}t)$, we can rewrite the previous differential equation as a first-order vector differential equation:

$$\underbrace{\begin{pmatrix} \mathrm{d}x_1(t)/\,\mathrm{d}t \\ \mathrm{d}x_2(t)/\,\mathrm{d}t \end{pmatrix}}_{\mathrm{d}\mathbf{x}(t)/\mathrm{d}t} = \underbrace{\begin{pmatrix} 0 & 1 \\ -\nu^2 & -\gamma \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}}_{\mathbf{f}(\mathbf{x}(t))} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\mathbf{L}} w(t). \tag{2.4}$$

The preceding equation can be seen to be a special case of models of the form

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{L}(\mathbf{x}(t), t)\,\mathbf{w}(t), \tag{2.5}$$

where the vector-valued function $\mathbf{x}(t) \in \mathbb{R}^D$ is generally called the state of the system, $\mathbf{f}(\bullet, \bullet)$ and $\mathbf{L}(\bullet, \bullet)$ are arbitrary functions, and $\mathbf{w}(t) \in \mathbb{R}^S$ is some (vector-valued) forcing function, driving function, or input to the system. Note that we can absorb the second term on the right into the first term to yield

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}(t), t), \tag{2.6}$$

and in that sense Equation (2.5) is slightly redundant. However, the form (2.5) turns out to be useful in the context of stochastic differential equations, and thus it is useful to consider it explicitly.

The first-order vector differential equation representation of an nth-order differential equation is often called the state-space form of the differential equation. Because nth order differential equations can (almost) always be converted into equivalent n-dimensional vector-valued first-order differential equations, it is convenient to just consider such first-order equations instead of considering nth-order equations explicitly. Thus in this book, we develop the theory and solution methods (mainly) for first-order vector differential equations and assume that nth-order equations are always first converted into equations of this class.

The spring–mass model in Equation (2.4) is also a special case of *linear differential equations* of the form

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{F}(t)\,\mathbf{x}(t) + \mathbf{L}(t)\,\mathbf{w}(t), \tag{2.7}$$

which is a very useful class of differential equations often arising in applications. The usefulness of linear equations is that we can actually solve these equations, unlike general nonlinear differential equations. This kind of equations will be analyzed in the next sections.

2.2 Solutions of Linear Time-Invariant Differential Equations

Consider the following scalar linear homogeneous differential equation with a fixed initial condition at $t = 0$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F\,x, \qquad x(0) = \text{given}, \tag{2.8}$$

where F is a constant. This equation can now be solved, for example, via separation of variables, which in this case means that we formally multiply by $\mathrm{d}t$ and divide by x to yield

$$\frac{\mathrm{d}x}{x} = F\,\mathrm{d}t. \tag{2.9}$$

If we now integrate the left-hand side from $x(0)$ to $x(t)$ and right-hand side from 0 to t, we get

$$\log x(t) - \log x(0) = F\,t, \tag{2.10}$$

which can be solved for $x(t)$ to give the final solution:

$$x(t) = \exp(F\,t)\,x(0). \tag{2.11}$$

Another way of arriving at the same solution is by integrating both sides of the original differential equation from 0 to t. Because $\int_0^t \mathrm{d}x/\mathrm{d}t\;\mathrm{d}t = x(t) - x(0)$, we can express the solution $x(t)$ as

$$x(t) = x(0) + \int_0^t F\,x(\tau)\,\mathrm{d}\tau. \tag{2.12}$$

We can now substitute the right-hand side of the equation for $x(\tau)$ inside the integral, which gives

$$\begin{aligned} x(t) &= x(0) + \int_0^t F\left[x(0) + \int_0^\tau F\,x(\tau')\,\mathrm{d}\tau'\right]\mathrm{d}\tau \\ &= x(0) + F\,x(0)\int_0^t \mathrm{d}\tau + \int_0^t\left[\int_0^\tau F^2\,x(\tau')\,\mathrm{d}\tau'\right]\mathrm{d}\tau \\ &= x(0) + F\,x(0)\,t + \int_0^t\int_0^\tau F^2\,x(\tau')\,\mathrm{d}\tau'\,\mathrm{d}\tau. \end{aligned} \tag{2.13}$$

Doing the same substitution for $x(\tau')$ inside the last integral further yields

$$\begin{aligned} x(t) &= x(0) + F\,x(0)\,t + \int_0^t\int_0^\tau F^2\left[x(0) + \int_0^{\tau'} F\,x(\tau'')\,\mathrm{d}\tau''\right]\mathrm{d}\tau'\,\mathrm{d}\tau \\ &= x(0) + F\,x(0)\,t + F^2\,x(0)\int_0^t\int_0^\tau \mathrm{d}\tau'\,\mathrm{d}\tau \\ &\quad + \int_0^t\int_0^\tau\int_0^{\tau'} F^3\,x(\tau'')\,\mathrm{d}\tau''\,\mathrm{d}\tau'\,\mathrm{d}\tau \\ &= x(0) + F\,x(0)\,t + F^2\,x(0)\,\frac{t^2}{2} + \int_0^t\int_0^\tau\int_0^{\tau'} F^3\,x(\tau'')\,\mathrm{d}\tau''\,\mathrm{d}\tau'\,\mathrm{d}\tau. \end{aligned} \tag{2.14}$$

It is easy to see that repeating this procedure yields the solution of the form

$$\begin{aligned} x(t) &= x(0) + F\,x(0)\,t + F^2\,x(0)\,\frac{t^2}{2} + F^3\,x(0)\,\frac{t^3}{6} + \cdots \\ &= \left(1 + F\,t + \frac{F^2\,t^2}{2!} + \frac{F^3\,t^3}{3!} + \cdots\right) x(0). \end{aligned} \tag{2.15}$$

The series in the parentheses can be recognized to be the Taylor series for $\exp(F\,t)$. Thus, provided that the series actually converges (it does), we again arrive at the solution

$$x(t) = \exp(F\,t)\,x(0). \tag{2.16}$$

The multidimensional generalization of the homogeneous linear differential equation (2.8) is an equation of the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}\,\mathbf{x}, \qquad \mathbf{x}(0) = \text{given}, \tag{2.17}$$

where $\mathbf{F}$ is a constant (i.e., time-independent) matrix. For this multidimensional equation, we cannot use the separation of variables method, because it only works for scalar equations. However, the series-based approach works and yields a solution of the form

$$\mathbf{x}(t) = \left(\mathbf{I} + \mathbf{F}\,t + \frac{\mathbf{F}^2\,t^2}{2!} + \frac{\mathbf{F}^3\,t^3}{3!} + \cdots\right) \mathbf{x}(0). \tag{2.18}$$

The series in the parentheses can now be seen as a matrix generalization of the exponential function. This series indeed is the definition of the matrix exponential

$$\exp(\mathbf{F}\,t) = \mathbf{I} + \mathbf{F}\,t + \frac{\mathbf{F}^2\,t^2}{2!} + \frac{\mathbf{F}^3\,t^3}{3!} + \cdots \tag{2.19}$$

and thus the solution to Equation (2.17) can be written as

$$\mathbf{x}(t) = \exp(\mathbf{F}\,t)\,\mathbf{x}(0). \tag{2.20}$$

Note that the matrix exponential cannot be computed by computing scalar exponentials of the individual elements in matrix $\mathbf{F}\,t$. It is a completely different function. Sometimes the matrix exponential is written as $\mathrm{expm}(\mathbf{F}\,t)$ to distinguish it from the elementwise computation, but here we use the common convention to simply write it as $\exp(\mathbf{F}\,t)$. The matrix exponential function can be found as a built-in function in most commercial and open-source mathematical software packages such as MATLAB® and Python. In addition to this kind of numerical solution, the exponential can be evaluated

analytically, for example, by directly using the Taylor series expansion, by using the Laplace or Fourier transform, or via the Cayley–Hamilton theorem (Åström and Wittenmark, 1997).

Example 2.1 (Matrix exponential)**.** *To illustrate the difference between the matrix exponential and the elementwise exponential, consider the equation*

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = 0, \quad x(0) = \textit{given}, \quad (\mathrm{d}x/\mathrm{d}t)(0) = \textit{given}, \tag{2.21}$$

which in state-space form can be written as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{\mathbf{F}} \mathbf{x}, \qquad \mathbf{x}(0) = \textit{given}, \tag{2.22}$$

where $\mathbf{x} = (x, \mathrm{d}x/\mathrm{d}t)$*. Because* $\mathbf{F}^n = \mathbf{0}$ *for* $n > 1$*, the matrix exponential is simply*

$$\exp(\mathbf{F}\, t) = \mathbf{I} + \mathbf{F}\, t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \tag{2.23}$$

which is completely different from the elementwise matrix exponential:

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} \exp(0) & \exp(t) \\ \exp(0) & \exp(0) \end{pmatrix} = \begin{pmatrix} 1 & e^t \\ 1 & 1 \end{pmatrix}. \tag{2.24}$$

Let us now consider the following linear differential equation with an inhomogeneous term on the right-hand side:

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{F}\, \mathbf{x}(t) + \mathbf{L}\, \mathbf{w}(t), \tag{2.25}$$

where $\mathbf{x}(t_0)$ is given and the matrices $\mathbf{F}$ and $\mathbf{L}$ are constant. For inhomogeneous equations, the solution methods are numerous, especially if we do not want to restrict ourselves to specific kinds of forcing functions $\mathbf{w}(t)$. However, the following *integrating factor* method can be used for solving general inhomogeneous equations.

If we move the term $\mathbf{F}\, \mathbf{x}(t)$ in Equation (2.25) to the left-hand side and multiply with a term called integrating factor $\exp(-\mathbf{F}\, t)$, we get the following result:

$$\exp(-\mathbf{F}\, t)\, \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} - \exp(-\mathbf{F}\, t)\, \mathbf{F}\, \mathbf{x}(t) = \exp(-\mathbf{F}\, t)\, \mathbf{L}\, \mathbf{w}(t). \tag{2.26}$$

From the definition of the matrix exponential, we can derive the following property:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left[\exp(-\mathbf{F}\, t)\right] = -\exp(-\mathbf{F}\, t)\, \mathbf{F}. \tag{2.27}$$

The key thing is now to observe that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\exp(-\mathbf{F}\,t)\,\mathbf{x}(t)\right] = \exp(-\mathbf{F}\,t)\,\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} - \exp(-\mathbf{F}\,t)\,\mathbf{F}\,\mathbf{x}(t), \tag{2.28}$$

which is exactly the left-hand side of Equation (2.26). Thus we can rewrite the equation as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\exp(-\mathbf{F}\,t)\,\mathbf{x}(t)\right] = \exp(-\mathbf{F}\,t)\,\mathbf{L}\,\mathbf{w}(t). \tag{2.29}$$

Integrating from t_0 to t then gives

$$\exp(-\mathbf{F}\,t)\,\mathbf{x}(t) - \exp(-\mathbf{F}\,t_0)\,\mathbf{x}(t_0) = \int_{t_0}^{t} \exp(-\mathbf{F}\,\tau)\,\mathbf{L}\,\mathbf{w}(\tau)\,\mathrm{d}\tau, \tag{2.30}$$

which can be further rearranged to give the final solution

$$\mathbf{x}(t) = \exp(\mathbf{F}\,(t-t_0))\,\mathbf{x}(t_0) + \int_{t_0}^{t} \exp(\mathbf{F}\,(t-\tau))\,\mathbf{L}\,\mathbf{w}(\tau)\,\mathrm{d}\tau. \tag{2.31}$$

In the preceding solution, we have also used the identity $\exp(\mathbf{F}\,s)\,\exp(\mathbf{F}\,t) = \exp(\mathbf{F}\,(s+t))$, which is true because the matrices $\mathbf{F}\,s$ and $\mathbf{F}\,t$ commute. The expression (2.31) is the complete solution to Equation (2.25).

2.3 Solutions of General Linear Differential Equations

In this section, we consider solutions to more general, time-varying linear differential equations. The corresponding stochastic equations are a useful class of equations, because they can be solved in (semi)closed form quite analogously to the deterministic case considered in this section.

The solution presented in the previous section in terms of matrix exponential only works if the matrix $\mathbf{F}$ is constant. Thus for the time-varying homogeneous equation of the form

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(t)\,\mathbf{x}, \quad \mathbf{x}(t_0) = \text{given}, \tag{2.32}$$

the matrix exponential solution does not work. However, we can express the solution in the form

$$\mathbf{x}(t) = \boldsymbol{\Psi}(t, t_0)\,\mathbf{x}(t_0), \tag{2.33}$$