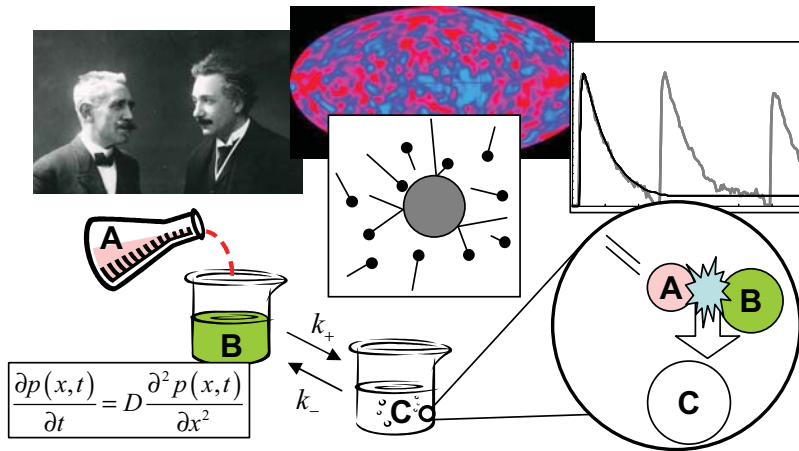




Applied Stochastic Processes in science and engineering

by

M. Scott

©2012

Objectives

This book is designed as an introduction to the ideas and methods used to formulate mathematical models of physical processes in terms of random functions. The first five chapters use the historical development of the study of Brownian motion as their guiding narrative. The remaining chapters are devoted to methods of solution for stochastic models. The material is too much for a single course – chapters 1-4 along with chapters 7 and 8 are ample for a senior undergraduate course offered to students with a suitably mathematical background (*i.e.* familiarity with most of the methods reviewed in Appendix B). For a graduate course, on the other hand, a quick review of the first three chapters, with a focus on later chapters should provide a good general introduction to numerical and analytic approximation methods in the solution of stochastic models.

The content is primarily designed to develop mathematical methods useful in the study of stochastic processes. Nevertheless, an effort has been made to tie the derivations, whenever possible, to the underlying physical assumptions that gave rise to the mathematics. As a consequence, very little is said about Itô formula and associated methods of what has come to be called *Stochastic Calculus*. If that comes as a disappointment to the reader, I suggest they consider C. W. Gardiner's book:

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004),

as a friendly introduction to Itô's calculus.

A list of references useful for further study appear at the beginning of some sections, and at the end of each chapter. These references are usually pedagogical texts or review articles, and are not meant to be an exhaustive tabulation of current results, but rather as a first step along the road of independent research in stochastic processes. A collection of exercises appear at the end of each chapter. Some of these are meant to focus the reader's attention on a particular point of the analysis, others are meant to develop essential technical skills. None of the exercises are very difficult, and all of them should be attempted (or at least read).

My hope is that the reader will find that the study of stochastic processes is not as difficult as it is sometimes made out to be.

Acknowledgements

This book is based, in part, upon the stochastic processes course taught by Pino Tenti at the University of Waterloo (with additional text and exercises provided by Zoran Miskovic), drawn extensively from the text by N. G. van Kampen “Stochastic process in physics and chemistry.” The content of Chapter 8 (particularly the material on parametric resonance) was developed in collaboration with Francis Poulin, and the *effective stability approximation* of Chapter 9 was developed in collaboration with Terry Hwa and Brian Ingalls. The material on stochastic delay equations was motivated by discussions with Lev Tsimring. Discussions with Stefan Klumpp led to clarification of a great deal of the material. I am grateful to Chris Ferrie, Killian Miller, David Stechlinski, Mihai Nica and all of my students in stochastic processes for correcting errors in the notes, and for trouble-shooting the exercises.

CONTENTS

1	Introduction	1
1.1	Stochastic Processes in Science and Engineering	1
1.2	Brownian motion	2
1.2.1	Einstein	4
1.2.2	Smolouchowski	9
1.2.3	Langevin	11
1.2.4	Ornstein and Uhlenbeck	15
1.3	Outline of the Book	20
1.4	Suggested References	22
2	Random processes	29
2.1	Random processes - Basic concepts	30
2.1.1	Moments of a Stochastic Process	34
2.2	Stationarity and the Ergodic Theorem	35
2.3	Spectral Density and Correlation Functions	39
2.3.1	Linear Systems	40
2.3.2	Fluctuation-Dissipation Relation	42
2.3.3	Johnson and Nyquist	44
3	Markov processes	50
3.1	Classification of Stochastic Processes	50
3.2	Chapman-Kolmogorov Equation	52
3.3	Master Equation	57
3.4	<i>Stosszahlansatz</i>	59

3.5	Example – One-step processes	62
3.6	Example – Bernoulli’s Urns and Recurrence	64
3.7	Example – Chemical Kinetics	68
4	Solution of the Master Equation	80
4.1	Exact Solutions	80
4.1.1	Moment Generating Functions	81
4.1.2	Matrix Iteration	84
4.2	Approximation Methods	87
4.2.1	Numerical Methods – Gillespie’s Algorithm	90
5	Perturbation Expansion of the Master Equation	100
5.1	Linear Noise Approximation (LNA)	100
5.1.1	Example – Nonlinear synthesis	105
5.1.2	Approximation of the Brusselator model	109
5.2	Example – ‘Burstiness’ in Protein Synthesis	113
5.3	Limitations of the LNA	116
6	Fokker-Planck Equation	120
6.1	Kolmogorov Equation	121
6.2	Derivation of the Fokker-Planck Equation	124
6.3	Macroscopic Equation	128
6.3.1	Coefficients of the Fokker-Planck Equation	131
6.4	Pure Diffusion and Ornstein-Uhlenbeck Processes	133
6.4.1	Wiener process	134
6.4.2	Ornstein-Uhlenbeck process	135
6.4.3	Heuristic Picture of the Fokker-Planck Equation	136
6.5	Connection to Langevin Equation	137
6.5.1	Linear Damping	137
6.5.2	Nonlinear Damping	140
6.6	Limitations of the Kramers-Moyal Expansion	143
6.7	Example – Kramer’s Escape Problem	143
7	Stochastic Analysis	152
7.1	Limits	152
7.2	Mean-square Continuity	154
7.3	Stochastic Differentiation	155
7.3.1	Wiener process is <i>not</i> differentiable	156
7.4	Stochastic Integration	158
7.4.1	Itô and Stratonovich Integration	160

8	Random Differential Equations	166
8.1	Numerical Simulation	168
8.2	Approximation Methods	173
8.2.1	Static Averaging	174
8.2.2	Bourret's Approximation	175
8.2.3	Cumulant Expansion	179
8.3	Example – Random Harmonic Oscillator	185
8.4	Example – Parametric Resonance	187
8.5	Optional – Stochastic Delay Differential Equations	188
9	Macroscopic Effects of Noise	199
9.1	Keizer's Paradox	200
9.2	Oscillations from Underdamped Dynamics	202
9.2.1	Example – Malthus Predator-Prey Dynamics	205
9.3	Effective Stability Analysis (ESA)	207
9.3.1	Example – Excitable Oscillator	209
10	Special Topics	216
10.1	Random Walks and Electric Networks	216
10.2	Fluctuations Along a Limit Cycle	217
10.3	Stochastic Resonance	220
10.4	Kalman Filter	222
10.5	Novikov-Furutsu-Donsker Relation	224
10.5.1	Variational Approximation	227
A	Review of Classical Probability	230
A.1	Statistics of a Random Variable	234
A.2	Examples of Distributions and Densities	236
A.3	Central limit theorem	241
A.4	Generating Random Numbers	242
A.4.1	Joint inversion generating method	243
A.5	Change of Variables	244
A.6	Monte-Carlo simulation	245
A.7	Bertrand's Paradox	245
B	Review of Mathematical Methods	254
B.1	Matrix Algebra and Time-Ordered Exponentials	254
B.2	Linear Stability Analysis	257
B.2.1	Routh-Hurwitz Criterion	258
B.3	Transforms Methods	260
B.3.1	Laplace Transform	260

B.3.2	Fourier Transform	262
B.3.3	z -Transform	264
B.4	Partial differential equations	264
B.4.1	Method of characteristics	264
B.4.2	Separation of variables	265
B.4.3	Transform methods	265
B.5	Some Results from Asymptotics and Functional Analysis .	266
B.5.1	Cauchy-Schwarz Inequality	266
B.5.2	Watson's Lemma	267
B.5.3	Stirling's Approximation	268
B.5.4	Eigenvalue perturbation	269
C	Itô Calculus	270
C.1	Itô's stochastic integral	270
C.1.1	Example – $\int_0^t W(t')dW(t')$	271
C.2	Change of Variables Formula	272
C.2.1	Example – Calculate $\langle \cos[\eta(t)] \rangle$	273
D	Sample Matlab code	275
D.1	Examples of Gillespie's direct method	275
D.1.1	Defining the propensity vector and stoichiometry matrix	276
D.1.2	Core stochastic simulation algorithm	276
D.1.3	Example - the Brusselator	278
D.2	Stochastic differential equations	278
D.2.1	White noise	279
D.2.2	Coloured noise	279

CHAPTER 1

INTRODUCTION

1.1 Stochastic Processes in Science and Engineering

Physics is the study of collective phenomena arising from the interaction of many individual entities. Even a cannonball dropped from a high tower will collide with some 10^{30} gas molecules on its way down. Part of the miracle of physics is that, as a rule, only a few variables are required to capture the behaviour of the system, and this behaviour can, in turn, be described by very simple physical laws. In the case of the falling cannonball, for example, only its position and velocity are important. In hydrodynamics, despite the incredible numbers of individual fluid molecules, the flow velocity, density and temperature are sufficient to describe the system under most circumstances.

Yet the interactions that are eliminated from large-scale¹ models make themselves felt in other ways: a ball of crumpled paper experiences a strong drag force in flight due to air-resistance, a constitutive equation must be provided to fix the *visco-elastic* properties of a fluid, *etc.* Drag forces, viscosity, electrical resistance – these are all vestiges of the microscopic dynamics left behind in the macroscopic models when the enormous degrees of freedom in the original many-body problem were integrated away to leave a simple deterministic description. These shadows

¹Also called *macroscopic* or *phenomenological* models.

of microscopic motion are often called *fluctuations* or *noise*, and their description and characterization will be the focus of this course.

Deterministic models (typically written in terms of systems of ordinary differential equations) have been very successfully applied to an endless variety of physical phenomena (*e.g.* Newton's laws of motion); however, there are situations where deterministic laws alone are inadequate, particularly as the number of interacting species becomes small and the individual motion of microscopic bodies has observable consequences.

The most famous example of observable fluctuations in a physical system is *Brownian motion*: the random, incessant meandering of a pollen grain suspended in a fluid, successfully described by Einstein in 1905. (We shall consider Brownian motion in greater detail throughout these notes.) The next major developments in the study of noise came in the 1950's and 1960's through the analysis of electrical circuits and radio wave propagation. Today, there is again renewed interest in the study of stochastic processes, particularly in the context of microfluidics, nanoscale devices and the study of genetic circuits that underlie cellular behaviour.

Why study Brownian motion?

Brownian motion will play a central role in the development of the ideas presented in these notes. This is partly historical because much of the mathematics of stochastic processes was developed in the context of studying Brownian motion, and partly pedagogical because Brownian motion provides a straightforward and concrete illustration of the mathematical formalism that will be developed. Furthermore, Brownian motion is a simple enough *physical* system that the limitations of the various assumptions employed in the modeling of physical phenomena are made obvious.

1.2 Brownian motion

Using a microscope, Robert Brown (1773-1858) observed and documented the motion of large pollen grains suspended in water. No matter what Brown did, the pollen grains moved incessantly in erratic and irregular motion (Figure 1.1). Brown tried different materials and different solvents, and still the motion of these particles continued. This was a time when most scientists did not believe in atoms or molecules, so the underlying mechanism responsible remained a mystery for nearly a century. In the words of S. G. Brush, "three quarters of a century of experiments produced almost no useful results in the understanding of Brownian mo-

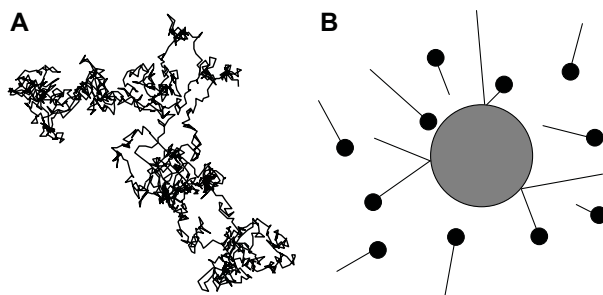


Figure 1.1: **Brownian motion** A) Sample path of Brownian motion. B) Schematic of the physical mechanism underlying Brownian motion. Notice the figure is not to scale – In reality, there would be about 10^9 solvent molecules in the view shown, and each molecule would be about 1/100 mm in diameter.

tion – simply because no theorist had told the experimentalists what to measure!”

Thanks mainly to the work of Guoy [*J. de Physique* **7**, 561 (1888)], several suggestive facts were known by century’s end:

1. The motion of the particles is *very* irregular, and the path appears to have no tangent at any point. (On an observable scale, the path appears non-differentiable, though of course it is continuous.)
2. The particles appear to move *independently* of one another, even when they approach closely.
3. The molecular composition and mass density of the particles has no effect.
4. As the solvent viscosity is decreased, the motion becomes more active.
5. As the particle radius is decreased, the motion becomes more active.
6. As the ambient temperature is increased, the motion becomes more active.
7. The motion never ceases.

Such was the state of affairs when Einstein published his seminal paper in 1905 [*Ann. Phys.* **17**, 549 (1905)], compelling him to write in the opening paragraph of that article, that “it is possible that the movements

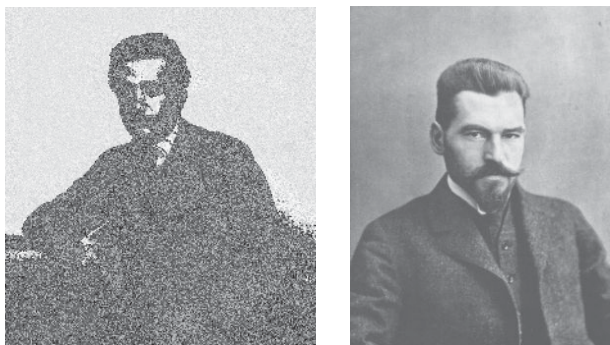


Figure 1.2: **Albert Einstein and Marian Smoluchowski**

to be discussed are identical with the so-called ‘Brownian molecular motion’; however, the information available to me regarding the latter is so lacking in precision, that I can form no judgment in the matter.”

1.2.1 Einstein

Einstein’s argument is a classic – and it contains the seeds of *all* of the ideas developed in this course. To appreciate the revolutionary nature of Einstein’s formulation of the problem, it is important to recall that at the time many physicists did not believe in atoms or molecules. The argument is built upon two key hypotheses. First, that the motion of the particle is caused by many frequent impacts of the incessantly moving solvent molecules. Second, that the effect of the complicated solvent motion can be described *probabilistically* in terms of very frequent *statistically independent* collisions.

Einstein’s approach follows two directions – the first is physical. Balancing the osmotic pressure with the body forces acting on the particle, Einstein derives a partial differential equation for the particle density $f(x, t)$,

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}.$$

Here, the diffusion coefficient D is a measure of how quickly the density distribution will spread along the X -axis. For a spherical Brownian particle of radius R , using Stokes’s theory for the drag force acting on the particle, Einstein arrives at the following expression for D ,

$$D = \frac{k_B T}{6\pi\nu R}. \quad (1.1)$$

Here ν is the shear viscosity of the suspending liquid, T is the absolute temperature, and k_B is Boltzmann's constant. The units of k_B are such that $k_B T$ has the units of energy. Knowledge of the value of k_B is equivalent to the knowledge of Avogadro's number, and hence of molecular size. Notice how the qualitative aspects of Brownian motion (items 4-6 from the table on p. 3) are reflected in the expression for the diffusion coefficient.

The second direction of Einstein's paper is mathematical, using simple assumptions to derive a partial differential equation governing the *probability* density of finding the Brownian particles at a given position x after a time t – a derivation that will be echoed throughout this course. The original article is a wonderful piece of scientific writing, not only because of the historical significance, but because of the clarity and force of the central ideas. For that reason, the portion of the paper describing the derivation of the partial differential equation is reproduced below.

Excerpt from Einstein's paper on Brownian motion (Author's translation)

- A. Einstein (1905) *Ann. Physik* **17**:549.

It must clearly be assumed that each individual particle executes a motion which is independent of the motions of all other particles; it will also be considered that the movements of one and the same particle in different time intervals are independent processes, as long as these time intervals are not chosen too small.

We introduce a time interval τ into consideration, which is very small compared to the observable time intervals, but nevertheless so large that in two successive time intervals τ , the motions executed by the particle can be thought of as events which are independent of each other.

Now let there be a total of n particles suspended in a liquid. In a time interval τ , the X -coordinates of the individual particles will increase by an amount Δ , where for each particle Δ has a different (positive or negative) value. There will be a certain *frequency law* for Δ ; the number dn of the particles which experience a shift which is between Δ and $\Delta + d\Delta$ will be expressible by an equation of the form

$$dn = n\phi(\Delta)d\Delta, \quad (1.2)$$

where

$$\int_{-\infty}^{\infty} \phi(\Delta) d\Delta = 1, \quad (1.3)$$

and ϕ is only different from zero for very small values of Δ , and satisfies the condition

$$\phi(\Delta) = \phi(-\Delta). \quad (1.4)$$

We now investigate how the diffusion coefficient depends on ϕ . We shall once more restrict ourselves to the case where the number ν of particles per unit volume depends only on x and t .

Let $\nu = f(x, t)$ be the number of particles per unit volume. We compare the distribution of particles at the time $t + \tau$ from the distribution at time t . From the definition of the function $\phi(\Delta)$, it is easy to find the number of particles which at time $t + \tau$ are found between two planes perpendicular to the x -axis and passing through the points x and $x + dx$. One obtains²

$$f(x, t + \tau) dx = dx \int_{-\infty}^{\infty} f(x - \Delta, t) \phi(\Delta) d\Delta. \quad (1.5)$$

But since τ is very small, we can set

$$f(x, t + \tau) = f(x, t) + \tau \frac{\partial f}{\partial t}. \quad (1.6)$$

Furthermore, we develop $f(x - \Delta, t)$ in powers of Δ :

$$f(x - \Delta, t) = f(x, t) - \Delta \frac{\partial f}{\partial x} + \frac{\Delta^2}{2!} \frac{\partial^2 f}{\partial x^2} + \dots \quad (1.7)$$

We can use this series under the integral, because only small values of Δ contribute to this equation. We obtain

$$f + \frac{\partial f}{\partial t} \tau = f \int_{-\infty}^{\infty} \phi(\Delta) d\Delta - \frac{\partial f}{\partial x} \int_{-\infty}^{\infty} \Delta \phi(\Delta) d\Delta + \frac{\partial^2 f}{\partial x^2} \int_{-\infty}^{\infty} \frac{\Delta^2}{2} \phi(\Delta) d\Delta. \quad (1.8)$$

²Einstein actually wrote $f(x + \Delta, t)$ in the integrand, which is incorrect (Why?). The sign has been corrected in all subsequent equations. In the case considered by Einstein, however, the collisions are unbiased: $\phi(\Delta) = \phi(-\Delta)$, so the sign of Δ makes no difference whatsoever.

Because $\phi(\Delta) = \phi(-\Delta)$, the second, fourth, etc., terms on the right-hand side vanish, while out of the 1st, 3rd, 5th, etc., terms, each one is very small compared with the previous. We obtain from this equation, by taking into consideration

$$\int_{-\infty}^{\infty} \phi(\Delta) d\Delta = 1, \quad (1.9)$$

and setting

$$\frac{1}{\tau} \int_{-\infty}^{\infty} \frac{\Delta^2}{2} \phi(\Delta) d\Delta = D, \quad (1.10)$$

and keeping only the 1st and third terms of the right-hand side,

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2} \dots \quad (1.11)$$

This is already known as the differential equation of diffusion and it can be seen that D is the diffusion coefficient. ...

The problem, which corresponds to the problem of diffusion from a single point (neglecting the interaction between the diffusing particles), is now completely determined mathematically; its solution is

$$f(x, t) = \frac{n}{\sqrt{4\pi D}} \frac{\exp\left[-\frac{x^2}{4Dt}\right]}{\sqrt{t}} \dots \quad (1.12)$$

We now calculate, with the help of this equation, the displacement λ_x in the direction of the X -axis that a particle experiences on the average or, more exactly, the square root of the arithmetic mean of the square of the displacement in the direction of the X -axis; it is

$$\lambda_x = \sqrt{\langle x^2 \rangle} = \sqrt{2Dt}. \quad (1.13)$$

(End of excerpt)

The temperature T and the viscosity of the surrounding fluid, ν , can be measured, and with care a suspension of spherical particles of fairly uniform radius R may be prepared. The diffusion coefficient D is then

available through the *statistics* of the Brownian particle (via Eq. 1.13). In this way, Boltzmann's constant (or, equivalently Avogadro's number) can be determined. This program was undertaken in a series of laborious experiments by Perrin and Chaudesaigues – resulting in a Nobel prize for Perrin in 1926. The coupling of Eq. 1.1 and Eq. 1.13 is one of the earliest examples of using fluctuations to quantify a physical parameter (as opposed to simply considering the averaged behaviour). Said another way, it is the *irreproducibility* in the experiment that is reproducible!

Einstein's argument does not give a dynamical theory of Brownian motion; it only determines the nature of the motion and the value of the diffusion coefficient on the basis of some assumptions. Smoluchowski, independently of Einstein, attempted a dynamical theory and arrived at Eq. 1.1 with an additional factor of $\frac{32}{27}$ on the right-hand side (Section 1.2.2). Smoluchowski also lay the foundation for a more extensive study of Brownian motion, including the possibility of various forces acting upon the particle. Langevin provided another derivation of Eq. 1.1 (Section 1.2.3) which was the starting point for the work of Ornstein and Uhlenbeck (Section 1.2.4).

Einstein's work was of great importance in physics, for it showed in a visible and concrete way that atoms were real. Quoting from Einstein's autobiography:

The agreement of these considerations with experience together with Planck's determination of the true molecular size from the law of radiation (for high temperature) convinced the skeptics, who were quite numerous at the time (Ostwald, Mach) of the reality of atoms. The antipathy of these scholars towards atomic theory can indubitably be traced back to their positivistic philosophical attitude. This is an interesting example of the fact that even scholars of audacious spirit and fine instinct can be obstructed in the interpretation of the facts by philosophical prejudices.

Relation between Einstein's derivation and the rest of the course

Einstein's work anticipated much of what was to come in the study of stochastic processes in physics. A few key steps will be highlighted, with reference to their appearance later in the course notes. Quoting from Gardiner's *Handbook*,

- *The Chapman-Kolmogorov Equation* (Eq. 1.5) states that the probability of the particle being at point x at time $t + \tau$ is given by

the sum of probabilities of all possible “pushes” Δ from positions $x - \Delta$, multiplied by the probability of being at $x - \Delta$ at time t . This assumption is based on the independence of the push Δ of any previous history of the motion: it is only necessary to know the initial position of the particle at time t – not at any previous time. This is the *Markov postulate* and the Chapman-Kolmogorov equation, of which Eq. 1.5 is a special form, is the central dynamical equation to all Markov processes. The Chapman-Kolmogorov equation and Markov processes will be studied in detail in Chapter 3.

- *The Fokker-Planck Equation:* (Eq. 1.11) is the diffusion equation, a special case of the Fokker-Planck equation, which describes a large class of very interesting stochastic processes in which the system has a continuous sample path. In this case, that means that the pollen grain’s position, if thought of as obeying a probabilistic law given by solving the diffusion equation, Eq. 1.11, in which time t is continuous (not discrete, as assumed by Einstein), can be written $x(t)$, where $x(t)$ is a *continuous function of time* – but a random function. This leads us to consider the possibility of describing the dynamics of the system in some direct probabilistic way, so that we would have a *random* or *stochastic differential equation for the path*. This procedure was initiated by Langevin with the famous equation that to this day bears his name. [see the following section]. We will discuss this in detail in at the end of Chapter 6.
- *The Kramers-Moyal* and similar expansions are essentially the same as that used by Einstein to go from Eq. 1.5 to Eq. 1.11. The use of this type of approximation, which effectively replaces a process whose sample paths need not be continuous with one whose paths are continuous, has been a topic of discussion in the last decade. Its use and validity will be discussed in Chapters 6 and 4.
- *The Fluctuation-Dissipation Relation:* Eqs. 1.1 and 1.13 connect the dissipation in the system with the magnitude of the fluctuations in thermal equilibrium (see Section 2.3.2 and Section 5.1, p. 105).

1.2.2 Smolouchowski

- M. von Smoluchowski (1906) *Ann. Physik* **21**:756.

Independently of Einstein, Smoluchowski developed a complementary model of Brownian motion. His original derivation has more in common

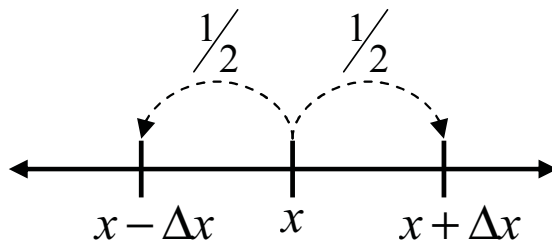


Figure 1.3: **Schematic illustration of an unbiased random walk in one dimension.** A particle at position x can move to neighbouring positions $x + \Delta x$ and $x - \Delta x$ with equal probability during each time-step $(t, t + \tau)$.

with Rayleigh's study of Brownian motion (see Exercise 9 on p. 149), although the essence of his argument, and his legacy to the modeling of stochastic processes, can be appreciated by considering a *spatially-discretized* version of Eq. 1.5.

In addition to a discretization of time $(t, t + \tau, t + 2\tau, \text{ etc.})$, imagine the motion of the particle is likewise constrained to discrete positions $\dots, x - \Delta x, x, x + \Delta x, \dots$ (Figure 1.3). To fully characterize the model, one assigns a starting position (generally $x = 0$ at $t = 0$ for convenience) and a transition probability between the lattice points.

Smoluchowski's model of Brownian motion assumes that a particle can move only one step to the right or to the left, and that these occur with equal probability. This type of process is called an *unbiased random walk*, an example of a more general class of *one-step processes*. Mathematically, the conditional probability density $P(x, t)$ for the unbiased random walk obeys a difference equation,

$$P(x, t + \tau) = \frac{1}{2} P(x - \Delta x, t) + \frac{1}{2} P(x + \Delta x, t), \quad P(x, 0) = \delta_{x,0}, \quad (1.14)$$

where δ_{ij} is the Kronecker delta. The two terms in the right-hand side of Eq. 1.14 represent the two ways in which a particle can move into position x during the interval $(t, t + \tau)$. Eq. 1.14 can be solved explicitly for the full time-dependent conditional distribution $P(x, t)$ using transform methods (see Exercise 7 on p. 25, and Chapter 4). Furthermore, in a certain limit, the Smoluchowski model reduces to Einstein's diffusion equation, Eq. 1.11 (see Exercise 6 on p. 24).

In many applications where x represents physical space, discretization is unnatural. In problems where the state space is inherently discrete, such

as molecule numbers or animal populations, Smoluchowski's methodology is particularly convenient. It is not difficult to imagine whole classes of more complicated processes that can be constructed along similar lines by allowing, for example, multiple steps, biased transition rates, state-dependent transition rates, multidimensional lattices, *etc.* These more general models will be explored in more detail in Section 3.3.

Relation between Smoluchowski's formulation and the rest of the course

- *The master equation:* Smoluchowski's conceptual framework is ideal for many physical applications. Many physical systems, with an appropriately chosen coarse-graining of the observation time-scale, can be represented by Markov processes (Chapter 3). The main dynamical equation for a Markov process is the Chapman-Kolmogorov equation (see p. 52). Under the same coarse-graining of the observation time-scale that permits the physical process to be described as Markovian (see p. 59), the Chapman-Kolmogorov equation reduces to the more simple *master equation*. The random walk (Eq. 1.14) is an example of a master equation with discrete states. In practice, the physics of the process are captured by the transition probabilities among lattice sites, then the master equation is used to predict how the state evolves in time. The master equation is rarely solvable exactly, so a large portion of the course will be devoted to simulating or approximating the solutions. The master equation and Markov processes will be studied in detail in Chapter 3.

1.2.3 Langevin

- P. Langevin (1908) *C. R. Acad. Sci. (Paris)* **146**:530.
- English translation and commentary: D. S. Lemons and A. Gythiel (1997) *Am. J. Phys.* **65**: 1079–1081.

A few years after Einstein and Smoluchowski's work, a new method for studying Brownian motion was developed by Paul Langevin. Langevin arrives at the same result for the mean-squared displacement as Einstein, although coming from a very different perspective and following [in his words] an “infinitely simpler” derivation. Beginning with the trajectory of a single particle, his work marks the first attempt to provide a *dynamical* theory of Brownian motion (to be contrasted with Einstein's *kinematical* theory). Langevin's argument is as follows.

Figure 1.4: **Paul Langevin with Einstein**

A single Brownian particle is subject to Newton's laws of motion. The Brownian particle moves around because it is subject to a resultant force F coming from the collision with the solvent molecules. With enough information about F , Brownian motion can be studied using Newton's second law, $F = ma$,

$$m \frac{d^2 x}{dt^2} = F. \quad (1.15)$$

The precise, microscopic form of the force acting on the particle is unknown; however, from a *macroscopic* point of view, it is reasonable to consider it as the resultant of two different forces:

1. **Viscous drag** $f_d = -\gamma \frac{dx}{dt}$ ($\gamma = 6\pi\nu R$) for a sphere of radius R in slow motion in a fluid with shear viscosity ν (Stokes flow).
2. **Fluctuating (random) force** $f_r(t)$ due to the incessant molecular impacts, about which next to nothing is known.

Under these assumptions, the equation of motion for the Brownian particle is,

$$m \frac{d^2 x}{dt^2} = -\gamma \frac{dx}{dt} + f_r(t), \quad (1.16)$$

or, after multiplication with x and re-arrangement,

$$\frac{m}{2} \frac{d^2}{dt^2}(x^2) - mv^2 = -\frac{\gamma}{2} \frac{d}{dt}(x^2) + x f_r(t), \quad (1.17)$$

Langevin remarks that “since the quantity $f_r(t)$ has great irregularity” Eq. 1.17 is not a differential equation in the ordinary sense. In fact, since the forcing $f_r(t)$ is a random function, so, too, will be the path $x(t)$. The path $x(t)$ was not Langevin’s objective, however. Assuming Eq. 1.17 has a well-defined meaning in a statistical sense, he proceeded to *average* the equation over a large number of similar (but different) particles in Brownian motion under the same conditions.

Denote this average by $\langle \cdot \cdot \cdot \rangle$, and assume that it commutes with the derivative operator $\frac{d}{dt}$, then from Eq. 1.17,

$$\frac{m}{2} \frac{d^2}{dt^2} \langle x^2 \rangle - \langle mv^2 \rangle = -\frac{\gamma}{2} \frac{d}{dt} \langle x^2 \rangle + \langle x f_r(t) \rangle. \quad (1.18)$$

From the equipartition of energy theorem derived in equilibrium statistical mechanics, Langevin was able to replace the average kinetic energy of the particle $\langle mv^2 \rangle$ by $k_B T$, where k_B is Boltzmann’s constant and T is the absolute temperature³. The cross-correlation $\langle x f_r(t) \rangle$ Langevin dismissed by saying “it can be set = 0, due to the great irregularity of $f_r(t)$.” The resulting differential equation for the mean-squared displacement $\langle x^2 \rangle$ is then simply,

$$m \frac{d^2}{dt^2} \langle x^2 \rangle + \gamma \frac{d}{dt} \langle x^2 \rangle = 2k_B T. \quad (1.19)$$

Integrating once,

$$\frac{d}{dt} \langle x^2 \rangle = \frac{2k_B T}{\beta} + C e^{-\Omega t}; \quad \left(C = \text{constant}, \Omega = \frac{\gamma}{m} \right). \quad (1.20)$$

Langevin proceeds by noting that experimentally $\Omega \approx 10^8$ Hz; so the exponential may be dropped after a time period of about $10^{-8} s$ has elapsed. Another integration yields the final result,

$$\langle x^2 \rangle - \langle x_0^2 \rangle = 2 \frac{k_B T}{\gamma} t = 2Dt, \quad (1.21)$$

if we identify the diffusion coefficient with Einstein’s result $D = \frac{k_B T}{\gamma} = \frac{k_B T}{6\pi\eta R}$.

It should be clear from Langevin’s equation that stochastic processes afford a tremendous simplification of the governing model. Instead of

³The equipartition of energy theorem comes from classical statistical mechanics (in the absence of quantum effects). It states that, at thermal equilibrium, energy is shared equally among all forms and all degrees of freedom. For an ideal gas, the average kinetic energy in thermal equilibrium of each molecule is $(3/2)k_B T$. In the x -component of the velocity, the energy is $(1/2)k_B T$ – this is the result used by Langevin.

including $\sim 10^{23}$ auxiliary variables describing the movement of the solvent molecules, Langevin includes their effect as the single random forcing function $f_r(t)$, leading to a closed equation for the particle position $x(t)$

– Quoting van Kampen:

Together with its surrounding fluid [the Brownian particle] constitutes a closed, isolated system. The “relevant” variable x is the position of the particle, which constitutes a stochastic process. . . [This] is due to the fact that the single variable x does not really obey a closed differential equation, but interacts with a host of fluid molecules. . . These variables are not visible in the equation for x but their effect shows up in the random Langevin force. Fluctuations in x are constantly being generated by the collisions.

Notice, in retrospect, the similarity between $\phi(\Delta)$ of Einstein and $f_r(t)$ of Langevin – both characterize the fluctuations. Nevertheless, while $\phi(\Delta)$ describes the *effect* of the force imparted by the collisions (a kinematic approach), $f_r(t)$ treats the force *itself* as a random variable (a dynamic approach). As a result, Einstein derives a partial differential equation governing the entire probability distribution, whereas Langevin derives an ordinary differential equation for the first two moments of the distribution.

Several critical remarks can be made regarding Langevin’s derivation, and criticisms were not long in coming.

1. What, precisely, is the function $f_r(t)$? We know what a random variable is, and we know what a function is – but $f_r(t)$ is a *random function*. What is the mathematical meaning of that statement?
2. If $f_r(t)$ is a random function, then $x(t)$ should inherit some of that randomness. We assume $x(t)$ obeys a differential equation, but how do we know the derivatives $\frac{dx}{dt}$ and $\frac{d^2x}{dt^2}$ even exist?
3. What kind of averaging are we using in moving from Eq. 1.17 to Eq. 1.18? Are we picking some *a priori* probability distribution and computing the expectation? If so, how do we choose the probability distribution? Are we taking an arithmetical average?
4. Setting $Ce^{-\Omega t} \approx 0$ because Ω is large only holds if t is large. What is going on over very short timescales?

Soon after the 1908 paper by Langevin, several outstanding contributions to the theory of Brownian motion appeared. Of particular interest is

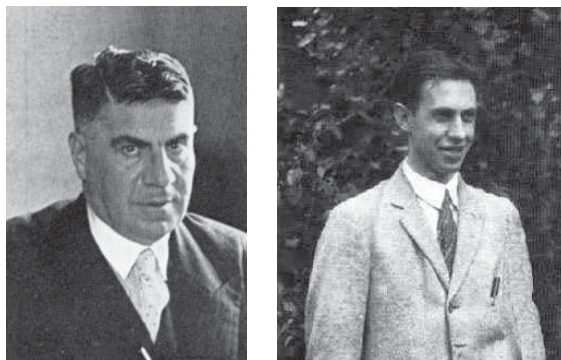


Figure 1.5: Leonard Ornstein and George Uhlenbeck

the 1930 paper by Uhlenbeck and Ornstein that sought to strengthen the foundations of Langevin's approach.

Relation between Langevin's approach and the rest of the course

- *Random differential equations:* Langevin's formulation of stochastic processes in terms of Newton's law's of motion still appeals to many physicists. For systems governed by linear dynamics in the absence of fluctuations, Langevin's approach is very intuitive and simple to formulate. For nonlinear systems, however, Langevin's approach cannot be used. Differential equations with stochastic forcing or stochastic coefficients are the subject of Chapter 8. The difficulties of Langevin's approach for nonlinear equations is discussed in N. G. van Kampen's article 'Fluctuations in Nonlinear Systems.'
- *Stochastic analysis:* The cavalier treatment of the analytic properties of the random force $f_r(t)$ and the resulting analytic properties of the stochastic process $x(t)$ generated extensive activity by mathematicians to solidify the foundations of stochastic calculus. Of particular note is the work of Kolmogorov (see p. 30), Doob (p. 20) and Itô (p. 270).

1.2.4 Ornstein and Uhlenbeck

The Ornstein-Uhlenbeck paper (G. E. Uhlenbeck and L. S. Ornstein (1930) *Phys. Rev.* **36**:823) begins by reminding the reader that the central result of both Einstein and Langevin is the mean-squared displacement of the Brownian particle; Einstein obtained the result (γ is the

friction coefficient),

$$\langle x^2 \rangle = 2Dt; \quad D = \frac{k_B T}{\gamma}, \quad (1.22)$$

kinematically, from the evolution equation of the probability distribution, while many workers preferred Langevin's *dynamical* approach based upon Newton's law,

$$m \frac{du}{dt} = -\gamma u + f_r(t); \quad u \equiv \frac{dx}{dt}, \quad (1.23)$$

along with some "natural" assumptions about the statistical properties of $f_r(t)$.

Next, Ornstein and Uhlenbeck go on to observe that the Brownian motion problem is still far from solved, and that the two following questions have yet to be answered:

- Einstein, Smoluchowski and Langevin all assumed a time scale such that

$$t \gg \frac{m}{\gamma} \quad (1.24)$$

Under what conditions is this assumption justified?

- What changes must be made to the formulation if the Brownian particle is subject to an external field?

The Ornstein and Uhlenbeck paper is devoted to answering these outstanding questions. Their paper, too, is a classic – We paraphrase it below, although the reader is encouraged to seek out the original.

The problem is to determine the probability that a free particle in Brownian motion after the time t has a velocity which lies between u and $u + du$, when it started at $t = 0$ with velocity u_0 .

They declare their intention of using Langevin's equation,

$$\frac{du}{dt} + \beta u = A(t); \quad \beta = \frac{\gamma}{m}, \quad A(t) = \frac{f(t)}{m}, \quad (1.25)$$

whose solution is,

$$u = u_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta \xi} A(\xi) d\xi. \quad (1.26)$$

(By *solution*, they mean an integration of the Langevin equation. Is that sensible?) On taking the average over the initial ensemble, using the assumption that $\langle A(t) \rangle = 0$ as Langevin did,

$$\langle u \rangle_{u_0} = u_0 e^{-\beta t}. \quad (1.27)$$

The subscript means ‘average of all u that began with velocity u_0 at time $t = 0$.’ Squaring Eq. 1.26 and taking the average,

$$\langle u^2 \rangle_{u_0} = u_0^2 e^{-2\beta t} + e^{-2\beta t} \int_0^t \int_0^t e^{\beta(\xi+\eta)} \langle A(\xi) A(\eta) \rangle d\xi d\eta. \quad (1.28)$$

To make progress, we now assume that,

$$\langle A(t_1) A(t_2) \rangle = \phi_1(t_1 - t_2) \quad (1.29)$$

where ϕ_1 is a function with values close to zero everywhere *except at the origin* where it has a *very sharp peak*. Physically, this is justifiable on the grounds that the fluctuating function $A(t)$ has values that are correlated only for very short time instants. As we will see in the coming chapters, ϕ_1 is called the *correlation function* or *autocorrelation* of $A(t)$. The dependence on time difference only, $t_1 - t_2$, is typical of a special class of stochastic processes.

In Ornstein and Uhlenbeck’s own words:

[W]e will naturally make the following assumptions... There will be correlation between the values of [the random forcing function $A(t)$] at different times t_1 and t_2 only when $|t_1 - t_2|$ is very small. More explicitly we shall suppose that:

$$\langle A(t_1) A(t_2) \rangle = \phi(t_1 - t_2),$$

where $\phi(x)$ is a function with a very sharp maximum at $x = 0$.

Making the change of variable $\xi + \eta = v$ and $\xi - \eta = w$, the integration can be carried out,

$$\langle u^2 \rangle_{u_0} = u_0^2 e^{-2\beta t} + \frac{\tau_1}{2\beta} (1 - e^{-2\beta t}); \quad \tau_1 \equiv \int_{-\infty}^{\infty} \phi_1(w) dw, \quad (1.30)$$

(where the integration limits in the definition of τ_1 extend to $\pm\infty$ because $\phi_1 \approx 0$ except near the origin). Recalling that the equipartition of energy theorem requires that,

$$\lim_{t \rightarrow \infty} \langle u^2 \rangle = \frac{k_B T}{m}, \quad (1.31)$$

we have – in that same limit – that $\frac{\tau_1}{2\beta} = \frac{k_B T}{m}$, so,

$$\tau_1 = \frac{2\beta k_B T}{m}, \quad (1.32)$$

and the final result for the mean-squared velocity of the Brownian particle is,

$$\langle u^2 \rangle_{u_0} = u_0^2 e^{-2\beta t} + \frac{k_B T}{m} (1 - e^{-2\beta t}). \quad (1.33)$$

If, in addition to Eq. 1.29, it is assumed that fluctuations $A(t)$ are *Gaussian* random variables at every time t , it is then possible to show that the random quantity $u - u_0 e^{-\beta t}$ also obeys a Gaussian (normal) probability distribution given by,

$$G(u, u_0, t) = \left[\frac{m}{2\pi k_B T (1 - e^{-2\beta t})} \right]^{\frac{1}{2}} \exp \left[-\frac{m}{2k_B T} \frac{(u - u_0 e^{-\beta t})^2}{(1 - e^{-2\beta t})} \right]. \quad (1.34)$$

Having derived the velocity distribution, Ornstein and Uhlenbeck proceed to calculate the distribution for the displacement.

The problem is to determine the probability that a free particle in Brownian motion which at $t = 0$ starts from $x = x_0$ with velocity u_0 will lie, after a time t , between x and $x + dx$. It is clear that this probability will depend on $s = x - x_0$ only, and on t .

In analogy with the velocity, they begin with Eq. 1.26,

$$u = u_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta \xi} A(\xi) d\xi, \quad (1.35)$$

and integrate, with the result,

$$x = x_0 + \frac{u_0}{\beta} (1 - e^{-\beta t}) + \int_0^t e^{-\beta \eta} \left[\int_0^\eta e^{\beta \xi} A(\xi) d\xi \right] d\eta; \quad (1.36)$$

integrating partially this may be written in the form,

$$s = x - x_0 = \frac{u_0}{\beta} (1 - e^{-\beta t}) - \frac{1}{\beta} e^{-\beta t} \int_0^t e^{\beta \xi} A(\xi) d\xi + \frac{1}{\beta} \int_0^t A(\xi) d\xi. \quad (1.37)$$

Taking the mean gives,

$$\langle s \rangle_{u_0} = \frac{u_0}{\beta} (1 - e^{-\beta t}). \quad (1.38)$$

By squaring, averaging, and computing the double integral as above, the mean-squared displacement is,

$$\langle s^2 \rangle_{u_0} = \frac{\tau_1}{\beta^2} t + \frac{u_0^2}{\beta^2} (1 - e^{-\beta t})^2 + \frac{\tau_1}{2\beta^3} (-3 + 4e^{-\beta t} - e^{-2\beta t}), \quad (1.39)$$

where τ_1 is as above. This result was first obtained by Ornstein (1917). In the limit of long times, it goes over to Einstein's result,

$$\langle s^2 \rangle_{u_0} \sim \frac{\tau_1}{\beta^2} t \sim 2 \frac{k_B T}{m\beta} t \quad \text{as } t \rightarrow \infty. \quad (1.40)$$

For short times, on the other hand, we have,

$$\langle s \rangle_{u_0} \sim u_0 t \text{ and } \langle s^2 \rangle_{u_0} \sim u_0^2 t^2 \quad \text{as } t \rightarrow 0, \quad (1.41)$$

i.e. the motion is uniform with velocity u_0 . Taking a second average over u_0 (using the distribution derived above), and since $\langle u_0^2 \rangle = \frac{k_B T}{m}$, we have,

$$\langle \langle s \rangle \rangle = 0 \quad \langle \langle s^2 \rangle \rangle = 2 \frac{k_B T}{m\beta^2} (\beta t - 1 + e^{-\beta t}). \quad (1.42)$$

The calculation of higher powers proceeds similarly. In this way, one can show that the variable,

$$S = s - \frac{u_0}{\beta} (1 - e^{-\beta t}), \quad (1.43)$$

is Gaussian distributed, too, with

$$F(x, x_0, t) = \left[\frac{m\beta^2}{2\pi k_B T (2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t})} \right]^{\frac{1}{2}} \times \exp \left[-\frac{m\beta^2}{2k_B T} \frac{\left(x - x_0 - \frac{u_0}{\beta} (1 - e^{-\beta t}) \right)^2}{(2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t})} \right], \quad (1.44)$$

under the assumption that $A(t)$ is Gaussian. The Ornstein-Uhlenbeck paper contains more than the results described here. For instance, they re-derive the equation using what is called the Fokker-Planck equation

(which we shall meet later, in Chapter 6). They also consider the effect of an external force, by solving the problem of an harmonically-bound particle,

$$\frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + \omega^2 x = A(t). \quad (1.45)$$

Although the approach of Langevin was improved and expanded upon by Ornstein and Uhlenbeck, some more fundamental problems remained, as Doob pointed out in the introduction to his famous paper of 1942 (J. L. Doob (1942) *Annals of Math.* **43**:351),

The purpose of this paper is to apply the methods and results of modern probability theory to the analysis of the Ornstein-Uhlenbeck distribution, its properties, and its derivation... A stochastic differential equation is introduced in a rigorous fashion to give a precise meaning to the Langevin equation for the velocity function. *This will avoid the usually embarrassing situation in which the Langevin equation, involving the second-derivative of x , is used to find a solution $x(t)$ not having a second-derivative.*

Obviously there is something deep going on in terms of the differentiability and integrability of a stochastic process. In particular, to make sense of Doob's statement, we need to know what is meant by *differentiability* in this context. That will be the focus of Chapter 7.

Relation between Ornstein and Uhlenbeck's work and the rest of the course

- *Fokker-Planck equation*: In addition to clarifying the assumptions left implicit by Langevin, and permitting a generalization of Langevin's approach, Ornstein and Uhlenbeck provided a mathematical connection between the kinematic approach of Einstein (using a partial differential equation to describe the probability distribution) and the dynamical approach of Langevin (using an ordinary differential equation with random forcing to describe a single trajectory). That connection will be made explicit in Chapter 6.

1.3 Outline of the Book

In broad strokes, the outline of book is as follows: First, the fundamental concepts are introduced –

Chapter 2 - Random processes: The ideas of classical probability theory (Appendix A) are extended from random variables to the concept of random *functions*. We focus on a particular class of stochastic processes – called *stationary processes* – that will be used throughout the course. Particular attention is paid to the second moment of a stochastic function evaluated at different times, called the *correlation function*.

After the fundamentals have been established, we examine how stochastic processes are classified, with particular focus on the most common process appearing in the modeling of physical systems: *The Markov process*.

Chapter 3 - Markov processes: The most commonly used stochastic descriptor with the property that the future state of a system is determined by the present state, and not on any states in the past. This is a strong requirement, but it is often approximately satisfied by many physical systems and allows a useful coarse-graining of the microscopic dynamics. The evolution equation for the probability distribution of a Markov process is often called the *master equation*, a difficult equation to solve exactly. The remaining chapters are, to a large extent, methods of approximating the master equation so that it is more amenable to solution.

Once Markov processes and their evolution equations have been derived, we examine in more detail solution methods and the kinematic formulation of Einstein.

Chapter 4 - Solution of the Master Equation: Briefly discuss exact solution methods. In most cases of interest, the master equation is too difficult to solve exactly. We examine the most popular approximation method – stochastic simulation.

Chapter 5 - Perturbation Expansion of the Master Equation: The linear noise approximation of van Kampen is a robust and widely-applicable analytic approximation method for the solution of the Master equation.

Chapter 6 - Fokker-Planck Equation: This is a partial differential diffusion equation with drift that governs the entire probability distribution – a generalization of the diffusion equation used by Einstein in his study of Brownian motion. A popular analytic method of approximating the evolution of a Markov process. There are some ambiguities and essential difficulties associated with this method when applied to systems with nonlinear transition probabilities.

Before returning to the dynamic (stochastic differential equation) meth-

ods of Langevin, Ornstein and Uhlenbeck, we must first formulate the calculus of random functions.

Chapter 7 - Stochastic Analysis: We shall be interested in the *mean-square limits* of sequences as the foundation of continuity and the operations of differentiation and integration. More refined ideas of integration will be required to provide a formal meaning to expressions involving white noise as the canonical noise source (Itô's calculus).

Chapter 8 - Random Differential Equations: The differential equation-based modeling of Langevin, Ornstein and Uhlenbeck has many appealing features. We examine generalizations of their analysis, with particular focus on numerical simulation and analytic approximation of solutions.

Combining the ideas of the previous chapters, we characterize the effect of fluctuations on the macroscopic behaviour of nonlinear systems.

Chapter 9 - Macroscopic effects of noise: There are situations where stochastic models exhibit behaviour that is in sharp contrast with deterministic models of the same system. We examine several such cases, and the analytic tools that can be used to study these effects.

Chapter 10 - Special Topics: Some additional topics are included which are beyond the scope of the course, though perhaps of interest to some readers.

1.4 Suggested References

Much of the content on Brownian motion comes from C. W. Gardiner's book:

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004).

Beyond the first chapter, it is difficult reading. As claimed, it is a handbook, so it is a great starting point if you're looking for specific information about a specific stochastic process.

One of the finest books on stochastic processes is by N. G. van Kampen:

- *Stochastic processes in physics and chemistry (2nd Ed.)*, N. G. van Kampen (North-Holland, 2001).

It, too, is difficult reading, but well-worth the effort. He doesn't shy away from making emphatic and unambiguous statements that help in learning

the material. The exercises are excellent – difficult and rewarding.

A good supplemental text is the statistical mechanics book by Reif:

- *Fundamentals of statistical and thermal physics*, F. Reif (McGraw-Hill, 1965).

Exercises

1. Fill in the steps that connect Eq. 1.28 to Eq. 1.30. It may be helpful to treat $\phi_1(t_1 - t_2)$ in Eq. 1.29 as an idealized delta function $\delta(t_1 - t_2)$ (see Section B.3.2).
2. **Harmonically-bound particle:** Repeat Ornstein and Uhlenbeck's analysis to calculate $\langle x(t) \rangle$ and $\langle x^2(t) \rangle$ for the harmonically-bound particle, Eq. 1.45.
3. **Benveniste:** Read the paper “Human basophil degranulation triggered by very dilute antiserum against IgE,” by Davenas *et al.* Notice the disclaimer appended to the end of the article by the editor. Briefly (one or two sentences) summarize what the authors claim. Why does the editor say that “[t]here is no physical basis for such an activity?” Specifically, if the results reported in this paper are correct, what parts of the course would need to be revised.
4. **Fick's law:** A substance with concentration $c(x, t)$ undergoing diffusive transport in one-dimension obeys the diffusion equation:

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} = -\frac{\partial}{\partial x} J, \quad (1.46)$$

to a good level of approximation. Here, the flux term $J = -D\partial c/\partial x$ was discovered empirically by Fick (1855). At the time, the dependence of the flux upon the concentration gradient was interpreted as meaning particles bump into each other in locations of high concentration and consequently move to areas of lower concentration. That interpretation persists in many textbooks. How does the physical interpretation change given Einstein's work on Brownian motion (1905)?

5. **Delbrück and Luria.** Read S.E. Luria and M. Delbrück (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491–511.

- (a) What is the problem Luria and Delbrück set out to study? How will they distinguish among the possible hypotheses?
- (b) What relation can you see between this article and Einstein's article on Brownian motion? What parameter in the work of Luria and Delbrück plays the role of Avogadro's number in Einstein's analysis?
- (c) It was primarily this work that resulted in Luria and Delbrück winning the Nobel prize in 1969 (along with Hershey) – why do you think this work is so significant?

6. **Brownian motion – Smoluchowski's model:** Smoluchowski formulated Brownian motion as a random walk along a discrete lattice. In a certain limit, Smoluchowski's formulation coincides with Einstein's diffusion equation.

- (a) Suppose each step has length Δ and is taken after a time τ , with equal probability of moving to the left or to the right. Using a centered-difference scheme in Δ , find an expression for,

$$\frac{p_m(s+1) - p_m(s)}{\tau},$$

where $p_m(s)$ is the probability the particle, beginning at lattice site $m = 0$ at $t = 0$, is at lattice site m after s steps (m and s are integers, with $s \geq 0$). Taking the limit $\Delta, \tau \rightarrow 0$ in such a way that,

$$\frac{\Delta^2}{2\tau} = D, \quad m\Delta \rightarrow x, \quad s\tau = t,$$

show that the discrete random walk reduces to Einstein's diffusion equation,

$$\frac{\partial P}{\partial t} = D \frac{\partial^2 P}{\partial x^2}.$$

- (b) Suppose the movement is biased, with probability $p = \frac{1}{2} + \beta\Delta$ of moving to the left and $q = 1 - p = \frac{1}{2} - \beta\Delta$ of moving to the right after each step (where Δ must be small enough that $q > 0$). Under the same limiting conditions as above, how does the partial differential equation for $P(x, t)$ change?

Solve this equation for $P(x, t)$ subject to $P(x, t) \rightarrow 0$ and $\partial P / \partial x \rightarrow 0$ as $x \rightarrow \infty$, and a *reflecting barrier* at $x = 0$. That is, when the particle reaches the point $x = 0$ it must, in the next step, move Δ to the right.

- (c) Suppose a state-dependent transition rate, so that the probability of moving in either direction depends upon the position of the particle. More precisely, if the particle is at $k\Delta$ the probabilities of moving right or left are,

$$\frac{1}{2} \left(1 - \frac{k}{R}\right) \quad \text{or} \quad \frac{1}{2} \left(1 + \frac{k}{R}\right),$$

respectively. R is a given integer, and possible positions of the particle are limited by the condition $-R \leq k \leq R$. Derive the difference equation for the probability $p_m(s)$ for occupying the site m after s moves (if the initial site was $m = 0$).

Under the same limiting conditions as above, and with $R \rightarrow \infty$ and $1/R\tau \rightarrow \gamma$, derive the partial differential equation for $P(x, t)$. Solve for $P(x, t)$, subject to the boundary conditions $P(x, t) \rightarrow 0$ and $\partial P / \partial x \rightarrow 0$ as $x \rightarrow \pm\infty$.

7. **One-dimensional random walk with hesitation:** Consider a random walk along a discrete lattice with unit steps, with equal probabilities for moving to the right and to the left, given by $p = q = \lambda\tau$ (so that the probability to remain on a site is $r = 1 - 2\lambda\tau$), where τ is the time between consecutive steps and λ is a given constant.

- (a) Write a balance equation for the probability $p_m(s)$ for occupying the site m after s moves (if the initial site was $m = 0$). Take the $\tau \rightarrow 0$ limit in that equation to derive the following continuous-time equation,

$$\frac{dp_m(t)}{dt} = \lambda [p_{m-1}(t) + p_{m+1}(t) - 2p_m(t)],$$

where $t = s\tau$ is a finite observation time. (Note that throughout, m is an integer.)

- (b) Show that,

$$p_m(t) = \frac{1}{2\pi i} \oint z^{-m-1} Q(z, t) dz,$$

where $Q(z, t)$ is the probability-generating function,

$$Q(z, t) = \sum_{m=-\infty}^{\infty} z^m p_m(t), \quad (1.47)$$

with z a complex number, and the integral $\oint$ running along the unit circle $|z| = 1$ in the complex plane.

- (c) Derive a differential equation for the probability-generating function and show that its solution, subject to the appropriate initial condition, is given by $Q(z, t) = \exp[(z + z^{-1} - 2)\lambda t]$.
- (d) Show that the solution of the equation from part 7a is given by,

$$p_m(t) = e^{-2\lambda t} I_{|m|}(2\lambda t),$$

where $I_n(x)$ is a modified-Bessel function of integer order n .

- (e) Deduce that the leading term of the asymptotic series for the probability density function from part 7d, when $t, m \rightarrow \infty$ with m^2/t fixed, has the form,

$$p_m(t) = \frac{1}{\sqrt{4\pi\lambda t}} \exp\left(-\frac{m^2}{4\lambda t}\right).$$

Note: In parts 7d and 7e, you may use any standard handbook of mathematical methods and formulas, such as the book by Abramowitz and Stegun.

8. **Monte Carlo simulation of random walks:** Often various properties of stochastic processes are not amenable to direct computation and simulation becomes the only option. Estimate the solution of the following problems by running a Monte Carlo simulation.

- (a) Consider a one-dimensional unbiased random walk, with a reflecting barrier at $x = 0$ and an absorbing barrier at $x = n$ (any particle that reaches $x = n$ disappears with probability 1). For a system that starts at $x = 0$, on average how many steps must it take before it is absorbed at $x = n$? Run your simulations for various n , and deduce a general formula. This is a simple enough example that the answer can be checked analytically – verify your formula obtained from simulation against the analytic solution.

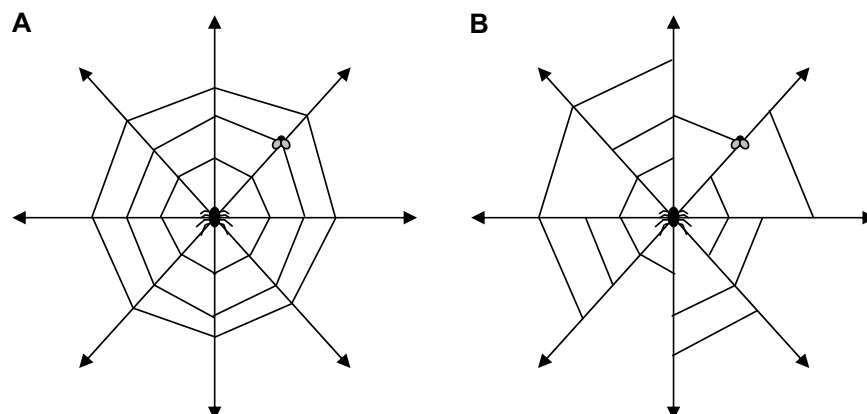


Figure 1.6: From a given vertex, the probability to move along any adjacent strand is equal. If the fly can make it to one of the arrows, then it is free. If the spider catches the fly, then the fly is eaten. **A.** Spider web for Exercise 8(b)i, 8(b)ii and 8(b)iii. **B.** Spider web for Exercise 8(b)iv.

- (b) Consider a two-dimensional random walk along a spider web (Fig. 1.6). A fly lands on the web at some vertex.
- i. Suppose the fly dies of fright, and the spider moves blindly along the web to find it. On average, how many steps must the spider take in order to find the fly? How does that number change as the number of radial vertices increases? (In Fig. 1.6, there are 3 radial vertices.) Can you imagine a mapping of this problem to a one-dimensional random walk? Verify your simulation results with an analytic estimate.
 - ii. Suppose the web is very dusty so that the fly can easily move along every strand (for simplicity, assume the time to traverse along a strand is independent of the strand length). For a spider asleep at the center of the web, what is the probability of escape for a fly landing on any given vertex of the web? Compute the escape probability at every vertex for 3, 4, and 5 radial vertices. *Hint:* Use symmetry arguments to cut down the number of simulations you must run.
 - iii. Now assume the spider can move, but slowly. Repeat 8(b)ii with a spider that moves 1 step for every 2 steps taken by the fly.

- iv. Compute the escape probability for the fly that lands in the tattered web shown in Fig. 1.6B, if the spider is asleep at the center. Suppose the fly dies of fright, and the spider moves blindly along the web to find it. On average, how many steps must the spider take in order to find the fly?
- (c) Suppose that in a coin-toss game with unit stakes, Albert bets on heads, and Paul bets on tails. The probability that Albert leads in $2r$ out of $2n$ tosses is called the *lead probability* $P_{2r,2n}$. For a coin tossed 20 times, what is the most likely number of tosses for which Albert is in the lead? Guess first, then run Monte Carlo simulations to make a table of $P_{2r,2n}$ for $n = 10$ and $r = 0, 1, 2, \dots, 10$. Follow the convention that in the event of a tie, the previous leader retains the lead.

CHAPTER 2

RANDOM PROCESSES

By the time the Ornstein and Uhlenbeck paper appeared, it had become clear that Brownian motion was a phenomenon that could not be handled in a rigorous fashion by classical probability theory. For one thing, although Brownian motion is certainly a *random* affair, the variables that describe it are not the ordinary random variables of classical probability theory, but rather they are *random functions*. Roughly speaking, these are functions which are specified by the results of an observation, and which can take on different values when the observation is repeated many times. For another, although Wiener (1922) had put a special case of Brownian motion on a sound mathematical basis, it was clear that Brownian motion was just one example of this kind of random phenomenon (with many more cases popping in science and engineering), and it was not obvious whether Wiener's methods could be generalized.

A phenomenon similar to Brownian motion is that of fluctuations in electric circuits. Strictly speaking, the current through a conductor is always a random function of time, since the thermal motion of the electrons produces uncontrollable fluctuations (thermal “noise”). This kind of noise was becoming increasingly important in the 1920's in electrical engineering due to the increasing importance of the vacuum tube – recall the pioneering work of van der Pol (1924) in this area. These vacuum tubes are always a source of considerable noise because of fluctuations in the number of electrons passing through the tube during identical time intervals (shot effect) and because of fluctuations in the cathode emission

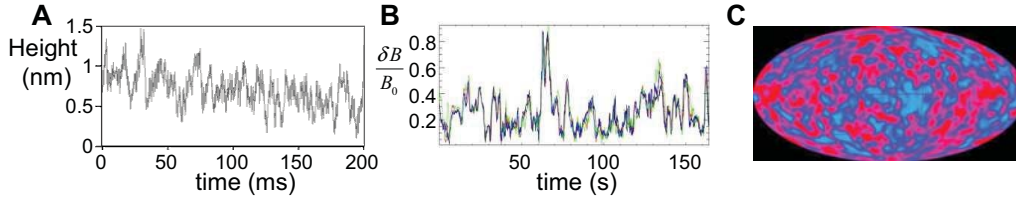


Figure 2.1: **Examples of Random Functions** A) Height of a protein molecule above a catalytic surface. B) Fluctuations in the magnetic field around the North pole. C) Example of a random field – The cosmic background radiation.

(flicker noise). In radio receivers, one not only observes noise arising in the electrical circuits themselves, but also random changes in the level of the received signals due to scattering of electromagnetic waves caused by inhomogeneities in the refractive index of the atmosphere (fading) and the influence of random electrical discharges (meteorological and industrial noise).

Electrical noise, albeit very important, is far from a unique case. As other examples, we may cite the pressure, temperature and velocity vector of a fluid particle in a turbulent flow – in particular, at a point in the Earth’s atmosphere. These random functions depend on four variables (the three space variables and time): they are called random *fields*. The list goes on...

To illustrate the appearance of the observed values of random functions, see Figure 2.1. Curves (and contours) of this type are obtained by experiment, that is, they are the result of observations. They represent *realizations* of the random functions; they are also called *sample functions* or *sample paths*.

Their general appearance shows that a deterministic description of the phenomenon would be so complicated that the resulting mathematical model (if possible at all) would be practically impossible to solve.

2.1 Random processes - Basic concepts

Suppose we are given an experiment specified by its outcomes $\omega \in \Omega$ (Ω = the set of all possible outcomes), and by the probability of occurrence of certain subsets of Ω . For example, in the tossing of a fair die, $\Omega = \{1, 2, 3, 4, 5, 6\}$, the probability of each outcome is independent of the

previous toss and equal to $P = \frac{1}{6}$. A sample experiment might be $\omega = \{5, 4, 6, 4, 3, 2\}$.

To every outcome ω we now assign – according to a certain rule – a function of time $\xi(\omega, t)$, real or complex. We have thus created a family of functions, one for each ω . This family is called a *stochastic process* (or a *random function*). Usually, $t \in \mathbb{R}$, although it could also be that $t \in [0, T]$.

A stochastic process is a function of two variables, ω and t . There are two points of view:

1. **Fix ω** , then $\xi(\omega, t) = \xi^{(\omega)}(t)$ is a (real-say) function of time, depending upon the *parameter* ω . To each outcome ω , there corresponds a function of t . This function is called a *realization*, or *sample function* of the stochastic process.
2. **Fix t** , then $\xi(\omega, t) = \xi_t(\omega)$ is a family of random variables depending upon the parameter t .

In that way, a stochastic process can be regarded as either a family of realizations $\xi^{(\omega)}(t)$, or as a family of random variables $\xi_t(\omega)$. This may seem a pedantic distinction, but notice that Einstein's point of view was to treat Brownian motion as a distribution of a random variable describing position ($\xi_t(\omega)$), while Langevin took the point of view that Newton's law's of motion apply to an individual realization ($\xi^{(\omega)}(t)$).

Remarks:

1. From a purely mathematical point of view, in order to specify a stochastic process we have to provide the probability (or probability density) of occurrence of the various realizations. This leads to the definition of a particular measure P (the probability measure) on the function space of realizations; by specifying this measure, we specify the stochastic process.

This approach, originating in Kolmogorov's work, has been *mathematically very successful and fruitful*; in particular, it can be shown *to include as special cases all other ways of specifying a random process*. Unfortunately, for us, it requires the use of advanced ideas from set and measure theory (see STAT 901/902); since the use of such techniques is not essential for our subsequent considerations, we shall develop the theory from a more physical point of view.

2. We shall demand that, for a fixed t , $\xi_t(\omega)$ be a random variable in the classical sense. Moreover, we will denote the stochastic process by $\xi(t)$, with the dependence on ω understood.

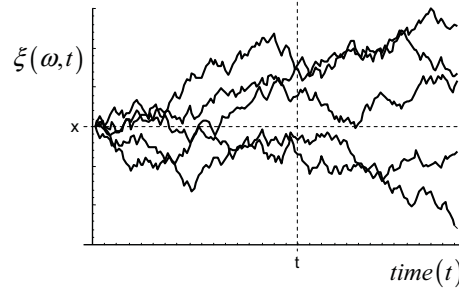


Figure 2.2: **Frequency interpretation of an ensemble of sample paths.**

3. Equality of two stochastic processes will mean their respective functions are identical for any outcome ω_i ; *i.e.*

$$\xi(t) = \zeta(t) \quad \text{means} \quad \xi_t(\omega_i) = \zeta_t(\omega_i).$$

Similarly, we define the operations of addition and multiplication.

Because of the above interpretation, various probability concepts may be immediately generalized to stochastic processes. For example, if $\xi(t)$ is a real stochastic process, then its distribution function is given by,

$$F(x; t) = P\{\xi(t) \leq x\} \quad (2.1)$$

(see Eq A.1 on page 230). The meaning of Eq. 2.1 is: Given $x, t \in \mathbb{R}$, $F(x; t)$ is the probability of the event $\{\xi(t) \leq x\}$ consisting of all outcomes ω such that $\xi_t(\omega) \leq x$. More intuitively, we can adopt the *frequency interpretation* of Eq. 2.1: An experiment is repeated n times and a sample path observed each time (see Figure 2.2). Now given two numbers (x, t) , we count the total number $n_t(x)$ of times that (at a given t), the ordinates of the observed functions do not exceed x ; then $F(x; t) \approx \frac{n_t(x)}{n}$.

As in classical probability theory, the probability density corresponding to the distribution $F(x; t)$ is given by,

$$f(x; t) = \frac{\partial F(x; t)}{\partial x}. \quad (2.2)$$

Next, given two instants t_1 and t_2 , consider the random variables $\xi(t_1)$ and $\xi(t_2)$ and define their *joint probability distribution* by,

$$F(x_1, x_2; t_1, t_2) = P\{\xi(t_1) \leq x_1, \xi(t_2) \leq x_2\} \quad (2.3)$$

with the probability density defined analogously,

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}, \quad (2.4)$$

and so on for higher-order joint distributions. In general,

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{\xi(t_1) \leq x_1, \dots, \xi(t_n) \leq x_n\} \quad (2.5)$$

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^n F(x_1, \dots, x_n; t_1, \dots, t_n)}{\partial x_1 \dots \partial x_n}. \quad (2.6)$$

Note that the n^{th} -order distribution function determines *all* lower-order distribution functions, and in fact, the n^{th} -order distribution function completely determines the stochastic process.

Remarks:

1. The distribution function Eq. 2.5 is not arbitrary, but must satisfy the following,
 - *Symmetry condition.* For every permutation $(j_1, j_2, \dots, j_n)$ of $(1, 2, \dots, n)$, we must have

$$F(x_{j_1}, x_{j_2}, \dots, x_{j_n}; t_{j_1}, t_{j_2}, \dots, t_{j_n}) = F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n).$$

This condition follows from the AND logic of the distribution function: the statements ‘A AND B AND C’ and ‘A AND C AND B’ are equivalent.

- *Compatibility condition.* For $m < n$, we must have

$$F(x_1, \dots, x_m, \infty, \dots, \infty; t_1, \dots, t_m, t_{m+1}, \dots, t_n) = F(x_1, \dots, x_m; t_1, \dots, t_m).$$

This statement again follows from the AND logic: the statements ‘A AND B AND C’ and ‘A AND TRUE AND C’ are equivalent.

2. Kolmogorov (1931) proved that given any set of density functions satisfying the two conditions above, there exists a probability space Ω and a stochastic process $\xi(\omega, t)$ such that $F(x_1, \dots, x_n; t_1, \dots, t_n)$ gives the joint probability distribution of $\xi_{t_1}(\omega), \dots, \xi_{t_n}(\omega)$.

As a further extension, we introduce the conditional probability density,

$$f(x_2, t_2 | x_1, t_1) = \frac{f(x_1, x_2; t_1, t_2)}{f(x_1, t_1)}, \quad (2.7)$$

$$f(x_{k+1}, \dots, x_n; t_{k+1}, \dots, t_n | x_1, \dots, x_k; t_1, \dots, t_k) = \frac{f(x_1, \dots, x_n; t_1, \dots, t_n)}{f(x_1, \dots, x_k; t_1, \dots, t_k)}.$$

In particular, we have,

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = f(x_n, t_n | x_{n-1}, \dots, x_1; t_{n-1}, \dots, t_1) \times \dots \times f(x_2, t_2 | x_1, t_1) \times f(x_1, t_1).$$

Intuitively, the relationship between the conditional and joint probabilities is very straightforward. Focusing upon the two variable case,

$$f(x_1, x_2; t_1, t_2) = f(x_2, t_2 | x_1, t_1) \times f(x_1, t_1),$$

simply means the probability that the state is x_2 at time t_2 and x_1 at t_1 is equal to the probability that the state is x_2 at time t_2 given that it was x_1 at t_1 multiplied by the probability that the state actually was x_1 at t_1 .

2.1.1 Moments of a Stochastic Process

We define the mean of the stochastic process by,

$$\mu(t) = \langle \xi(t) \rangle = \int_{-\infty}^{\infty} x f(x; t) dx; \quad (2.8)$$

the correlation function by,

$$B(t_1, t_2) = \langle \xi(t_1) \xi(t_2) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2. \quad (2.9)$$

It will be useful in coming chapters to re-write the correlation function using the definition from the previous section – re-writing the joint distribution using the conditional distribution,

$$B(t_1, t_2) = \langle \xi(t_1) \xi(t_2) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 [f(x_2, t_2 | x_1, t_1) \times f(x_1, t_1)] dx_1 dx_2$$

$$B(t_1, t_2) = \langle \xi(t_1) \xi(t_2) \rangle = \int_{-\infty}^{\infty} x_1 \langle x_2 \rangle_{x_1} f(x_1; t_1) dx_1, \quad (2.10)$$

where $\langle x_2 \rangle_{x_1}$ is the *conditional average* of x_2 ,

$$\langle x_2 \rangle_{x_1} = \int_{-\infty}^{\infty} x_2 f(x_2, t_2 | x_1, t_1) dx_2,$$

(conditioned so that $\langle x_2(t_1) \rangle = x_1$).

Related to the correlation is the covariance, given by

$$C(t_1, t_2) = \langle \{ \xi(t_1) - \mu(t_1) \} \cdot \{ \xi(t_2) - \mu(t_2) \} \rangle \equiv \langle \langle \xi(t_1) \xi(t_2) \rangle \rangle; \quad (2.11)$$

and so on for higher-order correlations. The double-angled brackets denote the *cumulants* of the process: Recall that for a scalar random variable ξ , the *cumulants* κ_m of ξ are defined via a generating function,

$$\langle e^{-i\xi t} \rangle = \exp \left[\sum_{m=1}^{\infty} \frac{(-it)^m}{m!} \kappa_m \right],$$

(see Eq. A.9 on page 236). Similarly, the cumulants of a *stochastic function* $\xi(t)$, denoted by $\kappa_m \equiv \langle \langle \xi^m \rangle \rangle$, are given by

$$\begin{aligned} & \langle e^{-i \int_0^t \xi(t') dt'} \rangle = \\ & \exp \left[\sum_{m=1}^{\infty} \frac{(-i)^m}{m!} \int_0^t \cdots \int_0^t \langle \langle \xi(t_1) \xi(t_2) \cdots \xi(t_m) \rangle \rangle dt_1 dt_2 \cdots dt_m \right]. \end{aligned} \quad (2.12)$$

The general rule for relating cumulants to the *moments* of $\xi(t)$ is to partition the digits of the moments into all possible subsets; for each partition, one writes the product of cumulants, then adds all such products. The first few relations will make the prescription clear – Writing 1, 2, ... for $\xi(t_1), \xi(t_2), \dots$,

$$\begin{aligned} \langle 1 \rangle &= \langle \langle 1 \rangle \rangle, \\ \langle 1 \ 2 \rangle &= \langle \langle 1 \rangle \rangle \langle \langle 2 \rangle \rangle + \langle \langle 1 \ 2 \rangle \rangle, \\ \langle 1 \ 2 \ 3 \rangle &= \\ & \langle \langle 1 \rangle \rangle \langle \langle 2 \rangle \rangle \langle \langle 3 \rangle \rangle + \langle \langle 1 \rangle \rangle \langle \langle 2 \ 3 \rangle \rangle + \langle \langle 2 \rangle \rangle \langle \langle 1 \ 3 \rangle \rangle + \langle \langle 3 \rangle \rangle \langle \langle 1 \ 2 \rangle \rangle + \langle \langle 1 \ 2 \ 3 \rangle \rangle, \dots \end{aligned}$$

2.2 Stationarity and the Ergodic Theorem

For systems characterized by linear equations, there is a very convenient relationship between the probability distribution of the fluctuations in *equilibrium* and the parameters in the model governing the dynamics

called the ‘fluctuations-dissipation relation’. The fluctuation-dissipation relation underlies many results in stochastic processes (including Einstein and Langevin’s treatment of Brownian motion) and its derivation will be the focus of this section. As a prerequisite, we introduce a very useful notion – stationarity.

Definition: The stochastic process $\xi(t)$ is called *stationary* if all finite dimensional distribution functions defining $\xi(t)$ remain unchanged when the whole group of points is shifted along the time axis; *i.e.*, if

$$F(x_1, x_2, \dots, x_n; t_1 + \tau, t_2 + \tau, \dots, t_n + \tau) = F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n), \quad (2.13)$$

for any $n, t_1, t_2, \dots, t_n$ and τ . In particular, all one-dimensional distribution functions must be identical (*i.e.* $F(x, t) = F(x)$ cannot depend upon t); all 2-dimensional distribution functions can only depend upon $t_1 - t_2$; and so on. It follows that the autocorrelation B depends upon time difference only,

$$\langle \xi(t)\xi(s) \rangle = B(t - s) = B(\tau).$$

It turns out that there exists a whole class of phenomena where the underlying stochastic process is completely characterized by the mean $\langle \xi(t) \rangle = \mu = \text{constant}$, and by the correlation function $B(\tau)$. Such stochastic processes are said to be *stationary in the wide sense* (or in *Khinchin’s sense*). It should be clear that, in general, Eq. 2.13 above may not hold for stochastic processes which are only stationary in the wide sense.

A very useful property of stationary stochastic processes is the fact that they obey the so-called *Ergodic Theorem*. Consider a stochastic process and suppose we know how to compute mathematical expectations; the question arises as to how to compare mathematical averages with those obtained by experiment.

Experimentally, we may proceed as follows:

1. Take a large number N of records (realizations) of $\xi(t)$, say,

$$\xi^{(1)}(t), \xi^{(2)}(t), \dots, \xi^{(j)}(t), \dots, \xi^{(N)}(t);$$

2. Then we take the arithmetic mean, with respect to j , of $\xi^{(j)}(t)$ for

every t and claim this equals $\langle \xi(t) \rangle$,

$$\langle \xi(t) \rangle \approx \frac{1}{N} \sum_{j=1}^N \xi^{(j)}(t).$$

3. Next, we take the arithmetic mean of the products $\xi^{(j)}(t)\xi^{(j)}(s)$ for every pair (t, s) , and claim that this equals $\langle \xi(t)\xi(s) \rangle = B(t, s)$;
4. Proceed analogously with higher-order moments ...

This program is fine in principal, but in practice taking all of these records and processing them as above is a very complicated affair – impossible to carry out in many cases. In fact, it is far easier to take a few *long* records of the phenomenon; in such a case, we *cannot* compute the arithmetic mean, but rather we can compute the time averages of the quantities of interest, *e.g.*,

$$\begin{aligned} \langle \xi(t) \rangle &\approx \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) dt, \\ \langle \xi(t) \xi(t + \tau) \rangle &\approx \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) \xi(t + \tau) dt, \end{aligned}$$

and so on for higher-order correlations. Here, T is a finite, but *very large* time. In general, this time-average will not equal the mathematical expectation; however, if the process is *stationary*, the ergodic theorem guarantees that they coincide.

If $\xi(t)$ is a *stationary* stochastic process (with t a continuous variable) satisfying certain general conditions, then

$$\begin{aligned} \langle \xi(t) \rangle &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) dt, \\ \langle \xi(t) \xi(t + \tau) \rangle &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) \xi(t + \tau) dt, \text{ etc.} \end{aligned}$$

A nice pictorial way of thinking about the content of the ergodic theorem is the following. Say we have a single, very long record $\xi^*(t)$ of

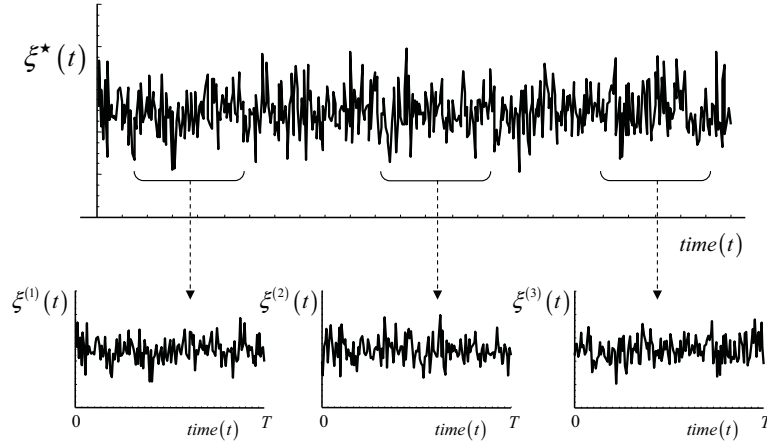


Figure 2.3: A long sample of a stationary process can be used in place of many (shorter) repeated samples.

our stochastic process $\xi(t)$ – something like Figure 2.3. Now suppose this *single* record is cut into a *large number* N of pieces of length T ; since the process is stationary, each $\xi^{(j)}(t)$ may be regarded as a separate measurement, and the arithmetic average can be performed,

$$\langle \xi(t) \rangle \approx \frac{1}{N} \sum_{j=1}^N \xi^{(j)}(t),$$

$$\langle \xi(t) \xi(t + \tau) \rangle = B(\tau) \approx \frac{1}{N} \sum_{j=1}^N \xi^{(j)}(t) \xi^{(j)}(t + \tau),$$

and so on. The ergodic theorem says that in the limit $N \rightarrow \infty$, the time-averages and the arithmetic averages coincide.

In practice, since $\xi(t)$ cannot (in general) be expressed by a closed formula, the calculation of the averages is performed as follows. The record $\xi^*(t)$ is digitized and the time integrals replaced by their approximating sums,

$$\int_0^T \xi(t) dt \approx \frac{T}{N} \sum_{k=1}^N \xi\left(k \frac{T}{N}\right), \quad N \rightarrow \infty,$$

so that,

$$\langle \xi(t) \rangle \approx \frac{1}{N} \sum_{k=1}^N \xi^*(k\Delta),$$

where Δ is a small time interval, and N is chosen to make $N\Delta = T$ sufficiently large. Usually, Δ is chosen in such a way that the function $\xi^*(t)$ does not change appreciably during Δ , while the number N should be so large that a further increase has a negligible effect on the value computed via Eq. 2.2.

2.3 Spectral Density and Correlation Functions

In practice, it is often much easier to measure the Fourier transform of a signal rather than trying to compute the autocorrelation function directly. The two representations are closely related, as we shall now show. First, we define,

$$\hat{X}(\omega) = \int_0^T \xi(t) e^{-i\omega t} dt.$$

Then, the *fluctuation spectrum*, or *spectral density of the fluctuations*, is defined as,

$$S(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2\pi T} |\hat{X}(\omega)|^2.$$

After some manipulation, the connection between the (time-averaged) autocorrelation function and the fluctuation spectrum can be made explicit,

$$S(\omega) = \lim_{T \rightarrow \infty} \left[\frac{1}{\pi} \int_0^T \cos(\omega\tau) \frac{1}{T} \int_0^{T-\tau} \xi(t) \xi(t+\tau) dt d\tau \right],$$

or,

$$S(\omega) = \frac{1}{\pi} \int_0^\infty \cos(\omega\tau) B(\tau) d\tau.$$

Written another way,

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty e^{-i\omega\tau} B(\tau) d\tau \iff B(\tau) = \int_{-\infty}^\infty e^{i\omega\tau} S(\omega) d\omega. \quad (2.14)$$

The conjugal relationship between the autocorrelation function $B(\tau)$ and the fluctuation spectrum $S(\omega)$ is called the *Wiener-Khinchin theorem* (see Fig. 2.4). Furthermore, for a *stationary* process $\xi(t)$, the time-average may be replaced by the ensemble average via the ergodic theorem, and we write,

$$\langle \xi(t)\xi(t+\tau) \rangle = B(\tau).$$

Using the Fourier transform of $\xi(t)$,

$$\hat{\Xi}(\omega) = \int_{-\infty}^{\infty} \xi(t)e^{-i\omega t} dt,$$

assuming $\langle \xi(t) \rangle = 0$, we have the following,

$$\langle \Xi(\omega) \rangle = \frac{1}{2\pi} \int \langle \xi(t) \rangle e^{-i\omega t} dt = 0,$$

$$\langle \hat{\Xi}(\omega)\hat{\Xi}(\omega') \rangle = \delta(\omega - \omega')S(\omega). \quad (2.15)$$

In the next chapter, we use the fluctuation spectrum $S(\omega)$ to derive the very useful *fluctuation-dissipation relation*.

2.3.1 Linear Systems

For linear systems with stochastic forcing $\eta(t)$,

$$\left[a_n \frac{d^n}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + \dots + a_0 \right] y(t) = \eta(t), \quad (2.16)$$

the correlation function of the process $y(t)$ is related in a very simple way to the correlation of the forcing $\eta(t)$. To show this relationship explicitly, we begin with the function $H(\omega)$ which is called the *transfer function* of the linear operator acting on $y(t)$ and is given by the Fourier transform of the solution $h(t)$ of the auxiliary equation,

$$\left[a_n \frac{d^n}{dt^n} + a_{n-1} \frac{d^{n-1}}{dt^{n-1}} + \dots + a_0 \right] h(t) = \delta(t),$$

(where $\delta(t)$ is the Dirac-delta function and $h(t)$ is often called the *impulse response* of the system). Writing $H(\omega)$ explicitly,

$$H(\omega) = [a_n (i\omega)^n + a_{n-1} (i\omega)^{n-1} + \dots + a_0]^{-1}.$$

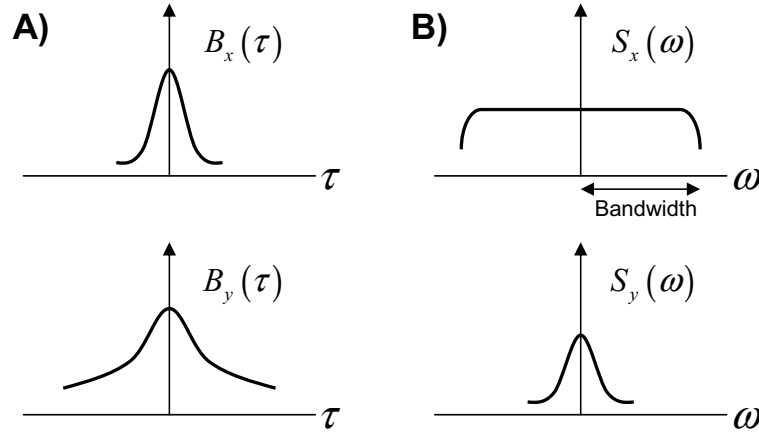


Figure 2.4: **Correlation functions and spectra** A) Example correlation functions. B) The corresponding fluctuation spectra. As the correlation time decreases, the autocorrelation becomes more narrowly peaked, while the spectrum broadens – this is a general feature of correlation functions and spectra, implied by the Wiener-Khinchin theorem.

The utility of the impulse response $h(t)$ is that the particular solution of an inhomogeneous differential equation can always be written as a convolution of the forcing function with $h(t)$. Eq. 2.16, for example, has the formal solution,

$$y(t) = \int_{-\infty}^{\infty} \eta(t - \tau') h(\tau') d\tau'.$$

If we multiply $\eta(t + \tau)$ by $y^*(t)$ (where the $*$ denotes the complex conjugate) and take the average, then

$$\langle \eta(t + \tau) y^*(t) \rangle = \int_{-\infty}^{\infty} \langle \eta(t + \tau) \eta^*(t - \tau') \rangle h^*(\tau') d\tau'.$$

After a change of variables, and a Fourier transform,

$$S_{\eta y}(\omega) = S_{\eta\eta}(\omega) H^*(\omega). \quad (2.17)$$

In a similar way, we also have,

$$S_{yy}(\omega) = H(\omega) S_{\eta\eta}(\omega). \quad (2.18)$$

Combining Eqs. 2.17 and 2.18, yields the fundamental relation,

$$\boxed{S_{yy}(\omega) = S_{\eta\eta}(\omega)H(\omega)H^*(\omega) = S_{\eta\eta}(\omega) \times |H(\omega)|^2}. \quad (2.19)$$

As an example, consider a linear system with damping driven by a stochastic function with zero mean and a delta-correlation function, $\langle \eta(t)\eta(t') \rangle = \Gamma\delta(t - t')$,

$$\frac{dy}{dt} = -\beta y + \eta(t), \quad (2.20)$$

the impulse response $H(\omega)$ is simply $H(\omega) = 1/(i\omega + \beta)$, so that

$$S_{yy}(\omega) = \frac{S_{\eta\eta}(\omega)}{\omega^2 + \beta^2} = \frac{1}{2\pi} \frac{\Gamma}{\omega^2 + \beta^2}. \quad (2.21)$$

Taking the inverse Fourier transform,

$$\langle y(t)y(t - \tau) \rangle = \Gamma \frac{e^{-\beta|\tau|}}{2\beta}. \quad (2.22)$$

Eq. 2.20 is Langevin's equation for the velocity of a Brownian particle, so (2.22) is the autocorrelation function for the Ornstein-Uhlenbeck process at steady-state.

2.3.2 Fluctuation-Dissipation Relation

For linear systems, we can obtain the relation between the correlation function of the process $y(t)$ and the correlation function of the forcing function $\eta(t)$ by directly calculating the autocorrelation of $y(t)$. If this is combined with the fundamental result above, we obtain the so-called *fluctuation-dissipation relation*, relating the noise magnitude Γ to the other physical parameters in the model. We can write a generic version of Langevin's equation as the following phenomenological damping law governing the variable y , subject to random forcing $\eta(t)$:

$$\frac{dy}{dt} = -\beta y + \eta(t), \quad (2.23)$$

where β is the damping constant and the random forcing $\eta(t)$ has the statistical properties,

$$\langle \eta(t) \rangle = 0, \quad (2.24)$$

and,

$$\langle \eta(t)\eta(t') \rangle = \Gamma \delta(t - t'). \quad (2.25)$$

(Compare with Eq. 1.29 on page 17.) Noise with a Dirac-delta correlation function is called *white noise* because one can easily show that the fluctuation spectrum of the Dirac-delta is a constant; *i.e.* all frequencies are equally represented in the spectrum, in analogy with *white light* that contains all frequencies of visible light (see Eq. B.34 on p. 263). For the linear system (2.23), it is possible to compute the fluctuation spectrum of y *without* using Eq. 2.25! To that end, note that the general solution of Eq. 2.23 is

$$y(t) = y_0 e^{-\beta t} + \int_0^t e^{-\beta(t-t')} \eta(t') dt'. \quad (2.26)$$

Taking the average, conditioned by the initial value $y(0) = y_0$,

$$\langle y(t) \rangle_{y_0} = y_0 e^{-\beta t}, \quad (2.27)$$

since the average of $\eta(t)$ vanishes by Eq. 2.24. Calculating the autocorrelation function,

$$\langle y(0) \langle y(t) \rangle_{y_0} \rangle^{eq} = \langle y^2 \rangle^{eq} e^{-\beta t}, \quad (2.28)$$

where $\langle y^2 \rangle^{eq}$ is given by the variance of the equilibrium distribution, as calculated by equilibrium statistical mechanics (equipartition of energy). The spectral density is simply,

$$S_{yy}(\omega) = \frac{1}{\pi} \langle y^2 \rangle^{eq} \frac{\beta}{\beta^2 + \omega^2}. \quad (2.29)$$

Notice that up to this point, Eq. 2.25 has not been used. From the fundamental relation above (Eq. 2.19),

$$(\beta^2 + \omega^2) S_{yy}(\omega) = S_{\eta\eta}(\omega). \quad (2.30)$$

Now using Eq. 2.25 and Eq. 2.30, we arrive at,

$$\Gamma = 2 \langle y^2 \rangle^{eq} \beta, \quad (2.31)$$

relating magnitude of the fluctuations (Γ) to the magnitude of the phenomenological dissipation (β). This is a statement of the *fluctuation-dissipation relation*. It says that in order to achieve thermal equilibrium,

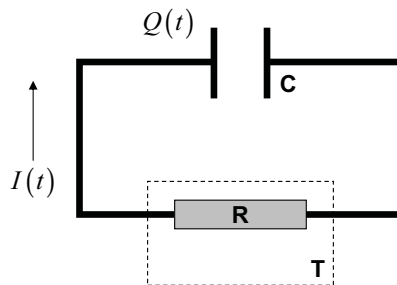


Figure 2.5: **Johnson-Nyquist circuit.** Capacitance C shorted through a resistance R at temperature T . The current $I(t)$ is dQ/dt . Redrawn from D. Lemons “An introduction to stochastic processes in physics,” (John Hopkins University Press, 2002).

the diffusive effects of the fluctuations must be balanced by the dissipative effects of the drag. A version of this relation was used by Einstein in his study of Brownian motion to relate the diffusion coefficient D to the drag force experienced by the particle, thereby obtaining an explicit relation between the mean-squared displacement and Boltzmann’s constant (*cf.* Eq. 1.13 on page 7). The fluctuation-dissipation relation was later used by Nyquist and Johnson in the study of thermal noise in a resistor. That work is described in more detail below and provides another example of using the fluctuations to quantify phenomenological parameters (in this case, again, Boltzmann’s constant).

2.3.3 Johnson and Nyquist

- J. B. Johnson (1928) “Thermal agitation of electricity in conductors,” *Physical Review* **32**: 97–109.
- H. Nyquist (1928) “Thermal agitation of electric charge in conductors,” *Physical Review* **32**: 110–113.

In 1928, J. B. Johnson observed random fluctuations of current through resistors of various materials. Most importantly, he observed that the power in the fluctuations scaled linearly with temperature. In an article immediately following, Nyquist provided a theoretical framework for Johnson’s observations. His main tools were, like Langevin, the equipartition of energy theorem from equilibrium statistical mechanics and the fluctuation-dissipation theorem. The consequence of Nyquist’s analysis

was that the linear dependence of the fluctuation power on the ambient temperature yields Boltzmann's constant (which is proportional to the slope of the line). We shall consider their work in more detail below, starting with the work of Nyquist and applying the fluctuation-dissipation relation to a simple RC-circuit (Figure 2.5).

Nyquist uses essentially a completely verbal argument based on equilibrium thermodynamics to arrive at the following conclusions:

- To a very good approximation, the fluctuations should be uniformly distributed among all frequencies (*i.e.*, the noise is *white*).
- At equilibrium, the average energy in each degree of freedom will be $k_B T$, where k_B is Boltzmann's constant and T is the absolute temperature of the system. Of that energy, one half is magnetic and the other half is electric.

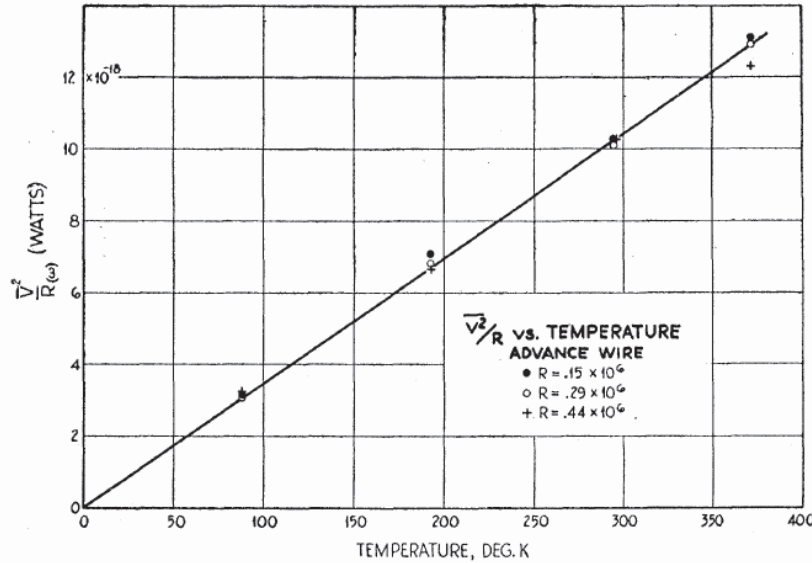


Figure 2.6: **Johnson's measurement of $\langle v^2 \rangle / R$ as a function of T .** From the fluctuation-dissipation relation (Eq. 2.34), the slope is proportional to Boltzmann's constant. Copied from Figure 6 of Johnson's paper.

We consider a specific circuit – A charged capacitor C discharging through a resistance R , all at a fixed temperature T (Figure 2.5). In the

absence of fluctuations, Kirchoff's law gives,

$$IR + \frac{Q}{C} = 0,$$

where I is the current in the circuit and Q is the charge on the capacitor. Johnson observed variability in the current, represented by a white noise source $\eta(t)$. Writing $I = dQ/dt$ leads to the Langevin equation

$$\frac{dQ}{dt} = -\frac{1}{RC} Q + \eta(t). \quad (2.32)$$

This equation has the canonical form discussed in the previous section, with damping coefficient $\beta = 1/RC$. From the equipartition of energy, we know that the mean electrical energy stored in the capacitor is,

$$\frac{\langle Q^2 \rangle}{2C} = \frac{k_B T}{2},$$

or, $\langle Q^2 \rangle = Ck_B T$. From the fluctuation-dissipation relation (Eq. 2.31), we immediately have the variance of the fluctuations Γ ,

$$\langle \eta^2 \rangle = \Gamma = 2\langle Q^2 \rangle \cdot \beta = \frac{2k_B T}{R}. \quad (2.33)$$

This result is more commonly written as a fluctuating *voltage* ν applied to the circuit. By Ohm's law, we have, $V = IR$, so that,

$$\langle \nu^2 \rangle = \langle \eta^2 \rangle \cdot R^2 = 2Rk_B T, \quad (2.34)$$

called *Nyquist's theorem*. Notice $\langle \nu^2 \rangle / R$ should be *linear* in T – that is precisely what Johnson observed (Figure 2.6). Using his setup, Johnson obtained an averaged estimate for Boltzmann's constant of $1.27 \pm 0.17 \times 10^{-23}$ J/K, as compared to the currently accepted value of 1.38×10^{-23} J/K.

Suggested References

Two superb books dealing with correlation functions and spectra from an applied perspective are,

- *Probability, random variables, and stochastic processes (2nd Ed.)*, A. Papoulis (McGraw-Hill, 1984).

- *Introduction to random processes with applications to signals and systems (2nd Ed.)*, W. A. Gardner (McGraw-Hill, 1990).

The second book (by Gardner) is particularly recommended for electrical engineering students or those wishing to use stochastic processes to model signal propagation and electrical networks.

Exercises

1. **White noise:** It is often useful, in cases where the time-scale of the fluctuations is much shorter than any characteristic time scale in the system of interest, to represent the fluctuations as *white noise*.
 - (a) Show that for δ -correlated fluctuations, $B(\tau) = \delta(\tau)$, the fluctuation spectrum is constant. Such a process is called *white-noise* since the spectrum contains all frequencies (in analogy with *white light*).
 - (b) Assume ξ has units such that $\langle \xi^2 \rangle \propto \text{Energy}$, then,

$$\int_{\omega_1}^{\omega_2} S(\omega) d\omega = E(\omega_1, \omega_2),$$

is the energy lying between frequencies ω_1 and ω_2 . For a white-noise process, show that the energy content of the fluctuations is *infinite*! That is to say, *no physical process can be exactly described by white noise*.

2. **Linear systems:** Systems such as Langevin's model of Brownian motion, with linear dynamic equations, filter fluctuations in a particularly simple fashion.
 - (a) Fill in the details to derive Eqs. 2.17 and 2.18.
 - (b) For systems characterized by several variables, the correlation function is generalized to the *correlation matrix*, defined either by the dot product: $\langle \mathbf{y}(t) \cdot \mathbf{y}(t')^T \rangle = \mathbf{C}(t, t')$ (where T is the *transpose*), or element-wise: $\langle y_i(t) y_j(t') \rangle = C_{ij}(t, t')$. For a stationary process, $C_{ij}(t, t') = C_{ij}(t - t')$, use the same line of reasoning as above to generalize Eq. 2.19 to higher-dimensions, and show that,

$$\boxed{\mathbf{S}_{yy}(\omega) = \mathbf{H}(\omega) \cdot \mathbf{S}_{\eta\eta}(\omega) \cdot \mathbf{H}^T(-\omega)}, \quad (2.35)$$

where $\mathbf{S}_{yy}(\omega)$ is the matrix of the spectra of $C_{ij}(t - t')$ and $\mathbf{H}$ is the matrix Fourier transform.

3. **MacDonald's theorem:** Let $Y(t)$ be a fluctuating current at equilibrium (*i.e.* $Y(t)$ is stationary). It is often easier to measure the transported charge $Z(t) = \int_0^t Y(t')dt'$. Show that the spectral density of Y , $S_{YY}(\omega)$, is related to the transported charge fluctuations by MacDonald's theorem,

$$S_{YY}(\omega) = \frac{\omega}{\pi} \int_0^\infty \sin \omega t \frac{d}{dt} \langle Z^2(t) \rangle dt.$$

Hint: First show $(d/dt)\langle Z^2(t) \rangle = 2 \int_0^t \langle Y(0)Y(t') \rangle dt'$.

4. Let the stochastic process $X(t)$ be defined by $X(t) = A \cos(\omega t) + B \sin(\omega t)$, where ω is constant and A and B are random variables. Show that,
- (a) If $\langle A \rangle = \langle B \rangle = \langle AB \rangle = 0$, and $\langle A^2 \rangle = \langle B^2 \rangle$, then $X(t)$ is stationary in the wide-sense.
 - (b) If the joint-probability density function for A and B has the form $f_{AB}(a, b) = h(\sqrt{a^2 + b^2})$, then $X(t)$ is stationary in the strict sense.
5. **Integrated Ornstein-Uhlenbeck process:** Define $Z(t) = \int_0^t Y(t')dt'$ ($t \geq 0$), where $Y(t)$ is an Ornstein-Uhlenbeck process with,

$$\langle Y(t) \rangle = 0 \text{ and } \langle Y(t)Y(t - \tau) \rangle = \Gamma \frac{e^{-\beta|\tau|}}{2\beta}.$$

$Z(t)$ is Gaussian, but neither stationary or Markovian.

- (a) Find $\langle Z(t_1)Z(t_2) \rangle$.
 - (b) Calculate $\langle \cos[Z(t_1) - Z(t_2)] \rangle$. It may be helpful to consider the cumulant expansion of the exponential of a random process, Eq. 2.12 on page 35.
6. **Reciprocal spreading and the Wiener-Khinchin theorem:** Suppose $g(t)$ is any non-periodic, real-valued function with Fourier transform $G(\omega)$ and finite energy:

$$W = \int_{-\infty}^{\infty} g^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 d\omega < \infty.$$

For the following, it may be useful to use the Cauchy-Schwarz inequality (p. 266) and consider the integral,

$$\int_{-\infty}^{\infty} t g(t) \frac{dg}{dt} dt$$

(a) Writing the ‘uncertainty’ in time as σ_t ,

$$\sigma_t = \sqrt{\int_{-\infty}^{\infty} t^2 \frac{g^2(t)}{W} dt},$$

and the ‘uncertainty’ in frequency as σ_ω ,

$$\sigma_\omega = \sqrt{\int_{-\infty}^{\infty} \omega^2 \frac{|G(\omega)|^2}{2\pi W} d\omega},$$

show that

$$\sigma_t \cdot \sigma_\omega \geq \frac{1}{2}.$$

(b) For what function $g(t)$ does $\sigma_t \cdot \sigma_\omega = \frac{1}{2}$?

CHAPTER 3

MARKOV PROCESSES

3.1 Classification of Stochastic Processes

Because the n^{th} -order distribution function $F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$, or density $f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$, completely determines a stochastic process, this leads to a natural classification system. Assuming the time ordering $t_1 < t_2 < \dots < t_n$, we then identify several examples of stochastic processes:

1. **Purely Random Process:** Successive values of $\xi(t)$ are statistically *independent*, *i.e.*

$$f(x_1, \dots, x_n; t_1, \dots, t_n) = f(x_1, t_1) \cdot f(x_2, t_2) \cdots f(x_n, t_n);$$

in other words, all the information about the process is contained in the 1st-order density. Obviously, if the n^{th} -order density factorizes, then all lower-order densities do as well, *e.g.* from

$$f(x_1, x_2, x_3; t_1, t_2, t_3) = f(x_1, t_1) \cdot f(x_2, t_2) \cdot f(x_3, t_3),$$

it follows that,

$$f(x_1, x_2; t_1, t_2) = f(x_1, t_1) \cdot f(x_2, t_2).$$

However, the converse is not true.

2. **Markov Process:** Defined by the fact that the *conditional probability density* enjoys the property,

$$f(x_n, t_n | x_1, \dots, x_{n-1}; t_1, \dots, t_{n-1}) = f(x_n, t_n | x_{n-1}; t_{n-1}). \quad (3.1)$$

That is, the conditional probability density at t_n , given the value at x_{n-1} at t_{n-1} , is not affected by the values at earlier times. In this sense, the process is “without memory.”

A Markov process is fully determined by the two functions $f(x_1, t_1)$ and $f(x_2, t_2 | x_1, t_1)$; the whole hierarchy can be reconstructed from them. For example, with $t_1 < t_2 < t_3$,

$$\begin{aligned} f(x_1, x_2, x_3; t_1, t_2, t_3) = \\ f(x_3, t_3 | x_1, x_2; t_1, t_2) \cdot f(x_2, t_2 | x_1, t_1) \cdot f(x_1, t_1). \end{aligned}$$

But,

$$f(x_3, t_3 | x_1, x_2; t_1, t_2) = f(x_3, t_3 | x_2, t_2),$$

by the Markov property, so,

$$f(x_1, x_2, x_3; t_1, t_2, t_3) = f(x_3, t_3 | x_2, t_2) \cdot f(x_2, t_2 | x_1, t_1) \cdot f(x_1, t_1). \quad (3.2)$$

The algorithm can be continued. This property makes Markov processes manageable, and in many applications (for example Einstein’s study of Brownian motion), this property is approximately satisfied by the process over the coarse-grained time scale (see p. 59).

Notice, too, the analogy with ordinary differential equations. Here, we have a *process* $f(x_1, x_2, x_3, \dots; t_1, t_2, t_3, \dots)$ with a *propagator* $f(x_{i+1}, t_{i+1} | x_i, t_i)$ carrying the system forward in time, beginning with the initial distribution $f(x_1, t_1)$. This viewpoint is developed in detail in the book by D. T. Gillespie (1992) *Markov Processes*.

3. Progressively more complicated processes may be defined in a similar way, although usually very little is known about them. In some cases, however, it is possible to add a set of auxiliary variables to generate an augmented system that obeys the Markov property. See N. G. van Kampen (1998) Remarks on non-Markov processes. *Brazilian Journal of Physics* **28**:90–96.

WARNING: In the physical literature, the adjective *Markovian* is used with regrettable looseness. The term seems to have a magical appeal, which invites its use in an intuitive sense not covered by the definition. In particular:

- When a physicist talks about a “process,” a certain phenomenon involving time is usually what is being referred to. It is meaningless to say a “phenomenon” is Markovian (or not) unless one specifies the variables (especially the time scale) to be used for its description.
- Eq. 3.1 is a condition on *all* the probability densities; one simply cannot say that a process is Markovian if only information about the first few of them is available. On the other hand, if one knows the process *is* Markovian, then of course $f(x_1, t_1)$ and $f(x_2, t_2 | x_1, t_1)$ *do* suffice to specify the entire process.

3.2 Chapman-Kolmogorov Equation

First some remarks about terminology before deriving this fundamental result.

1. We shall call those variables appearing on the left (right) of the conditional line in a conditional probability distribution the *left variables* (*right variables*). For example,

$$f \left(\underbrace{x_1, x_2, x_3, x_4, x_5}_{\text{Left Variables}} \mid \underbrace{x_6, x_7, x_8, x_9, x_{10}}_{\text{Right Variables}} \right)$$

Sometimes one wants to remove a left (or right) variable; this can be done according to the following rules,

- To remove a number of left variables, simply integrate with respect to them,

$$f(x_1 | x_3) = \int_{-\infty}^{\infty} f(x_1, x_2 | x_3) dx_2.$$

- To remove a number of right variables, multiply by their conditional density with respect to the remaining right variables and integrate,

$$f(x_1 | x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1 | x_2, x_3, x_4) f(x_2, x_3 | x_4) dx_2 dx_3.$$

2. From now on, the ensemble average (or $\mathbf{E}\{\cdot\}$) will be denoted by $\langle \cdots \rangle$,

$$\langle x(t) \rangle = \int_{-\infty}^{\infty} x f(x, t) dx.$$

3. Recall the definition of the distribution function of n random variables $\xi_1, \xi_2, \dots, \xi_n$:

$$F(x_1, \dots, x_n) = P\{\xi_1 \leq x_1, \dots, \xi_n \leq x_n\}.$$

This, of course, is the *joint* distribution function, and the joint density function is obtained by differentiation with respect to $\{x_1, \dots, x_n\}$. Obviously,

$$\begin{aligned} f(x_1, \dots, x_n) &\geq 0, \\ F(\infty, \dots, \infty) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1. \end{aligned}$$

If we substitute in the distribution function certain variables by ∞ , we obtain the joint distribution function of the remaining variables; *e.g.*

$$F(x_1, x_3) = F(x_1, \infty, x_3, \infty).$$

If we integrate the joint density function $f(x_1, \dots, x_n)$ with respect to certain variables, we obtain the joint density of the remaining variables (called the *marginal density*); *e.g.*

$$f(x_1, x_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_2 dx_4.$$

We now proceed with the derivation of the Chapman-Kolmogorov equation. As we have seen, a Markov process is fully determined by the two functions $f(x_1, t_1)$ and $f(x_2, t_2 | x_1, t_1)$. This, however, *does not* mean that these two functions can be chosen arbitrarily, for they must also obey two important identities. The first one follows from the definition of the *conditional density*,

$$f(x_1, x_2; t_1, t_2) = f(x_2, t_2 | x_1, t_1) \cdot f(x_1, t_1), \quad (3.3)$$

(where $f(x_2, t_2|x_1, t_1)$ can be thought of as the *transition probability* for reasons that will become clear below when we discuss the master equation). By integration over x_1 , we immediately have,

$$f(x_2, t_2) = \int_{-\infty}^{\infty} f(x_2, t_2|x_1, t_1) f(x_1, t_1) dx_1. \quad (3.4)$$

The second identity is obtained from Eq. 3.2,

$$f(x_1, x_2, x_3; t_1, t_2, t_3) = f(x_3, t_3|x_2, t_2) \cdot f(x_2, t_2|x_1, t_1) \cdot f(x_1, t_1) \quad (\text{Eq. 3.2}).$$

With $t_1 < t_2 < t_3$, integration over x_2 gives,

$$f(x_1, x_3; t_1, t_3) = f(x_1, t_1) \int_{-\infty}^{\infty} f(x_3, t_3|x_2, t_2) f(x_2, t_2|x_1, t_1) dx_2,$$

and, using $f(x_1, x_3; t_1, t_3) = f(x_3, t_3|x_1, t_1) \cdot f(x_1, t_1)$ (Eq. 3.3),

$$f(x_3, t_3|x_1, t_1) = \int_{-\infty}^{\infty} f(x_3, t_3|x_2, t_2) f(x_2, t_2|x_1, t_1) dx_2, \quad (3.5)$$

which is known as the *Chapman-Kolmogorov Equation* (Figure 3.1). It is a functional equation relating all conditional probability densities $f(x_i, t_i|x_j, t_j)$ for a Markov process, where the time ordering in the integrand is essential. The converse is also true: if $f(x_1, t_1)$ and $f(x_2, t_2|x_1, t_1)$ obey the consistency condition, Eqs. 3.4 and 3.5, then they uniquely define a Markov process.

Remarks:

1. The Chapman-Kolmogorov equation is a *functional equation* for the transition probability $f(x_i, t_i|x_j, t_j)$; its solution would give us a complete description of *any* Markov process – Unfortunately, no general solution to this equation is known.
2. From the meaning of $f(x, t|x_0, t_0)$, it is clear that we must have,

$$f(x, t|x_0, t_0) \rightarrow \delta(x - x_0) \quad \text{as } t \rightarrow t_0.$$

For example, an important Markov process is the one whose transition probability is normal; *i.e.* the Gaussian,

$$f(x_2, t_2|x_1, t_1) = \frac{1}{\sqrt{4\pi D(t_2 - t_1)}} \exp \left[-\frac{1}{4} \frac{(x_2 - x_1)^2}{D(t_2 - t_1)} \right].$$

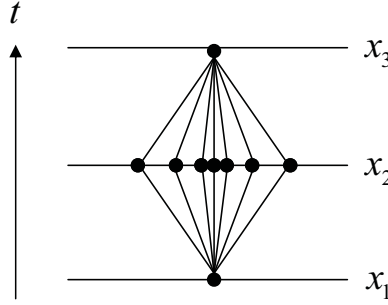


Figure 3.1: **Chapman-Kolmogorov Equation.** The intermediate variable x_2 is integrated over to provide a connection between x_1 and x_3 .

Convince yourself that it obeys the Chapman-Kolmogorov equation; and – if one chooses $f(x_1, 0) = \delta(x_1)$ and uses Eq. 2 – it follows from Eq. 3.4 that,

$$f(x, t) = \frac{1}{\sqrt{4D\pi t}} \exp \left[-\frac{1}{4} \frac{x^2}{Dt} \right], \quad (t \geq 0). \quad (3.6)$$

This is the probability density of the so-called *Wiener-Lévy process*. It is a Markov process, but despite appearances, it is *not* a stationary stochastic process, because

$$\langle \xi(t_1)\xi(t_2) \rangle = 2D \min(t_1, t_2),$$

(which is not a function of time-difference only – Exercise 1). Notice, however, that the transition probability $f(x_2, t_2|x_1, t_1)$ does depend on $(x_2 - x_1, t_2 - t_1)$ only. This process describes the *position* of a Brownian particle according to Einstein.

3. For a stationary Markov process, it is convenient to use a special notation. We let,

$$f(x_2, t_2|x_1, t_1) \equiv p(x_2|x_1, \tau), \quad \tau = t_2 - t_1,$$

in terms of which the Chapman-Kolmogorov equation reads,

$$p(x_3|x_1, \tau + \tau') = \int_{-\infty}^{\infty} p(x_3|x_2, \tau') p(x_2|x_1, \tau) dx_2, \quad (3.7)$$

where $\tau' = t_3 - t_2$, and $(\tau', \tau) > 0$ because of the time-ordering. We also have,

$$\begin{aligned} \int_{-\infty}^{\infty} p(x_2|x_1, \tau) dx_2 &= 1, \\ \int_{-\infty}^{\infty} p(x_2|x_1, \tau) f(x_1) dx_1 &= f(x_2), \\ B(\tau) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_2|x_1, \tau) f(x_1) dx_1 dx_2, \end{aligned}$$

the latter applying when the stationary Markov process has zero mean.

The Ornstein-Uhlenbeck process is the best-known example of a stationary Markov process; it is defined by:

$$\begin{aligned} f(x_1) &= \frac{1}{\sqrt{\pi(D/\beta)}} \exp \left[-\frac{x_1^2}{(D/\beta)} \right], \\ p(x_2|x_1, \tau) &= \frac{1}{\sqrt{\pi(D/\beta)(1 - e^{-2\beta\tau})}} \exp \left[-\frac{1}{(D/\beta)} \frac{(x_2 - x_1 e^{-\beta\tau})^2}{(1 - e^{-2\beta\tau})} \right]. \end{aligned} \quad (3.8)$$

This process describes the *velocity* of a Brownian particle according to Ornstein and Uhlenbeck.

It is straight-forward to verify that this process satisfies the Chapman-Kolmogorov equation, and that the correlation function is exponential (Exercise 2),

$$B(\tau) \propto e^{-\beta|\tau|}.$$

The Ornstein-Uhlenbeck process is *stationary, Gaussian and Markovian*; Doob (1942) proved that it is “essentially” the only process with these three properties.

3.3 Master Equation

Aside from providing a consistency check, the real importance of the Chapman-Kolmogorov equation is that it enables us to build up the conditional probability densities over the “long” time interval (t_1, t_3) from those over the “short” intervals (t_1, t_2) and (t_2, t_3) . It turns out this is an incredibly useful property. In physical applications, we generally have well-developed theories describing in detail (microscopically) the evolution of a system out of equilibrium. These theories are formulated in terms of differential equations describing the trajectories of the many ($\sim 10^{23}$) particles constituting the system – such as Hamilton’s equations in classical mechanics or Schrödinger’s equation in quantum mechanics. These descriptions are deterministic, and if we *could* solve the initial value problems for them, we would not need to think of approximations such as the splitting of time scales mentioned above, for we would have the solution for all time. [*Even* in that case, however, we would have a big puzzle on our hands; for while our fundamental microscopic theories are time-reversal invariant, such symmetry is lost at the macroscopic level and irreversibility dominates the evolution of physical phenomena. How does this irreversibility come about? This is one of the classic problems of statistical mechanics (see M. C. Mackey, *Time’s Arrow: The Origins of Thermodynamic Behavior*, (Dover, 2003), for an interesting discussion).]

There are several methods in physics that provide the microscopic dynamics over very short time scales, generating the transition probability between two states during the time interval Δt , as $\Delta t \rightarrow 0$ (*e.g.*, time-dependent perturbation theory in quantum mechanics). This is not enough, of course, for we need to know the evolution of the system on a time scale of the order of the time it takes to perform an experiment.

It is here, namely in bridging these two time scales, that the Markovian assumption helps: from our knowledge of the transition probability at small times we can build up our knowledge of the transition probability at all time iteratively from the Chapman-Kolmogorov equation.

For a large class of systems, it is possible to show that over very short time, the transition probability is,

$$p(x|z, \tau') = (1 - a_0\tau')\delta(x - z) + \tau'w(x|z) + o(\tau'), \quad (3.9)$$

where $w(x|z)$ is the transition probability per unit time, and a_0 is the

zeroth-jump moment,

$$a_0(z) = \int_{-\infty}^{\infty} w(x|z) dx. \quad (3.10)$$

The physical content of Eq. 3.9 is straightforward – it simply says that: the probability that a transition ($z \rightarrow x$) occurs + the probability that no transition occurs during that time (*i.e.* $z = x$) = the transition probability of moving from z to x during time τ' .

This will be the case in systems where the fluctuations arise from approximating a discrete stochastic process by a continuous deterministic model – for example, models of chemical reaction kinetics, predator-prey dynamics, particle collisions, radioactive decay, *etc.*... any model where it is the particulate nature of the variables that necessitates a stochastic formulation.

Substitute Eq. 3.9 into the Chapman-Kolmogorov equation,

$$\begin{aligned} p(x_3|x_1, \tau + \tau') &= \\ \int_{-\infty}^{\infty} [(1 - a_0(x_2)\tau')\delta(x_3 - x_2) + \tau'w(x_3|x_2)]p(x_2|x_1, \tau) dx_2 \\ &= \int_{-\infty}^{\infty} (1 - a_0(x_2)\tau')\delta(x_3 - x_2)p(x_2|x_1, \tau) dx_2 + \\ &\quad + \tau' \int_{-\infty}^{\infty} w(x_3|x_2)p(x_2|x_1, \tau) dx_2 \\ &= (1 - a_0(x_3)\tau')p(x_3|x_1, \tau) + \tau' \int_{-\infty}^{\infty} w(x_3|x_2)p(x_2|x_1, \tau) dx_2. \end{aligned}$$

Re-arranging, dividing by τ' , and recalling the definition of a_0 (*cf.* Eq. 3.10), we write

$$\begin{aligned} \frac{p(x_3|x_1, \tau + \tau') - p(x_3|x_1, \tau)}{\tau'} &= - \int_{-\infty}^{\infty} w(x_2|x_3)p(x_3|x_1, \tau) dx_2 \\ &\quad + \int_{-\infty}^{\infty} w(x_3|x_2)p(x_2|x_1, \tau) dx_2. \end{aligned}$$

Passing to the limit $\tau' \rightarrow 0$, we have¹,

$$\frac{\partial}{\partial \tau} p(x_3|x_1, \tau) = \int_{-\infty}^{\infty} [w(x_3|x_2) p(x_2|x_1, \tau) - w(x_2|x_3) p(x_3|x_1, \tau)] dx_2, \quad (3.11)$$

which is generally called the *Master Equation*. Note that it is a conservation equation of the gain-loss type, and there is of course a discrete version,

$$\frac{d}{dt} p_n(t) = \sum_m [w_{nm} p_m(t) - w_{mn} p_n(t)], \quad (3.12)$$

where the label n refers to the possible states of the stochastic process $\xi(t)$. The transition probabilities w_{nm} denote the probability of a transition from state m to state n in a small time increment dt .

The big difference between the Chapman-Kolmogorov equation and the master equation is that the Chapman-Kolmogorov equation is a *non-linear* equation (in the transition probabilities) that expresses the Markov character of the process, but containing no information about any *particular* Markov process. In the master equation, by contrast, one considers the transition probability at short times, $w(x_j|x_i)$, as a given function determined by the *specific* physical system, and the resulting equation is *linear* in the conditional probability density which determines the (mesoscopic) state of that system. The derivation and utility of the master equation formalism is best appreciated by example. We first consider the simplest master equation, the birth-and-death process, which is how Smoluchowski formulated his study of Brownian motion.

3.4 *Stosszahlansatz*

At this point, it is well worth taking a step back and examining the various assumptions that have been made so far (Figure 6.3). We begin with a general functional equation expressing a consistency condition for Markov processes (the Chapman-Kolmogorov equation). Assuming the transition probabilities are stationary, we are able to express the

¹This limit is taken in an asymptotic sense: τ' is much smaller than the characteristic time scale over which $p(x_3|x_1, \tau)$ varies, yet it is much larger than the time required for all of the microscopic variables in the system to relax to their equilibrium distribution; see Section 3.4 on page 59.

Chapman-Kolmogorov equation as a discrete-differential equation (the master equation) that turns out to be far more amenable to analysis. The mathematical derivation is straightforward, but it is important that the underlying *physical* assumptions are clear.

The master equation rests upon the assumption that on the time scale Δt over which the observable state evolves, all of the microscopic auxiliary variables (for example, the motion of the solvent molecules colliding with a Brownian particle) assume their stationary equilibrium distributions. Furthermore, the equilibrium distribution of the microscopic variables at time $t + \Delta t$ depends only upon the state of the system at time t . Thus at every time step Δt , the microscopic variables are perturbed from equilibrium, then rapidly relax to their new equilibrium distribution. This is called the *repeated randomness* assumption, or the *Stosszahlansatz*.

What is essential to bear in mind is that the limit $\Delta t \rightarrow 0$ used in the derivation of the master equation is *not* a true limit in the mathematical sense (*cf.* Eq. 3.11 on page 59) – Although $\Delta t \rightarrow 0$, we must remember that Δt is still long enough on the microscopic scale that the microscopic variables are allowed to relax to their equilibrium state so that the *microscopic* state is (approximately) independent at times t and $t + \Delta t$. A more detailed description of the *physical* derivation of the master equation is provided by van Kampen in “Fluctuations in Nonlinear Systems,” appended to these notes, and quoted in part below:

We are concerned with systems that consist of a very large number $\mathcal{N}$ of particles. In classical theory, the precise *microscopic* state of the system is described by $6\mathcal{N}$ variables $x_1, \dots, x_{3\mathcal{N}}, p_1, \dots, p_{3\mathcal{N}}$. They obey the $6\mathcal{N}$ microscopic equations of motion. The gross, *macroscopic* aspect of the state is described by a much smaller number of variables $Q_1, \dots, Q_n$, which are functions of $x_1, \dots, p_{3\mathcal{N}}$. For convenience we suppose that apart from the energy there is just one other Q , and drop the label. Experience tells us the remarkable fact that this macroscopic variable $Q(x_1, \dots, p_{3\mathcal{N}})$ obeys again a differential equation

$$\dot{Q} = F(Q), \quad (3.13)$$

which permits to uniquely determine its future values from its value at some initial instant. The phenomenological law (3.13) is *not* a purely mathematical consequence of the microscopic equations of motion. The reason why it exists can be roughly

understood as follows. Using the equations of motion one has

$$\dot{Q} = \sum_{k=1}^{3\mathcal{N}} \left(\frac{\partial Q}{\partial x_k} \dot{x}_k + \frac{\partial Q}{\partial p_k} \dot{p}_k \right) \equiv g(x_1, \dots, p_{3\mathcal{N}}).$$

The variables in g may be expressed in Q and the energy (which we do not write explicitly), and $6\mathcal{N} - 2$ remaining variables, $\vartheta_\lambda(x_1, \dots, p_{3\mathcal{N}})$ say. Hence

$$\dot{Q} = f(Q; \vartheta_1, \dots, \vartheta_{6\mathcal{N}-2}).$$

This may also be written

$$Q(t + \Delta t) - Q(t) = \int_t^{t+\Delta t} f[Q(t'); \vartheta(t')] dt'.$$

Now suppose that $Q(t)$ varies much more slowly than the ϑ_λ (which is the reason it is microscopic). It is then possible to pick Δt such that $Q(t)$ does not vary much during Δt , while the ϑ_λ practically run through all their possible values (ergodic theorem with fixed value for Q). Hence one may substitute in the integral $Q(t)$ for $Q(t')$ and replace the time integration by an average over that part of the phase space that corresponds to given values of the energy and Q :

$$Q(t + \Delta t) - Q(t) = \Delta t \cdot \langle f[Q(t); \vartheta] \rangle_{Q(t)} = \Delta t \cdot F[Q(t)].$$

It should be emphasized that this implies that at each time t the ϑ_λ vary in a sufficiently random way to justify the use of a phase space average ("repeated randomness assumption").

Fluctuations arise from the fact that, in the relevant part of phase space, f is not exactly equal to its average F , but has a probability distribution around it. Hence $Q(t + \Delta t)$ is no longer uniquely determined by $Q(t)$, but instead there exists a transition probability $W(q'|q)$. More precisely, $\Delta t W(q'|q) dq'$ is the probability that, if Q has the value q at time t , the value of $Q(t + \Delta t)$ will lie between q' and $q' + dq'$. The probability distribution $P(q, t)$ of Q at any time t then obeys the rate equation

$$\frac{\partial P(q, t)}{\partial t} = \int \{W(q|q')P(q', t) - W(q'|q)P(q, t)\} dq'. \quad (3.14)$$

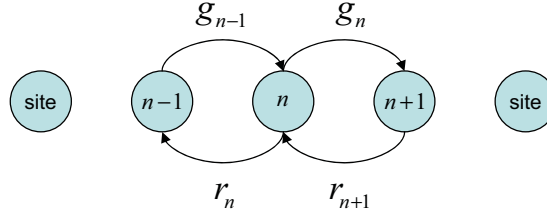


Figure 3.2: **Schematic illustration of a birth-death process**

This is the general form of the master equation ... Again a repeated randomness assumption is involved, namely that at each time the ϑ_λ are sufficiently random to justify the identification of probability with measure in phase space.

– N. G. van Kampen, *Fluctuations in Nonlinear Systems*.

In the following examples, it is the *derivation* of the master equation that is emphasized; discussion of actual *solution* methods is postponed to Chapter 4 – but by all means, read ahead if you're curious.

3.5 Example – One-step processes

Many stochastic processes are of a special type called *one-step process*, *birth-and-death process* or *generation-recombination process*. They are continuous-time Markov processes whose range consists of the integers n , and whose transition probability per unit time (*i.e.* w_{nm}) permits only jumps between adjacent sites,

$$\begin{aligned} w_{nm} &= r_m \delta_{n,m-1} + g_m \delta_{n,m+1}, \quad (m \neq n) \\ w_{nn} &= 1 - (r_n + g_n), \end{aligned}$$

where r_n is the probability per unit time that, being at site n , a jump occurs to site $n - 1$. Conversely, g_n is the probability per unit time that, being at site n , a jump occurs to site $n + 1$ (Figure 3.2).

Smoluchowski applied this model to the study of Brownian motion by setting the generation and recombination probabilities to $1/2$: $g_n = r_n = 1/2$. Defining the probability density $p_{n|m,s}$ as the probability that a random walker beginning at n at time $t = 0$ will be at site m after s steps, the master equation for this unbounded random walk is,

$$p_{n|m,s+1} = \frac{1}{2} p_{n|m-1,s} + \frac{1}{2} p_{n|m+1,s}. \quad (3.15)$$

Let $\nu = |m - n|$ be the net distance traveled by the random walker, then one can show that $p_{n|m,s}$ is given by,

$$p_{n|m,s} = \begin{cases} \frac{1}{2^s} \frac{s!}{\left(\frac{s+\nu}{2}\right)! \left(\frac{s-\nu}{2}\right)!}; & \nu \leq s \text{ and } \nu + s \text{ even,} \\ 0; & \text{otherwise.} \end{cases} \quad (3.16)$$

It is possible to show that this discrete random walk formulation is equivalent with Einstein's diffusion representation of free Brownian motion (Exercise 6a).

Note that it is often the case in literature that the authors will not explicitly write the conditioning of the distribution by the initial state, preferring instead the short-hand: $p_{n|m,s} \equiv p_{m,s}$ or $p_m(s)$. One must always remember, however, that any distribution governed by a master equation (or any other dynamical equation) is *necessarily conditioned by the initial distribution*.

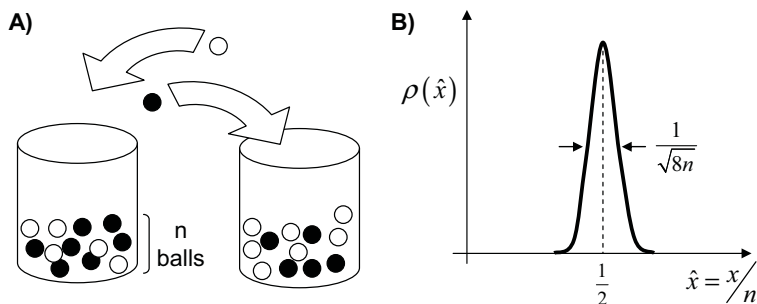


Figure 3.3: **Bernoulli's urn model.** **A)** Two urns, A and B, each contain n balls, with n of the $2n$ balls white and n black. A ball is drawn at random from each urn and then the ball that came from A is placed in urn B, and the ball that came from B is placed in urn A. **B)** Laplace's approximation of the fraction of white balls in urn A ($\hat{x} = x/n$) after a very long time. The distribution is Gaussian, with standard deviation $1/\sqrt{8n}$.

3.6 Example – Bernoulli's Urns and Recurrence

- D. Bernoulli (1770) "Disquisitiones analyticae de novo problemate conjecturali," *Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae* **XIV**: 3–25.
- P. S. Laplace (1812) *Théorie Analytique des Probabilités. Livre II. Théorie Générale des Probabilités*, Chapter 3.
- M. Jacobsen (1996) "Laplace and the origin of the Ornstein-Uhlenbeck process," *Bernoulli* **2**: 271–286.

The following problem was posed by Bernoulli (see Figure 3.3):

There are n white balls and n black balls in two urns, and each urn contains n balls. The balls are moved cyclically, one by one, from one urn to another. What is the probability $z_{x,r}$ that after r cycles, urn A will contain x white balls?

In attempting to solve this problem, Laplace derived the following differ-

ence equation for the evolution of $z_{x,r}$ – in words,

$$\begin{aligned}
 & \left[\begin{array}{c} \text{Prob. } x \text{ white balls} \\ \text{in urn A after} \\ r + 1 \text{ cycles} \end{array} \right] = \\
 & \left[\begin{array}{c} \text{Prob. } x + 1 \text{ white balls} \\ \text{in urn A after} \\ r \text{ cycles} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{white ball} \\ \text{from urn A} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{black ball} \\ \text{from urn B} \end{array} \right] \\
 & + \left[\begin{array}{c} \text{Prob. } x - 1 \text{ white balls} \\ \text{in urn A after} \\ r \text{ cycles} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{black ball} \\ \text{from urn A} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{white ball} \\ \text{from urn B} \end{array} \right] \\
 & + \left[\begin{array}{c} \text{Prob. } x \text{ white balls} \\ \text{in urn A after} \\ r \text{ cycles} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{white ball} \\ \text{from urn A} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{white ball} \\ \text{from urn B} \end{array} \right] \\
 & + \left[\begin{array}{c} \text{Prob. } x \text{ white balls} \\ \text{in urn A after} \\ r \text{ cycles} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{black ball} \\ \text{from urn A} \end{array} \right] \left[\begin{array}{c} \text{Prob. of} \\ \text{drawing a} \\ \text{black ball} \\ \text{from urn B} \end{array} \right], \tag{3.17}
 \end{aligned}$$

where the first term represents the *loss* of a white ball from urn A, the second term represents the *gain* of a white ball to urn A, and the remaining terms contribute no change in the number of white balls in urn A. The difference equation for this process is (much) more concisely written as,

$$z_{x,r+1} = \left(\frac{x+1}{n} \right)^2 z_{x+1,r} + 2 \frac{x}{n} \left(1 - \frac{x}{n} \right) z_{x,r} + \left(1 - \frac{x-1}{n} \right)^2 z_{x-1,r}. \tag{3.18}$$

Notice that in contrast to the master equation for the simple birth-death process described in the preceding section, Eq. 3.18 has *nonlinear* transition rates. If the transition rates are nonlinear, it is very difficult to find an explicit solution for the probability $z_{x,r}$, so Laplace sought an approximate solution instead. His approach will be described in some detail

because the method is essentially the same as van Kampen's *linear noise approximation*, which we will meet again in Section 5.1.

First, Laplace assumes the number of balls n is large, so that changes $x \pm 1$ are almost infinitesimal with respect to the *density* of white balls x/n , allowing the differences to be replaced with differentials,

$$\begin{aligned} z_{x \pm 1, r} &\approx z_{x, r} \pm \frac{\partial}{\partial x} z_{x, r} + \frac{1}{2} \frac{\partial^2}{\partial x^2} z_{x, r} \\ z_{x, r+1} &\approx z_{x, r} + \frac{\partial}{\partial r} z_{x, r}. \end{aligned} \quad (3.19)$$

Second, Laplace makes a change of variables, $r = nr'$, and

$$x = \frac{1}{2}n + \sqrt{n} \mu; \quad (3.20)$$

we shall see in Section 5.1 that this change of variables is tantamount to assuming that the *density* of white balls, x/n , is distributed about the “deterministic” value $1/2$, with some “fluctuations” parameterized by μ , and that the magnitude of these fluctuations scale with $\sqrt{n}$. With Eq. 3.20, the nonlinear transition rates can be expanded in powers of $1/n$, which, along with Eq. 3.19, leads to a linear partial differential equation governing the distribution of μ , $U(\mu, r)$,

$$\frac{\partial U(\mu, r')}{\partial r'} = 2 \frac{\partial}{\partial \mu} (\mu U(\mu, r')) + \frac{1}{2} \frac{\partial^2}{\partial \mu^2} \left(\frac{1}{2} U(\mu, r') \right). \quad (3.21)$$

This is an example of a *Fokker-Planck equation* (cf. Chapter 6; Eq. 6.28), and in particular, it characterizes what is called an *Ornstein-Uhlenbeck process* which we met in Chapter 1 with Ornstein and Uhlenbeck's study of the velocity of a Brownian particle (cf. Section 1.2.4). Assuming that n is large enough that the range of μ is unbounded ($\mu \in (-\infty, \infty)$), one can easily verify that the steady-state solution of Eq. 3.21 is the Gaussian²,

$$\lim_{r' \rightarrow \infty} U(\mu, r') = U(\mu) = \frac{2}{\sqrt{\pi}} e^{-4\mu^2}. \quad (3.22)$$

Reverting to the original variables by writing a continuous probability density $\rho(x/n)$ for the fraction of white balls in urn A ($\hat{x} = x/n$), and using Eq. 3.22,

$$\rho(\hat{x}) = 2 \sqrt{\frac{n}{\pi}} \exp \left[-4n \left(\hat{x} - \frac{1}{2} \right)^2 \right], \quad (3.23)$$

²The connection between the work of Ornstein and Uhlenbeck becomes apparent if one makes the substitution $\frac{m}{k_B T} = 8$ in Eq. 1.34.

which is a Gaussian distribution, with mean $1/2$ and variance $1/8n$. As the number of balls increases ($n \rightarrow \infty$), the density becomes more narrowly peaked at $\hat{x} = 1/2$, approaching a delta-function. In Section 4.1.2 (on page 84), we shall see that Laplace's approximation is extremely good all the way down to $n \approx 10$.

A model very similar to Bernoulli's urn model was proposed at the beginning of the 20th century by Paul and Tatiana Ehrenfest to clarify some foundational concerns that continued to undermine Boltzmann's formulation of statistical mechanics (see Exercise 7). The trouble lay with the precise meaning of *irreversibility* in thermodynamics. For example, a hot liquid will spontaneously, and irreversibly, lose heat to the surrounding environment until the temperature of the liquid is in equilibrium with the surroundings. Yet, if we treat the system as a microscopic dynamical system, Poincaré proved that almost every state of the system will be revisited eventually, to an arbitrarily prescribed degree of accuracy (this is Poincaré's *Wiederkehrrsatz*). Zermelo argued that the irreversibility of thermodynamics and the recurrence properties of dynamical systems are incompatible. Boltzmann replied that Poincaré's *Wiederkehrrsatz* is a mathematical result, true in the limit of infinite time, but that in practice the time-span between the recurrence of an unlikely state is unimaginably long, and so the statistical mechanical processes underlying thermodynamics are irreversible for all intents and purposes. One can make Boltzmann's argument quantitative using the following simple urn model,

Imagine $2R$ balls, numbered consecutively from 1 to $2R$, distributed in two boxes (I and II), so that at the beginning there are $R + n$ ($-R \leq n \leq R$) balls in box I. We choose a random integer between 1 and $2R$ (all integers are supposed to be equiprobable) and move the ball, whose number has been drawn, from the box in which it is to the other box. This process is repeated s times and we ask for the probability $Q_{R+m,s}$ that after s drawings there will be $R + m$ balls in box I.

M. Kac proves in his classic paper on Brownian motion (M. Kac (1947) "Random walk and the theory of Brownian motion," *The American Mathematical Monthly* **54**: 369–391) that irrespective of how the initial state is prepared, *every* possible state is visited with probability one – this is Poincaré's *Wiederkehrrsatz*. Nevertheless, Kac also shows that excursions from equilibrium return exponentially quickly (Newton's law of cooling) and that if the number of balls is large and the initial state is far from equilibrium, then the recurrence time is enormously long. Specifically, the number of draws, s_{recur} , required on average for the recurrence of a

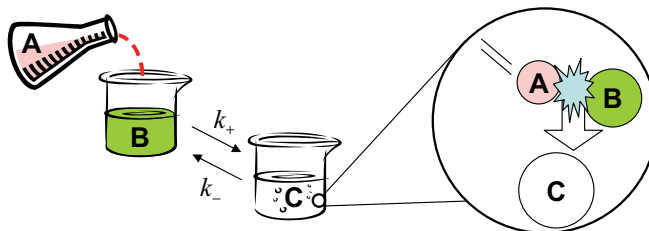


Figure 3.4: Noise in chemical reaction kinetics.

state beginning with $R + n$ balls in one urn is given by,

$$s_{recur} = \frac{(R+n)!(R-n)!}{(2R)!} 2^{2R}.$$

If, for example, we begin with $R = 10000$ and $n = 10000$ (*i.e.* all 20000 balls in Urn I), and each drawing takes 1 second, then on average we will have to wait more than 10^{6000} years (!) for this state to re-occur. On the other hand, close to equilibrium, neighbouring states are visited often: if we begin with $R = 10000$ and $n = 0$ (*i.e.* half of the 20000 balls in Urn I), then on average we must wait about $100\sqrt{\pi} \sim 175$ seconds for the state to re-occur.

3.7 Example – Chemical Kinetics

- D. McQuarrie (1967) “Stochastic approaches to chemical kinetics,” *Journal of Applied Probability* **4**: 413.
- J. Elf and M. Ehrenberg (2003) “Fast evaluation of fluctuations in biochemical networks with the linear noise approximation,” *Genome Research* **13**: 2475.
- D. T. Gillespie (1977) “Exact simulation of coupled chemical reactions,” *Journal of Chemical Physics* **81**: 2340.

Very often, the kinetics of chemical reactions are described in terms of deterministic chemical rate equations, which take the form of a system of coupled nonlinear differential equations. Underlying that formulation is the implicit assumption that the concentration of the reactants varies both continuously and differentially. For moles of reactants (*i.e.* molecule numbers of the order 10^{23}), these assumptions are perfectly justified since

a change of one or two molecules in a population of 10^{23} is, for all intents and purposes, infinitesimal. That accounts for the great success of deterministic models in most macroscopic systems, including freshman chemistry labs (Figure 3.4). For small pools of reactants, however, the mathematical formulation becomes more delicate.

Inside living cells, reactant numbers tend to be of the order 1-1000. A reaction altering the population by one or two therefore generates a large relative change, and the molecule numbers no longer evolve differentiably (Figure 3.5). Furthermore, reactions no longer occur ‘continuously’ over an infinitely small time interval, but rather progress in a series of steps of finite time width. By way of analogy, one can imagine the national birth rate as compared to the chances my next-door neighbor will have a baby. One often hears statements such as: “Every 10 minutes, a baby is born in this country.” That clearly cannot be true of my next-door neighbor. Evolution of the population of an entire country can be well-described using differential equations, but the evolution of the population of a small town occurs in a stochastic, step-wise manner. This example illustrates the dichotomy of *discrete* evolution of individuals on the one hand, and (nearly) *continuous* evolution of the population density on the other. We can make this relationship explicit by writing the number of individuals n as being proportional to a density X ; with the constant of proportionality Ω being a measure of the system size,

$$n = \Omega \cdot X.$$

In the urn example above, Ω was the total number of balls in one urn. In chemical reaction dynamics, Ω is usually the *volume* of the reaction vessel.

It is straightforward to argue that in the gas phase (or a very dilute solution), the dynamics of chemical reaction networks can be reasonably described by a master equation governing the probability density for the molecule *numbers* $\mathbf{n}$,

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_{\mathbf{n}'} w_{\mathbf{n}\mathbf{n}'} P(\mathbf{n}', t) - w_{\mathbf{n}'\mathbf{n}} P(\mathbf{n}, t). \quad (3.24)$$

As above, $w_{\mathbf{n}\mathbf{n}'}$ denotes the transition probability *from* the state $\mathbf{n}'$ *to* the state $\mathbf{n}$. In contrast with examples studied so far, $P(\mathbf{n}, t)$ is usually a *multivariate* probability distribution, depending upon state variables for all the products and reactants of interest in the network. The transition probabilities w_{ij} are generalizations of the deterministic reaction rates, and in fact we shall find that the coefficients appearing in the deterministic

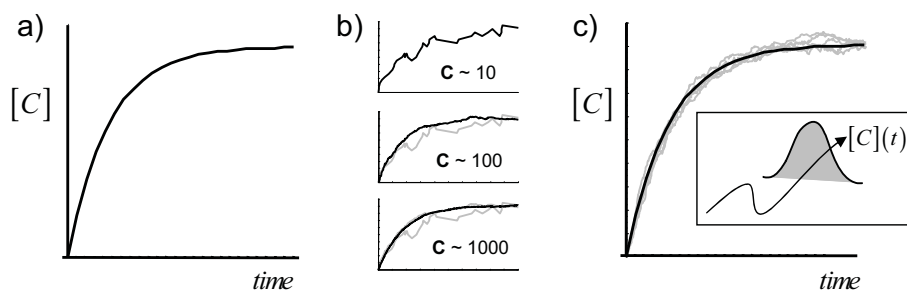


Figure 3.5: **Noise in chemical reaction kinetics.** **a)** For many reactant molecules, the species concentrations evolve both continuously and differentiably. **b)** When small numbers of reactants are involved, due to the probabilistic nature of individual reaction events and the finite change in molecule numbers incurred, the concentration evolves step-wise. Although as the reactant numbers increase, the relative size of the jumps decreases. **c)** Repeating an experiment many times, we typically obtain some repeatable averaged behavior that conforms very nearly to the deterministic description and some envelope around the average that accounts for the fluctuations in individual trajectories. *Inset:* We can imagine the full evolution of the system as composed of two parts: the deterministic evolution of $[C](t)$ and a probability distribution for the fluctuations that moves along $[C](t)$. The width of the probability distribution scales roughly as $\frac{1}{\sqrt{N}}$ where N is the number of molecules.

rate equations are simply proportional to the mean of an exponential distribution.

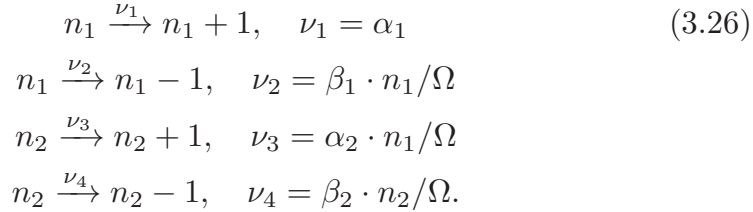
For multidimensional master equations evolving over a *discrete* state space, it is convenient to introduce several new objects: the step-operator $\mathbf{E}$, the stoichiometry matrix $\mathbf{S}$ and the propensity vector $\boldsymbol{\nu}$.

The step-operator $\mathbf{E}$ is short-hand for writing evolution over discrete space. The action of the operator is defined in the following way: $\mathbf{E}_i^k$ increments the i^{th} variable by an integer k . That is, for a function $f(n_1, n_2, \dots, n_i, \dots)$ depending upon several variables,

$$\mathbf{E}_i^k f(n_1, n_2, \dots, n_i, \dots) = f(n_1, n_2, \dots, n_i + k, \dots). \quad (3.25)$$

The stoichiometry matrix $\mathbf{S}$ describes *how much* each species changes with the completion of a given reaction, while the propensity vector $\boldsymbol{\nu}$ describes *the rate at which a particular reaction proceeds*. An example should make this notation more clear.

Example: Coupled Poisson processes – To introduce the stoichiometry matrix and the propensity vector, consider a simple linear two-state model:



All of the transition rates are linear or constant making it possible to solve for the probability distribution $P(n_1, n_2, t)$ exactly, at least in principle (see Section 5.2 on page 113). We shall not attempt to do so here, but focus instead on how to represent the reaction network in a convenient manner.

Generally, we record the reaction propensities in a vector $\boldsymbol{\nu}$ and the stoichiometries in a matrix $\mathbf{S}$ defined such that when the j^{th} reaction occurs it increments the i^{th} reactant by an integer S_{ij} : $n_i \xrightarrow{\nu_j} n_i + S_{ij}$. The collection of the elements S_{ij} compose the stoichiometry matrix. In the present example,

$$\mathbf{S} = \begin{array}{cccc} & \nu_1 & \nu_2 & \nu_3 & \nu_4 \\ \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} & n_1 \\ & & & & n_2 \end{array}$$

where each column of the stoichiometry matrix corresponds to a particular reaction and each row to a particular reactant. Using this notation, in the limit $\Omega \rightarrow \infty$ with $n_i/\Omega = x_i$ held constant, the deterministic rate equations can be written in terms of the stoichiometry matrix and the propensity vector,

$$\frac{d\mathbf{x}}{dt} = \lim_{\Omega \rightarrow \infty} \mathbf{S} \cdot \boldsymbol{\nu}, \quad (3.27)$$

or, explicitly as a system of coupled linear ordinary differential equations,

$$\begin{aligned} \frac{dx_1}{dt} &= \alpha_1 - \beta_1 \cdot x_1, \\ \frac{dx_2}{dt} &= \alpha_2 \cdot x_1 - \beta_2 \cdot x_2. \end{aligned}$$

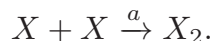
The real convenience of describing the reaction network in terms of $\mathbf{S}$ and $\boldsymbol{\nu}$ comes in writing the master equation. Piecing together each term in the equation from the various reactions can be tedious, but an explicit expression exists – For a system of R reactions and N reactants (*i.e.*, $\mathbf{n} \in \mathbb{R}^N$), the master equation is

$$\frac{\partial P}{\partial t} = \Omega \sum_{j=1}^R \left[\left(\prod_{i=1}^N \mathbf{E}_i^{-S_{ij}} \right) - 1 \right] \nu_j(\mathbf{n}) P(\mathbf{n}, t), \quad (3.28)$$

where ν_j is the *microscopic* reaction propensity (with units of concentration per time) explained below. Furthermore, for numerical and analytic approximations of the solution of the master equations, most schemes are concisely written in terms of $\mathbf{S}$ and $\boldsymbol{\nu}$.

Brief note about microscopic reaction propensities $\boldsymbol{\nu}$

For nonlinear transition rates, the propensity appearing in the master equation is *not* identical with the propensity in the deterministic rate equations (Eq. 3.27). The difference is not difficult to understand. Consider, for example, the dimerization reaction,



In a deterministic system, where the number of molecules is very large, the rate of accumulation of the product X_2 is written,

$$\frac{d}{dt} [X_2] = a [X]^2. \quad (3.29)$$

Microscopically, what we mean by a reaction event is that two molecules of X find one another with sufficient energy that they form the dimer X_2 . The probability for the reaction to occur is then proportional to the number of ways two molecules can collide,

$$\frac{d}{dt} [X_2] \propto \frac{1}{2} (n_X) (n_X - 1), \quad (3.30)$$

where n_x is the *number* of molecules of X in a reaction vessel of volume Ω . The last term on the right-hand side is $(n_X - 1)$ because we need at least two molecules to have a reaction, and the $1/2$ factor comes from not double-counting each reactant - one molecule colliding with another is the same as the other colliding with the one. Here, and throughout, the fraction accounting for different permutations of the reactants will be absorbed into the rate of reaction so that we can write, in units of concentration per time,

$$\underbrace{\nu \left(\frac{n_X}{\Omega} \right) = \frac{a'}{2} \cdot \frac{n_X}{\Omega} \cdot \frac{n_X - 1}{\Omega}}_{\text{microscopic reaction rate}} \xrightarrow{n_X \rightarrow \infty} \underbrace{\tilde{\nu}([X]) = a \cdot [X] \cdot [X]}_{\text{macroscopic reaction rate}}. \quad (3.31)$$

(with $a'/2 = a$). In numerical simulation of chemical reaction kinetics (Section 4.2.1), the microscopic reaction propensities ν are used, while in asymptotic solutions of the master equation (Section 5.1), to the order of approximation that we will be concerned with in this course, the macroscopic propensities $\tilde{\nu}$ are sufficient.

Suggested References

The text by Gardiner,

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004).

has a thorough discussion of Markov processes. The text by Gillespie,

- *Markov Processes*, D. T. Gillespie (Academic Press, Inc, 1992).

uses a distinctive approach to Markov processes that some readers may enjoy. The mathematics is a bit heavy, but the connection among many aspects of Markov processes are united into a single framework.

The pedagogical article by Kac (pronounced ‘Katz’),

- M. Kac (1947) “Random walk and the theory of Brownian motion,” *The American Mathematical Monthly* **54**: 369–391,

is a classic, as is the lengthy review by Chandrasekhar,

- S. Chandrasekhar (1943) “Stochastic problems in physics and astronomy,” *Review of Modern Physics* **15**: 1–89.

Finally, Elf and Ehrenberg present the stoichiometry matrix and propensity vector in the context of several applications,

- J. Elf and M. Ehrenberg (2003) “Fast evaluation of fluctuations in biochemical networks with the linear noise approximation,” *Genome Research* **13**: 2475.

Exercises

1. **Wiener process:** Use the defining features of the Wiener process $X(t)$, *i.e.* the stationary-independent increments and the 1st-order probability density given by (Eq. 3.6),

$$f(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp \left[-\frac{x^2}{4Dt} \right] \quad (t \geq 0),$$

to show the following:

- (a) The autocovariance is given by,

$$\langle \langle X(t_1)X(t_2) \rangle \rangle = 2D \min(t_1, t_2).$$

- (b) The increments of the Wiener process are un-correlated on disjoint time-intervals,

$$\langle [X(t_4) - X(t_3)] [X(t_2) - X(t_1)] \rangle = 0, \quad \text{for any } t_4 > t_3 > t_2 > t_1.$$

2. **Ornstein-Uhlenbeck process:** In the following, use the defining expressions for the Ornstein-Uhlenbeck process $X(t)$ (Eq. 3.8),

$$f(x_1) = \frac{1}{\sqrt{\pi(D/\beta)}} \exp \left[-\frac{x_1^2}{(D/\beta)} \right],$$

$$p(x_2|x_1, \tau) = \frac{1}{\sqrt{\pi(D/\beta)(1 - e^{-2\beta\tau})}} \exp \left[-\frac{1}{(D/\beta)} \frac{(x_2 - x_1 e^{-\beta\tau})^2}{(1 - e^{-2\beta\tau})} \right],$$

(where $\tau \equiv t_2 - t_1 > 0$).

- (a) Compute explicitly the autocorrelation function, $\langle X(t_1)X(t_2) \rangle$.
 (b) Show that the compatibility condition is satisfied for $t_2 > t_1$,

$$\int_{-\infty}^{\infty} f(x_1, x_2; t_1, t_2) dx_1 = f(x_2; t_2).$$

- (c) Show that the increments of the Ornstein-Uhlenbeck process are negatively correlated on disjoint time intervals,

$$\langle [X(t_4) - X(t_3)][X(t_2) - X(t_1)] \rangle < 0, \quad \text{for any } t_4 > t_3 > t_2 > t_1.$$

- (d) If $Z(t)$ is the integral of the Ornstein-Uhlenbeck process $X(t)$,

$$Z(t) = \int_0^t X(\tau) d\tau,$$

then find, $\langle Z(t_1)Z(t_2) \rangle$.

- (e) If $Z(t)$ is the integral of the Ornstein-Uhlenbeck process $X(t)$, then find $\langle \cos [Z(t_1) - Z(t_2)] \rangle$. *Hint:* It may be convenient to use cumulants.

3. Chapman-Kolmogorov equation: Consider a stochastic process $X(t)$ which starts from $X(0) = 0$, takes continuous values x and is homogeneous in time and space (*i.e.*, has stationary independent increments), such that $f_{1|1}(x_2, t_2 | x_1, t_1) = f(x_2 - x_1, t_2 - t_1)$ for $t_2 \geq t_1 \geq 0$, where $f(x, t)$ is the first-order probability density.

- (a) Show that the Chapman-Kolmogorov equation,

$$f(x, t) = \int_{-\infty}^{\infty} f(x - y, t - \tau) f(y, \tau) dy,$$

is satisfied if the first-order cumulant-generating function has the form: $\ln G(k, t) = tg(k)$, where $g(k)$ is an arbitrary function of k .

Hint: Use the Fourier transform to convert the Chapman-Kolmogorov equation into a functional equation for the characteristic function $G(k, t)$ of the process.

- (b) Assuming that the time evolution of the process is governed by the (known) transition probability per time of the form $w(x'|x) = w(x' - x)$, apply the Fourier transform to the master equation for $f(x, t)$ and show that,

$$f(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[-ikx + t \int_{-\infty}^{\infty} w(x') (e^{ikx'} - 1) dx' \right] dk.$$

4. **Dichotomic Markov process (Random telegraph process):**

The function $Y(t)$ is a Markov process with range $\{-1, 1\}$, switching between the two at a rate γ .

- (a) Construct the master equation for the dichotomic Markov process.
- (b) Solve the master equation with the initial condition $y(t_0) = y_0$ to obtain the conditional probability density $P(y, t|y_0, t_0)$.
- (c) Show that this is consistent with the single-point probability $P(y, t) = \frac{1}{2}(\delta_{y,-1} + \delta_{y,1})$, where δ_{ij} is the Kronecker delta.
- (d) Repeat the above for a Markov process $Y(t)$ with a range $\{a, b\}$ and asymmetric transition probabilities: $a \xrightarrow{\alpha} b$ and $b \xrightarrow{\beta} a$. Compute the steady-state autocorrelation function for this process.

5. **Lead-probability:** Suppose that in a coin-toss game with unit stakes, Albert bets on heads, and Paul bets on tails. The probability that Albert leads in $2r$ out of $2n$ tosses is called the *lead probability* $P_{2r,2n}$. One can show that,

$$P_{2r,2n} = \binom{2r}{r} \binom{2n-2r}{n-r} 2^{-2n}.$$

- (a) For a coin tossed 20 times, what is the most likely number of tosses for which Albert is in the lead? Guess first, then make a table of $P_{2r,2n}$ for $n = 10$ and $r = 1, 2, \dots, 10$.
- (b) Show that the probability that $\frac{r}{n} < x$ is given by,

$$f(x) \sim \frac{2}{\pi} \arcsin \sqrt{x},$$

as $n \rightarrow \infty$. For a game where the coin is tossed every second for one year, what is the probability that one of the players will be in the lead less than one day out of 365?

- (c) For a one-dimensional random walk, would you expect a given realization to spend most of its time on the negative axis, the positive axis, or to hover near the origin, spending equal time on the positive and negative axes?

6. **Bernoulli's urn model and the approximation of Laplace:**

The urn model of Bernoulli has nonlinear transition rates, and therefore cannot be solved in general. Laplace, by a clever substitution, derived an approximate evolution equation.

- (a) Solve Laplace's partial differential equation, Eq. 3.21, at steady-state. That is, set the left-hand side equal to 0, and solve the resulting ordinary differential equation for $U(\mu)$ assuming $U(\mu), U'(\mu) \rightarrow 0$ exponentially as $\mu \rightarrow \pm\infty$. Comparing the solution with the canonical Gaussian probability density (*cf.* Eq. A.19 on page 239), and using the change of variables Eq. 3.20, what is the variance of the steady-state distribution for the white ball density $\hat{x} = x/n$? What can you say about the equilibrium state as $n \rightarrow \infty$?
 - (b) Multiply Eq. 3.21 by μ and integrate from $\mu \in (-\infty, \infty)$ to obtain an evolution equation for the *average* $\langle\mu\rangle$. (Use the same boundary conditions as above: $U(\mu), U'(\mu) \rightarrow 0$ exponentially as $\mu \rightarrow \pm\infty$.) From the resulting equation, show that $\langle\mu\rangle$ returns to $\langle\mu\rangle = 0$ exponentially quickly; *i.e.*, deviations from the equilibrium state $x = \frac{1}{2}n$ are restored exponentially quickly (Newton's law of cooling). *Hint:* Integrate by parts, assuming that the probability distribution and its first derivative both vanish as $\mu \rightarrow \pm\infty$.
 - (c) Repeat the above, but this time multiply by μ^2 and integrate from $\mu \in (-\infty, \infty)$ to obtain an evolution equation for the *variance* $\langle\mu^2\rangle$. Show that the variance asymptotically approaches a non-zero steady-state. How does this steady-state compare with the results obtained in question 6a?
7. **Ehrenfests' urn model:** The urn model proposed by Paul and Tatiana Ehrenfest to clarify the idea of irreversibility in thermodynamics is the following:

Imagine $2R$ balls, numbered consecutively from 1 to $2R$, distributed in two boxes (I and II), so that at the beginning there are $R + n$ ($-R \leq n \leq R$) balls in box I. We choose a random integer between 1 and $2R$ (all integers are supposed to be equiprobable) and move the ball, whose number has been drawn, from the box in which it is to the other box. This process is repeated s times and we ask for the probability $Q_{R+m,s}$ that after s drawings there will be $R + m$ balls in box I.

- (a) Write out the master equation governing the evolution of the probability distribution $Q_{R+m,s}$.

- (b) Following the method of Laplace discussed in Section 3.6, define a suitable continuous fluctuation variable μ and derive a partial differential equation for the evolution of the probability distribution for μ .
- (c) Multiply the partial differential equation derived in 7b by μ and integrate from $\mu \in (-\infty, \infty)$ to obtain an evolution equation for the *average* $\langle \mu \rangle$. From the resulting equation, show that $\langle \mu \rangle$ returns to $\langle \mu \rangle = 0$ exponentially quickly; *i.e.*, deviations from the equilibrium state are restored exponentially quickly (Newton's law of cooling). *Hint:* Integrate by parts, assuming that the probability distribution and its first derivative both vanish exponentially as $\mu \rightarrow \pm\infty$.
8. **Malthus's Law:** Malthus's law for population growth assumes that the birth and death rates are proportional to the number of individuals, respectively, $b(n) = \beta n$ and $d(n) = \alpha n$, with α and β given constants.
- (a) Use the probability generating function, Eq. 1.47, to solve the master equation for $f(n, t) = P\{N(t) = n | N(0) = n_0\}$ and show that, in the case $\alpha = \beta$, the probability of extinction goes as,

$$f(0, t) = \left[1 + (\beta t)^{-1}\right]^{-n_0},$$

where n_0 is the initial population.

- (b) Solve the deterministic rate equation for $\langle N(t) \rangle$ and $\langle N^2(t) \rangle$. Describe the time-dependence of the variance $\sigma^2(t)$ for the cases when $\beta > \alpha$, $\beta < \alpha$, and $\beta = \alpha$.

9. **Step-operator (Eq. 3.25):**

- (a) Prove the identity

$$\sum_{N=0}^{\infty} g(N) \mathbf{E} f(N) = \sum_{N=1}^{\infty} f(N) \mathbf{E}^{-1} g(N) \quad (3.32)$$

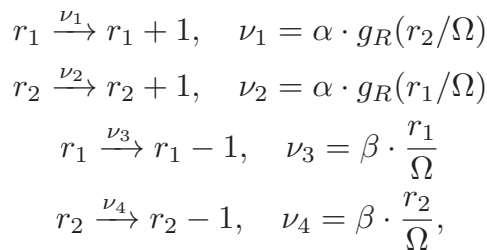
for any pair of functions f, g such that the sum converges.

- (b) Consider the decay of a radioactive isotope

$$X \xrightarrow{a} A, \quad (3.33)$$

where A is inert and does not enter into the equations.

- i. What is the deterministic equation describing this process?
 - ii. Write out the master equation governing the probability distribution of X .
 - iii. Compute the equation governing the mean $\langle X \rangle$ by multiplying the master equation by X and summing over all X . Use Eq. 3.32 to show that the mean satisfies the deterministic rate equation. This is a general characteristic of master equations with linear transition rates.
10. **Chemical reaction networks:** The time-evolution of species in a chemical reaction network is often represented by a Markov process, characterized by the master equation.
- (a) Write out the master equation for the simple linear network (3.26). Try to do the same without using the explicit formula provided.
 - (b) A “toggle switch” network consists of two mutually repressing species, r_1 and r_2 – If r_1 is high, synthesis of r_2 is low, and, conversely, if r_2 is high, r_1 is kept low. A simple network describing this system is the following four reactions:



where the function $g_R(x)$ is high if x is low and low if x is high. Write out the stoichiometry matrix $\mathbf{S}$ and the propensity vector $\boldsymbol{\nu}$ for the toggle switch. Write out the deterministic reaction rate equations and the master equation.

CHAPTER 4

SOLUTION OF THE MASTER EQUATION

The master equation derived in Chapter 3 provides a foundation for most applications of stochastic processes. Although it is more tractable than the Chapman-Kolmogorov equation, it is still rare to find an exact solution. For linear transition rates, the master equation can be transformed to a first-order partial differential equation from which it is sometimes possible to extract an exact solution. More often, the transition rates are nonlinear, and approximation methods are the only recourse. We shall explore two popular and powerful approximation methods: numerical simulation algorithms (Section 4.2.1) and a perturbation expansion called the linear noise approximation (Chapter 5).

4.1 Exact Solutions

There are very few general methods for solving the master equation. We shall discuss two of these in detail. The first is called *the moment generating function* and is used to transform the master equation into a linear partial differential equation. The method only works if the transition probabilities are *linear* in the state variables. As such, it is restricted to use on rather artificial and uninteresting examples. Furthermore, if the dimensionality of the system is high, the algebra is formidable and the auxiliary partial differential equation may not be amenable to solution either. The second method we shall discuss relies upon re-writing the master equation in vector-matrix notation. The steady-state solution

can then be computed as the eigenvector of a given transfer matrix. The method is semi-exact since the eigenvector is usually computed by numerical matrix iteration. Furthermore, the method becomes difficult to implement if the dimensionality of the system is large or the range of the state space is unbounded. Nevertheless, for one dimensional systems, even with nonlinear transition probabilities, the method can be very useful.

4.1.1 Moment Generating Functions

The moment generating function $Q(z, t)$, associated with the probability $P(n, t)$, is a discrete version of the Laplace transform,

$$Q(z, t) = \sum_{n=0}^{\infty} z^n P(n, t), \quad (4.1)$$

very similar to the z -transform used by electrical engineers¹. The moment generating function is so-named because the moments of $P(n, t)$ are generated by subsequent *derivatives* of $Q(z, t)$,

$$Q(1, t) = 1, \quad [\text{Normalization condition on } P(n, t)] \quad (4.2)$$

$$\left. \frac{\partial Q(z, t)}{\partial z} \right|_{z=1} = \sum_{n=0}^{\infty} n z^{n-1} P(n, t) \Big|_{z=1} = \langle n(t) \rangle, \quad (4.3)$$

$$\left. \frac{\partial^2 Q(z, t)}{\partial z^2} \right|_{z=1} = \sum_{n=0}^{\infty} n(n-1) z^{n-2} P(n, t) \Big|_{z=1} = \langle n^2(t) \rangle - \langle n(t) \rangle, \dots \quad (4.4)$$

Multiplying both sides of the master equation by z^n and summing over all n allows the discrete-differential master equation for $P(n, t)$ to be transformed into a partial differential equation for $Q(z, t)$. A simple example should clarify the procedure.

Example – Poisson process. Consider a simple birth-death process, with constant birth rate $g_n = \alpha$ and linear death rate $r_n = \beta \times n$. Call $P(n, t)$ the probability of finding the system in state n at time t (conditioned, as always, by some initial distribution). The master equation

¹The z -transform is usually defined over an unbounded domain, with z^n appearing in the denominator, *i.e.* for the sequence $\{x_k\}_{k=-\infty}^{\infty}$, the z -transform $X(z)$ is defined as $X(z) = \sum_{n=-\infty}^{\infty} \frac{x_n}{z^n}$. See Chapter 3 of G. James “Advanced modern engineering mathematics (3rd Ed.),” (Prentice-Hall, 2004), and Section B.3.3 on p. 264.

corresponding to this process is,

$$\frac{dP(n, t)}{dt} = [\alpha P(n-1, t) + \beta \underset{GAIN}{(n+1) P(n+1, t)}] - [\beta \underset{LOSS}{n P(n, t)} + \alpha P(n, t)], \quad (4.5)$$

or, rearranging slightly,

$$\frac{dP(n, t)}{dt} = \alpha [P(n-1, t) - P(n, t)] + \beta [(n+1) P(n+1, t) - n P(n, t)]. \quad (4.6)$$

A particularly elegant short-hand, which shall be useful in the following section on approximation methods, involves the *step-operator* $\mathbf{E}_i^k$ (recall Eq. 3.25 on page 71). The operator acts by finding the i^{th} entry of $\mathbf{n}$ and incrementing it by an integer k :

$$\mathbf{E}_i^k f(\dots, n_i, \dots) = f(\dots, n_i + k, \dots). \quad (4.7)$$

Using the step-operator, Eq. 4.6 is written,

$$\frac{dP(n, t)}{dt} = \alpha [\mathbf{E}_1^{-1} - 1] P(n, t) + \beta [\mathbf{E}_1^1 - 1] (n P(n, t)). \quad (4.8)$$

Multiplying by z^n ,

$$\frac{dz^n P(n, t)}{dt} = \alpha [\mathbf{E}_1^{-1} z - 1] z^n P(n, t) + \beta [\mathbf{E}_1^1 - z] (z^{n-1} n P(n, t)). \quad (4.9)$$

Under the transformation Eq. 4.1, Eq. 4.9 becomes a simple partial differential equation for $Q(z, t)$,

$$\frac{\partial Q(z, t)}{\partial t} = \alpha (z - 1) Q(z, t) - \beta (z - 1) \frac{\partial Q(z, t)}{\partial z}. \quad (4.10)$$

We have transformed a discrete-differential equation (difficult) into a linear first-order partial differential equation (easier). The full time-dependent solution $Q(z, t)$ can be determined using what is called the *method of characteristics*. Instead of the full time-dependent distribution, we shall focus upon the first two moments in steady-state. Solving the transformed Eq. 4.10 at steady-state $\frac{\partial Q}{\partial t} = 0$, we have,

$$Q^s(z) = \exp \left[\frac{\alpha}{\beta} (z - 1) \right]. \quad (4.11)$$

The steady-state moments follow immediately,

$$\left. \frac{\partial Q^s}{\partial z} \right|_{z=1} = \frac{\alpha}{\beta} = \langle n \rangle^s \quad (4.12)$$

$$\left. \frac{\partial^2 Q^s}{\partial z^2} \right|_{z=1} = \left(\frac{\alpha}{\beta} \right)^2 = \langle n^2 \rangle^s - \langle n \rangle^s, \quad (4.13)$$

with steady-state variance,

$$\sigma^2 = \frac{\alpha}{\beta}. \quad (4.14)$$

Having mean equal to the variance is the footprint of a *Poisson process*. In fact, from the definition of the moment generating function (Eq. 4.1), expanding $Q^s(z)$ as an infinite series,

$$Q^s(z) = \exp \left[\frac{\alpha_m}{\beta_m} (z - 1) \right] = \exp \left[-\frac{\alpha_m}{\beta_m} \right] \sum_n z^n \frac{(\alpha_m/\beta_m)^n}{n!},$$

we recognize the coefficients of the series as the probability distribution of a Poisson process,

$$p_n = e^{-\mu} \frac{\mu^n}{n!},$$

with $\mu = \alpha_m/\beta_m$. We can measure how close to Poisson distributed a given process is by considering the *Fano factor*, $\frac{\sigma^2}{\langle n \rangle}$. In our case (since our process is Poisson distributed), the Fano factor is 1,

$$\frac{\sigma^2}{\langle n \rangle} = 1. \quad (4.15)$$

Alternately, the fractional deviation $\eta = \sqrt{\frac{\sigma^2}{\langle n \rangle^2}}$ is a *dimensionless* measure of the fluctuations and often provides better physical insight than the Fano factor. In the example above,

$$\eta = \frac{1}{\sqrt{\langle n \rangle}}, \quad (4.16)$$

substantiating the rule-of-thumb that relative fluctuations scale roughly as the *square-root* of the number of reactants (*cf.* Laplace's solution to Bernoulli's urn problem, Eq. 3.20 on page 66). We shall exploit this scaling in Section 5.1, discussing the linear noise approximation.

The moment generating function method will only work if the transition rates are linear in the state variable – in complete analogy with the Laplace transform for ordinary differential equations. A (semi)-exact method that can be used in the case of nonlinear transition rates is matrix iteration; a method we shall discuss in the next section.

4.1.2 Matrix Iteration

Discrete master equations derived for systems with low dimensionality (*i.e.* one or two state variables) can often be usefully re-written in vector-matrix notation, with the probability of finding the system in a (discrete) state $n \in (0, N)$, written p_n , occupying elements in a $1 \times (N + 1)$ vector $\mathbf{p}$, and the transition probabilities occupying the elements of an $(N + 1) \times (N + 1)$ matrix $\mathbb{W}$,

$$\mathbf{p}(t + 1) = \mathbb{W} \cdot \mathbf{p}(t). \quad (4.17)$$

Notice, also, that the continuous time master equation,

$$\frac{\partial p_n}{\partial t} = \sum_{n'} w_{nn'} p_{n'} - w_{n'n} p_n$$

can likewise be expressed in terms of the transition matrix $\mathbb{W}$,

$$\frac{\partial p_n}{\partial t} = \mathbb{W} \cdot p_n$$

where

$$\mathbb{W}_{nn'} = w_{nn'} - \delta_{nn'} \left(\sum_{n''} w_{n''n} \right).$$

Although Eq. 4.17 could be generalized to an unbound state space $n \in (-\infty, \infty)$, it is clearly most useful for events restricted to a bounded domain. The evolution of the probability distribution is defined as an iterative matrix multiplication,

$$\mathbf{p}(t) = \mathbb{W}^t \cdot \mathbf{p}(0). \quad (4.18)$$

The steady-state distribution is then the eigenvector corresponding to the $\lambda = 1$ eigenvalue of the matrix $\mathbb{W}_s$ obtained from multiplying $\mathbb{W}$ by itself an infinite number of times,

$$\mathbb{W}_s = \lim_{t \rightarrow \infty} \mathbb{W}^t. \quad (4.19)$$

That is, the steady-state probability distribution $\mathbf{p}_s$ corresponds to the unit eigenvector of $\mathbb{W}_s$;

$$\mathbb{W}_s \cdot \mathbf{p}_s = \mathbf{p}_s.$$

A theorem by Perron and Frobenius states that if $\mathbb{W}$ is irreducible, then there is a unique stationary distribution $\mathbf{p}_s$. Furthermore, if the system is ‘aperiodic’, that is, systems whose deterministic counterparts have a single asymptotically stable fixed point, then once $\mathbb{W}_s$ is calculated, multiplication with *any* unit vector yields $\mathbf{p}_s$. If $\mathbb{W}$ is a separable transition matrix,

$$\mathbb{W} = \begin{bmatrix} \mathbb{A} & 0 & 0 \\ 0 & \mathbb{B} & 0 \\ 0 & 0 & \ddots \end{bmatrix},$$

different initial conditions converge to different equilibrium distributions (Exercise 1).

For many processes, the transition matrix $\mathbb{W}$ does not depend upon time. Furthermore, for one-step processes, the transition matrix takes the particularly simple tri-diagonal form,

$$\mathbb{W} = \begin{bmatrix} T_0 & T_- & 0 & 0 & \cdots \\ T_+ & T_0 & T_- & 0 & 0 \\ 0 & T_+ & T_0 & T_- & 0 \\ 0 & 0 & T_+ & T_0 & \ddots \\ \vdots & 0 & 0 & \ddots & \ddots \end{bmatrix},$$

where it should be understood that the elements T_0, T_+ and T_- are themselves functions of the row in which they appear. Matrix iteration provides a particularly convenient method to estimate the accuracy of Laplace’s approximate solution of the Bernoulli urn problem (see Section 3.6 on page 64).

Example – Bernoulli’s urn model. The following problem was posed by Bernoulli (see Figure 3.3 on page 64):

There are n white balls and n black balls in two urns, and each urn contains n balls. The balls are moved cyclically, one by one, from one urn to another. What is the probability $z_{x,r}$ that after r cycles, urn A will contain x white balls?

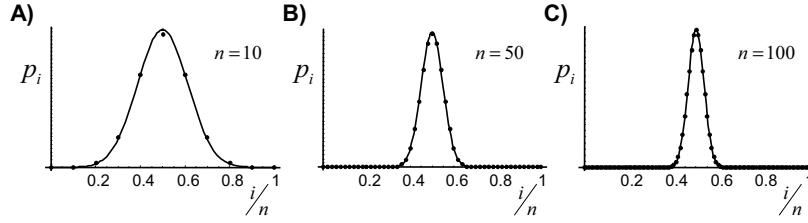


Figure 4.1: **Steady-state distribution for the fraction of white balls in one urn of Bernoulli's urn model.** The eigenvector of the repeated multiplication of the transition matrix is shown as filled circles, with Laplace's approximation (*cf.* Eq. 3.23 on page 66) shown as a solid line, for increasing numbers of balls. **A)** Number of balls in one urn, $n = 10$. **B)** $n = 50$. **C)** $n = 100$. As the size of the transition matrix becomes larger ($n \rightarrow \infty$), repeated multiplication becomes computationally demanding; however, as $n \rightarrow \infty$, Laplace's approximation becomes indistinguishable from the 'exact' numerical solution.

In attempting to solve this problem, Laplace derived the following difference equation for the evolution of $z_{x,r}$,

$$z_{x,r+1} = \left(\frac{x+1}{n}\right)^2 z_{x+1,r} + 2 \frac{x}{n} \left(1 - \frac{x}{n}\right) z_{x,r} + \left(1 - \frac{x-1}{n}\right)^2 z_{x-1,r}. \quad (4.20)$$

Assuming the number of balls is large ($n \rightarrow \infty$) so that changes in the *density* x/n is nearly continuous, along with the ansatz that $x = n/2 + \mu\sqrt{n}$, Laplace arrived at a partial differential equation governing the probability distribution for μ , $U(\mu, r)$,

$$\frac{\partial U(\mu, r')}{\partial r'} = 2 \frac{\partial}{\partial \mu} (\mu U(\mu, r')) + \frac{1}{2} \frac{\partial^2}{\partial \mu^2} \left(\frac{1}{2} U(\mu, r') \right). \quad (4.21)$$

The steady-state solution for the density $\hat{x} = x/n$ is the Gaussian, centered at $\hat{x} = 1/2$ with variance $1/8n$,

$$\rho(\hat{x}) = 2 \sqrt{\frac{n}{\pi}} \exp \left[-4n \left(\hat{x} - \frac{1}{2} \right)^2 \right]. \quad (4.22)$$

Re-writing Laplace's difference equation in vector-matrix notation allows us to use matrix iteration to estimate the error in Laplace's approximation.

Writing the probability $z_{x,r}$ as a vector $\mathbf{z}(r) \in \mathbb{R}^{(n+1)}$, where $z_i(r) \equiv z_{i-1,r}$, the transition rates appearing in the i^{th} row are given by,

$$W_{i,i} = 2\frac{i-1}{n} \left(1 - \frac{i-1}{n}\right), \quad W_{i,i+1} = \left(\frac{i}{n}\right)^2, \quad W_{i,i-1} = \left(1 - \frac{i-2}{n}\right)^2. \quad (4.23)$$

For example, with $n = 4$ balls in each urn, $\mathbb{W}$ is the 5×5 matrix,

$$\mathbb{W} = \begin{bmatrix} 0 & \frac{1}{16} & 0 & 0 & 0 \\ 1 & \frac{3}{8} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{9}{16} & \frac{1}{2} & \frac{9}{16} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{8} & 1 \\ 0 & 0 & 0 & \frac{1}{16} & 0 \end{bmatrix}.$$

The steady-state transition matrix $\mathbb{W}_s$ is given by,

$$\mathbb{W}_s = \frac{1}{70} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 16 & 16 & 16 & 16 & 16 \\ 36 & 36 & 36 & 36 & 36 \\ 16 & 16 & 16 & 16 & 16 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

so that the steady-state probability distribution is $\mathbf{z} = \frac{1}{70}[1 \ 16 \ 36 \ 16 \ 1]$. Continuing in the same fashion for larger n , we see that Laplace's approximation of the distribution is quite accurate even down to $n = 10$ (Figure 4.1).

4.2 Approximation Methods

There are two broad classes of approximation methods - numerical simulation algorithms and perturbation methods. Each has clear advantages and disadvantages.

1. Numerical Simulation: The classic reference for these types of algorithms is the paper by Gillespie: D. T. Gillespie (1977) "Exact simulation of coupled chemical reactions," *Journal of Chemical Physics* **81**: 2340.

The method simulates a *single* trajectory $\mathbf{n}(t)$ that comes from the unknown probability distribution $P(\mathbf{n}, t)$ characterized by the master equation.

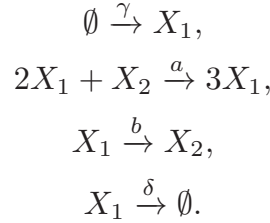
2. Perturbation Methods: The classic reference for the perturbation scheme we will consider is the paper by van Kampen: N. G. van Kampen (1976) “Expansion of the Master equation,” *Advances in Chemical Physics* **34**: 245.

For perturbation methods, the discrete jump in $\mathbf{n}(t)$ that occurs with each reaction is treated as a *nearly* deterministic process. The ‘nearly’ is what we use as a perturbation parameter.

The Brusselator

- S. Strogatz, *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry and Engineering* (Perseus Books Group, 2001).

We shall develop both the stochastic simulation algorithm and the linear noise approximation in the context of a specific example – the Brusselator. The Brusselator is a often-used pedagogical example of a system exhibiting limit cycle behaviour over a range of parameter values. The model is a two-species chemical network described by the following reactions,



Without loss of generality, the time and volume are scaled so that $\gamma = \delta = 1$ to give the deterministic rate equations,

$$\begin{aligned}\frac{dX_1}{dt} &= 1 + aX_1^2X_2 - (b+1)X_1, \\ \frac{dX_2}{dt} &= -aX_1^2X_2 + bX_1,\end{aligned}\tag{4.24}$$

and consequently, the re-scaled master equation governing the probability density $P(n_1, n_2, t)$,

$$\begin{aligned}\frac{\partial P}{\partial t} &= \Omega \left(\mathbf{E}_1^{-1} - 1 \right) P + \frac{a}{\Omega^2} \left(\mathbf{E}_1^{-1} \mathbf{E}_2^1 - 1 \right) n_1 (n_1 - 1) n_2 P + \\ &\quad + \left(\mathbf{E}_1^1 - 1 \right) n_1 P + b \left(\mathbf{E}_1^1 \mathbf{E}_2^{-1} - 1 \right) n_1 P.\end{aligned}\tag{4.25}$$

In this form, the deterministic rate equations admit a single stable equilibrium point $(X_1, X_2) = (1, \frac{b}{a})$ for all choices of parameters with $b < 1 + a$.

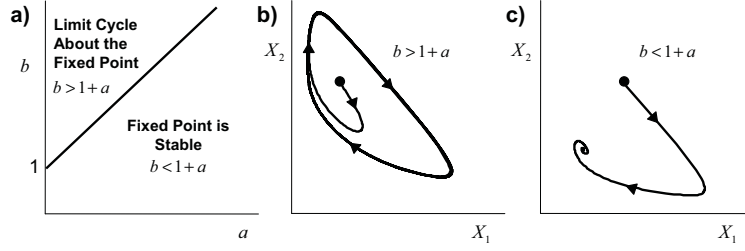


Figure 4.2: **The stability of the Brusselator.** **a)** With suitably chosen units, there are two parameters in the Brusselator model. The stability of the trajectory in phase space is determined by the relationship between a and b in parameter space. **b)** For $b > 1 + a$, a trajectory will move away from the fixed point, and orbit on a closed limit cycle. **c)** For $b < 1 + a$, a trajectory will move toward the stable fixed point.

Along the critical line $b = 1 + a$, the system undergoes a *Hopf* bifurcation, and the steady-state is a stable limit cycle (Figure 4.2). Notice that the transition rates are *nonlinear* in $\mathbf{n}$ – as a consequence, there is no general solution method available to determine $P(\mathbf{n}, t)$. Nevertheless, it is precisely the nonlinearity of the transition rates that makes this model interesting, and therefore the Brusselator is a terrific example to showcase numerical and analytic approximation methods.

We shall express both the stochastic simulation algorithm of Gillespie and the linear noise approximation of van Kampen in terms of the propensity vector $\boldsymbol{\nu}$ and the stoichiometry matrix $\mathbf{S}$ introduced in Chapter 3 (Section 3.7 on page 71). From the reaction network shown above, with $\gamma = \delta = 1$, the propensity vector and stoichiometry matrix are given by,

$$\boldsymbol{\nu} = \begin{bmatrix} 1 \\ a \cdot \frac{n_1 \cdot (n_1 - 1) \cdot n_2}{\Omega^3} \\ b \cdot \frac{n_1}{\Omega} \\ \frac{n_1}{\Omega} \end{bmatrix}, \quad (4.26)$$

and,

$$\mathbf{S} = \begin{bmatrix} \nu_1 & \nu_2 & \nu_3 & \nu_4 \\ 1 & 1 & -1 & -1 \\ 0 & -1 & 1 & 0 \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix} \quad (4.27)$$

4.2.1 Numerical Methods – Gillespie’s Algorithm

For master equations with nonlinear transition probabilities, the full distribution $P(\mathbf{n}, t)$ can rarely be solved exactly. Gillespie’s algorithm is a method by which an individual sample path, starting at a given initial point, can be simulated in time such that it conforms to the unknown probability distribution we seek; that is, for a sufficiently large population of sample paths, the inferred probability distribution is as near to the exact solution as we wish (analogous with Langevin’s modeling of Brownian motion). The algorithm proceeds in 3 steps:

1. The propensities ν_j are used to generate a probability distribution for the next reaction *time*, τ and τ is drawn from this distribution.
2. The propensities are used to generate a probability distribution for *which reaction in the network will occur next*, *i.e.* which of the ν_j ’s is completed at time $t + \tau$. Call the reaction index μ .
3. The time is advanced $t \rightarrow t + \tau$ and the state is updated using the stoichiometry matrix - for each reactant, $n_i \rightarrow n_i + S_{i\mu}$. Repeat ...

In this way, we generate a discrete time series for the reactant numbers. We shall go over each of these steps in greater detail (see Figure 4.3), and it will be clear that in developing his algorithm, Gillespie built upon some fundamental properties of stochastic processes. It is also instructive to see Gillespie’s algorithm in action, so we generate some stochastic simulations of the Brusselator model introduced on page 88.

Details of the stochastic simulation algorithm (Figure 4.3)

- D. T. Gillespie (1977) “Exact simulation of coupled chemical reactions,” *Journal of Chemical Physics* **81**: 2340.
- *Markov Processes*, D. T. Gillespie (Academic Press, Inc, 1992).

To advance the state with each reaction event, we need two random variables – the time for the completion of the next reaction τ , and the index of the reaction that fires μ . It is Gillespie’s deep insight that allows us to determine the probability distribution for each of these, and in fact how to generate the pair (τ, μ) using a unit uniform random number generator.

Ultimately, we are after the conditional probability distribution,

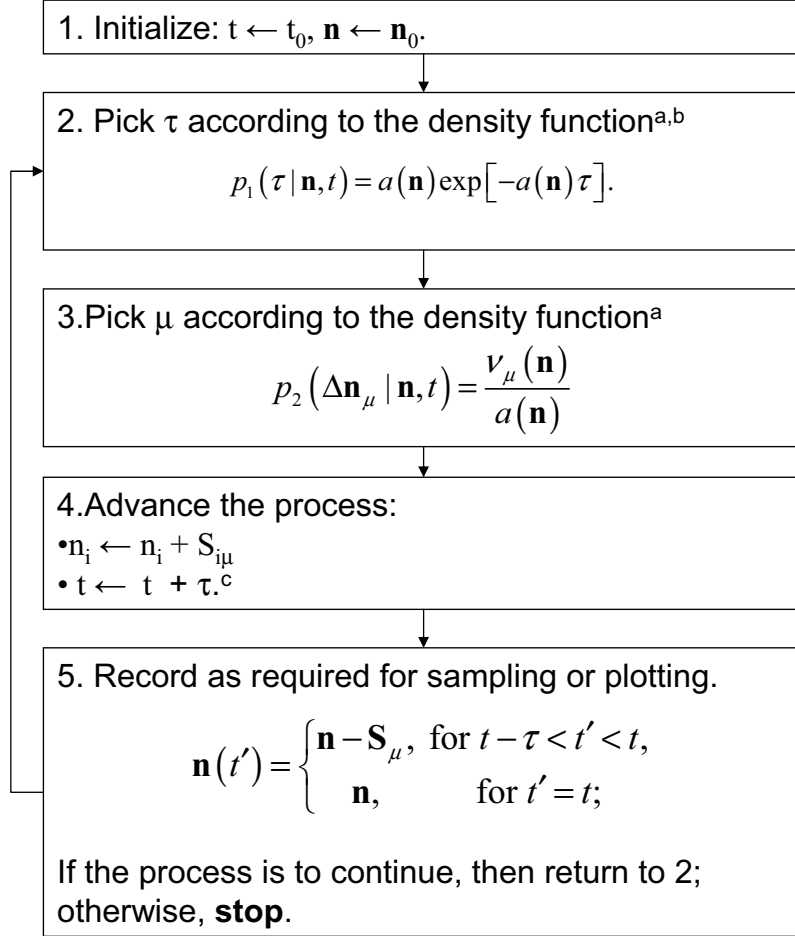


Figure 4.3: **Gillespie's algorithm for stochastic simulation of the master equation.** **a)** Use the inversion method for generating the random number τ (see Section A.4). If $a(\mathbf{n}, t) \equiv a(\mathbf{n})$, then the inversion is easy: Draw a unit uniform random number r_1 , and take $\tau = [1/a(\mathbf{n})] \ln(1/r_1)$. **b)** Draw a unit uniform random number r_2 and take μ to be the *smallest* integer for which the sum over $\nu_j(\mathbf{n})/a(\mathbf{n})$ from $j = 1$ to $j = \mu$ exceeds r_2 . Notice the jump in the state $\Delta \mathbf{n}_\mu = \mathbf{S}_\mu$, where $\mathbf{S}_\mu$ is the μ^{th} column of the stoichiometry matrix. **c)** In a simulation run containing $\sim 10^K$ jumps, the sum $t + \tau$ should be computed with at least $K + 3$ digits of precision. Taken from Gillespie (1992), p. 331.

$p(\mathbf{n} + \Delta\mathbf{n}, t + \tau | \mathbf{n}, t) d\tau$: the probability that, given the system is in state $\mathbf{n}$ at time t , the next jump occurs between $t + \tau$ and $t + \tau + d\tau$, carrying the state from $\mathbf{n}$ to $\mathbf{n} + \Delta\mathbf{n}$,

from which we draw the random numbers $(\tau, \Delta\mathbf{n})$ to advance the system. In practice, because the stoichiometry matrix $\mathbf{S}$ records how the state advances with each reaction event, it is sufficient that we simply generate the random variable specifying which reaction has fired, (τ, μ) .

Following Gillespie, we introduce the probability $q(\mathbf{n}, t; \tau)$ that the system in state $\mathbf{n}(t)$ *will jump* at some instant between t and $t + \tau$. From the microscopic transition rates $\nu(\mathbf{n})$, we know that over an *infinitesimal* interval dt ,

$$\begin{aligned} q(\mathbf{n}, t; dt) &\equiv \\ &[\text{Prob. Reaction 1 Occurs} + \text{Prob. Reaction 2 Occurs} + \dots] \\ &= [\nu_1(\mathbf{n})dt + \nu_2(\mathbf{n})dt \dots] = \left[\sum_{j=1}^N \nu_j(\mathbf{n}) \right] dt \equiv a(\mathbf{n}) dt \end{aligned} \quad (4.28)$$

Over an infinitesimal interval, at most one jump can occur, so we have that the probability that *no* jump occurs, $q^*(\mathbf{n}, t; d\tau)$, is $q^*(\mathbf{n}, t; d\tau) = 1 - q(\mathbf{n}, t; d\tau)$. Over a non-infinitesimal interval, the probability that *no* jump occurs is,

$$q^*(\mathbf{n}, t; \tau) = \exp[-a(\mathbf{n})\tau], \quad (4.29)$$

(Exercise 6). In that way, the probability $p(\mathbf{n} + \Delta\mathbf{n}, t + \tau | \mathbf{n}, t) d\tau$ is written,

$$\begin{aligned} p(\mathbf{n} + \Delta\mathbf{n}, t + \tau | \mathbf{n}, t) &= \\ &\underbrace{q^*(\mathbf{n}, t; \tau)}_{\text{Probability the state will NOT jump during } [t, t + \tau]} \times \underbrace{a(\mathbf{n}) d\tau}_{\text{Probability the state WILL jump in } [t + \tau, t + \tau + d\tau]} \times \underbrace{w(\Delta\mathbf{n} | \mathbf{n}, t + \tau)}_{\text{Probability that, given the state jumps at } t + \tau, \text{ it will land in } \mathbf{n} + \Delta\mathbf{n}} \end{aligned}$$

The first two terms on the right-hand side determine the next reaction *time* τ , while the last term determines the next reaction *index*. Therefore, we can factor the conditional probability into two parts $p_1(\tau)$ and $p_2(\Delta\mathbf{n})$,

$$p(\mathbf{n} + \Delta\mathbf{n}, t + \tau | \mathbf{n}, t) = \underbrace{a(\mathbf{n}) \exp[-a(\mathbf{n})\tau] d\tau}_{p_1(\tau | \mathbf{n}, t)} \times \underbrace{w(\Delta\mathbf{n} | \mathbf{n}, t + \tau)}_{p_2(\Delta\mathbf{n} | \mathbf{n}, t + \tau)}.$$

Probability $p_1(\tau)$ is exponentially distributed, consequently a unit uniform random number r_1 can be used to simulate τ via the inversion (see Section A.4),

$$\tau = \frac{1}{a(n)} \ln(1/r_1). \quad (4.30)$$

The index of the next reaction μ is a little more subtle – Notice from Eq. 4.28 that the probability of reaction μ to occur is proportional to the rate of reaction ν_μ . The normalization condition ensures that the probability it is reaction μ that has caused the state to jump is,

$$p_2(\Delta \mathbf{n}_\mu) = \frac{\nu_\mu(\mathbf{n})}{\sum_j \nu_j(\mathbf{n})} \equiv \frac{\nu_\mu(\mathbf{n})}{a(\mathbf{n})}, \quad (4.31)$$

where $\Delta \mathbf{n}_\mu$ is the μ^{th} column of the stoichiometry matrix $\mathbf{S}$. The next reaction index μ is simulated using a unit uniform random number r_2 via the integer inversion method (see Exercise 4b on p. 250). That is, the index μ drawn from p_2 is the first integer for which

$$\frac{1}{a(\mathbf{n})} \sum_{j=1}^{\mu} \nu_j(\mathbf{n}) > r_2. \quad (4.32)$$

With (τ, μ) , the system is updated,

$$\begin{aligned} t &\rightarrow t + \tau, \\ n_i &\rightarrow n_i + S_{i\mu}, \end{aligned}$$

and the algorithm is repeated as long as desired, each time drawing two random numbers (r_1, r_2) from a unit uniform random number generator. An implementation of Gillespie's algorithm coded in Matlab is annotated in the Appendix (see Section D.1.2 on p. 276).

Stochastic simulation of the Brusselator model

We shall produce stochastic simulations of the Brusselator model in both the stable and limit-cycle regimes of the model parameter space (see Section D.1.3 on p. 278 for example Matlab code). We shall find that both regimes are equally accessible to the stochastic simulation algorithm. In contrast, the linear noise approximation described in the next section requires a major modification to treat fluctuations around the limit cycle. On the other hand, it is very difficult to describe the simulation results

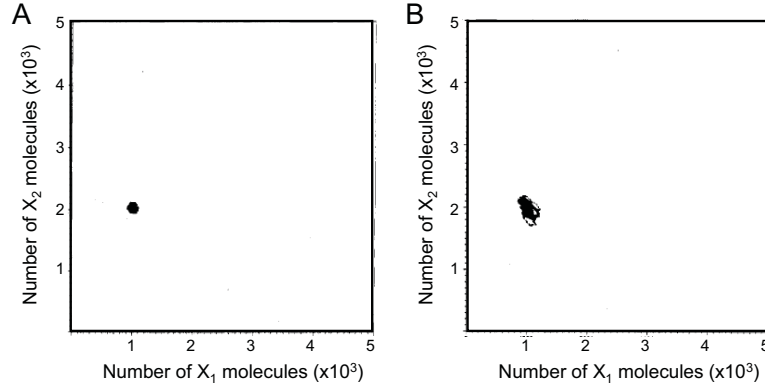


Figure 4.4: **Gillespie’s stochastic simulation algorithm - The Brusselator in the stable regime.** **a)** $(a, b) = (0.1, 0.2)$, ($\alpha = 10$ in Figure 23 of Gillespie (1977)). The system is very stable. **b)** $(a, b) = (0.5, 1)$, ($\alpha = 2$ in Figure 22 of Gillespie (1977)). The system is still stable, but fluctuations are more appreciable.

any more than qualitatively, while the perturbation methods provide a connection between system behaviour and model parameters.

Stable regime, $b < 1 + a$: In the stable regime of the model parameter space, the system admits a single, stable equilibrium point - In Figure 4.4, that equilibrium is $(x_1, x_2) = (1000, 2000)$. The number of reactant molecules is large, so the intrinsic fluctuations are correspondingly small. As the parameters get closer to the Hopf bifurcation, the fluctuations become somewhat larger (Figure 4.4b). Neither plot illustrates particularly interesting behaviour.

Limit cycle, $b \geq 1 + a$: At the bifurcation (Figure 4.5), the system parameters are on the threshold of stability and the fluctuations carry the state on long excursions away from the fixed point. From the time-series plot (Figure 4.5a), it appears as if the fluctuations are generating nearly regular oscillations (see Exercise 5 on p. 213). In phase-space (Figure 4.5b), the system seems confined to an elongated ellipse with a negatively-sloped major axis.

Beyond the Hopf bifurcation, the system exhibits regular limit cycle oscillations (Figure 4.6). As Gillespie notes in his original article, the horizontal leg of the cycle seems to travel along grooves of negative slope, rather than straight from right to left. There is some spread along the diagonal leg, but both the horizontal and vertical legs are little influenced by the fluctuations (for analytic insight into why that is, see Section 10.2).

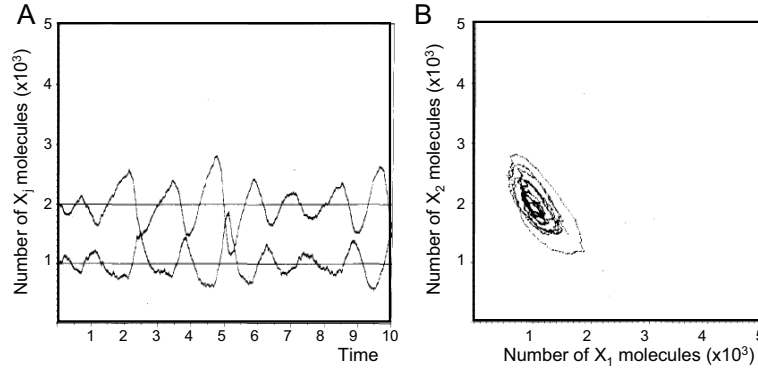


Figure 4.5: **Gillespie's stochastic simulation algorithm - The Brusselator at the Hopf bifurcation.** a) Simulation of the system *at* the Hopf bifurcation, $(a, b) = (1, 2)$ ($\alpha = 1$ in Figure 21 of Gillespie (1977)). The fluctuations generate what appear to be almost regular oscillations. b) In phase-space, the fluctuations are confined to a large ellipse, angled with negative slope indicating the strong cross-correlation between fluctuations in x_1 and x_2 .

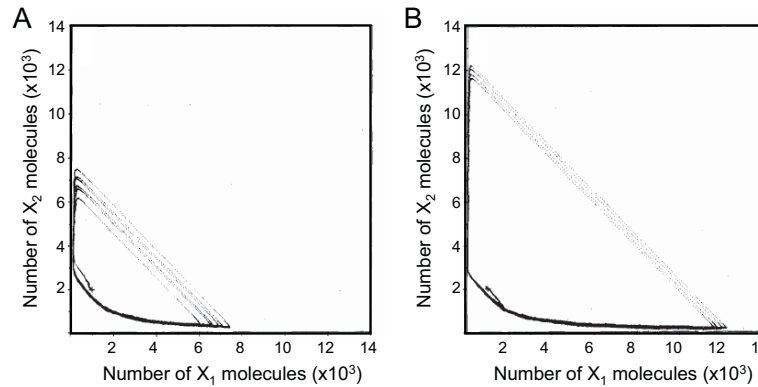


Figure 4.6: **Gillespie's stochastic simulation algorithm - The Brusselator in the unstable regime.** a) $(a, b) = (5, 10)$ ($\alpha = 0.2$ in Figure 19 of Gillespie (1977)). b) $(a, b) = (10, 20)$ ($\alpha = 0.1$ in Figure 18 of Gillespie (1977)).

The advantages of the stochastic simulation algorithm is that it is simple to program and provides an output trajectory that *exactly* conforms to the solution distribution of the master equation. The disadvantages are that the original algorithm is computationally expensive and the method

does not scale well as the number of molecules gets large (although there are approximate algorithms that alleviate some of the computational burden). Most importantly, the method suffers from the same limitations as any numerical scheme – there is a lack of deep insight into the model and it is difficult to systematically explore different regions of parameter space. Nevertheless, Gillespie’s algorithm is the benchmark against which all other methods of solving the master equation are measured.

Suggested References

For numerical simulation methods, the textbook by Gillespie is unsurpassed, along with his seminal article on the stochastic simulation algorithm,

- D. T. Gillespie, *Markov processes: An introduction for physical scientists* (Academic Press, 1992).
- D. T. Gillespie (1977) “Exact simulation of coupled chemical reactions,” *Journal of Chemical Physics* **81**: 2340.

Exercises

1. **Separable transition matrices:** Write out the most general 2×2 transition matrix $\mathbb{W}$. Show that for this case, any initial condition $\mathbf{p}(0)$ converges to a well-defined steady-state distribution $\mathbf{p}^s$, i.e. $\lim_{t \rightarrow \infty} \mathbb{W}^t \cdot \mathbf{p}(0) = \mathbf{p}^s$, with two exceptions.
2. **Time-dependent reaction rates:** A chemical reaction is described by the following deterministic rate equation,

$$\frac{dA}{dt} = k(t)A(t),$$

where $k(t)$ is a time-dependent reaction rate. Solve the associated chemical master equation using a moment generating function.

3. **Bursty Poisson model:** Consider the following generalization of the Poisson model: synthesis occurs with a constant rate α , but in ‘bursts’ of size b , and degradation is linear, with rate βn . The

master equation corresponding to this process is,

$$\frac{dP(n, t)}{dt} = \alpha [P(n - b, t) - P(n, t)] + \beta [(n + 1) P(n + 1, t) - n P(n, t)].$$

- (a) For $b = 1$, solve the characteristic function $Q(z, t)$ for all time. What do you notice about the distribution if $n_0 = 0$?
 - (b) Repeat 3a, but for arbitrary burst size.
4. **Bernoulli's urn model:** In describing Bernoulli's urn model, Laplace derived a difference equation with nonlinear transition rates (Eq. 4.20). The nonlinearity of the transition rates make the equation difficult to solve even at steady-state.
- (a) Using matrix iteration for the cases $n = 2, 3$, and 4, multiply the steady-state probability distribution $\mathbf{z}$ by its first element z_1 to get a vector of integers. Can you spot a pattern in the individual entries? Postulate a general solution for the steady-state probability distribution and verify that it satisfies the difference equation.
 - (b) Using Laplace's approximate solution (Eq. 4.22), calculate the mean-squared error between the exact solution and the continuous approximation as a function of the number of balls.
5. **Asymmetric cell division:** Most bacteria divide with surprising symmetry – *E. coli* for example, typically divides into daughter cells that differ in length by less than 5%. Suppose a bacterium divided *unequally*, what would the age distribution (*i.e.*, the time to next division) look like for a population?
- (a) Divide the lifetime of the cell into time-steps Δt – Assume the larger daughter lives $10\Delta t$ before division, while the smaller daughter lives $15\Delta t$. Scale the time so that the transition rates are 1. Write out the transition matrix $\mathbb{W}$.
 - (b) Find $\mathbb{W}^s$. What does the equilibrium lifetime distribution look like?
 - (c) Repeat the above, but on a finer scale. That is, assume the large daughter lives $100\Delta t$ and the small daughter lives $150\Delta t$.
 - (d) Derive a deterministic model for the process. How do the results above compare?

- (e) Suppose the size after division was itself a stochastic process. How would the transition matrix change?

6. **Gillespie's simulation algorithm:** Gillespie draws upon two very important ideas in stochastic processes – the evolution of probability for a Markov process, and the simulation of a random variable by a unit uniform random number.

- (a) Making use of the Markov property, show that for time independent reaction events $\nu(\mathbf{n}, t) = \nu(\mathbf{n})$, integration of the probability $q^*(\mathbf{n}, t; dt) \equiv 1 - q(\mathbf{n}, t; dt)$ gives

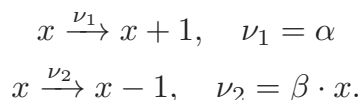
$$q^*(\mathbf{n}, t; \tau) = \exp[-a(\mathbf{n})\tau],$$

as quoted in the main text (Eq. 4.29).

- (b) Repeat part 6a for *time dependent* reaction events $\nu(\mathbf{n}, t)$.
- (c) Write a stochastic simulation algorithm to generate realizations of the stochastic process that describes the Brusselator in a growing cell. That is, repeat the example in the text, but with Ω a time-dependent quantity. In bacterial cell growth, the volume grows approximately exponentially over a cell-cycle, then divides more or less symmetrically.
- i. As a first approximation, assume perfect division of the cell volume and perfect partitioning of the cell contents into daughter cells.
 - ii. Code a routine to allow both the volume after division and the partitioned contents to be narrowly-peaked random variables. What distribution will you choose for these two variables?

7. **Exploring stochastic dynamics:** Stochastic models exhibit features that do not appear in their deterministic counter-parts. Some of these features are straight-forward, others are quite surprising.

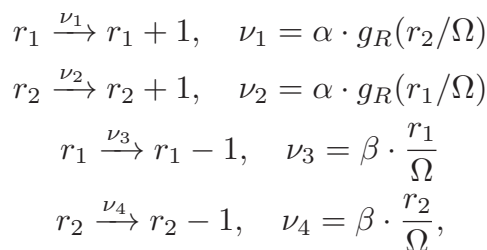
- (a) *Stable system.* Consider a very simple system with constant synthesis and linear degradation,



Starting with $x(0) = 0$, compute the deterministic trajectory for $x(t)$. Generate stochastic simulation data for molecule

numbers from about $10 - 10^3$, and plot these runs normalized to the same steady-state value. How does the relative magnitude of the fluctuations scale with the number of molecules? Plot a histogram of the fluctuations around the steady-state – what distribution does it resemble? How does the half-width scale with the number of molecules?

- (b) *Multi-stable system.* A “toggle switch” network consists of two mutually repressing species, r_1 and r_2 – If r_1 is high, synthesis of r_2 is low, and, conversely, if r_2 is high, r_1 is kept low. A simple network describing this system is the following four reactions:



where the function $g_R(x)$ is high if x is low and low if x is high. Suppose $g_R(x)$ takes the simple Hill-form,

$$g_R(x) = \frac{1 + f \cdot \left(\frac{x}{K_R}\right)^n}{1 + \left(\frac{x}{K_R}\right)^n},$$

where f is the *capacity*, K_R measures the *repressor strength* (smaller the K_R , the less repressor necessary to reduce g_R), and n is the *cooperativity* determining how abrupt the transition is from the high to low state. Nondimensionalize the deterministic rate equations corresponding to this system, and estimate a range of parameters for which the system exhibits bistability. Perform stochastic simulations of the model for parameters in the bistable regime, along with varying system size Ω . What differences do you see comparing the stochastic and deterministic models?

- (c) *Noise-induced oscillations.* Read J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler (2002) Mechanisms of noise-resistance in genetic oscillators, *Proceedings of the National Academy of Sciences USA* **99**: 15988–15992. Repeat the simulation of their model for the parameter choice leading to a deterministically stable system.

CHAPTER 5

PERTURBATION EXPANSION OF THE MASTER EQUATION

The master equation derived in Chapter 3 provides a foundation for most applications of stochastic processes to physical phenomena. Although it is more tractable than the Chapman-Kolmogorov equation, it is still rare to find an exact solution. One possibility is to adopt the Fokker-Planck equation as an approximate evolution equation, as in Chapter 6. We shall show in this Chapter that this is the first-step in a systematic analytic approximation scheme.

5.1 Linear Noise Approximation (LNA)

- N. G. van Kampen (1976) “Expansion of the Master equation,” *Advances in Chemical Physics* **34**: 245.

Often we can gain a better sense of a particular model by examining certain limiting regimes. The approximation method that we describe in this section examines system behavior in the limit of large numbers of reactant molecules.

We have already seen that as the number of molecules increases, the system evolution becomes more smooth and the deterministic formulation becomes more appropriate (Figure 3.5). The linear noise approximation exploits this behavior and rests upon the supposition that the deterministic evolution of the reactant concentrations, call them $\mathbf{x}$, can be mean-

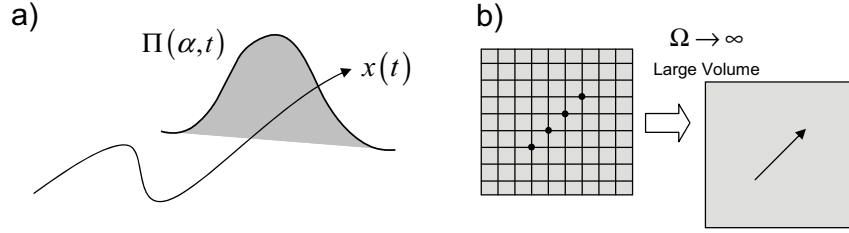


Figure 5.1: **The Linear Noise Approximation.** **a)** The microscopic fluctuations are separated from the macroscopic evolution of the system by re-writing the probability density for the whole state $P(\mathbf{n}, t)$ as a distribution for the fluctuations $\Pi(\alpha, t)$ centered on the macroscopic trajectory $\mathbf{x}(t)$. **b)** The discrete state space is smeared into a continuum by replacing the discrete step-operator $\mathbf{E}$ by a continuous *differential* operator.

ingfully separated from the fluctuations, call them α , and that these fluctuations scale roughly as the *square-root* of the number of molecules. We introduce an extensive parameter Ω that carries the units of volume and is directly proportional to the molecule numbers, allowing the molecule numbers to be written

$$n_i = \Omega x_i + \sqrt{\Omega} \alpha_i. \quad (5.1)$$

We are lead to the square-root scaling of the fluctuations by the suggestion from Poisson statistics (see Eq. 4.16 on page 83). Recall that for a Poisson process, the fractional deviation is inversely proportional to the square-root of the number of molecules (see Eq. 4.16). The picture that underlies Eq. 5.1 is that of a deterministic, reproducible trajectory surrounded by a cloud of fluctuations. We would like a set of equations that govern the change in the deterministic part $\mathbf{x}$ and an equation that governs the change in the probability distribution of the fluctuations, call it $\Pi(\alpha, t)$, centered upon $\mathbf{x}$ (Figure 5.1a).

With the propensity vector $\boldsymbol{\nu}$ and the stoichiometry matrix $\mathbf{S}$, we are in a position to write the master equation in a compact, and convenient manner: for a network of R reactions involving N species the master equation is,

$$\frac{dP(\mathbf{n}, t)}{dt} = \Omega \sum_{j=1}^R \left[\left(\prod_{i=1}^N \mathbf{E}_i^{-S_{ij}} \right) - 1 \right] \nu_j(\mathbf{n}, \Omega) P(\mathbf{n}, t), \quad (5.2)$$

(where we have used the step-operator $\mathbf{E}_i^k$ defined in Eq. 4.7). We have repeatedly emphasized that if the transition probabilities ν_j are *nonlinear* functions, then there is no systematic way to obtain an exact solution of the master equation, and we must resort to approximation methods. The linear noise approximation, which is the subject of this section, proceeds in three steps.

1. First, we replace the full probability distribution $P(\mathbf{n}, t)$ by the probability distribution for the fluctuations $\Pi(\alpha, t)$ centered on the macroscopic trajectory $\mathbf{x}$,

$$P(\mathbf{n}, t) \mapsto \Omega^{-\frac{N}{2}} \Pi(\alpha, t). \quad (5.3)$$

The pre-factor $\Omega^{-\frac{N}{2}}$ comes from the normalization of the probability distribution.

2. Recall that what makes the master equation difficult to solve exactly is the discrete evolution over state-space characterized by the step-operator $\mathbf{E}_i^k$. To make headway, we must find some *continuous* representation of the action of the operator. To that end, consider the action of the operator - it increments the i^{th} species by an *integer* k . Using the assumption above (Eq. 5.1), we write,

$$\begin{aligned} \mathbf{E}_i^k f(\dots, n_i, \dots) &= f(\dots, n_i + k, \dots) \\ &= f(\dots, \Omega x_i + \sqrt{\Omega} \left(\alpha_i + \frac{k}{\sqrt{\Omega}} \right), \dots). \end{aligned} \quad (5.4)$$

The term $\frac{k}{\sqrt{\Omega}}$ becomes negligibly small as $\Omega \rightarrow \infty$, suggesting a Taylor series around $\frac{k}{\sqrt{\Omega}} = 0$,

$$\begin{aligned} f(\dots, \Omega x_i + \sqrt{\Omega} \left(\alpha_i + \frac{k}{\sqrt{\Omega}} \right), \dots) &\approx \\ f(\dots, n_i, \dots) + \frac{k}{\sqrt{\Omega}} \frac{\partial f}{\partial \alpha_i} + \frac{k^2}{2 \Omega} \frac{\partial^2 f}{\partial \alpha_i^2} + \dots \end{aligned} \quad (5.5)$$

allowing us to approximate the *discrete* step-operator by a *continuous* differential operator (Figure 5.1b),

$$\mathbf{E}_i^k \approx \left[1 + \frac{k}{\sqrt{\Omega}} \frac{\partial}{\partial \alpha_i} + \frac{k^2}{2 \Omega} \frac{\partial^2}{\partial \alpha_i^2} + \dots \right]. \quad (5.6)$$

(Compare this with the *Kramers-Moyal expansion*, Eq. 6.11 on page 126).

3. Finally, to remain consistent in our perturbation scheme, we must likewise expand the propensities in the limit $\Omega \rightarrow \infty$,

$$\nu_j \left(\frac{\mathbf{n}}{\Omega} \right) \approx \tilde{\nu}_j(\mathbf{x}) + \frac{1}{\sqrt{\Omega}} \sum_{i=1}^N \frac{\partial \tilde{\nu}_j}{\partial x_i} \alpha_i + \dots, \quad (5.7)$$

where $\tilde{\nu}_j(\mathbf{x})$ are the macroscopic propensities defined in the limit that the molecule numbers go to infinity ($\Omega \rightarrow \infty$), while the concentration remains fixed ($\mathbf{n}/\Omega$ constant) (see p. 72),

$$\tilde{\nu}_j(\mathbf{x}) = \lim_{\Omega \rightarrow \infty} \nu_j \left(\frac{\mathbf{n}}{\Omega} \right). \quad (5.8)$$

It is the expansion of the propensities that distinguishes the linear noise approximation from the Kramers-Moyal expansion. The result is a consistent approximation scheme with a particularly simple Fokker-Planck equation governing the probability distribution of the fluctuations, as we show below.

Putting all of this together, taking care to write $\frac{\partial \Pi}{\partial t}$ using the chain rule¹,

$$\frac{\partial P}{\partial t} = \Omega^{\frac{1-N}{2}} \left[\Omega^{-\frac{1}{2}} \frac{\partial \Pi}{\partial t} - \sum_{i=1}^N \frac{dx_i}{dt} \frac{\partial \Pi}{\partial \alpha_i} \right], \quad (5.9)$$

we collect Eq. 5.2 in like powers of $\sqrt{\Omega}$ taking the limit $\Omega \rightarrow \infty$. To zero'th order (Ω^0), we have,

$$\Omega^0 : \quad \frac{dx_i}{dt} \frac{\partial \Pi}{\partial \alpha_i} = [\mathbf{S} \cdot \boldsymbol{\nu}]_i \frac{\partial \Pi}{\partial \alpha_i}. \quad (5.10)$$

This system of equations is identically satisfied if $\mathbf{x}$ obeys the deterministic rate equations,

$$\frac{dx_i}{dt} = [\mathbf{S} \cdot \boldsymbol{\nu}]_i \equiv f_i(\mathbf{x}). \quad (5.11)$$

At the next order, $\sqrt{\Omega}^{-1}$, we have the equation characterizing the probability distribution for the fluctuations,

$$\sqrt{\Omega}^{-1} : \quad \frac{\partial \Pi}{\partial t} = - \sum_{i,j} \Gamma_{ij} \partial_i (\alpha_j \Pi) + \frac{1}{2} \sum_{i,j} D_{ij} \partial_{ij} \Pi, \quad (5.12)$$

¹There is a missing step here. The chain rule will give an expression involving the time derivatives of $\boldsymbol{\alpha}$. To obtain the expression involving $\frac{dx_i}{dt}$, note that the time derivative of P is taken with $\mathbf{n}$ fixed; i.e., $-\frac{d\mathbf{x}}{dt} = \frac{1}{\sqrt{\Omega}} \frac{d\boldsymbol{\alpha}}{dt}$.

where $\partial_i \equiv \frac{\partial}{\partial \alpha_i}$ and,

$$\Gamma_{ij}(t) = \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathbf{x}(t)} \quad \mathbf{D} = \mathbf{S} \cdot \text{diag}[\nu] \cdot \mathbf{S}^T. \quad (5.13)$$

We now have in hand a system of ordinary differential equations that govern the deterministic evolution of the system, which happen to coincide with the macroscopic reaction rate equations. We also have a partial differential equation that characterizes the probability distribution of the fluctuations. Some comments are in order:

1. The equation describing $\Pi(\alpha, t)$, Eq. 5.12, is a special sub-class of Fokker-Planck equations since the coefficient matrices $\mathbf{\Gamma}$ and $\mathbf{D}$ are *independent* of α . You can prove (see Exercise 1) that for linear drift and constant diffusion coefficient matrices, the solution distribution is *Gaussian* for all time. Furthermore, the moments of α are easily computed from Eq. 5.12 by multiplying with α_i and integrating by parts to give,

$$\frac{d\langle \alpha \rangle}{dt} = \mathbf{\Gamma} \cdot \langle \alpha \rangle \quad (5.14)$$

If we choose the initial condition of $\mathbf{x}$ to coincide with the initial state of the system $\mathbf{n}_0$ (*i.e.* $\langle \alpha(0) \rangle = 0$), then $\langle \alpha \rangle = 0$ for all time. Without loss of generality, then, we set $\langle \alpha \rangle = 0$. For the covariance $\Xi_{ij} = \langle \alpha_i \alpha_j \rangle - \langle \alpha_i \rangle \langle \alpha_j \rangle = \langle \alpha_i \alpha_j \rangle$, multiplication of the Fokker-Planck equation, Eq. 5.12, by $\alpha_i \alpha_j$ and integration by parts gives,

$$\frac{d\Xi}{dt} = \mathbf{\Gamma} \cdot \Xi + \Xi \cdot \mathbf{\Gamma}^T + \mathbf{D}. \quad (5.15)$$

The covariance determines the width of $\Pi(\alpha, t)$ as it moves along $\mathbf{x}(t)$. The full distribution satisfying Eq. 5.12 is the Gaussian,

$$\Pi(\alpha, t) = [(2\pi)^N \det \Xi(t)]^{1/2} \exp \left[-\frac{1}{2} \alpha^T \cdot \Xi^{-1}(t) \cdot \alpha \right], \quad (5.16)$$

with covariance matrix $\Xi(t)$ determined by Eq. 5.15.

2. We have used Ω as an ordering parameter in our perturbation expansion of the master equation. Notice, however, that it is not the single parameter Ω that determines the reliability of the expansion, particularly in higher-dimensional systems. The condition

that must be satisfied is that the standard deviation of the fluctuations is small compared with the mean. After non-dimensionalizing the model equations, a more convenient measure is the requirement that the elements of the diffusion coefficient matrix $\mathbf{D}$ be small.

3. The Fokker-Planck equation provides an even deeper insight into the physics of the process. Notice $\mathbf{\Gamma}$ is simply the *Jacobian* of the deterministic system, evaluated pointwise along the macroscopic trajectory. As such, it represents the *local damping* or dissipation of the fluctuations. The diffusion coefficient matrix $\mathbf{D}$ tells us how much the microscopic system is changing at each point along the trajectory. As such, it represents the *local fluctuations*. The balance of these two competing effects - dissipation and fluctuation - occurs at steady-state and is described by the fluctuation-dissipation relation (Eq. 5.15 with $\frac{d\Xi}{dt} = 0$),

$$\mathbf{\Gamma}_s \cdot \mathbf{\Xi}_s + \mathbf{\Xi}_s \cdot \mathbf{\Gamma}_s^T + \mathbf{D}_s = 0 \quad (5.17)$$

where each of the matrices is evaluated at a stable equilibrium point of the deterministic system. Compare Eq. 5.12 with Kramer's equation for a Brownian particle trapped in a potential well (Section 6.7 on page 143). You will see that $\mathbf{\Gamma}$ plays the role of the curvature of the potential (the spring constant), and $\mathbf{D}$ plays the role of temperature.

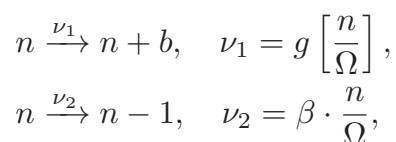
4. It is straightforward to show that the autocorrelation of the fluctuations about the steady-state are exponentially distributed (Exercise 3),

$$\langle \boldsymbol{\alpha}(t) \cdot \boldsymbol{\alpha}^T(0) \rangle = \exp[\mathbf{\Gamma}_s t] \cdot \mathbf{\Xi}_s. \quad (5.18)$$

5. Notice the role stoichiometry plays in the magnitude of the fluctuations through the coefficient matrix $\mathbf{D} = \mathbf{S} \cdot \text{diag}[\nu] \cdot \mathbf{S}^T$. Although $\mathbf{\Gamma}$ is unchanged by lumping together the propensity and stoichiometry, the fluctuation matrix $\mathbf{D}$ is *not*! (See Exercise 2.)

5.1.1 Example – Nonlinear synthesis

Consider a one-state model with a general synthesis rate, $g[x]$, but linear degradation,



where n is the number of molecules and Ω is the volume of the reaction vessel (so that $X = n/\Omega$ is the concentration). The master equation for this process is,

$$\frac{dP(n, t)}{dt} = \{\mathbf{E}^{-b} - 1\} g\left[\frac{n}{\Omega}\right] \cdot P(n, t) + \beta \{\mathbf{E}^1 - 1\} nP(n, t). \quad (5.19)$$

Substituting the ansatz (Eq. 5.1), along with preceding approximations (Eqs. 5.3, 5.6 and 5.8), into the master equation (Eq. 5.19),

$$\begin{aligned} & \frac{1}{\sqrt{\Omega}} \left\{ \frac{d\Pi}{dt} - \sqrt{\Omega} \frac{\partial \Pi}{\partial \alpha} \frac{dx}{dt} \right\} = \\ & \left\{ \frac{-b}{\sqrt{\Omega}} \frac{\partial}{\partial \alpha} + \frac{b^2}{2\Omega} \frac{\partial^2}{\partial \alpha^2} \right\} \left\{ g[x] + \frac{1}{\sqrt{\Omega}} g'[x] \alpha \right\} \frac{1}{\sqrt{\Omega}} \Pi(\alpha, t) + \\ & \beta \left\{ \frac{1}{\sqrt{\Omega}} \frac{\partial}{\partial \alpha} + \frac{1}{2\Omega} \frac{\partial^2}{\partial \alpha^2} \right\} \left\{ \Omega x + \sqrt{\Omega} \alpha \right\} \frac{1}{\sqrt{\Omega}} \Pi(\alpha, t). \end{aligned}$$

Cancelling a common factor of $1/\sqrt{\Omega}$ from both sides, and collecting terms in like powers of $1/\sqrt{\Omega}$, to leading-order,

$$-\frac{dx}{dt} \frac{\partial \Pi}{\partial \alpha} = -\{b \cdot g[x] - \beta x\} \frac{\partial \Pi}{\partial \alpha}.$$

This equation is satisfied if the deterministic trajectory $x(t)$ obeys the deterministic rate equation,

$$\frac{dx}{dt} = b \cdot g[x] - \beta \cdot x. \quad (5.20)$$

Although the approximation is called the linear noise approximation, the full nonlinearity in the deterministic trajectory is retained. To first-order in $1/\sqrt{\Omega}$, the fluctuations $\Pi(\alpha, t)$ obey a linear Fokker-Planck equation,

$$\frac{\partial \Pi}{\partial t} = -\{b \cdot g'[x] - \beta\} \frac{\partial (\alpha \Pi)}{\partial \alpha} + \frac{\{b^2 \cdot g[x] + \beta \cdot x\}}{2} \frac{\partial^2 \Pi}{\partial \alpha^2}. \quad (5.21)$$

The solution to this equation is straightforward – it is simply a Gaussian centered on the trajectory $x(t)$ determined by Eq. 5.20, with covariance given by,

$$\frac{d\langle \alpha^2 \rangle}{dt} = 2 \{b \cdot g'[x] - \beta\} \langle \alpha^2 \rangle + \{b^2 \cdot g[x] + \beta \cdot x\}.$$

The two equations, Eq. 5.20 and 5.21, provide an approximation of the full time-dependent evolution of the probability distribution $P(n, t)$

obeying the master equation (Eq. 5.19). The series neglects terms of order $1/\Omega$ and higher. If these terms are retained, then the coefficients in Eq. 5.21 become nonlinear in α , and higher-order partial derivatives appear, *i.e.* the distribution $\Pi(\alpha, t)$ is no longer Gaussian (see Exercise 5).

At steady-state, the mean x^* satisfies the algebraic condition,

$$x^* = \frac{b}{\beta} \cdot g[x],$$

and the variance is given by,

$$\langle\langle\alpha^2\rangle\rangle = \frac{(b+1)}{2} \frac{\beta x^*}{b \cdot g'[x^*] - \beta}. \quad (5.22)$$

What is the system size?

In the derivation of the linear noise approximation, the parameter Ω is used to order the terms in the perturbation expansion. The validity of the expansion depends upon the magnitude of Ω only indirectly; what is required is that $\Omega \mathbf{x} \gg \sqrt{\Omega} \boldsymbol{\alpha}$. More informally, it is essential to find an extensive parameter that quantifies the relative magnitude of the jump size in the reaction *numbers* (and take the limit that this jump magnitude becomes small). What precisely that parameter is will depend upon the problem at hand – for many electrical applications it is proportional to the capacitance in the circuit, for many chemistry problems it is proportional to the volume.

For example, consider a model of an autoactivator at the level of protein (after time-averaging the promoter and mRNA kinetics), and further suppose the synthesis of the activator is expressed as a Hill function so that the deterministic rate equation governing the concentration of A is,

$$\frac{dA}{dt} = A_0 \cdot g[A/K_A] - A,$$

where for convenience time is scaled relative to the degradation rate of A , and A_0 is the maximal synthesis rate of A . This model is a specific example of the nonlinear synthesis model described in detail above. The function $g[A/K_A]$ depends upon the dissociation constant K_A (in units of concentration). To be specific, let

$$g[x] = x^2/(1+x^2).$$

The relevant concentration scale in the deterministic model is the dissociation constant K_A , and the unitless parameter that fully characterizes

the behaviour of the deterministic model is the ratio A_0/K_A . Physically, the dissociation constant K_A determines the threshold beyond which a bistable positive feedback loop switches from the low state to the high state (or vice-versa). In an electronic feedback loop with a microphone pointed at a speaker, K_A would correspond to the sensitivity of the microphone to noise from the speaker, determining how loud a disturbance from the speaker must be in order to initiate a feedback whine.

For the linear noise approximation to be valid, we need the fractional deviation of the molecule numbers to be small. It is straightforward to analyze this condition around steady-state. The fractional deviation is,

$$\sqrt{\frac{\langle\langle n^2 \rangle\rangle}{\langle n \rangle^2}} = \sqrt{\frac{\langle\langle \alpha^2 \rangle\rangle}{\Omega x^2}}.$$

For clarity, we make the simplifying assumption that $g'[x^*] \ll 1$ after scaling time relative to the degradation rate. Then, using Eq. 5.22,

$$\sqrt{\frac{\langle\langle n^2 \rangle\rangle}{\langle n \rangle^2}} = \sqrt{\frac{(b+1)}{2} \frac{1}{\Omega x^*}}.$$

Non-dimensionalizing the mean x^* using the characteristic concentration K_A , the fractional deviation will remain small provided the unitless parameter,

$$\Delta = \frac{(b+1)}{2} \frac{1}{\Omega K_A},$$

remains small. What is the physical interpretation of Δ ? The first factor, $(b+1)/2$, is the average change in numbers when a reaction occurs (1 for degradation, b for synthesis). The second factor is perhaps the more important factor – it is the dissociation constant (in units of concentration) multiplied by the system volume. The factor $1/(\Omega K_A)$ expresses a characteristic *number* scale in the system. If K_A was, for example, 25 nanomolar and you were working in *E. coli* ($\Omega \sim 10^{-15}\text{L}$), then $K_A\Omega$ is about 25 molecules. So long as the average jump size $(b+1)/2$ is much less than 25, the linear noise approximation provides a useful approximation.

In the autoactivator example, the state space is one-dimensional and so the number scale is obvious. For a more complicated model, it may not be so obvious what dimensionless parameters must be small for the approximation to remain valid. One strategy is to identify in the diffusion matrix $\mathbf{D}$ (Eq. 5.13) dimensionless parameters like Δ that will control the magnitude of the variance.

5.1.2 Approximation of the Brusselator model

To illustrate the linear noise approximation, we return to the Brusselator example introduced on page 88. We use the ansatz $n_i = \Omega x_i + \sqrt{\Omega} \alpha_i$ to re-write the molecule numbers n_i as concentrations x_i and fluctuations α_i . Expanding the step-operator and propensity functions in the master equation for the Brusselator, Eq. 4.25, then collecting terms in powers of $\Omega^{-\frac{1}{2}}$, at Ω^0 we have,

$$-\frac{dx_1}{dt} \frac{\partial \Pi}{\partial \alpha_1} - \frac{dx_2}{dt} \frac{\partial \Pi}{\partial \alpha_2} = [-ax_1^2 x_2 + (1+b)x_1 - 1] \frac{\partial \Pi}{\partial \alpha_1} + [ax_1^2 x_2 - bx_1] \frac{\partial \Pi}{\partial \alpha_2}.$$

Identifying x_1 and x_2 with the macroscopic trajectory of the system, this term vanishes identically, because from Eq. 4.24,

$$\begin{aligned} \frac{dx_1}{dt} &= 1 + ax_1^2 x_2 - (1+b)x_1, \\ \frac{dx_2}{dt} &= -ax_1^2 x_2 + bx_1. \end{aligned}$$

The next term in the expansion, at $\Omega^{-\frac{1}{2}}$, is

$$\begin{aligned} \frac{\partial \Pi}{\partial t} &= -\frac{\partial}{\partial \alpha_1} \{ [(2ax_1 x_2 - (b+1)) \alpha_1 + ax_1^2 \alpha_2] \Pi \} \\ &\quad - \frac{\partial}{\partial \alpha_2} \{ [(b - 2ax_1 x_2) \alpha_1 - ax_1^2 \alpha_2] \Pi \} + \\ &\quad + \frac{1}{2} \left\{ [(b+1)x_1 + ax_1^2 x_2 + 1] \frac{\partial^2 \Pi}{\partial \alpha_1^2} + [bx_1 + ax_1^2 x_2] \frac{\partial^2 \Pi}{\partial \alpha_2^2} \right\} \\ &\quad + [-bx_1 - ax_1^2 x_2] \frac{\partial^2 \Pi}{\partial \alpha_1 \partial \alpha_2}. \end{aligned}$$

This is a Fokker-Planck equation with coefficients linear in α_i . Writing the above in more compact notation,

$$\frac{\partial \Pi}{\partial t} = - \sum_{i,j} \Gamma_{ij} \frac{\partial}{\partial \alpha_i} (\alpha_j \Pi) + \frac{1}{2} \sum_{i,j} D_{ij} \frac{\partial^2 \Pi}{\partial \alpha_i \partial \alpha_j} \quad (5.23)$$

with,

$$\begin{aligned} \mathbf{\Gamma} &= \begin{bmatrix} 2ax_1 x_2 - (b+1) & ax_1^2 \\ b - 2ax_1 x_2 & -ax_1^2 \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} (b+1)x_1 + ax_1^2 x_2 & -bx_1 - ax_1^2 x_2 \\ -bx_1 - ax_1^2 x_2 & bx_1 + ax_1^2 x_2 \end{bmatrix}. \end{aligned}$$

In this form, the mean and variance of α_1 and α_2 are easily derived, since:

$$\begin{aligned}\frac{d}{dt} \langle \alpha_i \rangle &= \sum_j \Gamma_{ij} \langle \alpha_j \rangle \\ \frac{d}{dt} \langle \alpha_i \alpha_j \rangle &= \sum_k \Gamma_{ik} \langle \alpha_k \alpha_j \rangle + \sum_k \Gamma_{kj} \langle \alpha_i \alpha_k \rangle + D_{ij}\end{aligned}$$

to give:

$$\begin{aligned}\frac{d}{dt} \langle \alpha_1 \rangle &= [2ax_1x_2 - (b+1)] \langle \alpha_1 \rangle + ax_1^2 \langle \alpha_2 \rangle, \\ \frac{d}{dt} \langle \alpha_2 \rangle &= [b - 2ax_1x_2] \langle \alpha_1 \rangle - ax_1^2 \langle \alpha_2 \rangle,\end{aligned}$$

and,

$$\begin{aligned}\frac{d}{dt} \langle \alpha_1^2 \rangle &= 2[2ax_1x_2 - (b+1)] \langle \alpha_1^2 \rangle + 2ax_1^2 \langle \alpha_1 \alpha_2 \rangle + 1 + (b+1)x_1 + ax_1^2x_2, \\ \frac{d}{dt} \langle \alpha_1 \alpha_2 \rangle &= [b - 2ax_1x_2] \langle \alpha_1^2 \rangle + [2ax_1x_2 - (b+1) - ax_1^2] \langle \alpha_1 \alpha_2 \rangle \\ &\quad + ax_1^2 \langle \alpha_2^2 \rangle - bx_1 - ax_1^2x_2, \\ \frac{d}{dt} \langle \alpha_2^2 \rangle &= 2[b - 2ax_1x_2] \langle \alpha_1 \alpha_2 \rangle - 2ax_1^2 \langle \alpha_2^2 \rangle + bx_1 + ax_1^2x_2.\end{aligned}$$

Stable regime, $b < 1 + a$: In the parameter regime where the macroscopic solutions are asymptotically stable, and tend to $(x_1^{ss}, x_2^{ss}) \rightarrow (1, \frac{b}{a})$, the mean of the fluctuations will vanish asymptotically ($\langle \alpha_1 \rangle^{ss}, \langle \alpha_2 \rangle^{ss} \rightarrow (0, 0)$), though the variance remains bounded and non-zero,

$$\langle \alpha_1^2 \rangle^{ss} = \frac{b + (1 + a)}{(1 + a) - b} \quad \langle \alpha_1 \alpha_2 \rangle^{ss} = \frac{-2b}{(1 + a) - b} \quad \langle \alpha_2^2 \rangle^{ss} = \frac{b}{a} \left[\frac{b + (1 + a)}{(1 + a) - b} \right].$$

These expressions obviously become meaningless for a choice of parameters outside the stable region, *i.e.*, $b \geq 1 + a$. If the macroscopic solution is unstable, the variance diverges and after a transient period, the variance will exceed Ω , so that the ordering implied by our original ansatz $n_i = \Omega x_i + \Omega^{\frac{1}{2}} \alpha_i$ breaks down. This is certainly true for exponentially divergent macroscopic trajectories, but for orbitally stable trajectories, it is only the uncertainty in the *phase* that is divergent. To examine the fluctuations in a system on a periodic orbit, more sophisticated techniques are required (see Section 10.2 on page 217).

It is instructive to compare the linear noise approximation with the results of stochastic simulation in order to contrast the two approaches.

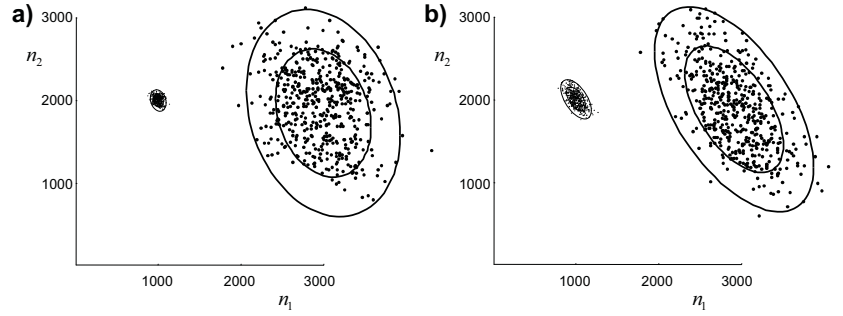


Figure 5.2: **The stable system - far from the *Hopf* bifurcation.** Below the critical line $b = 1 + a$, the macroscopic system is stable, with microscopic trajectories fluctuating about the fixed point, shown here at $(n_1, n_2) = (1000, 2000)$. **a)** $(a, b) = (0.1, 0.2)$, ($\alpha = 10$ in Figure 23 of Gillespie (1977)). The system is very stable - the inset shows the first and second standard deviations of the Gaussian variations predicted by the linear noise approximation. **b)** $(a, b) = (0.5, 1)$, ($\alpha = 2$ in Figure 22 of Gillespie (1977)). The system is still stable, but fluctuations are more appreciable. In both figures, $\Omega = 1000$, and the linear noise approximation describes the statistics of the fluctuations very well.

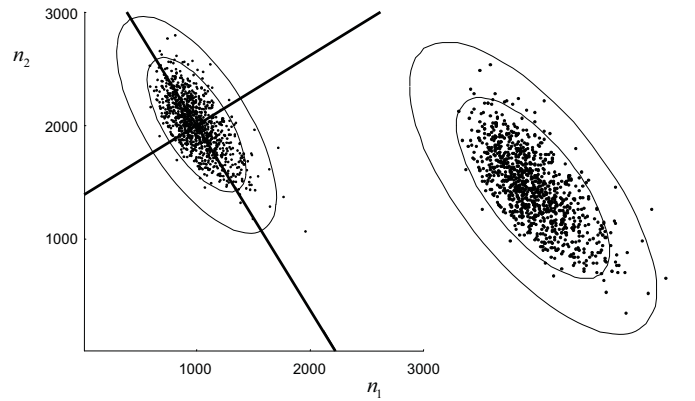


Figure 5.3: **The stable system - near to the *Hopf* bifurcation.** As the parameters approach the critical line $b = 1 + a$, the fluctuations are enhanced substantially. Here $a = 0.95, b = 1.90$, very close to the parameters used in Figure 4.5.

Since retention of terms up to first order in $\Omega^{-\frac{1}{2}}$ results in a linear Fokker-Planck equation, the probability density $\Pi(\alpha_1, \alpha_2, t)$ must necessarily be Gaussian (to this order in Ω). In fact, at steady-state, the probability density of α_i can be explicitly written as the bivariate Gaussian distribution,

$$\Pi^{ss}(\alpha_1, \alpha_2) = \frac{1}{2\pi\sqrt{\det \Xi_s}} \exp \left[-\frac{1}{2} (\alpha_1, \alpha_2) \cdot \Xi_s^{-1} \cdot (\alpha_1, \alpha_2)^T \right],$$

with

$$\Xi_s = \begin{bmatrix} \langle \alpha_1^2 \rangle^{ss} & \langle \alpha_1 \alpha_2 \rangle^{ss} \\ \langle \alpha_1 \alpha_2 \rangle^{ss} & \langle \alpha_2^2 \rangle^{ss} \end{bmatrix} = \frac{1}{(1+a)-b} \begin{bmatrix} b+(1+a) & -2b \\ -2b & \frac{b}{a}[b+(1+a)] \end{bmatrix}, \quad (5.24)$$

centred upon $(x_1^{ss}, x_2^{ss}) = (1, \frac{b}{a})$. The fluctuations are therefore constrained to lie within some level curve of $\Pi^{ss}(\alpha_1, \alpha_2)$, described by the ellipse

$$\begin{aligned} & [(b+(1+a))((1+a)-b)] \alpha_1^2 + \frac{a}{b} [(b+(1+a))((1+a)-b)] \alpha_2^2 \\ & + 4a[(1+a)-b] \alpha_1 \alpha_2 \\ & = K \sqrt{\frac{b(b^2+2b-2ab+2a+a^2+1)^{\frac{3}{2}}}{a[(1+a)-b]^2}}. \end{aligned}$$

The constant K determines the fraction of probability contained under the surface $\Pi^{ss}(\alpha_1, \alpha_2)$ (Figure 5.2). Notice that the major and minor axes of the level curve ellipse are the eigenvectors of Ξ^{-1} . As the system parameters approach the bifurcation line $b \rightarrow 1+a$, the fluctuations become enhanced and the eigenvectors approach a limiting slope (Figure 5.3 and Exercise 4).

In the stable regime, the linear noise approximation provides an excellent approximation of the statistics of the fluctuations. For fluctuations around a limit-cycle, the linear noise approximation must be modified (see Section 10.2 on page 217). For the tracking of a trajectory through a multi-stable phase space, the linear noise approximation cannot be used because it relies on linearization about a unique stable macroscopic trajectory. In systems exhibiting a limit cycle or multistability, stochastic simulation can be performed easily. Nevertheless, as the number of molecules increases ($\Omega \rightarrow \infty$), stochastic simulation becomes prohibitively time-consuming, though the analytic approximation becomes more reliable.

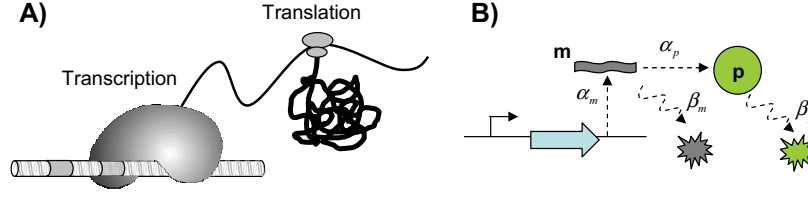


Figure 5.4: **Central Dogma of Molecular Biology** **A)** RNA-polymerase transcribes DNA into messenger RNA (mRNA). Ribosomes translate mRNA into protein. **B)** The process of protein synthesis can be represented by a simplified model. Here, the block arrow denotes the DNA of the gene of interest. The gray ribbon denotes mRNA (m), and the green circle denotes protein (p).

5.2 Example – ‘Burstiness’ in Protein Synthesis

- M. Thattai and A. van Oudenaarden (2001) “Intrinsic noise in gene regulatory networks,” *Proceedings of the National Academy of Sciences USA* **98**: 8614-8619.

In cells, protein is synthesized via a two-step process - DNA is *transcribed* to messenger RNA (mRNA) and that mRNA is *translated* to protein. Deterministically, this process can be represented as a pair of linear, coupled differential equations for the concentrations of mRNA (m) and protein (p),

$$\begin{aligned}\frac{dm}{dt} &= \alpha_m - \beta_m m, \\ \frac{dp}{dt} &= \alpha_p m - \beta_p p.\end{aligned}\tag{5.25}$$

where α_m is the rate of transcription, α_p is the rate of translation, and β_m and β_p are the rates of mRNA and protein degradation, respectively (Figure 5.4). Notice the steady-state values of the mRNA and protein levels are,

$$m^s = \frac{\alpha_m}{\beta_m}, \quad \text{and} \quad p^s = \frac{\alpha_p \cdot m^s}{\beta_p} = \frac{\alpha_m \cdot \alpha_p}{\beta_m \cdot \beta_p}.\tag{5.26}$$

On a mesoscopic level, the deterministic model above is recast as a master

equation for the *number* of mRNA (n_1) and protein (n_2) molecules,

$$\begin{aligned} \frac{\partial P(n_1, n_2, t)}{\partial t} = & \alpha_m (\mathbf{E}_1^{-1} - 1) P(n_1, n_2, t) + \beta_m (\mathbf{E}_1^1 - 1) n_1 P(n_1, n_2, t) + \\ & + \alpha_p (\mathbf{E}_2^{-1} - 1) n_1 P(n_1, n_2, t) + \beta_p (\mathbf{E}_2^1 - 1) n_2 P(n_1, n_2, t), \end{aligned} \quad (5.27)$$

where as above $\mathbf{E}$ is the *step-operator*: $\mathbf{E}_i^k f(n_i, n_j) = f(n_i + k, n_j)$. All of the transition rates are linear in the variables n_1, n_2 , so in principle the full distribution $P(n_1, n_2, t)$ can be obtained. Using the moment generating function $Q(z_1, z_2, t)$ as in Section 4.1, the partial differential equation for Q is,

$$\begin{aligned} \frac{\partial Q(z_1, z_2, t)}{\partial t} = & \alpha_m (z_1 - 1) Q + \beta_m (1 - z_1) \frac{\partial Q}{\partial z_1} + \alpha_p z_1 (z_2 - 1) \frac{\partial Q}{\partial z_1} + \beta_m (1 - z_2) \frac{\partial Q}{\partial z_2}. \end{aligned} \quad (5.28)$$

Despite being a linear partial differential equation, even at steady-state it is difficult to determine $Q^s(z_1, z_2)$ exactly. Often Q is expanded as a Taylor series about the point $z_1 = z_2 = 1$, and using Eqs. 4.3 and 4.4, an evolution equation for the first two moments of the distribution are obtained by treating $\varepsilon_1 = (z_1 - 1)$ and $\varepsilon_2 = (z_2 - 1)$ as small parameters and collecting terms of like powers in ε_i . The calculation is algebraically tedious – you are encouraged to try.

Alternatively, one can use the linear noise approximation. We make the usual ansatz that the concentration can be partitioned into deterministic and stochastic components,

$$m(t) = x_1(t) + \Omega^{-\frac{1}{2}} \alpha_1(t) \quad p(t) = x_2(t) + \Omega^{-\frac{1}{2}} \alpha_2(t).$$

Focusing upon the steady-state, the mean mRNA and protein levels correspond to the deterministic values,

$$\begin{aligned} x_1^s = m^s \quad x_2^s = p^s, \\ (\alpha_1^s = \alpha_2^s = 0). \end{aligned}$$

The steady-state covariance is determined by the fluctuation-dissipation relation, Eq. 5.17. Solving for the steady-state covariance (writing $\phi = \beta_m / \beta_p$),

$$\Xi^s = \begin{bmatrix} \langle m^2 \rangle & \langle mp \rangle \\ \langle mp \rangle & \langle p^2 \rangle \end{bmatrix} = \begin{bmatrix} m^s & \frac{p^s}{1+\phi} \\ \frac{p^s}{1+\phi} & \left(p^s + \frac{(p^s)^2}{m^s(1+\phi)} \right) \end{bmatrix}. \quad (5.29)$$

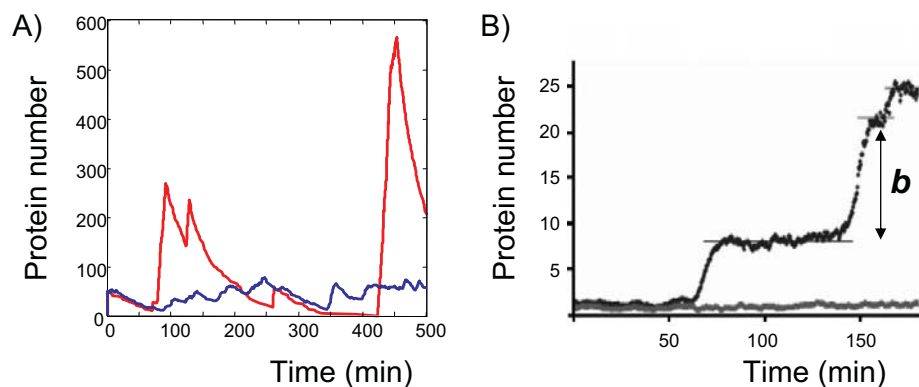


Figure 5.5: **Burstiness of protein synthesis.** **A)** Stochastic simulation the minimal model of protein synthesis (Eq. 5.27). Both the red and the blue curve have the same mean ($\langle p \rangle = 50$), but distinctly different variance. The difference is captured by the burstiness; $(\alpha_m, \alpha_p) = (0.1, 2)$ [blue] and $(0.01, 20)$ [red]. The degradation rates are the same for both: $\beta_m^{-1} = 5$ mins and $\beta_P^{-1} = 50$ mins. **B)** Burstiness can be observed experimentally. The Xie lab has developed sophisticated methods to observe bursts of protein off of individual mRNA transcribed from a highly repressed *lac* promoter. Trapping individual *E. coli* cells in a microfluidic chamber, it is possible to observe the step-wise increase of β -gal off the *lacZ* gene. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440:358-362.

The steady-state values m^s and p^s are simply the steady-state values calculated from the deterministic rate equations (Eq. 5.26). Focusing upon the protein, the fractional deviation of the fluctuations $\sqrt{\langle p^2 \rangle} / \langle p \rangle$ is,

$$\frac{\langle p^2 \rangle}{\langle p \rangle^2} = \frac{1}{p^s} \left[1 + \frac{\alpha_p}{\beta_p} \frac{1}{(1 + \phi)} \right].$$

In prokaryotes, the degradation rate of the mRNA is typically much greater than the degradation of the protein, so $\phi \gg 1$, and the fractional deviation becomes,

$$\frac{\langle p^2 \rangle}{\langle p \rangle^2} \approx \frac{1}{p^s} \left[1 + \frac{\alpha_p}{\beta_m} \right] \equiv \frac{1}{p^s} [1 + b], \quad (5.30)$$

where $b = \alpha_p / \beta_m$ is called the *burstiness* of protein synthesis. The burstiness is a measure of the amplification of transcription noise by translation, since each errant transcript is amplified by $b = \frac{\alpha_p}{\beta_m} = (\text{protein molecules} / \text{mRNA}) \times (\text{average mRNA lifetime})$ (Figure 5.5). What is surprising is that we can observe this burstiness experimentally (Figure 5.5b)!

The moment generating functions only work with linear transition rates and can be algebraically formidable. The linear noise approximation, by contrast, is calculated algorithmically and works regardless of the form of the transition rates, and in fact is *exact* for linear transition rates. Having said that, it is a wonder that moment generating functions are used at all.

5.3 Limitations of the LNA

The linear noise approximation is built upon the ansatz that the full state can be separated into deterministic and stochastic components, with the fluctuations scaling with the square-root of the system size (*cf.* Eq. 5.1),

$$n_i = \Omega x_i + \sqrt{\Omega} \alpha_i.$$

The limitations of the linear noise approximation are implicit in the ansatz. First, the approximation is a perturbation expansion that is valid so long as terms of order $1/\sqrt{\Omega}$ remain sub-dominant to the leading-order deterministic trajectory. Second, the linear noise approximation is a *local* expansion of the master equation, with the probability distribution of the fluctuations $\Pi(\boldsymbol{\alpha}, t)$ evaluated along the trajectory $\boldsymbol{x}(t)$.

The most conservative application of the approximation is to systems with deterministic dynamics exhibiting a single asymptotically stable fixed-point. For multistable systems, the approximation can still be applied, but only within a basin of attraction thereby providing an estimation of the fluctuations along a trajectory close to a fixed point. Furthermore, with slightly more elaborate analysis, the fluctuations about a stable *limit cycle* can also be estimated (see Section 10.2 on p. 217). The linear noise approximation cannot, however, be used to compute *global* properties of the system, such as switching rates between fixed-points or splitting probabilities among multiple equilibria. For those types of computation, stochastic simulation is the only consistent method presently available.

Suggested References

The review by van Kampen of his linear noise approximation is excellent,

- N. G. van Kampen (1976) “Expansion of the Master equation,” *Advances in Chemical Physics* **34**: 245.

For a critique of the Kramers-Moyal expansion as a reliable approximation of the master equation, see the two articles by van Kampen,

- N. G. van Kampen (1981) “The validity of non-linear Langevin-equations,” *Journal of Statistical Physics* **25**: 431.
- N. G. van Kampen, “The diffusion approximation for Markov processes,” in *Thermodynamics and kinetics of biological processes*, Ed. I. Lamprecht and A. I. Zotin (Walter de Gruyter & Co., 1982).

Exercises

1. **Linear Fokker-Planck equation:** Show that the solution of the linear Fokker-Planck equation is Gaussian for all time (conditioned by a delta-peak at $t = 0$).
2. **Stoichiometry vs. propensity:** The deterministic rate equations are written in terms of the dot-product of the stoichiometry matrix $\mathbf{S}$ and the propensity vector $\boldsymbol{\nu}$, Eq. 5.11. From the perspective of the phenomenological model, it is permissible to re-scale the entries in the stoichiometry to ± 1 , and absorb the step-size into each entry in

the propensity vector. From the point of view of characterizing the fluctuations, it is very important to distinguish between stoichiometry and propensity. To that end, examine the following process with variable step-size:

$$\begin{aligned} n &\xrightarrow{\nu_1} n + a, & \nu_1 &= \frac{\alpha}{a}, \\ n &\xrightarrow{\nu_2} n - b, & \nu_2 &= \frac{\beta}{b} \cdot n. \end{aligned}$$

- (a) Show that the deterministic rate equation is *independent* of the steps a and b .
 - (b) Calculate the variance in n . Show that it is *strongly dependent* upon a and b .
3. **Exponential autocorrelation in the linear noise approximation:** Derive the expression for the steady-state autocorrelation function in the linear noise approximation, Eq. 5.18. There are several ways to do this; perhaps the most straightforward is to generalize the ideas developed in the section on the fluctuation-dissipation relation, Section 2.3.2 on p. 42.
4. **Brusselator near the bifurcation:** The major and minor axes of the stationary probability distribution equiprobability contours are given by the eigenvectors of the inverse-covariance matrix, Ξ^{-1} . In the Brusselator example, calculate these eigenvectors using Eq. 5.24. What is the slope of the major axis as the parameters approach the bifurcation $a \rightarrow b - 1$? Do you recognize this number?
5. **Higher-order corrections:** For a model with nonlinear degradation,

$$\begin{aligned} n &\xrightarrow{\nu_1} n + 1, & \nu_1 &= \alpha, \\ 2n &\xrightarrow{\nu_2} n, & \nu_2 &= \frac{\beta n(n-1)}{\Omega^2}, \end{aligned}$$

determine the following, including terms of order $1/\Omega$ in the linear noise approximation:

- (a) The partial differential equation for $\Pi(\alpha, t)$.
- (b) The mean and variance of n .
- (c) The auto-correlation function at steady-state $\langle\langle n(t)n(t-\tau) \rangle\rangle$.

6. **Negative feedback:** Consider the nonlinear negative feedback model,

$$\begin{aligned} r &\xrightarrow{\nu_1} r + b, & \nu_1 &= \alpha \cdot g(r/(\Omega \cdot K_R)), \\ r &\xrightarrow{\nu_2} r - 1, & \nu_2 &= \beta \frac{r}{\Omega}, \end{aligned}$$

where $g(x)$ is a Hill-type function,

$$g(x) = (1 + x^n)^{-1}.$$

In the regime where the system is *homeostatic* ($r^* \gg \Omega \cdot K_R$):

- (a) Find explicit estimates for the steady-state mean $\langle r^* \rangle$, and the steady-state variance $\langle \langle r^2 \rangle \rangle$.
- (b) How does the fractional deviation about the steady-state, $\sqrt{\langle \langle r^2 \rangle \rangle} / \langle r^* \rangle$, change as the cooperativity n increases?
- (c) Identify a unitless parameter that determines the validity of the linear noise approximation (see Section 5.1.1).

CHAPTER 6

FOKKER-PLANCK EQUATION

The work of Ornstein and Uhlenbeck on Brownian motion suggested that there is a general connection between a certain class of stochastic differential equations for a given trajectory (as studied by Langevin) and a certain family of partial differential equations governing the probability distribution for an ensemble of trajectories (as studied by Einstein). In this Chapter, we formalize the connection between the two approaches, and find that for stochastic differential equations with multiplicative white noise, a new calculus is required.

The master equation (*cf.* Eq. 3.11 on page 59),

$$\frac{\partial}{\partial \tau} p(x_3|x_1, \tau) = \int_{-\infty}^{\infty} [w(x_3|x_2) p(x_2|x_1, \tau) - w(x_2|x_3) p(x_3|x_1, \tau)] dx_2,$$

is integro-differential equation (or, in the case of a discrete state-space, a discrete-differential equation, Eq. 3.12). This family of equations is difficult to solve in general. The Kolmogorov, or Fokker-Planck, equation is an attempt to approximate the master equation by a (one hopes) more amenable partial differential equation governing $p(x_3|x_1, \tau)$. Kolmogorov derived his equation by imposing abstract conditions on the moments of the transition probabilities w in the Chapman-Kolmogorov equation, while the derivation of the same equation by Fokker and Planck proceeds as a continuous approximation of the discrete master equation. We shall see in Chapter 5 that in systems where the noise is intrinsic to the

dynamics, both derivations are inconsistent approximations, and that a systematic approximation method is needed. (See N. G. van Kampen (1982) ‘The diffusion approximation for Markov processes’ in *Thermodynamics and kinetics of biological processes*, Ed. I. Lamprecht and A. I. Zotin, p. 181.)

6.1 Kolmogorov Equation

- A. N. Kolmogorov (1931) *Math. Ann.* **104**: 415.

Consider a stationary, Markov process – one dimensional, for simplicity. Write the Chapman-Kolmogorov equation as,

$$p(x|y, t + \Delta t) = \int_{-\infty}^{\infty} p(x|z, t) p(z|y, \Delta t) dz, \quad (6.1)$$

and define the *jump moments* by,

$$a_n(z, \Delta t) = \int_{-\infty}^{\infty} (y - z)^n p(z|y, \Delta t) dy.$$

Furthermore, assume that for $\Delta t \rightarrow 0$, only the 1st and 2nd moments become proportional to Δt (assume all higher moments vanish in that limit). Then, the following limits exist,

$$A(z) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} a_1(z, \Delta t) \quad B(z) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} a_2(z, \Delta t). \quad (6.2)$$

Now, let $R(y)$ be a suitable test function, possessing all properties required for the following operations to be well-defined, and note that,

$$\begin{aligned} \int_{-\infty}^{\infty} R(y) \frac{\partial p}{\partial t} dy &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} R(y) [p(x|y, t + \Delta t) - p(x|y, t)] dy \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{-\infty}^{\infty} R(y) p(x|y, t + \Delta t) dy - \int_{-\infty}^{\infty} R(y) p(x|y, t) dy \right\}. \end{aligned}$$

Using Eq. 6.1 to replace $p(x|y, t + \Delta t)$,

$$\int_{-\infty}^{\infty} R(y) \frac{\partial p}{\partial t} dy = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{-\infty}^{\infty} R(y) \int_{-\infty}^{\infty} p(x|z, t) p(z|y, \Delta t) dz dy - \int_{-\infty}^{\infty} R(y) p(x|y, t) dy \right\}.$$

We expand $R(y)$ around the point $y = z$ and interchange the order of integration. Since we have assumed all jump moments beyond the second vanish, we are left with,

$$\begin{aligned} & \int_{-\infty}^{\infty} R(y) \frac{\partial p}{\partial t} dy = \\ & \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{-\infty}^{\infty} R'(z) p(x|z, t) \left(\int_{-\infty}^{\infty} (y - z) p(z|y, \Delta t) dy \right) dz \right\} \\ & + \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \left\{ \int_{-\infty}^{\infty} R''(z) p(x|z, t) \left(\int_{-\infty}^{\infty} \frac{(y - z)^2}{2} p(z|y, \Delta t) dy \right) dz \right\}. \end{aligned}$$

(Prove to yourself that this is true.) In terms of the functions $A(z)$ and $B(z)$ defined above in Eq. 6.2,

$$\int_{-\infty}^{\infty} R(y) \frac{\partial p}{\partial t} dy = \int_{-\infty}^{\infty} R'(z) p(x|z, t) A(z) dz + \int_{-\infty}^{\infty} R''(z) p(x|z, t) B(z) dz. \quad (6.3)$$

Recall that $R(y)$ is an arbitrary test function, with all the necessary properties for the following to hold. In particular, choose $R(y)$ so that $R(y)$ and $R'(y) \rightarrow 0$ as $y \rightarrow \pm\infty$ as fast as necessary for the following to be true,

$$\begin{aligned} \lim_{z \rightarrow \pm\infty} R(z) p(x|z, t) A(z) &= 0, \\ \lim_{z \rightarrow \pm\infty} R'(z) p(x|z, t) B(z) &= 0. \end{aligned}$$

Integrating Eq. 6.3 by parts, we obtain:

$$\int_{-\infty}^{\infty} R(z) \left[\frac{\partial p(x|z, t)}{\partial t} + \frac{\partial}{\partial z} \{A(z) p(x|z, t)\} - \frac{1}{2} \frac{\partial^2}{\partial z^2} \{B(z) p(x|z, t)\} \right] dz = 0.$$

By the du Bois-Reymond lemma from analysis, since $R(y)$ is an *arbitrary* test function, we must conclude that,

$$\frac{\partial p(x|y, t)}{\partial t} = -\frac{\partial}{\partial y} \{A(y) p(x|y, t)\} + \frac{1}{2} \frac{\partial^2}{\partial y^2} \{B(y) p(x|y, t)\}. \quad (6.4)$$

This is the *Kolmogorov Equation*, often called the *Fokker-Planck Equation*, though sometimes appearing as the *Smoluchowski Equation*, or the *Generalized Diffusion Equation*. The first term on the right, involving $A(y)$, is called the *convection* or *drift* term; the second term, involving $B(y)$, is called the *diffusion* or *fluctuation* term.

Remarks:

1. Eq. 6.4 may be written as a *continuity equation*,

$$\frac{\partial p}{\partial t} = -\frac{\partial J}{\partial y},$$

where J is the *probability flux* obeying the constitutive equation,

$$J(y) = A(y)p - \frac{1}{2} \frac{\partial}{\partial y} \{B(y)p\}.$$

2. The derivation above may be repeated (with minor modifications) for an n -dimensional process; the result is then given by,

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^n \frac{\partial}{\partial y_i} \{A_i(\mathbf{y}) p\} + \frac{1}{2} \sum_{k,l=1}^n \frac{\partial^2}{\partial y_k \partial y_l} \{B_{kl}(\mathbf{y}) p\}. \quad (6.5)$$

3. The derivation of the Kolmogorov equation was rightly considered a breakthrough in the theory of stochastic processes. It is obvious that having a partial differential equation that is equivalent (under the assumed conditions) to the impossibly difficult Chapman-Kolmogorov equation is great progress. This success has been exploited in a large number of applications since 1931; but as work in this area grew, subtle questions of consistency arose, and the debate is far from over today. The crux of the problem is the following: Despite the impeccable mathematics of Kolmogorov, the derivation does not hold any information regarding what physical processes actually obey the necessary mathematical assumptions – in particular, for which processes one might expect the higher jump moments to vanish. See N. G. van Kampen (1982) ‘The diffusion approximation for Markov processes’ in *Thermodynamics and kinetics of biological processes*, Ed. I. Lamprecht and A. I. Zotin, p. 181.

4. Note that the Kolmogorov equation is an evolution equation for the *conditional* probability density, and that it must be solved subject to the initial condition,

$$p(x|y, 0) = \delta(x - y).$$

5. The initial-value problem can be solved in a few cases (see H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, (Springer-Verlag, 1996), for details). Notice, however, that before any attempt at solving the initial value problem is made, one must find an explicit expression for the coefficients $A(y)$ and $B(y)$.

6.2 Derivation of the Fokker-Planck Equation

Kolmogorov's derivation is formal, with no reference to any particular physical process, nor even an indication of what kinds of physical processes obey the necessary mathematical conditions. Fokker and Planck derived the same equation based on assumptions tied directly to particular features of physical systems under consideration. Nevertheless, it will become clear in Chapter 5 that Fokker's derivation (described below) is only consistent if the transition probabilities are *linear* in the state variables. Despite this very strict limitation, the Fokker-Planck equation is the favoured approximation method for many investigators (although rarely applied to systems with linear transition rates).

The labeled states x_1, x_2, x_3 appearing in the master equation may be interpreted as representing the initial, intermediate and final state, respectively. Let them be re-labeled as y_0, y', y , so that the master equation reads,

$$\frac{\partial}{\partial \tau} p(y|y_0, \tau) = \int_{-\infty}^{\infty} [w(y|y') p(y'|y_0, \tau) - w(y'|y) p(y|y_0, \tau)] dy'. \quad (6.6)$$

This is an *approximate evolution equation for the conditional probability density*, and it is *linear* in $p(y|y_0, \tau)$ since $w(y|y')$ is supposed to be provided by the “physics” of the system.

If the dependence on the initial state is understood, then Eq. 6.6 may

be written in a short-hand form as,

$$\frac{\partial}{\partial \tau} p(y, \tau) = \int_{-\infty}^{\infty} [w(y|y') p(y', \tau) - w(y'|y) p(y, \tau)] dy'. \quad (6.7)$$

This must *not* be confused with an evolution equation for the single-state probability distribution, since p is still conditioned upon the initial distribution $p(y_0, t_0)$. We introduce the jump-size $\Delta y = y - y'$; in that way, we can re-write the transition probabilities in terms of the jump-size Δy . The transition probability $w(y|y')$ is the probability per unit time of a transition $y' \rightarrow y$, *i.e.*, that starting at y' , there is a jump of size Δy . We write this as, $w(y|y') = w(y', \Delta y) \equiv w(y - \Delta y, \Delta y)$. Similarly, we re-write the transition probability $w(y'|y)$ as $w(y'|y) = w(y, -\Delta y)$. With this change in notation, the master equation becomes,

$$\begin{aligned} \frac{\partial}{\partial \tau} p(y, \tau) &= \int_{-\infty}^{\infty} [w(y - \Delta y, \Delta y) p(y - \Delta y, \tau) - w(y, -\Delta y) p(y, \tau)] d\Delta y \\ &= \int_{-\infty}^{\infty} w(y - \Delta y, \Delta y) p(y - \Delta y, \tau) d\Delta y - p(y, \tau) \int_{-\infty}^{\infty} w(y, -\Delta y) d\Delta y. \end{aligned} \quad (6.8)$$

Next, we make two assumptions,

1. *Only small jumps occur.* That is, we assume that there exists a $\delta > 0$ such that $w(y', \Delta y) \approx 0$ for $|\Delta y| > \delta$ and $w(y' + \Delta y, \Delta y) \approx w(y', \Delta y)$ for $|\Delta y| < \delta$.
2. *$p(y, \tau)$ varies slowly with y .* That is, we assume that $p(y + \Delta y, \tau) \approx p(y, \tau)$ for $|\Delta y| < \delta$.

Under these conditions, one is able to Taylor expand w and p in y' about y ,

$$\begin{aligned} w(y', \Delta y) p(y', \tau) &\approx \\ wp|_{y'=y} + \left[\frac{\partial}{\partial y'} (wp) \right]_{y'=y} (y' - y) + \left[\frac{\partial^2}{\partial y'^2} (wp) \right]_{y'=y} \frac{1}{2} (y' - y)^2. \end{aligned}$$

Or, since $\Delta y = y - y'$, to $O(\Delta y^3)$,

$$\begin{aligned} w(y', \Delta y) p(y', \tau) &= \\ w(y, \Delta y) p(y, \tau) - \Delta y \frac{\partial}{\partial y} [w(y, \Delta y) p(y, \tau)] + \frac{\Delta y^2}{2} \frac{\partial^2}{\partial y^2} [w(y, \Delta y) p(y, \tau)]. \end{aligned}$$

Eq. 6.8 is then approximated by,

$$\begin{aligned} \frac{\partial}{\partial \tau} p(y, \tau) = & \int_{-\infty}^{\infty} \cancel{w(y, \Delta y) p(y, \tau) d\Delta y} - \int_{-\infty}^{\infty} \Delta y \frac{\partial}{\partial y} [w(y, \Delta y) p(y, \tau)] d\Delta y \\ & + \frac{1}{2} \int_{-\infty}^{\infty} \Delta y^2 \frac{\partial^2}{\partial y^2} [w(y, \Delta y) p(y, \tau)] d\Delta y - \int_{-\infty}^{\infty} \cancel{w(y, -\Delta y) p(y, \tau) d\Delta y}. \end{aligned} \quad (6.9)$$

Note the dependence of $w(y, \Delta y)$ on the second argument is fully maintained; an expansion with respect to Δy is not possible since w varies rapidly with Δy . Furthermore, since

$$\begin{aligned} \int_{-\infty}^{\infty} \Delta y \frac{\partial}{\partial y} [w(y, \Delta y) p(y, \tau)] d\Delta y = & \frac{\partial}{\partial y} \left[p(y, \tau) \int_{-\infty}^{\infty} \Delta y w(y, \Delta y) d\Delta y \right] = \\ & \frac{\partial}{\partial y} [p(y, \tau) a_1(y)], \end{aligned}$$

we finally obtain the *Fokker-Planck Equation*,

$$\frac{\partial}{\partial \tau} p(y, \tau) \approx -\frac{\partial}{\partial y} [a_1(y) p(y, \tau)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [a_2(y) p(y, \tau)]. \quad (6.10)$$

Remarks:

1. The Fokker-Planck equation has the same mathematical form as the Kolmogorov equation, *viz.* a linear, 2^{nd} -order partial differential equation of parabolic type. Nevertheless, the assumptions made in deriving the Fokker-Planck equation make it possible to answer the question: For a given w , how good an approximation is provided by this equation?
2. It is not difficult to include all terms in the Taylor expansion used above and obtain the *Kramers-Moyal Expansion*,

$$\frac{\partial}{\partial \tau} p(y, \tau) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial y^n} [a_n(y) p(y, \tau)], \quad (6.11)$$

where,

$$a_n(y) = \int_{-\infty}^{\infty} \Delta y^n w(y, \Delta y) d\Delta y.$$

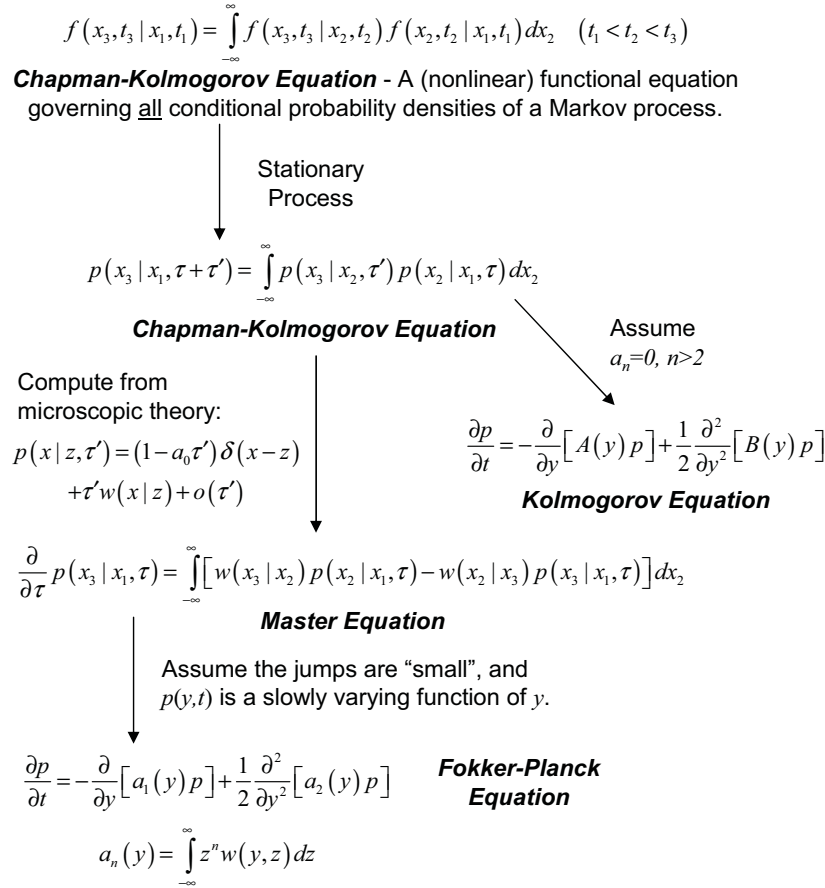


Figure 6.1: Review of Markov Processes

While this is *formally* equivalent to the master equation, to be practical the expansion must be truncated at a certain point – the Fokker-Planck equation, for example, is the result of truncation after two terms. Breaking-off the series in a consistent fashion requires the introduction of a well-defined extensive variable quantifying the relative size of the individual jump events (as we shall see in section 5.1). The truncated Kramers-Moyal expansion is what Einstein used in his derivation of the diffusion equation for a Brownian particle (*cf.* Eq. 1.7 on page 6).

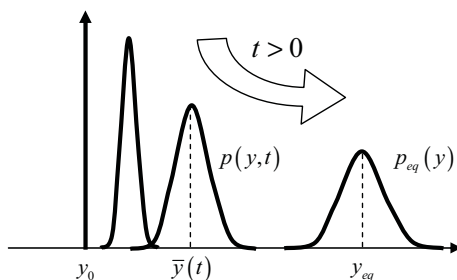


Figure 6.2: **Evolution of $p(y, t)$ toward its equilibrium distribution $p_{eq}(y)$.**

6.3 Macroscopic Equation

- R. Kubo, K. Matsuo and K. Kitahara (1973) Fluctuation and relaxation of macrovariables. *Journal of Statistical Physics* **9**:51–96

We know that the study of physical phenomena always starts from a phenomenological approach, reflecting the initial stage of organization of experimental facts. This stage often leads to a mathematical model in terms of differential equations where the variables of interest are macroscopic variables, in which the microscopic fluctuations are neglected (averaged over), resulting in a deterministic theory. Examples include Ohm's law, chemical rate equations, population growth dynamics, *etc.*

At the next, more fundamental (mesoscopic) level, the fluctuations are taken into account – by the master equation (or the Fokker-Planck equation, if it applies) for stationary Markov processes. As the latter determines the entire probability distribution, it must be possible to derive from it the macroscopic equations as an approximation for the case that the fluctuations are negligible.

Let Y be a physical quantity with Markov character, taking the value y_0 at $t = 0$; *i.e.*, $P(y, t|y_0, 0) \rightarrow \delta(y - y_0)$ as $t \rightarrow 0$. For definiteness, say the system is *closed* and *isolated*. Then we have from equilibrium statistical mechanics that the probability distribution $p(y, t)$ tends toward some equilibrium distribution $p_{eq}(y)$ as $t \rightarrow \infty$,

$$\lim_{t \rightarrow \infty} p(y, t) = p_{eq}(y),$$

(see figure 6.2). We know from experience that the fluctuations remain small during the whole process, and so $p(y, t)$ for each t is a sharply-peaked function. The location of this peak is a fairly well-defined number,

having an uncertainty of the order of the width of the peak, and it is to be identified with the macroscopic value $\bar{y}(t)$. Usually, one takes,

$$\bar{y}(t) = \langle Y \rangle_t = \int_{-\infty}^{\infty} y p(y, t) dy. \quad (6.12)$$

As t increases, the peak slides bodily along the y -axis from its initial location $\bar{y}(0) = y_0$ to its final location $\bar{y}(\infty) = y_{eq}$; the width of the distribution growing to the equilibrium value.

Bearing this picture in mind, we have:

$$\frac{d}{dt} \bar{y}(t) = \int_{-\infty}^{\infty} y \frac{\partial}{\partial t} p(y, t) dy.$$

Using either the master equation or the Fokker-Planck equation, we have,

$$\frac{d}{dt} \bar{y}(t) = \int_{-\infty}^{\infty} a_1(y) p(y, t) dy = \langle a_1(y) \rangle_t, \quad (6.13)$$

which we identify as the *macroscopic equation*.

Example: One-step process (see p. 62). The discrete master equation reads,

$$\frac{d}{dt} p_n = r_{n+1} p_{n+1} + g_{n-1} p_{n-1} - (r_n + g_n) p_n. \quad (6.14)$$

Multiplying both sides by n , and summing over all $n \in (-\infty, \infty)$, we obtain the evolution equation for the average n ,

$$\frac{d\langle n \rangle}{dt} = \langle g_n \rangle - \langle r_n \rangle. \quad (6.15)$$

For example, with $r_n = \beta n$ and $g_n = \alpha$, Eq. 6.15 reduces to:

$$\frac{d\langle n \rangle}{dt} = \alpha - \beta \langle n \rangle.$$

Note in this last example, we had $\langle r_n \rangle_t = \langle r(n) \rangle_t = r(\langle n \rangle_t)$ and similarly for g_n . The situation illustrated by this example holds in general:

Proposition: If the function $a_1(y)$ is *linear* in y , then the macroscopic variable $\bar{y}(t)$ satisfies the macroscopic equation,

$$\frac{d}{dt} \bar{y}(t) = a_1(\bar{y}), \quad (6.16)$$

which follows exactly from the master equation.

If, however, $a_1(y)$ is *nonlinear* in y , then Eq. 6.16 $\neq$ 6.13, and the macroscopic equation is no longer determined uniquely by the initial value of $\bar{y}$. Not only, but one must also specify what “macroscopic equation” means. *In the literature, one finds Eq. 6.16 used even when $a_1(y)$ is nonlinear*; the (usually-not-stated-) assumption is that since the fluctuations are small, and $a_1(y)$ is smooth, we may expand it about $\langle y \rangle_t = \bar{y}$,

$$a_1(y) = a_1(\bar{y}) + a_1'(\bar{y}) \cdot (y - \bar{y}) + \frac{1}{2} a_1''(\bar{y}) \cdot (y - \bar{y})^2 + \dots$$

Taking the average, we have,

$$\begin{aligned} \langle a_1(y) \rangle_t &= a_1(\bar{y}) + \frac{1}{2} a_1''(\bar{y}) \cdot \langle (y - \bar{y})^2 \rangle_t + \dots \\ \langle a_1(y) \rangle_t &\approx a_1(\bar{y}), \end{aligned}$$

on the grounds that the fluctuations $\langle (y - \bar{y})^2 \rangle_t$ are small. In that case, we read Eq. 6.16 as an approximation,

$$\frac{d}{dt} \bar{y}(t) \approx a_1(\bar{y}), \quad (6.17)$$

if $a_1(y)$ is nonlinear – this is the “meaning” that is assigned to the macroscopic equation.

It is also possible to deduce from the master equation (or the Fokker-Planck equation) an approximate evolution for the width of the distribution; first, one shows that

$$\frac{d\langle y^2 \rangle_t}{dt} = \langle a_2(y) \rangle_t + 2\langle y a_1(y) \rangle_t,$$

from which it follows that the variance $\sigma^2(t) = \langle y^2 \rangle_t - \langle y \rangle_t^2$ obeys,

$$\frac{d\sigma^2}{dt} = \langle a_2(y) \rangle_t + 2\langle (y - \bar{y}) a_1(y) \rangle_t. \quad (6.18)$$

Again, if a_1 and a_2 are *linear* in y , then Eq. 6.18 is identical with,

$$\frac{d\sigma^2}{dt} = a_2(\bar{y}) + 2\sigma^2 a_1'(\bar{y}), \quad (6.19)$$

though in general, Eq. 6.19 will be only an approximation. This equation for the variance may now be used to compute corrections to the macroscopic equation, Eq. 6.17,

$$\frac{d}{dt} \bar{y}(t) = a_1(\bar{y}) + \frac{1}{2} \sigma^2 a_1''(\bar{y}), \quad (6.20)$$

$$\frac{d\sigma^2}{dt} = a_2(\bar{y}) + 2\sigma^2 a_1'(\bar{y}). \quad (6.21)$$

Note that, by definition, $a_2(y) > 0$, and for the system to evolve toward a *stable* steady state, one needs $a'_1(y) < 0$. It follows that σ^2 tends to increase at the rate a_2 , but this tendency is kept in check by the second term; hence,

$$\sigma^2 \rightarrow \frac{a_2}{2|a'_1|}, \quad (6.22)$$

and so the condition for the approximate validity of Eq. 6.17 – namely that the 2^{nd} term in Eq. 6.20 be small compared to the first – is given by,

$$\frac{a_2}{2|a'_1|} \frac{1}{2} |a''_1| \ll |a_1|, \quad (6.23)$$

which says that it is the second derivative of a_1 (responsible for the departure from linearity) which must be small. The linear noise approximation, which we shall consider in Chapter 4 (Section 5.1) provides a more satisfying derivation of the macroscopic evolution from the master equation, proceeding as it does from a systematic expansion in some well-defined small parameter.

6.3.1 Coefficients of the Fokker-Planck Equation

The coefficients of the Fokker-Planck equation,

$$\frac{\partial}{\partial \tau} p(y, \tau) \approx -\frac{\partial}{\partial y} [a_1(y) p(y, \tau)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [a_2(y) p(y, \tau)], \quad (6.24)$$

are given by,

$$a_1(y) = \int_{-\infty}^{\infty} \Delta y w(y, \Delta y) d\Delta y$$

$$a_2(y) = \int_{-\infty}^{\infty} \Delta y^2 w(y, \Delta y) d\Delta y.$$

In theory, they can be computed explicitly since $w(y, \Delta y)$ is known from an underlying microscopic theory. In practice, that is not always easy to do, and an alternative method would be highly convenient. The way one usually proceeds is the following:

Let $p(y, t|y_0, t_0)$ be the solution of Eq. 6.24 with initial condition $\delta(y - y_0)$. According to Eq. 3.4, *i.e.*,

$$p(y_2, t_2) = \int_{-\infty}^{\infty} p(y_2, t_2|y_1, t_1) p(y_1, t_1) dy_1,$$

one may construct a Markov process with transition probability $p(y_2, t_2|y_1, t_1)$ whose one-time distribution $p(y_1, t_1)$ may still be chosen arbitrarily at one initial time t_0 . If one chooses the “steady-state” solution of Eq. 6.24,

$$p_s(y) = \frac{\text{constant}}{a_2(y)} \exp \left[2 \int^y \frac{a_1(y')}{a_2(y')} dy' \right], \quad (6.25)$$

then the resulting Markov process is stationary. This, of course, is only possible if p_s is integrable. For closed systems, Eq. 6.25 may be identified with the equilibrium probability distribution p_{eq} known from statistical mechanics.

To “derive” the coefficients of the Fokker-Planck equation, we proceed as follows. From Eq. 6.24, the macroscopic equation is,

$$\frac{d}{dt} \langle y \rangle_t = \langle a_1(y) \rangle_t \approx a_1(\langle y \rangle_t),$$

(neglecting fluctuations). Since this equation *must coincide* with the equation known from the phenomenological theory, the function a_1 is determined. Next, we know p_{eq} from statistical mechanics, identified with Eq. 6.25; hence a_2 is also determined. Note the various (for the most part uncontrolled) assumptions made in this derivation.

The simplest of all procedures, however, is the derivation based upon the Langevin equation. Recall that this begins with the dynamical description,

$$\frac{d\bar{y}}{dt} + \beta\bar{y} = f(t) \quad (6.26)$$

with assumptions about the statistics of the fluctuating force $f(t)$,

$$\langle f(t) \rangle = 0, \langle f(t_1)f(t_2) \rangle = 2D\delta(t_1 - t_2). \quad (6.27)$$

Integrating Eq. 6.26 over a short time interval,

$$\Delta\bar{y} = -\beta\bar{y}\Delta t + \int_t^{t+\Delta t} f(\xi) d\xi,$$

so that,

$$a_1(\bar{y}) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle \Delta \bar{y} \rangle = -\beta \bar{y} + \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle f(\xi) \rangle d\xi \rightarrow 0$$

Similarly,

$$\langle (\Delta \bar{y})^2 \rangle = \beta^2 \bar{y}^2 (\Delta t)^2 + \int_t^{t+\Delta t} \int \langle f(\xi) f(\eta) \rangle d\xi d\eta,$$

from which it follows – using Eq. 6.27 – that

$$a_2(\bar{y}) = 2D,$$

and the corresponding Fokker-Planck equation reads,

$$\frac{\partial p}{\partial t} = \beta \frac{\partial}{\partial y} (yp) + D \frac{\partial^2 p}{\partial y^2}. \quad (6.28)$$

This coincides with the appropriate Fokker-Planck equation for the position of a Brownian particle in the absence of an external field, as derived by Ornstein and Uhlenbeck.

The identification of $a_1(y)$ as the macroscopic rate law is really only valid when $a_1(y)$ is a *linear* function of y – as it is in the Ornstein-Uhlenbeck process described above. For *nonlinear* systems, derivation of the Fokker-Planck equation in the manner described here can lead to serious difficulties, and so the systematic expansion scheme described in Section 5.1 is indispensable. See N. G. van Kampen (1965) *Fluctuations in Nonlinear Systems*.

6.4 Pure Diffusion and Ornstein-Uhlenbeck Processes

As an example of the Fokker-Planck equation, we shall consider two families of diffusive equations - pure diffusion, and diffusion with drift (which is the same as the Ornstein-Uhlenbeck process shown above, Eq. 6.28).

6.4.1 Wiener process

First consider the simple case with $a_1 = 0, a_2 = k$, where k is some (positive) constant. Then we have the problem,

$$\frac{\partial p}{\partial t} = k \frac{\partial^2 p}{\partial y^2}; \quad \text{with} \quad p(y|y_0, 0) = \delta(y - y_0). \quad (6.29)$$

This is a parabolic partial differential equation with constant coefficients, so it is readily solved using a Fourier transform in y (see section B.3.2 on page 262),

$$\begin{aligned} \phi(s, t) &= \int_{-\infty}^{\infty} e^{isy} p(y, t) dy; \\ \phi(s, 0) &= e^{isy_0}. \end{aligned}$$

The transformed Eq. 6.29 leads to the ordinary differential equation in ϕ ,

$$\frac{d\phi(s, t)}{dt} = -s^2 k \phi(s, t),$$

whose solution is,

$$\phi(s, t) = \exp[isy_0 - ks^2 t].$$

This is a very famous Fourier transform that appears in many models of heat conduction. The inverse Fourier transform back to $p(y, t)$ yields what is sometimes called the ‘heat kernel’,

$$p(y, t) = \frac{1}{\sqrt{4\pi kt}} \exp \left[-\frac{(y - y_0)^2}{4kt} \right]. \quad (6.30)$$

In particular, for $k = D$, this coincides with Einstein’s result for a *free* particle in Brownian motion (*i.e.* in the absence of a damping force),

$$\frac{dW}{dt} = \sqrt{2D} \eta(t), \quad (6.31)$$

sometimes also called a *Wiener process* (denoted $W(t)$), or a *purely diffusive process*, or sometimes even called *Brownian motion*, although that designation is not very descriptive.

Because the forcing function $\eta(t)$ is delta-correlated, $\langle \eta(t_1)\eta(t_2) \rangle = \delta(t_1 - t_2)$, the autocorrelation function of the Wiener process is straightforward to compute, *viz.*,

$$\begin{aligned} \langle W(t_1)W(t_2) \rangle &= 2D \int_0^{t_1} \int_0^{t_2} \delta(u_1 - u_2) du_2 du_1, \\ \langle W(t_1)W(t_2) \rangle &= 2D \min(t_1, t_2). \end{aligned} \quad (6.32)$$

6.4.2 Ornstein-Uhlenbeck process

Now consider the more general Fokker-Planck equation, $a_1 = -ky$, $a_2 = D$, Eq. 6.28,

$$\frac{\partial p}{\partial t} = k \frac{\partial}{\partial y}(yp) + \frac{D}{2} \frac{\partial^2 p}{\partial y^2}, \quad (6.33)$$

where k and D are (positive) constants. Proceeding as above, taking the Fourier transform, we are left with a first-order linear *partial* differential equation,

$$\frac{\partial \phi}{\partial t} + ks \frac{\partial \phi}{\partial s} = -\frac{D}{2} s^2 \phi. \quad (6.34)$$

This is easily solved by the method of characteristics to give the partial solution,

$$\phi(s, t) = e^{-Ds^2/4k} g(se^{-kt}), \quad (6.35)$$

in terms of an arbitrary function g . From the initial distribution $p(y|y_0, 0) = \delta(y - y_0)$, the arbitrary function g is fully determined:

$$g(x) = \exp \left[\frac{Dx^2}{4k} + iy_0 x \right]. \quad (6.36)$$

The complete solution is therefore,

$$\phi(s, t) = \exp \left[-\frac{Ds^2}{4k} (1 - e^{-2kt}) + isy_0 e^{-kt} \right]. \quad (6.37)$$

Comparison with the solution above for the purely diffusive process shows that the density is a Gaussian with

$$\begin{aligned} \langle Y(t) \rangle &= y_0 e^{-kt}, \\ \sigma^2 &= \frac{D}{2k} (1 - e^{-2kt}), \end{aligned} \quad (6.38)$$

as derived by Ornstein and Uhlenbeck (see Chapter 1).

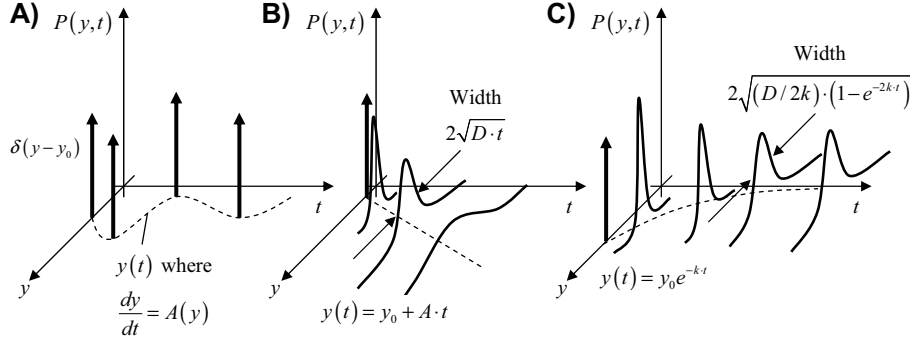


Figure 6.3: **Some solutions of the Fokker-Planck equation, (6.39).** **A) Liouville equation, $a_2 = 0$.** Without diffusion, the initial delta-distribution follows the deterministic trajectory. **B) Wiener process, $a_1 = A, a_2 = D$, both constant.** The solution is a Gaussian, but the variance spreads $\propto t$. **C) Ornstein-Uhlenbeck process, $a_1 = -ky, a_2 = D$.** The solution is Gaussian, relaxing to an equilibrium distribution around $y = 0$. It should be understood that in this figure $P(y, t) \equiv P(y, t|y_0, 0)$.

6.4.3 Heuristic Picture of the Fokker-Planck Equation

Examining several limiting cases of the Fokker-Planck equation is useful in developing intuition regarding the behaviour of the solution. Here, again for convenience, is the general Fokker-Planck equation,

$$\frac{\partial}{\partial \tau} p(y, \tau) = -\frac{\partial}{\partial y} [a_1(y) p(y, \tau)] + \frac{1}{2} \frac{\partial^2}{\partial y^2} [a_2(y) p(y, \tau)]. \quad (6.39)$$

We begin with the same initial condition $p(y, 0) = \delta(y - y_0)$, then by choosing different coefficients $a_1(y)$ and $a_2(y)$, certain characteristics of the solution can be made clear (Figure 6.3).

1. **Liouville equation** ($a_1(y) = A(y), a_2(y) = 0$). The probability remains a delta-peak, localized about $y(t)$, *i.e.*, $p(y, t|y_0, 0) = \delta(y - y(t))$, where $y(t)$ obeys the deterministic ordinary differential equation,

$$\frac{dy}{dt} = A(y); \quad y(0) = y_0.$$

2. **Wiener process** ($a_1(y) = A$, $a_2(y) = D$). This is the equation governing pure diffusion. The mean follows the line $y(t) = y_0 + A \cdot t$, and the variance grows linearly with t , $\langle y^2 \rangle - \langle y \rangle^2 = 2Dt$.
3. **Ornstein-Uhlenbeck process** ($a_1(y) = -ky$, $a_2(y) = D$). The coefficient $a_1(y) = -ky$ is like a Hooke's force restoring the mean state to $y = 0$ exponentially quickly,

$$\langle y(t) \rangle = y_0 e^{-kt}.$$

The variance grows until it reaches the equilibrium value $\langle y^2 \rangle = D/2k$. The equilibrium distribution is the Gaussian centered at $y = 0$ with variance $\langle y^2 \rangle = D/2k$, viz.

$$p^\infty(y) = \frac{1}{\sqrt{2\pi \cdot (D/2k)}} \exp \left[-\frac{y^2}{2 \cdot (D/2k)} \right].$$

6.5 Connection to Langevin Equation

From the work of Einstein and Langevin, there is clearly a direct analogy between the Fokker-Planck and Langevin equations in the case of Brownian motion. That direct correspondence holds in general and can be uniquely assigned for Langevin equations where the coefficient of the noise term is constant. For non-constant coefficients, the correspondence is no longer unique and an additional interpretation rule for the Langevin equation must be provided.

We shall outline the connection between the Fokker-Planck and Langevin equations below, for the cases of linear and non-linear drift coefficients with the requirement that the noise coefficient is constant. Interpretation rules for the case where the noise coefficient is a function of the state variables will be postponed to Chapter 7.

6.5.1 Linear Damping

We shall review the general procedure for the *linear case*. Langevin begins with the dynamical equation

$$\frac{dy}{dt} = -\beta y + \eta(t), \quad (6.40)$$

where the random force $\eta(t)$ is supposed to have the properties

$$\langle \eta(t) \rangle = 0, \quad (6.41)$$

$$\langle \eta(t)\eta(t') \rangle = \Gamma \delta(t - t'). \quad (6.42)$$

Note that Eqs. 6.41 and 6.42 are not enough to fully specify the stochastic process $\eta(t)$ – they only specify the first two moments. Also note that because of the random forcing $\eta(t)$, the variable $y(t)$ becomes a stochastic process. Suppose that $y(0) = y_0$, we have from Eq. 6.40,

$$y(t) = y_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta t'} \eta(t') dt'. \quad (6.43)$$

Averaging over the ensemble and using Eq. 6.41,

$$\langle y(t) \rangle_{y_0} = y_0 e^{-\beta t}. \quad (6.44)$$

Moreover, after squaring, averaging and using Eq. 6.42, one gets,

$$\langle y^2(t) \rangle_{y_0} = y_0^2 e^{-2\beta t} + \frac{\Gamma}{2\beta} (1 - e^{-2\beta t}). \quad (6.45)$$

The constant Γ is unknown so far; however, for $t \gg 1/\beta$ the effect of the initial conditions must disappear and the system reaches equilibrium. Equilibrium statistical mechanics (the equipartition of energy) can be used to derive,

$$\frac{1}{2} m \langle y^2(\infty) \rangle_{y_0} = \frac{1}{2} k_B T, \quad (6.46)$$

where k_B is Boltzmann's constant and T is the equilibrium temperature. Taking $t \rightarrow \infty$ in Eq. 6.45 and comparing with the equilibrium result above, we have

$$\Gamma = \frac{2\beta}{m} k_B T, \quad (6.47)$$

an example of the so-called *Fluctuation-Dissipation Relation* (see Eq. 2.31 on page 43). Physically, the random force $\eta(t)$ creates a tendency for y to spread out over an ever broadening range of values, while the damping term tries to bring y back to the origin. The equilibrium distribution is the resulting compromise between these two opposing tendencies.

To derive the relation between the Langevin equation and the Fokker-Planck equation, take for t in Eq. 6.44 a *small* time interval Δt . Then,

$$\langle \Delta y \rangle_{y_0} = \langle y(t) \rangle_{y_0} - y_0 = -\beta y_0 \Delta t + O(\Delta t^2). \quad (6.48)$$

Therefore, the drift coefficient in the Fokker-Planck equation is,

$$a_1(y_0) = \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta y \rangle_{y_0}}{\Delta t} = -\beta y_0. \quad (6.49)$$

Similarly, from the square of Eq. 6.43,

$$\langle \Delta y^2 \rangle_{y_0} = \Gamma \Delta t + O(\Delta t^2), \quad (6.50)$$

to give $a_2(y) = \Gamma$. The resulting Fokker-Planck equation reads,

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial y} [(\beta y)p] + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\Gamma p). \quad (6.51)$$

This Fokker-Planck equation returns the same values for the first- and second-moments of $y(t)$ as does the Langevin equation (6.40), along with Eqs. 6.41 and 6.42; *however*, it cannot yet be said that they are equivalent because the higher moments do not agree. In fact, whereas the Fokker-Planck equation provides a definite expression for each moment, the Langevin equation does not unless higher moments of $\eta(t)$ are specified. Customarily, the assumption is

- All odd moments of $\eta(t)$ vanish.
- All even moments obey the same rule as for Gaussian distributions, *e.g.* decomposition into pair-wise products,

$$\begin{aligned} \langle \eta(t_1)\eta(t_2)\eta(t_3)\eta(t_4) \rangle &= \langle \eta(t_1)\eta(t_2) \rangle \langle \eta(t_3)\eta(t_4) \rangle + \\ &+ \langle \eta(t_1)\eta(t_3) \rangle \langle \eta(t_2)\eta(t_4) \rangle + \langle \eta(t_1)\eta(t_4) \rangle \langle \eta(t_2)\eta(t_3) \rangle = \\ &\Gamma^2 [\delta(t_1 - t_2)\delta(t_3 - t_4) + \delta(t_1 - t_3)\delta(t_2 - t_4) + \delta(t_1 - t_4)\delta(t_2 - t_3)]. \end{aligned} \quad (6.52)$$

Alternatively, one may stipulate that all higher *cumulants* of η vanish beyond the second. The equivalence between the Fokker-Planck and Langevin descriptions then holds: According to Eq. 6.43, the value of $y(t)$ is a linear combination of the values that $\eta(t)$ takes at all previous times ($0 \geq t' \geq t$). Since the joint distribution of all quantities $\eta(t')$ is Gaussian, it follows that $y(t)$ is Gaussian. By the same argument, the joint distribution of $y(t_1), y(t_2), \dots$, is Gaussian. Hence the process $y(t)$ determined by Eq. 6.40 with initial condition y_0 is Gaussian. On the other hand, we know that the solution of the Fokker-Planck equation (6.51) with initial value y_0 is also a Gaussian. Since its coefficients have been chosen so that the first- and second-moments of both Gaussians are the same, it follows that the two Gaussians are identical. (Note that noise $\eta(t)$ defined as above is called *Gaussian white-noise*.)

6.5.2 Nonlinear Damping

Another case occurring frequently in applications is when the Langevin equation contains a nonlinear damping term,

$$\frac{dy}{dt} = A(y) + \eta(t), \quad (6.53)$$

where the random force is still assumed to be Gaussian white-noise, but the damping term $A(y)$ is a nonlinear function of y . Although in this case an explicit solution for $y(t)$ is not usually available, it can still be argued that the Langevin equation is equivalent to the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y} [A(y)p] + \frac{1}{2} \frac{\partial^2}{\partial y^2} (\Gamma p). \quad (6.54)$$

Outline of the proof: First, it is clear that for each sample function $\eta(t)$, Eq. 6.53 along with the initial condition $y(0) = y_0$ uniquely determines $y(t)$ from the existence/uniqueness theorem for differential equations. Since the values of $\eta(t)$ at different times are independent, it follows that $y(t)$ is Markovian. Hence, it obeys a master equation which may be written in the Kramers-Moyal form (see Eq. 6.11 on page 126).

Next, for a very small Δt , Eq. 6.53 gives

$$\Delta y = \int_t^{t+\Delta t} A(y(t')) dt' + \int_t^{t+\Delta t} \eta(t') dt'; \quad (6.55)$$

hence, upon averaging,

$$\langle \Delta y \rangle_{y_0} = A(y_0) \Delta t + O(\Delta t^2), \quad (6.56)$$

to yield $a_1(y) = A(y)$ as before. Taking the square of Eq. 6.55 and averaging,

$$\begin{aligned} \langle \Delta y^2 \rangle &= \left\langle \left(\int_t^{t+\Delta t} A(y(t')) dt' \right)^2 \right\rangle \\ &+ 2 \int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle A(y(t')) \eta(t'') \rangle dt'' \right) dt' \\ &+ \int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle \eta(t') \eta(t'') \rangle dt'' \right) dt'. \end{aligned}$$

The first line is $O(\Delta t^2)$, so it does not contribute to $a_2(y)$. The third line gives,

$$\int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle \eta(t') \eta(t'') \rangle dt'' \right) dt' = \int_t^{t+\Delta t} \Gamma dt' = \Gamma \Delta t, \quad (6.57)$$

as in the linear case. This agrees with the last term of Eq. 6.54, provided the second term above does not contribute. To untangle the second term, expand $A(y(t'))$ about $y(t)$,

$$A(y(t')) = A(y(t)) + A'(y(t))(y(t') - y(t)) + \dots$$

Substituting this in the second term above,

$$2 \int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle A(y(t')) \eta(t'') \rangle dt'' \right) dt' = \quad (6.58)$$

$$2A(y(t))\Delta t \int_t^{t+\Delta t} \langle \eta(t'') \rangle dt'' + 2A'(y(t)) \int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle [y(t') - y(t)] \cdot \eta(t'') \rangle dt'' \right) dt' + \dots, \quad (6.59)$$

the last term of which is $O(\Delta t^2)$, and therefore does not contribute to $a_2(y)$. Similarly, one can show that all higher-order moments contribute nothing, since $\langle (\Delta y)^n \rangle = o(\Delta t)$. That concludes the outline of the proof.

In summary, the Langevin equation,

$$\frac{dy}{dt} = A(y) + c\eta(t),$$

(with $\langle \eta(t)\eta(t') \rangle = \Gamma\delta(t-t')$ and the higher-moments of $\eta(t)$ suitably defined) is equivalent to the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y} [A(y)p] + \frac{c^2\Gamma}{2} \frac{\partial^2 p}{\partial y^2}.$$

irrespective of the form of $A(y)$. Difficulties emerge once $c(y)$ is no longer constant. We can proceed naïvely, and introduce the change of variables,

$$\hat{y} = \int \frac{dy}{c(y)}, \quad \frac{A(y)}{c(y)} = \hat{A}(\hat{y}), \quad \hat{P}(\hat{y}) = P(y)c(y), \quad (6.60)$$

transforming the nonlinear Langevin equation¹,

$$\frac{dy}{dt} = A(y) + c(y)\eta(t), \quad (6.61)$$

to the Langevin equation described above,

$$\frac{d\hat{y}}{dt} = \hat{A}(\hat{y}) + \eta(t). \quad (6.62)$$

This equation, in turn, is equivalent to the Fokker-Planck equation,

$$\frac{\partial \hat{P}}{\partial t} = -\frac{\partial}{\partial \hat{y}} \hat{A}(\hat{y}) \hat{P} + \frac{\Gamma}{2} \frac{\partial^2 \hat{P}}{\partial \hat{y}^2}. \quad (6.63)$$

In the original variables,

$$\frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y} \left[A(y) + \frac{\Gamma}{2} c(y) c'(y) \right] P(y, t) + \frac{\Gamma}{2} \frac{\partial^2}{\partial y^2} [c^2(y) P(y, t)]. \quad (6.64)$$

There is a problem – Eq. 6.61 as it stands *has no meaning*! The difficulty comes in trying to interpret the product $c(y)\eta(t)$. Since $\eta(t)$ causes a jump in y , what value of y should be used in $c(y)$? Specifically, how is one to interpret the $\int c(y(t'))\eta(t')dt'$ term in the equivalent integral equation,

$$y(t + dt) - y(t) = \int_t^{t+dt} A(y(t')) dt' + \int_t^{t+dt} c(y(t')) \eta(t') dt'?$$

Stratonovich chose to replace $c(y(t'))$ by its mean value over the interval, giving,

$$y(t + dt) - y(t) = \int_t^{t+dt} A(y(t')) dt' + c\left(\frac{y(t) + y(t + dt)}{2}\right) \int_t^{t+dt} \eta(t') dt'. \quad (6.65)$$

One can prove that this interpretation indeed leads to Eq. 6.64. Other choices are possible, generating different transformation laws. The calculus of stochastic functions is the focus of the next chapter, but first we shall turn attention to a classic application of the Fokker-Planck equation to the modeling of a physical system – Kramers escape over a potential barrier.

¹*Nonlinear Langevin equation* is the designation usually given to any equation of the Langevin type where $c(y)$ is non-constant, irrespective of the (non)linearity of $A(y)$.

6.6 Limitations of the Kramers-Moyal Expansion

The Fokker-Planck equation, as derived in Chapter 6, remains the most popular approximation method to calculate a solution for the probability distribution of fluctuations in nonlinear systems. Typically, the Fokker-Planck equation is generated through a two-term Kramers-Moyal expansion of the master equation (see Eq. 6.11 on page 126). The result is a Fokker-Planck equation with nonlinear drift and diffusion coefficients. There are several problems with this approach – First, the expansion is not *consistent* since the step-operator is approximated by a continuous operator, but the transition probabilities *remain as nonlinear functions of the microscopic variables*. Second, the nonlinear Fokker-Planck equation has no general solution for state dimensions greater than one. That is, as soon as the system of interest has more than one variable, the nonlinear Fokker-Planck equation must be solved numerically or by some specialized approximation methods. Since the nonlinear Fokker-Planck equation derived by the Kramers-Moyal expansion is consistent only insofar as it agrees with the linear Fokker-Planck equation derived using the linear noise approximation, and since the nonlinear Fokker-Planck equation is so difficult to solve, the Kramers-Moyal expansion seems an undesirable approximation method.

For a more detailed discussion, see:

- N. G. van Kampen (1981) “The validity of non-linear Langevin-equations,” *Journal of Statistical Physics* **25**: 431.
- N. G. van Kampen, “The diffusion approximation for Markov processes,” in *Thermodynamics and kinetics of biological processes*, Ed. I. Lamprecht and A. I. Zotin (Walter de Gruyter & Co., 1982).

6.7 Example – Kramer’s Escape Problem

- Kramers, HA (1940) “Brownian motion in a field of force and the diffusion model of chemical reactions,” *Physica* **7**: 284.
- D. ter Haar, *Master of Modern Physics: The Scientific Contributions of H. A. Kramers*. Princeton University Press, 1998. Chapter 6.
- P. Hänggi, P. Talkner and M. Borkovec (1990) “Reaction-rate theory: 50 years after Kramers,” *Reviews of Modern Physics* **62**: 251.

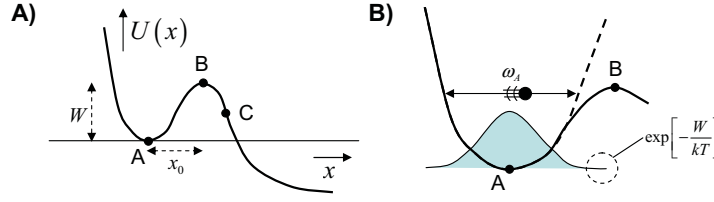


Figure 6.4: **Kramers Escape from a Potential Well.** **A)** Potential field with smooth barrier. Redrawn from Kramers (1940). **B)** The steady-state probability distribution is used, normalized as though the barrier at $x = B$ were not there. The escape time is computed by calculating the rate at which particles pass $X = B$ with positive velocity.

Briefly, Kramers escape problem is the following: Consider a particle trapped in a potential well, subject to Brownian motion (Figure 6.4). What is the probability that the particle will escape over the barrier? Among the many applications of this model, Kramers suggested that the rates of chemical reactions could be understood via this mechanism.

Imagine a Brownian particle subject to a position-dependent force $F(x)$, in addition to the damping and fluctuations of Brownian motion,

$$m \frac{d^2 X}{dt^2} = -\gamma \frac{dX}{dt} + F(X) + \eta(t), \quad (6.66)$$

Here, $\eta(t)$ is the same Gaussian white-noise forcing that appears in Langevin's equation. Suppose the force can be written as the gradient of a potential $U(X)$ (as is always the case in one-dimension),

$$F(X) = -U'(X).$$

Kramers had in mind a two-welled potential, as depicted in Figure 6.4a. To derive the associated Fokker-Planck equation, it is convenient to re-write Eq. 6.66 explicitly in terms of the position X and the velocity V ,

$$\begin{aligned} \frac{dX}{dt} &= V, \\ m \frac{dV}{dt} &= -\gamma V - U'(X) + \eta(t). \end{aligned}$$

We are after a Fokker-Planck equation that describes the bivariate probability distribution $P(X, V, t)$. The coefficients for the drift are straight-

forward to calculate (see Section 6.3.1),

$$\lim_{\Delta t \rightarrow 0} \frac{\langle \Delta X \rangle}{\Delta t} = V, \quad \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta V \rangle}{\Delta t} = \left\{ \frac{F(X)}{m} - \frac{\gamma}{m} V \right\}. \quad (6.67)$$

It is likewise straightforward to show that the diffusion coefficients $\langle \Delta X^2 \rangle$ and $\langle \Delta X \Delta V \rangle$ vanish,

$$\lim_{\Delta t \rightarrow 0} \frac{\langle \Delta X^2 \rangle}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta X \Delta V \rangle}{\Delta t} = 0. \quad (6.68)$$

For the $\langle \Delta V^2 \rangle$ -term, we must appeal to statistical mechanics which says that the equilibrium distribution for the velocity is given by the Maxwell distribution,

$$P^e(V) = \left(\frac{m}{2\pi kT} \right)^{1/2} \exp \left[-\frac{m}{2kT} V^2 \right]. \quad (6.69)$$

In that way, one can show (Exercise 9a),

$$\lim_{\Delta t \rightarrow 0} \frac{\langle \Delta V^2 \rangle}{\Delta t} = \gamma \frac{kT}{m}. \quad (6.70)$$

The full Fokker-Planck equation is then,

$$\frac{\partial P(X, V, t)}{\partial t} = -V \frac{\partial P}{\partial X} + \frac{U'(X)}{m} \frac{\partial P}{\partial V} + \frac{\gamma}{m} \left\{ \frac{\partial}{\partial V} V P + kT \frac{\partial^2 P}{\partial V^2} \right\}, \quad (6.71)$$

solved subject to the initial condition $P(X, V, 0) = \delta(X - A)\delta(V - 0)$. As a first estimate, we assume the well-depth W is large compared with kT . We are then justified in using the *stationary* distribution for $P(X, V, t)$ around the point $X = A$,

$$P^s(X, V) = \mathcal{N} \cdot \exp \left[-\frac{U(X) + \frac{1}{2}V^2}{kT} \right]. \quad (6.72)$$

There is escape over the barrier, so obviously $P^s(X, V)$ is not correct for long times, but over short times, the probability is *very* small near the top of the barrier $X = B$ (Figure 6.4b). We can then set $P^s(X', V) = 0$ for $X' > B$. The normalization constant $\mathcal{N}$ is approximately,

$$\mathcal{N}^{-1} = \sqrt{2\pi \frac{kT}{m}} \int_{-\infty}^B e^{-U(x)/kT} dx \approx \frac{2\pi kT}{\omega_a} e^{-U(A)/kT},$$

where $\omega_a^2 = U''(A)$ is the *curvature* at the base of the well, $X = A$. The escape rate comes from calculating the outward flow across the top of the barrier, at $X = B$,

$$\frac{1}{\tau} = \int_0^\infty V P^s(X, V) dV = \mathcal{N} e^{-U(B)/kT} \int_0^\infty V e^{-V^2/2kT} dV = \frac{\omega_a^2}{2\pi} e^{-W/kT}. \quad (6.73)$$

The underlying assumption is that if the particle is at the top of the barrier with positive velocity, then it will escape – as if there were an absorbing barrier at $X = B$. A heuristic interpretation of Eq. 6.73 is that the particle oscillates in a potential well $\frac{1}{2}\omega_a^2 X^2$ and therefore hits the wall $\omega_a^2/2\pi$ times per second, each time with a probability of $e^{-W/kT}$ to get across (Figure 6.4b).

In a more sophisticated treatment, the Fokker-Planck Eq. 6.71 is solved with an initial distribution inside the well and the escape rate is determined by the flow past a point $X = C$ *sufficiently far from the top of the barrier that the probability of return is negligible* (Figure 6.4a). This is Kramer's escape problem. There is no known solution for a potential well of the shape drawn in Figure 6.4a, so clever approximation methods are the only recourse for estimating the escape time. In Exercise 11, an approximate solution is developed in the limit of large friction, $\gamma \gg 1$.

Suggested References

The encyclopedic reference for applications of the Fokker-Planck equation along with an exhaustive detailing of existing solution methods is the following text by Risken:

- *The Fokker-Planck Equation: Methods of Solutions and Applications (2nd Ed.)*, H. Risken (Springer, 1996).

For all kinds of first-passage and escape problems,

- *A Guide to First-Passage Processes*, S. Redner (Cambridge University Press, 2001).

is recommended.

Exercises

1. **One-step process:** Fill in the details in the derivation of Eq. 6.15 from Eq. 6.14.
2. **Autocorrelation of the Wiener process.** Fill in the details leading to Eq. 6.32. It may be helpful to write the Dirac delta function as the derivative of the unit-step function, and integrate by parts.
3. **Deterministic equations:** Starting from the Fokker-Planck equation,

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} [\alpha_1(x, t)f(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\alpha_2(x, t)f(x, t)],$$

- (a) Derive the deterministic (macroscopic) rate equations for $\langle X(t) \rangle$ and $\langle X^2(t) \rangle$.
- (b) If $\alpha_1(x) = a(x) + \frac{1}{4}\alpha'_2(x)$ and $\alpha_2(x) = b^2(x)$, derive

$$\frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} [a(x)f(x, t)] + \frac{1}{2} \frac{\partial}{\partial x} \left\{ b(x) \frac{\partial}{\partial x} [b(x)f(x, t)] \right\}.$$

4. **Deterministic dynamics as a Markov process:** Consider the deterministic differential equation $dy/dt = f(y)$.
 - (a) Show that the process $y(t)$ is a Markov process.
 - (b) Derive the associated Fokker-Planck equation (called the Liouville equation). Show that if $y(t)$ satisfies the Liouville equation, then it likewise satisfies the original differential equation.
5. **Stratonovich Fokker-Planck equation.** Fill in the missing details that lead from the change of variables, Eq. 6.60, to the Fokker-Planck equation, Eq. 6.64.
6. **Diffusion in a wedge:** For a Brownian particle diffusing in an infinite wedge with absorbing edges (Figure 6.5), compute the probability that a particle starting from (r_0, θ_0) (where θ_0 is the angle with the horizontal) is absorbed by the horizontal edge of the wedge.
7. **Langevin equation with nonlinear drift:** In Eq. 6.59, show that the last term,

$$2A'(y(t)) \int_t^{t+\Delta t} \left(\int_t^{t+\Delta t} \langle [y(t') - y(t)] \cdot \eta(t'') \rangle dt'' \right) dt'$$

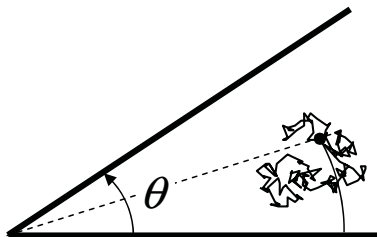


Figure 6.5: **Diffusion in a wedge.** An infinite wedge forming angle θ with the horizontal. The edges are absorbing. What is the probability that a Brownian particle beginning at (r_0, θ_0) is absorbed by the horizontal edge?

is indeed $O(\Delta t^2)$ as claimed in the text.

8. **First-passage time:** Consider the diffusion process $X(t)$ in one-dimension on a finite-interval $(0, L)$.

(a) Solve the Fokker-Planck equation for the process,

$$\frac{\partial f(x, t)}{\partial x} = D \frac{\partial^2 f(x, t)}{\partial x^2},$$

for $t \geq 0$, on the interval $0 \leq x \leq L$, with the initial condition $f(x, 0) = \delta(x - x_0)$ (where $x_0 \in (0, L)$), with *reflecting* boundary conditions at $x = 0$ and $x = L$ (no-flux boundary conditions). *Hint:* Your answer will be given in terms of a cosine Fourier series.

- (b) What is the stationary probability density function for the process $X(t)$?
- (c) For *absorbing* boundary conditions ($f(0, t) = f(L, t) = 0$), the *trapping-probability* $T_i(t)$ is the flux of probability through the boundary point at i . Find the *splitting probability* S_i that a particle beginning at x_0 will be eventually trapped at boundary point i . Find the average lifetime of the particle starting at $x = x_0$ before it is trapped at $x = 0$, $x = L$, or unconditionally trapped at either boundary.
- (d) Show that for $L \rightarrow \infty$, the probability distribution for diffusion on the positive half of the x -axis, with *reflecting* boundary

conditions at $x = 0$, can be written as

$$f(x, t) = \frac{1}{\sqrt{4\pi Dt}} \left[e^{-\frac{(x-x_0)^2}{4Dt}} + e^{-\frac{(x+x_0)^2}{4Dt}} \right].$$

If the boundary condition is *absorbing* at $x = 0$, how is the solution changed? Provide a physical interpretation of the distribution with absorbing conditions at $x = 0$.

- (e) In the semi-infinite domain ($L \rightarrow \infty$), suppose the $x = 0$ boundary is absorbing. Show that for a particle starting at $x = x_0$ trapping at $x = 0$ is certain, but that the average lifetime is infinite! This is an example of Zermelo's paradox (see p. 67).
9. **Rayleigh particle:** In 1891 (some 15 years before Einstein's work), Lord Rayleigh published his study of the motion of a particle buffeted by collisions with the surrounding gas molecules. The motion is studied on a timescale over which the *velocity* relaxes (a finer time scale than Einstein's study of Brownian motion). In one dimension, the macroscopic equation for the velocity is given by the damping law,

$$\frac{dV}{dt} = -\gamma V. \quad (6.74)$$

- (a) Derive the Fokker-Planck equation governing the distribution for the velocity $P(V, t)$. The drift coefficient is easily found,

$$\lim_{\Delta t \rightarrow 0} \frac{\langle \Delta V \rangle}{\Delta t} = -\gamma V.$$

From statistical mechanics, the equilibrium distribution for the velocity is given by the Maxwell distribution,

$$P^e(V) = \left(\frac{m}{2\pi kT} \right)^{1/2} \exp \left[-\frac{m}{2kT} V^2 \right]. \quad (6.75)$$

Use this to derive the diffusion coefficient, and thereby the full Fokker-Planck equation governing $P(V, t)$.

- (b) For the initial condition $V(0) = V_0$, find $\langle V(t) \rangle$ and $\langle \langle V(t)V(t+\tau) \rangle \rangle$.
10. **Brownian motion in a magnetic field:** A Brownian particle with mass m and charge q moves in three-dimensional space spanned

by a rectangular coordinate system xyz under the influence of a constant, homogeneous magnetic field $\vec{B}$ directed along the z -axis, $\vec{B} = (0, 0, B)$. The velocity vector $\vec{V}(t) = (V_x(t), V_y(t), V_z(t))^T$ satisfies the 3-component Langevin equation,

$$m \frac{d\vec{V}}{dt} = -\beta\vec{V}(t) - q\vec{B} \times \vec{V}(t) + \sqrt{2\beta k_B T} \vec{\Gamma}(t), \quad (6.76)$$

where β is the friction coefficient, and $\vec{\Gamma}(t) = (\Gamma_x(t), \Gamma_y(t), \Gamma_z(t))^T$ is a vector of Gaussian white noise with independent components.

- (a) Assuming over-damped motion, neglect the inertial term in Eq. 6.76 and derive the Langevin equation for the position $\vec{R}(t) = (X(t), Y(t), Z(t))^T$ in matrix form,

$$\frac{d\vec{R}}{dt} = \mathbf{C} \cdot \vec{\Gamma}(t), \quad (6.77)$$

where $\mathbf{C}$ is a 3×3 matrix.

- (b) Derive from Eq. 6.77 the coefficients of the Fokker-Planck equation governing the joint probability distribution $f(x, y, z, t)$ for the position process $\vec{R}(t)$.
- (c) Solve the Fokker-Planck equation in 10b subject to the initial condition $f(x, y, z, 0) = \delta(x)\delta(y)\delta(z)$ by using a 3-component Fourier transform. Show that the components $X(t)$ and $Y(t)$ are independent Wiener processes, each with zero mean and variance $2D_B t$, where $D_B = D / [1 + (qB/\beta)^2]$ and $D = k_B T / \beta$. What can you say about the motion $Z(t)$?

11. **Large damping ($\gamma \gg 1$) limit of Kramers Fokker-Planck equation:** For large damping, the velocity relaxes very quickly, and so we would like an equation that governs the marginal distribution for the position X alone, $P(X, t)$.

- (a) In Eq. 6.71, re-scale the variables,

$$X = x\sqrt{kT/m}, \quad V = v\sqrt{kT/m}, \quad F(X) = f(x)\sqrt{kT/m}, \quad (6.78)$$

to derive a re-scaled Fokker-Planck equation.

- (b) We seek a perturbation expansion of $P(x, v, t)$ in powers of γ^{-1} . To that end, make the ansatz,

$$P(x, v, t) = P^{(0)}(x, v, t) + \gamma^{-1} P^{(1)}(x, v, t) + \gamma^{-2} P^{(2)}(x, v, t) + \dots$$

Substitute into the re-scaled Fokker-Planck equation, and find $P(x, v, t)$ up to $O(\gamma^{-2})$.

- (c) Integrate $P(x, v, t)$ over v to obtain the marginal distribution $P(x, t)$ and show that it obeys the Fokker-Planck equation,

$$\frac{\partial P(x, t)}{\partial t} + \gamma^{-1} \left(\frac{d}{dx} f(x) P - \frac{\partial^2 P}{\partial x^2} \right) = O(\gamma^{-2}),$$

or, in the original variables,

$$\frac{\partial P(X, t)}{\partial t} = -\frac{\partial}{\partial X} \frac{F(X)}{m\gamma} P(X, t) + \frac{kT}{m\gamma} \frac{\partial^2}{\partial X^2} P(X, t). \quad (6.79)$$

CHAPTER 7

STOCHASTIC ANALYSIS

7.1 Limits

We return now to Doob's criticism of Ornstein and Uhlenbeck's study of Brownian motion (*cf.* page 20),

The purpose of this paper is to apply the methods and results of modern probability theory to the analysis of the Ornstein-Uhlenbeck distribution, its properties, and its derivation... A stochastic differential equation is introduced in a rigorous fashion to give a precise meaning to the Langevin equation for the velocity function. *This will avoid the usually embarrassing situation in which the Langevin equation, involving the second-derivative of x , is used to find a solution $x(t)$ not having a second-derivative.*

—J. L. Doob (1942) *Annals of Math.* **43**:351.

To understand Doob's objection, we must extend the definition of the derivative to include random functions. This definition, in turn, requires a definition for the *limit* of a sequence of random variables, of which there are several. We choose to use the mean-square limit here because of the close association between limits defined in this way and limits from ordinary calculus. For more details, M. Loève "Probability theory," Van Nostrand (1966) is recommended.

Definition: The random variable ξ is the limit of the sequence of random variables $\{\xi_1, \xi_2, \dots, \xi_n\}$, if

$$\lim_{n \rightarrow \infty} \langle |\xi - \xi_n|^2 \rangle = 0; \quad (7.1)$$

i.e. if for any $\varepsilon > 0$, there exists an $N = N(\varepsilon)$ such that for all $n > N$, $\langle |\xi - \xi_n|^2 \rangle < \varepsilon$. A limit defined in this way is usually called a *mean-square limit* (or a limit in the mean-square) and $\{\xi_n\}$ is said to converge to ξ *in the mean-square*.

Using Chebyshev's inequality,

$$P\{|\xi_n - \xi| \geq \varepsilon\} \leq \frac{\langle |\xi - \xi_n|^2 \rangle}{\varepsilon^2},$$

it is straightforward to verify that Eq. 7.1 implies that the sequence also *converges in probability*,

$$\lim_{n \rightarrow \infty} P\{|\xi_n - \xi| \geq \varepsilon\} = 0.$$

This means, in turn, that given any $\varepsilon > 0$ and $\eta > 0$, there exists an $N_0 = N_0(\varepsilon, \eta)$ such that for all $n > N_0$,

$$P\{|\xi_n - \xi| < \varepsilon\} > 1 - \eta.$$

(If ξ is the limit in the mean-square, then it is also the limit in probability.) Using these definitions, the statement of the ergodic theorem from Chapter 2 can be made more precise. We will not do so here; instead, we shall discuss some conditions which have to be imposed on a stationary stochastic process to guarantee the validity of the ergodic theorem.

One question that arises is: Under what conditions do we have,

$$m = \langle \xi(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) dt = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \xi^{(j)}(t)? \quad (7.2)$$

This question was answered by Slutski (1938) (see Doob's "Stochastic Processes," Wiley (1960) for more details):

Consider the *centered correlation function*,

$$C(\tau) = \langle [\xi(t + \tau) - m][\xi(t) - m] \rangle = B(\tau) - m^2,$$

then Eq. 7.2 holds if, and only if,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T C(\tau) d\tau = 0, \quad (7.3)$$

where the limits are defined as above.

In practice, the function $C(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$ since the dependence between $\xi(t + \tau)$ and $\xi(t)$ usually weakens as $\tau \rightarrow 0$. Of course, in this case Eq. 7.3 is satisfied. Another case of practical importance occurs when $C(\tau) = (\text{function} \rightarrow 0 \text{ as } \tau \rightarrow \infty) + (\text{periodic terms})$, as when $\xi(t)$ contains a purely harmonic component of the type $\xi = \xi_0 e^{i\omega t}$, where ξ_0 is a random variable and ω is a real number. Then, too, Eq. 7.3 is satisfied.

7.2 Mean-square Continuity

Throughout this chapter, we shall modify the standard definitions of calculus (continuity, differentiability and integrability) to include stochastic processes. The idea of convergence in the mean-square will be crucial. Likewise, we shall find that wherever mean-squared convergence appears, reformulation of the definition in terms of the autocorrelation is natural. We begin with the definition of mean-squared continuity.

Definition: A process $\xi(t)$ is *mean-square continuous* at t if, and only if, the limit

$$\lim_{h \rightarrow 0} \langle (\xi(t) - \xi(t - h))^2 \rangle = 0, \quad (7.4)$$

exists.

The formal definition can be recast in a more useful form involving the correlation function $B(t_1, t_2) = \langle \xi(t_1)\xi(t_2) \rangle$. The above limit exists if, and only if, $B(t_1, t_2)$ is continuous in t_1 and t_2 at the point $t_1 = t_2 = t$; that is, the limit

$$\lim_{h_1, h_2 \rightarrow 0} [B(t - h_1, t - h_2) - B(t, t)] = 0, \quad (7.5)$$

must exist. Notice this is *not the same* as setting $h_1 = h_2 = h$ and taking the limit as $h \rightarrow 0$. If $\xi(t)$ is *wide-sense stationary*, then $B(t, t - \tau) = B(\tau)$ is a function of the time-difference only and the condition for continuity is satisfied if, and only if, $B(\tau)$ is continuous at $\tau = 0$; that is, the limit

$$\lim_{h \rightarrow 0} [B(h) - B(0)] = 0, \quad (7.6)$$

must exist. From Eq. 7.6 it should be clear that if a wide-sense stationary process is continuous at one time t , it is continuous for all time.

7.3 Stochastic Differentiation

The derivative of a stochastic process is defined in a formal way:

Definition: A process $\xi(t)$ is *mean-square differentiable* at t if, and only if, there exists a random variable, denoted by $\xi'(t)$, such that the limit

$$\lim_{h \rightarrow 0} \left\langle \left(\frac{\xi(t+h) - \xi(t)}{h} - \xi'(t) \right)^2 \right\rangle = 0, \quad (7.7)$$

exists.

The definition, as usual, is difficult to use in practice. As a more useful corollary, one can show that Eq. 7.7 is satisfied if, and only if, the autocorrelation function $B(t_1, t_2) = \langle \xi(t_1)\xi(t_2) \rangle$ is differentiable at the point $t = t_1 = t_2$, *i.e.* the limit

$$\lim_{h_1, h_2 \rightarrow 0} \frac{1}{h_1 h_2} [B(t - h_1, t - h_2) - B(t, t - h_2) - B(t - h_1, t) + B(t, t)], \quad (7.8)$$

must exist. If the process $\xi(t)$ is *wide-sense stationary*, then $B(t, t - \tau) = B(\tau)$ is a function of the time-difference only and the condition for differentiability is simplified: The limit

$$\lim_{h \rightarrow 0} \frac{1}{h^2} [B(h) - 2B(0) + B(-h)], \quad (7.9)$$

must exist. Moreover, the autocorrelation function of the derivative $\xi'(t)$ is given by,

$$B_{\xi'(t)}(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} B(t_1, t_2),$$

or, for a stationary process,

$$B_{\xi'(t)}(\tau) = -\frac{d^2}{d\tau^2} B(\tau).$$

With these conditions in hand, it is straightforward to show that the Wiener process and the Ornstein-Uhlenbeck process *are not differentiable* (precisely Doob's objection), but we shall also show that this does not matter in the least for the modeling of physical systems.

7.3.1 Wiener process is *not* differentiable

Recall from Chapter 6 (*cf.* Eq. 6.31 on page 134) that in the absence of a damping force, the Langevin equation reduces to the equation characterizing the Wiener process $W(t)$,

$$\frac{dW}{dt} = \eta(t).$$

We will now show that $W(t)$ is *not* differentiable – and so the equation above, as it stands, is meaningless. In Chapter 6 (Eq. 6.32 on page 135), we derived the autocorrelation function for the Wiener process,

$$\langle W(t_1)W(t_2) \rangle = \min(t_1, t_2).$$

Clearly, the autocorrelation function indicates that the process $W(t)$ is *nonstationary*, so we must use Eq. 7.8 to check its differentiability:

$$\begin{aligned} & \lim_{h_1, h_2 \rightarrow 0} \frac{1}{h_1 h_2} [\min(t - h_1, t - h_2) - (t - h_2) - (t - h_1) + t] \\ &= \lim_{h_1, h_2 \rightarrow 0} \frac{1}{h_1 h_2} [\min(t - h_1, t - h_2) + h_2 + h_1 - t]. \end{aligned}$$

The limit does not exist, and we must conclude that the Wiener process is *not differentiable* (although one can easily show that it is mean-square *continuous*).

The Ornstein-Uhlenbeck process is stationary at steady-state, and the autocorrelation function,

$$B(\tau) = \frac{\Gamma}{2\tau_c} e^{-|\tau|/\tau_c}$$

is a function of the time-difference τ – We can therefore use Eq. 7.9 to check its differentiability. One finds that the Ornstein-Uhlenbeck process is likewise *not differentiable* (Exercise 3a). That was Doob's point in his criticism of the work of Ornstein and Uhlenbeck (quoted on page 20).

BUT (and this is critically important)

We shall now show that if the forcing function $F(t)$ is not strictly delta-correlated, if the process has a non-zero correlation time (however small), then the differential equation,

$$\frac{dy}{dt} = -\gamma y + F(t), \tag{7.10}$$

is well-defined and $y(t)$ is differentiable. For sake of example, suppose $F(t)$ is an Ornstein-Uhlenbeck process. As shown above, the steady-state correlation function for the Ornstein-Uhlenbeck process is the exponential,

$$B_F(\tau) = \frac{\Gamma}{2\tau_c} e^{-|\tau|/\tau_c},$$

where we have written the correlation time explicitly, and made the pre-factor proportional to $1/\tau_c$ to clarify the correspondence between the present example and the Wiener process studied above. Notice, in the limit of vanishing correlation time,

$$\lim_{\tau_c \rightarrow 0} \frac{1}{2\tau_c} e^{-|\tau|/\tau_c} \rightarrow \delta(\tau).$$

From Chapter 2 (Eq. 2.19 on page 42), the spectral density of the process $y(t)$ (as characterized by Eq. 7.10) is simply,

$$S_{yy}(\omega) = \frac{\Gamma}{\omega^2 + \gamma^2} \frac{(1/\tau_c)^2}{\omega^2 + (1/\tau_c)^2}, \quad (7.11)$$

from which the autocorrelation function follows,

$$B_y(\tau) = \frac{\Gamma}{2\gamma} \frac{e^{-\gamma|\tau|} - \gamma\tau_c e^{-|\tau|/\tau_c}}{1 - \gamma^2\tau_c^2}. \quad (7.12)$$

Taking the limit (*cf.* Eq. 7.9),

$$\lim_{h \rightarrow 0} \frac{B_y(h) - 2B_y(0) + B_y(-h)}{h^2} = -\frac{1}{2\tau_c} \frac{\Gamma}{1 + \gamma\tau_c}. \quad (7.13)$$

The limit exists, so we conclude that the process $y(t)$ defined by the differential equation (7.10) is differentiable. Obviously, as $\tau_c \rightarrow 0$, the limit above becomes undefined, but *for any non-zero correlation time, however small, the derivative of $y(t)$ is well-defined and can be manipulated using the rules of ordinary calculus.* Incidentally, Ornstein and Uhlenbeck never explicitly state that their forcing function is delta-correlated, merely that it is very narrow (G. E. Uhlenbeck and L. S. Ornstein (1930) *Physical Review* **36**: 823–841):

[W]e will naturally make the following assumptions... There will be correlation between the values of [the random forcing

function $A(t)$] at different times t_1 and t_2 only when $|t_1 - t_2|$ is very small. More explicitly we shall suppose that:

$$\langle A(t_1)A(t_2) \rangle = \phi(t_1 - t_2),$$

where $\phi(x)$ is a function with a very sharp maximum at $x = 0$.

As shown above, under these assumptions the derivation of Ornstein and Uhlenbeck is perfectly correct and no questions of inconsistency arise. This led Wang and Uhlenbeck to write as a footnote in a later publication (M. C. Wang and G. E. Uhlenbeck (1945) *Reviews of Modern Physics* **17**: 323–342),

The authors are aware of the fact that in the mathematical literature (especially in papers by N. Wiener, J. L. Doob, and others; *cf.* for instance Doob, *Ann. Math.* **43**, 351 (1942), also for further references) the notion of a random (or stochastic) process has been defined in a much more refined way. This allows for instance to determine in certain cases the probability that a random function $y(t)$ is of bounded variation, or continuous or differentiable, etc. However, it seems to us that these investigations have not helped in the solution of problems of direct physical interest, and we will, therefore, not try to give an account of them.

7.4 Stochastic Integration

Similar ideas can be used to define the integral of a stochastic process $\xi(t)$:

Definition: A process $\xi(t)$ is *mean-square integrable* on the interval $(0, t)$ if, and only if, there exists a random variable, denoted by $\xi^{(-1)}(t)$ or $\int_0^t \xi(u)du$, such that the limit

$$\lim_{\varepsilon \rightarrow 0} \left\langle \left(\varepsilon \sum_{i=0}^{\lfloor t/\varepsilon \rfloor} \xi(i\varepsilon) - \int_0^t \xi(u)du \right)^2 \right\rangle = 0, \quad (7.14)$$

exists (where the floor function $\lfloor \cdot \rfloor$ rounds down to the nearest integer).

As above with differentiation, the definition can be recast as a condition on the correlation function $B(t_1, t_2)$. Specifically, the limit Eq. 7.14

exists if and only if $B(t_1, t_2)$ is *Riemann-integrable* on the square $(0, t) \times (0, t)$; that is, the limit

$$\int_0^t \int_0^t B(t_1, t_2) dt_1 dt_2 \equiv \lim_{\varepsilon \rightarrow 0} \left[\varepsilon^2 \sum_{i,j=0}^{\lfloor t/\varepsilon \rfloor} B(i\varepsilon, j\varepsilon) \right], \quad (7.15)$$

must exist. If the process $\xi(t)$ is wide-sense stationary, then the corresponding limit for the stationary correlation function must exist,

$$\int_0^t B(\tau) d\tau \equiv \lim_{\varepsilon \rightarrow 0} \left[\varepsilon \sum_{i=0}^{\lfloor t/\varepsilon \rfloor} B(i\varepsilon) \right]. \quad (7.16)$$

For a test function $f(t)$, a necessary condition for

$$\int_a^b f(t) \xi(t) dt, \quad (7.17)$$

to be integrable in the mean-squared sense is that $\xi(t)$ is integrable; or, if the integral is defined in the Lebesgue-Stieltjes sense, then measure $\xi(t) dt$ must be of *bounded variation*. That means, very roughly, that the infinitesimal $\xi(t) dt$ must not be infinitely large in the mean-squared sense.

Recall that the correspondence between the Langevin equation and the Fokker-Planck equation is unambiguous, *unless* the white-noise forcing is multiplied by a nonlinear function of the state variable,¹

$$\frac{dy}{dt} = A(y) + c(y)\eta(t).$$

The question comes down to how one defines the terms in the equivalent integrated equation,

$$\int dy = \int A(y) dt + \int c(y)\eta(t) dt. \quad (7.18)$$

In particular, what precisely is the meaning of the last term, $\int c(y)\eta(t) dt$? Notice that because the correlation function for white noise is delta-correlated, *i.e.*, $B(t_1, t_2) = \delta(t_1 - t_2)$, this process is not integrable. Therefore, $\int c(y)\eta(t) dt$ is *not* integrable in the ordinary sense and some interpretation must be provided, beyond the ordinary rules of calculus. There

¹For convenience, we call any Langevin equation (irrespective of whether $A(y)$ is nonlinear) a *nonlinear Langevin equation* if the forcing function $c(y)$ is a nonconstant function of the state variable $y(t)$.

are two interpretations that dominate the literature: the Stratonovich interpretation and the Itô interpretation. It is useful to compare the two viewpoints, contrasting how they were developed, and most importantly, how they come in to solving actual problems.

7.4.1 Itô and Stratonovich Integration

First, some notation. As we showed in Section 7.3, the Langevin equation is not well-defined since the process it characterizes is non-differentiable. The problem comes from the singularity of the forcing function's correlation $\langle \eta(t_1)\eta(t_2) \rangle = \delta(t_1 - t_2)$.² Nevertheless, the integrated form of the equation is well-defined, *once an interpretation rule has been attached*. It is therefore more correct mathematically to write the Langevin equation as,

$$dy = A(y)dt + c(y)\eta(t)dt. \quad (7.19)$$

Furthermore, recall from Section 6.4 on page 133 that a purely diffusive process (*i.e.* a Langevin equation without damping), is called *Brownian motion* or a *Wiener process* $W(t)$,

$$\frac{dW}{dt} = \eta(t).$$

In the integrated form,

$$dW(t) = \eta(t)dt.$$

Very often, Eq. 7.19 is written using the increment of the Wiener process, thus,

$$dy = A(y)dt + c(y)dW(t). \quad (7.20)$$

It should be emphasized once again that Eqs. 7.19 and 7.20 have no meaning until an interpretation rule is provided for $\int c(y(t'))\eta(t')dt' \equiv \int c(y(t'))dW(t')$. The two dominant interpretations are discussed below.

1. **Stratonovich interpretation.** Stratonovich viewed the correlation $\langle \eta(t_1)\eta(t_2) \rangle = \delta(t_1 - t_2)$ as an idealization of a narrowly peaked correlation function. As we saw in Section 7.3, so long as the correlation time τ_c is non zero, the correlation function is non-singular,

²For simplicity, we set the variance of the white noise to 1, *i.e.*, $\Gamma = 1$.

and the Langevin equation can be manipulated using ordinary calculus. In particular, with the change of variables (*cf.* Eq. 6.60 on page 141),

$$\hat{y} = \int \frac{dy}{c(y)}, \quad \frac{A(y)}{c(y)} = \hat{A}(\hat{y}), \quad \hat{P}(\hat{y}) = P(y)c(y), \quad (7.21)$$

Stratonovich arrived at the following correspondence between the Langevin and Fokker-Planck equations,

$$dy = A(y)dt + c(y)dW(t) \quad (\text{STRATONOVICH}) \quad (7.22)$$

$$\Updownarrow$$

$$\frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y} \left[A(y) + \frac{1}{2}c(y)c'(y) \right] P(y, t) + \frac{1}{2} \frac{\partial^2}{\partial y^2} [c^2(y)P(y, t)].$$

Typically, the function $c(y) \neq 0$ for any y – if it does, more sophisticated methods are required to patch together solutions in the vicinity of the singular point. The above amounts to setting $c(y)$ equal to its *mean* value over the infinitesimal range of integration in Eq. 7.20,

$$y(t + dt) - y(t) = \int_t^{t+dt} A(y(t'))dt' + c \left(\frac{y(t) + y(t + dt)}{2} \right) \int_t^{t+dt} \eta(t')dt'. \quad (7.23)$$

2. **Itô interpretation.** Itô began from an entirely different viewpoint. In the same way that Lebesgue developed the ideas of set theory to provide a more general definition of Riemann's integral, Itô extended the ideas of Lebesgue to include integration with the Wiener process as the weighting function $dW(t')$. One can show that $dW(t)$ has *unbounded variation* which means, roughly speaking, that the mean-squared length of the infinitesimal $dW(t)$ is infinite – so Itô's taming of the wild Wiener process into a consistent definition of stochastic integration is certainly a mathematical tour-de-force. His development leads to the correspondence,

$$dy = A(y)dt + c(y)dW(t) \quad (\text{ITÔ}) \quad (7.24)$$

$$\Updownarrow$$

$$\frac{\partial P(y, t)}{\partial t} = -\frac{\partial}{\partial y} [A(y)] P(y, t) + \frac{1}{2} \frac{\partial^2}{\partial y^2} [c^2(y)P(y, t)].$$

This amounts to setting $c(y)$ equal to its *initial* value over the infinitesimal range of integration in Eq. 7.20,

$$y(t + dt) - y(t) = \int_t^{t+dt} A(y(t')) dt' + c(y(t)) \int_t^{t+dt} \eta(t') dt'. \quad (7.25)$$

It is important to note that in contrast to Stratonovich's interpretation, the Itô interpretation *cannot be formulated unless the noise is strictly delta-correlated*. Transformation of variables under Itô's calculus are not the same as ordinary calculus, and this is often useful in proving various properties of the solution of Eq. 7.24 (see Appendix C).

Notice that the Fokker-Planck equation does not require an interpretation rule – it is simply a partial differential equation governing the probability density. Furthermore, the Stratonovich and Itô interpretations can be made to generate the same form for the Fokker-Planck equation at the expense of re-defining the drift $A(y)$,

$$(\text{STRATONOVICH}) \bar{A}(y) = [A(y) - \frac{1}{2}c(y)c'(y)] \Leftrightarrow \bar{A}(y) = A(y) \text{ (ITÔ)}.$$

Obviously, if $c(y)$ is constant ($c'(y) = 0$), then the two interpretations coincide. A great deal is made about the distinction between Stratonovich and Itô calculus in some circles, but in the end, what it all comes down to is that *no natural process is described by white noise*, so the nonlinear Langevin equation is, by construction, a gross simplification of whatever process it is meant to represent. In the words of van Kampen ((1981) *Journal of Statistical Physics* **24**: 175–187),

The final conclusion is that a physicist cannot go wrong by regarding the Itô interpretation as one of those vagaries of the mathematical mind that are of no concern to him. It merely served to point out that [the nonlinear Langevin equation] is not a meaningful equation, and thereby warn him against glib applications of the Langevin approach to nonlinear equations.

Nevertheless, white noise is a useful construction in the development of approximation methods, particularly in the numerical simulation of stochastic differential equations (see Section 8.1). For further details about Itô's calculus, Gardiner's "Handbook of stochastic methods" is recommended.

Afterword

As mentioned above, $dW(t)$ has *unbounded variation* so that the mean-squared length of the infinitesimal $dW(t)$ is infinite! That is clearly problematic, and leads to $\int c(y(t'))dW(t')$ being *undefined* in any ordinary sense. Nevertheless, Itô did for Lebesgue integration what Lebesgue did for Riemann integration; namely, by providing a foundation built upon the very deep ideas of set theory, pathological integrals such as $\int c(y(t'))dW(t')$ are endowed with useful and self-consistent meaning. Furthermore, using Itô's calculus, various transformations and theorems have been established that greatly simplify the analysis of nonlinear Langevin equations. Said another way, Itô's calculus is a useful construction of pure mathematics that streamlines the proof of theorems using the Wiener process as a canonical noise source. This streamlining is so significant that often white noise sources are used to expedite the analysis of various models.

From a physical point of view, however, white noise *does not exist*! Any equation resembling the nonlinear Langevin equation is *necessarily an approximation of dynamics forced by a noise source with very narrow (though nonzero) correlation time τ_c* . The formal limit of vanishing correlation time τ_c from whence white noise in the Stratonovich sense is defined,

$$\lim_{\tau_c \rightarrow 0} \frac{1}{2\tau_c} \exp[-|t|/\tau_c] \rightarrow \delta(t),$$

is best thought of as an asymptotic limit because our Markov assumption relies upon a full relaxation of the microscopic variables over time scales much shorter than τ_c (see Section 3.4 on page 59). In summary, Itô's calculus is lovely mathematics and can sometimes provide a short-cut to lengthy computations, but it is a drastic approximation of real processes and therefore the physical relevance of results derived in this fashion must be viewed with skepticism.

Suggested References

Much of the early content of this chapter comes from

- *Introduction to random processes (2nd Ed.)*, W. A. Gardner (McGraw-Hill, 1990),

which is an excellent text on stochastic signal analysis directed toward an electrical engineering audience.

Gardiner's *Handbook* has a large collection of examples illustrating the range and applicability of Itô calculus,

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004),

and devotes several pages to the derivations outlined in this chapter.

N. G. van Kampen's master work,

- *Stochastic processes in physics and chemistry (2nd Ed.)*, N. G. van Kampen (North-Holland, 2001),

spares no punches in a characteristic savaging of Itô calculus as it relates to the modeling of physical processes (see p. 232–237 in that book). For a more self-contained and detailed version of his argument,

- N. G. van Kampen (1981) "The validity of non-linear Langevin-equations," *Journal of Statistical Physics* **25**: 431–442,

is highly recommended, along with

- N. G. van Kampen (1981) "Itô versus Stratonovich," *Journal of Statistical Physics* **24**: 175–187.

Exercises

1. **Wide-sense stationary processes:** For wide-sense stationary processes, the fundamental definitions of continuity, differentiability and integrability were re-cast in terms of the correlation function $B(\tau)$.
 - (a) Show that Eq. 7.6 follows from the definition of continuity for a wide-sense stationary process.
 - (b) Show that Eq. 7.9 follows from the definition of differentiability for a wide-sense stationary process.
 - (c) Show that Eq. 7.16 follows from the definition of integrability for a wide-sense stationary process.
 - (d) Show that mean-squared differentiability implies mean-squared continuity.
2. **Orthogonality of a wide-sense stationary process:** For the wide-sense stationary process $\xi(t)$, show that

$$\left\langle \xi(t) \frac{d\xi(t)}{dt} \right\rangle = 0.$$

3. Non-differentiability of the Ornstein-Uhlenbeck process.

- (a) Show that the Ornstein-Uhlenbeck process is not differentiable.
- (b) Explain, in words, why the damped differential equation for $y(t)$ (Eq. 7.10),

$$\frac{dy}{dt} = -\gamma y + F(t),$$

forced with a *nondifferentiable* Ornstein-Uhlenbeck process $F(t)$ still results in $y(t)$ being differentiable.

4. **Moments of an integrated process:** Show that for the integrable wide-sense stationary process $\xi(t)$, the integrated process $Y(t) = \int_0^t \xi(u) du$ has mean and variance given by,

$$\begin{aligned} \langle Y(t) \rangle &= \langle \xi \rangle \cdot t, \\ \langle (Y(t) - \langle Y(t) \rangle)^2 \rangle &= t \cdot \int_{-t}^t \left(1 - \frac{|\tau|}{t}\right) C_\xi(\tau) d\tau, \end{aligned}$$

where $C_\xi(\tau) = \langle \xi(t)\xi(t-\tau) \rangle - \langle \xi \rangle^2$.

5. Itô and Stratonovich interpretations:

- (a) Show that with Itô's interpretation, $\langle c(y)\eta(t) \rangle = 0$, while the Stratonovich interpretation gives $\langle c(y)\eta(t) \rangle = \frac{1}{2} \langle c'(y) c(y) \rangle$.
- (b) For the following Langevin equation,

$$\frac{dx}{dt} = -x \eta(t), \quad x(0) = x_0, \quad (7.26)$$

where $\eta(t)$ is Gaussian white noise, what is $\langle x(t) \rangle$ with Eq. 7.26,

- i. interpreted in the Stratonovich sense.
- ii. interpreted in the Itô sense.

CHAPTER 8

RANDOM DIFFERENTIAL EQUATIONS

The ‘repeated-randomness assumption’ (or the *Stosszahlansatz*, section 3.4 on page 59) allows many degrees of microscopic freedom in the model dynamics to be eliminated, leaving behind a Markov process governed by a master equation. In some case, this coarse-graining is continued, and it is possible to identify a subset of variables whose stochastic properties are not influenced by the variables of interest. In the extreme case, all of the stochastic character comes from these external variables, and the master equation reduces to a differential equation with random coefficients (a *random differential equation*). Noise of this type is usefully denoted ‘*extrinsic*’ or ‘*external*’ noise (in contrast with ‘*intrinsic*’ noise where the fluctuations are inherent to the dynamics, and the master equation cannot be further reduced). Examples of extrinsic noise are Brownian motion, where the motion of the solvent molecules give rise to a random forcing of the solute particle, although their motion is not affected by the solute. Another example is an electromagnetic wave passing through a turbulent troposphere – the wave obeys Maxwell’s equations with a random dielectric constant (leading to a random coefficient in the differential equation), justified by the observation that the passing wave has negligible effect on the turbulent atmosphere.

In much of the current literature, the term *stochastic differential equation* is used almost exclusively to denote Itô’s version of the nonlinear

Langevin equation,

$$dy(t) = A(y)dt + c(y)dW(t),$$

where $dW(t) = \eta(t)dt$ is the Wiener process coming from a Gaussian white noise source $\eta(t)$. When the random force $d\xi(t)$ is *not* white noise, the problem of interpretation (Itô versus Stratonovich) does not occur because $\int c(y)d\xi(t)$ is well-defined as an ordinary (Lebesgue-Stieltjes) integral.

If one's focus is upon problems arising from physical systems (*i.e.* problems in science and engineering), the emphasis on Itô's equation and Itô calculus is misdirected because it is the fact that $dW(t)$ comes from a white noise source that creates all the problems. In a manner of speaking, these difficulties are artificial and disappear when one takes into account that a random force in physics is *never* really white noise, but has (at best) a very short correlation time. Once that is accepted, one also gets rid of Doob's objection arising from the non-existence of the derivative in the Langevin equation, since a non-white noise forcing does not affect the differentiability of the process. Consequently, stochastic differential equations can be formally written as ordinary differential equations,

$$\frac{du}{dt} = F(u, t; Y(t)), \quad u(t_0) = u_0, \quad (8.1)$$

where u and F may be vectors, and $Y(t)$ is a random function whose stochastic properties are given and whose correlation time is non-zero.

Roughly speaking, random differential equations can be subdivided into several classes,

1. linear random differential equations where only the forcing term is a random function, as in the Langevin equation (additive noise).
2. linear random differential equations where one or more of the coefficients are random functions (multiplicative noise).
3. nonlinear random differential equations.
4. other variations, random partial differential equations, *etc.*

Additive noise was dealt with in great detail in Chapter 1. Multiplicative noise will be dealt with in the following sections. Nonlinear stochastic differential equations and partial differential equations are more advanced than the level of this course.

8.1 Numerical Simulation

- C. W. Gardiner, *Handbook of stochastic methods (3rd Ed.)*, (Springer, 2004).
- D. T. Gillespie, *Markov processes: An introduction for physical scientists*, (Academic Press, 1992).
- D. S. Lemons, *An introduction to stochastic processes in physics*, (John Hopkins University Press, 2002).

The typical situation in physics is that the physical world, or at least some large system, is subdivided into a subsystem and its environment (see Section 3.4 on page 59). The influence of the environment on a subsystem is treated like a heat bath, *viz.* as a random force whose stochastic properties are given (or assumed). Hence the equations of motion of the total system are reduced to those of the subsystem at the expense of introducing random coefficients.

The system described by Eq. 8.1 determines for each particular realization $y(t)$ of the random function $Y(t)$ a functional $U([y], t, u_0)$ which depends upon *all* values $y(t')$ for $0 \leq t' \leq t$. The ensemble of solutions $U([y], t, u_0)$ for all possible realizations $y(t')$ constitutes a stochastic process. The random differential equation is *solved* when the stochastic properties of this process have been found – this can rarely be done in practice, and so approximation methods are necessary.

As an example of the treatment of random differential equations, we concentrate on the class of multiplicative noise processes, *i.e.* those random differential equations with random coefficients. Even for this comparatively simple class, the theory is very incomplete and a general solution is out of the question. Instead, we focus upon the first two moments, and specifically on the behaviour of the average $\langle u(t) \rangle$. In the following sections, we briefly discuss how numerical solutions of random differential equations should be computed, then turn attention to analytic approximation schemes for generating closed equations for the moments of $u(t)$.

White noise

Numerical simulation of random differential equations begins with a discretization of the Langevin equation,

$$dy = A(y, t)dt + c(y, t)\eta(t)dt,$$

where $\eta(t)$ is *Gaussian white noise*. One can show that the integrated process dy (interpreted in the *Itô* sense) is equivalent to the equation

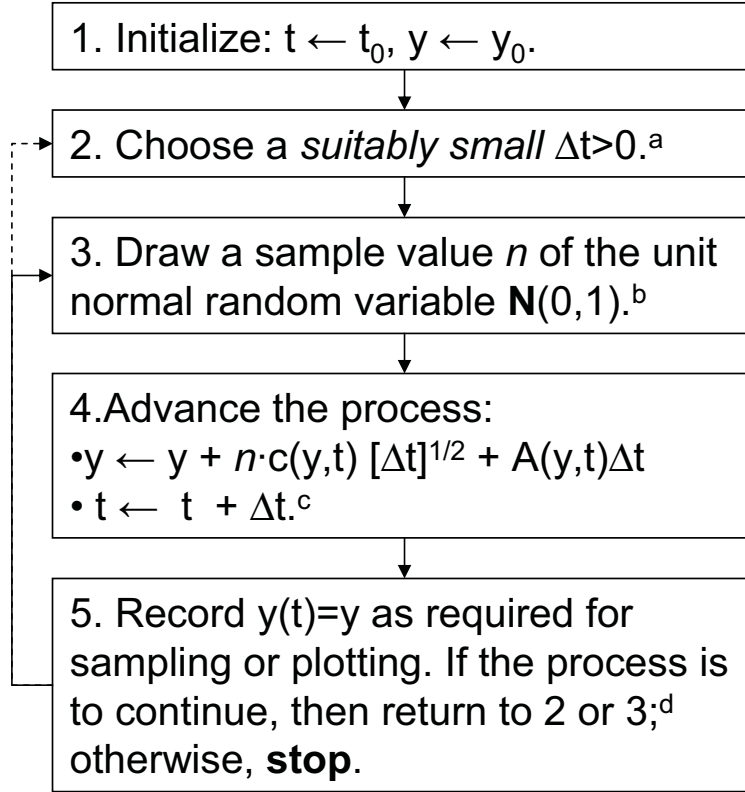


Figure 8.1: **Simulation of a trajectory drawn from $p(y, t | y_0, t_0)$ satisfying the Fokker-Planck equation $\partial p / \partial t = -\partial_y [A(y, t)p] + \frac{1}{2} \partial_y^2 [c^2(y, t)p]$.** **a)** Should satisfy the *first-order accuracy condition* (Eq. 8.5) for some $\varepsilon_1 \ll 1$. Optionally, may satisfy the *plotting condition* (Eq. 8.6) for some $\varepsilon_2 < 1$. **b)** Use a suitable method to generate the normally distributed random number (see Exercise 3). **c)** If $(t_{max} - t_0) / \Delta t \sim 10^K$, then the sum $t + \Delta t$ should be computed with at least $K + 3$ digits of precision. **d)** The value of Δt may be reset at the beginning of any cycle. Taken from Gillespie (1992), p. 194.

(Exercise 2a),

$$dy = A(y, t)dt + \mathbf{N}(0, c^2(y, t)dt), \quad (8.2)$$

where $\mathbf{N}(\mu, \sigma^2)$ is a Gaussian (normal) distribution with mean μ and variance σ^2 (see Section A.2 on page 236). From the properties of the Gaussian distribution, we can re-write Eq. 8.2 as,

$$dy = A(y, t)dt + n \cdot c(y, t)\sqrt{dt}, \quad (8.3)$$

where now n is a number drawn from a unit normal distribution $\mathbf{N}(0, 1)$. Several remarks are in order—

1. For the Wiener process ($A = 0, D = 1$), Eq. 8.3 reduces to,

$$dW(t) = n \cdot dt^{1/2},$$

a relation that is exploited in many theorems using Itô's calculus (see Gardiner (2004) and Appendix C for examples).

2. The Wiener process is continuous, so

$$\begin{aligned} W(t + \Delta t) - W(t) = \\ \left[W(t + \Delta t) - W\left(t + \frac{\Delta t}{2}\right) \right] + \left[W\left(t + \frac{\Delta t}{2}\right) - W(t) \right]. \end{aligned}$$

From Eq. 8.3,

$$\sqrt{\Delta t} \mathbf{N}_t^{t+\Delta t}(0, 1) = \sqrt{\frac{\Delta t}{2}} \mathbf{N}_{t+\Delta t/2}^{t+\Delta t}(0, 1) + \sqrt{\frac{\Delta t}{2}} \mathbf{N}_t^{t+\Delta t/2}(0, 1).$$

The sub- and super-scripts of $\mathbf{N}_t^{t+\Delta t}$ indicate that the random variable is *explicitly associated with the interval* $(t, t + \Delta t)$. This relation is only satisfied if the normal distributions are independent of one another for non-overlapping intervals. We say, therefore, that the Wiener process has *independent increments*.

The integration of Eq. 8.3 over a small, but finite, time interval Δt provides a practical numerical simulation algorithm,

$$y(t + \Delta t) = y(t) + A(y, t)\Delta t + n \cdot c(y, t)\sqrt{\Delta t}, \quad (8.4)$$

(see Figure 8.1). The deviation of the numerical scheme from the Chapman-Kolmogorov equation quantifies the error involved; to minimize this type

of consistency error (*i.e.* to keep the error higher order in Δt), the time step Δt must satisfy the *accuracy condition* ($\varepsilon_1 \ll 1$),

$$\Delta t \leq \min \left[\frac{\varepsilon_1}{|\partial_y A(y, t)|}, \left(\frac{\varepsilon_1}{\partial_y c(y, t)} \right)^2 \right]. \quad (8.5)$$

Furthermore, to generate a representative plot of the simulation results, the data must be stored with a resolution such that the fluctuations are captured on a time-scale much shorter than the evolution due to the drift, resulting in an additional *plotting condition* on the storage time step ($\varepsilon_2 < 1$),

$$\Delta t \leq \frac{\varepsilon_2^2 c^2(y, t)}{A^2(y, t)}. \quad (8.6)$$

Higher-order schemes

- S. Asmussen and P. W. Glynn, *Stochastic simulation: Algorithms and analysis*, (Springer, 2007).

The simulation algorithm, Eq. 8.4, in the absence of diffusion ($c(y, t) = 0$), reduces to the simple forward Euler method for the numerical solution of ordinary differential equations. One can show that the scheme described above exhibits $\frac{1}{2}$ -order strong convergence and 1^{st} -order weak convergence. The question that immediately presents itself is whether higher-order schemes, such as Runge-Kutta or Adams-Bashford methods, which have proved so useful in numerical solution of ordinary differential equations, can be extended to stochastic systems. For the Langevin equation, it would seem they cannot without considerable complication of the stepping algorithm. The major difficulty is not so much the stochasticity, but rather that the trajectory governed by the Langevin equation has no well-defined derivative, frustrating the derivation of higher-order methods. We will briefly outline how higher-order explicit schemes can be derived – although we shall see that very little is gained from the added complexity of the algorithm.

The main error associated with the Euler scheme comes from the approximation,

$$\int_0^{\Delta t} c(y(t), t) dW(t) \approx c(y(0), 0) W(\Delta t).$$

To improve upon the algorithm, higher-order terms are included in the estimate. Re-writing

$$\begin{aligned} \int_0^{\Delta t} c(y(t), t) dW(t) - c(y(0), 0)W(\Delta t) = \\ \int_0^{\Delta t} \{c(y(t), t) - c(y(0), 0)\} dW(t), \end{aligned} \quad (8.7)$$

we expand $c(y(t), t)$ using Itô's formula (see Appendix C, p. 272). The result is *Milstein's scheme*, (where again n_i are independent samples of a unit Normal distribution $\mathbf{N}(0, 1)$),

$$\begin{aligned} y(t + \Delta t) = \\ y(t) + A(y, t)\Delta t + n_1 \cdot c(y, t)\sqrt{\Delta t} - \frac{\Delta t}{2} c(y, t) \cdot c_y(y, t) \cdot (1 - n_2^2), \end{aligned} \quad (8.8)$$

which converges strongly with order 1. Higher-order schemes can be generated, but they become increasingly unwieldy with only moderate improvement of convergence. A pragmatic approach is to use *coloured* noise and higher-order stepping schemes, although convergence order can rarely be computed explicitly for these schemes.

Coloured noise

Coloured noise is often simulated by *augmenting* the model equations to include a Langevin equation governing an Ornstein-Uhlenbeck process. For example, the equation

$$\frac{dy}{dt} = A(y, t) + c(y, t) \cdot F(t), \quad (8.9)$$

where $F(t)$ is an Ornstein-Uhlenbeck process, is simulated by the augmented system,

$$\begin{aligned} \frac{dy}{dt} &= A(y, t) + c(y, t) \cdot F(t), \\ \frac{dF}{dt} &= -\frac{1}{\tau_c} F(t) + \frac{1}{\tau_c} \eta(t), \end{aligned}$$

where $\eta(t)$ is zero-mean, unit-variance Gaussian white noise, and τ_c is the correlation time of the coloured noise process. Written in this way, the steady-state autocorrelation function of $F(t)$ is

$$\langle F(t)F(t - \tau) \rangle = \frac{1}{2\tau_c} e^{-|\tau|/\tau_c} \xrightarrow{\tau_c \rightarrow 0} \delta(\tau).$$

An improved approach comes from noting that the Ornstein-Uhlenbeck process is Gaussian. We can therefore write the *exact* stepping formula for $F(t)$ as (Exercise 2b),

$$F(t + \Delta t) = F(t) \cdot e^{-\Delta t/\tau_c} + n \cdot \left[\left(\frac{1}{2\tau_c} \right) (1 - e^{-2\Delta t/\tau_c}) \right]^{1/2}, \quad (8.10)$$

where, again, n is a unit normal random variable. The process governed by Eq. 8.9 *is differentiable* (see Eq. 7.13 on page 157), so we are free to use *any stepping scheme we wish to advance y*. That is,

$$y(t + \Delta t) = h[y(t), \dots, y(t + \Delta t), F(t)],$$

$$F(t + \Delta t) = F(t) \cdot e^{-\Delta t/\tau_c} + n \cdot \left[\left(\frac{1}{2\tau_c} \right) (1 - e^{-2\Delta t/\tau_c}) \right]^{1/2},$$

where $h[\cdot]$ is any implicit or explicit stepping scheme used to integrate Eq. 8.9. Echoing the analysis of Chapter 7, we again find that the Langevin equation is difficult to deal with because it is a differential equation characterizing a non-differentiable process. This paradox disappears when we replace the canonical white noise forcing by a more physically relevant coloured noise. We are then able to use all the methods developed in ordinary calculus, including the numerical integration of differential equations. Although it should not need saying, notice that in the limit $\tau_c \rightarrow 0$, Eq. 8.9 reduces to the Langevin equation *interpreted in the Stratonovich sense!* In the next section, we shall discuss various approximation schemes for treating equations such as Eq. 8.9 under various assumptions about the correlation time τ_c and the magnitude of the noise (focusing exclusively on models with A and c *linear* in y).

8.2 Approximation Methods

There are several approximation methods that have been developed to derive evolution equations for the moments of a process governed by a linear stochastic differential equation. We shall consider approximation methods that assume either very long or very short correlation time for the fluctuations, and we shall limit the analysis to the derivation of leading order terms.

8.2.1 Static Averaging

Consider the stochastic differential equation,

$$\frac{du}{dt} = A(t, \xi)u,$$

where ξ is a stochastic process. It can happen that the fluctuations have very long correlation time compared with the rapid deterministic relaxation of the system. In that case, we may treat the fluctuating coefficients as a *time-independent* random variable. The differential equation is solved in the ordinary fashion to obtain a solution that depends parametrically upon the fluctuating term $u(t; \xi)$, then the solution is *averaged over* the probability distribution of the fluctuating coefficient $f(\xi)$,

$$\langle u(t) \rangle = \int_{-\infty}^{\infty} f(\xi) u(t; \xi) d\xi,$$

We shall call this the *static averaging approximation*. An example will make the procedure clear.

Example: The static averaging approximation is well-illustrated by the complex harmonic oscillator with random frequency,

$$\frac{du}{dt} = -i\xi u; \quad u(0) = u_0,$$

where ξ is a random variable. Because ξ does not depend on time, we can solve the random differential equation explicitly – *i.e.* solve for arbitrary ξ , and average over the ensemble. Remember, we assume from the outset that the entire probability distribution for ξ is known. The solution is,

$$u(t) = u_0 e^{-i\xi t}.$$

It is instructive to look at a few special cases for the probability distribution of $f(\xi)$; in the following list ξ_0 and γ are fixed parameters.

1. Cauchy/Lorentz:

$$f(\xi) = \frac{\gamma/\pi}{\gamma^2 + (\xi - \xi_0)^2}; \quad \langle u \rangle = u_0 e^{-i\xi_0 t - \gamma t}$$

2. Gaussian:

$$f(\xi) = \frac{1}{\sqrt{2\pi\gamma}} \exp \left[-\frac{1}{2\gamma} (\xi - \xi_0)^2 \right]; \quad \langle u \rangle = u_0 e^{-i\xi_0 t - \gamma t^2}$$

3. Laplace:

$$f(\xi) = \frac{1}{2}\gamma e^{-\gamma|\xi-\xi_0|}; \quad \langle u \rangle = u_0 e^{-i\xi_0 t} \frac{\gamma^2}{\gamma^2 + t^2}.$$

In each case, the averaged solution tends to zero as $t \rightarrow \infty$. This damping is due to the fact that the harmonic oscillators of the ensemble gradually lose the phase coherence they had at $t = 0$. In plasma physics this is called “phase mixing”, in mathematics the “Riemann-Lebesgue theorem”. The modulus of u is not subject to phase mixing and does not tend to zero; in fact,

$$|u(t)| = u_0, \text{ hence } \langle |u(t)|^2 \rangle = u_0^2.$$

The form of the damping factor is determined by $f(\xi)$. Only for one particular $f(\xi)$ does it have the form that corresponds to a complex frequency. Note that $\langle u(t) \rangle$ is identical with the characteristic function $\chi(t)$ of the distribution $f(\xi)$. The fact that $\langle u(t) \rangle = \chi(t)$ suggests the use of the cumulant expansion

$$\langle u(t) \rangle = \langle e^{-i\xi t} \rangle = \exp \left[\sum_{m=1}^{\infty} \frac{(-it)^m}{m!} \kappa_m \right],$$

where κ_m stands for the m^{th} -cumulant, and indeed it is this idea that underlies the approximation discussed in Section 8.2.3.

8.2.2 Bourret's Approximation

- R. C. Bourret (1965) “Fictitious theory of dynamical systems with noisy parameters,” *Canadian Journal of Physics* **43**: 619–639.
- A. Brissaud and U. Frisch (1974) “Solving linear stochastic differential equations,” *Journal of Mathematical Physics* **15**: 524–534.
- N. G. van Kampen (1976) “Stochastic differential equations,” *Physics Reports* **24**: 171–228.

Consider the linear random differential equation,

$$\frac{du}{dt} = [\mathbf{A}_0 + \alpha \mathbf{A}_1(t)]u; \quad u(0) = u_0, \quad (8.11)$$

where u is a vector, $\mathbf{A}_0$ is a constant matrix, $\mathbf{A}_1(t)$ is a random matrix, and α is a parameter measuring the magnitude of the fluctuations in the coefficients. If the mathematical model is chosen properly, then α is small. We would like to find *deterministic* equations for the moments of u . Notice if we simply average the stochastic differential equation,

$$\frac{d\langle u \rangle}{dt} = \mathbf{A}_0 \langle u \rangle + \alpha \langle \mathbf{A}_1(t) u \rangle.$$

But we have no obvious method to evaluate the cross-correlation $\langle \mathbf{A}_1(t) u \rangle$. The methods derived in this section are approximations of that cross-correlation. Concerning the random matrix $\mathbf{A}_1(t)$, we make the following assumptions:

1. $\langle \mathbf{A}_1(t) \rangle = 0$: This is not strictly necessary, but it simplifies the discussion. It is, for example, true if $\mathbf{A}_1(t)$ is *stationary*, for then $\langle \mathbf{A}_1(t) \rangle$ is independent of t and can be absorbed into $\mathbf{A}_0$.
2. $\mathbf{A}_1(t)$ has a finite (non-zero) correlation time τ_c : This means that for any two times t_1, t_2 , such that $|t_1 - t_2| \gg \tau_c$, one may treat all matrix elements of $\mathbf{A}_1(t_1)$ and of $\mathbf{A}_1(t_2)$ as statistically independent.

First, we eliminate $\mathbf{A}_0$ from the random differential equation by setting,

$$u(t) = e^{t\mathbf{A}_0} v(t).$$

Substitution into Eq. 8.11 yields the new equation for $v(t)$,

$$\frac{dv}{dt} = \alpha \mathbf{V}(t) v(t), \quad (8.12)$$

where $\mathbf{V}(t) = e^{-t\mathbf{A}_0} \mathbf{A}_1(t) e^{t\mathbf{A}_0}$, and $v(0) = u(0) = u_0$. Since α is small, the obvious method for solving this problem seems to be a perturbation series in α , *i.e.* assume a solution of the form,

$$v(t) = v_0(t) + \alpha v_1(t) + \alpha^2 v_2(t) + \dots$$

Substituting into Eq. 8.12,

$$v(t) = u_0 + \alpha \left(\int_0^t \mathbf{V}(t_1) dt_1 \right) u_0 + \alpha^2 \left(\int_0^t \int_0^{t_1} \mathbf{V}(t_1) \mathbf{V}(t_2) dt_2 dt_1 \right) u_0 + \dots \quad (8.13)$$

Upon taking the average, with fixed u_0 ,

$$\langle v(t) \rangle = u_0 + \alpha^2 \left(\int_0^t \int_0^{t_1} \langle \mathbf{V}(t_1) \mathbf{V}(t_2) \rangle dt_2 dt_1 \right) u_0 + \dots \quad (8.14)$$

Dropping all higher-order terms, we have *Born's iterative approximation*. One could claim to have an approximation for $\langle v \rangle$ to 2^{nd} -order in α ; however, this is *not* a suitable perturbation scheme because the successive terms are not merely of increasing order in α , but also in t . That is, the expansion is actually in powers of (αt) and is therefore valid only for a limited time, *i.e.* $\alpha t \ll 1$.

A different approach was taken by Bourret. Starting from Eq. 8.12, the equation can be re-written in an equivalent integrated form,

$$v(t) = u_0 + \alpha \int_0^t \mathbf{V}(t') v(t') dt'. \quad (8.15)$$

By iteration, this equation can be re-written as,

$$\begin{aligned} v(t) &= u_0 + \alpha \int_0^t \mathbf{V}(t') \left\{ u_0 + \alpha \int_0^{t'} \mathbf{V}(t'') v(t'') dt'' \right\} dt' \\ &= u_0 + u_0 \alpha \int_0^t \mathbf{V}(t') dt' + \alpha^2 \int_0^t \int_0^{t'} \mathbf{V}(t') \mathbf{V}(t'') v(t'') dt'' dt', \end{aligned}$$

and on taking the average,

$$\langle v(t) \rangle = u_0 + \alpha^2 \int_0^t \int_0^{t'} \langle \mathbf{V}(t') \mathbf{V}(t'') v(t'') \rangle dt'' dt'. \quad (8.16)$$

This equation is exact, but of no help in finding $\langle v(t) \rangle$ because it contains the higher-order correlation $\langle \mathbf{V}(t') \mathbf{V}(t'') v(t'') \rangle$. Suppose, however, that one could write,

$$\langle \mathbf{V}(t') \mathbf{V}(t'') v(t'') \rangle \approx \langle \mathbf{V}(t') \mathbf{V}(t'') \rangle \langle v(t'') \rangle. \quad (8.17)$$

Then Eq. 8.16 becomes an integral equation for $\langle v \rangle$ alone,

$$\langle v(t) \rangle = u_0 + \alpha^2 \int_0^t \int_0^{t'} \langle \mathbf{V}(t') \mathbf{V}(t'') \rangle \langle v(t'') \rangle dt'' dt', \quad (8.18)$$

or, in terms of the original variables, and after differentiation, we are left with the convolution equation for $\langle u(t) \rangle$

$$\frac{d}{dt} \langle u(t) \rangle = \mathbf{A}_0 \langle u(t) \rangle + \alpha^2 \int_0^t \langle \mathbf{A}_1(t) e^{\mathbf{A}_0(t-t')} \mathbf{A}_1(t') \rangle \langle u(t') \rangle dt'. \quad (8.19)$$

This is called *Bourret's convolution equation*, or sometimes in the physics literature, the *mode-coupling approximation*. It is a striking result because it asserts that the average of $u(t)$ obeys a *closed* equation; thus one can find $\langle u(t) \rangle$ without knowledge of the higher moments of $u(t)$. That is, without going through the procedure of solving the random differential equation $du/dt = A(t; \omega)u$ for each individual ω and averaging afterwards.

Bourret's convolution equation is formally solved using the Laplace transform, but the *structure* of the equation is sometimes not convenient. A convolution equation extends over the past values of $\langle u(t) \rangle$, and as such does not describe a Markov process. Nevertheless, with further manipulation, we are able to turn the convolution equation into an ordinary differential equation (albeit with modified coefficients), as was done by Kubo using a different argument. We use the *noiseless* dynamics ($\alpha = 0$) to re-write $\langle u(t') \rangle$,

$$\langle u(t') \rangle \approx e^{-\mathbf{A}_0(t-t')} \langle u(t) \rangle, \quad (8.20)$$

allowing $\langle u(t) \rangle$ to pass through the integral in Eq. 8.19,

$$\begin{aligned} & \int_0^t \langle \mathbf{A}_1(t) e^{\mathbf{A}_0(t-t')} \mathbf{A}_1(t') \rangle \langle u(t') \rangle dt' \approx \\ & \left(\int_0^t \langle \mathbf{A}_1(t) e^{\mathbf{A}_0(t-t')} \mathbf{A}_1(t') \rangle e^{-\mathbf{A}_0(t-t')} dt' \right) \langle u(t) \rangle. \end{aligned} \quad (8.21)$$

We have assumed that $\mathbf{A}_1(t)$ is a stationary random process, so the argument of the integral is a function of time-difference only,

$$\langle \mathbf{A}_1(t) e^{\mathbf{A}_0(t-t')} \mathbf{A}_1(t') \rangle e^{-\mathbf{A}_0(t-t')} \equiv \mathbf{K}'(t - t').$$

Making the change of variable $\tau = t - t'$,

$$\int_0^t \mathbf{K}'(t - t') dt' = \int_0^t \mathbf{K}'(\tau) d\tau.$$

The fluctuations are correlated only for small time separations ($\tau \ll \tau_c$), which means the upper limit of integration is not important and can be extended $t \rightarrow \infty$ incurring only exponentially small error,

$$\int_0^t \mathbf{K}'(\tau) d\tau \approx \int_0^\infty \mathbf{K}'(\tau) d\tau.$$

This last term is a constant (*i.e.*, independent of time), and so we arrive at an ordinary differential equation for $\langle u(t) \rangle$ where $\int_0^\infty \mathbf{K}'(\tau) d\tau$ renormalizes the constant coefficient matrix $\mathbf{A}_0$,

$$\frac{d}{dt} \langle u(t) \rangle = \left(\mathbf{A}_0 + \alpha^2 \int_0^\infty \mathbf{K}'(\tau) d\tau \right) \cdot \langle u(t) \rangle \quad (8.22)$$

$$= \left(\mathbf{A}_0 + \alpha^2 \int_0^\infty \langle \mathbf{A}_1(\tau) e^{\mathbf{A}_0 \tau} \mathbf{A}_1(0) \rangle e^{-\mathbf{A}_0 \tau} d\tau \right) \cdot \langle u(t) \rangle. \quad (8.23)$$

This is called *Kubo's renormalization equation*. In contrast with Bourret's convolution equation, Kubo's approximation results in an ordinary differential equation, and so $\langle u(t) \rangle$ characterized by this equation is a *Markov* process. Although the derivation may seem *ad hoc*, we can use cumulants to compute a very tight error estimate of the approximation, and in fact show that the error incurred in deriving Kubo's renormalization is of the same order as the error incurred in Bourret's approximation – so that Eq. 8.23 is as accurate as Eq. 8.19! The cumulant expansion method also points the way to calculation of higher-order terms (although the algebra becomes formidable), and will be the focus of the next two sections.

8.2.3 Cumulant Expansion

- N. G. van Kampen (1974) “A cumulant expansion for stochastic linear differential equations,” *Physica* **74**: 215–238.
- R. H. Terwiel (1974) “Projection operator method applied to stochastic linear differential equations,” *Physica* **74**: 248–265.

The hypothesis of Bourret, Eq. 8.17, is unsatisfactory because it implies *an uncontrolled neglect* of certain fluctuations without any *a priori* justification. Nonetheless, it is important to note that replacing an average of a product with a product of averages underlies the microscopic theory of transport phenomena, and, in a sense, it is what is meant by the *Stosszahlansatz* discussed on page 59, as we show below. In this section, we shall develop Bourret's approximation (and higher-order terms) with a careful estimate of the error incurred at each step.

Cumulants of a Stochastic Process

- R. Kubo (1962) “Generalized cumulant expansion method,” *Journal of the Physical Society of Japan* **17**: 1100–1120.

Let ξ be a scalar random variable. The *cumulants* κ_m of ξ are defined by means of the generating function,

$$\langle e^{-i\xi t} \rangle = \exp \left[\sum_{m=1}^{\infty} \frac{(-it)^m}{m!} \kappa_m \right],$$

(cf. Eq. A.9 on page 236). When discussing the cumulants of a *stochastic function* $\xi(t)$, it is convenient to adopt the notation $\kappa_m \equiv \langle \langle \xi^m \rangle \rangle$.

Let $\xi(t)$ be a scalar random function, then,

$$\begin{aligned} & \langle e^{-i \int_0^t \xi(t') dt'} \rangle = \\ & \exp \left[\sum_{m=1}^{\infty} \frac{(-i)^m}{m!} \int_0^t \cdots \int_0^t \langle \langle \xi(t_1) \xi(t_2) \cdots \xi(t_m) \rangle \rangle dt_1 dt_2 \cdots dt_m \right]. \end{aligned} \quad (8.24)$$

The general rule for relating cumulants to the *moments* of $\xi(t)$ is to partition the digits of the moments into all possible subsets; for each partition, one writes the product of cumulants, then adds all such products. The first few relations will make the prescription clear – Writing 1, 2, ... for $\xi(t_1), \xi(t_2), \dots$,

$$\begin{aligned} \langle 1 \rangle &= \langle \langle 1 \rangle \rangle, \\ \langle 1 \ 2 \rangle &= \langle \langle 1 \rangle \rangle \langle \langle 2 \rangle \rangle + \langle \langle 1 \ 2 \rangle \rangle, \\ \langle 1 \ 2 \ 3 \rangle &= \\ & \langle \langle 1 \rangle \rangle \langle \langle 2 \rangle \rangle \langle \langle 3 \rangle \rangle + \langle \langle 1 \rangle \rangle \langle \langle 2 \ 3 \rangle \rangle + \langle \langle 2 \rangle \rangle \langle \langle 1 \ 3 \rangle \rangle + \langle \langle 3 \rangle \rangle \langle \langle 1 \ 2 \rangle \rangle + \langle \langle 1 \ 2 \ 3 \rangle \rangle, \dots \end{aligned}$$

An expansion in cumulants is preferable over an expansion in moments for two main reasons – first, the expansion occurs in the *exponential*, greatly decreasing the risk of secular terms¹. The second advantage of cumulants is that higher-order cumulants *vanish* if *any* of the arguments are uncorrelated with *any* of the others. For example, suppose $\xi(t)$ has a short correlation time τ_c – specifically, that for $|t_1 - t_2| \gg \tau_c$, $\xi(t_1)$ and $\xi(t_2)$ are uncorrelated, then;

$$\langle 1 \ 2 \rangle = \langle 1 \rangle \langle 2 \rangle,$$

¹ *Secular terms* are terms arising in a perturbation approximation that diverge in time although the exact solution remains well-behaved. For example, after a short time $1 - t + t^2/2!$ is a *terrible* approximation of e^{-t} .

so we have,

$$\langle 1 \ 2 \rangle = \langle \cancel{1} \cancel{2} \rangle = \langle \cancel{\langle 1 \rangle} \cancel{\langle 2 \rangle} \rangle + \langle \langle 1 \ 2 \rangle \rangle,$$

which implies $\langle \langle 1 \ 2 \rangle \rangle = 0$. Similarly, one can show that having $\xi(t_1)$ and $\xi(t_2)$ uncorrelated is sufficient to ensure that *all* higher-order cumulants vanish, irrespective of their correlations with one another. That is to say, *the m^{th} cumulant vanishes as soon as the sequence of times $\{t_1, t_2, \dots, t_m\}$ contains a gap large compared to τ_c .*

Since the cumulants are approximately zero once the gap among any of the time points exceeds τ_c , each integral in Eq. 8.24 is an $(m-1)$ -dimensional sphere of radius $\sim \tau_c$ moving along the time axis. Said another way, each integral in Eq. 8.24 grows *linearly* in time (once $t \gg \tau_c$).

Cumulant Expansion

Returning again to our original linear stochastic differential equation,

$$\frac{dv}{dt} = \alpha \mathbf{V}(t)v(t), \quad (8.25)$$

(with $\mathbf{V}(t) = e^{-t\mathbf{A}_0} \mathbf{A}_1(t) e^{t\mathbf{A}_0}$), we see that equation contains *three* time scales:

1. The time τ_V over which $\mathbf{V}(t)$ varies, which is not relevant in what follows.
2. The time $1/\alpha$ over which $v(t)$ varies.
3. The correlation time τ_c over which $\mathbf{V}(t)$ is correlated.
That is, $\langle \mathbf{V}(t) \mathbf{V}(t') \rangle \approx 0$ for $|t - t'| \gg \tau_c$.

While the naive perturbation solution initially proposed failed because we expanded in powers of (αt) , an improved perturbation scheme should be possible if we expand in powers of $(\alpha \tau_c)$.² If $\alpha \tau_c \ll 1$, then it is possible to subdivide the interval $[0, t]$ so that $t \gg \tau_c$ and yet $\alpha t \ll 1$. This is precisely the *Stosszahlansatz* underlying the derivation of the master equation, and precisely the assumption Einstein made when constructing his model of Brownian motion – that the time axis can be subdivided into steps Δt which are long enough that the microscopic variables relax to their equilibrium state ($\Delta t \gg \tau_c$), though not long enough that the observable state $v(t)$ is appreciably changed ($1/\alpha \gg \Delta t$). The condition that $\alpha \tau_c \ll 1$ underlies the Bourret and Kubo approximations. As we now

²The dimensionless parameter $\alpha \tau_c$ is called the *Kubo number*.

show, it is possible to justify these approximations and obtain higher-order terms as a systematic expansion in powers of $(\alpha\tau_c)$ using the cumulants of a stochastic process.

We can write the formal solution of Eq. 8.25 as a time-ordered exponential (see Section B.1),

$$v(t) = \left[\exp \left\{ \int_0^t \alpha \mathbf{V}(t') dt' \right\} \right] v(0), \quad (8.26)$$

where the time-ordering operator $[]$ is short-hand for the iterated series, Eq. 8.13. For the *average* process, $\langle v(t) \rangle$, we have the suggestive expression,

$$\langle v(t) \rangle = \left[\left\langle \exp \left\{ \int_0^t \alpha \mathbf{V}(t') dt' \right\} \right\rangle \right] v(0), \quad (8.27)$$

since the average and the time-ordering commute. Eq. 8.24 shows how the cumulants of a *scalar* stochastic process can be used to express the averaged exponential. For the matrix $\mathbf{V}(t)$, it would seem at first sight that Eq. 8.24 is of little use, *but* we are saved by the time-ordering. Inside the time-ordering operator, we can freely commute the matrix operator $\mathbf{V}(t)$ since everything must be put in chronological order once the time-ordering operator is removed. We are therefore justified in writing,

$$\begin{aligned} \langle v(t) \rangle = & \\ \left[\exp \left\{ \alpha \int_0^t \langle \mathbf{V}(t_1) \rangle dt_1 + \frac{\alpha^2}{2} \int_0^t \int_0^t \langle \mathbf{V}(t_1) \mathbf{V}(t_2) \rangle dt_2 dt_1 + \dots \right\} \right] v(0). & \end{aligned} \quad (8.28)$$

This equation is exact, no approximations have been made up to this point. There are several important characteristics of Eq. 8.28, however, that make it well-suited as a starting point for approximations of an evolution equation for $\langle v(t) \rangle$. First, the cumulant expansion appears in the exponent, greatly reducing the risk of secular terms. Second, successive terms in the exponent are of order α , α^2 , $\dots$, and grow *linearly* in time. For the matrix $\mathbf{V}(t)$, we assume the fluctuations have finite correlation time τ_c in the sense that the cumulants of the matrix elements

$$\langle \langle V_{ij}(t_1) V_{kl}(t_2) \cdots V_{qr}(t_m) \rangle \rangle = 0$$

vanish as soon as a gap of order τ_c appears among *any* of the time-points. We then have a good estimate of the size of each term in the expansion – each term is of order $\alpha^m t \tau_c^{m-1}$.

To obtain from the exact Eq. 8.28 a practical approximation of the process $\langle v(t) \rangle$, we begin eliminating terms in the expansion. For example, if we eliminate all terms α^2 and higher, we have,

$$\langle v(t) \rangle \approx \left[\exp \left\{ \int_0^t \langle \langle \mathbf{V}(t_1) \rangle \rangle dt_1 \right\} \right] v(0),$$

which solves the differential equation,

$$\frac{d}{dt} \langle v(t) \rangle = \langle \mathbf{V}(t) \rangle \langle v(t) \rangle,$$

the simplest possible approximation where the fluctuations are replaced by their average. Suppose that $\langle \mathbf{V} \rangle = 0$ (see page 176), then truncating Eq. 8.28 after two terms,

$$\langle v(t) \rangle \approx \left[\exp \left\{ \frac{\alpha^2}{2} \int_0^t \int_0^t \langle \langle \mathbf{V}(t_1) \mathbf{V}(t_2) \rangle \rangle dt_2 dt_1 \right\} \right] v(0),$$

or, equivalently,

$$\langle v(t) \rangle = \left[\exp \left\{ \alpha^2 \int_0^t \int_0^{t_1} \langle \langle \mathbf{V}(t_1) \mathbf{V}(t_2) \rangle \rangle dt_2 dt_1 \right\} \right] v(0). \quad (8.29)$$

We introduce the operator,

$$\mathbf{K}(t_1) = \int_0^{t_1} \langle \langle \mathbf{V}(t_1) \mathbf{V}(t_2) \rangle \rangle dt_2,$$

and consider the differential equation

$$\frac{d}{dt} \langle v(t) \rangle = \alpha^2 \mathbf{K}(t) \langle v(t) \rangle. \quad (8.30)$$

The solution is given by,

$$\langle v(t) \rangle = \left[\exp \left\{ \alpha^2 \int_0^t \mathbf{K}(t_1) dt_1 \right\} \right] v(0). \quad (8.31)$$

This solution is not quite the same as Eq. 8.29, although the difference is of higher-order in $\alpha\tau_c$. We may therefore use Eq. 8.30 as an approximation of the evolution equation for $\langle v(t) \rangle$. In the original representation, Eq. 8.30 is identical to Eq. 8.21,

$$\frac{d}{dt} \langle u(t) \rangle = \left\{ \mathbf{A}_0 + \alpha^2 \int_0^t \langle \langle \mathbf{A}_1(t) e^{\mathbf{A}_0 \tau} \mathbf{A}_1(t-\tau) \rangle \rangle e^{-\mathbf{A}_0 \tau} \right\} \langle u(t) \rangle, \quad (8.32)$$

which reduces to Bourret's approximation once $\langle u(t) \rangle$ is brought back inside the integral, or Kubo's approximation once the upper limit of integration is extended $t \rightarrow \infty$. No *ad hoc* assumptions regarding the factoring of correlations have been made. Nor have any secular terms appeared in the derivation. Furthermore, we have a good estimate of the error incurred by the approximation – the error is of order $(\alpha^3 \tau_c^2)$.

Summary of approximation methods

Approximation	Condition on Kubo number	Condition on t
Born Iteration	$\alpha\tau_c < 1$	$\alpha^2\tau_c t < < 1$
	$\alpha\tau_c > 1$	$\alpha t < < 1$
Bourret Convolution	$\alpha\tau_c < < 1$	
Kubo Renormalization	$\alpha\tau_c < < 1$	$t > > \tau_c$
Static Averaging	$\alpha\tau_c > > 1$	or $t < < \tau_c$

Figure 8.2: **User's guide for various approximations of the evolution equation for $\langle y(t) \rangle$ derived from a linear random differential equation.** Redrawn after Brissaud and Frisch (1974).

With the many approximation methods discussed above, it is helpful to have a “user's guide” outlining the range of applicability of each approximation scheme (Figure 8.2). Loosely speaking, if the correlation time of the fluctuations is short, then the Born approximation is only good over short times, the Kubo renormalization is good for times longer than the correlation time, and the Bourret approximation is good over short and long times, although it is a convolution equation, so it is slightly

more cumbersome to deal with than the Kubo renormalization. If the correlation time of the noise is long, then the static averaging approximation is useful – if the Kubo number ($\alpha\tau_c$) is very large ($\alpha\tau_c \gg 1$), then the approximation is good for all time, while if the Kubo number is not large, then the approximation is good for times shorter than the noise correlation time.

8.3 Example – Random Harmonic Oscillator

- N. G. van Kampen (1976) “Stochastic differential equations,” *Physics Reports* **24**: 171.
- M. Gitterman, *The noisy oscillator: The first hundred years, from Einstein until now*, (World Scientific Publishing Company, 2005).

As an example of a simple linear system with stochastic coefficients, consider the random harmonic oscillator,

$$\frac{d^2x}{dt^2} + \omega_0^2 x = 0; \quad x(0) = a, \frac{dx}{dt}(0) = 0.$$

For constant frequency ω_0 , the solution is $x(t) = a \cos(\omega_0 t)$. Suppose instead that ω_0 is a random function,

$$\omega_0^2 \mapsto \omega_0^2(1 + \alpha\xi(t)),$$

where $\xi(t)$ is a zero-mean, stationary stochastic process. Re-scaling time, $t \rightarrow \omega_0 t$, we have the non-dimensionalized model,

$$\frac{d^2x}{dt^2} + (1 + \alpha\xi(t))x = 0; \quad x(0) = a, \frac{dx}{dt}(0) = 0. \quad (8.33)$$

The process $\xi(t)$ is assumed to have a finite correlation time τ_c such that,

$$\langle \xi(t)\xi(t - \tau) \rangle = 0, \quad |\tau| \gg \tau_c.$$

We now examine two limiting regimes for τ_c .

Long correlation time, $\tau_c \gg 1$

If the correlation time τ_c is long, then we can invoke the static averaging approximation by treating $\xi(t) \rightarrow \xi$ as a *time-independent random variable*. The formal solution for $x(t)$ is then given by,

$$x(t) = a \cos[(1 + \alpha\xi)t]. \quad (8.34)$$

To find $\langle x(t) \rangle$, we average Eq. 8.34 over the probability distribution for ξ , $f(\xi)$,

$$\langle x(t) \rangle = a \int_{-\infty}^{\infty} \cos[(1 + \alpha\xi)t] f(\xi) d\xi. \quad (8.35)$$

We tabulate several examples of $f(\xi)$ below:

1. Gaussian: ($\sigma \ll 1$)

$$f(\xi) = \mathbf{N}(0, \sigma^2); \quad \langle x(t) \rangle = ae^{-\frac{\alpha^2 \sigma^2 t^2}{2}}$$

2. Uniform: ($\beta \ll 1$)

$$f(\xi) = \mathbf{U}(-\beta, \beta); \quad \langle x(t) \rangle = a \frac{\sin[\alpha\beta t]}{(\alpha\beta t)}$$

3. Cauchy: ($\sigma \ll 1$)

$$f(\xi) = \mathbf{C}(0, \sigma); \quad \langle x(t) \rangle = ae^{-\alpha\sigma t}$$

In all cases, we see again that the super-position of an ensemble of out-of-phase oscillators results in the *damping* of the average $\langle x(t) \rangle$.

Short correlation time, $\tau_c \ll 1$

It is convenient to re-write Eq. 8.33 in matrix-vector notation,

$$\frac{d}{dt} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} + \alpha\xi(t) \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix}, \quad (8.36)$$

with,

$$\mathbf{A}_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{A}_1(t) = \alpha\xi(t) \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}. \quad (8.37)$$

Using Kubo's renormalization approximation, Eq. 8.23, we have,

$$\frac{d^2}{dt^2} \langle x \rangle + c_2 \frac{\alpha^2}{2} \frac{d}{dt} \langle x \rangle + (1 - \frac{\alpha^2}{2} c_1) \langle x \rangle = 0, \quad (8.38)$$

where,

$$c_1 = \int_0^\infty \langle \xi(t) \xi(t - \tau) \rangle \sin[2\tau] d\tau,$$

$$c_2 = \int_0^\infty \langle \xi(t) \xi(t - \tau) \rangle (1 - \cos[2\tau]) d\tau.$$

Comparing the two regimes, we find that for long correlation time it is the *distribution* $f(\xi)$ of the fluctuations that is important, while for short correlation time it is the *autocorrelation function* $\langle \xi(t) \xi(t - \tau) \rangle$ that determines the average state. Although outside the scope of this course, there are approximation methods that combine these ideas to develop an approximation method valid over *all* correlation times – see, for example, the review by Brissaud and Frisch,

- A. Brissaud and U. Frisch (1974) “Solving linear stochastic differential equations,” *Journal of Mathematical Physics* **15**: 524–534.

8.4 Example – Parametric Resonance

In contrast to the example above, we now consider a linear differential equation with *time-varying* coefficients – the Mathieu equation,

$$\ddot{u} + [a + 2q \cos 2t]u = 0. \quad (8.39)$$

Over a range of a and q , the periodic time-dependence in the frequency leads to instability via a mechanism called *parametric resonance*. This is the same effect that is exploited by children “pumping” a swing. To transform the Mathieu equation to a stochastic differential equation, we allow the parametric forcing to have a random amplitude $q \mapsto q(1 + \alpha\xi(t))$. In vector-matrix notation,

$$\frac{d}{dt} \begin{pmatrix} u \\ \dot{u} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -(a + 2q \cos 2t) & 0 \end{pmatrix} \begin{pmatrix} u \\ \dot{u} \end{pmatrix} + \alpha\xi(t) \begin{pmatrix} 0 & 0 \\ -2q \cos 2t & 0 \end{pmatrix} \begin{pmatrix} u \\ \dot{u} \end{pmatrix}. \quad (8.40)$$

Because the Mathieu equation has *time-dependent* coefficients, it is no longer possible to find a closed form for the solution, even in the absence

of fluctuations. Nevertheless, the formal solution can be written as a time-ordered exponential,

$$\begin{pmatrix} u \\ \dot{u} \end{pmatrix} = \left[\exp \left\{ \int_0^t \mathbf{A}_0(t_1) dt_1 \right\} \right] \cdot \begin{pmatrix} u(0) \\ \dot{u}(0) \end{pmatrix}; \quad \alpha = 0. \quad (8.41)$$

If we adopt the Kubo renormalization approximation, the correction to the deterministic dynamics is the integral,

$$\begin{aligned} & \int_0^\infty \left\langle \mathbf{A}_1(t) \cdot \left[\exp \left\{ \int_0^\tau \mathbf{A}_0(t_1) dt_1 \right\} \right] \times \mathbf{A}_1(t - \tau) \right\rangle \\ & \times \left[\exp \left\{ - \int_0^\tau \mathbf{A}_0(t_1) dt_1 \right\} \right] d\tau \end{aligned} \quad (8.42)$$

Here, $\left[\exp \left\{ - \int_0^\tau \mathbf{A}_0(t_1) dt_1 \right\} \right]$ is the *matrix inverse* of the deterministic propagator, Eq. 8.41, and

$$\mathbf{A}_1(t) = \alpha \xi(t) \begin{pmatrix} 0 & 0 \\ -2q \cos 2t & 0 \end{pmatrix}. \quad (8.43)$$

Let $\xi(t)$ be a colored noise process, with an exponential correlation function,

$$\langle \xi(t) \xi(t - \tau) \rangle = \exp \left[- \frac{|\tau|}{\tau_c} \right].$$

In that case, we can invoke Watson's lemma (Section B.5.2) to approximate the integral (8.42) as a series in powers of τ_c (Exercise 9).

8.5 Optional – Stochastic Delay Differential Equations

In many models of physical systems, several degrees of freedom are eliminated from the governing equations by introducing a time delay into the dynamics. The result is a system of delay differential equations. For linear delay differential equations with stochastic coefficients, it is straightforward to derive an approximation for the averaged process as we have done in the preceding sections. In particular, we shall derive Bourret's convolution approximation for a linear stochastic delay equation in the limit that the time-delay is large compared to the correlation time of the

fluctuating coefficients. To that end, consider the linear delay differential equation,

$$\frac{d}{dt}x = ax + bx^\tau, \quad x(t) = x_0 \text{ for } t \leq 0,$$

where τ_d is the delay time and $x^\tau = x(t - \tau_d)$. The one-sided Green's function for this system $g(t)$ is calculated using the Laplace transform. The auxiliary equation characterizing $g(t)$ is

$$\frac{d}{dt}g(t) - ag(t) - bg(t - \tau_d) = \delta(t).$$

Taking the Laplace transform $\mathcal{L}\{\cdot\}$, with $\hat{g}(s) = \mathcal{L}\{g(t)\}$,

$$s\hat{g}(s) - a\hat{g}(s) - be^{-s\tau_d}\hat{g}(s) = 1,$$

since $\mathcal{L}\{g(t - \tau_d)\} = e^{-s\tau_d}\mathcal{L}\{g(t)\}$. Explicitly, we write,

$$g(t) = \mathcal{L}^{-1}\{\hat{g}(s)\} = \mathcal{L}^{-1}\left\{\frac{1}{s - a - be^{-s\tau_d}}\right\}.$$

Consider the original delay equation, but now with zero-mean multiplicative noise $\eta(t)$,

$$\frac{d}{dt}x = ax + bx^\tau + \eta(t)x. \quad (8.44)$$

Using the one-sided Green's function $g(t)$, we write the formal solution of (8.44) as a convolution,

$$x(t) = x_0g(t) + \int_0^t g(t - t')\eta(t')x(t')dt'.$$

With substitution into the right-hand side of (8.44),

$$\frac{d}{dt}x = ax + bx^\tau + \int_0^t g(t - t')\eta(t)\eta(t')x(t')dt'.$$

Taking the ensemble average, we are left with the evolution equation for the first-moment,

$$\frac{d}{dt}\langle x \rangle = a\langle x \rangle + b\langle x^\tau \rangle + \int_0^t g(t - t')\langle \eta(t)\eta(t')x(t') \rangle dt'.$$

Invoking Bourret's approximation to factor the cross-correlation,

$$\langle \eta(t) \eta(t') x(t') \rangle \approx \langle \eta(t) \eta(t') \rangle \langle x(t') \rangle,$$

we have

$$\frac{d}{dt} \langle x \rangle = a \langle x \rangle + b \langle x^\tau \rangle + \int_0^t g(t-t') \langle \eta(t) \eta(t') \rangle \langle x(t') \rangle dt'.$$

We further assume that $\eta(t)$ is a *Gaussian distributed, stationary, Markov process* so that,

$$\langle \eta(t) \eta(t') \rangle = K(t-t') = \sigma^2 \exp \left[-\frac{|t-t'|}{\tau_c} \right],$$

and τ_c is the correlation time of the noise. The approximate evolution equation then simplifies to,

$$\frac{d}{dt} \langle x \rangle = a \langle x \rangle + b \langle x^\tau \rangle + \int_0^t g(t-t') K(t-t') \langle x(t') \rangle dt'.$$

Writing the convolution explicitly,

$$\frac{d}{dt} \langle x(t) \rangle = a \langle x(t) \rangle + b \langle x(t-\tau_d) \rangle + [g(t) K(t)] * \langle x(t) \rangle.$$

The Laplace transform $\langle \hat{x}(s) \rangle$ is readily obtained,

$$\langle \hat{x}(s) \rangle = \frac{x_0}{s - a - be^{-s\tau_d} - \sigma^2 \mathcal{L} \left\{ g(t) e^{-\frac{t}{\tau_c}} \right\}}.$$

Explicit inversion of the Laplace transform $\hat{g}(s)$ is difficult, but since

$$\mathcal{L} \left\{ g(t) e^{-\frac{t}{\tau_c}} \right\} = \hat{g} \left(s + \frac{1}{\tau_c} \right),$$

we have,

$$\langle \hat{x}(s) \rangle = \frac{x_0}{(s - a - be^{-s\tau_d}) - \sigma^2 [\tilde{s} - a - be^{-\tilde{s}\tau_d}]^{-1}}, \quad (8.45)$$

where $\tilde{s} = s + \frac{1}{\tau_c}$.

In the absence of noise, the asymptotic stability of $x(t)$ is determined by the real part of the *pole* of the Laplace transform $\hat{x}(s)$, or equivalently by the real part of s^* , where s^* is the root of $s^* - a - be^{-s^*\tau_d}$,

$$s^* - a - be^{-s^*\tau_d} = 0.$$

Similarly, the asymptotic stability of the first-moment $\langle x(t) \rangle$ is determined by the real part of s^* satisfying,

$$(s^* - a - be^{-s^*\tau_d}) - \sigma^2 \left[\left(s^* + \frac{1}{\tau_c} \right) - a - be^{-(s^* + \frac{1}{\tau_c})\tau_d} \right]^{-1} = 0.$$

To examine the white-noise limit (in the Stratonovich sense), it is convenient to make the substitution $\sigma^2 = \frac{\Gamma^2}{2\tau_c}$ so that,

$$\lim_{\tau_c \rightarrow 0} \sigma^2 e^{-\frac{|t|}{\tau_c}} = \lim_{\tau_c \rightarrow 0} \frac{\Gamma^2}{2\tau_c} e^{-\frac{|t|}{\tau_c}} = \Gamma^2 \delta(t),$$

with $\int_0^\varepsilon \delta(t) dt = \frac{1}{2}$, (the Stratonovich interpretation of the Dirac delta function). We then write the resolvent equation as,

$$\begin{aligned} & \left(s^* - a - be^{-s^*\tau_d} \right) - \frac{\Gamma^2}{2} + \\ & \tau_c \left(s^* - a - be^{-s^*\tau_d} e^{-\frac{\tau_d}{\tau_c}} \right) \left(s^* - a - be^{-s^*\tau_d} \right) = 0. \end{aligned} \quad (8.46)$$

In this form, the roots of the resolvent equation can be developed as an asymptotic series in a straightforward manner (see Exercise 11).

Suggested References

Much of the early part of this chapter comes from,

- A. Brissaud and U. Frisch (1974) “Solving linear stochastic differential equations,” *Journal of Mathematical Physics* **15**: 524–534.

Two excellent sources are van Kampen’s review,

- N. G. van Kampen (1976) “Stochastic differential equations,” *Physics Reports* **24**: 171–228,

and the review, translated from Russian,

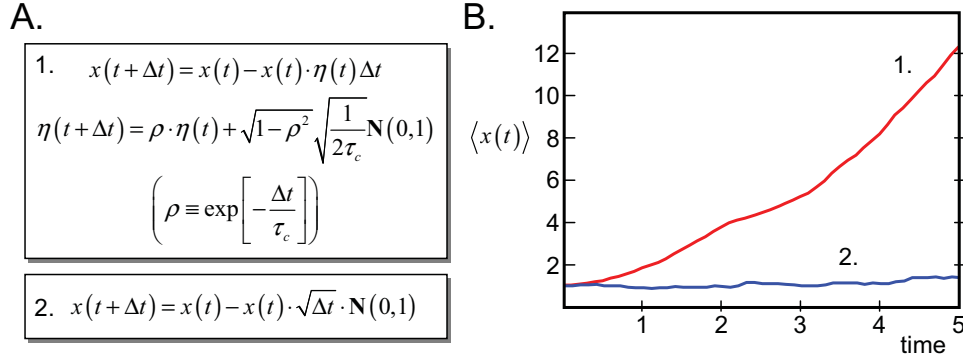


Figure 8.3: **Jenny’s paradox - Exercise 1.** **A)** Two different schemes to represent the differential equation $\dot{x} = -x \times \eta(t)$, where $\eta(t)$ is Gaussian white noise. **B)** The averaged behaviour using each of the schemes shown in panel A. The average is taken over an ensemble of 1000. The step-size is $\Delta t = 10^{-2}$ and the correlation time is $\tau_c = 10^{-1}$. $\mathbf{N}(\mu, \sigma^2)$ represents a random number drawn from a Normal distribution with mean μ and variance σ^2 .

- V. I. Klyatskin and V. I. Tatarskii (1974) “Diffusive random process approximation in certain nonstationary statistical problems of physics,” *Soviet Physics – Uspekhi* **16**: 494–511.

This second article is notable in its use of the Novikov-Furutsu relation (see Section 10.5 on page 224).

Exercises

1. **Jenny’s paradox:** In the course of testing her code to numerically integrate stochastic differential equations, Jenny noticed that the following differential equation,

$$\frac{dx}{dt} = -x\eta(t), \quad x(0) = 1,$$

(where $\eta(t)$ is Gaussian white noise) had different averaged behaviour $\langle x(t) \rangle$ depending upon how she coded the white noise (Figure 8.3).

- (a) Explain what each routine shown in Figure 8.3A is intended to do. What is the fundamental difference between the two routines?

- (b) Why don't both routines return the same result? What should Jenny expect the averages to look like? What will happen to the output of scheme 1 if the step-size Δt is reduced? What will happen to the output of scheme 1 if the correlation time τ_c is reduced (with $0 < \Delta t \ll \tau_c$)? Will the two schemes coincide in that limit?

2. Simulation of the Langevin equation (in the Itô's sense):

- (a) Show that the update scheme, Eq. 8.2, follows from the Langevin equation interpreted in the Itô sense.
- (b) Show that the update scheme, Eq. 8.10, provides the necessary correlation function for the Ornstein-Uhlenbeck process $F(t)$.

3. Gaussian distributed random number generator: In the simulation of random differential equations, it is often necessary to generate Gaussian-distributed random numbers. Since there are many tried and true routines for generating a unit uniform random number, we shall consider how to transform a realization $r \in \mathbf{U}(0, 1)$ to a number $n \in \mathbf{N}(0, 1)$. Though most coding packages include subroutines for just this purpose, explicit derivation of the algorithm is illuminating.

- (a) What makes it impossible to directly invert a unit uniform random variable into a normally distributed random number? Why is this not a problem for inversion to an exponential or Cauchy distribution?
- (b) Using *two* unit uniform realizations r_1 and r_2 , introduce two auxiliary random numbers $s = \sigma[2\ln(1/r_1)]^{1/2}$ and $\theta = 2\pi r_2$. Show that $x_1 = \mu + s \cos \theta$ and $x_2 = \mu + s \sin \theta$ are a pair of statistically independent sample values of the Gaussian distribution $\mathbf{N}(\mu, \sigma^2)$. *Hint:* Use the Joint Inversion Generating Method described on page 243, along with the subordinate density functions $Q_1(s)$ and $Q_2^{(1)}(\theta|s)$.

4. Cumulant expansion: In the derivation of the cumulant expansion, several details were left out.

- (a) What is the difference between Eq. 8.29 and Eq. 8.31?
- (b) From Eq. 8.32, derive the Bourret and Kubo approximations.

5. Show that for a zero-mean Gaussian process $F(t)$,

$$\left\langle \exp \left[i \int_0^t F(t') dt' \right] \right\rangle = \exp \left[-\sigma^2 \int_0^t (t - \tau) \psi(\tau) d\tau \right],$$

where $\langle F(t_1)F(t_2) \rangle = \sigma^2 \psi(|t_1 - t_2|)$.

6. **Eigenvalues of the averaged random harmonic oscillator:**

From the equation governing the averaged response of the random harmonic oscillator, Eq. 8.38, derive the eigenvalues of the system for the following choices of noise autocorrelation function:

- (a) Exponential correlation: $\langle \langle \xi(t) \xi(t - \tau) \rangle \rangle = \sigma^2 \exp \left[-\frac{|\tau|}{\tau_c} \right]$
- (b) Uniform correlation: $\langle \langle \xi(t) \xi(t - \tau) \rangle \rangle = \begin{cases} \frac{\sigma^2}{2\tau_c} & -\tau_c \leq \tau \leq \tau_c \\ 0 & \text{otherwise} \end{cases}$
- (c) Triangle correlation: $\langle \langle \xi(t) \xi(t - \tau) \rangle \rangle = \begin{cases} \frac{\sigma^2}{\tau_c} - \frac{\sigma^2}{\tau_c^2} |\tau| & -\tau_c \leq \tau \leq \tau_c \\ 0 & \text{otherwise} \end{cases}$
- (d) Damped sinusoidal correlation: $\langle \langle \xi(t) \xi(t - \tau) \rangle \rangle = \frac{\sigma^2}{\pi} \frac{\sin |\tau|}{|\tau|}$

Do you notice any commonalities among the eigenvalues resulting from all of these various choices? Explain.

7. **Redfield equations:** We can exploit the linearity of Eq. 8.11 to generate an equation for the higher moments.

- (a) Using tensor notation, each component of the N -dimensional differential equation,

$$\frac{d\mathbf{u}}{dt} = [\mathbf{A}_0 + \alpha \mathbf{A}_1(t)] \cdot \mathbf{u} \equiv \mathbf{A} \cdot \mathbf{u},$$

is written $u_i = \sum_{j=1}^N A_{ij} u_j$. Use this formulation to derive the differential equation for the products $u_i u_k$.

- (b) Use either Bourret's convolution equation or Kubo's renormalization approximation to derive an equation for the second moments $\langle u_i u_k \rangle$ for the damped random harmonic oscillator discussed in Section 8.3. Verify the approximate equation for the second moments using an ensemble of simulation data.

- (c) Determine the conditions for stability of the second moments (mean-squared stability) for the damped random harmonic oscillator discussed in Section 8.3. It may be helpful to review stability conditions for ordinary differential equations, *cf.* Appendix B.2.1. Verify the stability bounds using an ensemble of simulation data.
- (d) Use Kubo's renormalization approximation to derive an equation for the second moments $\langle u_i u_k \rangle$ for the parametric oscillator discussed in Section 8.4. What happens to the corrected dynamics in the limit of white noise $t_c \rightarrow 0$? Verify the approximate equation for the second moments using an ensemble of simulation data.
8. **Additive noise:** Consider the inhomogeneous random differential equation,

$$\frac{d\mathbf{u}}{dt} = [\mathbf{A}_0 + \alpha \mathbf{A}_1(t)] \cdot \mathbf{u} + \mathbf{f}(t),$$

where $\mathbf{A}_1(t)$ and $\mathbf{f}(t)$ are correlated, and $\mathbf{f}(t) = \mathbf{f}_0 + \mathbf{f}_1(t)$. We seek an approximate evolution equation for the mean $\langle \mathbf{u} \rangle$.

- (a) Notice that the inhomogeneous equation can be written as a *homogeneous* equation if the state space is expanded,

$$\frac{d}{dt} \begin{pmatrix} \mathbf{u} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{f} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ 1 \end{pmatrix},$$

allowing the methods developed in this section to be readily applied. To that end, show that,

$$\exp \left[\tau \begin{pmatrix} \mathbf{A}_0 & \mathbf{f}_0 \\ 0 & 0 \end{pmatrix} \right] = \begin{pmatrix} e^{\tau \mathbf{A}_0} & \frac{e^{\tau \mathbf{A}_0} - 1}{\mathbf{A}_0} \mathbf{f}_0 \\ 0 & 1 \end{pmatrix}.$$

- (b) Derive an equation for $\langle \mathbf{u} \rangle$.
9. **Parametric resonance:** It is challenging to compute the evolution equation for the average process $\langle u \rangle$ characterized by the stochastic Mathieu equation, Eq 8.40. The primary difficulty lies in the time dependence of the coefficients which results in the deterministic Mathieu equation not having a closed form-solution.
- (a) Write out the formal solution of the *deterministic* Mathieu equation (*i.e.*, $\alpha = 0$ in Eq. 8.40) as a series of iterated integrals (the Born approximation).

- (b) For narrowly-peaked correlation function, we can invoke Watson's lemma (Section B.5.2) to express integral (8.42) as a series in the correlation time τ_c . Write out the details of this approximation and thereby arrive at the renormalized equation for $\langle u \rangle$.
- (c) The stability of the *mean* of an oscillatory process is not particularly informative – much more important is the *energy stability* determined by the stability of $\langle u^2 \rangle + \langle \dot{u}^2 \rangle$. Use the Redfield equations (Exercise 7) to generate approximate equations governing $\langle u^2 \rangle$, $\langle u \dot{u} \rangle$ and $\langle \dot{u}^2 \rangle$. Under what conditions are these asymptotically stable? Verify the approximation results using an ensemble of simulation data.

10. **Random differential equations with Markov noise:** For the random differential equation $du/dt = F(u, t; Y(t))$, where $u \in \mathbb{R}^N$ and $Y(t)$ is a Markov process having probability density $\Pi(y, t|y_0, t_0)$ obeying the master equation $d\Pi/dt = \mathbb{W}\Pi$, the joint probability density $P(u, y, t|u_0, y_0, t_0)$ obeys its own master equation,

$$\frac{\partial P}{\partial t} = - \sum_{i=1}^N \frac{\partial}{\partial u_i} F_i(u, t; y) P + \mathbb{W}P. \quad (8.47)$$

This equation is often not very useful since $\mathbb{W}$ can be quite large (even infinite). Nevertheless, Eq. 8.47 is exact and it makes no assumptions about the correlation time of the fluctuations.

- (a) Show that if the random differential equation is linear (with coefficients that do not depend upon time),

$$\frac{du_i}{dt} = \sum_{j=1}^N A_{ij}(Y(t)) u_j,$$

then Eq. 8.47 reduces to a particularly simple form.

- (b) Define the marginal averages m_i ,

$$m_i(y, t) = \int u_i P(u, y, t) du.$$

For the linear system introduced in (10a), find the evolution equation for m_i . These equations must be solved with initial

values $m_i(y, 0) = u_{0i}\Pi(y, 0)$. The average of the process of interest is given by the integral,

$$\langle u_i(t) \rangle = \int m_i(y, t) dy.$$

- (c) If $Y(t)$ is a *dichotomic Markov process* (i.e. a process taking one of two values ± 1 with transition rate γ ; see Exercise 4 on page 76), then the joint probability is simply the two-component vector $P(u, \pm 1, t) \equiv [P_+(u, t) \ P_-(u, t)]$ with transition matrix,

$$\mathbb{W} = \begin{bmatrix} -\gamma & \gamma \\ \gamma & -\gamma \end{bmatrix}.$$

Consider the random differential equation $\dot{u} = -i\omega(t)u$ where u is complex and $\omega(t)$ is a random function of time. This example was introduced by Kubo to illustrate the effect of fluctuations on spectral line broadening. Let $\omega(t) = \omega_0 + \alpha\xi(t)$, where $\xi(t)$ is a dichotomic Markov process. Show that Eq. 8.47 gives two coupled equations,

$$\begin{aligned} \frac{\partial P_+}{\partial t} &= i \frac{\partial}{\partial u} (\omega_0 + \alpha)uP_+ - \gamma P_+ + \gamma P_-, \\ \frac{\partial P_-}{\partial t} &= i \frac{\partial}{\partial u} (\omega_0 - \alpha)uP_- - \gamma P_- + \gamma P_+. \end{aligned}$$

Find the equations for the two marginal averages $m_{\pm}(t)$ using the initial condition $m_{\pm}(0) = \frac{1}{2}a$. Find $\langle u(t) \rangle = m_+(t) + m_-(t)$. What can you say about the two limits $\gamma \ll \alpha$ (*slow* fluctuations) and $\gamma \gg \alpha$ (*fast* fluctuations)?

11. Stochastic delay differential equations: Read Section 8.5.

- (a) To determine the stability of the stochastic delay differential equation is not easy. Re-write the resolvent equation by non-dimensionalizing all of the parameters appearing in Eq. 8.46 with respect to the delay time τ_d , and introduce the perturbation parameter $\varepsilon = \frac{\tau_c}{\tau_d}$.
- (b) Assume the correlation time is short compared with the delay time ($\varepsilon = \frac{\tau_c}{\tau_d} \ll 1$), and develop a perturbation series for $s = s_0 + \varepsilon s_1 + \varepsilon^2 s_2 + \dots$. The zero-order term s_0 is given in terms of Lambert functions. Find an expression for s_1 .

- (c) Derive the analogue of Eq. 8.45 for a *system* of delay differential equations, *i.e.* find an expression for $\langle \hat{\boldsymbol{x}}(s) \rangle$ where $\boldsymbol{x}(t)$ is an n -dimensional vector.

CHAPTER 9

MACROSCOPIC EFFECTS OF NOISE

The great success of deterministic ordinary differential equations in mathematical modeling tends to foster the intuition that including fluctuations will simply distort the deterministic signal somewhat, leading to a distribution of states that follows the macroscopic solution. In that sense, studying stochastic processes seems to be an interesting hobby, but not very practical. We have seen, however, that the statistics of the fluctuations can be used to measure parameters that are not accessible on an observable scale – for example, Avagadro’s number as determined by Perrin from Einstein’s work on Brownian motion (page 8), Johnson’s determination of Avagadro’s number from thermal noise in resistors using Nyquist’s analysis (page 46), and even the mutation rate in bacteria (see Exercise 5 on page 23). In this Chapter, we consider some examples where stochastic models exhibit behaviour that is different from their deterministic counterparts¹: Unstable equilibrium points at 0 become stable (*Keizer’s paradox*, Section 9.1), or stable systems that exhibit regular oscillations when fluctuations are included (Sections 9.2 and 9.3.1). In these cases, stochastic models provide insight into system behaviour that is simply not available from deterministic formulations.

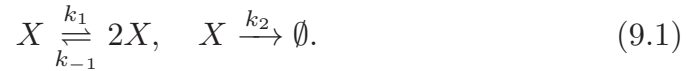
¹*Deterministic counterpart* means the deterministic ordinary differential equations obtained in the macroscopic limit of the master equation, *i.e.*, number of individuals and the volume go to infinity, while the *density* is held fixed. Alternatively, the zero’th order term in the linear noise approximation.

9.1 Keizer's Paradox

- J. Keizer *Statistical thermodynamics of nonequilibrium processes* (Springer-Verlag, 1987), p. 164–169.
- M. Vellela and H. Qian (2007) “A quasistationary analysis of a stochastic chemical reaction: Keizers paradox,” *Bulletin of Mathematical Biology* **69**: 1727.

The fluctuations described by the master equation come about, in part, because the systems under study are composed of *discrete* particles. It should be no surprise then that as the number of particles decreases, the discrete nature of the constituents will become manifest. So, for example, a fixed point that is unstable with respect to *infinitesimal* perturbations is stable if approached along a discrete lattice. This is Keizer's paradox. We shall examine his argument in more detail below, but it is important to emphasize that Keizer used his example to illustrate how correct macroscopic behaviour is sometimes not exhibited by the master equation, suggesting caution in interpreting asymptotic behaviour of the master equation literally. Some authors, however, have taken variations of Keizer's paradox to suggest that the master equation formalism is somehow fundamentally flawed, or that the macroscopic limit is problematic. Such claims should not be taken too seriously.

As a concrete example, consider the following (nonlinear) autocatalytic reaction scheme:



The deterministic equation governing the concentration of the species X is given by the differential equation,

$$\frac{dX}{dt} = (k_1 - k_2)X - k_{-1}X^2. \quad (9.2)$$

The two equilibrium points for the model are,

$$X^{ss} = 0 \quad \text{and} \quad X^{ss} = \frac{(k_1 - k_2)}{k_{-1}}. \quad (9.3)$$

It is straightforward to show that $X^{ss} = 0$ is *unstable*, while the equilibrium point $X^{ss} = \frac{(k_1 - k_2)}{k_{-1}}$ is *stable*. Re-cast as a master equation, the probability $p(n, t)$ of the system having n molecules of X at time t obeys the (nonlinear) master equation,

$$\frac{dp}{dt} = \hat{k}_1 (\mathbf{E}^{-1} - 1) np + \hat{k}_2 (\mathbf{E} - 1) np + \hat{k}_{-1} (\mathbf{E} - 1) n(n-1)p, \quad (9.4)$$

where we have absorbed the volume V into the definition of the rate constants: $\hat{k}_1 = k_1/V$, etc. Writing out each component of $p(n)$ as a separate element in the vector p_n , the master equation can be expressed in terms of the transition matrix $\mathbb{W}$ (Exercise 1b),

$$\frac{dp}{dt} = \mathbb{W} \cdot p. \quad (9.5)$$

It is more convenient to separate each element in Eq. 9.5 as a coupled system of ordinary differential equations,

$$\begin{aligned} \frac{dp_0}{dt} &= \hat{k}_2 p_1, \\ \frac{dp_1}{dt} &= 2(\hat{k}_{-1} + \hat{k}_2) p_2 - (\hat{k}_1 + \hat{k}_2) p_1, \\ \frac{dp_2}{dt} &= \hat{k}_1 p_1 - 2(\hat{k}_{-1} + \hat{k}_1 + \hat{k}_2) p_2 + 3(2\hat{k}_{-1} + \hat{k}_2), \dots \end{aligned}$$

Quoting from Vellela and Qian,

By induction, all probabilities p_n for $n > 0$ are zero. Because the sum of all the probabilities must add to one, this forces $p_0 = 1$. So we have,

$$p^*(0) = 1 \quad \text{and} \quad p^*(n) = 0, \quad n > 0, \quad (9.6)$$

as the probability distribution for the unique steady state of the stochastic model. Note that the steady state of the deterministic model are fixed points, and the steady state of the stochastic model has a distribution. The stochastic model shows that eventually there will be no X left in the system. This is in striking contrast to the previous deterministic model, which predicts that the concentration of X will stabilize at the nonzero x^* , while $x^* = 0$ is unstable. This is the Keizers paradox.

This ‘paradox’ is of the same form as Zermelo’s paradox discussed on page 67 and is resolved in precisely the same fashion. For systems with a reasonable number of molecules, say $X(0) = 1000$, the paradox disappears. Quoting from Keizer (p. 166),

Since n_X^0 is large, the probability of n_X taking on other values will grow, and in a time $\tau_1 = 1/\hat{k}_1$ it will become a sharp

Gaussian centered at $n_X^\infty = V\hat{k}_1/2k_2$ [*the macroscopic equilibrium point*]. On a much longer time scale, the order of $\tau_2 = \tau_1 e^{V\hat{k}_1/2k_2}$, the systems of the ensemble in the low- n_X tail of the Gaussian will fall into the absorbing state at $n_X = 0$. This creates a new peak near $n_X = 0$ and, if one waits for the period of time τ_2 , almost all members of the ensemble will have no X molecules. The important point here is the difference in the two time scales τ_1 and τ_2 . For a large system, V becomes infinite. This has no effect on $\hat{k}_1$, which is independent of the volume, whereas τ_2 depends strongly on the volume and becomes infinite like e^V . Thus to witness the probability in Eq. 9.6 one would have to wait a time of the order of $10^{10^{23}}$ times longer than it takes to achieve the distribution centered at $n_X^\infty = V\hat{k}_1/2k_2$. This is tantamount to the probability in Eq. 9.6 being unobservable.

We will generally be interested in systems where we are able to apply the linear noise approximation. Under the conditions of applicability of this approximation, the paradoxical distinction between the behaviour of the master equation and its deterministic counterpart disappears. As Keizer says, “if used uncritically, [the master equation] can lead to meaningless results.” There *are* situations where the behaviour of a stochastic system, even with a large number of molecules, is obviously in conflict with the predictions of the deterministic model, coming from mechanisms distinct from the separation of time scales underlying Keizer’s paradox. We shall examine two examples in the following sections. Other deviant effects are described in:

- M. S. Samoilov and A. P. Arkin (2006) Deviant effects in molecular reaction pathways. *Nature biotechnology* **24**: 1235-1240.

9.2 Oscillations from Underdamped Dynamics

Before we consider quantification of noise-induced oscillations, we will collect a review of relevant background scattered through previous chapters. Looking back to the multivariate linear noise approximation (Eq. 6.10 on page 126), we found that to leading order in the system size Ω , the

fluctuations α obey the linear Fokker-Planck equation,

$$\sqrt{\Omega}^{-1} : \quad \frac{\partial \Pi}{\partial t} = - \sum_{i,j} \Gamma_{ij} \partial_i (\alpha_j \Pi) + \frac{1}{2} \sum_{i,j} D_{ij} \partial_{ij} \Pi, \quad (9.7)$$

where $\partial_i \equiv \frac{\partial}{\partial \alpha_i}$ and,

$$\Gamma_{ij}(t) = \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathbf{x}(t)} \quad \mathbf{D} = \mathbf{S} \cdot \text{diag}[\nu] \cdot \mathbf{S}^T, \quad (9.8)$$

where, again, $\mathbf{f}$ are the macroscopic reaction rates, $\mathbf{S}$ is the stoichiometry matrix and $\mathbf{nu}$ is the vector of reaction propensities. The matrices $\mathbf{\Gamma}$ and $\mathbf{D}$ are independent of α , which appears only linearly in the drift term. As a consequence, the distribution $\Pi(\alpha, t)$ will be Gaussian for all time. In particular, at equilibrium $\mathbf{\Gamma}_s$ and $\mathbf{D}_s$ will be constant and the fluctuations are distributed with density,

$$\Pi_s(\alpha) = \left[(2\pi)^d \det \mathbf{\Xi}_s \right]^{\frac{1}{2}} \exp \left[-\frac{1}{2} \alpha^T \cdot \mathbf{\Xi}_s^{-1} \cdot \alpha \right],$$

and variance $\mathbf{\Xi}_s = \langle \langle \alpha \cdot \alpha^T \rangle \rangle$ determined by,

$$\mathbf{\Gamma}_s \cdot \mathbf{\Xi}_s + \mathbf{\Xi}_s \cdot \mathbf{\Gamma}_s^T + \mathbf{D}_s = 0. \quad (9.9)$$

Finally, the steady-state autocorrelation is exponentially distributed (see Eq. 5.18 on page 5.18),

$$\mathbf{B}(t) = \langle \alpha_s(t) \cdot \alpha_s^T(0) \rangle = \exp [\mathbf{\Gamma}_s t] \cdot \mathbf{\Xi}_s. \quad (9.10)$$

In Chapter 2, we introduced the fluctuation spectrum $S(\omega)$ (see Section 2.3 on page 39),

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} B(\tau) d\tau \iff B(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega. \quad (9.11)$$

which is the Fourier transform of the autocorrelation function, and provides the distribution of the frequency content of the fluctuations. In multivariate systems, it is straightforward to show that the Fourier transform of Eq. 9.10 is (Exercise 2),

$$\mathbf{S}_\alpha(\omega) = \frac{1}{2\pi} [\mathbf{\Gamma}_s + \mathbf{I}i\omega]^{-1} \cdot \mathbf{D}_s \cdot [\mathbf{\Gamma}_s^T - \mathbf{I}i\omega]^{-1}. \quad (9.12)$$

Since $\mathbf{S}_\alpha(\omega)$ contains the distribution of the frequency content of the fluctuations, narrow peaks in $\mathbf{S}_\alpha(\omega)$ indicate coherent noise-induced oscillations, as we discuss below.

Finally, in Chapter 6 (on page 6.5), it was shown that the linear Fokker-Planck equation is equivalent to the Langevin equation. For a multivariate system, the same result holds, and we find that Eq. 9.7 is equivalent to,

$$\frac{d\alpha}{dt} = \mathbf{\Gamma}_s \cdot \alpha + \mathbf{B} \cdot \eta, \quad (9.13)$$

where $\mathbf{D}_s = \mathbf{B}\mathbf{B}^T$ and η is a vector of uncorrelated unit-variance white noise with the same dimensions as the reaction propensity vector ν .

Physical insight into the origin of noise-induced oscillations comes from re-writing this equation more suggestively,

$$\frac{d\alpha}{dt} - \mathbf{\Gamma}_s \cdot \alpha = \mathbf{B} \cdot \eta, \quad (9.14)$$

where now the dynamics of α about the deterministic steady-state are mapped to a *harmonic oscillator with white-noise forcing*. It should be clear that the frequency response of the system is given by the eigenvalues of the deterministic Jacobian $\mathbf{\Gamma}_s$, and that if any of the eigenvalues of $\mathbf{\Gamma}_s$ have a non-zero imaginary part, then the system will selectively amplify the components of the white-noise forcing that lie near to the resonance frequency of the system.

Said another way, looking at the spectrum at the level of species *concentration*,

$$S_\mathbf{x}(\omega) = \frac{1}{2\pi} [\mathbf{\Gamma}_s + \mathbf{I}i\omega]^{-1} \cdot \frac{\mathbf{D}_s}{\Omega} \cdot [\mathbf{\Gamma}_s^T - \mathbf{I}i\omega]^{-1}, \quad (9.15)$$

if the eigenvalues of $\mathbf{\Gamma}_s$ are complex (i.e., underdamped), then noise-induced oscillations are inevitable and these oscillations will have amplitude proportional to $\frac{\mathbf{D}_s}{\Omega}$. In the deterministic limit ($\Omega \rightarrow \infty$; constant concentration), the amplitude of the oscillations will of course vanish, recovering the deterministic stability implied by the negative real parts of the eigenvalues of $\mathbf{\Gamma}_s$.

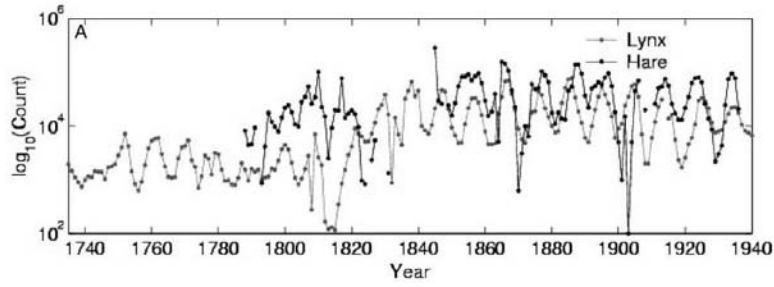


Figure 9.1: **Oscillations in Predator-Prey Populations.** Sample of lynx and hare populations taken from Pineda-Krch, Blok, Dieckmann and Doebeli (2007) *Oikos* **116**: 53–64; Figure 5A. Data is from Elton and Nicholson (1942) *J. Anim. Ecol.* **11**: 215–244. Notice the vertical axis is *logarithmic*.

9.2.1 Example – Malthus Predator-Prey Dynamics

McKane AJ, Newman TJ (2005) “Predator-prey cycles from resonant amplification of demographic stochasticity,” *Physical Review Letters* **94** (21): Art. No. 218102.

Following McKane and Newman, we consider noise-induced oscillations in predatory-prey models used in theoretical ecology. Some very brief historical context helps to illuminate the significance of McKane and Newman’s analysis.

Predatory-prey populations often exhibit oscillatory dynamics (Figure 9.1). One of the earliest models of predator-prey dynamics is the *Lotka-Volterra model* (1925-1926), governing the birth, death and predation among a population of predators Z and their prey P ,

$$\begin{aligned}\frac{dP}{dt} &= \alpha \cdot P - \beta \cdot Z \cdot P, \\ \frac{dZ}{dt} &= \gamma \cdot Z \cdot P - \delta Z.\end{aligned}\tag{9.16}$$

Here:

α : birth rate of prey

β : rate of predation of P by Z

γ : efficiency of the predator to turn food into offspring

δ : death rate of predators.

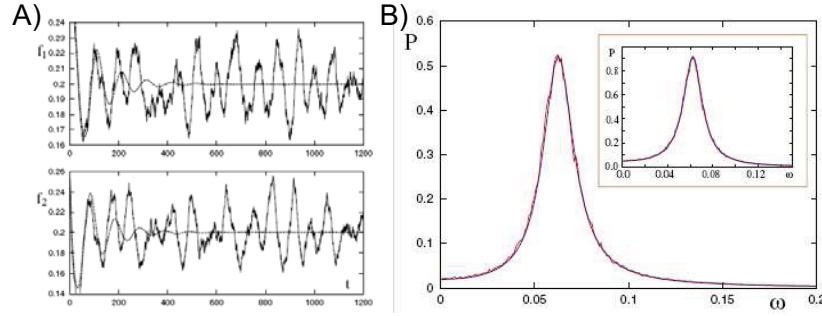


Figure 9.2: **Noise-induced oscillations in predator-prey models.** **A)** Sample stochastic simulation of the modified Lotka-Volterra model, Eq. 9.17. The deterministic trajectory rapidly approaches equilibrium (dashed line). Noise-induced oscillations are evident in the simulation data. **B)** Fluctuation spectrum for the predator population extracted from an ensemble of 500 stochastic simulations, compared with the analytic approximation, Eq. 9.15. The inset corresponds to the power spectrum for the prey. Taken from McKane and Newman (2005).

While these equations do indeed exhibit oscillatory dynamics, the eigenvalues are *pure imaginary*, so the magnitude of the oscillations depends precisely upon the initial conditions. Physically, it makes more sense for the model to evolve toward some stable limit cycle, irrespective of the initial conditions.

A simple modification of the Lotka-Volterra model that attempts to increase the relevance of the equations is to limit the exponential growth of the prey P by introducing a death term corresponding to overcrowding,

$$\begin{aligned}\frac{dP}{dt} &= \alpha \cdot P \cdot \left(1 - \frac{P}{K}\right) - \beta \cdot Z \cdot P, \\ \frac{dZ}{dt} &= \gamma \cdot Z \cdot P - \delta Z,\end{aligned}\tag{9.17}$$

where K is called the “carrying capacity of the environment,” and corresponds to the maximum prey population that can be sustained by the environment. Unfortunately, this modified set of equations *no longer oscillates*. (Exercise 4). Despite the fact that Eqs. 9.17 do not oscillate *deterministically*, they do exhibit noise-induced oscillations!

Figure 9.2A shows an example trajectory of the predator-prey dynamics from a stochastic simulation. Figure 9.2B compares the theoretical spectrum (Eq. 9.15) to the spectrum extracted from an ensemble of 500

stochastic simulations. Obviously, the agreement is quite good.

9.3 Effective Stability Analysis (ESA)

- M. Scott, T. Hwa and B. Ingalls (2007) “Deterministic characterization of stochastic genetic circuits,” *Proceedings of the National Academy of Sciences USA* **104**: 7402.

Oscillations in the predator-prey model above comes from driving with full-spectrum white noise a system with complex eigenvalues. Deviations from deterministic behaviour can occur in systems with all-real eigenvalues, as well. Here, the interplay between intrinsic noise and nonlinear transition rates leads to a noise-induced loss of stability. Depending upon the underlying phase-space, the loss of stability may simply bounce the state among multiple fixed points, or, in some cases, generate coherent noise-induced oscillations.

The effective stability approximation is a method by which the effect of the noise is used to renormalize the eigenvalues of the system, linearized about a deterministic fixed point, providing conditions on the model parameters for which stability is lost in the stochastic model. In analogy with linear stability analysis of ordinary nonlinear differential equations, the effective stability approximation will demonstrate that the system is unstable, but cannot say anymore than that – in particular, it cannot be used to provide the fluctuation spectrum of the noise-induced oscillations.

The approximation method combines the linear noise approximation of Chapter 4 (page 100) to characterize the fluctuations, and Bourret’s approximation from Chapter 8 (page 177) to renormalize the eigenvalues. Both will be briefly reviewed below.

To calculate the stability of the macroscopic model $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$ to small perturbations, the system is linearized about the equilibrium point: $\mathbf{x} = \mathbf{x}_s + \mathbf{x}_p$,

$$\frac{d}{dt}\mathbf{x}_p = \mathbf{J}^{(0)} \cdot \mathbf{x}_p. \quad (9.18)$$

The eigenvalues of the Jacobian $\mathbf{J}^{(0)} = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_s}$ provide the decay rate of the exponential eigenmodes; if all the eigenvalues have negative real part, we say the system is *locally asymptotically stable*.

To accommodate fluctuations on top of the small perturbation $\mathbf{x}_p$, we set $\mathbf{x} = \mathbf{x}_s + \mathbf{x}_p + \omega \boldsymbol{\alpha}(t)$, where we have written $\omega \equiv \sqrt{\Omega}^{-1}$ to keep the

notation compact. The Jacobian

$$\mathbf{J} \equiv \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_s+\omega\boldsymbol{\alpha}},$$

will then be a (generally) nonlinear function of the fluctuations about the steady-state $\boldsymbol{\alpha}(t)$. In the limit $\omega \rightarrow 0$, we can further linearize $\mathbf{J}$ with respect to ω ,

$$\mathbf{J} \approx \mathbf{J}|_{\omega \rightarrow 0} + \omega \left. \frac{\partial \mathbf{J}}{\partial \omega} \right|_{\omega \rightarrow 0} \equiv \mathbf{J}^{(0)} + \omega \mathbf{J}^{(1)}(t).$$

The stability equation is then given by,

$$\frac{d}{dt} \mathbf{x}_p = [\mathbf{J}^{(0)} + \omega \mathbf{J}^{(1)}(t)] \cdot \mathbf{x}_p + \text{zero mean terms.} \quad (9.19)$$

This is a linear stochastic differential equation with random coefficient matrix $\mathbf{J}^{(1)}(t)$ composed of a linear combination of the steady-state fluctuations $\boldsymbol{\alpha}(t)$ which have *non-zero* correlation time (*cf.* Eq. 9.10).

Our present interest is in the *mean stability* of the equilibrium point. Taking the ensemble average of Eq. 9.19,

$$\frac{d}{dt} \langle \mathbf{x}_p \rangle = \mathbf{J}^{(0)} \cdot \langle \mathbf{x}_p \rangle + \omega \left\langle \mathbf{J}^{(1)}(t) \cdot \mathbf{x}_p \right\rangle.$$

The right-most term is the cross-correlation between the process $\mathbf{x}_p$ and the coefficient matrix $\mathbf{J}^{(1)}(t)$. Since the correlation time of $\mathbf{J}^{(1)}(t)$ is not small compared with the other time scales in the problem, it cannot be replaced by white noise, and an approximation scheme must be developed to find a closed evolution equation for $\langle \mathbf{x}_p \rangle$.

By assumption, the number of molecules is large so the parameter ω is small, although not so small that intrinsic fluctuations can be ignored. To leading-order in ω , the trajectory $\mathbf{x}_p(t)$ is a random function of time since it is described by a differential equation with random coefficients. Derivation of the entire probability distribution of $\mathbf{x}_p(t)$ is usually impossible, and we must resort to methods of approximation. We shall adopt the closure scheme of Bourret (Eq. 8.19, page 177) to arrive at a deterministic equation for the evolution of the averaged process $\langle \mathbf{x}_p(t) \rangle$ in terms of only the first and second moments of the fluctuations. In that approximation, provided $\mathbf{J}^{(0)} \gg \omega \mathbf{J}^{(1)}$, the dynamics of $\langle \mathbf{x}_p \rangle$ are governed by the convolution equation,

$$\frac{d}{dt} \langle \mathbf{x}_p(t) \rangle = \mathbf{J}_0 \langle \mathbf{x}_p(t) \rangle + \omega^2 \int_0^t \mathbf{J}_c(t-\tau) \langle \mathbf{x}_p(\tau) \rangle d\tau, \quad (9.20)$$

where $\mathbf{J}_c(t - \tau) = \langle \mathbf{J}^{(1)}(t) e^{\mathbf{J}^{(0)}(t-\tau)} \mathbf{J}^{(1)}(\tau) \rangle$ is the time autocorrelation matrix of the fluctuations. The equation can be solved formally by Laplace transform,

$$\langle \hat{\mathbf{x}}_{\mathbf{p}}(s) \rangle = \left[s\mathbf{I} - \mathbf{J}^{(0)} - \omega^2 \hat{\mathbf{J}}_c(s) \right]^{-1} \langle \mathbf{x}_p(0) \rangle,$$

where now $\hat{\mathbf{J}}_c(s) = \int_0^t \mathbf{J}_c(t) e^{-st} dt$. A necessary and sufficient condition for asymptotic stability of the averaged perturbation modes $\langle \mathbf{x}_p(t) \rangle$ is that the roots λ' of the resolvent,

$$\det \left[\lambda' \mathbf{I} - \mathbf{J}_0 - \omega^2 \hat{\mathbf{J}}_c(\lambda') \right] = 0, \quad (9.21)$$

all have negative real parts ($Re(\lambda') < 0$). Some insight into the behavior of the system can be gained by considering a perturbation expansion of the effective eigenvalues λ' in terms of the small parameter ω . We further diagonalize $\mathbf{J}^{(0)}$, $\text{diag}[\lambda_i] = \mathbf{P}^{-1} \cdot \mathbf{J}^{(0)} \cdot \mathbf{P}$, and provided the eigenvalues are distinct, we can explicitly write λ'_i in terms of the unperturbed eigenvalues λ_i to $O(\omega^4)$ as,

$$\lambda'_i = \lambda_i + \omega^2 \left[\mathbf{P}^{-1} \cdot \hat{\mathbf{J}}_c(\lambda_i) \cdot \mathbf{P} \right]_{ii}, \quad (9.22)$$

where $[\cdot]_{ii}$ denotes the i^{th} diagonal entry of the matrix. Notice the matrix product $\mathbf{J}_c(t)$ contains linear combinations of the correlation of the fluctuations $\langle \alpha_i(t) \alpha_j(0) \rangle$. These are simply given by the exponential autocorrelation function derived above, Eq. 9.10. In the next section, we will apply the effective stability approximation to a model of an excitable oscillator, using the method to construct a phase-plot in parameter space illustrating regions of noise-induced oscillations.

9.3.1 Example – Excitable Oscillator

- Vilar, J., H.Y. Kueh, N. Barkai, and S. Leibler (2002) “Mechanisms of noise-resistance in genetic oscillators,” *Proceedings of the National Academy of Sciences USA* **99**, 5988-5992.

We consider the generic model proposed by Vilar and co-workers to describe circadian rhythms in eukaryotes, with a transcriptional autoactivator driving expression of a repressor that provides negative control by sequestering activator proteins through dimerization. The repressor and activator form an inert complex until the activator degrades, recycling

repressor back into the system. In their model, the degradation rate of the activator, δ_A , is the same irrespective of whether it is bound in the inert complex or free in solution. We simplify their original model somewhat, and assume fast activator/DNA binding along with rapid mRNA turnover, leading to a reduced set of rate equations governing the concentration of activator A , repressor R and the inert dimer C ,

$$\begin{aligned}\frac{dA}{dt} &= \gamma_A \cdot g\left(\frac{A}{K_A}, f_A\right) - \delta_A \cdot A - \kappa_C \cdot A \cdot R \\ \frac{dR}{dt} &= \gamma_R \cdot g\left(\frac{A}{K_R}, f_R\right) - \delta_R \cdot R - \kappa_C \cdot A \cdot R + \delta_A \cdot C \\ \frac{dC}{dt} &= \kappa_C \cdot A \cdot R - \delta_A \cdot C.\end{aligned}\tag{9.23}$$

Here, the function g ,

$$g(x, f) = \frac{1 + f \cdot x}{1 + x},\tag{9.24}$$

characterizes the response of the promoter to the concentration of the regulatory protein A . The fold-change in the synthesis rate, f , is $\gg 1$ because A is an activator. In this complicated system, there are many dimensionless combinations of parameters that characterize the system dynamics. The scaled repressor degradation rate $\epsilon = \delta_R/\delta_A$ is a key control parameter in the deterministic model since oscillations occur only for an intermediate range of this parameter. For the nominal parameter set used in the Vilar paper, the deterministic model exhibits oscillations over the range $0.12 < \epsilon < 40$ (Figure 9.3a, black region). We shall focus on the parameter regime near to the phase boundary at $\epsilon \approx 0.12$ and examine the role intrinsic noise plays in generating regular oscillations from a deterministically stable system.

Applying the ESA to the oscillator model, the parameter $\Delta_{b_A} = (b_A + 1)/(2 \cdot K_A \cdot V_{cell})$ emerges as an important measure quantifying the discreteness in activator synthesis. Here, b_A is the burst size in the activator synthesis (see p. 113), K_A is the activator/DNA dissociation constant and V_{cell} is the cell volume (with $V_{cell} = 100\mu m^3$ as is appropriate for eukaryotic cells.)

The nominal parameter set of Vilar leads to a burstiness in activator synthesis of $b_A = 5$ (giving $\Delta_{b_A} = 6 \times 10^{-2}$) and a burstiness in repressor synthesis of $b_R = 10$. The phase boundary predicted by the ESA is shown as a solid line in Figure 9.3a, bounding a region of parameter space between the deterministic phase boundary where qualitatively different behavior is expected from the stochastic model. We examine the

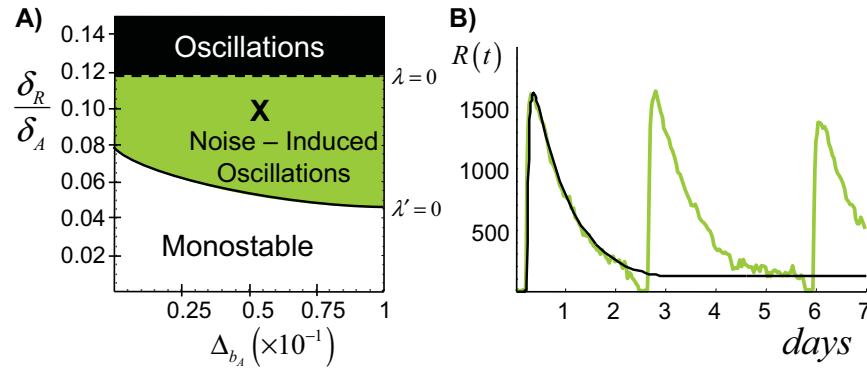


Figure 9.3: **Oscillations in the excitable system.** **A)** Phase plot shows a region of noise induced oscillations. **B)** Stochastic simulation of a deterministically stable system (black line; denoted by a cross in the left panel) shows oscillations (grey line).

system behavior in this region by running a stochastic simulation using the parameter choice $\epsilon = 0.1$ and $\Delta_{b_A} = 6 \times 10^{-2}$ (denoted by a cross in Figure 9.3a). With this choice, the deterministic model is stable (Figure 9.3b, black line). Nevertheless, a stochastic simulation of the same model, with the same parameters including protein bursting and stochastic dimerization, clearly shows oscillations (Figure 9.3b, green line).

Suggested References

Beyond the references cited above, Keizer's paradox is discussed by Gillespie in his seminal paper on stochastic simulation:

- D. T. Gillespie (1977) "Exact simulation of coupled chemical reactions," *Journal of Chemical Physics* **81**: 2340.

He takes the view that the stochastic simulation trajectories show very clearly the immense time scales involved before the system reaches the anomalous stability of the fixed point at the origin.

Another method of characterizing noise-induced oscillations is the multiple-scales analysis of Kuske and coworkers:

- R. Kuske, L. F. Gordillo and P. Greenwood. (2006) "Sustained oscillations via coherence resonance in SIR," *Journal of Theoretical Biology* **245**: 459-469.

Like the method of McKane and Newman discussed in the text, the method of Kuske allows the noise-induced oscillations to be fully characterized in terms of fluctuation spectra and oscillation amplitude.

Exercises

1. **Keizer's paradox:** Stochastic models often exhibit behaviour that is in contrast to their deterministic counter-parts. In Section 9.1, we discussed a particular autocatalytic network. The details left out of the main text will be filled in below.
 - (a) Use linear stability analysis to prove the claim that the equilibrium point $X^{ss} = 0$ is *unstable* in the deterministic model.
 - (b) Derive the master equation that corresponds to the autocatalytic network shown in Eq. 9.1. What is the explicit form of the transition matrix $\mathbb{W}$ appearing in Eq. 9.5?
 - (c) Write a stochastic simulation of the example discussed in the text. Initializing the system at the macroscopic equilibrium point, show sample trajectories to argue that for all intents and purposes, the macroscopic equilibrium point is stable. Do this for various choices of the system size.
2. **Derivation of the multivariate frequency spectrum:** Derive Eq. 9.12 by taking the multivariate Fourier transform of the correlation function $\mathbf{B}(t)$. Notice that $\mathbf{B}(t)$, as written in Eq. 9.10, is valid for $t > 0$. Find the expression for the autocorrelation function for $t < 0$ in order to be able to compute the Fourier integral over the interval $t \in (-\infty, \infty)$. Use the fluctuation-dissipation relation to put the spectrum in the form shown in Eq. 9.12.
3. **Resonance in driven linear systems:** Some systems can be excited into oscillation through periodic forcing.
 - (a) *Damped harmonic oscillator.* Solve

$$\frac{dy}{dt} + (a + ib)y = e^{i\omega t} \quad (a > 0, i \equiv \sqrt{-1}),$$

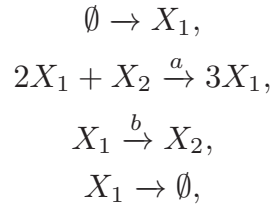
with an arbitrary initial condition. Show that in the limit $t \rightarrow \infty$, the influence of the initial condition vanishes, and the amplitude of y is maximum when $b = \omega$.

- (b) *2D Langevin equation.* Consider the harmonic oscillator driven by white noise $\eta(t)$,

$$\frac{dy}{dt} + (a + ib)y = \eta(t) \quad (a > 0).$$

Write out the white noise as a Fourier series. What can you say about the $t \rightarrow \infty$ amplitude maximum of y in this case?

4. **Predator-prey models:** In the predator-prey models discussed in Section 9.2, several details were suppressed – these will be made more explicit below.
- (a) Find the fixed points of the Lotka-Volterra model, Eq. 9.16, and determine the stability of the system linearized about those fixed points. What can you say about stochastic trajectories in the predator-prey phase space?
 - (b) Find the fixed points of the modified Lotka-Volterra model, Eq. 9.17, and determine the stability of the system linearized about those fixed points. What can you say about stochastic trajectories in the predator-prey phase space for this model?
 - (c) For what choice of parameters would you expect noise-induced oscillations in the modified Lotka-Volterra model (Eq. 9.17)?
5. **Noise-induced oscillations in the Brusselator:** In Section 4.2 on p. 88, the *Brusselator* model was introduced,



corresponding to the deterministic rate equations,

$$\begin{aligned} \frac{dX_1}{dt} &= 1 + aX_1^2X_2 - (b+1)X_1, \\ \frac{dX_2}{dt} &= -aX_1^2X_2 + bX_1. \end{aligned}$$

Compute the power spectrum for the fluctuations about the steady-state for $\langle X^2 \rangle$, $\langle XY \rangle$ and $\langle Y^2 \rangle$. What are the conditions for noise-induced oscillations in this model?

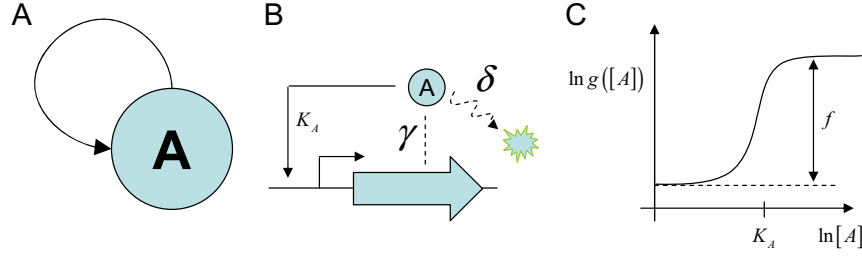


Figure 9.4: **Exercise 6 – The autoactivator.** **A)** Schematic of the positive feedback loop. **B)** More detailed view of a genetic circuit that encodes a protein that activates its own rate of transcription. **C)** The activation function $g([A])$ increases (nonlinearly) with increasing activator concentration.

6. **Effective stability approximation of the autoactivator:** The vector-matrix formalism of the effective stability approximation leads to unwieldy expressions for the effective eigenvalues in systems with several state variables. The one-species autoactivator model, by contrast, allows the approximation to be computed explicitly.

The autoactivator is a genetic regulatory motif with a product A that stimulates its own synthesis through a positive feedback loop (see Figure 9.4). Each transcription event triggers the translation of b activator proteins, where b is the burst parameter (see Section 5.2). The deterministic rate equation for this system is,

$$\frac{dA}{dt} = \gamma \cdot g(A) - \delta \cdot A,$$

where,

$$g(A) = \frac{1 + f \cdot (A/K_A)^n}{1 + (A/K_A)^n} \quad (n > 1).$$

- (a) Drawing the synthesis and degradation rates on the same log-log plot, decide how many stable fixed points the deterministic system has. What is the eigenvalue of the system? Identify the dimensionless combinations of parameters that fully characterize this system, and draw a phase plot showing the boundaries where the number of stable fixed points changes. Hypothesize about what noise will do to this system.

- (b) Write out explicitly the propensity vector $\boldsymbol{\nu}$ and the stoichiometry matrix $\mathbf{S}$ for the stochastic version of this model.
- (c) Write out the diffusion matrix $D = \mathbf{S} \cdot \text{diag}[\boldsymbol{\nu}] \cdot \mathbf{S}^T$, and identify an additional dimensionless combination that characterizes the *stochastic behaviour* of the system. Provide a physical interpretation of this parameter.
- (d) Compute the effective eigenvalue, written in terms of the dimensionless parameters and without writing out $g(A)$ explicitly. What can you say about the noise-induced correction?

CHAPTER 10

SPECIAL TOPICS

This final chapter contains supplemental topics that don't really fit in the rest of the notes. They are either suggested extensions of topics covered in previous chapters, or very short discussions of topics that may be of interest to some readers, along with a list of relevant references.

10.1 Random Walks and Electric Networks

- P. G. Doyle and J. L. Snell, *Random Walks and Electrical Networks* (Mathematical Association of America, 1984).
- S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, 2001).

There is a direct analogy between the escape probability for a random walk along a lattice (Figure 10.1a) and the potential along the nodes of an array of resistors (Figure 10.1b). The equivalence is made precise by observing the correspondence of the steady-state master equation governing the probability and the Kirchoff laws governing the potential.

Beyond the analogy, recasting a high-dimensional random walk in terms of the electrical properties of a high-dimensional array of resistors allows many difficult theorems to be proved with ease. Consider, for example, the question of recurrence in high-dimensional random walks. If a random walker is guaranteed to return to the origin at some point in the wandering, then the walk is called *recurrent*. If there is some probability

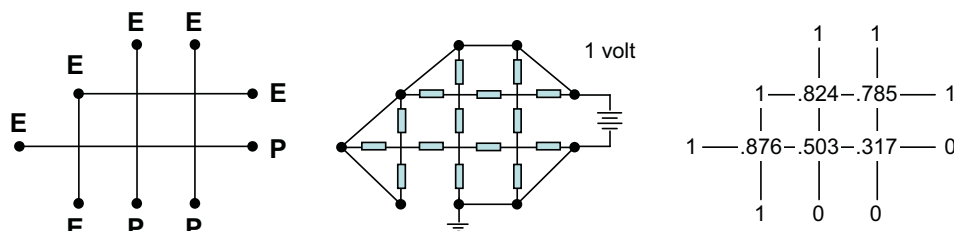


Figure 10.1: **Equivalence of a random walk and the potential along a resistor array.** A) Escape probability from a two-dimensional lattice for a random walker. The nodes marked E are escape points, while the nodes marked P are police. B) An array of $1\ \Omega$ resistors, with boundary points held at 1 volt, or grounded. C) The escape probability, or equivalently the potential, at the interior nodes. Redrawn from Figures 2.1 and 2.2 of Doyle and Snell *Random Walks and Electrical Networks* (1984).

of never returning, then the walk is *transient*. Pólya proved the following theorem:

PÓLYA'S THEOREM: A simple random walk on a d -dimensional lattice is recurrent for $d = 1, 2$ and transient for $d > 2$.

In the language of resistor arrays, Pólya's theorem can be restated as: the random walk in d -dimensions is recurrent if and only if the resistance to infinity is infinite. Estimating the resistance at infinity is a far simpler approach than Pólya's original proof of the problem.

10.2 Fluctuations Along a Limit Cycle

- K. Tomita, T. Ohta and H. Tomita (1974) "Irreversible circulation and orbital revolution," *Progress of Theoretical Physics* **52**: 1744.
- F. Ali and M. Menzinger (1999) "On the local stability of limit cycles," *Chaos* **9**: 348.
- Scott M, Ingalls B, Kaern M (2006) Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos* **16**: 026107.

In Section 5.1.2, we used the linear noise approximation to estimate the statistics of the steady-state fluctuations in the Brusselator model. What makes the Brusselator model interesting is that over a range of parameter values, the system exhibits a stable *limit cycle*. With a change

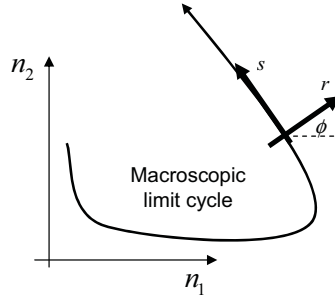


Figure 10.2: **Rotating change of basis along the limit cycle.** To separate the fluctuations tangent to the limit cycle from those perpendicular, we make a rotating change of basis from n_1 - n_2 space to r - s space.

of basis, the linear noise approximation can again be used to characterize the fluctuations around the limit cycle. That change of basis is the subject of the present section.

Limit cycle regime $b > 1 + a$

From the Fokker-Planck equation for $\Pi(\alpha_1, \alpha_2, t)$ far away from the critical line $b = 1 + a$, Eq. 5.23, we obtain the evolution equations for the variance of the fluctuations,

$$\frac{dC_{ij}}{dt} = \sum_m \Gamma_{im} C_{mj} + \sum_n \Gamma_{jn} C_{in} + D_{ij},$$

where we have written $C_{ij} = C_{ji} = \langle \alpha_i \alpha_j \rangle$. In the parameter regime where the macroscopic system follows a limit cycle, the coefficients $\mathbf{\Gamma}$ and $\mathbf{D}$ will be periodic functions of time, and the equations governing the variance will not admit a stable steady-state. Physically, there is no mechanism to control fluctuations tangent to the limit cycle, and the variance grows unbounded in that direction. Trajectories perturbed away from the limit cycle are drawn back, however, and so we expect the variance of fluctuations *perpendicular* to the limit cycle to reach a steady-state. We introduce a change of basis, with s tangent to the limit cycle, and r perpendicular (Figure 10.2). The transformation matrix to the new coordinate system is given by the rotation:

$$\mathbf{U}(t) = \begin{bmatrix} \cos \phi(t) & \sin \phi(t) \\ -\sin \phi(t) & \cos \phi(t) \end{bmatrix},$$

where $\phi(t)$ is related to the macroscopic rates of x_1 and x_2 , (denoted by

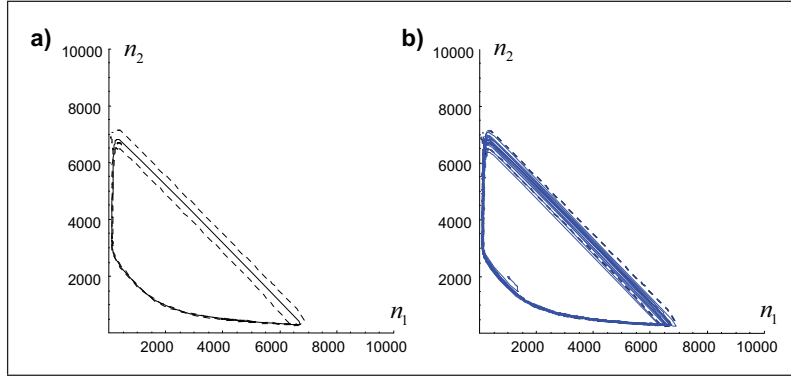


Figure 10.3: **Fluctuations about the limit cycle.** **a)** Although the variance of the fluctuations tangent to the limit cycle grows without bound, the perpendicular fluctuations are confined to a gully of width C_{rr} , shown as a dashed line. **b)** The blue curve is the result of a stochastic simulation of the system. The trajectory is confined to the a region very close to the macroscopic limit cycle.

f_1 and f_2 , respectively),

$$\begin{aligned} -\sin \phi(t) &= \frac{f_1(t)}{\sqrt{f_1^2(t) + f_2^2(t)}}, \\ \cos \phi(t) &= \frac{f_2(t)}{\sqrt{f_1^2(t) + f_2^2(t)}}. \end{aligned}$$

With the change of basis, the evolution equation for the variance $\mathbf{C}'$ in the r - s coordinate system is given by,

$$\frac{d\mathbf{C}'}{dt} = (\mathbf{\Gamma}' + \mathbf{R}) \mathbf{C}' + [(\mathbf{\Gamma}' + \mathbf{R}) \mathbf{C}']^T + \mathbf{D}',$$

where $\mathbf{\Gamma}'$ and $\mathbf{D}'$ are the transformed matrices,

$$\begin{aligned} \mathbf{\Gamma}' &= \mathbf{U}(t) \cdot \mathbf{\Gamma}(t) \cdot \mathbf{U}^T(t), \\ \mathbf{D}' &= \mathbf{U}(t) \cdot \mathbf{D}(t) \cdot \mathbf{U}^T(t), \end{aligned}$$

and $\mathbf{R}$ is the rate of rotation of the coordinate frame itself,

$$\mathbf{R} = \frac{d\phi}{dt} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

In this new coordinate system, the variance C_{rr} decouples from the tangential fluctuations, and is characterized by the *single* evolution equation,

$$\frac{dC_{rr}}{dt} = 2\Gamma'_{rr}C_{rr} + D'_{rr},$$

that converges to a stable limit cycle. After the transients have passed, the trajectory in the original n_1 - n_2 coordinate system can be anywhere along the limit cycle, but tangential fluctuations are confined to a narrow gully with perpendicular variance C_{rr} , shown in Figure 10.3a as a dashed curve confining the macroscopic limit cycle. Here, $a = 5$, $b = 10$, and $\Omega = 1000$. The blue curve (Figure 10.3b) is a stochastic simulation of the same system (compare with Gillespie (1977), Figure 19). Again, the linear noise approximation captures the statistics of the randomly generated trajectories very well.

10.3 Stochastic Resonance

- B. McNamara and K. Wiesenfeld (1989) “Theory of stochastic resonance,” *Physical Review A* **39**: 4854.

Recall from Section 6.7 on page 143 that the escape probability from a potential well depends *exponentially* upon the well depth. As a consequence, small changes in the well-depth can have exponentially larger effects on the hopping rate between wells.

For example, suppose in addition to a symmetric, two-well potential there is a small, periodic perturbation to the potential (Figure 10.4),

$$U(x) \rightarrow U(x) + \varepsilon \cos \omega_p t \quad (\varepsilon \ll 1).$$

The particle in the well is subject to Brownian motion of variance D , and for a fixed fixed signal magnitude ε and frequency ω_s , the system will exhibit a maximum in the signal-to-noise ratio as the white noise variance D is varied (Figure 10.5). Here, the *signal* is the component of the power spectrum centered on ω_s .

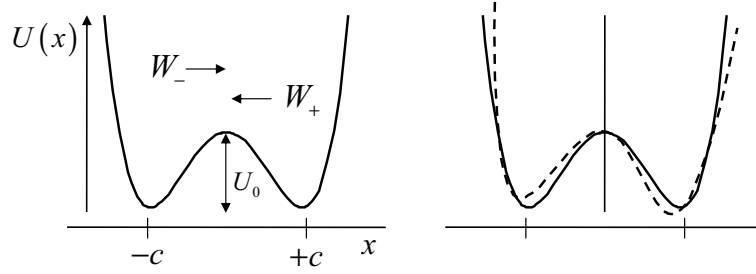


Figure 10.4: **Stochastic resonance.** A) Consider the hopping of a particle between the wells of a *symmetric* double-well potential. The rate of hopping from $-c$ to c is the same as the rate of hopping from c to $-c$. B) Impose upon this potential a second, very small perturbation that results in the *tilting* of the double-well with a prescribed frequency ω_p – called the *signal*. Over a range of noise strengths, the hopping probability will become slaved to the external perturbation. As the magnitude of the noise is increased, the signal-to-noise ratio will attain a maximum. This ability of the system to turn noise into an amplification of a weak signal (rather than a corruption) is called *stochastic resonance*. Redrawn from Figure 2 of McNamara and Wiesenfeld (1989).

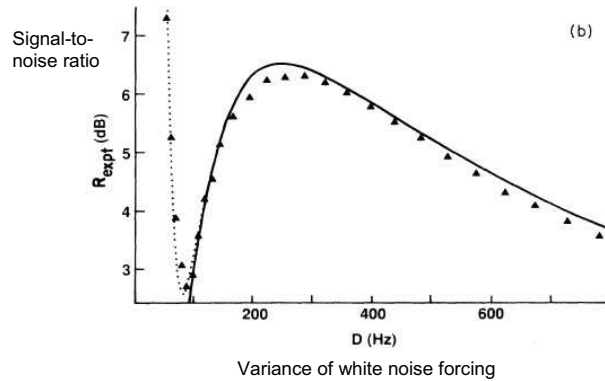


Figure 10.5: **Stochastic resonance.** As the variance in the white noise forcing (D) is varied, the component of the power spectrum lying at frequency ω_s is enhanced well above the background white noise. This effect is manifest as a maximum in the signal-to-noise ratio. Redrawn from Figure 9b of McNamara and Wiesenfeld (1989).

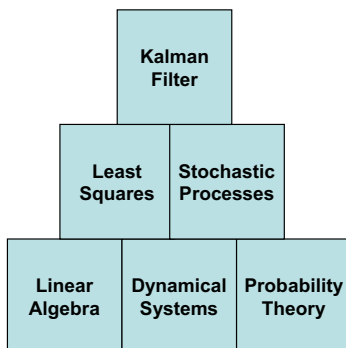


Figure 10.6: **Mathematical foundations of the Kalman filter.** The Kalman filter combines fundamental ideas spread across several areas to obtain an estimate of the state in the presence of noise in the system *and* noise in the observations. Redrawn from Figure 1.1 of Grewal and Andrews *Kalman Filtering* (2001).

10.4 Kalman Filter

- R. E. Kalman (1960) “A new approach to linear filtering and prediction problems,” *Transaction of the ASME Journal of Basic Engineering* **82**: 35.
- M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and practice using MATLAB*, 2nd Ed. (Wiley Interscience, 2001)

A problem that is of central concern in control theory is how to estimate the trajectory of a system, perturbed by noise, given a history of observations corrupted by noise. For example, given the following model for the trajectory x with known input u , and the set of observations z ,

$$\frac{dx}{dt} = Fx + G\eta(t) + Cu \quad (10.1)$$

$$z = Hx + \xi(t) + Du, \quad (10.2)$$

where $\eta(t)$ and $\xi(t)$ are zero-mean (uncorrelated) white noise,

$$\langle \eta(t_1)\eta(t_2) \rangle = Q(t_1)\delta(t_1 - t_2), \quad (10.3)$$

$$\langle \xi(t_1)\xi(t_2) \rangle = R(t_1)\delta(t_1 - t_2), \quad (10.4)$$

$$\langle \eta(t_1)\xi(t_2) \rangle = 0, \quad (10.5)$$

how can the state of the system x be estimated? Obviously in the absence of noise, $\hat{x} = H^{-1}z$ provides an exact estimate of the state. The question

is: with the state dynamics and the observation data corrupted by noise, how can we optimally extract the estimate $\hat{x}$?

First, we must define in what sense an estimate is optimal. We seek an estimate $\hat{x}(t)$ which is a linear function of the observations $z(t)$ which minimizes

$$\hat{x}(t) = \min_{\hat{x}} \langle [x(t) - \hat{x}(t)]^T \cdot M \cdot [x(t) - \hat{x}(t)] \rangle,$$

where M is a given symmetric positive-definite matrix. We say that $\hat{x}(t)$ is optimal in *the mean-squared sense*, or that $\hat{x}(t)$ is a least-squares estimate of $x(t)$. To be precise, the continuous-time filter is called the *Kalman-Bucy* filter, while the discrete-time filter is called the *Kalman* filter. The discrete-time implementation is more straightforward to explain.

Discrete Kalman filter

We have a state x_k that evolves through (known) deterministic dynamics Φ and some white noise forcing w_k ,

$$x_{k+1} = \Phi \cdot x_k + w_k,$$

where w_k is a vector of random variables drawn from a normal distribution $\mathbf{N}(0, Q_k)$. Our knowledge of the true state x_k comes from observations z_k that are also corrupted by white noise v_k

$$z_k = H \cdot x_k + v_k,$$

where v_k is a vector of random variables drawn from a normal distribution $\mathbf{N}(0, R_k)$, that are all independent of w_k .

Given some initial estimate of the state $\hat{x}_0$ (here, the $\hat{}$ indicates an *estimate* of the unknown state x_k) and the variance of our estimate $P_0 = \langle \hat{x}_0 \cdot \hat{x}_0^T \rangle$, the Kalman filter provides an algorithm for updating our estimate of the state $\hat{x}_0$ and the variance P_0 using an auxiliary quantity called the *Kalman gain matrix*. The Kalman gain matrix K_k tells us how much to trust the observation z_k in refining our estimate $\hat{x}_k$ beyond the deterministic dynamics Φ - This is done in such a way that the estimate $\hat{x}_k$ is optimal in the mean-squared sense described above.

Specifically, we use Φ to form an initial estimate of the state $\hat{x}_k(-)$ and of the variance $P_k(-)$,

$$\begin{aligned} \hat{x}_k(-) &= \Phi_{k-1} \cdot \hat{x}_{k-1}(+), \\ P_k(-) &= \Phi_{k-1} \cdot P_{k-1}(+) \cdot \Phi_{k-1}^T + Q_{k-1}. \end{aligned}$$

Then we use the Kalman gain matrix K_k ,

$$K_k = P_k(-) \cdot H_k^T \cdot [H_k \cdot P_k(-) \cdot H_k^T + R_k]^{-1},$$

and the observation z_k to refine our original estimates,

$$\begin{aligned}\hat{x}_k(+) &= \hat{x}_k(-) + K_k \cdot [z_k - H_k \cdot \hat{x}_k(-)], \\ P_k(+) &= [I - K_k \cdot H_k] \cdot P_k(-),\end{aligned}$$

and the process is repeated at each time-step.

10.5 Novikov-Furutsu-Donsker Relation

- E. A. Novikov (1965) “Functionals and the random-force method in turbulence theory,” *Soviet Physics – JETP* **20**: 1290–1294.
- K. Furutsu (1963) “On the statistical theory of electromagnetic waves in a fluctuating medium,” *Journal of the Research of the National Bureau of Standards* **D67**: 303.
- M. D. Donsker M.D (1964) “On function space integrals” in *Analysis in Function Space*, eds. W.T. Martin & I. Segal, MIT Press; pp 17–30.

For a Gaussian-distributed multiplicative noise source $A_1(t)$,

$$\frac{dy}{dt} = A_0 y + A_1(t)y, \quad (10.6)$$

the Novikov theorem allows the correlation of the process $y(t)$ depending implicitly upon $A_1(t)$ to be calculated in terms of the autocorrelation of $A_1(t)$ and the functional derivative of $y(t)$ with respect to the fluctuations. To leading order, we simply recover Bourret’s approximation, although the Novikov theorem does open the way to a variational estimate of higher-order terms as shown in Section 10.5.1.

If we average the evolution equation (10.6) directly, we have,

$$\frac{d}{dt} \langle y(t) \rangle = A_0 \langle y(t) \rangle + \alpha \langle A_1(t) y(t) \rangle, \quad (10.7)$$

for the scalar function $y(t)$. The correlation of the coefficient $A_1(t)$ and the process $y(t)$ can be calculated using Novikov’s theorem,

$$\langle A_1(t) y(t) \rangle = \int_0^t \langle A_1(t) A_1(t') \rangle \left\langle \frac{\delta y(t)}{\delta A_1(t')} \right\rangle dt', \quad (10.8)$$

where $\left\langle \frac{\delta y(t)}{\delta A_1(t')} \right\rangle$ is the *functional derivative* of $y(t)$ with respect to the Gaussian random function $A_1(t')$. The Green's function $G_0(t, t')$ for the noiseless operator $\left[\frac{d}{dt} - A_0\right]$ allows $y(t)$ to be re-written as an integral equation,

$$y(t) = G_0(t, 0) y(0) + \alpha \int_0^t G_0(t, \tau) A_1(\tau) y(\tau) d\tau.$$

Taking the functional derivative with respect to $A_1(t')$,

$$\begin{aligned} \frac{\delta y(t)}{\delta A_1(t')} &= 0 + \alpha \int_0^t G_0(t, \tau) \left[\delta(\tau - t') y(\tau) + A_1(\tau) \frac{\delta y(\tau)}{\delta A_1(t')} \right] d\tau \\ \frac{\delta y(t)}{\delta A_1(t')} &= \alpha G_0(t, t') y(t') + \alpha \int_0^t G_0(t, \tau) A_1(\tau) \frac{\delta y(\tau)}{\delta A_1(t')} d\tau, \end{aligned} \quad (10.9)$$

which is an integral equation for the functional derivative that can be solved iteratively if α is small. Retaining only the leading term, and averaging

$$\left\langle \frac{\delta y(t)}{\delta A_1(t')} \right\rangle \approx \alpha G_0(t, t') \langle y(t') \rangle.$$

[In the Section 10.5.1, we consider a variational approximation of the higher-order terms]. With substitution into (10.8), (10.7) becomes,

$$\frac{d}{dt} \langle y(t) \rangle = A_0 \langle y(t) \rangle + \alpha^2 \int_0^t \langle A_1(t) A_1(t') \rangle G_0(t, t') \langle y(t') \rangle dt'$$

which is identical to Bourret's approximation of the evolution equation for the first-moment (8.19) written in the original notation with $G_0(t, t') = e^{A_0(t-t')}$ and $y(t) = e^{A_0(t-t')} y(t')$.

* * * * *

This proof of Eq. 10.8 follows the original by Novikov, with annotations. In full, the Novikov-Furutsu-Donsker relation reads:

The correlation of a Gaussian random variable $\gamma(t)$ with a function depending upon that variable $x[\gamma(t); t]$ is:

$$\langle \gamma(t) x(t) \rangle = \int \langle \gamma(t) \gamma(t_1) \rangle \left\langle \frac{\delta x(t)}{\delta \gamma(t_1)} \right\rangle dt_1.$$

Proof :

The functional $x[\gamma]$ is expanded as a functional Taylor series about the point $\gamma(t) = 0$:

$$\begin{aligned} x[\gamma(t)] &= \\ x[0] + \int \left(\frac{\delta x}{\delta \gamma_1} \right) \Big|_{\gamma=0} \gamma(t_1) dt_1 + \frac{1}{2} \int \left(\frac{\delta^2 x}{\delta \gamma_1 \delta \gamma_2} \right) \Big|_{\gamma=0} \gamma(t_1) \gamma(t_2) dt_1 dt_2 + \dots \\ &= x[0] + \sum_{n=1}^{\infty} \frac{1}{n!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right) \Big|_{\gamma=0} \gamma(t_1) \dots \gamma(t_n) dt_1 \dots dt_n. \end{aligned} \quad (10.10)$$

Here, for brevity,

$$\left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \equiv \left(\frac{\delta^n x(t)}{\delta \gamma(t_1) \dots \delta \gamma(t_n)} \right)_{\gamma(t_1)=\dots=\gamma(t_n)=0}.$$

The functional derivative is evaluated at $\gamma = 0$ (zero noise) and as such constitutes a non-random function so that

$$\left\langle \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \right\rangle = \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0}.$$

Multiplying (10.10) by $\gamma(t)$, and taking the ensemble average,

$$\langle \gamma(t) x(t) \rangle = \sum_{n=1}^{\infty} \frac{1}{n!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \langle \gamma(t) \gamma(t_1) \dots \gamma(t_n) \rangle dt_1 \dots dt_n. \quad (10.11)$$

For a Gaussian random variable, the mean value of the product of an odd number of terms vanishes, while for an even number of terms, the mean value is equal to the sum of products of all pair-wise combinations:

$$\langle \gamma(t) \gamma(t_1) \dots \gamma(t_n) \rangle = \sum_{\alpha=1}^n \langle \gamma(t) \gamma(t_\alpha) \rangle \langle \gamma(t_1) \dots \gamma(t_{\alpha-1}) \gamma(t_{\alpha+1}) \dots \gamma(t_n) \rangle,$$

(where $\langle \gamma(t_1) \dots \gamma(t_{\alpha-1}) \gamma(t_{\alpha+1}) \dots \gamma(t_n) \rangle$ can obviously be further divided into pair-wise combinations). The functional derivatives in the integrals of (10.10) are all symmetric with respect to the labelling of the arguments $t_1, \dots, t_n$, so that there are n equivalent ways to choose α above,

$$\langle \gamma(t) \gamma(t_1) \dots \gamma(t_n) \rangle = n \langle \gamma(t) \gamma(t_1) \rangle \langle \gamma(t_2) \dots \gamma(t_n) \rangle.$$

Substituting into the cross-correlation (10.11),

$$\begin{aligned} \langle \gamma(t) x(t) \rangle &= \\ \int \langle \gamma(t) \gamma(t_1) \rangle &\left[\sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \langle \gamma(t_2) \dots \gamma(t_n) \rangle dt_2 \dots dt_n \right] dt_1. \end{aligned} \quad (10.12)$$

On the other hand, taking the functional derivative of (10.10) directly, again taking advantage of the symmetry,

$$\begin{aligned} \frac{\delta x[\gamma(t)]}{\delta \gamma(t')} &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \frac{\delta \gamma(t_1)}{\delta \gamma(t')} \gamma(t_2) \dots \gamma(t_n) dt_1 \dots dt_n \\ &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \dots \delta \gamma_n} \right)_{\gamma=0} \delta(t_1 - t') \gamma(t_2) \dots \gamma(t_n) dt_1 \dots dt_n \\ &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int \left(\frac{\delta^n x}{\delta \gamma' \delta \gamma_2 \dots \delta \gamma_n} \right)_{\gamma=0} \gamma(t_2) \dots \gamma(t_n) dt_2 \dots dt_n. \end{aligned}$$

Call $t' \equiv t_1$, and take the ensemble average,

$$\left\langle \frac{\delta x(t)}{\delta \gamma(t_1)} \right\rangle = \left[\sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int \left(\frac{\delta^n x}{\delta \gamma_1 \delta \gamma_2 \dots \delta \gamma_n} \right)_{\gamma=0} \langle \gamma(t_2) \dots \gamma(t_n) \rangle dt_2 \dots dt_n \right]. \quad (10.13)$$

Substituting the left-side of (10.13) into (10.12) yields the Novikov-Furutsu-Donsker relation. For a causal process, $x(t)$ will only depend upon values of $\gamma(t_1)$ that precede it, so,

$$\boxed{\langle \gamma(t) x(t) \rangle = \int_0^t \langle \gamma(t) \gamma(t_1) \rangle \left\langle \frac{\delta x(t)}{\delta \gamma(t_1)} \right\rangle dt_1.} \quad (10.14)$$

10.5.1 Variational Approximation

- F. J. Dyson (1949) “The radiation theories of Tomonago, Schwinger, and Feynman,” *Physical Review* **75**: 486–502.

- L. Rosenberg and S. Tolchin (1973) “Variational methods for many body problems,” *Physical Review A* **8**: 1702–1709.

Beginning with equation (10.9),

$$\frac{\delta y(t)}{\delta A_1(t')} = \alpha G_0(t, t') y(t') + \alpha \int G_0(t, \tau) A_1(\tau) \frac{\delta y(\tau)}{\delta A_1(t')} d\tau,$$

it is convenient to assume a form for the functional derivative

$$\frac{\delta y(t)}{\delta A_1(t')} = \alpha G(t, t') y(t'), \quad (10.15)$$

for some noiseless function $G(t, t')$. With substitution,

$$G(t, t') = G_0(t, t') + \alpha \int G_0(t, \tau) A_1(\tau) G(\tau, t') d\tau. \quad (10.16)$$

This is sometimes called *Dyson's equation*, and a variational principle derived from this equation has a long history (see Rosenberg and Tolchin). A variational estimate of $G(t, t')$ correct to first order is:

$$\begin{aligned} G_v(t, t') &= G_0(t, t') + \alpha \int G_1(t, \tau) A_1(\tau) G_0(\tau, t') d\tau \\ &+ \alpha \int G_0(t, \tau) A_1(\tau) G_2(\tau, t') d\tau - \alpha \int G_1(t, \tau) A_1(\tau) G_2(\tau, t') d\tau \\ &+ \alpha^2 \int \int G_1(t, \tau_1) A_1(\tau_1) G_0(\tau_1, \tau_2) A_1(\tau_2) G_2(\tau_2, t') d\tau_1 d\tau_2. \end{aligned} \quad (10.17)$$

Here, $G_1(t, t')$ and $G_2(t, t')$ are two trial functions. The stationarity of this estimate with respect to independent variations in the trial functions about the exact solution can easily be demonstrated using (10.16),

$$\begin{aligned} &\left. \frac{\delta G_v(t, t')}{\delta G_1(t_1, t'_1)} \right|_{G_1=G_2=G} = \\ &= -\alpha \delta(t - t_1) A_1(t'_1) \times \\ &\times \left[G(t'_1, t') - G_0(t'_1, t') - \alpha \int G_0(t'_1, \tau) A_1(\tau) G(\tau, t') d\tau \right] \\ &= 0, \end{aligned}$$

and similarly for variations in $G_2(t, t')$. The variational approximation can be used as the starting point for a Raleigh-Ritz scheme to determine

the optimal trial functions via some expansion of $G_1(t, t')$ and $G_2(t, t')$ in a suitable orthonormal basis. Here, however, we shall use $G_v(t, t')$ directly.

By assumption, $G(t, t')$ is a noiseless function, while $G_v(t, t')$ is not, so we take the ensemble average of (10.17) (assuming the trial functions are deterministic),

$$\begin{aligned} G(t, t') &\approx \langle G_v(t, t') \rangle = \\ &G_0(t, t') + \alpha \int G_1(t, \tau) \langle A_1(\tau) \rangle G_0(\tau, t') d\tau \\ &+ \alpha \int G_0(t, \tau) \langle A_1(\tau) \rangle G_2(\tau, t') d\tau - \alpha \int G_1(t, \tau) \langle A_1(\tau) \rangle G_2(\tau, t') d\tau \\ &+ \alpha^2 \int \int G_1(t, \tau_1) G_0(\tau_1, \tau_2) G_2(\tau_2, t') \langle A_1(\tau_1) A_1(\tau_2) \rangle d\tau_1 d\tau_2. \end{aligned}$$

$\langle G_v(t, t') \rangle$ still obeys the variational principle, as one can show by multiplying (10.16) by $A_1(\tau_1)$, and taking the ensemble average (provided the trial functions are not random). Furthermore, the averaged solution should provide a superior estimate of $G(t, t')$ than the one obtained by successive iteration of the original integral equation (10.9).

With the convenient choice $G_1(t, t') = G_2(t, t') = G_0(t, t')$ and for zero-mean noise, the averaged solution is considerably simplified:

$$\begin{aligned} G(t, t') &\approx G_0(t, t') + \\ &\alpha^2 \int \int G_0(t, \tau_1) G_0(\tau_1, \tau_2) G_0(\tau_2, t') \langle A_1(\tau_1) A_1(\tau_2) \rangle d\tau_1 d\tau_2 \end{aligned}$$

Substituting back into the variational derivative (10.15),

$$\begin{aligned} \frac{\delta y(t)}{\delta A_1(t')} &= \alpha G_0(t, t') y(t') + \\ &+ \alpha^3 \left[\int \int G_0(t, \tau_1) G_0(\tau_1, \tau_2) G_0(\tau_2, t') \langle A_1(\tau_1) A_1(\tau_2) \rangle d\tau_1 d\tau_2 \right] y(t'). \end{aligned}$$

Finally, the Novikov relation reads:

$$\begin{aligned} \langle A_1(t) y(t) \rangle &= \alpha \int G_0(t, t') \langle A_1(t) A_1(t') \rangle \langle y(t') \rangle dt' + \\ &+ \alpha^3 \int \left[\int \int G_0(t, \tau_1) G_0(\tau_1, \tau_2) G_0(\tau_2, t') \langle A_1(\tau_1) A_1(\tau_2) \rangle d\tau_1 d\tau_2 \right] \times \\ &\quad \times \langle A_1(t) A_1(t') \rangle \langle y(t') \rangle dt' \end{aligned}$$

The first term is Bourret's approximation, while the second term is the higher-order correction obtained from the variational principle.

APPENDIX A

REVIEW OF CLASSICAL PROBABILITY

The mathematical theory of probability deals with phenomena *en masse*, *i.e.* observations (experiments, trials) which can be repeated many times under similar conditions. Its principal concern is with the numerical characteristics of the phenomenon under study, *i.e.* quantities which take on various numerical values depending upon the result of the observation. Such quantities are called *random variables*; *e.g.*

- the number of points which appear when a die is tossed.
- the number of calls which arrive at a telephone station during a given time interval.
- the lifetime of an electric light bulb.
- the error made in measuring a physical quantity...

A random variable ξ is regarded as specified if one knows the *distribution function*,

$$F(x) = P\{\xi \leq x\}, \quad x \in (-\infty, \infty) \quad (\text{A.1})$$

where P is the probability of occurrence of the relation in $\{\cdot\}$. The distribution function $F(x)$ can be shown to have the following properties,

1. $F(-\infty) = 0; \quad F(\infty) = 1;$
2. $x_1 < x_2 \Rightarrow F(x_1) < F(x_2)$

3. $F(x^+) = F(x)$ (continuity from the right)

It is convenient to introduce the *probability density function*,

$$f(x) = \frac{dF}{dx}, \quad (\text{A.2})$$

of the random variable ξ . Since $F(x)$ might not have a derivative for every x , one can distinguish several types of random variables – in particular,

1. **Continuous random variables.** We assume that the number of points where $f(x)$ doesn't exist is a countable set (*i.e.* points of discontinuity are comparatively “few”). Then it can be shown that,

- $2 \Rightarrow f(x) \geq 0$.
- 1 and A.2 $\Rightarrow \int_{-\infty}^{\infty} f(x) dx = 1$.
- A.2 $\Rightarrow F(x) = \int_{-\infty}^x f(x') dx'$, *i.e.* $F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$.

From this, and from $P\{x_1 \leq \xi \leq x_2\} = F(x_2) - F(x_1)$, it follows that,

$$P\{x_1 \leq \xi \leq x_2\} = \int_{x_1}^{x_2} f(x) dx.$$

In particular, if Δx is sufficiently small, then,

$$P\{x \leq \xi \leq x + \Delta x\} \approx f(x)\Delta x,$$

and so one also has the definition,

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P\{x \leq \xi \leq x + \Delta x\}}{\Delta x}.$$

Note this definition leads to a *frequency interpretation* for the probability density function. To determine $f(x)$ for a given x , we perform the experiment n times, and count the number of trials $\Delta n(x)$ such that $x \leq \xi \leq x + \Delta x$, so

$$f(x)\Delta x \approx \frac{\Delta n(x)}{n}$$

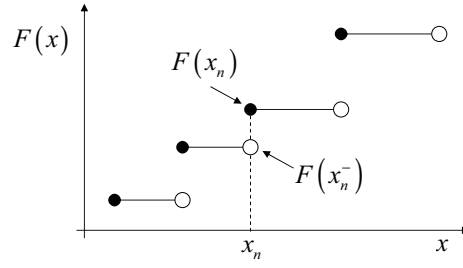


Figure A.1: **Typical distribution function for a discrete process.**

2. **Discrete (and mixed) random variables.** These have a distribution function $F(x)$ that resembles a staircase (figure A.1). We have,

$$p_n = P\{\xi = x_n\} = F(x_n) - F(x_n^-) \quad (\text{A.3})$$

$$\sum_n p_n = F(\infty) - F(-\infty) = 1 \quad (\text{A.4})$$

In general,

$$F(x) = \sum_n P\{\xi = x_n\}, \quad \text{such that } x_n \leq x. \quad (\text{A.5})$$

If a random variable is not of continuous type, then one doesn't usually associate with it a density function, since $F(x)$ is not differentiable in the ordinary sense. Nevertheless, we extend the concept of function to include distributions – in particular, the Dirac δ -function, specified by the integral property,

$$\int_{-\infty}^{\infty} \phi(x) \delta(x - x_0) dx = \phi(x_0),$$

where ϕ is continuous at x_0 , though otherwise arbitrary.

Then one can show that if $F(x)$ is discontinuous at x_0 , we have,

$$\left. \frac{dF}{dx} \right|_{x=x_0} = k \cdot \delta(x - x_0); \quad k = F(x_0^+) - F(x_0^-) \quad (\text{the step height}).$$

In particular, if

$$F(x) = \theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

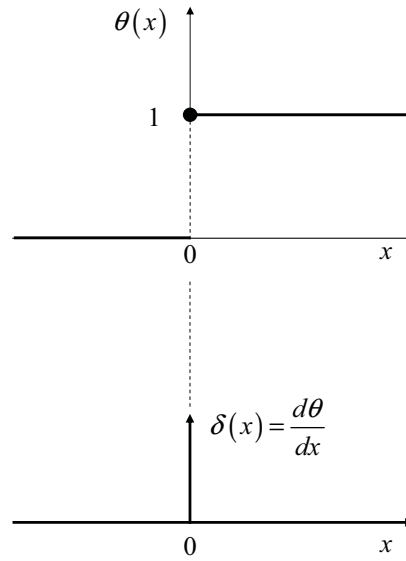


Figure A.2: The Dirac Delta function $\delta(x)$ is a distribution that can be thought of as the *derivative* of the Heaviside step function $\theta(x)$.

then,

$$\frac{d\theta}{dx} = \delta(x)$$

(see figure A.2). Using this, Eqs. A.4 and A.5 give,

$$f(x) = \sum_n p_n \delta(x - x_n).$$

An important, and subtle, result is the following.

Existence theorem: Given a function $G(x)$ such that,

- $G(-\infty) = 0; \quad G(\infty) = 1;$
- $x_1 < x_2 \Rightarrow G(x_1) < G(x_2)$
- $G(x^+) = G(x)$ (continuity from the right)

one can find an experiment, and an associated random variable, such that its distribution function $F(x) = G(x)$.

Similarly, one can determine a random variable having as a density a given function $g(x)$, provided $g(x) \geq 0$ and $\int_{-\infty}^{\infty} g(x) dx = 1$. This is done by determining first,

$$G(x) = \int_{-\infty}^x g(x') dx',$$

and then invoking the existence theorem.

A.1 Statistics of a Random Variable

1. Expectation value:

$$\mathbf{E}\{\xi\} = \begin{cases} \int_{-\infty}^{\infty} x f(x) dx, & \text{if } \xi \text{ is continuous} \\ \sum_n x_n p_n, & \text{if } \xi \text{ is discrete} \end{cases}$$

2. Variance:

$$\sigma^2 = \mathbf{E}\{(\xi - \mathbf{E}\{\xi\})^2\} = \begin{cases} \int_{-\infty}^{\infty} (x - \mathbf{E}\{\xi\})^2 f(x) dx \\ \sum_n (x_n - \mathbf{E}\{\xi\})^2 p_n \end{cases}$$

from which we have the important result:

$$\sigma^2 = \mathbf{E}\{\xi^2\} - \mathbf{E}^2\{\xi\}.$$

Note that σ^2 is also called the *dispersion*, and σ , the *standard deviation*. If the variance is vanishingly small, then the random variable ξ becomes a deterministic, or *sure*, variable:

$$\sigma^2 = 0 \iff \xi \text{ is the sure variable } \langle \xi \rangle. \quad (\text{A.6})$$

3. Moments:

$$m_k = \mathbf{E}\{\xi^k\} = \begin{cases} \int_{-\infty}^{\infty} x^k f(x) dx \\ \sum_n x_n^k p_n \end{cases}.$$

Note that $m_0 = 1$ and $m_1 = \mathbf{E}\{\xi\}$. Actually, one frequently uses the central moments,

$$\mu_k = \mathbf{E}\left\{(\xi - \mathbf{E}\{\xi\})^k\right\} = \begin{cases} \int_{-\infty}^{\infty} (x - \mathbf{E}\{\xi\})^k f(x) dx \\ \sum_n (x_n - \mathbf{E}\{\xi\})^k p_n \end{cases},$$

where, again, we note that $\mu_0 = 1$, $\mu_1 = 0$ and $\mu_2 = \sigma^2$. When we discuss multivariate distributions, it will be convenient to introduce the following notation – We denote the second central moment by the *double* angled brackets, *viz.*

$$\langle\langle XY \rangle\rangle \equiv \langle XY \rangle - \langle X \rangle \langle Y \rangle. \quad (\text{A.7})$$

4. **Correlation Coefficient:** It is sometimes useful to express the degree of correlation among variables by some dimensionless quantity. For that purpose, we introduce the *correlation coefficient*,

$$\rho = \frac{\langle\langle XY \rangle\rangle}{\sigma_X \sigma_Y}. \quad (\text{A.8})$$

Notice that the correlation coefficient is bounded, $-1 \leq \rho \leq 1$. Furthermore, if $\rho < 1$ we say the two processes X and Y are *anti-correlated*, while if $\rho > 0$, we say X and Y are *correlated*. For $\rho = 0$, the two process are obviously *uncorrelated*.

5. **Characteristic function:** The characteristic function is a generating function for the moments of a distribution. There is a unique one-to-one correspondence between the characteristic function $\phi(\omega)$ and a given probability density $f(x)$. The characteristic function is defined as

$$\phi(\omega) = \mathbf{E}\{e^{i\omega\xi}\} = \int_{-\infty}^{\infty} e^{i\omega x} f(x) dx.$$

Notice this is precisely the *Fourier transform* of the probability density $f(x)$. A useful result is the moment theorem,

$$\left. \frac{d^n}{d\omega^n} \phi(\omega) \right|_{\omega=0} = i^n m_n.$$

The characteristic function is also useful as an intermediate step in computing the moments of a nonlinear transformation of a given distribution.

6. **Cumulants:** The characteristic function also generates the *cumulants* κ_n of the random variable,

$$\left. \frac{d^n}{d\omega^n} \ln \phi(\omega) \right|_{\omega=0} = i^n \kappa_n. \quad (\text{A.9})$$

The cumulants are combinations of the lower moments, *e.g.*

$$\begin{aligned} \kappa_1 &= m_1, \\ \kappa_2 &= m_2 - m_1^2 = \sigma^2, \\ \kappa_3 &= m_3 - 3m_2m_1 + 2m_1^3, \dots \end{aligned}$$

For a Gaussian distribution, one can show that all cumulants beyond the second are zero. The utility of cumulants comes from the identity,

$$\langle e^{a\xi} \rangle = \sum_{n=0}^{\infty} \frac{a^n}{n!} m_n = \exp \left\{ \sum_{n=1}^{\infty} \frac{a^n}{n!} \kappa_n \right\}, \quad (\text{A.10})$$

where a is a constant and ξ is a random variable. In that way, cumulants appear in approximations of the average of an exponentiated noise source.

A.2 Examples of Distributions and Densities

1. **Uniform** The uniform random variable has density function,

$$f(x) = \begin{cases} (b-a)^{-1} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}, \quad (\text{A.11})$$

(see Figure A.3A). In the domain $x \in [a, b]$, the distribution function is correspondingly given by,

$$F(x) = \int_a^x (b-a)^{-1} dx' = \frac{x-a}{b-a}, \quad (\text{A.12})$$

(see Figure A.4A), with $F(x) = 0$ for $x < a$, and $F(x) = 1$ for $x > b$. A random variable drawn from this distribution is said to

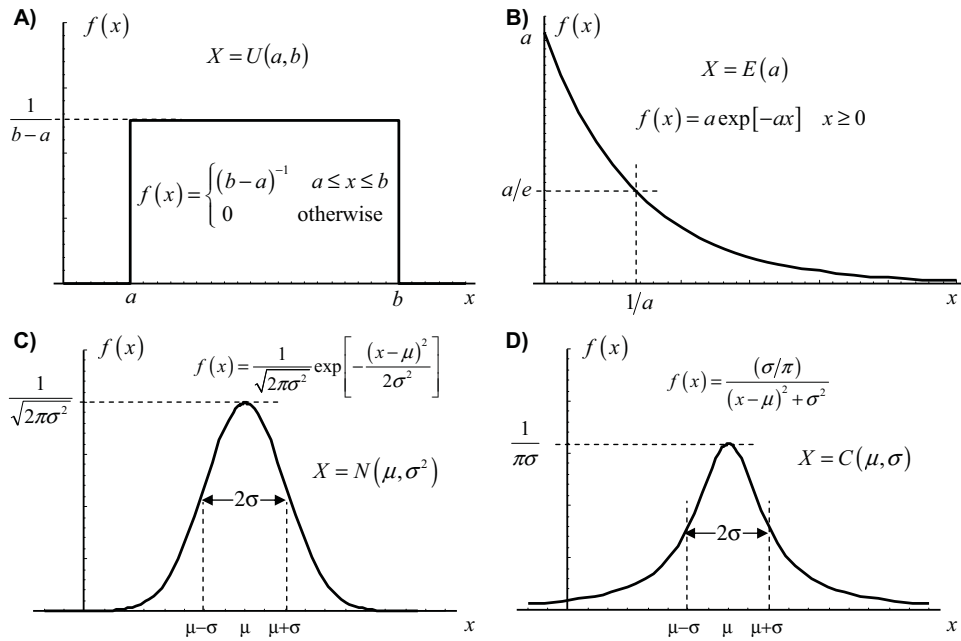


Figure A.3: **Common probability *density* functions.** **A)** Uniform. **B)** Exponential. **C)** Normal (Gaussian). **D)** Cauchy (Lorentz). Redrawn from Gillespie (1992) *Markov Processes*, Academic Press, Inc.

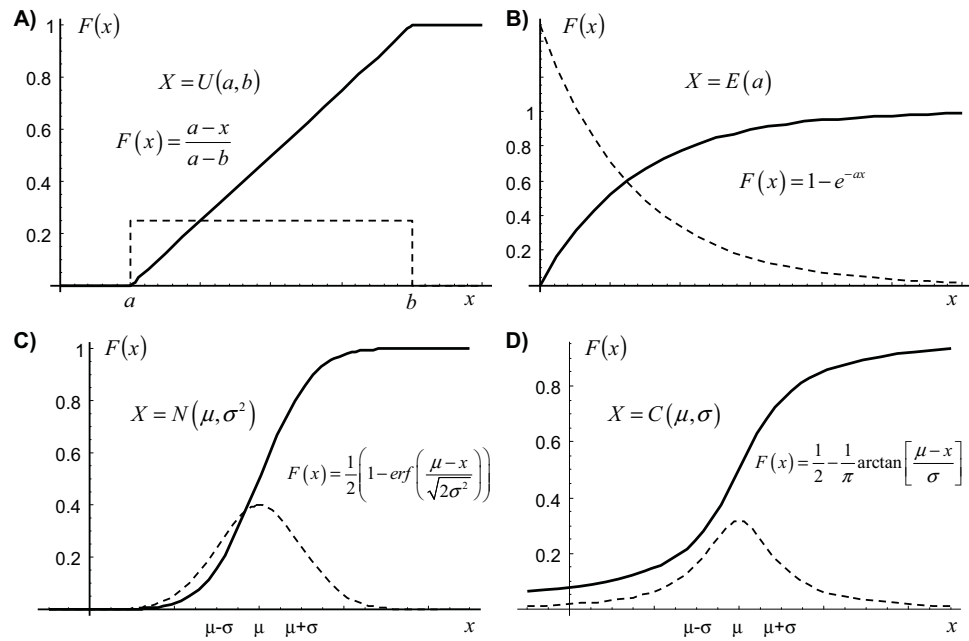


Figure A.4: **Common probability *distribution* functions.** **A)** Uniform. **B)** Exponential. **C)** Normal (Gaussian). **D)** Cauchy (Lorentz). The corresponding density functions are shown as dashed lines.

be “uniformly distributed on $[a, b]$,” written as $X = \mathbf{U}(a, b)$. The mean and standard deviation are,

$$\langle X \rangle = \frac{a+b}{2}; \quad \langle (X - \langle X \rangle)^2 \rangle = \frac{b-a}{2\sqrt{3}}. \quad (\text{A.13})$$

As a consequence,

$$\lim_{b \rightarrow a} \mathbf{U}(a, b) = \text{the sure variable } a. \quad (\text{A.14})$$

2. **Exponential** The exponential random variable has density function,

$$f(x) = a \exp(-ax) \quad x \geq 0, \quad (\text{A.15})$$

(see Figure A.3B). For $x \geq 0$, the distribution function is correspondingly given by,

$$F(x) = \int_0^x a \exp(-ax') dx' = 1 - \exp(-ax), \quad (\text{A.16})$$

(see Figure A.4B), with $F(x) = 0$ for $x < 0$. A random variable drawn from this distribution is said to be “exponentially distributed with decay constant a ,” written as $X = \mathbf{E}(a)$. The mean and standard deviation are,

$$\langle X \rangle = \langle (X - \langle X \rangle)^2 \rangle = \frac{1}{a}. \quad (\text{A.17})$$

As a consequence,

$$\lim_{a \rightarrow \infty} \mathbf{E}(a) = \text{the sure number } 0. \quad (\text{A.18})$$

3. **Normal/Gaussian** The normal random variable has density function,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \quad (\text{A.19})$$

(see Figure A.3C). The distribution function is correspondingly given by,

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x' - \mu)^2}{2\sigma^2} \right] dx' = \frac{1}{2} \left(1 - \operatorname{erf} \left[\frac{\mu - x}{\sqrt{2\pi\sigma^2}} \right] \right), \quad (\text{A.20})$$

where

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt, \quad (\text{A.21})$$

is the *error function* (see Figure A.4C). A random variable drawn from this distribution is said to be “normally (or Gaussian) distributed with mean μ and variance σ^2 ,” written as $X = \mathbf{N}(\mu, \sigma^2)$. The mean and standard deviation are,

$$\langle X \rangle = \mu; \quad \langle (X - \langle X \rangle)^2 \rangle = \sigma^2. \quad (\text{A.22})$$

As a consequence,

$$\lim_{\sigma \rightarrow 0} \mathbf{N}(\mu, \sigma^2) = \text{the sure variable } \mu. \quad (\text{A.23})$$

The multivariate Gaussian distribution is a straightforward generalization of the above – for the n -dimensional vector $\mathbf{x} \in \mathbb{R}^n$, the joint Gaussian distribution is,

$$P(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} (\det \mathbf{C})^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \langle \mathbf{x} \rangle)^T \cdot \mathbf{C}^{-1} \cdot (\mathbf{x} - \langle \mathbf{x} \rangle) \right], \quad (\text{A.24})$$

where $\langle \mathbf{x} \rangle = \mu$ is the vector of averages and $\mathbf{C}$ is called the *cross-correlation matrix* with elements given by $C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$. Obviously, by definition $\mathbf{C}$ is a *symmetric, positive-semidefinite matrix*.

4. **Cauchy/Lorentz** The Cauchy random variable has density function,

$$f(x) = \frac{(\sigma/\pi)}{(x - \mu)^2 + \sigma^2}, \quad (\text{A.25})$$

(see Figure A.3D). The distribution function is correspondingly given by,

$$F(x) = \int_{-\infty}^x \frac{(\sigma/\pi)}{(x' - \mu)^2 + \sigma^2} dx' = \frac{1}{2} - \frac{1}{\pi} \arctan \left[\frac{\mu - x}{\sigma} \right], \quad (\text{A.26})$$

(see Figure A.4D). A random variable drawn from this distribution is said to be “Cauchy distributed about μ with half-width σ ,”

written as $X = \mathbf{C}(\mu, \sigma)$. Although the distribution satisfies the normalization condition, the tails of the distribution vanish so slowly that one can show that *no higher moments of the distribution are defined!* Nevertheless, since the density function is one of many representations of the Dirac-delta function, it follows that,

$$\lim_{\sigma \rightarrow 0} \mathbf{C}(\mu, \sigma) = \text{the sure variable } \mu. \quad (\text{A.27})$$

A.3 Central limit theorem

I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the “Law of Frequency of Error”. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

—Sir Francis Dalton (1889).

In its most restrictive form, the *Central Limit Theorem* states,

Let $X_1, X_2, X_3, \dots, X_n$ be a sequence of n independent and identically distributed random variables having each finite values of expectation μ and variance $\sigma^2 > 0$. As the sample size n increases ($n \rightarrow \infty$), the distribution of the random variable

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}, \quad (\text{A.28})$$

approaches the standard normal distribution $\mathbf{N}(0, 1)$.

This result was significantly generalized by Lyapunov (1901), who showed that it applies to independent random variables *without* identical distribution:

Let $\{X_i\}$ be a sequence of independent random variables defined on the same probability space. Assume that X_i has finite

mean μ_i and finite standard deviation σ_i . We define

$$s_n^2 = \sum_{i=1}^n \sigma_i^2.$$

Assume that the third central moments

$$r_n^3 = \sum_{i=1}^n \langle |X_i - \mu_i|^3 \rangle$$

are finite for every n , and that

$$\lim_{n \rightarrow \infty} \frac{r_n}{s_n} = 0.$$

Under these conditions, writing the sum of random variables $S_n = X_1 + \cdots + X_n$, and the variable defined by

$$Z_n = \frac{S_n - m_n}{s_n},$$

then the distribution of Z_n converges to the standard normal distribution $\mathbf{N}(0, 1)$.

A.4 Generating Random Numbers

Notice the distribution function $F(x)$ by definition always lies between 0 and 1. It is possible, then, to obtain a sample x from *any given distribution* by generating a unit uniform random variable $r \in \mathbf{U}(0, 1)$, and solving for $F(x) = r$, *viz.*,

$$x = F^{-1}(r). \quad (\text{A.29})$$

By spraying the vertical axis with a uniformly distributed random sampling, the distribution function *reflects* the sampling onto the horizontal axis thereby transforming r into the random variable x with the desired statistics. Since $F'(x) = f(x)$, and $f(x) \geq 0$, the function $F(x)$ is strictly increasing and the inversion F^{-1} is uniquely defined for any distribution.

For example, the exponentially distributed random variable $\mathbf{E}(a)$ has the distribution function,

$$F(x) = 1 - \exp(-ax). \quad (\text{A.30})$$

If we set $F(x) = 1 - r$, where r is a unit uniform random variable, then

$$x = (1/a) \ln(1/r), \quad (\text{A.31})$$

is a random variable, exponentially distributed with decay constant a . This relation will be of particular interest in Chapter 4 when we consider numerical simulation of random processes.

A.4.1 Joint inversion generating method

- *Markov Processes*, D. T. Gillespie (Academic Press, Inc, 1992), p. 53.

The following procedure is used to generate joint random variables, and is particularly useful in the generating of Gaussian-distributed random variables (see Exercise 3 on page 193).

Let X_1 , X_2 and X_3 be three random variables with joint density function P . Define the functions F_1 , F_2 and F_3 by

$$\begin{aligned} F_1(x_1) &= \int_{-\infty}^{x_1} P_1(x'_1) dx'_1, \\ F_2(x_2; x_1) &= \int_{-\infty}^{x_2} P_2^{(1)}(x'_2 | x_1) dx'_2, \\ F_3(x_3; x_1, x_2) &= \int_{-\infty}^{x_3} P_3^{(1,2)}(x'_3 | x_1, x_2) dx'_3, \end{aligned}$$

where the subordinate densities $P_k^{(i,j)}$ are defined as follows,

$$\begin{aligned} P_1(x_1) &= \int_{-\infty}^{\infty} P(x_1, x_2, x_3) dx_2 dx_3, \\ P_2^{(1)}(x_2 | x_1) &= \frac{\int_{-\infty}^{\infty} P(x_1, x_2, x_3) dx_3}{\int_{-\infty}^{\infty} P(x_1, x_2, x_3) dx_2 dx_3}, \\ P_3^{(1,2)}(x_3; x_1, x_2) &= \frac{P(x_1, x_2, x_3)}{\int_{-\infty}^{\infty} P(x_1, x_2, x_3) dx_3}. \end{aligned}$$

Then if r_1 , r_2 and r_3 are three independent unit uniform random numbers, the values x_1 , x_2 and x_3 obtained by successively solving the three equations

$$\begin{aligned} F_1(x_1) &= r_1, \\ F_2(x_2; x_1) &= r_2, \\ F_3(x_3; x_1, x_2) &= r_3, \end{aligned}$$

are simultaneous sample values of X_1 , X_2 and X_3 .

A.5 Change of Variables

- D. T. Gillespie (1983) “A theorem for physicists in the theory of random variables,” *American Journal of Physics* **51**: 520.

Given a random variable X with density distribution $P(x)$, and a new random variable Y defined by the transformation,

$$y = g(x), \quad (\text{A.32})$$

(where $g(\cdot)$ is a given deterministic function), the probability density $Q(y)$ of this new variable is given by,

$$Q(y) = \int_{-\infty}^{\infty} P(x) \delta(y - g(x)) dx. \quad (\text{A.33})$$

Here, as always, $\delta(\cdot)$ is the Dirac-delta function. In general, for a vector of random variables $\{X_i\}_{i=1}^n$ with density $P(\mathbf{x})$, and transformed variables $\{y_j = g_j(\mathbf{x})\}_{j=1}^m$, the joint density for $\mathbf{Y}$ is,

$$Q(\mathbf{y}) = \int_{-\infty}^{\infty} P(\mathbf{x}) \prod_{j=1}^m \delta(y_j - g_j(\mathbf{x})) d\mathbf{x}. \quad (\text{A.34})$$

If $m = n$, then making the change of variables $z_j = g_j(\mathbf{x})$ in the integral above, we arrive at the formula for a transformation of probability density functions,

$$Q(\mathbf{y}) = P(\mathbf{x}) \left| \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(y_1, y_2, \dots, y_n)} \right|, \quad (\text{A.35})$$

where $|\partial\mathbf{x}/\partial\mathbf{y}|$ is the determinant of the Jacobian of the transformation $x_i = h_i(\mathbf{y})$.

A.6 Monte-Carlo simulation

- P. J. Nahin *Digital dice*, (Princeton, 2008).

There are situations when exact solutions to probability questions are very difficult to calculate, or perhaps the approach to a solution is not clear. Often, Monte Carlo simulations may be the only solution strategy available, providing a straightforward method to gain some insight into the behaviour of the system.

The main idea is to use a random number generator and a program loop to estimate the averaged outcome of a particular experiment. For example, if I want to know the average number of times I will get 4 heads in a row flipping a fair coin, I could sit down and flip a coin 4 times, recording the outcome each time, and repeat this hundreds or thousands of times, then take the number of times that 4 heads appear in my list, and divide this by the number of 4-flip trials I have done. The larger the number of trials, the better my estimate for the average will be. The data in the table below comes from a random number generator and a *for* loop in place of flipping an actual coin. For comparison, the Matlab script used to generate the data is also shown (Table A.5).

A more sophisticated example comes from trying to estimate the value of an integral using Monte Carlo methods. Suppose we would like to estimate the area of a quarter-circle of radius 1. One approach is to randomly generate points in the 1×1 square in the first quadrant, and count how many of these lie inside the circle (*i.e.* $x^2 + y^2 \leq 1$) divided by the number of total points generated. We then assume that fraction is an adequate representation of the *fraction* of the area of the 1×1 square that lies within the quarter-circle (Figure A.6). Again, a table of data and the routine used to generate that data are shown below (Table A.7). More challenging applications of Monte Carlo methods appear in the exercises.

A.7 Bertrand's Paradox

The classical theory of probability is – from a mathematical point of view – nothing but transformations of variables. Some probability distribution is given *a priori* on a set of elementary events; the problem is then to transform it into a probability distribution for the possible outcomes each of which corresponds to a collection of elementary events. For example, when 2 dice are cast there are 36 elementary events and they are *assumed* to have *equal a priori* probability; the problem is to find the probability

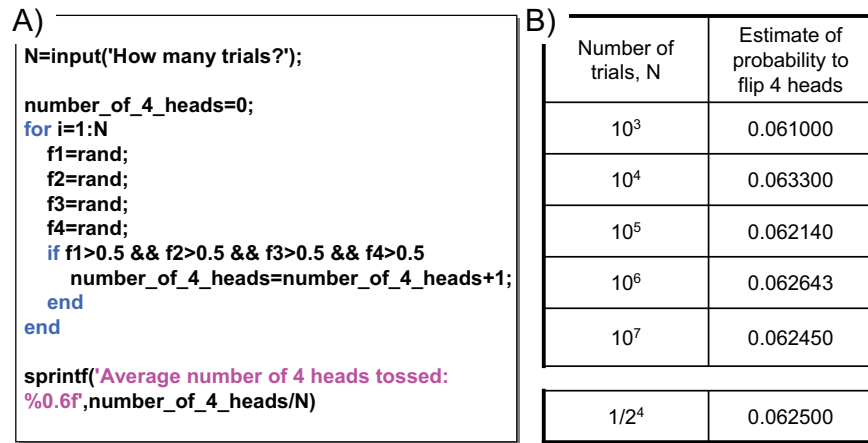


Figure A.5: **A)** Matlab routine to simulate a four-toss experiment, recording the number of times 4 heads are flipped. The only idiosyncratic commands are *rand* which calls a unit uniform random number generator and *sprintf* which displays the output as a formatted string, in particular the `%0.6f` command ensures that the output is displayed with 6 decimal points. **B)** Output simulation data. As the number of trials increases, the Monte Carlo estimate approaches the analytic value.

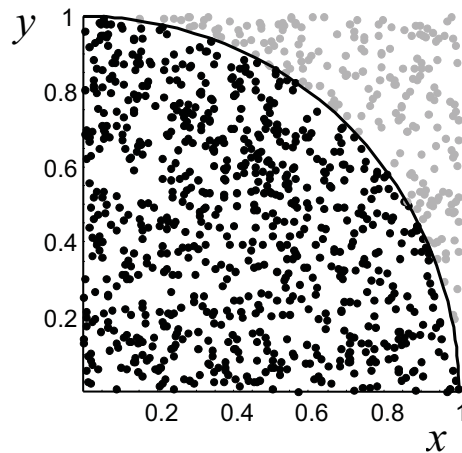


Figure A.6: The fraction of randomly generated points lying inside the quarter circle provide an estimate of the area.

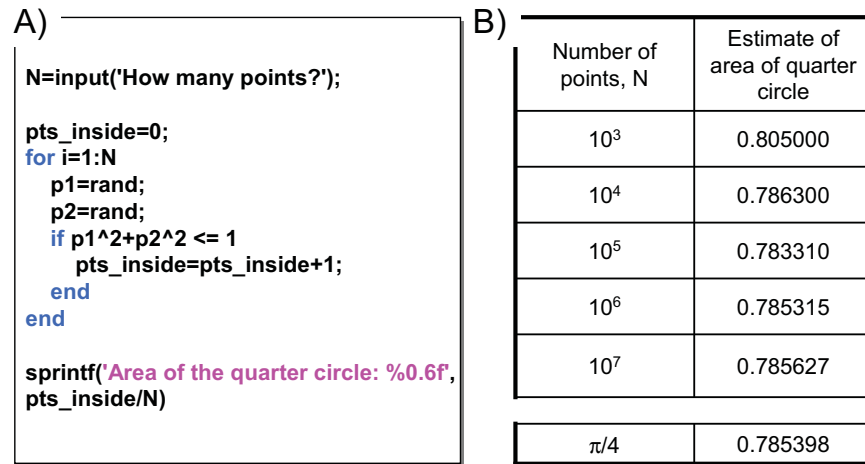


Figure A.7: **A)** Matlab routine to simulate points inside the 1×1 square, recording the number of points laying inside the quarter circle of radius 1. **B)** Output simulation data. As the number of points increases, the Monte Carlo estimate approaches the analytic value.

for the various totals by counting the number of elementary events that make up each total.

Mathematics can only derive probabilities of outcomes from a given *a priori* distribution. In applications to the real world, *one must therefore decide which a priori distribution correctly describes the actual situation.* (This is not a mathematical problem.) In problems of gambling, or balls in urns, the correct choice (or at least the one meant by the author) is usually clear, so is frequently not stated explicitly. This has led to the erroneous view that pure mathematics can conjure up the correct probability for actual events to occur, with an enormous literature of semi-philosophical character. For a review, see for example J. R. Lucas “The concept of probability,” Clarendon, Oxford 1970.

One attempt to avoid an explicit assumption of the *a priori* probability distribution is the so-called *Principle of Insufficient Reason*. It states that two elementary events have the same probability when there is no known reason why they should not. At best, this can be considered a working hypothesis, no philosophical principle can tell whether a die is loaded or not!

The danger of intuitive ideas about equal probabilities has been beautifully illustrated by Bertrand (*Calcul des probabilités*, Gauthiers-Villars, Paris 1889):

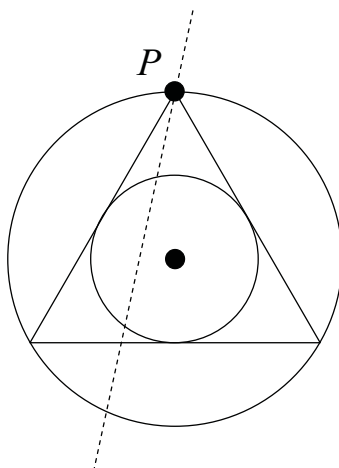


Figure A.8: **A circle of radius 1, with an inscribed equilateral triangle with sides of length $\sqrt{3}$.**

Take a fixed circle of radius 1, and draw *at random* a straight line intersecting it. What is the probability that the chord has length $> \sqrt{3}$ (the length of the side of the inscribed equilateral triangle)?

Answer 1: Take all random lines through a fixed point P on the edge of the circle (Figure A.8). Apart from the tangent (zero probability), all such lines will intersect the circle. For the chord to be $> \sqrt{3}$, the line must lie within an angle of 60° out of a total 180° . Hence, $p = \frac{1}{3}$.

Answer 2: Take all random lines perpendicular to a fixed diameter. The chord is $> \sqrt{3}$ when the point of intersection lies on the middle half of the diameter. Hence, $p = \frac{1}{2}$.

Answer 3: For a chord to have length $> \sqrt{3}$, its center must lie at a distance less than $\frac{1}{2}$ from the center of the circle. The area of a circle of radius $\frac{1}{2}$ is a quarter of that of the original circle. Hence, $p = \frac{1}{4}$.

Each *solution* is based upon a different assumption about equal *a priori* probabilities. The loose phrase *at random* does not sufficiently specify the *a priori* probability to choose among the solutions.

Suggested References

Much of this chapter was taken from,

- *Stochastic Processes in Physics and Chemistry (2nd Ed.)*, N. G. van Kampen (North-Holland, 2001),

and

- *Markov Processes*, D. T. Gillespie (Academic Press, Inc, 1992).

Gillespie's book is particularly useful if one wants to develop practical schemes for simulating random processes. In that direction,

- *An Introduction to Stochastic Processes in Physics*, D. S. Lemons (Johns Hopkins University Press, 2002).

is also very useful.

Exercises

1. **Flipping a biased coin:** Suppose you have a biased coin - showing heads with a probability $p \neq 1/2$. How could you use this coin to generate a random string of binary outcomes, each with a probability of exactly $1/2$?

2. **Characteristic functions:**

- (a) Derive the characteristic function $\phi(\omega)$ for the normal distribution $\mathbf{N}(\mu, \sigma^2)$. Use this characteristic function to show that, for two independent normal random variables,

$$a\mathbf{N}(\mu_1, \sigma_1^2) + b\mathbf{N}(\mu_2, \sigma_2^2) = \mathbf{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2).$$

- (b) Use the characteristic function to show that for an arbitrary random variable the third-order cumulant κ_3 is expressed in terms of the moments μ_1 , μ_2 , and μ_3 , as follows,

$$\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3.$$

- (c) Use the characteristic function for an arbitrary random variable to prove the weak version of the central limit theorem – That is, show that for a series of independent, identically

distributed random variables X_i with mean μ and standard deviation σ , the characteristic function of the random variable

$$Y_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}, \quad (\text{A.36})$$

converges to the characteristic function of the unit normal distribution $\mathbf{N}(0, 1)$ as $n \rightarrow \infty$. *Hint:* Let

$$Z_i = \frac{X_i - \mu}{\sigma},$$

and write out the Taylor series for the characteristic function of Z_i . Use this to derive an expression for the characteristic function of Y_n in the limit $n \rightarrow \infty$.

- (d) Write a Monte Carlo simulation to verify the central limit theorem. Start with any kind of nonlinear transformation of a unit uniform random number, and plot a histogram of the distribution of the sum Eq. A.28. What do you notice for small n ? Compare the histogram to a unit normal distribution – beyond what values of n are the plots indistinguishable?
3. **Multivariate Gaussian distribution:** Show that if the correlation matrix $\mathbf{C}$ is diagonal ($C_{ij} = 0$; $i \neq j$), then $\mathbf{x}$ is a vector of *independent* Gaussian random variables.
4. **Generating random numbers:** Very powerful and reliable schemes have been developed to generate unit uniform random numbers. But we often want random numbers drawn from more exotic distributions.
 - (a) Write the generation formulas for x drawn from $\mathbf{U}(a, b)$ and $\mathbf{C}(\mu, \sigma)$ in terms of the unit uniform variable r . Do you foresee a problem trying to do the same for x drawn from $\mathbf{N}(\mu, \sigma^2)$?
 - (b) Given the discrete probability distribution $P(n)$, show that for r drawn from a unit uniform distribution $\mathbf{U}(0, 1)$, the integer n satisfying,

$$\sum_{n'=-\infty}^{n-1} P(n') \leq r < \sum_{n'=-\infty}^n P(n'),$$

is a realization drawn from $P(n)$. *Hint:* What is the probability that $a \leq r < b$, for $0 \leq a < b < 1$.

5. **Monte Carlo simulation:** Often analytic solutions to probability questions are prohibitively difficult. In those cases, or simply to gain some intuition for the process, stochastic simulation is indispensable. Estimate the solution of the following problems by running a Monte Carlo simulation.

(a) Suppose you have to match members of two lists – For example, a list of architects and their famous buildings. Each member of one list maps uniquely to a member of the other list. For a list of 15 members, what is the average number of correct connections that will be made if each member of one list is randomly connected to another member of the other list (in a one-to-one fashion)? How does this average change for a list with 5 members? With 100 members?

(b) Suppose you have a dozen eggs. You take them out of the carton and wash them. When you put them back, what is the probability that none of the eggs end up in their original location? What if you have 100 eggs? Do you recognize the reciprocal of this probability?

The probability that after permutation no member returns to where it began is called the *derangement* probability, and was computed explicitly for an arbitrary number of elements by Euler (1751) using a very crafty argument.

(c) Two (independent) bus lines operate from a stop in front of your apartment building. One bus arrives every hour, on the hour, the other an unknown, but constant x fraction of an hour later (where, for lack of additional information, we assume x is a unit uniform random variable). What is the average wait time for a rider arriving at the stop at random? What about the case for n independent lines running from the same stop? By running simulations for different n , can you speculate about a general formula for all n ?

6. **Change of variables:** To a large extent, classical probability theory is simply a change of variables. The following test that notion.

(a) **Ballistics:** (van Kampen, 2001) A cannon shoots a ball with initial velocity ν at an angle θ with the horizontal. Both ν and θ have uncertainty given by Gaussian distributions with variance σ_ν^2 and σ_θ^2 , centered on ν_0 and θ_0 , respectively. The variance is narrow enough that negative values of ν and θ can

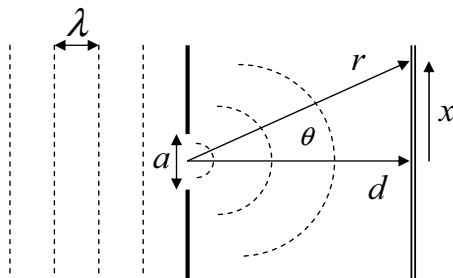


Figure A.9: **Single-slit diffraction.** From Lemons (2002), p. 30.

be ignored. Find the probability distribution for the distance traveled by the cannon ball.

- (b) **Single-slit diffraction:** (Lemons, 2002) According to the probability interpretation of light, formulated by Max Born in 1926, light intensity at a point is proportional to the probability that a photon exists at that point.
- i. What is the probability density $f(x)$ that a single photon passes through a narrow slit and arrives at position x on a screen parallel to and at a distance d beyond the barrier? Each angle of forward propagation θ is the uniform random variable $\mathbf{U}(0, \pi/2)$ (see Figure A.9).
 - ii. The light intensity produced by diffraction through a single, narrow slit, as found in almost any introductory physics text, is proportional to

$$I \propto \frac{1}{r^2} \frac{\sin^2 [(\pi a/\lambda) \sin \theta]}{\sin^2 \theta},$$

where r is the distance from the center of the slit to an arbitrary place on the screen, a is the slit width, and λ the light wavelength. Show that for slits so narrow that $\pi a/\lambda \ll 1$, the above light intensity is proportional to the photon probability density derived in part 6(b)i.

- (c) **Linear noise approximation:** If we define the random variables $\mathbf{n}(t)$ by the linear (though time-dependent) transformation,

$$n_i(t) = \Omega x_i(t) + \Omega^{1/2} \alpha_i(t), \quad i = 1, 2, \dots, N;$$

where $\boldsymbol{\alpha}(t)$ is a random variable with density $\Pi(\boldsymbol{\alpha}, t)$, Ω is a constant, and $\mathbf{x}(t)$ is a deterministic function of time, then show that the density $P(\mathbf{n}, t)$ is simply,

$$P(\mathbf{n}, t) = \Omega^{N/2} \Pi(\boldsymbol{\alpha}, t).$$

APPENDIX B

REVIEW OF MATHEMATICAL METHODS

There are several mathematical methods we return to again and again in this course. In the main notes, it is assumed that the reader is familiar with matrix algebra, linear stability analysis Laplace/Fourier transforms, and some elementary results from asymptotic analysis. As a refresher, more background is provided in this appendix.

B.1 Matrix Algebra and Time-Ordered Exponentials

The main application of matrix algebra in this course will to codify a system of differential equations in a compact way. As such, most of the methods outlined in this appendix will be written in vector-matrix notation.

Matrix form of differential equations

A system of high-order differential equations can always be written as a larger system of first-order differential equations by the simple expedient of introducing new state variables for each of the derivatives. For example, the equation describing simple harmonic motion can be written either as a second-order differential equation or as a 2×2 -system of first-order

differential equations—

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \Leftrightarrow \frac{d}{dt} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix}, \quad (\text{B.1})$$

where the first line of the matrix equation defines the new state variable $\dot{x}$. In many ways, the matrix form of the equation is preferable since solution methods immediately generalize to higher-order systems. For a fully nonlinear system, where $\mathbf{x} \in \mathbb{R}^n$,

$$\frac{d}{dt}\mathbf{x} = \mathbf{F}(\mathbf{x}), \quad (\text{B.2})$$

where $\mathbf{F} = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_n(\mathbf{x}))$ is the reaction rate vector. There is no general solution for such systems. If, however, the reaction rate vector is *linear* in the state variables, there are formal solutions.

Autonomous linear differential equations

Autonomous means that the coefficients appearing in the differential equation are constant. For a homogeneous (*i.e.* zero right-hand side), the system reads,

$$\frac{d\mathbf{x}}{dt} - \mathbf{A} \cdot \mathbf{x} = 0; \quad \mathbf{x}(0) = \mathbf{x}_0. \quad (\text{B.3})$$

In direct analogy with the one-dimensional system, the fundamental solutions of this equation are simply exponentials, albeit with matrix arguments. We define the matrix exponential as,

$$\exp \mathbf{A} = \mathbf{I} + \mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \dots = \sum_{n=0}^{\infty} \frac{\mathbf{A}^n}{n!}. \quad (\text{B.4})$$

Here, $\mathbf{I}$ is the identity matrix,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

With the matrix exponential, the formal solution of Eq. B.3 is simply,

$$\mathbf{x}(t) = \mathbf{x}_0 \cdot \exp [\mathbf{A}t]. \quad (\text{B.5})$$

Time-order exponentials for non-autonomous differential equations

- R. Feynman (1951) “An operator calculus having applications in quantum electrodynamics,” *Physical Review* **84**: 108 – 128.
- N. G. van Kampen (1974) “A cumulant expansion for stochastic linear differential equations. 1,” *Physica* **74**: 215–238.

Non-autonomous means the coefficients in the differential equation are no longer constant. Consider, for example, the differential equation with *time-dependent* coefficients¹,

$$\frac{d}{dt}\mathbf{Y}(t) = \mathbf{A}(t) \cdot \mathbf{Y}(t); \quad \mathbf{Y}(0) = \mathbf{1}.$$

By iteration, the solution can be expressed as an infinite series,

$$\mathbf{Y}(t) = \mathbf{I} + \int_0^t \mathbf{A}(t_1) dt_1 + \int_0^t \int_0^{t_1} \mathbf{A}(t_1) \mathbf{A}(t_2) dt_2 dt_1 + \dots \quad (\text{B.6})$$

$$= \sum_{n=0}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \mathbf{A}(t_1) \mathbf{A}(t_2) \cdots \mathbf{A}(t_n). \quad (\text{B.7})$$

If the matrices $\mathbf{A}(t_n)$ commute with one another, this series can be simply written as an exponential of an integral (as in the autonomous case discussed above),

$$\begin{aligned} \mathbf{Y}(t) &= \exp \left[\int_0^t \mathbf{A}(t') dt' \right] = \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \mathbf{A}(t_1) \mathbf{A}(t_2) \cdots \mathbf{A}(t_n), \end{aligned}$$

where the $1/n!$ is included to accommodate the extended domain of integration from $[0, t]$ for each iterate. If the matrices $\mathbf{A}(t_n)$ *do not* commute with one another (as is typically the case), then a *time-ordering operator* $[\]$ is used to allow the infinite series, Eq. B.7, to be written as an exponential,

$$\begin{aligned} \mathbf{Y}(t) &= \sum_{n=0}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \mathbf{A}(t_1) \mathbf{A}(t_2) \cdots \mathbf{A}(t_n) \\ &\equiv \left[\exp \left[\int_0^t \mathbf{A}(t') dt' \right] \right]. \end{aligned} \quad (\text{B.8})$$

¹Equations of this sort *cannot* be readily solved using Laplace or Fourier transform methods.

Many equations, for example the Mathieu equation discussed in Section 8.4, can be written formally as a time-ordered exponential, although the explicit solution cannot be written in terms of elementary functions.

B.2 Linear Stability Analysis

- L. Edelstein-Keshet *Mathematical models in biology*, (SIAM, 2005). Chapter 4.

Suppose we have the nonlinear system of differential equations,

$$\frac{d}{dt}\mathbf{x} = \mathbf{F}(\mathbf{x}). \quad (\text{B.9})$$

We can determine the equilibrium, or fixed, points of this system by looking for values of the state variables such that,

$$\frac{d\mathbf{x}}{dt} = 0,$$

or, equivalently, those values of $\mathbf{x}$ that satisfy the condition,

$$\mathbf{F}(\mathbf{x}^*) = \mathbf{0}. \quad (\text{B.10})$$

Once these equilibrium points have been determined, either analytically or numerically, the question arises as to whether or not they are stable in the sense that small perturbations away from $\mathbf{x}^*$ return to $\mathbf{x}^*$.

Eigenvalues and the fundamental modes

For very small perturbations $\mathbf{x}_p$, the local dynamics near to $\mathbf{x}^*$ can be adequately described by a *linearization* of the rates. Starting from the governing equation,

$$\frac{d}{dt} [\mathbf{x}^* + \mathbf{x}_p] = \frac{d}{dt} \mathbf{x}_p = \mathbf{F}(\mathbf{x}^* + \mathbf{x}_p),$$

we use the Taylor series to approximate $\mathbf{F}$ for small $\mathbf{x}_p$,

$$\mathbf{F}(\mathbf{x}^* + \mathbf{x}_p) \approx \mathbf{F}(\mathbf{x}^*) + \mathbf{J} \cdot \mathbf{x}_p = \mathbf{J} \cdot \mathbf{x}_p,$$

where $\mathbf{J}$ is called the *Jacobian* or response matrix of $\mathbf{F}$,

$$J_{ij} = \frac{\partial F_i(\mathbf{x}^*)}{\partial x_j}.$$

The dynamics of the small perturbation modes $\mathbf{x}_p$ are then governed by the homogeneous linear equation,

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{J} \cdot \mathbf{x}_p. \quad (\text{B.11})$$

For a one-dimensional system, we know that the solution is an exponential $x_p \propto \exp \lambda t$. Assuming the same holds true in multiple dimensions, we make the ansatz,

$$\mathbf{x}_p(t) = \mathbf{v} e^{\lambda t}. \quad (\text{B.12})$$

Substituting into both sides of Eq. B.11, and cancelling $e^{\lambda t}$, we arrive at the constraint on λ ,

$$\lambda \mathbf{v} = \mathbf{J} \cdot \mathbf{v}. \quad (\text{B.13})$$

Or, writing $\lambda \mathbf{v} = \lambda \mathbf{I} \cdot \mathbf{v}$,

$$[\lambda \mathbf{I} - \mathbf{J}] \cdot \mathbf{v} = 0. \quad (\text{B.14})$$

Non-trivial solutions, $\mathbf{v} \neq \mathbf{0}$, are only possible if the matrix on the left-hand side is not invertible, *i.e.*,

$$\det [\lambda \mathbf{I} - \mathbf{J}] = 0. \quad (\text{B.15})$$

This is called the resolvent equation for the linear operator $\mathbf{J}$, and values λ that satisfy this condition are called the *eigenvalues* of $\mathbf{J}$. The importance of the eigenvalues comes from what they tell us about the long-time stability of the perturbation modes $\mathbf{x}_p$. Since we have $\mathbf{x}_p \propto e^{\lambda t}$, if any of the eigenvalues have a positive real part, the perturbations will grow and we say that the equilibrium point $\mathbf{x}^*$ is *unstable*.

B.2.1 Routh-Hurwitz Criterion

- L. Edelstein-Keshet *Mathematical models in biology*, (SIAM, 2005). pp. 233–234.

Notice that the resolvent equation, Eq. B.15, is a polynomial in λ with degree equal to the dimensionality of our system – That is, if $\mathbf{J} \in \mathbb{R}^n \times \mathbb{R}^n$, then the resolvent equation is an n^{th} -order polynomial. It would be nice, (particularly when the order is greater than 4 or some of the coefficients are free parameters), if we had a condition for stability that did not require that the λ be computed explicitly. The Routh-Hurwitz criterion

provides just such a test, determining the stability from the coefficients of the resolvent equation without explicit computation of λ .

Suppose our resolvent equation is the k^{th} -order polynomial,

$$\lambda^k + a_1\lambda^{k-1} + \dots + a_k = 0, \quad (\text{B.16})$$

then define the following matrices,

$$\begin{aligned} \mathbf{H}_1 &= [a_1], \quad \mathbf{H}_2 = \begin{bmatrix} a_1 & 1 \\ a_3 & a_2 \end{bmatrix}, \dots, \\ \mathbf{H}_k &= \begin{bmatrix} a_1 & 1 & 0 & \dots & 0 \\ a_3 & a_2 & a_1 & 1 & 0 & \dots & \vdots \\ \vdots & & & & & & 0 \\ 0 & 0 & \dots & & & & a_k \end{bmatrix} \end{aligned} \quad (\text{B.17})$$

where the (l, m) element of the matrix $\mathbf{H}_j$ is

$$\begin{aligned} a_{2l-m} & \text{ for } 0 < 2l - m < k, \\ 1 & \text{ for } 2l = m, \\ 0 & \text{ for } \text{otherwise.} \end{aligned} \quad (\text{B.18})$$

It looks cumbersome written out for a general system, but it is easily coded into symbolic mathematics packages like Maple, Mathematica or Matlab. The stability criterion is the following,

A necessary and sufficient condition for all of the roots of the resolvent Eq. B.16 to have negative real parts is that the determinants of the matrices $\mathbf{H}_k$ are all positive (> 0).

For characteristics polynomials of low-degree, the Routh-Hurwitz criteria for stability can be stated simply and explicitly.

For a characteristic equation of degree n ,

$$\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n = 0,$$

the eigenvalues λ all have negative real parts if, and only if,

For $n = 2$: $a_1 > 0, a_2 > 0$.

For $n = 3$: $a_1 > 0, a_3 > 0; a_1a_2 > a_3$.

For $n = 4$: $a_1 > 0, a_3 > 0, a_4 > 0; a_1a_2a_3 > a_3^2 + a_1^2a_4$.

B.3 Transforms Methods

- G. James, *Advanced Modern Engineering Mathematics*, (Prentice Hall, 2005).

Laplace and Fourier transforms are a convenient method of exchanging the search for the solution of a differential equation (difficult) with the search for the solution of an algebraic equation (easier). The simplification of solving the problem in Laplace (or Fourier) space comes at the expense of a sometimes difficult (or impossible) inversion back to the original problem space.

In the context of the present course, we will rarely be interested in the inverting of the Laplace or Fourier transform back to the original problem space since we will be using the properties of the transformed quantities directly - the poles of the Laplace transform will be used to determine the asymptotic stability (Section 8.5 on page 188 and Section 9.3 on page 207), and the Fourier transform of the correlation function is related to the power spectrum of the fluctuations (Section 2.3 on page 39).

B.3.1 Laplace Transform

The Laplace transform of the vector function $\mathbf{f}(t) \in \mathbb{R}^n$ is given by the integration of the function multiplied by $\exp[s\mathbf{I}t]$,

$$\mathcal{L}[\mathbf{f}(t)](s) \equiv \hat{\mathbf{F}}(s) = \int_0^{\infty} \mathbf{f}(t) \cdot e^{s\mathbf{I}t} dt, \quad (\text{B.19})$$

where $\mathbf{I}$ is the $n \times n$ identity matrix (for the transform of a scalar function, set $n = 1$). Our interest in the Laplace transform comes from two useful properties. First, using the definition of the transform and integration by parts, it is possible to derive the relation:

$$\mathcal{L}\left[\frac{d\mathbf{x}}{dt}\right] = s\mathbf{I} \cdot \hat{\mathbf{X}}(s) - \mathbf{x}(0). \quad (\text{B.20})$$

Then, for a system of linear differential equations,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A} \cdot \mathbf{x}; \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (\text{B.21})$$

the formal solution can always be expressed as a Laplace transform,

$$\hat{\mathbf{X}}(s) = [s\mathbf{I} - \mathbf{A}]^{-1} \cdot \mathbf{x}_0. \quad (\text{B.22})$$

The second useful property is that the Laplace transform of a convolution integral,

$$\mathbf{g}(t) * \mathbf{f}(t) = \int_0^t \mathbf{g}(t - \tau) \cdot \mathbf{f}(\tau) d\tau \quad (\text{B.23})$$

is simply the *product* of the individual Laplace transforms,

$$\mathcal{L}[\mathbf{g}(t) * \mathbf{f}(t)] = \mathcal{L}[\mathbf{g}(t)] \cdot \mathcal{L}[\mathbf{f}(t)] = \hat{\mathbf{G}}(s) \cdot \hat{\mathbf{F}}(s). \quad (\text{B.24})$$

In that way, the formal solution of the linear convolution equation,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A} \cdot \mathbf{x} + \int_0^t \mathbf{g}(t - \tau) \cdot \mathbf{x}(\tau) d\tau; \quad \mathbf{x}(0) = \mathbf{x}_0, \quad (\text{B.25})$$

is,

$$\hat{\mathbf{X}}(s) = [s\mathbf{I} - \mathbf{A} - \hat{\mathbf{G}}(s)]^{-1} \cdot \mathbf{x}_0. \quad (\text{B.26})$$

Asymptotic Stability

- S. I. Grossman and R. K. Miller (1973) “Nonlinear Volterra integrodifferential systems with L^1 -kernels,” *Journal of Differential Equations* **13**: 551–566.

Notice from Eq. B.22 that $[s\mathbf{I} - \mathbf{A}]$ is not invertible precisely at those values of s for which,

$$\det[s\mathbf{I} - \mathbf{A}] = 0. \quad (\text{B.27})$$

This equation, called the *resolvent equation*, likewise determines the eigenvalues of the matrix $\mathbf{A}$ (Section B.2). Therefore, the solution of the differential equation Eq. B.29, is asymptotically stable ($\mathbf{x}(t) \rightarrow \mathbf{0}$ as $t \rightarrow \infty$) if, and only if, those values of s that solve the resolvent equation all have negative real parts:

$$\det[s\mathbf{I} - \mathbf{A}] = 0 \quad \text{Re}(s) < 0. \quad (\text{B.28})$$

A more general theorem (Grossman and Miller, 1973) states that asymptotic stability of the convolution equation,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A} \cdot \mathbf{x} + \int_0^t \mathbf{g}(t - \tau) \cdot \mathbf{x}(\tau) d\tau, \quad (\text{B.29})$$

is also guaranteed, provided the roots of the resolvent equation have negative real parts,

$$\det[s\mathbf{I} - \mathbf{A} - \hat{\mathbf{G}}(s)] = 0 \quad \text{Re}(s) < 0. \quad (\text{B.30})$$

Finally, a useful feature of the Laplace transform is that the $s \rightarrow 0$ limit provides the $t \rightarrow \infty$ limit of the un-transformed function $\mathbf{x}(t)$,

$$\lim_{t \rightarrow \infty} \mathbf{x}(t) = \lim_{s \rightarrow 0} s \cdot \hat{\mathbf{x}}(s). \quad (\text{B.31})$$

B.3.2 Fourier Transform

The Fourier transform is very similar to the Laplace transform, but now with an explicitly complex argument $i\omega$ and an unbounded domain of integration,

$$\mathcal{F}[\mathbf{f}(t)](i\omega) \equiv \hat{\mathbf{F}}(i\omega) = \int_{-\infty}^{\infty} \mathbf{f}(t) \cdot e^{i\omega \mathbf{I}t} dt. \quad (\text{B.32})$$

The Fourier transform, in contrast to the Laplace or z -transform, is easily inverted. That is,

$$\mathbf{f}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mathbf{F}}(i\omega) \cdot e^{i\omega \mathbf{I}t} d\omega.$$

Up to a factor of 2π , the Fourier transform is its own inverse. Furthermore, note the duality property of the Fourier transform,

$$\mathcal{F}[\hat{\mathbf{F}}(it)] = 2\pi \mathbf{f}(-\omega). \quad (\text{B.33})$$

The Fourier transform can be thought of as a projecting of the frequency content of a signal $\mathbf{f}(t)$ onto the basis functions $\cos \omega t$ and $\sin \omega t$, for a continuous distribution of frequencies. To illustrate this property, we use the Dirac delta function to derive generalized Fourier transforms.

Dirac delta function

The Dirac delta function $\delta(t)$ is a distribution with the following sifting property,

$$\int_a^b f(t) \delta(t - c) dt = f(c) \quad \text{for } a < c < b.$$

What happens for $c = a$ or $c = b$ is a matter of convention – If $\delta(t)$ is defined as the limit of a sequence of functions, for example,

$$\lim_{n \rightarrow \infty} \frac{n}{2} e^{-|\tau|^n} = \delta(\tau),$$

then it follows that,

$$\int_a^b f(t) \delta(t - c) dt = \frac{1}{2} f(c) \quad c = a, b.$$

One can show that this convention is consistent with Stratonovich's interpretation of the nonlinear Langevin equation. Itô's interpretation, on the other hand, is consistent with the convention,

$$\int_a^b f(t) \delta(t - c) dt = \begin{cases} f(c) & c = a \\ 0 & c = b. \end{cases}$$

For the purposes of deriving generalized Fourier transforms, these distinctions are immaterial, as we will be considering an unbounded domain of integration.

Generalized Fourier transforms

By the sifting property,

$$\mathcal{F}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t) e^{-i\omega t} dt = 1, \quad (\text{B.34})$$

and

$$\mathcal{F}[\delta(t - t_0)] = \int_{-\infty}^{\infty} \delta(t - t_0) e^{-i\omega t} dt = e^{-i\omega t_0}.$$

In and of themselves, these transform pairs are unremarkable. By using the duality property (Eq. B.33), however, we arrive at the decidedly more exotic pairs,

$$\begin{aligned} 1 &\Longleftrightarrow 2\pi\delta(\omega), \\ e^{i\omega_0 t} &\Longleftrightarrow 2\pi\delta(\omega - \omega_0). \end{aligned}$$

The trigonometric functions $\cos \omega t$ and $\sin \omega t$ can be written as a sum of complex exponentials, leading to the following Fourier transforms,

$$\begin{aligned} \mathcal{F}[\cos \omega_0 t] &= \pi [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)], \\ \mathcal{F}[\sin \omega_0 t] &= i\pi [\delta(\omega + \omega_0) - \delta(\omega - \omega_0)]. \end{aligned}$$

The Fourier transform can therefore be thought of as a portrait of the *frequency content of the signal* $f(t)$. A peak at some frequency ω_0 indicates a dominant oscillatory component in the signal. Conversely, a flat spectrum indicates very little structure in the signal (as is the case for white noise).

B.3.3 z -Transform

The z -transform is essentially a discrete form of the Laplace transform. Formally, given a sequence $\{x_k\}_{-\infty}^{\infty}$ of complex numbers x_k , the z -transform of the sequence is defined as

$$\mathcal{Z} \{x_k\}_{-\infty}^{\infty} = X(z) = \sum_{k=-\infty}^{\infty} \frac{x_k}{z^k},$$

whenever the sum exists.

A sequence is called *causal* if $x_k = 0$ for $k < 0$. For a causal sequence then, the z -transform reduces to,

$$\mathcal{Z} \{x_k\}_{-\infty}^{\infty} = X(z) = \sum_{k=0}^{\infty} \frac{x_k}{z^k}.$$

The z -transform is useful to solve linear difference equations, such as one finds in trying to solve for the equilibrium probability distribution of a random walk on a discrete lattice. It also appears in the guise of moment generating functions again used to solve linear difference equations.

B.4 Partial differential equations

Partial differential equations are equations characterizing the behaviour of a multivariable function in terms of partial derivatives. They are notoriously difficult to solve, particularly if the equation is nonlinear in the function of interest. For the purposes of this course, we will only consider *linear* partial differential equations that will be susceptible to solution through a variety of simple methods. For details, consult a text devoted to the solution of partial differential equations, for example the reasonably-priced

- S. J. Farlow, *Partial Differential Equations for Scientists and Engineers*, (Dover, 1993).

B.4.1 Method of characteristics

Notice that the chain-rule applied to the function $u(x(s), t(s))$ of parameterized curves $(x(s), t(s))$,

$$\frac{du}{ds} = \frac{\partial u}{\partial x} \frac{dx}{ds} + \frac{\partial u}{\partial t} \frac{dt}{ds},$$

implies that the first-order partial differential equation

$$a(x, t, u) \frac{\partial u}{\partial x} + b(x, t, u) \frac{\partial u}{\partial t} = c(x, t, u),$$

can be written as a system of *ordinary* differential equations,

$$\begin{aligned} \frac{dx}{ds} &= a(x(s), t(s), u(s)), \\ \frac{dt}{ds} &= b(x(s), t(s), u(s)), \\ \frac{du}{ds} &= c(x(s), t(s), u(s)). \end{aligned}$$

The method of characteristics is only applicable to first-order partial differential equations, although these occur in application when the moment generating function is used to transform a master equation with *linear* transition rates (see Section 4.1.1).

B.4.2 Separation of variables

Partial differential equations with particular underlying symmetry governing the function $u(x, t)$ can be transformed to a coupled system of ordinary differential equations by assuming a solution of the form $u(x, t) = X(x)T(t)$. For example, the diffusion equation,

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2},$$

is separable. Substitution of a solution of the form $u(x, t) = X(x)T(t)$ yields two ordinary differential equations for $X(x)$ and $T(t)$. It was in this context that Fourier derived his famous series.

B.4.3 Transform methods

The transform methods, Laplace and Fourier, because they convert derivatives into algebraic expressions involving the transformed functions, are particularly powerful methods for solving partial differential equations. The limitations are that they only work if the coefficients are constant in the dependent variable and inversion back to the original problem space can be difficult (or impossible). Typically, Laplace transforms are used to transform variables with domain defined on the positive half-line - for

example, Laplace transform of $u(x, t)$ with $t > 0$ yields an ordinary differential equation for $\hat{U}(x, s)$. Fourier transforms, on the other hand, because of the unbounded support of the transform integral, are used when the function of interest has an unbounded domain - for example, Fourier transform of $u(x, t)$ with $-\infty < x < \infty$ yields an ordinary differential equation for $\hat{U}(\omega, t)$.

B.5 Some Results from Asymptotics and Functional Analysis

Occasionally, use is made of some important results from functional analysis and asymptotic analysis. We will use only basic results that can be proved without much effort.

B.5.1 Cauchy-Schwarz Inequality

If $f(t)$ and $g(t)$ are *any* real-valued functions, then it is obviously true that for some real constant $\lambda \in \mathbb{R}$,

$$\int_L^U \{f(t) + \lambda g(t)\}^2 dt \geq 0,$$

since the integrand is nowhere negative. Expanding this expression,

$$\lambda^2 \int_L^U g^2(t) dt + 2\lambda \int_L^U f(t)g(t) dt + \int_L^U f^2(t) dt \geq 0. \quad (\text{B.35})$$

For fixed limits of integration U and L (possibly infinite), the value of each integral is simply a constant, call them a , b , and c . Eq. B.35 reduces to,

$$h(\lambda) = a\lambda^2 + 2b\lambda + c \geq 0, \quad (\text{B.36})$$

where,

$$\begin{aligned} a &= \int_L^U g^2(t) dt, \\ b &= \int_L^U f(t)g(t) dt, \\ c &= \int_L^U f^2(t) dt. \end{aligned}$$

The condition $h(\lambda) \geq 0$ means a plot of $h(\lambda)$ must lie above the λ -axis and cannot cross it. At most, it may *touch* the axis, in which case we have the double root $\lambda = -b/a$. For $h(\lambda)$ above the λ -axis, we must have that the roots of $h(\lambda)$ are imaginary. From the quadratic formula, we then have,

$$h(\lambda) \geq 0 \Leftrightarrow b^2 \leq ac. \quad (\text{B.37})$$

In terms of our original integrals,

$$\left\{ \int_L^U f(t)g(t)dt \right\}^2 \leq \left\{ \int_L^U f^2(t)dt \right\} \left\{ \int_L^U g^2(t)dt \right\}. \quad (\text{B.38})$$

This is the Cauchy-Schwarz inequality.

B.5.2 Watson's Lemma

For $0 < b < \infty$, with,

$$f(t) \sim f_0(t - t_0)^\alpha + f_1(t - t_0)^\beta, \quad -1 < \alpha < \beta, \quad (\text{B.39})$$

as $t \rightarrow t_0$, the estimate for the following integral holds,

$$\int_0^b f(t)e^{-xt}dt \sim f_0 \frac{\Gamma(1+\alpha)}{x^{1+\alpha}} + f_1 \frac{\Gamma(1+\beta)}{x^{1+\beta}}, \quad \text{as } x \rightarrow \infty. \quad (\text{B.40})$$

More generally, for $-\infty < a < b < \infty$, and,

$$g(t) \sim g_0 + g_1(t - t_0)^\lambda, \quad \lambda > 0, g_1 \neq 0, \quad (\text{B.41})$$

the estimate for the following integral holds,

$$\int_a^b f(t)e^{-xg(t)}dt \sim \frac{2f_0}{\lambda} \Gamma\left(\frac{1+\alpha}{\lambda}\right) \left(\frac{1}{xg_1}\right)^{\frac{1+\alpha}{\lambda}} e^{-xg_0}, \quad \text{as } x \rightarrow \infty. \quad (\text{B.42})$$

where the minimum of $g(t)$ occurs at t_0 ($a < t_0 < b$). If the minimum t_0 should fall on either of the end points a or b , then the factor of 2 no longer appears; *i.e.* for $t_0 = a$ or $t_0 = b$,

$$\int_a^b f(t)e^{-xg(t)}dt \sim \frac{f_0}{\lambda} \Gamma\left(\frac{1+\alpha}{\lambda}\right) \left(\frac{1}{xg_1}\right)^{\frac{1+\alpha}{\lambda}} e^{-xg_0}, \quad \text{as } x \rightarrow \infty. \quad (\text{B.43})$$

B.5.3 Stirling's Approximation

Stirling's approximation is:

$$\boxed{n! \approx \sqrt{2\pi n} \cdot n^n e^{-n} \quad \text{for } n \gg 1} \quad (\text{B.44})$$

which is incredibly accurate. To justify the expression, we write the factorial in its equivalent form as an improper integral,

$$n! = \int_0^{\infty} x^n e^{-x} dx. \quad (\text{B.45})$$

Note that the integrand $F(x) = x^n e^{-x}$ is a sharply-peaked function of x , and so we seek an approximation of the integrand near the maximum. Actually, $F(x)$ is so sharply-peaked that it turns out to be more convenient to consider the logarithm $\ln F(x)$. It is straightforward to show that $\ln F(x)$ (and hence, $F(x)$), has a maximum at $x = n$. We make the change of variable $x = n + \varepsilon$, and expand $\ln F(x)$ in a power-series in ε ,

$$\ln F = n \ln x - x = n \ln(n + \varepsilon) - (n + \varepsilon) \quad (\text{B.46})$$

$$\ln F \approx n \ln n - n - \frac{1}{2} \frac{\varepsilon^2}{n}.$$

Or, in terms of the original integrand,

$$F \approx n^n e^{-n} e^{-\frac{\varepsilon^2}{2n}}. \quad (\text{B.47})$$

From the integral for the factorial, Eq. B.45,

$$n! \approx \int_{-n}^{\infty} n^n e^{-n} e^{-\frac{\varepsilon^2}{2n}} d\varepsilon \approx n^n e^{-n} \int_{-\infty}^{\infty} e^{-\frac{\varepsilon^2}{2n}} d\varepsilon, \quad (\text{B.48})$$

where we have replaced the lower-limit of integration by $-\infty$. The remaining integral is simply the integral of a Gaussian probability distribution. Therefore,

$$n! \approx \sqrt{2\pi n} n^n e^{-n} \quad \text{for } n \gg 1, \quad (\text{B.49})$$

as above.

B.5.4 Eigenvalue perturbation

Suggested references

One of the finest books on general mathematical methods is

- M. L. Boas, *Mathematical Methods in the Physical Sciences*, (Wiley, 2005).

It has gone through several editions, and no harm comes from picking up a cheap used copy of an earlier edition.

For asymptotic and perturbation methods,

- C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory*, (Springer, 1999),

is excellent. It contains a huge survey of methods by masters of the field. It, too, has gone through various editions and can be picked up used. The short book by Hinch,

- E. J. Hinch, *Perturbation methods*, (Cambridge University Press, 1991),

provides a more unified presentation than Bender and Orszag, and may be more useful for self-study.

from the point of in the system, and no effect on the

APPENDIX C

ITÔ CALCULUS

Itô's calculus is used very commonly in financial applications of stochastic processes, presumably because it allows very concise proofs of boundedness, continuity, *etc.*, that would be difficult or impossible using Stratonovich's interpretation of a white-noise stochastic differential equation. A very useful formula developed by Itô is his rule for a change of variables. We will derive the change of variable formula and other related properties of the Itô stochastic integral in this brief appendix.

C.1 Itô's stochastic integral

We would like to assign a meaning to the integral $\int_{t_0}^t G(t')dW(t')$, where $dW(t)$ is the increment of a Wiener process, and $G(t)$ is an arbitrary function. Following the formulation of the Riemann-Stieltjes integral, we divide the domain into N subintervals $[t_0, t_1] \cup [t_1, t_2] \cup \dots \cup [t_{n-1}, t]$ and choose intermediate points τ_i such that $\tau_i \in [t_{i-1}, t_i]$. The integral $\int_{t_0}^t G(t')dW(t')$ is then defined as the (mean-square) limit of the partial sum,

$$S_n = \sum_{i=1}^n G(\tau_i) [W(t_i) - W(t_{i-1})].$$

as $n \rightarrow \infty$. In contrast with the Riemann-Stieltjes integral, however, the limit depends upon the choice of intermediate point τ_i ! One can show that

the two interpretations of the stochastic integral discussed in Section 7.4.1 correspond to particular choices of τ_i – For the Itô interpretation, τ_i is the initial point in the interval $\tau_i = t_{i-1}$, while the Stratonovich interpretation amounts to choosing τ_i as the mid-point $\tau_i = \frac{t_i + t_{i-1}}{2}$.

C.1.1 Example – $\int_0^t W(t')dW(t')$

Writing $W_i \equiv W(t_i)$ and $\Delta W_i \equiv W_i - W_{i-1}$, in the Itô interpretation ($\tau_i = t_{i-1}$),

$$\begin{aligned} S_n &= \sum_{i=1}^n W_{i-1} \Delta W_i = \frac{1}{2} \sum_{i=1}^n \left[(W_{i-1} + \Delta W_i)^2 - (W_{i-1})^2 - (\Delta W_i)^2 \right] \\ &= \frac{1}{2} [W^2(t) - W^2(t_0)] - \frac{1}{2} \sum_{i=1}^n (\Delta W_i)^2. \end{aligned} \quad (\text{C.1})$$

The mean-square limit of the last term can be computed: Since

$$\left\langle \sum_i \Delta W_i^2 \right\rangle = t - t_0, \quad (\text{C.2})$$

we find,

$$\left\langle \left[\sum_i (W_i - W_{i-1})^2 - (t - t_0) \right]^2 \right\rangle = 2 \sum_i (t_i - t_{i-1})^2 \rightarrow 0, \quad (\text{C.3})$$

as $n \rightarrow \infty$. Or, equivalently,

$$\int_{t_0}^t W(t')dW(t') = \frac{1}{2} [W^2(t) - W^2(t_0) - (t - t_0)] \quad (\text{ITÔ}) \quad (\text{C.4})$$

In the Stratonovich interpretation, one can show that (Exercise 2),

$$\begin{aligned} \int_{t_0}^t W(t')dW(t') &= \lim_{n \rightarrow \infty} \sum_i \frac{W(t_i) + W(t_{i-1})}{2} \Delta W_i = \\ &= \frac{1}{2} [W^2(t) - W^2(t_0)] \quad (\text{STRATONOVICH}) \end{aligned} \quad (\text{C.5})$$

For an arbitrary function $G(t)$, the Stratonovich integral has *no relationship whatever to the Itô integral*, unless, of course, $G(t)$ is related to a stochastic differential equation, then there is an explicit relationship as explained in Section 7.4.1.

C.2 Change of Variables Formula

Computations using Itô's calculus are greatly facilitated by the following identities,

$$dW^2(t) = dt \quad \text{and} \quad dW^{2+N}(t) = 0 \quad (N > 0).$$

That is, written more precisely,

$$\int_{t_0}^t G(t') [dW(t')]^{2+N} = \begin{cases} \int_{t_0}^t G(t') dt' & N = 0, \\ 0 & N > 0. \end{cases}$$

These relations lead to a generalized chain rule in the Itô interpretation. For example,

$$\begin{aligned} d\{\exp[W(t)]\} &= \exp[W(t) + dW(t)] - \exp[W(t)] \\ &= \exp[W(t)] \left[dW(t) + \frac{1}{2}dW^2(t) \right] \\ &= \exp[W(t)] \left[dW(t) + \frac{1}{2}dt \right]. \end{aligned}$$

Written in general, for an arbitrary function $f[W(t)]$,

$$df[W(t), t] = \left(\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W^2} \right) dt + \frac{\partial f}{\partial W} dW(t). \quad (\text{C.6})$$

Suppose we have a function $X(t)$ characterized by the Langevin equation,

$$dX(t) = a(X, t)dt + b(X, t)dW(t), \quad (\text{C.7})$$

and $\phi(X)$ is a function of the solution of this stochastic differential equation. Using the generalized chain rule, Eq. C.6, and the Langevin equation, Eq. C.7, we obtain Itô's *change of variables* formula:

$$\begin{aligned} d\phi(X(t)) &= \phi(X(t) + dX(t)) - \phi(X(t)) = \\ &= \phi'(X(t)) dX(t) + \frac{1}{2}\phi''(X(t)) [dX(t)]^2 + \dots \\ &= \phi'(X(t)) [a(X(t), t) dt + b(X(t), t) dW(t)] \\ &\quad + \frac{1}{2}\phi''(X(t)) b^2(X(t), t) [dW(t)]^2 + \dots \\ &= \left\{ a(X(t), t) \phi'(X(t)) + \frac{1}{2} b^2(X(t), t) \phi''(X(t)) \right\} dt \\ &\quad + b(X(t), t) \phi'(X(t)) dW(t). \end{aligned} \quad (\text{C.8})$$

Notice that if $\phi(X)$ is *linear* in $X(t)$, then the anomalous term,

$$\frac{1}{2} b^2(X(t), t) \phi''(X(t)),$$

vanishes and the change of variables formula reduces to the chain rule of ordinary (Stratonovich) calculus.

C.2.1 Example – Calculate $\langle \cos[\eta(t)] \rangle$

Suppose we have white noise governed by the stochastic differential equation

$$d\eta = \Gamma dW, \quad \eta(0) = 0,$$

i.e. $\langle \eta(t) \rangle = 0$ and $\langle \eta(t)\eta(t') \rangle = \Gamma^2 \delta(t - t')$, and we would like to compute $\langle \cos[\eta(t)] \rangle$. Using Itô's change of variables or the generalized chain rule, we arrive at the equation governing $\cos[\eta(t)]$,

$$d \cos \eta = -\frac{\Gamma^2}{2} \cos \eta dt + \Gamma \sin \eta dW,$$

since $a(X) = 0$ and $b(X) = \Gamma$ in Eq. C.8. Taking the average, the last term on the right-hand side vanishes, leaving behind the ordinary differential equation

$$\frac{d\langle \cos \eta \rangle}{dt} = -\frac{\Gamma^2}{2} \langle \cos \eta \rangle, \quad \langle \cos \eta(0) \rangle = 1.$$

The solution is simply,

$$\langle \cos \eta \rangle = e^{-\frac{\Gamma^2}{2} t}.$$

Suggested References

This chapter is taken from Gardiner's *Handbook*,

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004),

that contains a large collection of examples illustrating the range and applicability of Itô calculus.

Exercises

1. Fill in the details in the derivation of Eq. C.4 from Eq. C.1.
2. Prove Eq. C.5.
3. Repeat the calculation of $\langle \cos[\eta(t)] \rangle$ using the average of the cumulant generating function (Eq. A.10 on p. 236) instead of Itô's change of variable formula. Could you use Itô's formula for Gaussian *coloured noise*, (*i.e.* noise with nonzero correlation time)? Could you use the cumulant generating function for Gaussian coloured noise? Repeat the calculation for $\langle \cos[F(t)] \rangle$ for the Ornstein-Uhlenbeck process $F(t)$ with autocorrelation function,

$$\langle \langle F(t)F(t-\tau) \rangle \rangle = \frac{\sigma^2}{2\tau_c} \exp \left[-\frac{|\tau|}{\tau_c} \right].$$

4. **Milstein simulation algorithm:** Using Itô's change of variables formula, and neglecting terms of $O(\Delta t^2)$, show that (Eq. 8.7),

$$\begin{aligned} \int_0^{\Delta t} \{c(y(t), t) - c(y(0), 0)\} dW(t) = \\ c(y(0), 0) \cdot c_y(y(0), 0) \left[\frac{W^2(\Delta t)}{2} - \frac{\Delta t}{2} \right], \end{aligned}$$

where $y(t)$ obeys the Langevin equation (interpreted in the Itô sense),

$$dy = A(y, t)dt + c(y, t)dW(t). \quad (\text{C.9})$$

In that way, justify Milstein's scheme for simulating Eq. C.9,

$$\begin{aligned} y(t + \Delta t) = y(t) + A(y, t)\Delta t + n_1 \cdot c(y, t)\sqrt{\Delta t} \\ - \frac{\Delta t}{2} c(y, t) \cdot c_y(y, t) \cdot (1 - n_2^2), \end{aligned}$$

where n_i are independent samples of a unit Normal distribution $\mathbf{N}(0, 1)$.

APPENDIX D

SAMPLE MATLAB CODE

In this chapter, there are several examples of code used to generate simulation data shown in the main notes. The routines are not fully annotated, and some familiarity with Matlab is assumed. If you are not familiar with Matlab, there are several good references, both print and web-based, including the following:

- *Numerical methods using MATLAB*, J. H. Mathews and K. D. Fink (Pearson, 2004).
- <http://www.cs.ubc.ca/spider/cavers/MatlabGuide/guide.html>

The most conspicuous absence is a lack of fully-assembled code – Only small steps in the simulation algorithms are shown. This is partly to allow the reader to assemble full programs in their own style, and partly because full routines require many tedious lines of code for pre-allocating memory, plotting results, etc, that only detract from the presentation of the algorithms.

D.1 Examples of Gillespie’s direct method

- D. T. Gillespie (1977) “Exact simulation of coupled chemical reactions,” *Journal of Chemical Physics* **81**: 2340.

D.1.1 Defining the propensity vector and stoichiometry matrix

For a system with N reactants and M reactions, we define two functions:

1. The first is run at the initialization of the program, and returns the transpose of the stoichiometry matrix Smat (an $M \times N$ matrix).

```
function Smat = defineReactions(N,M)
%Generates the stoichiometry matrix
Smat = zeros(M,N);
% reactant      1      2      ...      N
Smat(1,:) = [ 1  -1  ...  0 ] ;
Smat(2,:) = [ -1  1  ... -2 ] ;
      :      = [ 2  0  ...  1 ] ;
Smat(M,:) = [ 0  1  ...  1 ] ;
```

2. The second takes as an argument the present state of the system X and the system size OMEGA , returning a vector of the reaction rates Vvec (a $M \times 1$ vector). The microscopic reaction rates ν_j are written in terms of the reactant *concentrations* (see p. 72). The OMEGA multiplying each rate ν_j converts the units from *concentration per time* to *per time*.

```
function vVec=vVecIn(X,OMEGA)
% Function returning vector of M vVec which are functions of
% N reactants X

% Writing the reactions in terms of concentration
Xc=X/OMEGA;

vVec(1) = OMEGA * nu_1(Xc)
vVec(2) = OMEGA * nu_2(Xc)
      :      :
vVec(M) = OMEGA * nu_M(Xc)
```

D.1.2 Core stochastic simulation algorithm

Once the stoichiometry matrix and propensity vector are defined, the core simulation algorithm is straightforward. The current state of the system

X , the current time T , the stoichiometry matrix $Smat$ and the system size Ω are input, and the updated state and time, X_{new} and T_{new} , are output.

```
function [Xnew, Tnew]=CoreSSA(X,T,Smat,OMEGA)

[M,N]=size(Smat);% N reactants, M reactions

%Step 1: Calculate a_mu & a_0

    a_mu = vVecIn(X,OMEGA);
    a_0 = sum(a_mu) ;

%Step 2: Calculate tau and mu using random number generators
% Tau is the *time* the next reaction completes and mu is the
% *index* of
% the next reaction.

    r1 = rand;
    tau = (1/a_0)*log(1/r1);

    r2 = rand;
    next_mu = find(cumsum(a_mu)>=r2*a_0,1,'first');

%Step 3: Update the system:
% carry out reaction next_mu

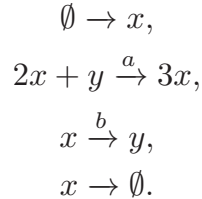
prod = Smat(next_mu,1:N) ; % Stoichiometry of update

    for i=1:N
        Xnew(i) = X(i)+prod(i);
    end

% update the time
Tnew = T + tau;
```

D.1.3 Example - the Brusselator

The Brusselator example (Section 4.2) is a simple model that exhibits limit cycle and stable behaviour over a range of parameter values. The reactions are,



In the Gillespie algorithm, the propensity vector is coded as,

```
function vVec=vVecIn(X,OMEGA)
global a b% if your parameter values are declared as global
variables
x=X(1)/OMEGA;xm1=(X(1)-1)/OMEGA;y=X(2)/OMEGA;

vVec(1) = OMEGA*(1)
vVec(2) = OMEGA*(a*x*xm1*y)
vVec(3) = OMEGA*(b*x)
vVec(4) = OMEGA*(x)
```

The stoichiometry matrix is,

```
function Smat = defineReactions(N,M)
Smat = zeros(M,N);
% reactant      x      y
Smat(1,:) = [ 1  0 ] ;
Smat(2,:) = [ 1 -1 ] ;
Smat(3,:) = [ -1 1 ] ;
Smat(4,:) = [ -1 0 ] ;
```

D.2 Stochastic differential equations

The update step for a simple forward-Euler method is shown. Higher-order methods are possible – see

- *Handbook of stochastic methods (3rd Ed.)*, C. W. Gardiner (Springer, 2004),

for details and references.

D.2.1 White noise

The Euler method for white noise is straight-forward, and since the white noise is uncorrelated, its history does not need to be saved, in contrast with the coloured noise algorithm below, where a past value is needed.

We are simulating a sample trajectory of the process $y(t)$ characterized by the Itô stochastic differential equation,

$$dy = A(y, t)dt + c(y, t)dW(t).$$

The input is the present state of the system $(y(t), t)$ and the time-step dt . The output is the updated state $(y(t + dt), t + dt)$.

```
function [yNew,tNew]=whiteEuler(y,t,dt)

yNew=y+A(y,t)*dt+b(y,t)*dt^(1/2)*randn;
tNew=t+dt;
```

D.2.2 Coloured noise

We are simulating a sample trajectory of the process $y(t)$ characterized by the stochastic differential equation,

$$\frac{dy}{dt} = A(y, t) + c(y, t)\eta(t).$$

where $\eta(t)$ is *coloured noise* with unit variance and correlation time $\tau_c = \text{tauC}$. The input is the present state of the system $(y(t), t, \eta(t))$ and the time-step dt . The output is the updated state $(y(t + dt), t + dt, \eta(t + dt))$.

```
function [yNew,tNew,nNew]=colouredEuler(y,t,n,dt)

rho=exp(-dt/tauC);
yNew=y+a(y,t)*dt+c(y,t)*n*dt;
nNew=rho*n+(1-rho^2)^(1/2)*(1/2/tauC)^(1/2)*randn;
tNew=t+dt;
```