Charles R. Doering
Department of Physics and Institute for Nonlinear Studies, Clarkson University, Potsdam, New York 13699-5820

Modeling Complex Systems: Stochastic Processes, Stochastic Differential Equations, and Fokker-Planck Equations

This course consists of a brief mathematical introduction to the theory and applications of Markov diffusion processes. We discuss Brownian motion, white noise, stochastic differential equations, and Fokker-Planck equations. We approach the subject with the aim of developing techniques for analyzing models of complex interactions as stochastic dynamic systems and obtaining reduced descriptions of those models in certain limits. Toward the latter end, we discuss adiabatic elimination of fast noise as a singular perturbation analysis of higher-dimensional stochastic processes. The distinction between equilibrium and nonequilibrium stationary states is stressed as we illustrate the technology by application to a simple example from the natural sciences.

1990 Lectures in Complex Systems, SFI Studies in the Sciences of Complexity, Lect. Vol. III, Eds. L. Nadel and D. Stein, Addison-Wesley, 1991

1. INTRODUCTION

Complex systems, by definition, consist of many interacting components. In some situations we are concerned with the behavior of the system as a whole, and sometimes we are concerned with the behavior of a small part of it, focusing on just a few specific aspects of the phenomena at hand. The canonical example is perhaps equilibrium thermodynamics, where the behavior of large numbers of atoms or molecules are, for many purposes, adequately described by just a few macroscopic variables—energy, entropy, pressure, etc. Moreover, those macroscopic variables obey a compact set of fundamental laws independent of the precise constituents of the system or the precise nature of their interactions.

The task of reducing the number of variables in a complex dynamical system to some manageable level, and the search for some general rules controlling the behavior of complex systems, cuts across all disciplinary boundaries—thermodynamics, for example, is fundamental to all branches of physics, as well as chemistry, biology, and engineering. More specifically, even in the best of cases, it is impossible to completely divorce a real sub-system from its environment, whether that sub-system is a solid-state physics experiment, a chemical reaction, or a living entity. The influence of many factors are brought to bear on any "isolated" system, and we really have no hope of completely accounting for them theoretically. It is often useful then, in the modeling process, to lump all of these unknown, uncontrollable, and essentially stochastic factors into some noise acting on the system. In this way at least some of the overwhelming complexity of the whole is accounted for in the description of the part. The cost of our parody of these effects as a random noise is the sacrifice of complete predictability. Once we start talking about random variables, we necessarily restrict our predictions to probabilities and averages, rather than to the definite outcome of any specific experiment or observation.

The more general goal of obtaining faithful reduced descriptions of complicated networks of interacting components doesn't end with the introduction of a stochastic model. For example, the approximately 10^{23} coupled differential equations describing the motion of particles that make up a fluid can reduced to the Boltzmann equation by making some probabilistic assumptions about the dynamics. The Boltzmann equation description is then further reduced to one of mesoscopic average motions, or fluctuating hydrodynamics. On larger and slower scales, i.e., at the macroscopic level, fluctuations are accounted for by simple dissipation and pressure terms in a Navier-Stokes equation. Even the dynamics in the Euler and Navier-Stokes equations can, in certain situations, be accounted for in simpler amplitude equations like the Korteweg-deVries, nonlinear Schrödinger, or Ginzburg-Landau equations. Although few of these simplifying steps are taken with full mathematical rigor, there are some tested methods which are used over and over at different levels. Chief among them is the "averaging over" of fast variables, leaving a smaller set of the remaining slowly changing variables to describe the system.

Several aspects of stochastic modeling will be discussed here. Besides representing complex influences by random noise and exploring some elementary implications of noise in a simple nonlinear model, we will also take the next step and see how one can, in certain situations, reduce the level of description even further by the identification, separation, and elimination of rapidly evolving dynamical variables.

In these lectures we will present some basic ideas and methods in the theory of stochastic processes and stochastic differential equations. The aim is to study the behavior of noise-driven dynamical systems to gain some insight into the variety of phenomena possible in (generally nonlinear) systems coupled to a complex environment that can be modeled as random fluctuations, or noise. It is *not* the mission of these lectures either to present a rigorous mathematical formulation or to exhaust the subject of mathematical modeling. Rather, some fundamental concepts and derivations in the theory of continuous-time Markov processes will be introduced and developed to the point that these techniques may become part of the students' technical arsenal for the modeling and analysis of stochastic dynamic systems. Calculations are presented in detail here, with missing steps or natural extensions stated explicitly as exercises. The assumed mathematical background is a post-calculus-level undergraduate course in probability theory. The reader who can complete Exercise 1 (below) should be adequately prepared. I have very much tried to present the theory with as little mystery and in as much detail as the space and time allow.

Most of the material discussed herein may be found in greater mathematical detail in monographs by Arnold,[1] van Kampen,[6] Horsthemke and Lefever,[3] and Risken.[5] These books also develop myriad applications of stochastic processes and stochastic differential equations in engineering and the sciences—from signal processing to superconductivity to the electrical activity of nerve membranes. Recent reviews of many aspects of the interplay of noise and nonlinear dynamics are given in Moss and McClintock[4] and Doering, Brand, and Ecke.[2] Readers who wish to continue their studies in this area are referred to these works as a starting point into the current literature.

Beginning with the next section, we develop the mathematical framework for the analysis of continuous-time stochastic processes, focussing on the Wiener process, a.k.a. Brownian motion. We then discuss the concept of white noise as a prelude to the study of stochastic differential equations driven by gaussian white noise. We derive the relationship between the coefficients of a stochastic differential equation with white noise and those of the associated Fokker-Planck equation for the transition density, interpreting the stochastic differential equation (in the sense of Itô) as the continuous time limit of a discrete time problem. The Fokker-Planck equation is a linear partial differential equation, amenable to comprehensive analysis and, at times, exact solution.

Armed with these preliminaries, we introduce the notion of detailed balance in the stationary state of a stochastic dynamic system, and discuss the concepts of equilibrium vs. non-equilibrium stationary states. We then go on to present the general mathematical ideas of the reduction of the number of variables via adiabatic elimination and the overdamped approximation. This procedure is developed as a singular perturbation analysis of a multidimensional Fokker-Planck equation. The difference between the Itô and Stratonovich interpretations of stochastic differential equations with white noise is presented in this way by comparing the white-noise limit of a continuous-time real noise problem with the continuous time limit of the discrete time problem. Along the way these concepts are illustrated with a specific example, the Verhulst model of population biology.

EXERCISE 1. Let X_1 and X_2 be independent identically distributed (i.i.d) random variables, uniformly distributed over the interval (0,1). Let

$$G_1 = \sqrt{-2\ln[X_1]}\cos(2\pi X_2) \text{ and}$$
$$G_2 = \sqrt{-2\ln[X_1]}\sin(2\pi X_2).$$

Show that G_1 and G_2 are i.i.d. random variables, with a mean zero, unit variance gaussian distribution.

2. STOCHASTIC PROCESSES

Stochastic processes, also known as random processes, are to be thought of as random functions of a variable which we will call *time*. That is, if $X(\bullet)$ is a random process, then for each value t of its argument, $X(t)$ is a random variable characterized by a probability density function, $\rho(x,t)$. The argument, or *index*, of the

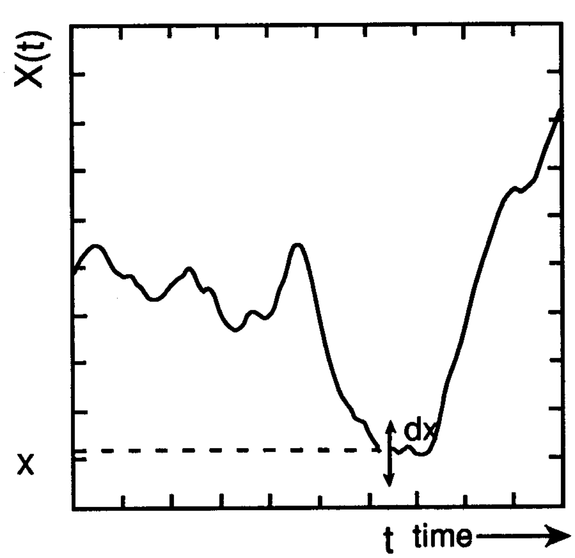


FIGURE 1 A "random" function. The single time probability density $\rho(x,t)$, times dx, gives the probability that the random function passes through a window of width dx about x.

ay by comparing the white-noise
the continuous time limit of the
pts are illustrated with a specific
y.

tically distributed (i.i.d) random
(0,1). Let

X_2) and

X_2).

, with a mean zero, unit variance

esses, are to be thought of as ran-
ime. That is, if $X(\bullet)$ is a random
$X(t)$ is a random variable charac-
. The argument, or *index*, of the

FIGURE 1 A "random" function. The single time probability density $\rho(x,t)$, times dx, gives the probability that the random function passes through a window of width dx about x.

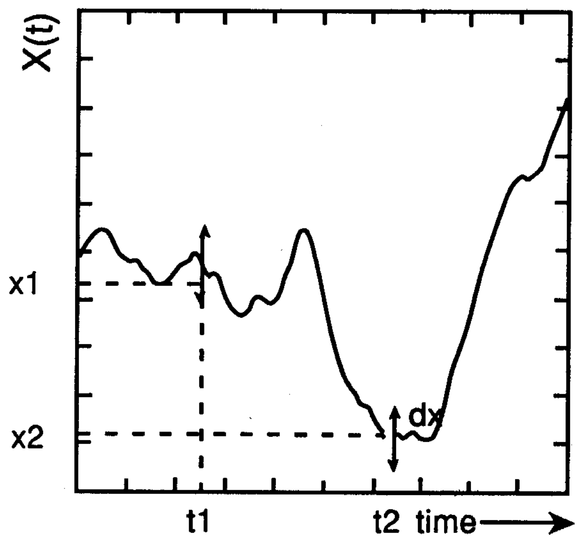


FIGURE 2 The joint density $\rho(x_2,t_2;x_1,t_1)$, times dx^2, gives the probability that the random function falls through both windows of width dx about x_1 at time t_1, and about x_2 at time t_2.

process t may be a discrete or a continuous variable; we concentrate on the case where t is a continuous variable. An actual realization of the process is called a sample path. The probability density function times the width of some small window dx around the value x at each instant, i.e., $\rho(x,t)dx$, gives the relative frequency that a random function falls within that window, as illustrated in Figure 1. The probability density is positive and for each time t is related to the probability of $X(t)$ taking some values by

$$\int_a^b \rho(x,t)dx = \mathrm{Prob}[a < X(t) < b], \tag{2.1}$$

for $a < b$, and where the Prob[$\bullet$] means the probability of the specified event.

It is not enough to know the characteristics of the random variables $X(t)$ for each t alone. For example, in order to contemplate the probability that one of the random functions passes through two windows, one of width dx_1 around x_1 at time t_1 *and* one of width dx_2 around x_2 at time t_2, as illustrated in Figure 2, we must consider the *joint* probability distribution $\rho(x_2,t_2;x_1,t_1)$. Our notation is that the semicolon (;) is read "and" in the Boolean sense. Only if the values of the random functions are *independent* at different times, i.e., if

$$\rho(x_2,t_2;x_1,t_1) = \rho(x_2,t_2)\rho(x_1,t_1), \tag{2.2}$$

does the single-time density function specify the statistics of the process as a whole. Processes where the values at different times are statistically independent are called *white-noise* processes.

In order to answer more complicated questions about a stochastic process we must know the joint probability distributions for the values of the functions at an arbitrary number of times. That is, we must consider the complete set of joint densities, or the finite-dimensional distributions, $\rho(x_1, t_1; \ldots; x_n, t_n)$, for all n. The joint distribution functions cannot be specified arbitrarily, but must satisfy the compatibility conditions

$$\int \rho(x_n, t_n; \ldots; x_i, t_i; \ldots; x_1, t_1) dx_i = \rho(x_n, t_n; \ldots; x_{i+1}, t_{i+1}; x_{i-1}, t_{i-1}; \ldots; x_1, t_1) \tag{2.3}$$

where the integral is over the allowed values of X, and the ith window has been "removed" on the right-hand side. This condition merely states that the random functions fall somewhere—anywhere—between $t = t_{i-1}$ and $t = t_{i+1}$.

A fundamental theorem of Kolmogorov states that this hierarchy of joint density functions is just what is needed to completely specify the stochastic process. This is not a trivial statement because it asserts that we need only know the probabilities that the random functions fall through any *discrete* number of windows, rather than the uncountable number of possibilities allowed by the continuous time variable. More precisely, Kolmogorov's theorem states that, for each consistent hierarchy of finite-dimensional distributions, there exists a probability space equipped with a sigma algebra, a measure, and a family of random variables $X(t)$, defined for each t, whose joint probability density functions are those originally given.

The *average*, or *expectation*, of a random variable X with probability density $\rho(x)$ is denoted $E\{X\}$,[1] and is defined as

$$E\{X\} = \int x \rho(x) dx . \tag{2.4}$$

The average of the nth power of X, $E\{X^n\}$, is called the nth *moment* of X, and the expectation of a function $f\{X\}$ is

$$E\{f(X)\} = \int f(x) \rho(x) dx . \tag{2.5}$$

For a stochastic process $X(t)$, the average of n products of the process at various times is called the *n-point correlation function:*

$$E\{X(t_n) \ldots X(t_1)\} = \int x_n \ldots x_1 \rho(x_n, t_n; \ldots; x_1, t_1) dx_n \ldots dx_1 . \tag{2.6}$$

The expectation of the product of the process at two times, i.e., the two-point correlation function

$$E\{X(t)X(s)\} = \int xy \rho(x, t; y, s) dx dy , \tag{2.7}$$

[1] A common notation for the expectation of a random variable X is $\langle X \rangle$, with the exact same meaning as $E\{X\}$.

will be referred to simply as the *correlation function* or the *covariance*. Note that while the density functions determine the moments and the correlation functions, the converse is not true; i.e., the process is *not* uniquely specified by the moments and the correlation functions.

The probability density that a random function passes through x_n at time t_n *given* that it passed through x_{n-1} at time t_{n-1}, x_{n-2} at time t_{n-2}, etc., is a *conditional* density. In terms of the finite-dimensional distributions, this conditional density is defined by

$$\rho(x_n, t_n | x_{n-1}, t_{n-1}; \ldots; x_1, t_1) = \frac{\rho(x_n, t_n; x_{n-1}, t_{n-1}; \ldots; x_1, t_1)}{\rho(x_{n-1}, t_{n-1}; \ldots; x_1, t_1)} . \tag{2.8}$$

for an ordered set of times $t_1 < \ldots < t_n$. Our notation is that the vertical bar ($|$) is read "given." In particular, the *transition density* for the process to go from x_{n-1} at t_{n-1} to x_n at t_n is

$$\rho(x_n, t_n | x_{n-1}, t_{n-1}) = \frac{\rho(x_n, t_n; x_{n-1}, t_{n-1})}{\rho(x_{n-1}, t_{n-1})} . \tag{2.9}$$

For the rest of these lectures we will be concerned with a restricted class of stochastic processes known as *Markov* processes. A Markov process is a stochastic process with the property that, in plain language, "the future is independent of the past given the present." In terms of the joint density functions, this means that

$$\rho(x_n, t_n | x_{n-1}, t_{n-1}; x_{n-2}, t_{n-2}; \ldots; x_1, t_1) = \rho(x_n, t_n | x_{n-1}, t_{n-1}), \tag{2.10}$$

and this property will serve as our definition of a Markov process. Once the value of the process $X(t)$ is specified at time t_{n-1}, its future evolution for times $t > t_{n-1}$ is statistically independent of its history before time t_{n-1}. The joint distribution functions for a Markov process may be written in terms of the single-time density and the transition density:

$$\begin{aligned}\rho(x_n, t_n; x_{n-1}, t_{n-1}; x_{n-2}, t_{n-2}; \ldots; x_2, t_2; x_1, t_1)& \\ = \rho(x_n, t_n | x_{n-1}, t_{n-1}) \rho(x_{n-1}, t_{n-1} | x_{n-2}, t_{n-2}) \ldots \rho(x_2, t_2 | x_1, t_1) \rho(x_1, t_1) .&\end{aligned} \tag{2.11}$$

Equation (2.11) says that, for Markov processes, the probability of passing through the n windows is the product of the probabilities to go from one to the next, times the probability of passing through the first window.

Markov processes are, in a practical sense, just one step more general than white-noise processes. That is, white-noise processes are defined by the single-time density function alone, from which all higher joint density functions may be built. Markov processes are defined by the single-time density function and the transition density function (or the two-time joint density alone), from which all higher joint density functions may be built. This extra degree of generality attributed to Markov processes makes a world of difference in applications. To a certain extent, white-noise processes are trivial; they lack any structure to justify their consideration as much more than just a collection of independent random variables. White noise is not useless, however, as some natural processes can be modeled as being nearly statistically independent at successive instants. Often the first step in modeling random phenomena is to determine the level on which some variables, or some aspect of the problem, can be modeled as white noise. On the other hand, Markov processes, with their one more degree of complexity, are a reasonable model of many stochastic systems where the variables are correlated at different times, but they "lose memory" of their history in the sense that knowledge of the current state is sufficient to predict the future (statistically) no matter how that current state was achieved. As we will discuss later in these lectures, Markov processes are the solutions of differential equations with white-noise coefficients and thus arise naturally in applications. More generally, as we will also discuss below, they can be good approximations to non-Markovian processes on long time scales.

EXERCISE 2. Using the definitions of the conditional density, Eq. (2.8), and Markov processes, Eq. (2.10), verify Eq. (2.11).

3. BROWNIAN MOTION AND WHITE NOISE

A fundamental example of a Markov process is *Brownian motion*, also referred to as the *Wiener process*, which we denote $W(t)$. It is the Markov process with a δ-function single-time density at $t = 0$,

$$\rho(w, t = 0) = \delta(w)\,, \tag{3.1}$$

and a gaussian transition density, of variance Δt, between times t and $t + \Delta t$,

$$\rho(w, t + \Delta t | w', t) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-\frac{1}{2}\frac{(w-w')^2}{\Delta t}}\,. \tag{3.2}$$

The single-time density is explicitly computed from these definitions according to

$$\begin{aligned} \rho(w,t) &= \int \rho(w, t; w', 0)\, dw' \\ &= \int \rho(w, t|w', 0)\rho(w', 0)\, dw' \\ &= \int \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{(w-w')^2}{t}} \delta(w')\, dw' \\ &= \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{w^2}{t}}\,. \end{aligned} \tag{3.3}$$

Hence $W(t)$ is a mean zer
finite-dimensional distribu

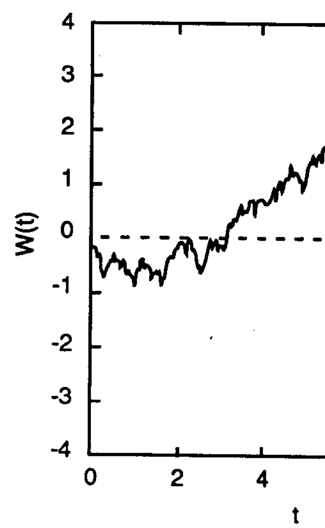
$\rho(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-}$
$= \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}}$
$\cdots \frac{1}{\sqrt{2\pi(t_2 - t_1)}}$

Brownian motion is a *ga*
$W(t_n)$ are jointly gaussian
process is shown in Figure
plotted in Figure 4 for sev
The moments of the W
distribution:

$E\{W(t)^{2}$

and

$E\{W(t)^{2n}\}$



ations. To a certain extent, white-
re to justify their consideration as
nt random variables. White noise
es can be modeled as being nearly
. Often the first step in modeling
on which some variables, or some
noise. On the other hand, Markov
plexity, are a reasonable model of
e correlated at different times, but
nse that knowledge of the current
cally) no matter how that current
hese lectures, Markov processes are
te-noise coefficients and thus arise
will also discuss below, they can be
es on long time scales.

onal density, Eq. (2.8), and Markov

NOISE

Brownian motion, also referred to
). It is the Markov process with a

$w)$, (3.1)

Δt, between times t and $t + \Delta t$,

$e^{-\frac{1}{2}\frac{(w-w')^2}{\Delta t}}$. (3.2)

from these definitions according to

dw'

$(w', 0)dw'$ (3.3)

$\delta(w')dw'$

Hence $W(t)$ is a mean zero, variance t, normally distributed random variable. The finite-dimensional distributions for Brownian motion are, according to Eq.(7),

$$\begin{aligned} &\rho(w_n, t_n; w_{n-1}, t_{n-1}; w_{n-2}, t_{n-2}; \ldots; w_2, t_2; w_1, t_1) \\ &= \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{1}{2}\frac{(w_n - w_{n-1})^2}{(t_n - t_{n-1})}} \frac{1}{\sqrt{2\pi(t_{n-1} - t_{n-2})}} e^{-\frac{1}{2}\frac{(w_{n-1} - w_{n-2})^2}{(t_{n-1} - t_{n-2})}} \\ &\quad \ldots \frac{1}{\sqrt{2\pi(t_2 - t_1)}} e^{-\frac{1}{2}\frac{(w_2 - w_1)^2}{(t_2 - t_1)}} \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{1}{2}\frac{w_1^2}{t_1}} . \end{aligned} \tag{3.4}$$

Brownian motion is a ***gaussian*** process because the random variables $W(t_1) \ldots W(t_n)$ are jointly gaussian random variables. A typical realization of the Wiener process is shown in Figure 3. The time-dependent probability density, Eq. (3.3), is plotted in Figure 4 for several time values.

The moments of the Wiener process are easy to compute because of its gaussian distribution:

$$E\{W(t)^{2n+1}\} = \int_{-\infty}^{\infty} w^{2n+1} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{w^2}{t}} dw = 0 \tag{3.5}$$

and

$$E\{W(t)^{2n}\} = \int_{-\infty}^{\infty} w^{2n} \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{w^2}{t}} dw = \frac{(2n)!}{2^n n!} t^n . \tag{3.6}$$

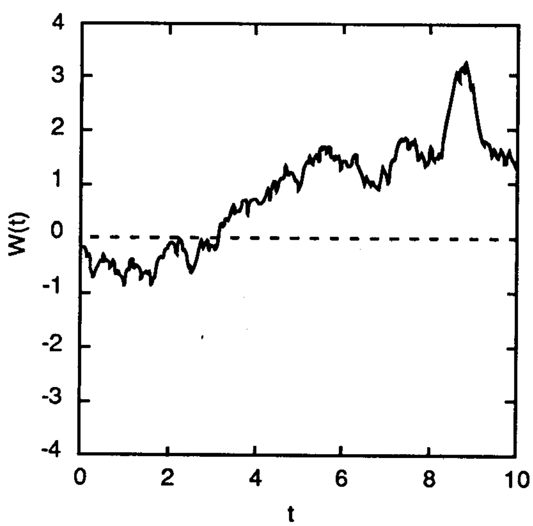


FIGURE 3 A typical realization of the Wiener process.

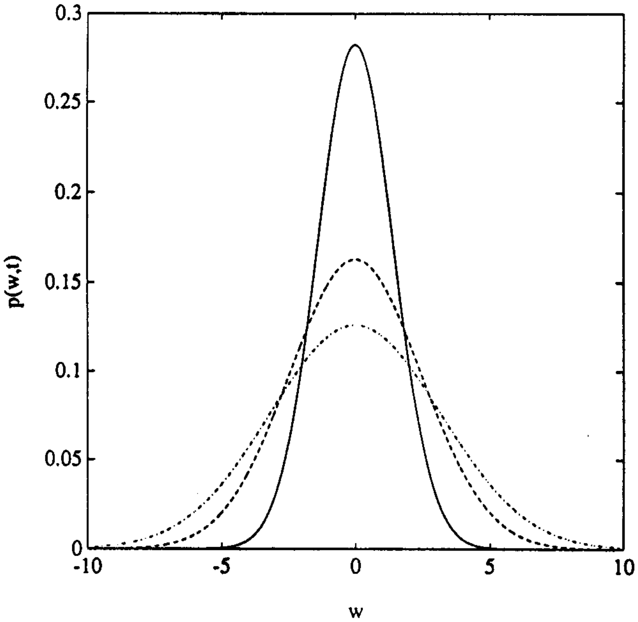


FIGURE 4 Several single-time probability densities for the Wiener process. These are the densities at times $t = 2$ (solid), $t = 6$ (dashed), and $t = 10$ (dash-dot).

The two-point correlation function—the covariance—is, for $t \geq s$,

$$\begin{aligned} E\{W(t)W(s)\} &= \int_{-\infty}^{\infty} dw \int_{-\infty}^{\infty} dw'\, ww' \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{1}{2}\frac{(w-w')^2}{(t-s)}} \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\frac{w'^2}{s}} \\ &= \int_{-\infty}^{\infty} \frac{dw'}{\sqrt{2\pi s}} w' e^{-\frac{1}{2}\frac{w'^2}{s}} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi(t-s)}} (z+w') e^{-\frac{1}{2}\frac{z^2}{(t-s)}} \\ &= \int_{-\infty}^{\infty} \frac{dw'}{\sqrt{2\pi s}} w'^2 e^{-\frac{1}{2}\frac{w'^2}{s}} \\ &= s\,. \end{aligned} \tag{3.7}$$

For an arbitrary ordering of t and s, the covariance is thus

$$E\{W(t)W(s)\} = \min\{t, s\}\,. \tag{3.8}$$

A quantity of particular interest is the change

$$\Delta W(t) = W(t + \Delta t) - W(t) \tag{3.9}$$

FIGURE 4 Several single-time probability densities for the Wiener process. These are the densities at times $t = 2$ (solid), $t = 6$ (dashed), and $t = 10$ (dash-dot).

of the Wiener process over a time interval Δt. The *increment* $\Delta W(t)$ is a gaussian random variable of mean zero, and variance

$$\begin{aligned} E\{\Delta W(t)^2\} &= E\{(W(t+\Delta t) - W(t))^2\} \\ &= E\{W(t+\Delta t)^2\} - 2E\{W(t+\Delta t)W(t)\} + E\{W(t)^2\} \\ &= t + \Delta t - 2t + t \\ &= \Delta t\,. \end{aligned} \tag{3.10}$$

The change in $W(t)$ over a short time interval Δt is, on average, of the order $\sqrt{\Delta t}$. This suggests that the sample paths of Brownian motion are continuous functions—which is in fact true—but that they are highly irregular. These properties of Brownian motion are seen in Figure 3. The irregularity over short time scales results from the fact that the increments of the Wiener process over disjoint intervals are independent random variables. Jointly gaussian (with mean zero) random variables are independent if their covariance vanishes, and for $t > s + \Delta t$ the jointly gaussian random variables $\Delta W(t)$ and $\Delta W(s)$ satisfy

$$\begin{aligned} E\{\Delta W(t)\Delta W(s)\} &= E\{[W(t+\Delta t) - W(t)][W(s+\Delta t) - W(s)]\} \\ &= E\{W(t+\Delta t)W(s+\Delta t) - W(t)W(s+\Delta t) \\ &\quad - W(t+\Delta t)W(s) + W(t)W(s)\} \\ &= (s+\Delta t) - (s+\Delta t) - s + s \\ &= 0\,. \end{aligned} \tag{3.11}$$

If we consider the approximate derivative $\Delta W(t)/\Delta t$ of the Wiener process as a stochastic process itself, with discrete index set (time), then it is a white-noise process as defined earlier. This white-noise process consists of independent, identically distributed, gaussian random variables with mean zero and variance

$$E\left\{\left(\frac{\Delta W}{\Delta t}\right)^2\right\} = \frac{\Delta t}{\Delta t^2} = \frac{1}{\Delta t}\,. \tag{3.12}$$

If we try to think of the $\Delta t \to 0$ limit of the approximate derivatives of the Wiener process as a stochastic process itself, then we run into trouble because the variance in Eq. (3.12) diverges as $\Delta t \to 0$. In Figure 5(a), (b), and (c) we plot the approximate derivatives—of the particular realization in Figure 3—for several values of the time interval Δt, showing their divergence. The derivative of Brownian motion does not make sense as an ordinary stochastic process as we discussed in the last section, because it is not a well-defined random variable for each time t; the calculation above shows that it is a "gaussian random variable with infinite variance." The sample paths of the Wiener process are continuous but nondifferentiable functions.

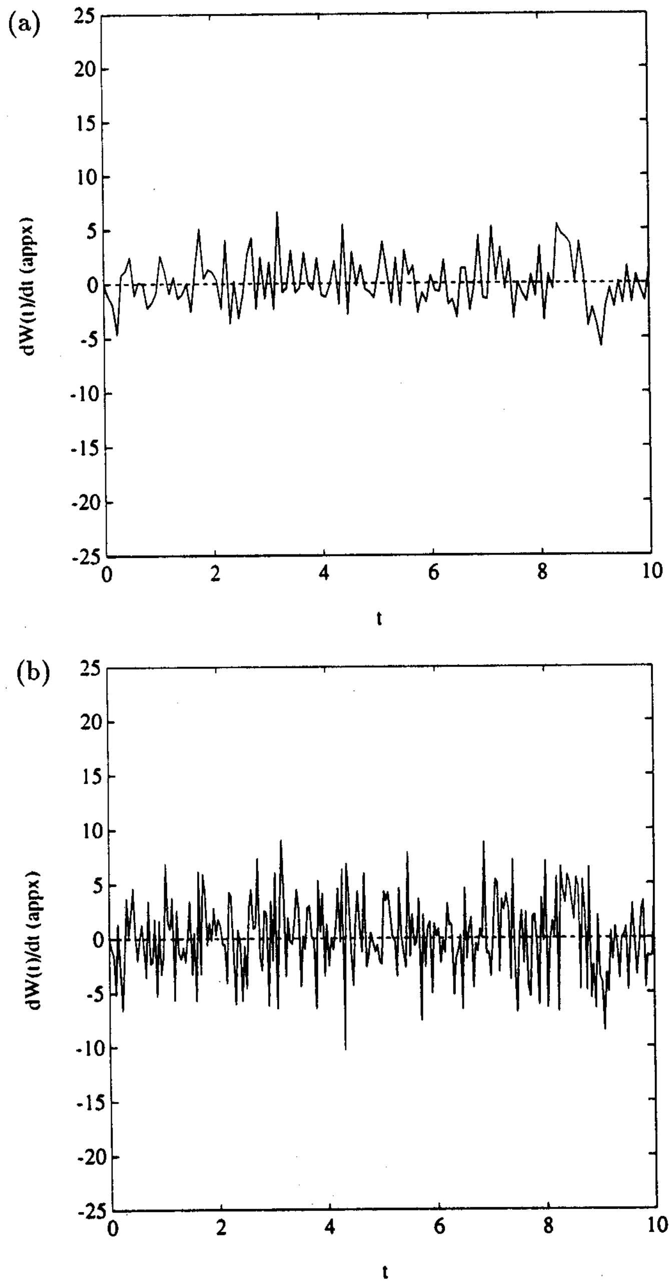


FIGURE 5
Approximate derivatives—finite differences—of the Wiener process. These are for the realization shown in Figure 3, and are for time steps (a) $\Delta t = 0.08$, (b) $\Delta t = 0.04$, (c) $\Delta t = 0.02$, and (d) $\Delta t = 0.01$.

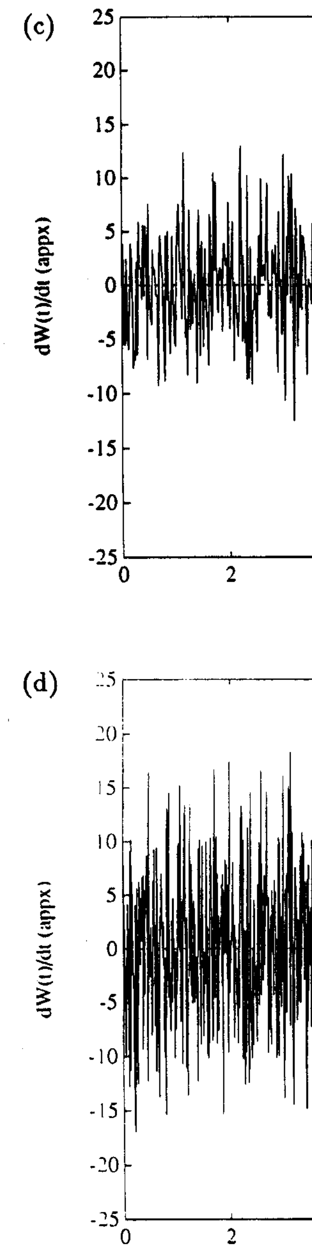
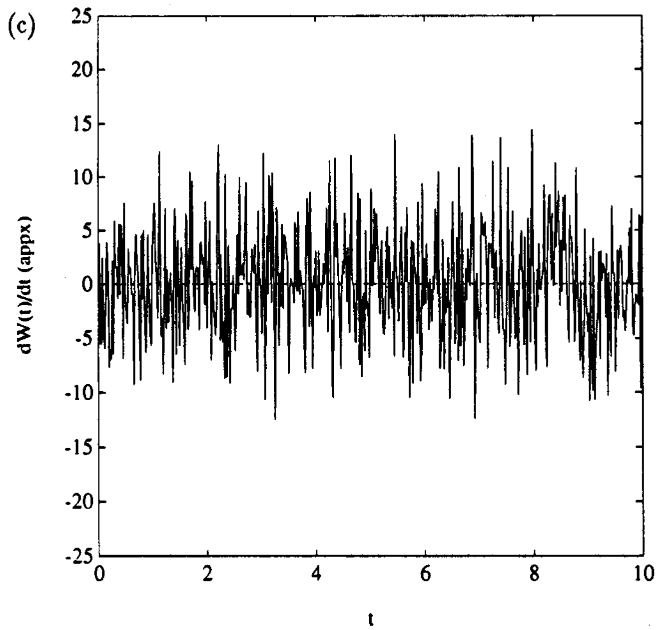


FIGURE 5
Approximate derivatives—finite differences—of the Wiener process. These are for the realization shown in Figure 3, and are for time steps (a) $\Delta t = 0.08$, (b) $\Delta t = 0.04$, (c) $\Delta t = 0.02$, and (d) $\Delta t = 0.01$.



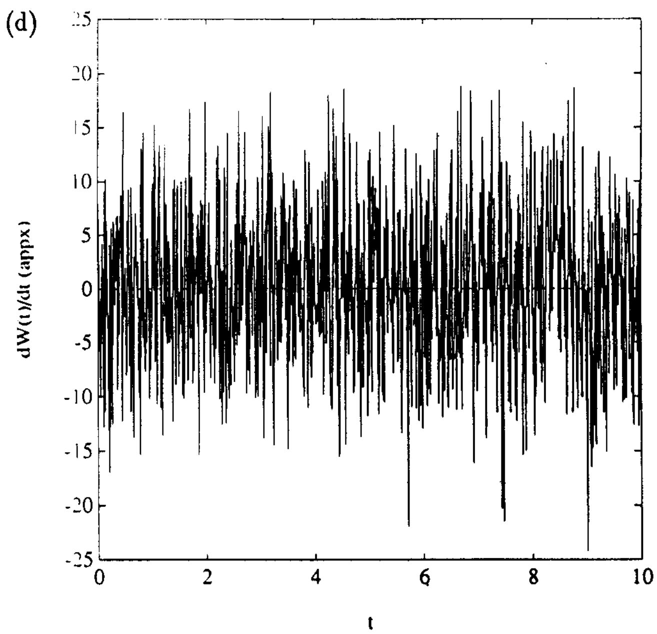


FIGURE 5
(continued)

We can make sense of the derivative of the Wiener process as a *generalized stochastic process.* That is, rather than considering the stochastic process as a collection of random variables indexed by time, we enlarge the index set to a space of functions. The idea here is that of a distribution-valued random process $\xi(\cdot)$, where in order to get a well-defined random variable we need to specify not just a number (like t) but a whole function $f(t)$. Think of the *test function* $f(t)$ as the linear response function of a probe which is sensitive to the value of ξ, and $\xi(f)$ as a measurement of ξ according to

$$\xi(f) = \int \xi(t) f(t) dt \,. \tag{3.13}$$

For smooth, bounded, and rapidly vanishing functions f, this integral makes perfect sense if the sample paths of $\xi(\cdot)$ are (almost surely) locally integrable functions of t; it even makes sense if the realizations of $\xi(\cdot)$ are more general objects like δ-functions. The more restricted the space of test functions, the more pathological the objects ξ may be. On the other hand, if the space of test functions includes distributions like the δ-functions, then ξ must be defined pointwise and we recover our original definition of the stochastic process.

The derivative of the Wiener process can be interpreted in this "smeared" way by noting that if $\xi = dW/dt$, then ξ should be gaussian and mean zero (because every approximation to it is) with variance

$$\begin{aligned}
E\{\xi(f)^2\} &= E\left\{\left[\int_0^\infty \frac{dW}{dt} f(t) dt\right]^2\right\} \\
&= E\left\{\left[-\int_0^\infty \frac{df}{dt} W(t) dt\right]^2\right\} \qquad \text{(integrating by parts)} \\
&= \int_0^\infty dt \int_0^\infty ds E\{W(t)W(s)\} f'(t) f'(s) \\
&\qquad \text{(interchanging integrals and the expectation)} \\
&= \int_0^\infty dt \int_0^\infty ds \min(t,s) f'(t) f'(s) \\
&= -\int_0^\infty f'(t) dt \int_0^t f(s) ds \qquad \text{(integrating by parts in the } s \text{ integral)} \\
&= \int_0^\infty f(t)^2 dt \,. \qquad \text{(integrating by parts once more)}
\end{aligned} \tag{3.14}$$

In the above we assume that the test function $f(t)$ vanishes fast enough as $t \to \infty$ to allow the integrations by parts without introducing any boundary terms. Thus, the variance of the derivative of Brownian motion *smeared* with a square integrable function $f(t)$ is finite and $\xi(f)$ may be interpreted as a perfectly well-behaved, gaussian random variable.

The white-noise process $\xi(\cdot)$, defined as the derivative of the Wiener process in the generalized sense described above, is known as *gaussian white noise.* Often we (and many others, to be sure) will write gaussian white noise as if it was an ordinary gaussian stochastic process, $\xi(t)$, with the moments

$$\begin{aligned} E\{\xi(t)\} &= 0\,, \\ E\{\xi(t)\xi(s)\} &= \delta(t-s)\,. \end{aligned} \tag{3.15}$$

The δ-function covariance indicates that $\xi(t)$ is not really a random variable for each t, but this is what we must put in to reproduce the calculation in Eq. (3.14) in a shorthand notation:

$$\begin{aligned} E\{\xi(f)^2\} &= E\left\{\left[\int_0^\infty \xi(t)f(t)dt)\right]^2\right\} \\ &= \int_0^\infty dt \int_0^\infty ds E\{\xi(t)\xi(s)\} f(t)f(s) \\ &= \int_0^\infty dt \int_0^\infty ds\delta(t-s)f(t)f(s) \\ &= \int_0^\infty f(t)^2 dt\,. \end{aligned} \tag{3.16}$$

Gaussian white noise is not just an abstract mathematical construction. We may think of it as the continuous time limit of a discrete time model of some stochastic process which is very rapidly varying in time, so much so that its values at successive instants are essentially independent. The gaussian statistics of the process makes it very useful as a model of processes which result from the sum of many nearly independent random effects: the central limit theorem then ensures, under very general conditions, that the resulting random variables will be normally distributed.

EXERCISE 3. Show that the transition density for the Wiener process, Eq. (3.2), for $t > s$, satisfies the diffusion equation

$$\partial_t \rho(x,t|y,s) = \frac{1}{2}\partial_x^2 \rho(x,t|y,s) \tag{3.17}$$

and that

$$\lim_{t\downarrow s} \rho(x,t|y,s) = \delta(x-y)\,. \tag{3.18}$$

4. STOCHASTIC DIFFERENTIAL EQUATIONS AND FOKKER-PLANCK EQUATIONS

We have discussed one specific example of a continuous time Markov process at this point, Brownian motion $W(t)$. Brownian motion is in the class of continuous time Markov processes whose sample paths are (almost surely) continuous but not differentiable functions, called ***diffusion processes.*** This is a natural name for these processes because, as we will see below, their transition densities satisfy partial differential equations related to the diffusion equation. For example, by direct calculation in Exercise 3, we found that the transition density for the Wiener process satisfies the diffusion equation

$$\partial_t \rho(x,t|y,s) = \frac{1}{2}\partial_x{}^2 \rho(x,t|y,s), \tag{4.1}$$

with the initial condition

$$\rho(x,s|y,s) = \delta(x-y) \tag{4.2}$$

at time $t = s$. In this section we will show how Markov diffusion processes arise naturally as the solutions of differential equations with white-noise coefficients, i.e., *stochastic differential equations.* Moreover, we will show how to write down the partial differential equation for the transition density, generally called the *Fokker-Planck equation*, starting from the stochastic differential equation.

Brownian motion may be described as the (one-dimensional) position as a function of time of a particle whose velocity is gaussian white noise. That is, by definition of gaussian white noise as the derivative of the Wiener process, $\xi(t) = dW(t)/dt$, we may consider the reverse logical order of this equation, writing

$$\frac{dW(t)}{dt} = \xi(t), \tag{4.3}$$

and think of $W(t)$ as the solution of this stochastic differential equation. The associated Fokker-Planck equation for transition density of the solution is exactly Eq. (4.1) and, given some initial conditions, its solution contains all possible information about the stochastic process (remember, from Eq. (2.11), that all the joint density functions are built up from the transition density and the initial one-point density for Markov processes).

A more general stochastic differential equation is one where the derivative of the solution depends on the solution itself and the white-noise process appears as a coefficient. Because gaussian white noise is really a generalized stochastic process, it makes no sense to perform nonlinear operations on it, so the most general such stochastic differential equation for a process $X(t)$ is

$$\frac{dX(t)}{dt} = f\big(X(t),t\big) + g\big(X(t),t\big)\xi(t)\,. \tag{4.4}$$

This is a model of a system subject to a deterministic driving force, $f(X,t)$, as well as a rapidly fluctuating random force, $g(X,t)\xi(t)$, whose influence depends both on the state of the system and on the time. For locations in the state space and intervals of time where g is rather large in absolute value, the fluctuations ξ have more influence on the dynamical variable's evolution than those locations and times where it is small in absolute value. Such white-noise-driven differential equations are often called *Langevin equations* in the physics and chemistry literature. In the specific case $f = 0$ and $g = 1$, Eq. (4.4) reduces to Eq. (4.3) and we know that its corresponding Fokker-Planck equation is Eq. (4.1). Our task now is to start from Eq. (4.4), stating somewhat more carefully how we will interpret it, and then to derive the Fokker-Planck equation for the transition density of the process $X(t)$.

We will consider the stochastic differential equation Eq. (4.4) to be the continuous time limit of the discrete time problem

$$\Delta X(t) = X(t+\Delta t) - X(t) = f(X(t),t)\Delta t + g(X(t),t)\Delta W(t)\,, \tag{4.5}$$

where $\Delta W(t) = W(t+\Delta t) - W(t)$ is the increment of the Wiener process in the time interval between t and $t+\Delta t$. This is the kind of discretization that we could contemplate solving numerically on a computer. If we divide through by Δt and formally take the limit $\Delta t \to 0$, we recover Eq. (4.4). This discrete formulation also explicitly displays the Markov character of the solution process $X(t)$: if the process is given (i.e., if its value is specified) at time t_0, then we may evolve the process some integral number of time steps into the future using only the statistically independent random variables $\Delta W(t)$ which we generate along the way, completely independently of the past history of the process. Thus the future of $X(t_0)$, i.e., $X(t_0 + n\Delta t)$ is independent of the past when we are given the present value of the process. As the Markov property holds for each discretization of $X(t)$ with ever decreasing intervals Δt, it is not surprising that this characteristic survives the limit to continuous time.

We use the time-sliced formulation in Eq. (4.5) to view the information contained in the stochastic differential equation from a slightly different angle. Here we need to introduce the concept of the *conditional expectation* of a function of a stochastic process. For a Markov process $X(t)$ with transition density $\rho(x,t|y,s)$, we define the "expectation of $F(X(t))$ given $X(s) = y$" for $t > s$, denoted $E\{F(X(t))|X(s) = y\}$, as the expectation computed using the conditional probability density:

$$E\{F(X(t))|X(s) = y\} = \int F(x)\rho(x,t|y,s)dx\,, \tag{4.6}$$

where, as usual, the integral covers the process' state space. The concept is simple; we ask for the average value of $F(X(t))$ over those realizations of the process that started at position y at time s. In particular, for $s = t$, the process is taking on the given value, $X(s) = y$, so that

$$\begin{aligned} E\{F(X(t))|X(t) = y\} &= \int F(x)\rho(x,t|y,t)dx \\ &= \int F(x)\delta(x-y)dx = F(y)\,. \end{aligned} \tag{4.7}$$

Now we may use the discrete formula Eq. (4.5) to find the conditional expectation of the jump in $X(t)$ given its starting point:

$$\begin{aligned} &E\{\Delta X(t)|X(t) = x\} \\ &\quad = E\{X(t+\Delta t)|X(t) = x\} - E\{X(t)|X(t) = x\} \\ &\quad = \int (x'-x)\rho(x',t+\Delta t|x,t)dx' \\ &\quad = E\{f(X(t),t)\Delta t|X(t) = x\} + E\{g(X(t),t)\Delta W(t)|X(t) = x\}\,. \end{aligned} \tag{4.8}$$

Clearly, $E\{f(X(t),t)\Delta t|X(t) = x\} = f(x,t)\Delta t$. Also, because the increment of the Wiener process $\Delta W(t)$ is *independent* of $X(t)$ and all the previous history of the process,

$$\begin{aligned} E\{g(X(t),t)\Delta W(t)|X(t) = x\} &= E\{g(X(t),t)|X(t) = x\}E\{\Delta W(t)|X(t) = x\} \\ &= g(x,t)E\{\Delta W(t)\} \\ &= 0\,. \end{aligned} \tag{4.9}$$

Hence, the expectation of the increment in the process $X(t)$ is

$$E\{\Delta X(t)|X(t) = x\} = f(x,t)\Delta t\,. \tag{4.10}$$

We also need the second moment of the increments of the process, and these can all be computed from the difference equation. Indeed, recalling that the variance of ΔW is Δt,

$$\begin{aligned} E\{\Delta X(t)^2|X(t) = x\} &= \int (x'-x)^2\rho(x',t+\Delta t|x,t)dx' \\ &= E\left\{\left[f(X(t),t)\Delta t + g(X(t),t)\Delta W(t)\right]^2|X(t) = x\right\} \\ &= f(x,t)^2\Delta t^2 + 2f(x,t)\Delta t g(x,t)E\{\Delta W(t)\} \\ &\quad + g(x,t)^2E\{\Delta W(t)^2\} \\ &= g(x,t)^2\Delta t + O(\Delta t^2)\,. \end{aligned} \tag{4.11}$$

All the higher jump moments are of higher than linear order in the time increment Δt. For our purposes, the stochastic differential equation $dX/dt = f + g\xi$ will be interpreted in terms of the jump moments derived above; $X(t)$ is a Markov process whose first and second jump moments are of order Δt and specified as in Eq. (4.10) and Eq. (4.11), with all higher moments of higher order in Δt.

To derive the Fokker-Planck equation, we begin with the assumption that the process $X(t)$ possesses a differentiable transition density, and the Chapman-Kolmogorov relation

$$\rho(x,t|y,s) = \int \rho(x,t|z,u)\rho(z,u|y,s)dz \qquad \text{for } s < u < t\,, \tag{4.12}$$

which is simply a rewriting of Eqs.(2.3) and (2.11). This says that the probability to go from y at time s, to x at time t, is the sum of the probabilities of the mutually exclusive events of making the transition via different points z at the intermediate time u. The probability of each of these events is the product of the probabilities of two independent events for Markov processes.

Introduce an arbitrary smearing function $R(x)$ on the state space which is smooth (so we have as many derivatives as we wish) and rapidly vanishing at the boundaries of the state space (so that no boundary terms arise from integrations by parts). Then, for $t > s$ and $\Delta t > 0$, the Chapman-Kolmogorov relation says that the conditional expectation of $R(X(t+\Delta t))$ given $X(s) = y$ is

$$\begin{aligned}\int R(x)\rho(x,t+\Delta t|y,s)dx &= \int dx \int dz \rho(x,t+\Delta t|z,t)\rho(z,t|y,s)R(x) \\ &= \int dz\rho(z,t|y,s)\int dx R(x)\rho(x,t+\Delta t|z,t)\,.\end{aligned} \tag{4.13}$$

Now expand $R(x)$ in a Taylor series around z. The motivation here is that for small Δt, $\rho(x,t+\Delta t|z,t)$ is "almost" $\delta(x-z)$, so that only the values of $R(x)$ near $x = z$ are significant in the second integral above. Using the normalization of the transition density,

$$\int \rho(x,t+\Delta t|z,t)dx = 1\,,$$

we have

$$\begin{aligned}&\int R(x)\rho(x,t+\Delta t|y,s)dx \\ &\quad = \int dz\rho(z,t|y,s)R(z) \\ &\quad\quad + \int dz\rho(z,t|y,s)R'(z)\int dx(x-z)\rho(x,t+\Delta t|z,t) \\ &\quad\quad + \int dz\rho(z,t|y,s)\frac{1}{2}R''(z)\int dx(x-z)^2\rho(x,t+\Delta t|z,t) + \ldots\end{aligned} \tag{4.14}$$

We have now cast the conditional expectation of $R(X(t+\Delta t))$ in terms of conditional expectations of $R(X(t))$ and its derivatives, and the jump moments for the process. The jump moments were computed from the stochastic differential equation above, and inserting Eqs.(4.10) and (4.11) into Eq. (4.14) above,

$$\int R(x)\rho(x,t+\Delta t|y,s)dx = \int dz\rho(z,t|y,s)\Big\{R(z) + R'(z)f(z,t)\Delta t + \frac{1}{2}R''(z)g(z,t)^2\Delta t + O(\Delta t^2)\Big\}. \tag{4.15}$$

Change the z to an x on the right-hand side above, place all the terms on one side of the equality sign, and divide through by Δt:

$$0 = \int dx\Big\{R(x)\frac{\rho(x,t+\Delta t|y,s) - \rho(z,t|y,s)}{\Delta t} - \left[R'(x)f(x,t) + \frac{1}{2}R''(x)g(x,t)^2 + O(\Delta t)\right]\rho(x,t|y,s)\Big\}. \tag{4.16}$$

In the limit $\Delta t \to 0$, we find that for any function $R(\cdot)$,

$$0 = \int dx\left\{R(x)\frac{\partial\rho(x,t|y,s)}{\partial t} - \left[R'(x)f(x,t) + \frac{1}{2}R''(x)g(x,t)^2\right]\rho(x,t|y,s)\right\}. \tag{4.17}$$

Integrate by parts in the last two terms above to find

$$0 = \int dx R(x)\left\{\partial_t\rho(x,t|y,s) + \partial_x[f(x,t)\rho(x,t|y,s)] - \frac{1}{2}\partial_x{}^2[g(x,t)^2\rho(x,t|y,s)]\right\}. \tag{4.18}$$

Because R is an arbitrary function, the term in brackets must vanish so the transition density $\rho(x,t|y,s)$ satisfies the Fokker-Planck equation

$$\partial_t\rho(x,t|y,s) = \left[-\partial_x f(x,t) + \frac{1}{2}\partial_x{}^2 g(x,t)^2\right]\rho(x,t|y,s). \tag{4.19}$$

(The reader is reminded at this point that the differential operator ∂_x acts on everything to their right.) Along with this evolution equation for $\rho(x,t|y,s)$ goes the initial condition at time $t = s$

$$\rho(x,t=s|y,s) = \delta(x-y). \tag{4.20}$$

This is the central goal of this section. The solution to the stochastic evolution in Eq. (4.4) is a Markov process defined by its transition density which satisfies the Fokker-Planck equation in Eq. (4.19). This is really how we are interpreting the random dynamical evolution law in the original stochastic differential equation. The solution of the associated Fokker-Planck equation provides us with the transition density defining the Markov process which we take to be the solution $X(t)$.

The natural generalization of these considerations allows us to formulate the same ideas for stochastic processes which are not just scalar valued, but also multicomponent objects. For example, let ξ_i, $i = 1, \ldots, n$, be n independent gaussian white-noise processes, and $\mathbf{X}(t) \in R^n$ be an n-dimensional vector-valued process with components $X_i(t)$ satisfying the stochastic differential equations

$$\frac{dX_i}{dt} = f_i(\mathbf{X}) + g_{ij}(\mathbf{X})\xi_j \,. \tag{4.21}$$

Here we use the summation convention on repeated indices; $\mathbf{f}(\mathbf{X})$ is a vector-valued function of $\mathbf{X}$ with components f_i and $g_{ij}(\mathbf{X})$ are the components of an $n \times n$ matrix-valued function of $\mathbf{X}$. The transition density $\rho(\mathbf{x}, t|\mathbf{y}, s)$ for the vector-valued Markov process $\mathbf{X}(t)$ then satisfies the Fokker-Planck equation

$$\frac{\partial \rho(\mathbf{x}, t|\mathbf{y}, s)}{\partial t} = \left\{ -\frac{\partial}{\partial x_i} f_i(\mathbf{x}) + \frac{1}{2} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} D_{ij}(\mathbf{x}) \right\} \rho(\mathbf{x}, t|\mathbf{y}, s) \tag{4.22}$$

where the positive semi-definite diffusion matrix $D_{ij}(\mathbf{x})$ is

$$D_{ij} = g_{ik} g_{jk} \,, \tag{4.23}$$

and the transition density satisfies the initial condition

$$\rho(\mathbf{x}, t|\mathbf{y}, s) = \delta(\mathbf{x} - \mathbf{y}) \,. \tag{4.24}$$

Along with the Fokker-Planck equation and its initial condition go some boundary conditions. The issue of boundary conditions for these processes is rather involved, and we will for the most part neglect the subtle issues here. Our practical criteria will be the condition that a proposed solution to the Fokker-Planck equation is positive and integrable, and thus interpretable as a probability density. We refer the reader to the discussion of boundary conditions in Horsthemke and Lefever[3] for details.

EXERCISE 4. The Ornstein-Uhlenbeck process, $U(t)$, is defined by the stochastic differential equation

$$\frac{dU}{dt} = -\gamma U + \sigma \xi \tag{4.25}$$

with γ and σ constant parameters. It is Newton's law for the acceleration (dU/dt) of a particle in some medium subject to frictional retarding force proportional to its velocity $(-\gamma U)$ and a rapidly fluctuating random force due to collisions with the "fluid" particles in the medium $(\sigma\xi)$.

a. Show that the associated Fokker-Planck equation is

$$\frac{\partial \rho(x, t|y, s)}{\partial t} = \left[\gamma \frac{\partial}{\partial x} x + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2} \right] \rho(x, t|y, s) \,. \tag{4.26}$$

b. Verify that the transition density is

$$\rho(x,t|y,s) = \frac{1}{\sqrt{2\pi\Sigma(t-s)}} \exp\left\{-\frac{(x-ye^{-\gamma(t-s)})^2}{2\Sigma(t-s)}\right\} \tag{4.27}$$

where

$$\Sigma(t) = \frac{\sigma^2}{2\gamma}(1-e^{-2\gamma t}). \tag{4.28}$$

c. The *stationary state* of the Ornstein-Uhlenbeck process is achieved as $t \to \infty$. It is a gaussian Markov process with the time-independent one-time density (the *stationary* density ρ_{stat})

$$\rho_{\text{stat}}(x) = \lim_{t\to\infty} \rho(x,t) = \sqrt{\frac{\gamma}{\pi\sigma^2}} \exp\left\{-\frac{\gamma x^2}{\sigma 2}\right\}. \tag{4.29}$$

Show that the covariance in the stationary state is

$$\begin{aligned} E\{U(t)U(s)\} &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dyxy\rho(x,t|y,s)\rho_{\text{stat}}(y) \\ &= \frac{\sigma^2}{2\gamma} e^{-\gamma|t-s|}. \end{aligned} \tag{4.30}$$

The exponential decay time $\tau = \gamma^{-1}$ of the correlation function is called the *correlation time* of the process.

5. APPLICATION: THE VERHULST MODEL OF POPULATION DYNAMICS

The example we will develop here is a simple model of population dynamics. Let $X(t) > 0$ be the population of the species, obeying the *Verhulst equation*

$$\frac{dX}{dt} = \mu X - X^2. \tag{5.1}$$

The linear term on the right-hand side is the net rate of change of population due to birth and death; the growth rate μ is the difference between the birth and death rates. The nonlinear saturation term roughly models the effect of overcrowding which limits the total population. We have chosen the units so that the coefficient of the nonlinear term is 1.

The deterministic dynamics of this equation are very simple. If $\mu < 0$ (i.e., death rate > birth rate), then the population always dies out because the right-hand side of Eq. (5.1) is always negative. Thus $X = 0$ is a stable steady state of the system for $\mu < 0$, and it is in fact the attractor. All solutions approach this attracting state at a uniform asymptotic rate:

$$X(t) \sim e^{\mu t} \to 0, \qquad \text{as } t \to \infty \text{ for } \mu < 0\,. \tag{5.2}$$

If $\mu > 0$, then birth outpaces death and a small population will grow—at first exponentially—until the nonlinear saturation takes over. On the other hand, if the population starts out very large, then the nonlinear drives it down until it balances with the positive growth rate. Both of these cases are illustrated in Figure 6. The solution $X = 0$ is linearly unstable when $\mu > 0$, and the steady new solution $X = \mu$ is a stable fixed point. All solutions are attracted to this state at a steady rate,

$$|X(t) - \mu| \sim e^{-\mu t} \to 0, \qquad \text{as } t \to \infty \text{ for } \mu > 0\,. \tag{5.3}$$

The long time dynamics are simply summarized in the bifurcation diagram in Figure 7, where we show the stable steady population as a function of μ.

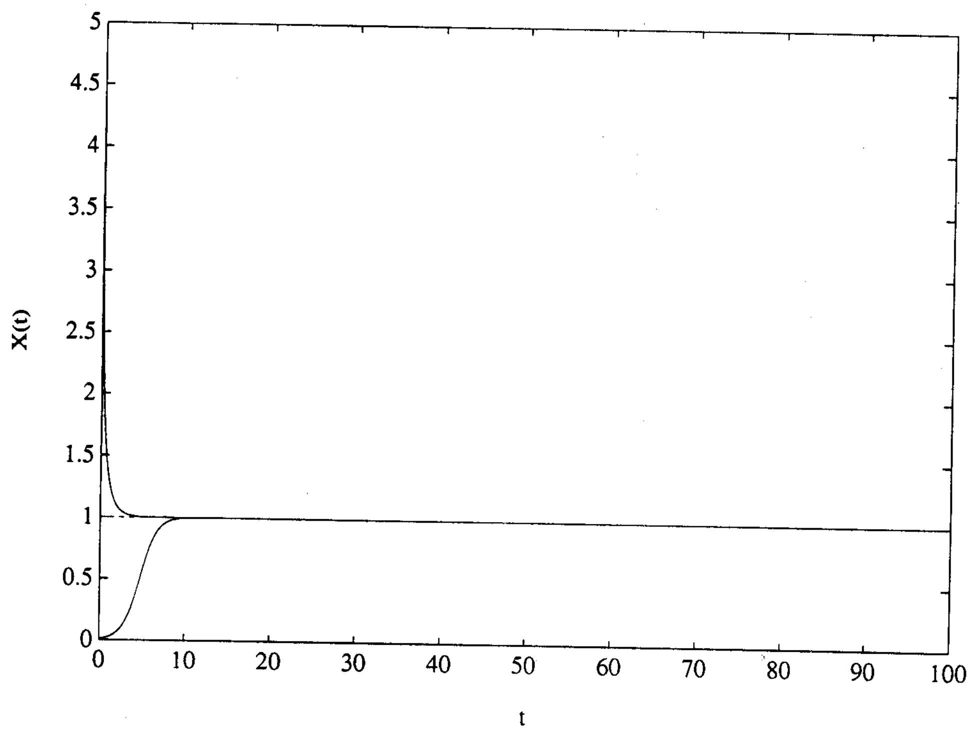


FIGURE 6 Two solutions of the deterministic Verhulst equation, starting from $X(0) = 5$ (top) and $X(0) = .01$ (bottom). The growth rate is $\mu = 1$. At long times the solutions approach the steady state $X = \mu$.

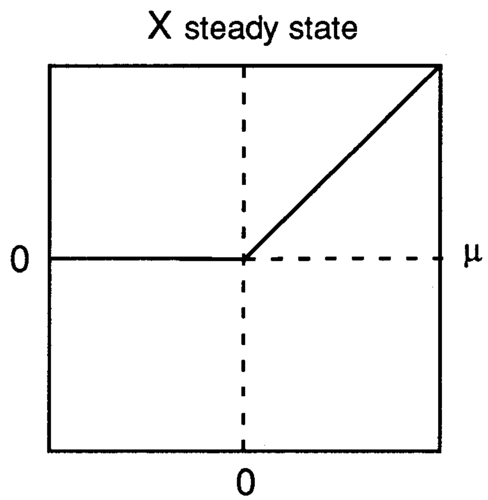


FIGURE 7 Bifurcation diagram of the stable steady state for the Verhulst equation.

Now the question is the following: what is the effect of environmental fluctuations on this population? That is, imagine that the birth and death rates are not constant (more specifically, their difference μ is not constant), but rather they fluctuate from generation to generation due to environmental effects like the weather, predator populations, disease, etc. We will model these fluctuations as essentially random from the point of view of the species under study, so the "noisy" growth rate $\mu(t)$ is a stochastic process. Specifically, we write

$$\mu(t) = \langle\mu\rangle + \sigma\xi(t)\,, \tag{5.4}$$

where $\langle\mu\rangle$ is the time averaged rate, σ is a parameter that we will refer to as the noise amplitude, and ξ is gaussian white noise (normalized as usual by $E\{\xi(t)\xi(s)\} = \delta(t-s)$). Physically, in modeling the fluctuations as white noise we are assuming that the variations are very fast on the system's deterministic time scale, given by the relaxation time $\langle\mu\rangle^{-1}$. Our stochastic differential equation for the population is then

$$\frac{dX}{dt} = \langle\mu\rangle X - X^2 + \sigma X\xi(t)\,, \tag{5.5}$$

which we interpret, as in the last section, as the continuous time limit of the discrete time process defined by

$$\Delta X(t) = (\langle\mu\rangle X - X^2)\Delta t + \sigma X \Delta W(t)\,. \tag{5.6}$$

This discrete time process is already a sensible model if we consider nonoverlapping generations with environmental fluctuations affecting only the birth rate. The continuous time limit then describes the population dynamics on time scales much longer than the life of any one generation.

Let us begin by mal
ical simulations of the d
small amount of noise si
solution, as illustrated i
behavior of the populatic
system after any initial t
steady-state population
similar to the low noise c
population is, at this noi
by the average growth ra
higher noise as shown i
lower than that expecte
substantial amount of ti
tions to high population s
as the environmental noi
when the noise is large e
time. In this case the ste

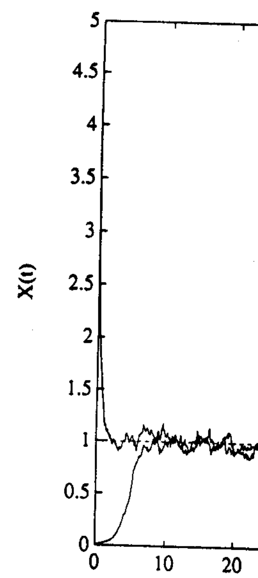


FIGURE 8 A small amount
system in Figure 6. The nois
$\langle\mu\rangle = 1$.

Let us begin by making some phenomenological observations based on numerical simulations of the discrete time process in Eq. (5.6). Not unexpectedly, a very small amount of noise simply causes the population to vary about the deterministic solution, as illustrated in Figure 8. We are interested in the long-term steady-state behavior of the population, characterized by the statistics of the stochastic dynamic system after any initial transients have died away. In Figures 9 and 10 we show the steady-state population for two higher noise values. The qualitative behavior is similar to the low noise dynamics, although it is clear in Figure 10 that the average population is, at this noise level, less than the deterministic steady population set by the average growth rate. The time series starts to look significantly different for higher noise as shown in Figures 11 and 12. Not only is the average population lower than that expected from the average growth rate, but the species spends a substantial amount of time near zero population, occasionally making large deviations to high population states. These deviations appear to be increasingly sporadic as the environmental noise amplitude is increased. In Figure 13 we observe that, when the noise is large enough, the population completely vanishes at some finite time. In this case the steady-state behavior is extinction!

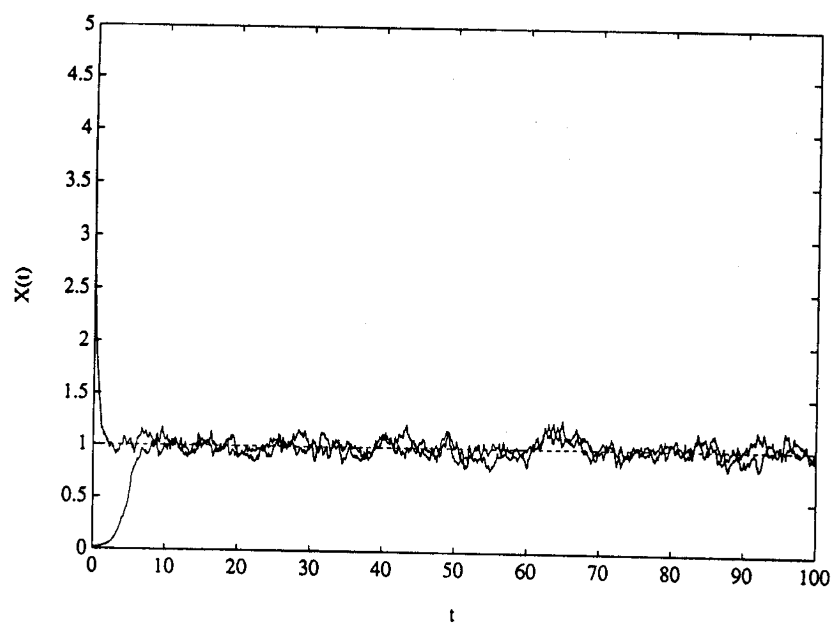


FIGURE 8 A small amount of white noise in the growth rate, a perturbation of the system in Figure 6. The noise amplitude is $\sigma^2 = .01$, and the average growth rate is $\langle\mu\rangle = 1$.

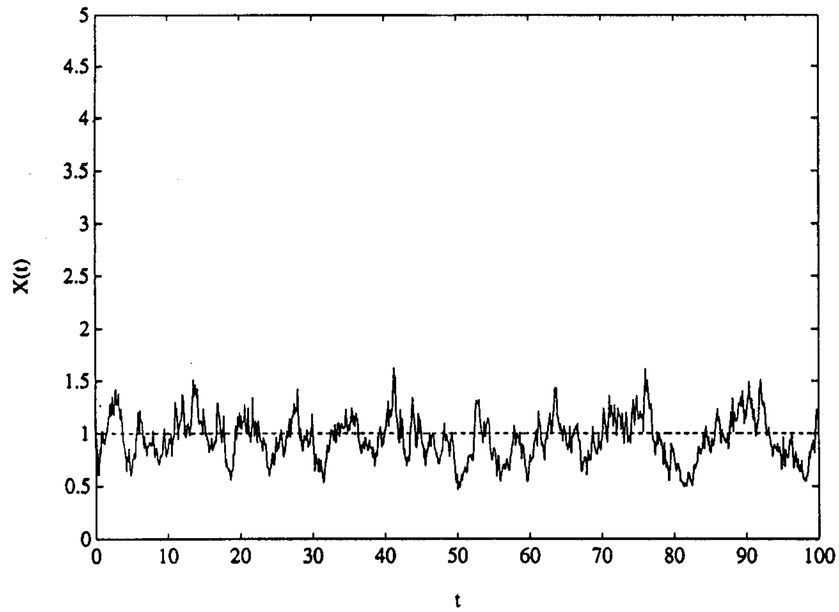


FIGURE 9 Steady-state time series for $\sigma^2 = .1$, average growth rate $\langle \mu \rangle = 1$.

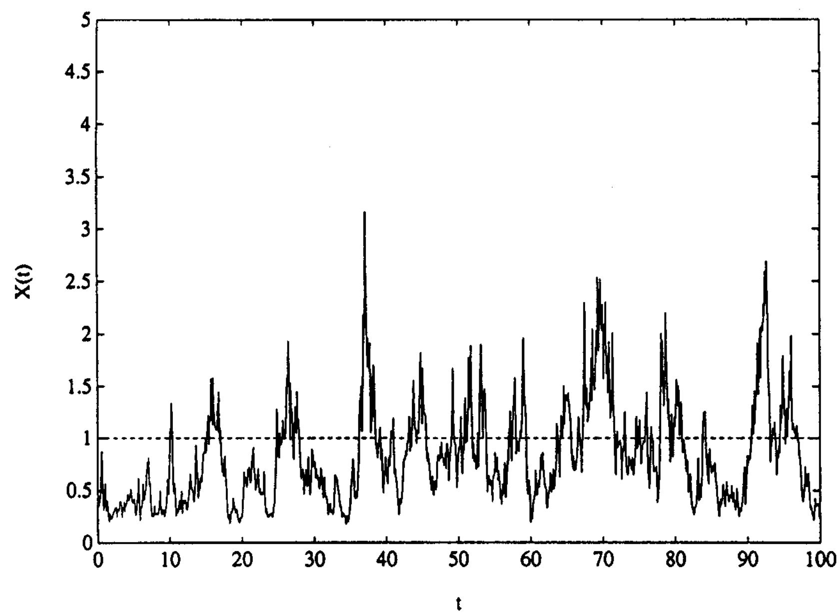


FIGURE 10 Steady-state time series for $\sigma^2 = .5$, average growth rate $\langle \mu \rangle = 1$.

FIGURE 11 Steady-state

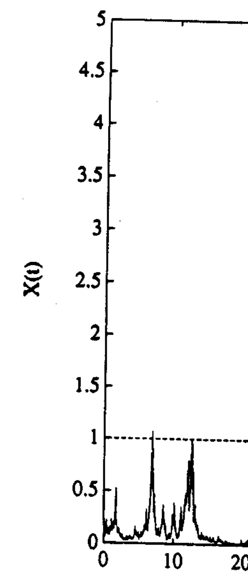


FIGURE 12 Steady-state t

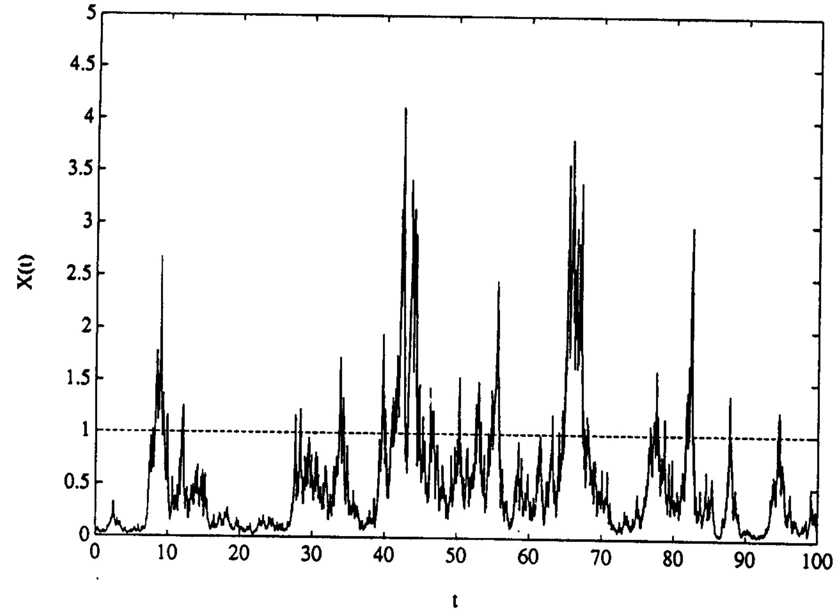


FIGURE 11 Steady-state time series for $\sigma^2 = 1$, average growth rate $\langle\mu\rangle = 1$.

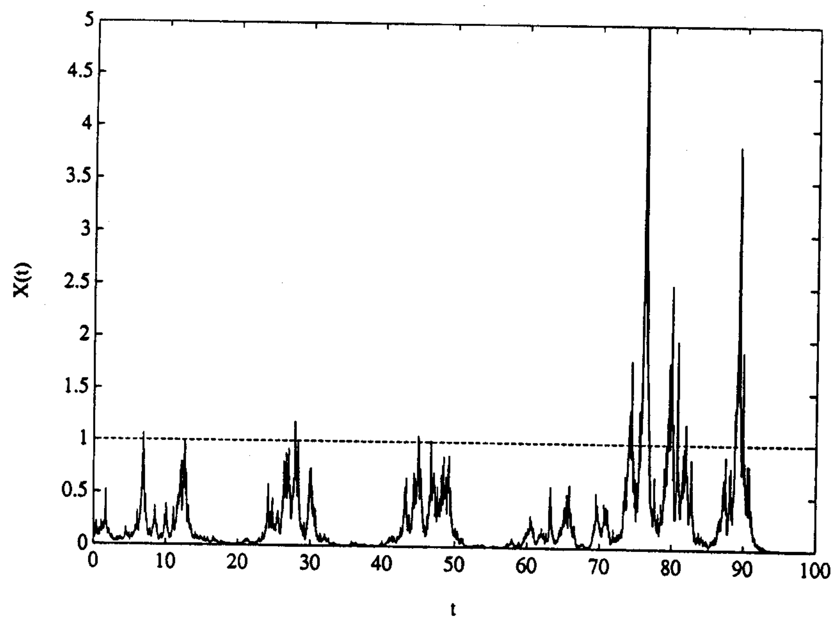


FIGURE 12 Steady-state time series for $\sigma^2 = 1.5$, average growth rate $\langle\mu\rangle = 1$.

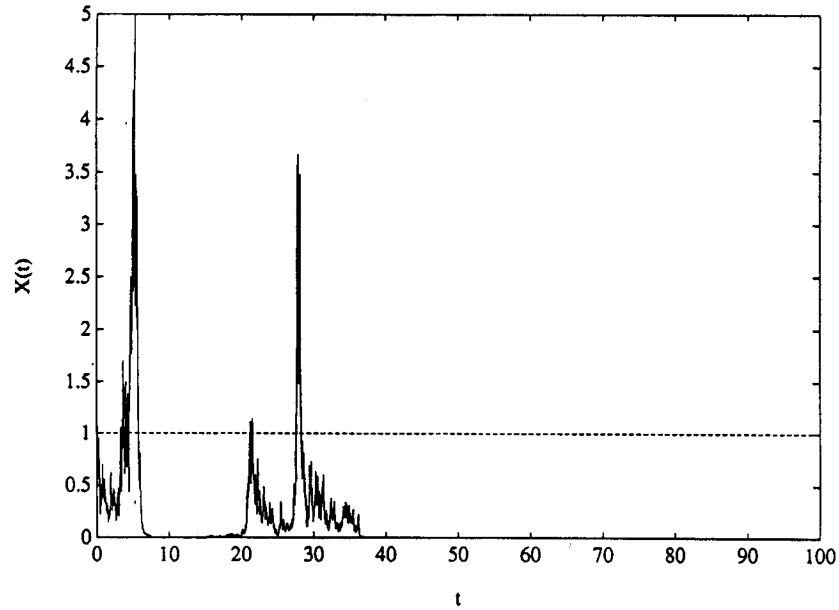


FIGURE 13 Time series for $\sigma^2 = 2$, average growth rate $\langle\mu\rangle = 1$. The population becomes extinct (in this simulation) at some time between $t = 30$ and $t = 40$.

Let us turn now to a more quantitative analysis of the process. The stochastic differential equation, Eq. (5.5), fits into the analytical framework of the last section: the drift function is $f(x) = (\langle\mu\rangle x - x^2)$, and the diffusion function is $g(x)^2 = \sigma^2 x^2$. The solution, $X(t)$, is a Markov diffusion process whose transition density, $\rho(x,t|y,s)$, obeys the Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial x}\left\{x^2 - \langle\mu\rangle x + \frac{\sigma^2}{2}\frac{\partial}{\partial x}x^2\right\}\rho\,. \tag{5.7}$$

The steady state of a stochastic process is characterized by (among other things) the stationary probability density, $\rho_{\text{stat}}(x)$, defined by

$$\rho_{\text{stat}}(x) = \lim_{t\to\infty}\rho(x,t|y,s)\,, \tag{5.8}$$

assuming that this limit makes sense as a probability distribution independent of the starting point $X(s) = y$. This steady-state probability density is an invariant density and so satisfies the time-independent Fokker-Planck equation

$$0 = \frac{\partial}{\partial x}\left\{x^2 - \langle\mu\rangle x + \frac{\sigma^2}{2}\frac{\partial}{\partial x}x^2\right\}\rho_{\text{stat}}\,. \tag{5.9}$$

This equation has the formal solution

$$\rho_{\text{stat}}(x) = N x^{2(\langle\mu\rangle/\sigma^2 - 1)} e^{-2x/\sigma^2}, \tag{5.10}$$

which is an acceptable probability density on the positive real line as long as it's integrable, i.e., so long as the normalization condition

$$1 = \int_0^\infty \rho_{\text{stat}}(x)dx = N \left(\frac{\sigma^2}{2}\right)^{(2\langle\mu\rangle/\sigma^2 - 1)} \Gamma\left(\frac{2\langle\mu\rangle}{\sigma^2} - 1\right) \tag{5.11}$$

can be satisfied for some finite, nonvanishing normalization constant N. Hence, the function in Eq. (5.10) is a stationary probability distribution for the process as long as the average growth rate and the noise amplitude satisfy

$$\frac{2\langle\mu\rangle}{\sigma^2} > 1\,. \tag{5.12}$$

This condition says that, if the environmental fluctuations are not too strong, on the scale of the average growth rate, the process may achieve a stochastic stationary state where the population $X(t)$ is described by the time-independent probability density in Eq. (5.10). This probability density is plotted for several values of the noise amplitude in Figures 14–17. The qualitative behavior of the process is clearly indicated by the features of the stationary density: at small noise the process fluctuates in a relatively tight region around its most probable value (compare Figures 9 and 10), and under the influence of larger variations, the most probable value is actually the zero population state (compare Figures 11 and 12). There is an interesting change in the population dynamics precisely when $\sigma^2 = \langle\mu\rangle$, the point at which the zero population state becomes the most probable value. This kind of qualitative change in a system's behavior, brought on by the influence of fluctuations alone—the mean growth rate is constant here—is called a *noise-induced transition*, akin to bifurcations in deterministic systems and phase transitions in equilibrium statistical mechanics, and is studied in detail in the monograph of Horsthemke and Lefever.[3]

What happens if the noise is even stronger, i.e., if the inequality in Eq. (5.12) is violated? We must keep in mind that the trivial stochastic process $X(t) = 0$ is an exact solution of the stochastic differential equation, corresponding to the stationary density $\rho_{\text{stat}}(x) = \delta(x)$. This singular distribution function will certainly describe the system if the initial population was zero, but it also describes the steady state if the population eventually dies out for any reason. In fact, this is just how this model behaves. If the amplitude of the growth rate fluctuations are too

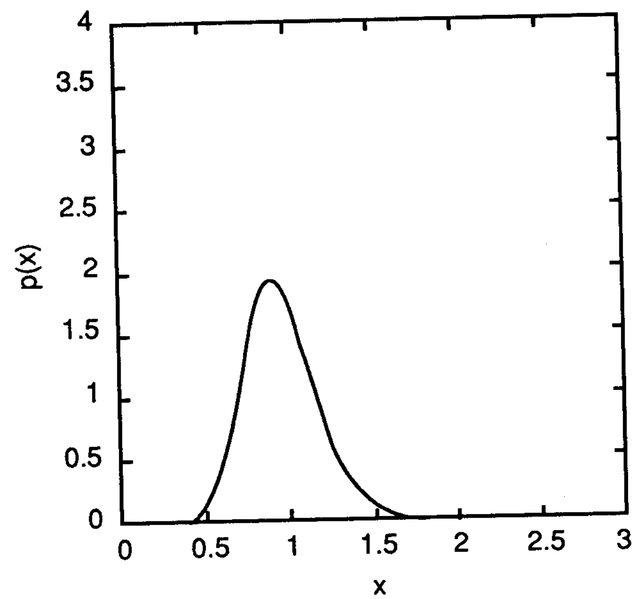


FIGURE 14 Stationary probability distribution for $\sigma^2 = .1$, average growth rate $\langle \mu \rangle = 1$.

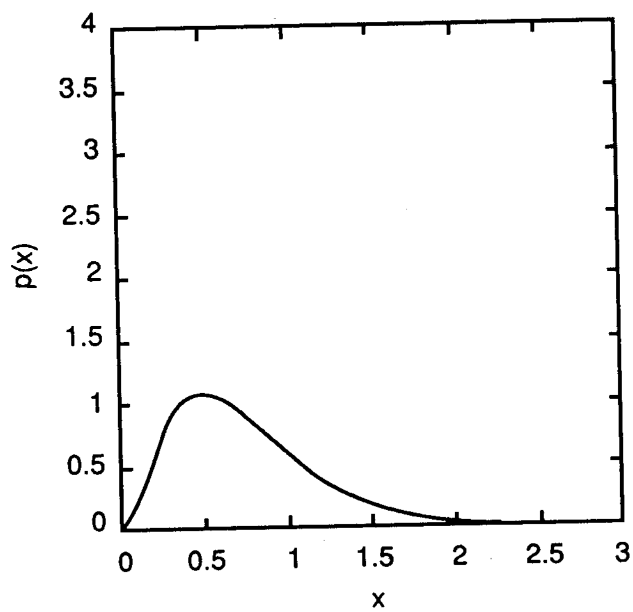


FIGURE 15 Stationary probability distribution for $\sigma^2 = .5$, average growth rate $\langle \mu \rangle = 1$.

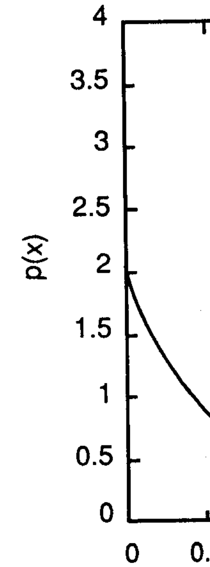


FIGURE 16 Stationary pro

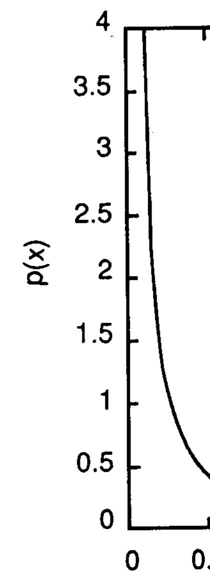
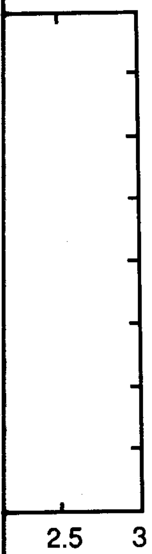
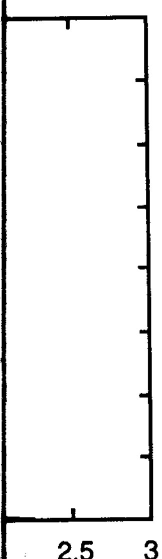


FIGURE 17 Stationary pr
$\langle \mu \rangle = 1$.



^{2} = .1, average growth rate



^{2} = .5, average growth rate

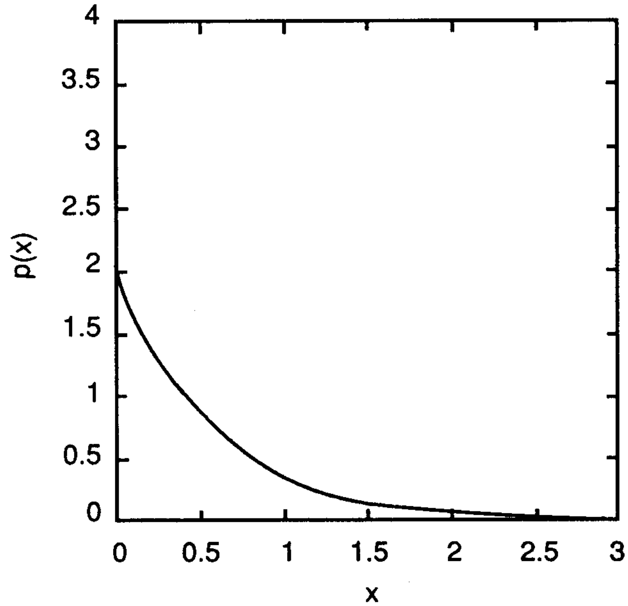


FIGURE 16 Stationary probability distribution for $\sigma^2 = 1$, average growth rate $\langle\mu\rangle = 1$.

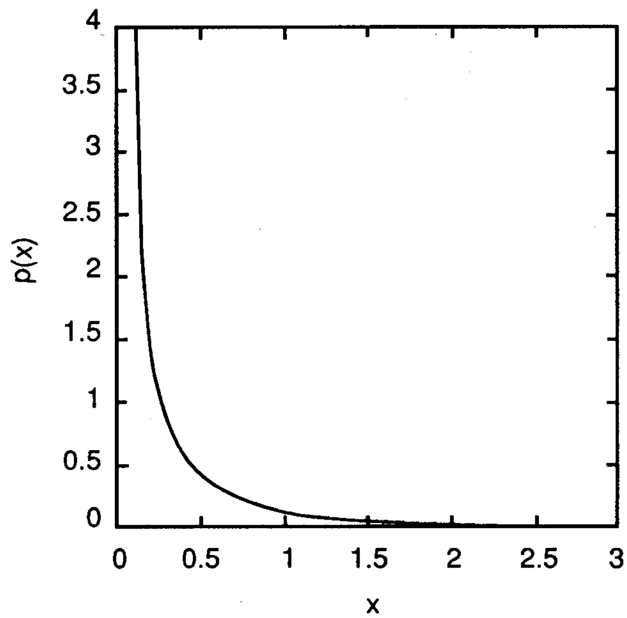


FIGURE 17 Stationary probability distribution for $\sigma^2 = 1.5$, average growth rate $\langle\mu\rangle = 1$.

strong, specifically if $\sigma^2 \geq 2\langle\mu\rangle$, then the population eventually becomes extinct and the steady state density is

$$\rho_{\text{stat}}(x) = \delta(x) \qquad (\text{for } \sigma^2 \geq 2\langle\mu\rangle)\,. \tag{5.13}$$

EXERCISE 5. Using the stationary probability density Eq. (5.10) or Eq. (5.13), compute the average steady-state population as a function of the mean growth rate and the noise amplitude.

6. EQUILIBRIUM VS. NONEQUILIBRIUM STATIONARY STATES

The Fokker-Planck equation is a continuity equation for the flow of the probability density of the variables of the stochastic process. Indeed, for the system

$$\frac{dX_i}{dt} = f_i(\mathbf{X}) + g_{ij}(\mathbf{X})\xi_j\,, \tag{6.1}$$

the Fokker-Planck equation

$$\frac{\partial\rho}{\partial t} = \left\{-\frac{\partial}{\partial x_i}f_i + \frac{1}{2}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}D_{ij}\right\}\rho\,, \tag{6.2}$$

where $D_{ij}(\mathbf{x}) = g_{ik}(\mathbf{x})g_{jk}(\mathbf{x})$, can be written in the form of the continuity equation

$$\frac{\partial\rho}{\partial t} + \frac{\partial J_i}{\partial x_i} = 0\,, \tag{6.3}$$

with the current vector field $\mathbf{J}$ whose components are

$$J_i(\mathbf{x},t) = \left\{f_i - \frac{1}{2}\frac{\partial}{\partial x_j}D_{ij}\right\}\rho(\mathbf{x},t)\,. \tag{6.4}$$

The current vector gives the flow of probability in the system's state space.

A stationary state of the stochastic process is achieved when the probability density of the variable $\mathbf{X}(t)$ becomes time independent. In terms of the Fokker-Planck equation, the stationary state is described by a time-independent probability distribution, $\rho_{\text{stat}}(\mathbf{x})$, that solves the equation

$$0 = \frac{\partial}{\partial x_i}\left\{-f_i + \frac{1}{2}\frac{\partial}{\partial x_j}D_{ij}\right\}\rho_{\text{stat}}(\mathbf{x})\,. \tag{6.5}$$

The approach to stationary behavior may also be of interest in some situations, but in this section we will restrict our considerations to the nature of time-independent solutions of the Fokker-Planck equation.

The stationary Fokker-Planck equation, Eq. (6.5), implies that the stationary current vector field, $\mathbf{J}_{\text{stat}}(\mathbf{x})$, defined by

$$J_{\text{stat},i}(\mathbf{x}) = \left\{ f_i(\mathbf{x}) - \frac{1}{2}\frac{\partial}{\partial x_j} D_{ij}(\mathbf{x}) \right\} \rho_{\text{stat}}(\mathbf{x}), \tag{6.6}$$

is a divergence-free vector field:

$$0 = \nabla \cdot \mathbf{J}_{\text{stat}}\,. \tag{6.7}$$

In general, the current need not vanish in a stationary state—only its divergence necessarily vanishes.

We define an *equilibrium* stationary state to be a solution of the stationary Fokker-Planck equation which satisfies not only the stationary Fokker-Planck equation, but also

$$\mathbf{J}_{\text{stat}} = 0\,. \tag{6.8}$$

Physically, the vanishing of the probability current in the stationary state implies that the rate at which probability flows from any one point in the state space to another is equal to the rate at which probability flows in the opposite direction between the two points. This is the condition of *detailed balance* in the stationary state, and this is a general feature of systems in true thermodynamic equilibrium. In fact, if the stochastic dynamics described by the Fokker-Planck equation is the reduced description of an underlying system with time reversal invariant dynamics, such as Hamiltonian dynamics of classical mechanics, then the condition of detailed balance follows for the stationary state. Stochastic dynamic systems that satisfy the detailed balance condition can for many purposes be considered as thermal equilibrium systems. Mathematically, the property of detailed balance in equilibrium stationary states facilitates solution of the Fokker-Planck equation. Indeed, rather than solving the second-order differential equation in Eq. (6.5), we need only solve the first-order equation in Eq. (6.8).

As an example, consider a diffusion process defined by the stochastic differential equations

$$\frac{dX_i}{dt} = f_i(\mathbf{X}) + \sigma\xi_i(t), \qquad i = 1, \ldots, n\,, \tag{6.9}$$

where the ξ_i are independent gaussian white-noise processes. These dynamics describe the effect of additive noise, of equal magnitude in each component, on an n-dimensional dynamical system. The stationary Fokker-Planck equation is

$$0 = \frac{\partial}{\partial x_i}\left\{ -f_i + \frac{\sigma^2}{2}\frac{\partial}{\partial x_i} \right\} \rho_{\text{stat}}(\mathbf{x})\,, \tag{6.10}$$

and if the stationary state is to be an equilibrium stationary state, then we must also have

$$0 = \left\{ -f_i + \frac{\sigma^2}{2}\frac{\partial}{\partial x_i} \right\} \rho_{\text{stat}}(\mathbf{x})\,. \tag{6.11}$$

What are the conditions on the dynamics so that this is possible? Because the stationary density is a positive function, we may write

$$\rho_{\text{stat}}(\mathbf{x}) = \exp\left\{-\frac{2\Phi(\mathbf{x})}{\sigma^2}\right\}, \tag{6.12}$$

introducing the "potential" $\Phi(\mathbf{x})$. Inserting Eq. (6.12) into Eq. (6.11), we find

$$f_i(\mathbf{x}) = -\frac{\partial\Phi(\mathbf{x})}{\partial x_i}. \tag{6.13}$$

This says that the drift vector field $\mathbf{f}(\mathbf{x})$ must be a gradient vector field—in particular it must be curl free—in order for the detailed balance to hold in the stationary state. This is a strong restriction on the unperturbed deterministic dynamics so that the stochastic version possesses an equilibrium stationary state; the deterministic dynamics is certainly not chaotic in this case, for example. Such finite-dimensional "gradient flow" systems generically display few interesting dynamics.

More generally, the condition of detailed balance in the stationary state places severe constraints on the functional forms of the drift vector and diffusion matrix, but allows for the solution of the stationary distribution up to quadratures. For systems with a positive definite diffusion matrix $D_{ij}(\mathbf{x})$, if there is a zero current solution for Eq. (6.6), then we may write $\rho_{\text{stat}}(\mathbf{x}) = \exp\{-\Phi(\mathbf{x})\}$ so that

$$\frac{\partial\Phi(\mathbf{x})}{\partial x_j} = D_{ij}{}^{-1}(\mathbf{x})\left\{2f_j(\mathbf{x}) - \frac{\partial}{\partial x_k}D_{jk}(\mathbf{x})\right\}. \tag{6.14}$$

Here the condition for detailed balance is that the vector field defined by the diffusion matrix and the drift vector above is a gradient, which is not true for arbitrary dynamics. Should this condition hold, then the generalized potential $\Phi(\mathbf{x})$ can be computed by just one integration (in each variable).

We refer to stationary states with a nonvanishing probability current as *non-equilibrium* stationary states. Physical examples are steady current flow in a wire, steady heat flow across a slab of material, or even steady convection rolls in a fluid heated from below. Mathematical examples are any time-independent solutions to our Fokker-Planck equation which do not satisfy the detailed balance condition. We do not generally know the functional form of the stationary probability distribution in terms of the drift and diffusion for these systems, and thus we are generally ignorant of their stationary properties. One of the fundamental open questions in statistical physics is to develop techniques for the analysis of nonequilibrium steady states with the goal of identifying some common—and if possible universal—rules by which they are organized. Further research in this area is left as a challenge to the reader.

7. ADIABATIC ELIMINATION: COLORED NOISE, ITÔ VS. STRATONOVICH, AND KRAMERS VS. SMOLUCHOWSKI

We have defined the solution of a stochastic differential equation with gaussian white noise as a Markov process defined by the transition density obtained as the solution of the Fokker-Planck equation. The Fokker-Planck equation associated with a given stochastic differential equation was derived based on a specific interpretation of the stochastic differential equation as the continuous time limit of a discrete time problem. The analysis in section 4 corresponds to what is called the *Itô interpretation* of the stochastic differential equation. Our interpretation of the stochastic differential equation is not unique, however, and other interpretations (which may be appropriate for some problems) may lead to different results—both qualitative as well as quantitative! In this section we will develop the *Stratonovich interpretation* of the stochastic differential equation as the white-noise limit of a "real" noise problem.

The approach we will take here is to start with the Itô interpretation of a multidimensional system, one of whose components corresponds to a rapidly fluctuating noise, and the other of which is the system variable. In the limit that the ratio of the time scale associated with the fast noise to the system time scale vanishes, we may consider the noise to be white. Using a singular perturbation analysis of the Fokker-Planck equation for the full system, we will perform an *adiabatic elimination* of the noise variable and derive a reduced Fokker-Planck equation for the system variable alone. We will see that this interpretation of the white-noise problem, as the "white-noise limit" of a continuous-time fast noise problem, may lead to a different Fokker-Planck equation than before. The adiabatic elimination procedure that we will develop may also be applied to other problems where there is a clear separation of time scales between several processes. The same mathematical approach will be used to reduce Kramers' equation, the Fokker-Planck equation associated with Newton's law for the velocity and position of a particle subject to both a random and a deterministic force, to Smoluchowski's equation—a Fokker-Planck equation for the position of the particle alone.

We begin with a somewhat general formulation. Consider a system variable $X(t)$ which obeys the stochastic differential equation

$$\frac{dX}{dt} = f(X) + g(X)\frac{1}{\sqrt{\tau}}\zeta(t) \tag{7.1}$$

where $f(X)$ is the deterministic component of the evolution, and $\tau^{-1/2}\zeta(t)$ is an approximate white noise, i.e., a stochastic process which we assume varies on a time scale much faster than $X(t)$. The state-dependent diffusion factor $g(X)$ describes the sensitivity of the system variable to the noise at different locations in the state space. The noise process $\zeta(t)$ evolves separately from the system variable, and we have to make some specific assumptions regarding it in order to proceed. It is not unreasonable for many applications to consider the noise to be gaussian because this would be the case (under very general conditions) if it was the result of the sum of many separate, nearly independent effects. Moreover, it is not unreasonable to model the noise as a Markov process, that is, we will assume that the future state of the noise process cannot be predicted any better by supplying more of its history than simply its current state. Finally, we assume that the noise is stationary. Then its one-time probability density is time independent and its transition density between times t and s is only a function of the time difference $|t-s|$. We may thus model $\zeta(t)$ as a stationary Ornstein-Uhlenbeck process as introduced in Exercise 4.

To establish notations and normalizations, we take $\zeta(t)$ to be a Markov process with the time-independent one-time probability density that is normal with mean zero and variance one,

$$\rho(z) = \frac{1}{\sqrt{\pi}} e^{-z^2}, \tag{7.2}$$

and transition density

$$\rho(z,t|z_0,s) = \frac{1}{\sqrt{2\pi\Sigma(t-s)}} \exp\left\{-\frac{(z - z_0 e^{-(t-s)/\tau})^2}{2\Sigma(t-s)}\right\} \tag{7.3}$$

with

$$\Sigma(t) = \frac{1}{2}(1 - e^{-2t/\tau}). \tag{7.4}$$

The stationary covariance function of $\zeta(t)$ is thus

$$E\{\zeta(t)\zeta(s)\} = \frac{1}{2} e^{-|t-s|/\tau}. \tag{7.5}$$

In Figure 18 we show a typical realization of the Ornstein-Uhlenbeck process on a time scale much shorter than the relaxation time τ, $\zeta(t)$ then looks like Brownian motion (compare Figures 18 and 3). The difference between the $\zeta(t)$ and the Wiener process starts to appear on time scales of the order of the relaxation time. The Ornstein-Uhlenbeck process does not tend to wander so far from its starting point, and this is the property that allows for the existence of a nontrivial stationary state for the process. On a time scale much longer than τ, the process looks very much like a white noise because it appears to almost instantly decorrelate from itself. Compare Figures 19 and 5. The ζ–ζ correlation function in Eq. (7.5) vanishes almost immediately for short τ so the correlation function of $\tau^{-1/2}\zeta$ is an approximate δ-function with the correct normalization,

$$\int_{-\infty}^{\infty} \tau^{-1} E\{\zeta(t)\zeta(s)\} dt = 1, \tag{7.6}$$

like the white-noise correlation function. The Ornstein-Uhlenbeck process $\zeta(t)$ is the solution of the stochastic differential equation

$$\frac{d\zeta}{dt} = -\frac{1}{\tau}\zeta + \frac{1}{\sqrt{\tau}}\xi \tag{7.7}$$

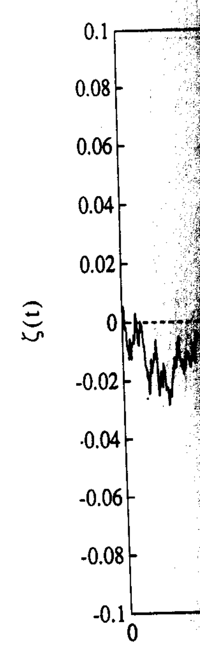


FIGURE 18 A realization of
The correlation time here is

with $\xi(t)$ the usual gaussi
white noise as can be seen

$\zeta(t) = e^{-}$

In the stationary state (as
response of a linear low-p
for ever decreasing relaxa
the unfiltered white-noise
The two-dimensional
and the noise simultaneou

ditions) if it was the result of the
s. Moreover, it is not unreasonable
s, we will assume that the future
ny better by supplying more of its
ssume that the noise is stationary.
pendent and its transition density
me difference $|t-s|$. We may thus
rocess as introduced in Exercise 4.
e take $\zeta(t)$ to be a Markov process
density that is normal with mean

$$, \qquad (7.2)$$

$$\frac{(z - z_0 e^{-(t-s)/\tau})^2}{2\Sigma(t-s)} \Big\} \qquad (7.3)$$

$$/\tau) . \qquad (7.4)$$

us

$$|t-s|/\tau . \qquad (7.5)$$

the Ornstein-Uhlenbeck process on
me τ, $\zeta(t)$ then looks like Brownian
ce between the $\zeta(t)$ and the Wiener
order of the relaxation time. The
nder so far from its starting point,
ence of a nontrivial stationary state
an τ, the process looks very much
t instantly decorrelate from itself.
unction in Eq. (7.5) vanishes almost
tion of $\tau^{-1/2}\zeta$ is an approximate δ-

$$\}dt = 1 , \qquad (7.6)$$

Ornstein-Uhlenbeck process $\zeta(t)$ is
ion

$$\frac{1}{\sqrt{\tau}}\xi \qquad (7.7)$$

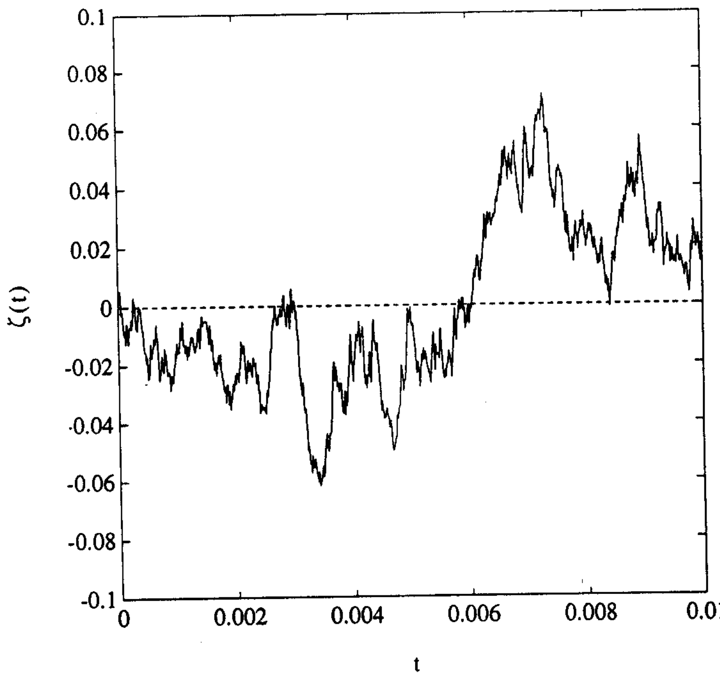


FIGURE 18 A realization of the Ornstein-Uhlenbeck process on s short time scale. The correlation time here is $\tau = 1$.

with $\xi(t)$ the usual gaussian white-noise process. In fact, $\zeta(t)$ is just a "filtered" white noise as can be seen by writing the solution to Eq. (7.7) in integral form:

$$\zeta(t) = e^{-(t-t_0)/\tau}\zeta(t_0) + \frac{1}{\sqrt{\tau}}\int_{t_0}^{t} e^{-(t-s)/\tau}\xi(s)ds . \qquad (7.8)$$

In the stationary state (as the initial time $t_0 \to -\infty$), $\zeta(t)$ becomes the steady-state response of a linear low-pass filter of bandwidth τ^{-1} to a white-noise signal. Thus, for ever decreasing relaxation time τ, the Ornstein-Uhlenbeck process approaches the unfiltered white-noise process itself.

The two-dimensional Markov process $(X(t), \zeta(t))$, describing the state variable and the noise simultaneously, satisfies the system of stochastic differential equations

$$\begin{aligned} \frac{dX}{dt} &= f(X) + g(X)\frac{1}{\sqrt{\tau}}\zeta \\ \frac{d\zeta}{dt} &= -\frac{1}{\tau}\zeta + \frac{1}{\sqrt{\tau}}\xi \end{aligned} \qquad (7.9)$$

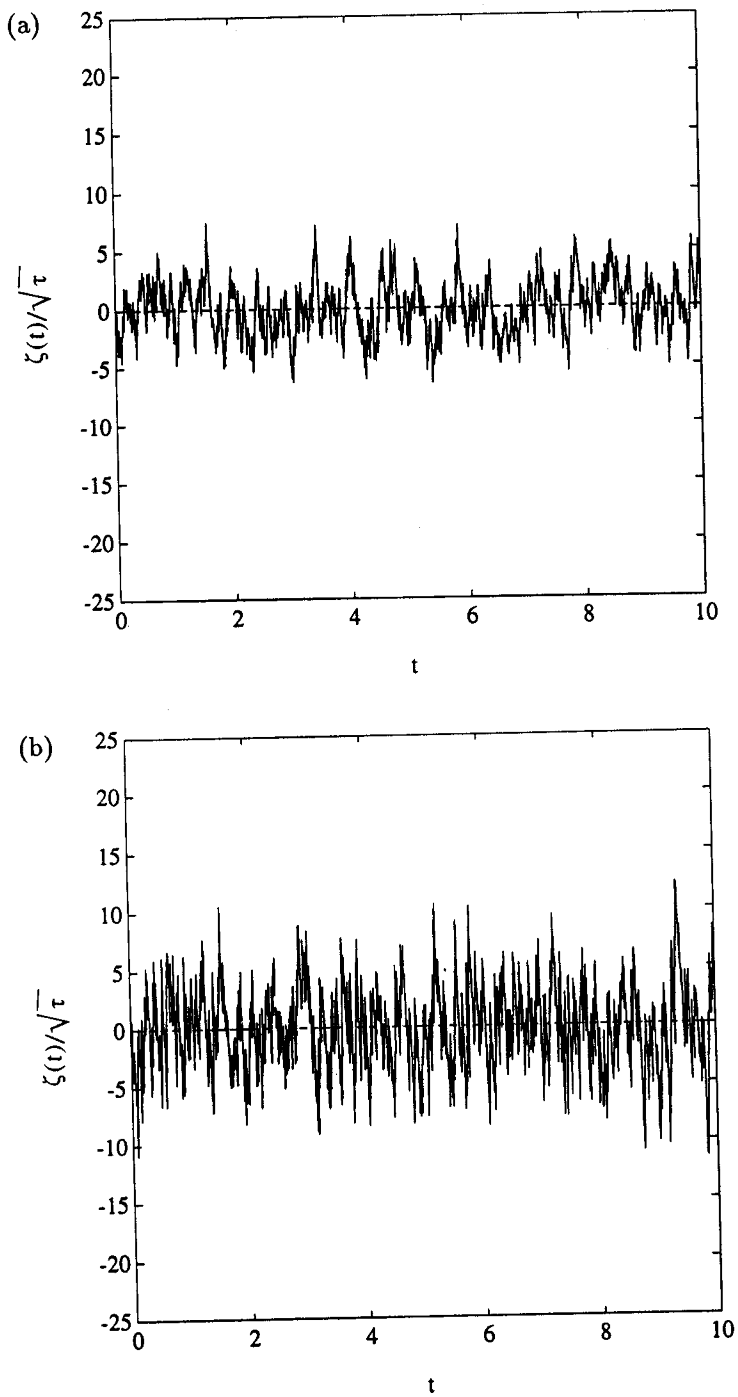


FIGURE 19 The approximate white-noise process $\zeta(t)/\sqrt{\tau}$, for several decreasing correlation times: (a) $\tau = .08$, (b) $\tau = .04$, (c) $\tau = .02$, and (d) $\tau = .01$. Compare Figure 5.

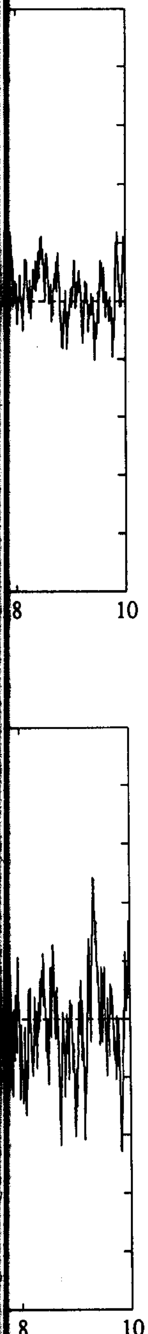


FIGURE 19 The approximate white-noise process $\zeta(t)/\sqrt{\tau}$, for several decreasing correlation times: (a) $\tau = .08$, (b) $\tau = .04$, (c) $\tau = .02$, and (d) $\tau = .01$. Compare Figure 5.

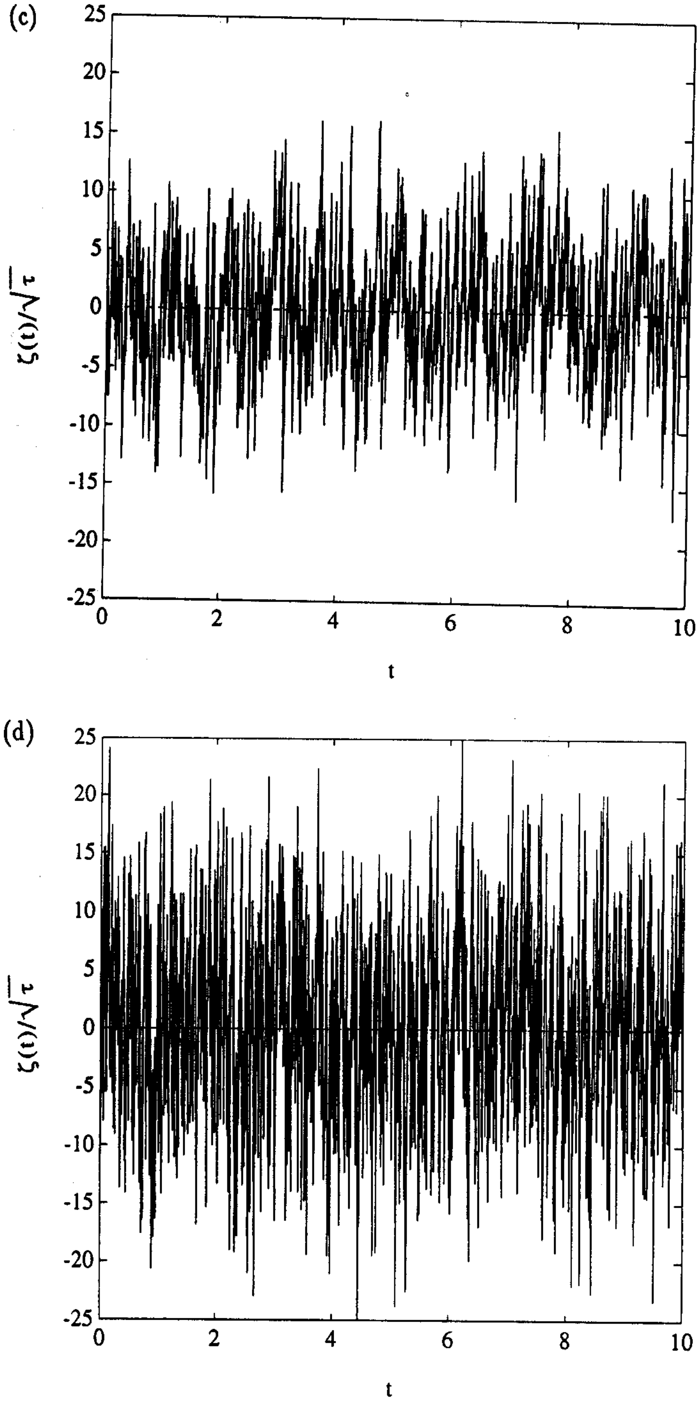


FIGURE 19 (continued)

According to Eqs.(4.21)–(4.23), the transition density $\rho(x, z, t|x_0, z_0, t_0)$ satisfies the Fokker-Planck equation

$$\partial_t \rho = \left\{ -\partial_x f(x) - \frac{1}{\sqrt{\tau}} z \partial_x g(x) + \frac{1}{\tau} \partial_z \left(z + \frac{1}{2} \partial_z \right) \right\} \rho . \tag{7.10}$$

Our goal now is to derive a Fokker-Planck equation for the "reduced," or "marginal" transition density $r(x, t|x_0, t_0)$ of the process $X(t)$ *alone* in the limit $\tau \to 0$. For each $\tau > 0$, the reduced density is defined as

$$r(x, t|x_0, t_0) = \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz_0 \rho(x, z, t|x_0, z_0, t_0) \frac{1}{\sqrt{\pi}} e^{-z_0^2} . \tag{7.11}$$

The marginal transition density does not satisfy a (closed) Fokker-Planck equation for $\tau > 0$, but it does in the limit $\tau \to 0$. We cannot generally solve the full Fokker-Planck problem in Eq. (7.10) even in a stationary state, because the process does not generically satisfy the condition for detailed balance. Hence, we are forced to resort to some kind of perturbation theory.

Let $\varepsilon = \sqrt{\tau}$, so that the Fokker-Planck equation in Eq. (7.10) is written

$$0 = \left\{ (-\partial_t + F_2) + \frac{1}{\varepsilon} F_1 + \frac{1}{\varepsilon^2} F_0 \right\} \rho , \tag{7.12}$$

where

$$\begin{aligned} F_2 &= -\partial_x f(x), \\ F_1 &= -z \partial_x g(x), \\ F_0 &= \partial_z \left(z + \frac{1}{2} \partial_z \right) . \end{aligned} \tag{7.13}$$

We make the ansatz that the full transition density can be expanded in a power series in ε:

$$\rho = \rho_0 + \varepsilon \rho_1 + \varepsilon^2 \rho_2 + \dots \tag{7.14}$$

The marginal density is then also assumed to be a power series in ε as we have

$$\begin{aligned} r &= \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} dz_0 \left\{ \rho_0 + \varepsilon \rho_1 + \varepsilon^2 \rho_2 + \dots \right\} \frac{1}{\sqrt{\pi}} e^{-z_0^2}, \\ &= r_0 + \varepsilon r_1 + \varepsilon^2 r_2 + \dots \end{aligned} \tag{7.15}$$

This is a *singular perturbation* analysis, because of the singular appearance of the expansion parameter in the partial differential equation.

Inserting this expansion for the transition density into the Fokker-Planck equation and collecting coefficients of like powers of ε, we find

$$\begin{aligned} O(\varepsilon^{-2}) &: \quad 0 = F_0 \rho_0 \\ O(\varepsilon^{-1}) &: \quad 0 = F_1 \rho_0 + F_0 \rho_1 \\ O(\varepsilon^{0}) &: \quad 0 = (-\partial_t + F_2) \rho_0 + F_1 \rho_1 + F_0 \rho_2 \end{aligned} \tag{7.16}$$

and so on. These terms above are all that we will need for the analysis.

Of primary importance in this calculation is the properties of the "noise evolution operator," $F_0 = \partial_z(z + (1/2)\partial_z)$. This operates only on the noise variable z, and has the following spectrum of eigenvalues and eigenfunctions $p_n(z)$:

$$F_0 p_n = -n p_n , \tag{7.17}$$

where

$$p_n = H_n(z)\frac{1}{\sqrt{\pi}}e^{-z^2} , \tag{7.18}$$

and the $H_n(z)$ are the Hermite polynomials defined by

$$H_n(z) = (-1)^n e^{z^2}\frac{d^n}{dz^n}e^{-z^2} . \tag{7.19}$$

The stationary distribution of $\zeta(t)$ is the $n = 0$ eigenfunction of the operator, p_0. The first few Hermite polynomials are

$$\begin{aligned} H_0(z) &= 1 \\ H_1(z) &= 2z \\ H_2(z) &= 2(2z^2 - 1). \end{aligned} \tag{7.20}$$

and they, as well as the eigenfunctions p_n, satisfy the recursion relation

$$zH_n = \frac{1}{2}H_{n+1} + nH_{n-1} . \tag{7.21}$$

The $O(\varepsilon^{-2})$ equation implies that ρ_0 must be decomposed as (suppressing the initial data)

$$\rho_0(x, z, t) = r_0(x, t)p_0(z), \tag{7.22}$$

where $r_0(x, t)$, whose evolution equation is yet to be determined, is the leading term in the reduced density as $\varepsilon \to 0$. Then, the second equation in the perturbation expansion, of order ε^{-1}, is rewritten

$$\begin{aligned} F_0\rho_1 &= - F_1\rho_0 \\ &= zp_0(z)\partial_x g(x)r_0(x, t) \\ &= \frac{1}{2}p_1(z)\partial_x g(x)r_0(x, t), \end{aligned} \tag{7.23}$$

where the recursion relation in Eq. (7.21) was used to write $zp_0(z) = p_1(z)/2$. Using Eq. (7.17) for $n = 0$ and 1, we see that the solution for ρ_1 is

$$\rho_1(x, z, t) = -\frac{1}{2}\partial_x g(x)r_0(x, t)p_1(z) + r_1(x, t)p_0(z) , \tag{7.24}$$

where $r_1(x,t)$ is also undetermined at this point. Now, the $O(\varepsilon^0)$ equation gives

$$
\begin{aligned}
F_0\rho_2 &= -F_1\rho_1 + (\partial_t - F_2)\rho_0 \\
&= z\partial_x g(x)\left\{-\frac{1}{2}\partial_x g(x) r_0(x,t)p_1(z) + r_1(x,t)p_0(z)\right\} \\
&\quad + \{\partial_t + \partial_x f(x)\} r_0(x,t)p_0(z) \\
&= -\frac{1}{2}\partial_x g \partial_x g r_0 z p_1(z) + \partial_x g r_1 z p_0(z) + \{\partial_t + \partial_x f\} r_0 p_0(z)\,.
\end{aligned}
\tag{7.25}
$$

Using the recursion relation for the terms zp_1 and zp_0 above, we thus have have

$$
F_0\rho_2 = \left\{\partial_t + \partial_x f - \frac{1}{2}\partial_x g \partial_x g\right\} r_0 p_0 + \frac{1}{2}\partial_x g r_1 p_1 - \frac{1}{4}\partial_x g \partial_x g r_0 p_2\,. \tag{7.26}
$$

To solve Eq. (7.26) for ρ_2, the coefficient of $p_0(z)$ on the right-hand side above must vanish. That is, because $F_0 p_0 = 0$, the operator F_0 is not invertible on the functional subspace spanned by p_0. This is the central logical step in this singular perturbation theory, and this integrability condition finally determines the evolution equation for the leading term in the reduced density:

$$
\partial_t r_0(x,t|x_0,t_0) = \left\{-\partial_x f + \frac{1}{2}\partial_x g \partial_x g\right\} r_0\,. \tag{7.27}
$$

This is the main result of this section. In the white-noise limit, $\tau \longrightarrow 0$, the state variable $X(t)$ has a transition probability $r(x,t|x_0,t_0) = r_0(x,t|x_0,t_0)$ that satisfies the Fokker-Planck equation above. This procedure of *adiabatic elimination* of the noise variable is a useful way of simplifying a problem by "averaging over" the fast variables, leaving an effective dynamics for the "slow" components. It should be clear that the procedure can be carried on and we may derive dynamical equations for succeeding corrections to the leading white-noise behavior of the marginal probability density.

Note that this Fokker-Planck equation is *not* the same that we would have written down for the white-noise problem

$$
\frac{dX}{dt} = f(X) + g(X)\xi, \tag{7.28}
$$

which is

$$
\partial_t r(x,t|x_0,t_0) = \left\{-\partial_x f + \frac{1}{2}\partial_x^2 g^2\right\} r. \tag{7.29}
$$

The Fokker-Planck equation in Eq. (7.27), regarded as the evolution equation for the transition density of the white-noise limit of a "colored," or "real," noise problem, corresponds to the *Stratonovich* interpretation of the stochastic differential equation. The Itô interpretation, and its Fokker-Planck equation in Eq. (7.29), that we have been using up to now, is valid as the continuous time limit of a discrete time problem. The two interpretations are not necessarily the same—if the solutions of the Fokker-Planck equations in Eqs.(7.27) and (7.29) are different, then we are talking about two different Markov processes. This means that it may be crucial in the modeling process to be aware of the exact sense in which fluctuations are to be regarded as white noise. The only case in which the evolution operators in Eqs.(7.27) and (7.29) are the same is if the diffusion coefficient $g(x)$ is constant as a function of x. In this case we say that the stochastic differential equation is driven by *additive* noise, as opposed to *multiplicative*, or *state-dependent*, noise.

The same philosophy of adiabatic elimination, or the elimination of fast variables, can be used to simplify the description of the effect of additive noise of systems with inertia. By this we mean systems where the *acceleration* of the state variable is proportional to a white noise. Consider a model of a particle of mass m moving (without loss of generality in one dimension) in a dissipative medium, and subject to a conservative force field with potential $U(x)$. The dissipative medium is a model of a gas or fluid whose effect on the particle is to provide a force due to the rapid bombardment of the particle by the particles that make up the medium. This force is made up of two components: (1) a viscous drag force proportional in magnitude to the velocity, and (2) a fast, mean zero, stochastic force due to fluctuations caused by the discrete nature of the medium. Writing the particle's position-velocity coordinates as $X(t)$ and $V(t)$, we have from Newton's third law

$$\begin{aligned} \frac{dX}{dt} &= V, \\ m\frac{dV}{dt} &= -U'(X) - \gamma V + \sigma\xi, \end{aligned} \tag{7.30}$$

where γ is the linear friction coefficient, and σ is the strength of the δ-correlated white-noise force ξ. We will analyze this model now without saying any more about its validity in specific situations. In fact, a great part of the field of statistical mechanics is concerned with the derivation of such a system of equations from first principles.

The Fokker-Planck equation for the phase-space probability density of the joint process is called the *Kramers' equation* or *Klein-Kramers equation*. Explicitly, it is

$$\partial_t \rho(x,v,t) = \left\{ -\partial_x v + \partial_v \frac{U'(x)}{m} + \frac{\gamma}{m}\partial_v\left(v + \frac{\sigma^2}{2m\gamma}\partial_v\right)\right\}\rho. \tag{7.31}$$

We require that the equilibrium (stationary) probability density of the process is the Gibbs distribution of equilibrium statistical mechanics,

$$\rho_{eq}(x,v) = Z^{-1}\exp\left\{\frac{-H}{kT}\right\}, \tag{7.32}$$

where the Hamiltonian function H is

$$H = \frac{mv^2}{2} + U(x), \tag{7.33}$$

the partition function Z is the normalization constant for the probability density, T is the temperature of the medium in which the particle lies, and k is Boltzmann's constant. Then, the friction and noise amplitude coefficients are not independent, but are connected by setting

$$0 = \left\{-\partial_x v + \partial_v \frac{U'(x)}{m} + \frac{\gamma}{m}\partial_v (v + \frac{\sigma^2}{2m\gamma}\partial_v)\right\} \rho_{eq}(x,v). \qquad (7.34)$$

This is true when

$$\frac{\sigma^2}{2\gamma} = kT. \qquad (7.35)$$

This relationship between the effective noise strength σ^2, the friction coefficient γ, and the temperature T, is known as a *fluctuation-dissipation* relation. In line with our intuition, the temperature T is proportional to the strength of the stochastic force from the medium (at fixed friction coefficient).

Redefining the noise amplitude in terms of the temperature and γ, we have the Fokker-Planck equation

$$\partial_t \rho(x,v,t) = \left\{-\partial_x v + \partial_v \frac{U'(x)}{m} + \frac{\gamma}{m}\partial_v \left(v + \frac{kT}{m}\partial_v\right)\right\} \rho. \qquad (7.36)$$

At a given temperature then, the frictional rate γ/m plays a similar mathematical role as the inverse time scale of the noise, τ^{-1}, in Eq. (7.10). For large friction coefficients, the noise—the velocity variable $V(t)$—is "fast" on the time scale of the evolution of the position variable. What we would like to do is to find a Fokker-Planck equation for the position variable alone in the high friction limit. To systematize the procedure, let us change variables to a dimensionless velocity measured in units of the thermal velocity $\sqrt{kT/m}$:

$$w = \sqrt{\frac{m}{kT}}v, \qquad (7.37)$$

and then let us change to a long time scale

$$s = \frac{m}{\gamma}t. \qquad (7.38)$$

Kramers' equation becomes

$$\partial_s \rho(x,w,s) = \varepsilon^{-1}\left\{-\sqrt{\frac{kT}{m}}w\partial_x + \frac{U'(x)}{\sqrt{mkT}}\partial_w\right\}\rho + \varepsilon^{-2}\partial_w(w+\partial_w)\rho, \qquad (7.39)$$

where $\varepsilon = (m/\gamma)$. In the limit $\varepsilon \to 0$, we would like to find an effective Fokker-Planck equation for the reduced density

$$r(x,t|x_0,t_0) = \int_{-\infty}^{\infty} dw \int_{-\infty}^{\infty} dw_0 \rho(x,w,s|x_0,w_0,s_0)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}w_0^2}, \qquad (7.40)$$

where we take the marginal distribution of the (fast) velocity process to be the equilibrium Boltzmann distribution, $(2\pi)^{-1/2}\exp\{-w^2/2\}$.

As before, we expand the joint density in a power series

$$\rho = \rho_0 + \varepsilon\rho_1 + \varepsilon^2\rho_2 + \ldots \tag{7.41}$$

and collect terms order by order in powers of ε:

$$\begin{aligned}
O(\varepsilon^{-2}):&\quad 0 = \partial_w(w+\partial_w)\rho_0\\
O(\varepsilon^{-1}):&\quad 0 = \left\{-\sqrt{\tfrac{kT}{m}}w\partial_x + \tfrac{U'(x)}{\sqrt{mkT}}\partial_w\right\}\rho_0 + \partial_w(w+\partial_w)\rho_1\\
O(\varepsilon^{0}):&\quad \partial_s\rho_0 = \left\{-\sqrt{\tfrac{kT}{m}}w\partial_x + \tfrac{U'(x)}{\sqrt{mkT}}\partial_w\right\}\rho_1 + \partial_w(w+\partial_w)\rho_2,
\end{aligned} \tag{7.42}$$

and so on. The role of F_0 above is played by the closely related operator $\partial_w(w+\partial_w)$. Its spectrum is the same as F_0's, with slightly different eigenfunctions:

$$\partial_w(w+\partial_w)R_n(w) = -nR_n(w) \tag{7.43}$$

where the eigenfunctions are

$$R_n = He_n(w)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}w^2}, \tag{7.44}$$

where the Hermite polynomials $He_n(w)$ are defined by

$$He_n(w) = (-1)^n e^{\frac{1}{2}w^2}\frac{d^n}{dw^n}e^{-\frac{1}{2}w^2}. \tag{7.45}$$

The first few of these Hermite polynomials are

$$\begin{aligned}
He_0(w) &= 1\\
He_1(w) &= w\\
He_2(w) &= w^2 - 1.
\end{aligned} \tag{7.46}$$

and they, as well as the eigenfunctions R_n, satisfy the recursion relation

$$wHe_n = He_{n+1} + nHe_{n-1}\,. \tag{7.47}$$

The $O(\varepsilon^{-2})$ equation tells us that

$$\rho_0(x,w,s) = r_0(x,s)R_0(w), \tag{7.48}$$

where, as before, the function $r_0(x,s)$ is the leading order reduced density whose evolution equation we seek. Using the recursion relation in Eq. (7.47), along with the fact that $\partial_w R_0 = -R_1$, the $O(\varepsilon^{-1})$ equation becomes

$$\begin{aligned}\partial_w(w+\partial_w)\rho_1 &= \sqrt{\frac{kT}{m}}\partial_x r_0(x,s) w R_0(w) - \frac{U'(x)}{\sqrt{mkT}} r_0(x,s)\partial_w R_0(w) \\ &= \left[\sqrt{\frac{kT}{m}}\partial_x r_0(x,s) + \frac{U'(x)}{\sqrt{mkT}} r_0(x,s)\right] R_1(w).\end{aligned} \tag{7.49}$$

In light of Eq. (7.43), we find the expression for ρ_1

$$\rho_1 = -\left[\sqrt{\frac{kT}{m}}\partial_x r_0(x,s) + \frac{U'(x)}{\sqrt{mkT}} r_0(x,s)\right] R_1(w) + r_1(x,s)R_0(w), \tag{7.50}$$

where, again as before, $r_1(x,s)$ is undetermined at this stage. Now the $O(\varepsilon^0)$ equation is

$$\begin{aligned}\partial_w(w+\partial_w)\rho_2 &= \partial_s\rho_0 + \left\{\sqrt{\frac{kT}{m}} w\partial_x - \frac{U'(x)}{\sqrt{mkT}}\partial_w\right\}\rho_1 \\ &= \partial_s r_0(x,s)R_0(w) - \left\{\sqrt{\frac{kT}{m}} w\partial_x - \frac{U'(x)}{\sqrt{mkT}}\partial_w\right\} \\ &\quad\times\left[\sqrt{\frac{kT}{m}}\partial_x r_0(x,s) + \frac{U'(x)}{\sqrt{mkT}} r_0(x,s)\right] R_1(w) \\ &\quad+\left\{\sqrt{\frac{kT}{m}} w\partial_x - \frac{U'(x)}{\sqrt{mkT}}\partial_w\right\} r_1(x,s)R_0(w).\end{aligned} \tag{7.51}$$

The idea is to isolate the coefficient of $R_0(w)$ on the right-hand side of Eq. (7.51); it must vanish as the integrability condition for us to solve for ρ_2. The recursion relation and the fact that $\partial_w R_1 = -R_2$ lead us to

$$\begin{aligned}\partial_w(w+\partial_w)\rho_2 &= R_0(w)\left\{\partial_s - \partial_x\frac{U'(x)}{m} - \frac{kT}{m}\partial_x^2\right\} r_0 \\ &\quad+ [\text{terms proportional to } R_1(w) \text{ and } R_2(w)],\end{aligned} \tag{7.52}$$

so, in the limit $\varepsilon \to 0$, we have the reduced dynamics for $r(x,s)$,

$$\partial_s r(x,s) = \partial_x\left\{\frac{U'(x)}{m} + \frac{kT}{m}\partial_x\right\} r. \tag{7.53}$$

This Fokker-Planck
$X(t)$ alone, in the high fr
Fokker-Planck equation

or in terms of the origina

EXERCISE 6. Show that
change of variables $t \to$

8. APPLICATION:
REVISITED

To illustrate the differen
white-noise stochastic di
section 6 interpreting the
noise process. In the St
differential equation

has a transition density s

$$\frac{\partial\rho}{\partial t} =$$

$$=$$

This Fokker-Planck equa
but for this example the
parameters as shown by
stationary probability de
$\langle\mu\rangle \to \langle\mu\rangle + \sigma^2/2$ in thos
The stationary proba

This Fokker-Planck equation for the marginal density of the position process $X(t)$ alone, in the high friction limit, is called the *Smoluchowski equation*. It is the Fokker-Planck equation corresponding to the stochastic differential equation

$$\frac{dX(s)}{ds} = -\frac{U'(X)}{m} + \sqrt{\frac{2kT}{m}}\xi(s), \tag{7.54}$$

or in terms of the original time variable $t = \gamma s/m$,

$$\frac{dX(t)}{dt} = -\frac{U'(X)}{\gamma} + \sqrt{\frac{2kT}{\gamma}}\xi(t). \tag{7.55}$$

EXERCISE 6. Show that white noise must rescale as $\xi(t) \to \sqrt{\alpha}\xi(\alpha t)$ under the change of variables $t \to \alpha t$.

8. APPLICATION: THE STOCHASTIC VERHULST EQUATION REVISITED

To illustrate the differences between the Itô and Stratonovich interpretations of white-noise stochastic differential equations, we will now reconsider the model of section 6 interpreting the stochastic dynamics as the white-noise limit of a colored-noise process. In the Stratonovich interpretation, the solution of the stochastic differential equation

$$\frac{dX}{dt} = \langle\mu\rangle X - X^2 + \sigma\xi \tag{8.1}$$

has a transition density satisfying the Fokker-Planck equation

$$\begin{aligned}\frac{\partial\rho}{\partial t} &= \frac{\partial}{\partial x}\left\{x^2 - \langle\mu\rangle x + \frac{\sigma^2}{2}x\frac{\partial}{\partial x}x\right\}\rho \\ &= \frac{\partial}{\partial x}\left\{x^2 - \langle\mu\rangle x - \frac{\sigma^2}{2}x + \frac{\sigma^2}{2}\frac{\partial}{\partial x}x^2\right\}\rho .\end{aligned} \tag{8.2}$$

This Fokker-Planck equation is different from that considered before, Eq. (5.7), but for this example the difference can be absorbed into a renormalization of the parameters as shown by the second line above. That is, our previous results for the stationary probability density all carry over from section 5 with the replacement $\langle\mu\rangle \to \langle\mu\rangle + \sigma^2/2$ in those formulæ.

The stationary probability density is thus

$$\rho_{\text{stat}}(x) = N x^{(2\langle\mu\rangle/\sigma^2 - 1)} e^{-2x/\sigma^2}, \tag{8.3}$$

so long as

$$\frac{\langle \mu \rangle}{\sigma^2} > 0. \tag{8.4}$$

This condition holds whenever the average birth rate exceeds the average death rate, for any finite value of the noise amplitude. The other solution,

$$\rho_{\mathrm{stat}}(x) = \delta(x), \tag{8.4}$$

holds true for the initial condition $X(0) = 0$, or for average death rate exceeding average birth rate ($\langle \mu \rangle \le 0$). The nontrivial density in Eq. (8.4) displays the noise-induced transition from the state where the most probable value is nonzero, to the state where it is zero, at the critical noise amplitude $\sigma^2 = 2\langle \mu \rangle$.

Contrary to Itô interpretation, in the Stratonovich interpretation there is never a "noise-induced extinction" phenomenon where the average growth rate is positive, but very strong fluctuations may drive the population to zero. These distinctions represent very different qualitative behaviors coming from the two interpretations of the model, highlighting the crucial role that the details of a white-noise model play in its predictions.

9. SUMMARY

We have introduced some of the fundamentals of stochastic processes, focusing on Markov diffusion processes as the solutions of stochastic differential equations driven by gaussian white noise. The three central achievements of these lectures has been (1) to show how one may consistently and practically interpret differential equations with white-noise coefficients, (2) to illustrate by example that some interesting problems can be completely and exactly solved within this formalism, and (3) to show that the question of modeling is a crucial one for these systems. In particular, the example illustrated explicitly that the interplay of noise and nonlinear dynamics is not a trivial one. Even using the idealization of white noise with its infinitely fast fluctuation time scale, the effects of the variations may not just "average out," but can deeply modify the system's qualitative behavior.

ACKNOWLEDGMENTS

I thank Benjamin Luce and Paul Laub for careful and thoughtful readings of the manuscript. I also thank all the students of the 1990 Complex Systems Summer school for their interest, their questions, and for the warm yet challenging reception that these lectures incited. Preparation of the lectures and these lecture notes was supported by NSF grants PHY-8958506 (Presidential Young Investigator Award)

and PHY-8907755. The nu
on the MATLAB software

REFERENCES

1. Arnold, L. *Stochastic I*
2. Doering, C., H. Brand,
with Spatial Degrees o
3. Horsthemke, W., and I
Springer, 1984.
4. Moss, F., and P. McCl
1–3. New York: Cambr
5. Risken, H. *The Fokker-*
6. van Kampen, N. *Stoch*
North Holland, 1981.

and PHY-8907755. The numerical simulations and computations were performed on the MATLAB software system, supplied by The Math Works, Inc.

REFERENCES

1. Arnold, L. *Stochastic Differential Equations.* New York: Wiley, 1974.
2. Doering, C., H. Brand, and R. Ecke, eds. "External Noise and Its Interaction with Spatial Degrees of Freedom." *J. Stat. Phys.* **54(5/6)** (1989).
3. Horsthemke, W., and R. Lefever. *Noise Induced Transitions.* New York: Springer, 1984.
4. Moss, F., and P. McClintock. *Noise in Nonlinear Dynamical Systems*, vols. 1–3. New York: Cambridge, 1989.
5. Risken, H. *The Fokker-Planck Equation,* 2nd ed. New York: Springer, 1989.
6. van Kampen, N. *Stochastic Processes in Physics and Chemistry.* Amsterdam: North Holland, 1981.