Stochastic Chaos and Thermodynamic Phase Transitions: Theory and Bayesian Estimation Algorithms

by

Zhi-De Deng

S.B. Physics
Massachusetts Institute of Technology, 2006

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Bachelor of Science

and

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2007
[June 2007]
© Zhi-De Deng, MMVII. All rights reserved.

The author hereby grants to MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part.

Author .....
Department of Electrical Engineering and Computer Science
May 25, 2007

Certified by .....
Chi-Sang Poon
Principle Research Scientist
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 0 3 2007
LIBRARIES

ARCHIVES

Stochastic Chaos and Thermodynamic Phase Transitions: Theory and Bayesian Estimation Algorithms

by

Zhi-De Deng

Submitted to the Department of Electrical Engineering and Computer Science
on May 25, 2007, in partial fulfillment of the
requirements for the degrees of
Bachelor of Science
and
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The chaotic behavior of dynamical systems underlies the foundations of statistical mechanics through ergodic theory. This putative connection is made more concrete in Part I of this thesis, where we show how to quantify certain chaotic properties of a system that are of relevance to statistical mechanics and kinetic theory. We consider the motion of a particle trapped in a double-well potential coupled to a noisy environment. By use of the classic Langevin and Fokker-Planck equations, we investigate Kramers' escape rate problem. We show that there is a deep analogy between kinetic rate theory and stochastic chaos, for which we propose a novel definition.

In Part II, we develop techniques based on Volterra series modeling and Bayesian nonlinear filtering to distinguish between dynamic noise and measurement noise. We quantify how much of the system's ergodic behavior can be attributed to intrinsic deterministic dynamical properties vis-a-vis inevitable extrinsic noise perturbations.

Thesis Supervisor: Chi-Sang Poon
Title: Principle Research Scientist

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> The diversity of the phenomena of Nature is so great, and the treasure hidden in the heavens so rich, precisely in order that the human mind shall never be lacking in fresh nourishment.
>
> —Johannes Kepler [154]

Statistical mechanics describes the macroscopic physical properties of a large system of particles, in terms of the average properties of a large ensemble of mechanically identical systems which satisfy the same macroscopic constraints as the particular system of interest [233] [251]. It has been applied successfully to a wide variety of equilibrium systems, from simple molecular gases to white dwarf stars [71]. In spite of its great success, nonequilibrium statistical mechanics suffers from a lack of firm foundations especially concerning the intrinsic dynamical properties which may justify the introduction of irreversible phenomenological equations. The paradox of irreversibility is well known, the macroscopic equations are irreversible due to transport and other dissipative processes, while Hamiltonian equations (microscopic laws of mechanics) are time-reversal invariant. The recognition of the second law of thermodynamics must be something more than a consequence of Newton's laws led to the introduction of probabilistic ideas into this branch of physics.

1.1 Statistical mechanics, thermodynamics, and chaos

James Clerk Maxwell (1831–1879) and Ludwig Eduard Boltzmann (1844–1906) opened the way to reconcile such paradoxes by their introduction of statistical ensembles of trajectories in phase space [24] [191]. They were aware of the need to provide a better mechanical foundation for the second law of thermodynamics beyond the stochastic arguments advanced by Boltzmann. This gave rise to the notion of the **ergodic behavior** of a mechanical system. Boltzmann argued that, for systems with a large number of degrees of freedom,

most of the phase space is taken up by regions where the macroscopic properties of systems in them have values very close to those we associate with thermodynamic equilibrium. Boltzmann then made the hypothesis that a mechanical system's trajectory in phase space will spend *equal times* in regions of *equal phase-space measure*. Granting for the moment the validity of Boltzmann's hypothesis, one can prove that the long-time average of some property of the system is then equal to an ensemble average, taken with respect to the so-called **microcanonical distribution**. From this, one can derive all of the results of statistical thermodynamics for classical systems. Of course, the problem of proving the ergodic character of a given mechanical system was left unsolved today.

After the work of Maxwell and Boltzmann, Josiah Willard Gibbs (1839–1903) [103] introduced the notion of a **mixing system**. Using the idea of mixing a drop of oil in an immiscible fluid, Gibbs argued that if one likened a small set of points in phase-space to the "drop of oil", then the dynamical evolution of such a set would lead to its becoming uniformly mixed over the entire phase space. That is, Gibbs argued that it's reasonable to try to characterize the mixing property of a dynamical system by saying that infinitesimally close points in any small set in phase space will separate so rapidly that the set becomes "strung out" over a large region of phase-space in a short amount of time. If so, we will say that the system is mixing. In later chapters, we will characterize the rate of separation of nearby trajectories in phase-space by quantities called **Lyapunov exponents**. If a system possesses trajectories that separate at an exponential rate, we call such systems **chaotic**.

The irregular behavior of dynamical systems which might allow for the mixing or ergodic behavior is not typical of the simple linear or integrable systems that we usually study in classical mechanics—such as harmonic oscillators or the two-body problems. Jules Henri Poincaré (1854–1912), in his analysis of the gravitational three-body problems, showed that the motion of such systems can be very complicated indeed, and not easily described by ordinary mathematical notions of continuity, differentiability and perturbation analysis. These systems characteristically exhibit the phenomenon of a **homoclinic tangle**, which is a complicated intersection of curves describing the system's motion in time, which remains complicated on an infinitesimally fine scale. Thus, in the early part of the 20th century there was a convergence of two lines of considerations for mechanical systems—one from statistical mechanics, and one from celestial mechanics—both of which suggested that most dynamical systems should have properties very different from linear systems. The advent of modern computers has opened an era of systematic exploration of phase-space structures that helped solidify our understanding of chaos in dynamical systems.

In the late 1930s and early 40s, visionaries like Norbert Wiener (1894–1964) [297], Stanslaw Ulam (1909–1986) and John von Neumann (1903–1957) [283] paved the way to the fusion of dynamical systems theory with probability theory. Major advances in the theory of dynamical systems were made by the group in Russia around Andrei N. Kolmogorov (1903–1987) [160], and including Dmitri V. Anosov [9], Yakov G. Sinai [261], Yakov B.

Pesin [221], and others. In the late 1940s, the connection between ergodic theory and information theory developed by Claude E. Shannon (1916–2001) [254] was a turning point which led to the widespread recognition that deterministic dynamical systems may be as random as a coin tossing process. Statistical formalism of dynamical chaos, elaborated in particular by Rufus Bowen and David P. Ruelle [27] from the analogy between the time randomness of chaotic trajectories and the space randomness of a system configuration in equilibrium statistical mechanics. These authors developed the mathematical foundations for a rigorous treatment of the properties of what are now known as hyperbolic dynamic systems, which, very loosely speaking, are those which exhibit exponential separation of phase-space trajectories. Moreover, Sinai, Ruelle, and Bowen were able to show that there is a deep analogy between equilibrium statistical mechanics and methods used to characterize the mathematical properties of hyperbolic dynamical systems. The formal results suggested by this analogy, now called the **thermodynamic formalism**, have proved to be very powerful in the applications of dynamical systems theory to statistical mechanics.

The chaotic behavior of dynamical systems underlies the foundations of statistical mechanics through ergodic theory. It is the purpose of this thesis to try to make these ideas more concrete, and show how to quantify some of the chaotic properties that are of interest for statistical mechanics and kinetic theory. We will see a number of deep connections between the dynamical properties of a system, such as its Lyapunov exponents, and its transport properties. We will discuss an even simpler dynamical system, namely the logistic map, and show how to calculate dynamical quantities from kinetic theory.

1.1.1 Nonlinear time series analysis

Takens' **Delay Embedding Theorem** [274] forms the basis of virtually all approaches to the analysis of time series generated by nonlinear deterministic dynamical systems. The theorem gives the conditions under which a chaotic dynamical system can be reconstructed from a sequence of observations of the state of a dynamical system. The reconstruction preserves the properties of the dynamical system that do not change under smooth coordinate changes, but it does not preserve the geometric shape of structures in phase space. This provides the theoretical foundation for many popular techniques, including those for the measurement of fractal dimensions and Lyapunov exponents, for the prediction of future behavior, noise reduction and signal separation, control and targeting; it gives rise to a virtually new branch of research of **chaotic time series analysis**. The original version of Takens' theorem assume that the underlying system is autonomous and noise-free, which has been recently extended to deterministically forced systems [268] as well as for arbitrarily and stochastically forced systems [269].

1.1.2 So what is the benefits of knowing that a system is chaotic?

Chaos and noise are two different types of irregularities. In practice, it would be interesting and important to distinguish between the two different sources of randomness. Philosophical implications apart, such a distinction would enable us to use the most appropriate tools for system analysis. If the source is from a chaotic system, then using techniques from nonlinear dynamics can increase sensitivity and specificity in predicting outcome. The nonlinear dynamics measure more accurately tracks output of the system because it is based on the presumption that the variation is deterministic rather than being random. The control of chaos in biological systems [296] may be important in the dynamics of gene expression and translation [262], reaction-diffusion models of blood clotting [178] [294], human heart-rate dynamics [108] and cardiac tissue [214], and nerve cells [33] [159]; see Ref. [105] for a discussion on the origin and significance of complex physiological rhythms. On the other hand, if the source is from a stochastic dynamical system, then such chaos-dynamics-based approaches are often inappropriate unless we make suitable modifications.

1.2 Noise-induced phenomena

Noise in physical systems is like the well-known supervillain in the Batman comics—Two-Face. Both halves of its face are relevant to the understanding of complex systems. The ugly, destructive face of noise—based mainly on our day to day experience —corresponds to the familiar blurring effect by random fluctuations of otherwise well-defined quantities, the perturbation of ordered systems, and the destruction of fine detail in intricate patterns. The ugly face of noise is often observed in relation to studies of chaotic phenomena in real systems, where it is usually the effect of internal (e.g. temperature fluctuations) or external noise that sets the practical lower limit on the range of coordinate scales over which, for example, fractal effects can persist.

It is perhaps less well known—but it is one of the central themes of this thesis—that noise can also exhibit a face that is beautiful, in the sense that its effect can also be positive and constructive. In contrast to the notion that noise in physical systems dominated by determinism has small effects, recent research has demonstrated otherwise. Recent studies have shown that noise may greatly enrich dynamics of nonlinear systems [257] [303] [305]. Specifically, noise in deterministic dynamical systems plays prominent role in global behavior due to large-scale effects governed by transport in phase space [20]. Examples of such behavior appear as **stochastic resonance** [58] [87], and phenomena of the so-called **noise-induced order** [189], of the **noise-induced instability** [34] [48], and of the noise-induced chaos-order transitions [100]. One particular problem is the identification of the cause of qualitatively new emergent dynamics that are not observable in deterministic systems, such as **stochastic chaos** [59]. Such chaotic stochastic processes are likely to be ubiquitous, especially in biological and chemical dynamics, because the ingredients of the underlying

mechanism are very common. Moreover, it is a more stable phenomenon than the existence of a deterministic chaotic attractor [228].

Noise-induced transitions in nonlinear dynamics far from equilibrium has become a rapidly growing area of modern statistical physics [95] [245] [302]. Practical complex systems where noise is playing a critical role are: complex physiology rhythmic processes [105]; critical point phenomena in materials including supercooling and glass formation [85]. Sethna et al. [253] show that the seemingly random, impulsive events by which many physical systems evolve exhibit universal—and, to some extend, predictable—behavior. These few examples illustrate the general observation that noise may indeed cause a global change in real dynamical systems, thus making it clear that stochastic effects play a definitive role in the dynamics and must be accounted for in our models and our analysis of them. Remarkably they amount to a symbiotic relationship of order and randomness in contrast to the commonly held view that order and randomness form an antagonistic pair. The existence of noise-induced transitions clearly forces us to reappraise the role of noise.

1.2.1 Models of dynamic noise and measurement noise

The study of the influence of noise on systems with purely temporal dependence was culminated in 1984 by Horsthemke and Lefever [138]. In time series problems, noise can be divided into two categories: **dynamic noise** that drives the process, and **observation noise**, which is added in the measurement process. Observation noise of a high level can pose a severe problem if it is not handled properly, leading to models that underestimate the functional relation between past and future values. A typical example of such observation noise is when an electrocardiogram recorder that measurements the electric activity of the heart over time: fluctuations in the electrode position, body movements, low-frequency drifts, etc., will affect the value of the displayed output, but will not influence the dynamics of the heart. In contrast, dynamic noise does influence future values of the system. For example, in an autoregressive process, the noise drives the system. And subsequent values are derived from that perturbed state. Fig. 1-1 illustrates how dynamic noise and measurement noise interact with a nonlinear system: dynamic noise may alter the behavior of the physical system, whereas measurement noise affects the observed signal without altering the system dynamics itself.

Some interesting questions to be addressed are: In what way would dynamic noise affect the characteristic of the nonlinear deterministic system? To what extent can dynamic noise induce complex (chaos-like) behavior in the output of the system, even though the underlying system is not chaotic by itself? How to distinguish deterministic chaos from stochastic chaos induced by dynamic noise? This knowledge can be exploited to serve good purposes, such as forecasting, control and better understanding of underlying mechanisms.

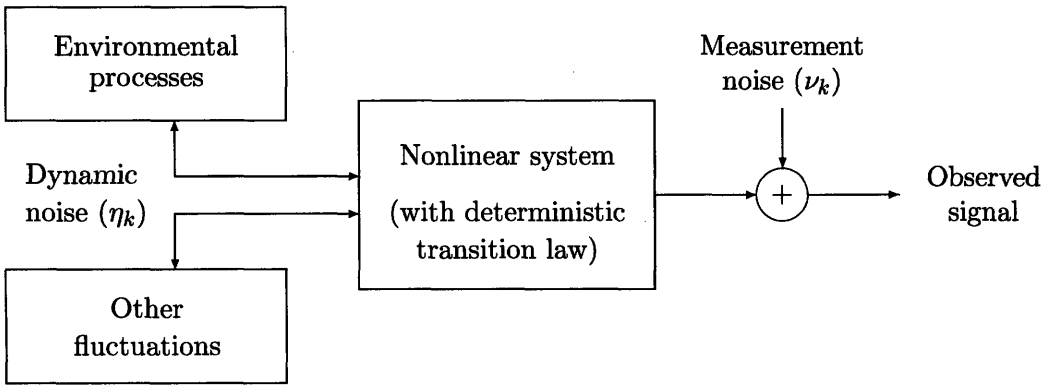


Figure 1-1: Block diagram of a complex system comprising a nonlinear deterministic system perturbed by: (a) dynamic noise, η_k, at the input; and (b) measurement noise, ν_k, at the output.

1.3 Organization of thesis

The outline of the thesis is the following: Part I presents the formalism for the description of noise-induced phenomena in dynamical systems coupled to a fluctuating environment and gives a detailed study of stochastic chaos and noise-induced transitions. Following this introduction, an overview of the noise-induced transitions is presented in Chapter 2 with a brief account of the mathematical tools which allow us to deal with fluctuations: the Langevin equation and the Fokker-Planck equation. Chapter 3 establishes the connection between kinetic rate theory and chaotic dynamics. In Chapter 4, we characterize stochastically perturbed nonlinear systems using an array of tools from time series analysis.

Part II is completely methodological, with emphasis on numerical techniques from nonlinear state estimation theory for the characterization of systems that exhibit stochastic chaos. Chapter 5 covers some nonlinear filtering techniques, including the Kalman filter and its extensions, and the particle filter. In Chapter 6, we present some additional topics in nonlinear filtering. Chapter 7 considers the problem of discriminating between dynamic and measurement noise using the nonlinear filtering tools developed in the previous chapters. Conclusions and future work are discussed in Chapter 8.

In Appendices A and B we give the derivation for some formulae and equations used in this thesis. Appendix C gives a simple statistical test for white noise. An introduction to support vector machine is presented in Appendix D.

Part I

Stochastic Chaos & Thermodynamic Phase Transitions

Chapter 2

Noise-Induced Transitions

> Thus we may say that all physical processes depending on the irregular motion of molecules go in the direction of increasing probability and the state of equilibrium, when nothing more happens, corresponds to the maximum of probability.
>
> —George Gamow [88]

The study of phase transitions is among the most fascinating fields in physics. Originally limited to transition phenomena in equilibrium systems, this field has outgrown its classical confines during the last few decades. The behavior of far-from-equilibrium systems has received more and more attention and has been an extremely active subject of research for physicists, chemists and biologists [101] [170]. The notion of phase transition has been proven fruitful in application to nonequilibrium instabilities known for about eight decades, like certain hydrodynamic instabilities [179] [207], as systems such as the laser [174]; in thermochemical bistabilities [36], as systems such as the Belousov-Zhabotinskii reaction [117] [118] [299] [304]; and in biological systems [67].

Phase transition can be extended further to a new class of transition phenomena which occur only in nonequilibrium systems subjected to a randomly fluctuating environment [140]. In other words, for these systems the environment is not constant in time as is usually assumed in the study of nonequilibrium phenomena but displays random temporal variations, also called external noise. These new transition phenomena present a fascinating subject of investigation since, contrary to all intuition, the environment randomness induces a more structured behavior of the system. This new type of nonequilibrium transition phenomena is called **noise-induced transitions** in order to stress the essential role of the noise.

In this chapter, we present the formalism for the description of nonlinear systems coupled to a random environment. Specifically, we examine systems whose dynamics is described by **stochastic partial differential equations (SPDEs)** of the Langevin type in Sec. 2.1. In Sec. 2.2, we study the Fokker-Planck equation, from which any averages of macroscopic

variables are obtained. Then finally in Sec. 2.3, we use the Fokker-Planck equation to solve a class of problems involving a particle in a bistable potential.

2.1 The Langevin equation

The state of many physical, chemical, and biological systems can be described by a single time-dependent variable. If we denote this variable by $\xi(t)$, consider that the system is autonomous (i.e., with no explicitly time-dependent forces) and neglect for the moment all fluctuations, the dynamical behavior of the system can be described in a general way by the ordinary differential equation

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = f(\xi) \equiv \frac{\mathrm{d}V}{\mathrm{d}\xi}, \tag{2.1}$$

where $f(\xi)$ is the force that drives the system and $V(\xi)$ is a "potential" function defined in such a ways that its value always decreases along any trajectory followed by the system ($\mathrm{d}V/\mathrm{d}\xi < 0$). Therefore, the system evolves in time toward one of the minima of $V(\xi)$, depending on the initial condition, but independently of whether these minima are relative or global. Relative minima correspond to metastable states, from which any system is well known to escape after a certain time. However, this situation is not described in model eq. (2.1), for which when the system falls in a relative minimum, it remains there forever afters. if we want the system to escape from metastable minima, we need to include fluctuations in the dynamical description. But is that the only effect of those fluctuations?

2.1.1 Brownian motion of a free particle

In the studies on Brownian motion we are primarily concerned with the perpetual irregular motions exhibited by small particles of colloidal size immersed in a fluid. The collision frequency is too high and the changes in the particle's path caused by each single impact are too fine to be discerned by the observer. But the superposition of many small interactions produces an observable effect, which we can describe by the theory of stochastic processes. The exact path of the particle cannot be followed in any detail and this is the only reason why we have to consider this problem in stochastic way.

As in well known, this problem was first solved by Einstein [72]. The modern theory of the Brownian motion of a free particle generally starts with the Langevin equation [96] [236] [282] [295]. From statistical mechanics, it was known that the mean kinetic energy of the particle should, in equilibrium, reach a value

$$\left\langle \frac{1}{2}mv^2 \right\rangle = \frac{1}{2}k_{\mathrm{B}}T, \tag{2.2}$$

where T is the absolute temperature, and k_{B} is Boltzmann's constant. The influence of the

surrounding medium on the motion of the particle can be split into two parts:

1. a viscous drag: this is derived from macroscopic hydrodynamics; the drag force is governed by Stokes' law which states that the frictional force decelerating a spherical particle of radius a and mass m is given by $6\pi\eta av/m$, where v is the velocity of the particle and η is the viscosity of the surrounding fluid.

2. a fluctuation force $\Gamma(t)$ which represents the incessant impacts of the molecules of the liquid on the particle. Since we expect this force to be positive and negative with equal probability, we naturally assume that its mean, over an ensemble of particles, is zero. In addition, we assume that there will be correlation between the value of $\Gamma(t)$ at different times t and t' only when $|t-t'|$ is very small. That is,

$$\begin{aligned} \langle \Gamma(t) \rangle &= 0, \\ \langle \Gamma(t)\Gamma(t') \rangle &= \sigma^2 \delta(t-t'). \end{aligned} \tag{2.3}$$

The equation of motion for the displacement of the particle is given by Newton's law as

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = \frac{-6\pi\eta a}{m}\frac{\mathrm{d}x}{\mathrm{d}t} + \Gamma(t). \tag{2.4}$$

Langevin's equation was the first example of the **stochastic differential equation**—a differential equation with a random term Γ. Each solution of the Langevin's equation represents a different random trajectory and, and therefore the probability distribution is the most general thing the theory can predict.

2.1.2 Nonlinear Langevin equation

For one stochastic variable ξ, the general Langevin equation has the form

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = h(\xi, t) + g(\xi, t)\Gamma(t). \tag{2.5}$$

The Langevin force $\Gamma(t)$ is again assumed to be a Gaussian random variable with zero mean and δ correlation function. The constant σ^2 in eq. (2.3) describing the noise strength may be absorbed into the function g.

For constant g, eq. (2.5) is called a Langevin equation with an additive noise force. For g depending on ξ one speaks of a Langevin equation with a multiplicative noise term [285]. This distinction between additive and multiplicative noise is not very significant because for the one variable equation (2.5), for time-independent h and g and for $g \neq 0$, the multiplicative noise always becomes an additive noise by a simple transformation of variables.

For ξ-dependent g the following difficulty arises. Because the noise $\Gamma(t)$ has no correlation time (2.3), it is not yet clear which ξ value one has to use in the function g in eq. (2.5).

If, for instance, $\Gamma(t)$ is an impulse train, the stochastic variable $\xi(t)$ will jump at every time τ when such a peaked function occurs. The question then arises: Which ξ value must one use in g? One may use the value ξ just before τ or just after τ or some value between these two values. From a purely mathematical point one cannot answer this question, but one has to use some additional specification, for instance, the Itô or the Stratonovich definition. Here we assume as it is usually done in physics that the $\delta(t)$ function is replaced by a function $\delta_\epsilon(t)$ with a very small finite width ϵ. In the final result one then has to take the limit $\epsilon \to 0$. With this procedure we see that the average value of $\langle g(\xi,t)\Gamma(t)\rangle$ is no longer zero if g depends on ξ. The above average leads to the "spurious" or "noise-induced" drift.

Usually a formal general solution of eq. (2.5) cannot be obtained. As shown in Sec. (2.2), we can set up a Fokker-Planck equation by which the probability density of the stochastic variable can be calculated.

2.1.3 Internal noise

There are several types of internal fluctuations, but we will focus here on the one that dominates in classical macroscopic systems: thermal fluctuations. Consider a particle that satisfies the Langevin's equation:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -6\pi\eta a\frac{\mathrm{d}x}{\mathrm{d}t} + m\Gamma(t). \tag{2.6}$$

We multiply by x on both sides of the equation, we get

$$\frac{m}{2}\frac{\mathrm{d}^2(x^2)}{\mathrm{d}t^2} - \left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 = -3\pi\eta a\frac{\mathrm{d}(x^2)}{\mathrm{d}t} + mx\Gamma(t), \tag{2.7}$$

We now average over the ensemble of particles and use eq. (2.2) to obtain an equation for $\langle x^2\rangle$:

$$\frac{m}{2}\frac{\mathrm{d}^2\langle x^2\rangle}{\mathrm{d}t^2} + 3\pi\eta a\frac{\mathrm{d}\langle x^2\rangle}{\mathrm{d}t} = k_\mathrm{B}T, \tag{2.8}$$

The term $\langle mx\Gamma(t)\rangle$ vanishes because Γ is zero mean. One then finds the general solution

$$\frac{\mathrm{d}\langle x^2\rangle}{\mathrm{d}t} = \frac{k_\mathrm{B}T}{3\pi\eta a} + C\mathrm{e}^{-\frac{6\pi\eta a t}{m}}, \tag{2.9}$$

where C is an arbitrary constant. Langevin estimated that the decaying exponential approaches zero with a time constant on the order of 10^{-8} s. Thus, for practical purposes, we can neglect this term and integrate once more to get

$$\langle x^2\rangle - \langle x_0^2\rangle = \frac{k_\mathrm{B}T}{3\pi\eta a}t. \tag{2.10}$$

From this, we can show that the intensity of the fluctuations, measured by the parameter σ^2, obeys the **Einstein relation** or the **fluctuation-dissipation relation**

$$\sigma^2 = \frac{k_B T}{6\pi\eta a}. \tag{2.11}$$

This relation ensures that the steady probability distribution $p_{st}(x, \dot{x})$ is the Boltzmann equilibrium distribution,

$$p_{st}(x, \dot{x}) \sim \exp\left\{-\frac{\mathcal{H}(x, \dot{x})}{k_B T}\right\}, \tag{2.12}$$

where $\mathcal{H}(x, \dot{x})$ is the classical Hamiltonian of the model

$$\mathcal{H}(x, \dot{x}) = \frac{\dot{x}^2}{2} + V(x). \tag{2.13}$$

As we will show in the Sec. 2.2, the temporal evolution of x is given by the **Fokker-Planck equation**, which can be found to be [96] [239]

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left[f(x)p(x,t)\right] + \sigma^2 \frac{\partial^2}{\partial x^2} p(x,t). \tag{2.14}$$

The maxima of the steady probability distribution $p_{st}(x)$ (i.e., the probability distribution in the limit $t \to \infty$) are assumed to correspond to the stable steady states of the system. They are indeed the most probable states, those in which the system stays for more time. The steady probability distribution is found by solving eq. (2.14) with the left-hand-side term set equal to zero. Assuming natural boundaries, the solution is simply

$$p_{st}(x) = N \exp\left\{\frac{1}{\sigma^2}\int^x \mathrm{d}x'\, f(x')\right\}, \tag{2.15}$$

where N is just a normalization constant. Taking into account the definition of the deterministic force $f(x)$ given in eq. (2.1) and the fluctuation-dissipation relation given by eq. (2.11), the steady-state probability distribution can be written in the following Boltzmann form:

$$p_{st}(x) \sim \exp\left\{-\frac{V(x)}{k_B T}\right\}. \tag{2.16}$$

Taking this result into account, we can generalize the concept of internal noise to those fluctuations that are measured by $k_B T$, and that leads to a Boltzmann like steady-state probability distribution with a suitable "potential."

The maxima of the probability distribution (2.16) correspond to the most probable states of the system. Since the potential $V(x)$ appearing in expression (2.16) coincides with the deterministic potential (2.1), one can conclude that internal noise does not change the deterministic steady state behavior of the system, its only effect being to broaden the probability peaks, which in the deterministic case are just Dirac deltas. Therefore, this

kind of noise is not able to induce any transition in the behavior of the system, except for trivial disordering transitions, such as the one that will be presented in what follows in the context of the Landau model.

Landau model

Prior to studying the effect of internal additive noise in a particular zero-dimensional system, namely, the Landau model, we present an example of a deterministic transition in this model. Transitions in zero-dimensional stochastic systems are reflected in changes in the shape of the probability distribution of the state variable (the number and position of its extrema change as the transition point is surpassed). Therefore, even though the system runs over the whole phase space in the ordered state (owing to the fluctuations), the shape of the probability distribution reveals the different behavior of the system in its time evolution.

Let us consider a system defined in terms of a free-energy function $F(x)$:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\mathrm{d}F}{\mathrm{d}x} + \eta(t), \tag{2.17}$$

with $\eta(t)$ a Gaussian white noise. The Landau model is based on the assumption that the free energy F can be expanded in a Taylor series around $x = 0$ in the following way:

$$F(x) = a_0 + a_2 x^2 + a_4 x^4, \tag{2.18}$$

where the odd terms are set equal to zero because we require F to be rest invariant under a reversal in the sign of the variable. With this form of the free energy, and after a suitable scale transformation, the evolution equation can be rewritten as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \alpha x - x^3 + \eta(t). \tag{2.19}$$

The linear coefficient α will be used as the control parameter of the system. The shape of the Landau free energy changes from having a minimum at $x = 0$ for $\alpha < 0$ to having two minima ($x = \pm\sqrt{\alpha}$) for $\alpha > 0$. The qualitative effect of the internal fluctuations consists of broadening the probability peaks around these steady states.

The steady probability density of x can be calculated exactly from eq. (2.15):

$$p_{\mathrm{st}}(x) = N \exp\left\{\frac{1}{\sigma^2}\left(\alpha\frac{x^2}{2} - \frac{x^4}{4}\right)\right\}, \tag{2.20}$$

whose maxima are $x_m = 0$ for $\alpha < 0$ and $x_m = \pm\sqrt{\alpha}$ for $\alpha > 0$, in agreement with the free-energy analysis presented previously. Hence, this model describes a continuous phase transition located at $\alpha = 0$. This transition is properly characterized by means of the probability density and its extrema. The value of the first-order statistical moment

$\langle x \rangle = \int \mathrm{d}x x p_{\mathrm{st}}(x)$, however, remains 0 for all α, due to the symmetric character of eq. (2.20).

It would be worth analyzing the effect of an increase in noise intensity (that is, in temperature) on the system. Will an ordered state ($\alpha > 0$) become disordered as expected according to intuition? The mean value $\langle x \rangle$ is still 0, whereas the quantity $\langle |x| \rangle$ behaves with an initial decrease showing the disordering character of the noise, followed by an increase reflecting the enhancement of fluctuations. More representative is the variance σ_x^2

$$\sigma_x^2 = \frac{\langle x^2 \rangle - \langle x \rangle^2}{\sigma^2}. \tag{2.21}$$

As shown, σ_x^2 presents a maximum for a positive value of the noise intensity σ^2, presumably indicating the existence of a disordering transition. Nevertheless, this presumed transition is intrinsically different from the previous transition controlled by α: here, unlike that case, a variation in the noise intensity parameter does not lead to a change in the shape of the stochastic potential, nor in the probability density and its extrema.

2.1.4 External noise

We have just seen that an additive noise is not able to modify the deterministic steady states of a zero-dimensional system. It would be interesting to analyze what happens in the case of an *external* noise, which consists of those fluctuations that are not of thermal origin, and whose steady probability distribution is not Boltzmann like.

As a very common example, let us consider the particular case of a multiplicative white noise. The Langevin equation is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x) + g(x)\eta(t). \tag{2.22}$$

And the corresponding Fokker-Planck equation, in the Stratonovich interpretation, can be found to be [96] [239]

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x} f(x) p(x,t) + \sigma^2 \frac{\partial}{\partial x} g(x) \frac{\partial}{\partial x} g(x) p(x,t). \tag{2.23}$$

The stationary solution of this equation with natural boundaries is

$$p_{\mathrm{st}}(x) = \frac{N}{g(x)} \exp\left\{ \int \mathrm{d}x \, \frac{f(x)}{\sigma^2 g^2(x)} \right\}, \tag{2.24}$$

and a "stochastic potential" $\Phi(x)$ can be defined in this case as

$$\Phi(x) \equiv -\int \mathrm{d}x \, \frac{f(x)}{g^2(x)}. \tag{2.25}$$

As in previous examples, the steady state of the system will be characterized by the minima

of this stochastic potential. But now, because $g(x)$ is no longer a constant function, $\Phi(x)$ is not directly proportional to the potential $V(x)$, eq. (2.1), and therefore the stable steady states can be different from the deterministic ones. Hence, external noise can be expected to have an important influence in zero-dimensional systems.

2.2 The Fokker-Planck equation

In the case of Brownian motion, the complete solution of a macroscopic system would consist in solving all the microscopic equations of the system. Because we cannot generally do this we use instead a stochastic description, i.e., we describe the system by macroscopic variables which fluctuate in a stochastic way. Macroscopically, for an ensemble of particles or systems, the variations can be modeled by a diffusion process. The distribution function of the random variables of the system will, therefore, fulfill a partial differential equation of the diffusion type. The Fokker-Planck equation is just an equation of motion for the distribution function of fluctuating macroscopic variables. Many books and reviews on the Fokker-Planck equation now exist [47] [171] [172] [173].

By solving the Fokker-Planck equation one obtains distribution functions from which any statistical moments of macroscopic variables are obtained by integration. Since the application of the Fokker-Planck equation is not restricted to systems near thermal equilibrium, we may as well apply it to systems far from thermal equilibrium [119]. The Fokker-Planck equation not only describes stationary properties, but also the dynamics of systems, if the appropriate time-dependent solution is used.

2.2.1 Kramers-Moyal expansion

We start with a Markov process W whose transition properties are dictated by P_{T}:

$$W(x,t+\tau) = \int \mathrm{d}x'\, P_{\mathrm{T}}(x,t+\tau|x',t)W(x',t), \tag{2.26}$$

To derive an expression for the differential $\partial W(x,t)/\partial t$, we must know the transition probability $P_{\mathrm{T}}(x,t+\tau|x',t)$ for small τ. We first assume that all the moments of P_{T} are given

$$M_n(x',t,\tau) = \langle[\xi(t+\tau)-\xi(t)]^n\rangle\big|_{\xi(t)=x'} = \int \mathrm{d}x\,(x-x')^n P_{\mathrm{T}}(x,t+\tau|x',t). \tag{2.27}$$

We now derive a general expansion of the transition probability by starting with

$$P_{\mathrm{T}}(x,t+\tau|x',t) = \int \mathrm{d}y\,\delta(y-x)P_{\mathrm{T}}(y,t-\tau|x',t) \tag{2.28}$$

and using the formal Taylor series expansion of the δ function in the form

$$\begin{aligned}\delta(y-x) &= \delta(x'-x+y-x') \\ &= \sum_{n=0}^{\infty} \frac{(y-x')^n}{n!} \left(\frac{\partial}{\partial x'}\right)^n \delta(x'-x) \\ &= \sum_{n=0}^{\infty} \frac{(y-x')^n}{n!} \left(-\frac{\partial}{\partial x}\right)^n \delta(x'-x),\end{aligned} \tag{2.29}$$

we get

$$\begin{aligned}P_{\mathrm{T}}(x,t+\tau|x',t) &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n \int \mathrm{d}y\, (y-x')^n P_{\mathrm{T}}(y,t+\tau|x',t)\delta(x'-x) \\ &= \left[1+\sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n M_n(x',t,\tau)\right] \delta(x'-x) \\ &= \left[1+\sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n M_n(x,t,\tau)\right] \delta(x-x').\end{aligned} \tag{2.30}$$

In deriving the last line, we use the fact that $\delta(x'-x) = \delta(x-x')$ and the sifting property of the delta function. Inserting eq. (2.30) into eq. (2.26) leads to

$$\begin{aligned}W(x,t+\tau) - W(x,t) &= \frac{\partial W(x,t)}{\partial t}\tau + \mathcal{O}(\tau^2) \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n \int \mathrm{d}x' \delta(x-x') M_n(x,t,\tau) W(x',t) \\ &= \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n \frac{M_n(x,t,\tau)}{n!} W(x,t).\end{aligned} \tag{2.31}$$

We now assume that moments M_n can be expanded into a Taylor series with respect to τ

$$\frac{M_n(x,t,\tau)}{n!} = \mathcal{D}^{(n)}(x,t)\tau + \mathcal{O}(\tau^2). \tag{2.32}$$

The term with τ^0 must vanish, because for $\tau = 0$ the transition probability P_{T} has the initial value

$$P_{\mathrm{T}}(x,t|x',t) = \delta(x-x'). \tag{2.33}$$

By taking into account only the linear terms in τ we have

$$\frac{\partial W(x,t)}{\partial t} = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n \mathcal{D}^{(n)}(x,t) W(x,t) = \mathcal{L}_{\mathrm{KM}} W(x,t), \tag{2.34}$$

where the **Kramers-Moyal operator** $\mathcal{L}_{\mathrm{KM}}$ is defined by

$$\mathcal{L}_{\mathrm{KM}}(x,t) = \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^n \mathcal{D}^{(n)}(x,t). \tag{2.35}$$

Eq. (2.34) is the **Kramers-Moyal expansion.**

2.2.2 Pawula theorem

Here we state the Pawula theorem [216], which states that for a positive transition probability P_{T}, the expansion (2.35) may stop either after the first term or after the second term; if it does not stop after the second term, then it must contain an infinite number of terms. If the expansion stops after the second term, then eq. (2.34) becomes the Fokker-Planck equation.

To derive the Pawula theorem we need the generalized Schwarz inequality

$$\left(\int \mathrm{d}x\, f(x)g(x)P(x)\right)^2 \leq \left(\int \mathrm{d}x\, f^2(x)P(x)\right)\left(\int \mathrm{d}x\, g^2(x)P(x)\right). \tag{2.36}$$

where $P(x)$ is a nonnegative function and $f(x)$ and $g(x)$ are arbitrary functions.

We now apply eq. (2.36) with

$$\begin{aligned} f(x) &= (x-x')^n; \\ g(x) &= (x-x')^{n+m}; \\ P(x) &= P(x,t+\tau|x',t') \end{aligned} \tag{2.37}$$

and thus obtain for the moments in eq. (2.27) the inequality

$$M_{2n+m}^2 \leq M_{2n} M_{2n+2m}. \tag{2.38}$$

For $n=0$ we have $M_m^2 \leq M_{2m}$. This relation is obviously fulfilled for $m=0$ ($M_0=1$). For $m \geq 1$ no restriction follows from this relation for the short time expansion coefficients $\mathcal{D}^{(n)}$ of M_n. For $m=0$, $M_{2n}^2 \leq M_{2n}^2$, which is trivially fulfilled for every n. Thus we need to consider eq. (2.38) only for $n \geq 1$ and $m \geq 1$. By inserting eq. (2.32) into eq. (2.38), dividing the resulting inequality by τ^2 and taking the limit $\tau \to 0$ we then obtain the following inequality for the expansion coefficients $\mathcal{D}^{(n)}$:

$$\left[(2n+m)!\mathcal{D}^{(2n+m)}\right]^2 \leq (2n)!(2n+2m)!\mathcal{D}^{(2n)}\mathcal{D}^{(2n+2m)}. \tag{2.39}$$

If $\mathcal{D}^{(2n)}$ is zero, $\mathcal{D}^{(2n+m)}$ must be zero, too, i.e.,

$$\mathcal{D}^{(2n)} = 0 \Rightarrow \mathcal{D}^{(2n+1)} = \mathcal{D}^{(2n+2)} = \cdots = 0 \quad \forall n \geq 1. \tag{2.40}$$

Furthermore if $\mathcal{D}^{(2n+2m)}$ is zero, $\mathcal{D}^{(2n+m)}$ must be zero, too, i.e.,

$$\begin{aligned} \mathcal{D}^{(2r)} = 0 \Rightarrow \mathcal{D}^{(r+n)} = 0 \quad \text{for } n = 1, \ldots, r-1; \\ \mathcal{D}^{(2r-1)} = \cdots = \mathcal{D}^{(r+1)} = 0 \quad \text{for } r \geq 1. \end{aligned} \tag{2.41}$$

From eq. (2.40) and the repeated use of eq. (2.41), one concludes that if any $\mathcal{D}^{(2r)} = 0$ for $r \geq 1$, all coefficients $\mathcal{D}^{(n)}$ with $n \geq 3$ must vanish, i.e.,

$$\mathcal{D}^{(2r)} = 0 \Rightarrow \mathcal{D}^{(3)} = \mathcal{D}^{(4)} = \cdots = 0 \quad \forall r \geq 1. \tag{2.42}$$

The Pawula theorem immediately follows from the last statement.

If the Kramers-Moyal expansion in eq. (2.34) truncates after the second term, we get the general Fokker-Planck equation for one variable x has the form

$$\frac{\partial W(x,t)}{\partial t} = \left[-\frac{\partial \mathcal{D}^{(1)}(x)}{\partial x} + \frac{\partial^2 \mathcal{D}^{(2)}(x)}{\partial x^2} \right] W(x,t), \tag{2.43}$$

where $\mathcal{D}^{(2)} > 0$ is called the **diffusion coefficient** and $\mathcal{D}^{(1)}$ the **drift coefficient.**

2.2.3 Diffusion and drift coefficients for the Langevin equation

To derive the diffusion and drift coefficients for the nonlinear Langevin equation, we first write the Langevin equation (2.5) as an integral equation

$$\xi(t+\tau) - x = \int_t^{t+\tau} \mathrm{d}t' \left[h\left(x(t'), t'\right) + g\left(x(t'), t'\right) \Gamma(t') \right] \tag{2.44}$$

and assume that $h(\cdot)$ and $g(\cdot)$ can be expanded according to

$$\begin{aligned} h\left(\xi(t'), t'\right) &= h\left(x, t'\right) + \frac{\partial h\left(x, t'\right)}{\partial x} \left[\xi(t') - x\right] + \cdots \\ g\left(\xi(t'), t'\right) &= g\left(x, t'\right) + \frac{\partial g\left(x, t'\right)}{\partial x} \left[\xi(t') - x\right] + \cdots \end{aligned} \tag{2.45}$$

Inserting eq. (2.45) into eq. (2.44) leads to

$$\begin{aligned} \xi(t+\tau) - x = &\int_t^{t+\tau} \mathrm{d}t'\, h\left(x, t'\right) + \int_t^{t+\tau} \mathrm{d}t' \frac{\partial h\left(x, t'\right)}{\partial x} \left[\xi(t') - x\right] + \cdots \\ &+ \int_t^{t+\tau} \mathrm{d}t'\, g\left(x, t'\right) \Gamma(t') + \int_t^{t+\tau} \mathrm{d}t' \frac{\partial g\left(x, t'\right)}{\partial x} \left[\xi(t') - x\right] \Gamma(t') + \cdots \end{aligned} \tag{2.46}$$

For $[\xi(t')-x]$ in the integrand we iterate eq. (2.46), producing

$$\begin{aligned}\xi(t+\tau)-x=&\int_t^{t+\tau}\mathrm{d}t'\,h\left(x,t'\right)+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial h\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,h(x,t'')\right]\\&+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial h\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,g\left(x,t''\right)\Gamma(t'')\right]\\&+\int_t^{t+\tau}\mathrm{d}t'\,g\left(x,t'\right)\Gamma(t')+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial g\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,h(x,t'')\Gamma(t'')\right]\\&+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial g\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,g(x,t'')\Gamma(t'')\Gamma(t')\right]+\cdots\end{aligned}\tag{2.47}$$

By repeated iterations, only Langevin forces and the known functions $h(x,t)$ and $g(x,t)$ and their derivatives appear on the right-hand side of eq. (2.47). If we now take the average of eq. (2.47), and use eq. (2.3), we get

$$\begin{aligned}\langle\xi(t+\tau)-x\rangle=&\int_t^{t+\tau}\mathrm{d}t'\,h(x,t')+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial h\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,h(x,t'')\right]+\cdots\\&+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial g\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,g(x,t'')\delta(t''-t')\right]+\cdots\\=&\int_t^{t+\tau}\mathrm{d}t'\,h(x,t')+\int_t^{t+\tau}\mathrm{d}t'\left[\frac{\partial h\left(x,t'\right)}{\partial x}\int_t^{t'}\mathrm{d}t''\,h(x,t'')\right]+\cdots\\&+\int_t^{t+\tau}\mathrm{d}t'\,\frac{\partial g\left(x,t'\right)}{\partial x}g(x,t')+\cdots\end{aligned}\tag{2.48}$$

We plug in the above expression into eq. (2.32), we get

$$\begin{aligned}\mathcal{D}^{(1)}(x,t)&=\lim_{\tau\to0}\frac{1}{\tau}\langle\xi(t+\tau)-x\rangle\Big|_{\xi(t)=x}\\&=h(x,t)+\frac{\partial g\left(x,t\right)}{\partial x}g(x,t).\end{aligned}\tag{2.49}$$

Using similar procedures we obtain the second coefficient

$$\begin{aligned}\mathcal{D}^{(2)}(x,t)&=\frac{1}{2}\lim_{\tau\to0}\frac{1}{\tau}\int_t^{t+\tau}\mathrm{d}t''\left[\int_t^{t+\tau}\mathrm{d}t'\,g(x,t')g(x,t'')\delta(t'-t'')\right]\\&=g^2(x,t).\end{aligned}\tag{2.50}$$

One can see that in addition to the deterministic drift $h(x,t)$, $\mathcal{D}^{(1)}$ contains a term which is called the **spurious drift** or the **noise-induced drift**

$$\mathcal{D}^{(1)}_{\text{spurious}}=\frac{\partial g(x,t)}{\partial x}g(x,t)=\frac{1}{2}\frac{\partial}{\partial x}\mathcal{D}^{(2)}(x,t).\tag{2.51}$$

It stems from the fact that when $\Gamma(t)$ varies, $\xi(t)$ changes with it, therefore $\langle g(\xi(t),t)\Gamma(t)\rangle$ is no longer zero.

2.2.4 Method of solution

We now want to discuss methods for solving the one-variable Fokker-Planck equation with time-independent drift and diffusion coefficients, assuming $\mathcal{D}^{(2)}(x) > 0$. The following method is similar to the ones found in Refs. [47] [152]. There are several other ways of solving the Fokker-Planck equation, for example, one method using the eigenfunctions of the Fokker-Planck operator can be found in Ref. [236].

Normalization

By a suitable transformation $x' \equiv y = y(x)$ the x-dependent diffusion coefficient can be transformed to an arbitrary constant $\mathcal{D} > 0$. For the one-variable case this transformation according to

$$\mathcal{D}'^{(2)} \equiv \mathcal{D} = \left(\frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\right)^2 \mathcal{D}^{(2)}(x). \tag{2.52}$$

Thus this transformation is given by

$$y = y(x) = \int_{x_0}^{x} \mathrm{d}\xi \sqrt{\frac{\mathcal{D}}{\mathcal{D}^{(2)}(\xi)}}\,. \tag{2.53}$$

The transformation drift coefficient then takes the form

$$\begin{aligned}\mathcal{D}'^{(1)}(y) &= \frac{\mathrm{d}y}{\mathrm{d}x}\mathcal{D}^{(1)}(x) + \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}\mathcal{D}^{(2)}(x) \\ &= \sqrt{\frac{\mathcal{D}}{\mathcal{D}^{(2)}(x)}}\left[\mathcal{D}^{(1)}(x) - \frac{1}{2}\frac{\mathrm{d}\mathcal{D}^{(2)}(x)}{\mathrm{d}x}\right]\end{aligned} \tag{2.54}$$

and the transformed Fokker-Planck equation reads ($\mathcal{D}$ = constant)

$$\frac{\partial W'(y,t)}{\partial t} = \left[-\frac{\partial}{\partial y}\mathcal{D}'^{(1)}(y) + \mathcal{D}\frac{\partial^2}{\partial y^2}\right] W'(y,t), \tag{2.55}$$

where W' is given by

$$W' = \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^{-1} W = \sqrt{\frac{\mathcal{D}^{(2)}(x)}{\mathcal{D}}} W. \tag{2.56}$$

Without loss of generality we may thus treat the equation with constant diffusion coefficient, i.e.,

$$\frac{\partial W}{\partial t} = \left[-\frac{\partial}{\partial x}f'(x) + \mathcal{D}\frac{\partial^2}{\partial x^2}\right] W = -\frac{\partial}{\partial x}J(x,t), \tag{2.57}$$

where $f(x) = -\int^x \mathrm{d}x'\, \mathcal{D}^{(1)}(x')$ is the potential, and J is the probability current.

Because $\mathcal{D}$ is arbitrary, we may use $\mathcal{D} = 1$. This normalization is, however, not very convenient if the low-noise limit $\mathcal{D} \to 0$ is considered and we therefore retained the constant $\mathcal{D}$. The transformation in eq. (2.53) can also be done in the Langevin equation.

Stationary solution

For stationary solutions the probability current in eq. (2.57) must be constant. Thus, if the probability current vanishes at some x the current must be identically equal to 0. Then for $S = 0$

$$\mathcal{D}^{(1)}(x)W_{\mathrm{st}}(x) = \frac{\mathcal{D}^{(1)}(x)}{\mathcal{D}^{(2)}(x)}\mathcal{D}^{(2)}(x)W_{\mathrm{st}}(x) = \frac{\partial}{\partial x}\mathcal{D}^{(2)}(x)W_{\mathrm{st}}(x). \tag{2.58}$$

We can integrate eq. (2.58), yielding

$$W_{\mathrm{st}}(x) = \frac{N_0}{\mathcal{D}^{(2)}(x)} \exp\left\{\int^x \mathrm{d}x' \, \frac{\mathcal{D}^{(1)}(x')}{\mathcal{D}^{(2)}(x')}\right\} = N_0 \mathrm{e}^{-\Phi(x)}, \tag{2.59}$$

where N_0 is the integration constant, which has to be chosen such that W_{st} is normalized. In eq. (2.59) we introduced the potential

$$\Phi(x) = \log \mathcal{D}^{(2)}(x) - \int^x \mathrm{d}x' \, \frac{\mathcal{D}^{(1)}(x')}{\mathcal{D}^{(2)}(x')}. \tag{2.60}$$

We may put $\Phi(x) = f(x)/\mathcal{D}$ for eq. (2.57) since the potential $\Phi(x)$ is defined only up to an additive constant and therefore the $\log \mathcal{D}^{(2)}(x)$ term may be omitted. Introducing this potential the probability current may be written in the form

$$J(x,t) = \mathcal{D}^{(2)}(x)\mathrm{e}^{-\Phi(x)}\frac{\partial}{\partial x}\left[\mathrm{e}^{\Phi(x)}W(x,t)\right]. \tag{2.61}$$

In the stationary case, where J is constant, we thus have for arbitrary J

$$W_{\mathrm{st}} = N\mathrm{e}^{-\Phi(x)} - J\mathrm{e}^{-\Phi(x)} \int^x \mathrm{d}x' \frac{\mathrm{e}^{\Phi(x')}}{\mathcal{D}^{(2)}(x')}. \tag{2.62}$$

One of the integration constants in eq. (2.62) is determined by the normalization condition

$$\int \mathrm{d}x \, W_{\mathrm{st}} = 1, \tag{2.63}$$

the other constant must be determined from the boundary conditions. For problems where x extends to $\pm\infty$, we require that the integral in eq. (2.63) exists. In that case, W and J must vanish at $\pm\infty$ (natural boundary conditions) and therefore $J = 0$ for every x.

An important question is whether every initial distribution finally converges to the stationary distribution. For some restrictions of the drift and diffusion coefficients and of the boundary conditions one can prove that any two solutions of the Fokker-Planck equation

agree for large times. Thus if a stationary solution exist, every solution must finally converge to that solution.

2.3 Bistability, metastability, and escape problems

This section is devoted to the asymptotic study of systems which can exist in at least two stable states. The mean first-passage time, i.e., the average time that elapses until a stochastic process switch spontaneously from prescribed domains, is an important characteristic of the system [231]. It is widely used for describing various dynamic features such as relative stability problems, exit problems, activation rates, relaxation times from unstable states, lifetimes of metastable states. These properties underlie a wide range of practical applications, such as neuron firing [104], gating transition of ion channels [112], and magnetic relaxation of systems composed of ferromagnetic nanoparticles [62]. The calculations in the subsequent sections closely follows the work of Hendrik Anthony Kramers (1894–1952) [123] [167]. An alternative derivation can be found in Ref. [259].

2.3.1 Surmounting a potential barrier

We first apply the Fokker-Planck equation to calculate escape rates over a potential barrier. We consider once again the model where the probability density $p(x,t)$ of a particle obeys the Fokker-Planck equation

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial \left[\Phi'(x)p(x,t)\right]}{\partial x} + \sigma^2 \frac{\partial^2 p(x,t)}{\partial x^2}. \tag{2.64}$$

The shape of $\Phi(x)$ is as shown in Fig. 2-1. There are two minima at a and c and in between, a local maximum. We suppose that motion is on an infinite range, which means the stationary solution is

$$p_{\mathrm{st}}(x) = N\mathrm{e}^{-\frac{\Phi(x)}{\sigma^2}} \tag{2.65}$$

and it is this that demonstrates the bistability. Corresponding to a, c and b are two maxima, and a central minimum as plotted in Fig. 2-1. There is a relatively high probability of being on the left or the right of b, but not near b.

(a) Behavior for $\sigma^2 = 0$ In this case, $x(t)$ obeys the differential equation

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\Phi'(x), \\ x(0) &= x_0. \end{aligned} \tag{2.66}$$

Since

$$\frac{\mathrm{d}\Phi(x)}{\mathrm{d}t} = \Phi'(x)\frac{\mathrm{d}x}{\mathrm{d}t} = -\left[\Phi'(x)\right]^2 < 0, \tag{2.67}$$

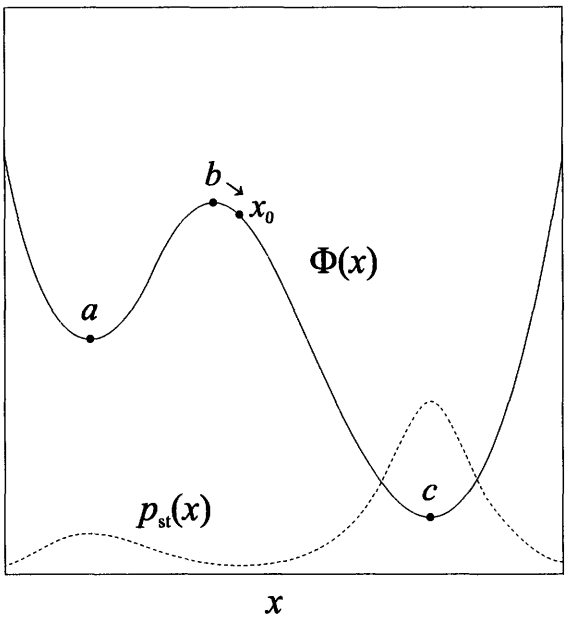


Figure 2-1: Plot of $p_{st}(x)$ and $\Phi(x)$ for a double well potential. In the potential $\Phi(x)$, there are two minima at $x = a$ and $x = c$ and in between, a local maximum. The stationary solution $p_{st}(x)$ demonstrates the bistability: there is a relatively high probability of being on the left or the right of b, but not near b.

$x(t)$ always moves in such a way as to minimize $\Phi(x)$, and stops only when $\Phi'(x)$ is zero. Thus, depending on whether x_0 is greater than or less than b, the particle ends up at c or a, respectively. The motion follows the arrows on the figure.

Once the particle is at a or c it stays there. If it starts exactly at b, it also stays there, though the slightest perturbation drives it to a or c. Thus, b is an unstable stationary point and a and c are stable. There is no question of *relative* stability of a and c.

(b) Behavior if σ^2 is very small With the addition of noise, the situation changes. The stationary state can be approximated asymptotically as follows. Assuming $\Phi(x)$ is everywhere sufficiently smooth, we can write

$$\Phi(x) \simeq \begin{cases} \Phi(a) + \dfrac{\Phi''(a)(x-a)^2}{2}, & |x-a| \text{ small;} \\ \Phi(c) + \dfrac{\Phi''(c)(x-c)^2}{2}, & |x-c| \text{ small.} \end{cases} \tag{2.68}$$

If σ^2 is very small, then we may approximate

$$p_{st}(x) \simeq \begin{cases} N\mathrm{e}^{-\frac{\Phi(a)}{\sigma^2} - \frac{1}{2}\frac{\Phi''(a)(x-a)^2}{\sigma^2}}, & |x-a| \text{ small;} \\ N\mathrm{e}^{-\frac{\Phi(c)}{\sigma^2} - \frac{1}{2}\frac{\Phi''(c)(x-c)^2}{\sigma^2}}, & |x-c| \text{ small;} \\ 0, & \text{elsewhere.} \end{cases} \tag{2.69}$$

with

$$\frac{1}{N} \simeq \mathrm{e}^{-\frac{\Phi(a)}{\sigma^2}} \sqrt{\frac{2\pi\sigma^2}{\Phi''(a)}} + \mathrm{e}^{-\frac{\Phi(c)}{\sigma^2}} \sqrt{\frac{2\pi\sigma^2}{\Phi''(c)}}. \tag{2.70}$$

Suppose, as drawn in the figure, $\Phi(a) > \Phi(c)$. Then for small enough σ^2, the second term is overwhelmingly larger than the first and $1/N$ can be approximated by the second term alone. Substituting into eq. (2.69) we find

$$p_{\mathrm{st}}(x) \simeq \begin{cases} \sqrt{\dfrac{\Phi''(c)}{2\pi\sigma^2}}\, \mathrm{e}^{-\frac{1}{2}\frac{\Phi''(c)(x-c)^2}{\sigma^2}}, & |x-a| \sim \sigma; \\ 0, & \text{otherwise.} \end{cases} \tag{2.71}$$

This means that in the limit of very small σ^2, the deterministic stationary state at which $\Phi(x)$ has an absolute minimum is the more stable state in the sense that in the stochastic stationary state, $p_{\mathrm{st}}(x)$ is very small everywhere except in its immediate vicinity. This means that in the case of a and c, fluctuations take over and the motion is given approximately by linearizing the stochastic differential equation around a or c. Around b, the linearized stochastic differential equation is unstable. The particle, therefore, follows the **Ornstein-Uhlenbeck process** [282] until it leaves the immediate neighborhood of $x = b$, at which stage the asymptotic expansion in σ takes over. The process that can occur is that of escape over the central barrier. The noise $\mathrm{d}W(t)$ can cause the particle to climb the barrier at b and reach the other side. This involves times of order $\exp(-\mathrm{const}/\sigma^2)$, which do not contribute to an asymptotic expansion in powers of σ^2 since they go to zero faster than any power of σ^2 as $\sigma^2 \to 0$.

2.3.2 First passage times

What is the mean escape time from the left hand well? By this we mean, what is the mean first passage time from a to x, where x is in the vicinity of b? The solution of this problem can be achieved by use of the **backward Fokker-Planck equation**. For a more in depth review of the backward-equation method, see Ref. [172].

(a) Two absorbing barriers Let the particle be initially at x at time $t = 0$ and let us ask how long it remains in the interval (a, b) which is assumed to contain x. We set absorbing barriers at $x = a$ and $x = b$ so that the particle is removed when it reaches a or b. Hence, if the particle is still in the interval (a, b), it has never left that interval.

Under these conditions, the probability that at time t the particle is still in (a, b) is

$$\int_a^b \mathrm{d}x'\, p(x', t|x, 0) \equiv G(x, t). \tag{2.72}$$

Let the time that the particle leaves (a, b) be $\mathcal{T}$. Then we can rewrite eq. (2.72) as

$$\Pr(\mathcal{T} \geq t) = \int_a^b \mathrm{d}x'\, p(x', t|x, 0), \tag{2.73}$$

which means that $G(x,t) = \Pr(\mathcal{T} \geq t)$. Since the system is time homogeneous, we can write

$$p(x', t|x, 0) = p(x', 0|x, -t) \tag{2.74}$$

and the backward Fokker-Planck equation can be written as

$$\frac{\partial}{\partial t} p(x', t|x, 0) = A(x) \frac{\partial}{\partial x} p(x', t|x, 0) + \frac{1}{2} B(x) \frac{\partial^2}{\partial x^2} p(x', t|x, 0) \tag{2.75}$$

and hence, $G(x,t)$ obeys the equation

$$\frac{\partial}{\partial t} G(x,t) = A(x) \frac{\partial}{\partial x} G(x,t) + \frac{1}{2} B(x) \frac{\partial^2}{\partial x^2} G(x,t) \tag{2.76}$$

The boundary conditions are clearly that

$$p(x', 0|x, 0) = \delta(x - x') \tag{2.77}$$

and hence,

$$G(x, 0) = \begin{cases} 1, & a \leq x \leq b; \\ 0, & \text{elsewhere,} \end{cases} \tag{2.78}$$

and if $x = a$ or $x = b$, the particle is absorbed immediately, so $\Pr(\mathcal{T} \geq t) = 0$ when $x = a$ or $x = b$, i.e., $G(a,t) = G(b,t) = 0$.

Since $G(x,t)$ is the probability that $\mathcal{T} \geq t$, the mean of any function of $\mathcal{T}$ is

$$\langle f(\mathcal{T}) \rangle = -\int_0^\infty \mathrm{d}G(x,t)\, f(\mathcal{T}). \tag{2.79}$$

Thus, the **mean first passage time** $\mathcal{T}(x) = \langle \mathcal{T} \rangle$ is given by

$$\begin{aligned} \mathcal{T}(x) &= -\int_0^\infty \mathrm{d}t\, t \frac{\partial G(x,t)}{\partial t} \\ &= -tG(x,t)\big|_o^\infty + \int_0^\infty \mathrm{d}t\, G(x,t) \\ &= \int_0^\infty \mathrm{d}t\, G(x,t). \end{aligned} \tag{2.80}$$

We can derive a simple ordinary differential equation for $\mathcal{T}(x)$ by using eq. (2.80) and

integrating eq. (2.76) over $(0, \infty)$. Noting that

$$\int_0^\infty \mathrm{d}t \, \frac{\partial G(x,t)}{\partial t} = G(x,\infty) - G(x,0) = -1, \tag{2.81}$$

we derive

$$A(x)\frac{\partial \mathcal{T}(x)}{\partial t} + \frac{1}{2}B(x)\frac{\partial^2 \mathcal{T}(x)}{\partial x^2} = -1 \tag{2.82}$$

with the boundary condition $\mathcal{T}(a) = \mathcal{T}(b) = 0$.

Eq. (2.82) can be solved directly by integration. The solution, after some manipulation, can be written as

$$\mathcal{T}(x) = \frac{2\left[\left(\int_a^x \frac{\mathrm{d}y}{\psi(y)}\right)\int_x^b \frac{\mathrm{d}y'}{\psi(y')}\int_a^{y'} \frac{\mathrm{d}z\,\psi(z)}{B(z)} - \left(\int_x^b \frac{\mathrm{d}y}{\psi(y)}\right)\int_a^x \frac{\mathrm{d}y'}{\psi(y')}\int_a^{y'} \frac{\mathrm{d}z\,\psi(z)}{B(z)}\right]}{\int_a^b \frac{\mathrm{d}y}{\psi(y)}} \tag{2.83}$$

with

$$\psi(x) = \exp\left\{\int_a^x \mathrm{d}x' \, \frac{2A(x')}{B(x')}\right\}. \tag{2.84}$$

(b) One absorbing barrier We consider motion still in the interval (a, b) but suppose the barrier at $x = a$ to be reflecting. The boundary conditions then become

$$\left.\frac{\partial G(x,t)}{\partial x}\right|_{x=a} = G(x,t)\big|_{x=b} = 0. \tag{2.85}$$

We solve eq. (2.82) with the corresponding boundary conditions and obtain

$$\mathcal{T}(x) = 2\int_x^b \frac{\mathrm{d}y}{\psi(y)}\int_a^y \frac{\mathrm{d}z\psi(z)}{B(z)} \qquad \begin{array}{l} a \text{ reflecting} \\ b \text{ absorbing} \\ a < b. \end{array} \tag{2.86}$$

Similarly, one finds

$$\mathcal{T}(x) = 2\int_a^x \frac{\mathrm{d}y}{\psi(y)}\int_y^b \frac{\mathrm{d}z\psi(z)}{B(z)} \qquad \begin{array}{l} b \text{ reflecting} \\ a \text{ absorbing} \\ a < b. \end{array} \tag{2.87}$$

(c) Escape over a potential barrier We use eq. (2.86) with the substitutions

$$b \to x_0, \quad a \to -\infty, \quad x \to a, \tag{2.88}$$

so that

$$\mathcal{T}(a \to x_0) = \frac{1}{\sigma^2}\int_a^{x_0} \mathrm{d}y \, \mathrm{e}^{\frac{\Phi(y)}{\sigma^2}} \int_{-\infty}^y \mathrm{d}z \, \mathrm{e}^{-\frac{\Phi(z)}{\sigma^2}}. \tag{2.89}$$

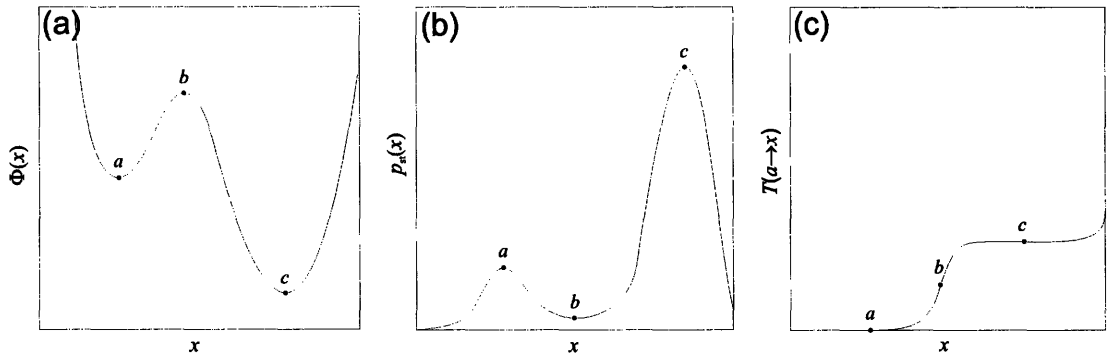


Figure 2-2: (a)Double well potential $\Phi(x)$; (b) Stationary distribution $p_{\rm st}(x)$; (c) Mean first passage time from a to x, $\mathcal{T}(a \to x)$. The plot of $\mathcal{T}(a \to x)$ shows that the mean first passage time is quite small for x in the left well and quite large for x in the right well.

If the central maximum of $\Phi(x)$ is large and σ^2 is small, then $\exp[\Phi(y)/\sigma^2]$ is sharply peaked at $x = b$, while $\exp[-\Phi(z)/\sigma^2]$ is very small near $z = b$. Therefore, $\int_{-\infty}^{y} \mathrm{d}z\ \exp[-\Phi(z)/\sigma^2]$ is a very slowing varying function of y near $y = b$. This means that the value of the integral $\int_{-\infty}^{y} \mathrm{d}z\ \exp[-\Phi(z)/\sigma^2]$ will be approximately constant for those values of y which yield a value of $\exp[\Phi(y)/\sigma^2]$ which is significantly different from zero. Hence, in the inner integral, we can set $y = b$ and remove the resulting constant factor from inside the integral with respect to y. Hence, we can approximate eq. (2.89) by

$$\mathcal{T}(a \to x_0) \simeq \left\{ \frac{1}{\sigma^2} \int_{-\infty}^{b} \mathrm{d}y\, \mathrm{e}^{-\frac{\Phi(z)}{\sigma^2}} \right\} \int_{z}^{x_0} \mathrm{d}y\, \mathrm{e}^{\frac{\Phi(y)}{\sigma^2}}. \tag{2.90}$$

Notice that by the definition of $p_{\rm st}$ in eq. (2.65), we can say that

$$\int_{-\infty}^{b} \mathrm{d}y\, \mathrm{e}^{\frac{-\Phi(z)}{\sigma^2}} = \frac{n_a}{N}, \tag{2.91}$$

where n_a is the probability that the particle is to the left of b when the system is stationary. A plot of $\mathcal{T}(a \to x_0)$ against x_0 is shown in Fig. 2-2 and shows that the mean first passage time to x_0 is quite small for x_0 in the left well and quite large for x_0 in the right well. This means that the particle, in going over the barrier to the right well, takes most of the time in actually surmounting the barrier. It is quite meaningful to talk of the *escape time* as that for the particle, initially at a, to reach a point near c, since this time is quite sensitive to the exact location of the initial and final points. We can evaluate this by further assuming that near b we can write

$$\Phi(x \approx b) \simeq \Phi(b) - \frac{1}{2}\left(\frac{x-b}{\delta}\right)^2 \tag{2.92}$$

and near a

$$\Phi(x \approx a) \simeq \Phi(a) + \frac{1}{2}\left(\frac{x-a}{\alpha}\right)^2. \tag{2.93}$$

The constant factor in eq. (2.90) is evaluated as

$$\begin{aligned}\int_{-\infty}^{b} \mathrm{d}z\, \mathrm{e}^{-\frac{\Phi(z)}{\sigma^2}} &\sim \int_{-\infty}^{\infty} \mathrm{d}z\, \mathrm{e}^{-\frac{\Phi(a)}{\sigma^2}-\frac{(z-a)^2}{2\sigma^2\alpha^2}} \\ &\sim \alpha\sqrt{2\pi\sigma^2}\, \mathrm{e}^{-\frac{\Phi(a)}{\sigma^2}} \end{aligned} \tag{2.94}$$

and the inner factor becomes, on assuming x_0 is far to the right of the central point b,

$$\begin{aligned}\int_{-\infty}^{b} \mathrm{d}y\, \mathrm{e}^{-\frac{\Phi(y)}{\sigma^2}} &\sim \int_{-\infty}^{\infty} \mathrm{d}y\, \mathrm{e}^{-\frac{\Phi(b)}{\sigma^2}-\frac{(y-b)^2}{2\sigma^2\delta^2}} \\ &\sim \delta\sqrt{2\pi\sigma^2}\, \mathrm{e}^{-\frac{\Phi(b)}{\sigma^2}}. \end{aligned}$$

Putting both of these in eq. (2.90), we get

$$\mathcal{T}(a \to x_0) \simeq 2\alpha\delta\, \mathrm{e}^{\frac{\Phi(b)-\Phi(a)}{\sigma^2}}. \tag{2.95}$$

This is the classical **Van't Hoff-Arrhenius law** of chemical reaction theory [13]. In a chemical reaction, we can model the reaction by introducing a coordinate such that $x = a$ is species A and $x = c$ is species C. The reaction is modeled by the above diffusion process and the two distinct chemical species are separated by the potential barrier at b. In the chemical reaction, statistical mechanics gives the value $\sigma^2 = k_{\mathrm{B}}T$, where k_{B} is Boltzmann's constant and T is the absolute temperature. We see that the most important dependence on temperature comes from the exponential factor

$$\exp\left(\frac{E_{\mathrm{A}}}{k_{\mathrm{B}}T}\right) \tag{2.96}$$

and predicts a very characteristic dependence on temperature. Intuitively, the exponential factor represents the probability that the energy will exceed that of the barrier when the system is in thermal equilibrium. Those molecules that reach this energy then react. Aside from chemical kinetic theory, the Arrhenius factor was also found in the switching rates in magnetic spin-torque systems, where T is the effective spin temperature [10] [31].

2.3.3 Probability of exit through a particular end of the interval

What is the probability that the particle, initially at x in (a, b), exits through a, and what is the mean exit time? The total probability that at time t the particle has exited through a is given by the time integral of the probability current at a

$$\begin{aligned} g_a(x,t) &= -\int_0^t \mathrm{d}t'\, J(a,t'|x,0) \\ &= \int_0^t \mathrm{d}t' \left\{ -A(a)p(a,t'|x,0) + \frac{1}{2}\frac{\partial}{\partial a}\left[B(a)p(a,t'|x,0)\right] \right\}, \end{aligned} \tag{2.97}$$

the negative sign is chosen since we need the current pointing to the left, and

$$g_b(x,t) = \int_0^t \mathrm{d}t' \left\{ -A(b)p(b,t'|x,0) + \frac{1}{2}\frac{\partial}{\partial a}\left[B(a)p(b,t'|x,0)\right] \right\}. \tag{2.98}$$

These quantities give the probabilities that the particle has exited through $x = a$ or b at time t, respectively. The probability that it existed before time t is

$$\Pr(\mathcal{T}_a < t) = \frac{g_a(x,t)}{g_a(x,\infty)}. \tag{2.99}$$

We now find an expression for $g_a(x,t)$. We use the fact that $p(a,t|x,0)$ satisfies a backward Fokker-Planck equation. Thus,

$$\begin{aligned} A(x)\frac{\partial g_a(x,t)}{\partial x} + \frac{1}{2}B(x)\frac{\partial^2 g_a(x,t)}{\partial x^2} &= -\int_0^t \mathrm{d}t' \, \frac{\partial J(a,t'|x,0)}{\partial t'} \\ &= -J(a,t'|x,0) \\ &= \frac{\partial g_a(x,t)}{\partial t}. \end{aligned} \tag{2.100}$$

The mean exit time, given that exit is through a, is

$$\mathcal{T}(a,x) = \int_0^\infty \mathrm{d}t \, t\frac{\partial \Pr(\mathcal{T}_a < t)}{\partial t} = -\int_0^\infty \mathrm{d}t \, \frac{g_a(x,t)}{g_a(x,\infty)}. \tag{2.101}$$

Simply integrating eq. (2.100) with respect to t, we get

$$A(x)\frac{\partial\left[\pi_a(x)\mathcal{T}(a,x)\right]}{\partial x} + \frac{1}{2}B(x)\frac{\partial^2\left[\pi_a(x)\mathcal{T}(a,x)\right]}{\partial x^2} = -\pi_a(x), \tag{2.102}$$

where we define

$$\pi_a(x) = \text{probability of exit through } a = g_a(x,\infty). \tag{2.103}$$

The boundary conditions on eq. (2.102) are

$$\pi_a(a)\mathcal{T}(a,a) = \pi_a(b)\mathcal{T}(a,b) = 0. \tag{2.104}$$

In the first of these, clearly $\mathcal{T}(a,a)$ is zero (the time to reach a from a is zero) and in the second, $\pi_a(b)$ is zero (the probability of existing through a, starting from b, is zero).

By letting $t \to \infty$ in eq. (2.100), we note that eventually the particle is no longer in (a,b). Hence, the right-hand side tends to zero and we get

$$A(x)\frac{\partial \pi_a(x)}{\partial x} + \frac{1}{2}B(x)\frac{\partial^2 \pi_a(x)}{\partial x^2} = 0, \tag{2.105}$$

with the boundary conditions

$$\pi_a(a) = 1,$$
$$\pi_a(b) = 0. \tag{2.106}$$

The solution of eq. (2.105) subject to this boundary condition and the conservation condition

$$\pi_a(x) + \pi_b(x) = 1 \tag{2.107}$$

is

$$\pi_a(x) = \frac{\int_x^b \mathrm{d}y\, \psi(y)}{\int_a^b \mathrm{d}y\, \psi(y)}, \quad \pi_b(x) = \frac{\int_a^x \mathrm{d}y\, \psi(y)}{\int_a^b \mathrm{d}y\, \psi(y)} \tag{2.108}$$

with $\psi(x)$ as defined in eq. (2.84). These formulae find application in the problem of relaxation of a distribution initially concentrated at an unstable stationary point, which we will present in the next section.

2.3.4 Splitting probability

Suppose we put the particle at x_0: what is the probability that it reaches a before c, or c before a? This can be related to the problem of exit through a particular end of an interval, studies in Sec. 2.3.3. We put absorbing barriers at $x = a$ and $x = c$, and using the results of that section, find π_a and π_c, the **splitting probabilities** for reaching a or c first. These are:

$$\pi_a(x_0) = \frac{\int_{x_0}^c \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1}}{\int_a^c \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1}}, \quad \pi_c(x_0) = \frac{\int_a^{x_0} \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1}}{\int_a^c \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1}}. \tag{2.109}$$

The splitting probabilities of π_a and π_c can be viewed more generally as simply the probability that the particle, started at x_0, will fall into the left or right-hand well, since the particle, having reached a, will remain on that side of the well for a time of the same order as the mean exit time to b.

We nose consider two possible asymptotic forms as $\sigma^2 \to 0$.

(a) x_0 a finite distance from b We first evaluate

$$\int_a^c \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1} = \frac{1}{N} \int_a^c \mathrm{d}x\, \mathrm{e}^{\frac{\Phi(x)}{\sigma^2}}. \tag{2.110}$$

This is dominated by the behavior at $x \sim b$. An asymptotic evaluation is correctly obtained by setting

$$\Phi(x) \simeq \Phi(b) - \frac{1}{2}\frac{|\Phi''(b)|}{(b-x)^2}. \tag{2.111}$$

As $\sigma^2 \to 0$, the limits at $x = a, c$ effectively recede to $\pm\infty$ and we find

$$\int_a^c \mathrm{d}x\, p_{\mathrm{st}}(x)^{-1} = \frac{1}{N}\sqrt{\frac{2\pi\sigma^2}{|\Phi''(b)|}}\, \mathrm{e}^{\frac{\Phi(b)}{\sigma^2}}. \tag{2.112}$$

Now suppose $x_0 < b$. Then $\int_a^{x_0} \mathrm{d}x\, p(x)^{-1}$ can be evaluated by the substitution $y = \Phi(x)$ with an inverse $x = W(y)$ and is asymptotically

$$\begin{aligned}\frac{1}{N}\int_{-\infty}^{\Phi(x_0)} \mathrm{d}y\, \mathrm{e}^{\frac{y}{\sigma^2}} W'(y) &\sim \frac{1}{N}\sigma^2\, \mathrm{e}^{\frac{\Phi(x_0)}{\sigma^2}} W'[\Phi(x_0)] \\ &= \frac{1}{N}\frac{\sigma^2\, \mathrm{e}^{\frac{\Phi(x_0)}{\sigma^2}}}{\Phi'(x_0)}.\end{aligned} \tag{2.113}$$

Thus,

$$\pi_c \sim \frac{1}{\Phi'(x_0)}\sqrt{\frac{|\Phi''(b)|\sigma^2}{2\pi}}\, \mathrm{e}^{\frac{\Phi(x_0)-\Phi(b)}{\sigma^2}} \tag{2.114}$$

and

$$\pi_a = 1 - \pi_c. \tag{2.115}$$

We see here that the splitting probability depends only on x_0 and b. thus, the probability of reaching c in this limit is governed entirely by the probability of jumping the barrier at b. The points at a and c are effectively infinitely distant.

(b) x_0 infinitesimally distant from b Suppose

$$x_0 = b - y_0\sigma. \tag{2.116}$$

In this case, we can make the approximation in both integrals. Defining

$$\mathrm{erf}(x) = \sqrt{\frac{\pi}{2}}\int_0^x \mathrm{d}t\, \mathrm{e}^{-t^2} \tag{2.117}$$

we find

$$\begin{aligned}\pi_c = 1 - \pi_a &\sim \frac{1}{2}\left\{1 - \mathrm{erf}\left[y_0\sqrt{|\Phi''(b)|}\right]\right\} \\ &= \frac{1}{2}\left\{1 - \mathrm{erf}\left[(b-x_0)\sqrt{\frac{|\Phi''(b)|}{\sigma^2}}\right]\right\}.\end{aligned} \tag{2.118}$$

Eq. (2.118) is the result that would be obtained if we replaced $\Phi(x)$ by its quadratic approximation eq. (2.111) over the whole range.

(c) Comparison of two regions The two regions give different results, and we find that a simple linearization of the SDE [which is what replacing $\Phi(x)$ by a quadratic approximation amounts to] gives the correct result only in the limit of large σ^2 and in a region of order of magnitude $\sqrt{\sigma^2}$ around the maximum b.

2.3.5 Decay from an unstable state

The mean time for a particle placed at a point on a potential to reach one well or the other is an object capable of being measured experimentally. If we use eq. (2.14) for the process, then the mean time to reach a or c from b can be computed exactly using the formulae of Sec. 2.3.3. The mean time to reach a from b is the solution of

$$-\Phi'(x)\frac{\partial\left[\pi_a(x)\mathcal{T}(a,x)\right]}{\partial x}+\sigma^2\frac{\partial^2[\pi_a(x)\mathcal{T}(a,x)]}{\partial x^2}=\pi_a(x) \tag{2.119}$$

with the boundary conditions

$$\pi_a(a)\mathcal{T}(a,a)=\pi_a(c)\mathcal{T}(a,c)=0 \tag{2.120}$$

and $\pi_a(x)$ is given by eq. (2.109).

The solution to eq. (2.119) is quite straightforward to obtain by direct integration, but it is rather cumbersome. The solution technique is exactly the same as that used for eq. (2.83) and the result is similar. Even the case covered by eq. (2.83) where we do not distinguish between exit and the right or at the left, is very complicated.

For the record, however, we set down that

$$\mathcal{T}(a,x)=\frac{\pi_c(x)\displaystyle\int_x^c \mathrm{d}x' p_{\mathrm{st}}(x')^{-1}\int_a^{x'}\mathrm{d}z\pi_a(z)p_{\mathrm{st}}(z)-\pi_a(x)\int_a^x \mathrm{d}x' p_{\mathrm{st}}(x')^{-1}\int_a^{x'}\mathrm{d}z\pi_a(z)p_{\mathrm{st}}(z)}{\sigma^2\pi_a(x)}, \tag{2.121}$$

where one considers that $\pi_a(x)$ is given by eq. (2.109) and $p_{\mathrm{st}}(z)$ by eq. (2.69). It can be seen that even for the simplest possible situation, namely,

$$\Phi(x)=-\frac{1}{2}kx^2, \tag{2.122}$$

the expression is almost impossible to comprehend. An asymptotic treatment is perhaps required. Fortunately, in the cases where $p_{\mathrm{st}}(z)$ is sharply peaked at a and c with a sharp minimum at b, the problem reduces essentially to the problem of relaxation of a or to c with a reflecting barrier at b. To see this note that

1. the explicit solution for $\pi_a(x)$ in eq. (2.109) means

$$\pi_a(x) \simeq \begin{cases} 1, & x < b; \\ \frac{1}{2}, & x = b; \\ 0, & x > b \end{cases} \tag{2.123}$$

and the transition takes place over a distance $\sim \sqrt{\sigma^2}$, the width of the peak in $p_{\rm st}(x)^{-1}$.

2. In the integrals with integrand $\pi_a(z)p_{\rm st}(z)$, we distinguish two cases.

$x' > b$: in this case the estimates allow us to say

$$\int_a^{x'} \mathrm{d}z\, \pi_a(z)p_{\rm st}(z) \simeq \frac{n_a}{2}, \tag{2.124}$$

where $n_a = \int_{-\infty}^{b} \mathrm{d}z p_{\rm st}(z)$ and represents the probability of being in the left-hand well.

However, when $x' < a$, we may still approximate $\pi_a(z)$ by 1, so we get

$$\int_a^{x'} \mathrm{d}z\, \pi_a(z)p_{\rm st}(z) \simeq \int_a^{x'} \mathrm{d}z\, p_{\rm st}(z) = \frac{n_a}{2} - \int_{x'}^{b} \mathrm{d}z\, p_{\rm st}(z). \tag{2.125}$$

Substituting these estimates into eq. (2.121) we obtain

$$\mathcal{T}(a,b) \simeq \mathcal{D}^{-1} \int_a^b \mathrm{d}x'\, p_{\rm st}(x)^{-1} \int_{x'}^{b} \mathrm{d}z\, p_{\rm st}(z), \tag{2.126}$$

which is the exact mean exit time from b to a with a reflecting barrier at b. Similarly,

$$\mathcal{T}(c,b) \simeq \mathcal{D}^{-1} \int_b^c \mathrm{d}x'\, p_{\rm st}(x)^{-1} \int_{b}^{x'} \mathrm{d}z\, p_{\rm st}(z). \tag{2.127}$$

2.4 Summary

In this chapter we presented a first-principle derivation of the Fokker-Planck equation, from which we obtained a complete probabilistic description of the motion of a Brownian particle in a fluctuating environment. We further investigated the escape time problem; we have shown that in the thermodynamical equilibrium limit, one can recover a host of interesting results, such as the Arrhenius law. Other situations where noise can have a *constructive* role in the behavior of a dynamical system is the phenomenon of **stochastic resonance**, in which the response of the system to an external signal under the presence of fluctuations can be enhanced by tuning the noise intensity to a particular value [43] [87] [195]. One would be interested in estimating the mean first passage time for the response to reach a target value. The dependence of the mean exit time on the noise intensity for metastable and unstable systems was revealed to have resonance character [134].

Chapter 3

Stochastic chaos

> The equations of dynamics completely express the laws of the historical method as applied to matter, but the application of these equations implies a perfect knowledge of all the data. But the smallest portion of matter which we can subject to experiment consists of millions of molecules, not one of which ever becomes individually sensible to us. We cannot, therefore, ascertain the actual motion of any one of these molecules, so that we are obliged to abandon the strict historical method, and to adopt the statistical method of dealing with large groups of molecules.
>
> —James Clerk Maxwell [190]

To model a physical system one route is to include, as realistically as possible, all the complicated many-body interactions. The system is deterministic, and so in principle could be completely described. Then we try to obtain a quantitative prediction of the behavior by solving the resultant equations of motion. The challenge we face here is that usually exact solution to the deterministic system is difficult, if not impossible to obtained. In practice, we make approximation or assumptions to simplify the model. For example, we can use stochastic processes as an approximate description of a deterministic system, which has unknown initial conditions, and which may have the high sensitivity to initial conditions. The theory of probability and stochastic processes was developed to describe such complicated irregular phenomena. For example, a coin toss, the quintessential example of a random system, could in principle be described adequately, if inconveniently, by Newton's laws. Adopting a stochastic model to describe a system became viewed as a matter of convenience: absent a low-dimensional deterministic model of the system, a stochastic model at least might provide the long-run frequency tendencies of the system states. Randomness is chaos, but calling it "noise" is a strategic modeling decision amounting to an admission that the system fluctuations remain high dimensional and unexplained.

The other extreme is to write down the simplest possible model that still includes the essential physics. For example, we can ask what configurations are important to the thermodynamic properties of a statistical mechanical system, and then include all configurations that contribution to the statistical average in finite systems. In some cases, we add random noise to our simple model and this noise represents nothing more than our lack of knowledge of the system structure or inadequacy of the identification procedure. And then we hope that it is tractable to analytic or precise numerical solution. The aim here is often to study universal behavior or to gain a qualitative understanding of the physics governing the behavior of the system.

In the previous chapter, we have started with the former approach, we solved the Brownian motion problem using stochastic methods. We take the latter approach in this chapter, namely, to study the same global effects in simpler models. Despite the apparent simplicity of the models, they show a rich mathematical structure. Moreover, and perhaps surprisingly at first sight, they do provide valid and useful representation of real world systems.

3.1 Thermodynamic and chaos

There are some ideas of transport and diffusion theory which have developed over the past decade [65], and which are based on the notions of chaos theory. These methods are of interest to us here since they directly relate macroscopic transport quantities, to microscopic dynamical quantities such as Lyapunov exponents and Kolmogorov-Sinai entropies.

3.1.1 Initial-value sensitivity and the Lyapunov exponents

Even when the mathematical properties of deterministic chaos gradually become better understood, the technicalities of chaos definition vary among textbooks, the property of sensitivity to initial conditions was a nearly universal ingredient. Typical additional requirements of chaos were that the attractor had to be bounded and densely embedded with periodic solutions. Consider a general one-dimensional map defined by the formula $y = f(x)$. Iterating this formula, we get, for $k = 1, 2, 3, \ldots$

$$x_{k+1} = f(x_k). \tag{3.1}$$

Also, we consider the simple and ideal case of two solutions with infinitesimally different initial conditions $x_0 = x$ and $x_0 = x + \mathrm{d}x$ so that $\mathrm{d}x_0 = \mathrm{d}x$. It follows from calculus that the infinitesimal difference between the two solutions after one iterate equals $\mathrm{d}x_1 = f'(x_0)\,\mathrm{d}x$, where f' denotes the first derivative of f. After m iterates, the chain rule of differentiation implies that the infinitesimal difference between the two solutions becomes

$$\left|\frac{\mathrm{d}x_m}{\mathrm{d}x_o}\right| = \prod_{k=0}^{m-1}\left|\frac{\mathrm{d}x_{k+1}}{\mathrm{d}x_k}\right| = \prod_{k=0}^{m-1}|f'(x_k)| \tag{3.2}$$

The product $f'(x_0)f'(x_1)\cdots f'(x_{m-1})$ is the "amplifying" factor. It is more convenient to study the average of the logarithm of the magnitude of this factor which becomes

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \log|f'(x_k)| = \langle \log f'(x) \rangle . \tag{3.3}$$

This amounts to computing the temporal average of the logarithm of the modulus of the derivative of the map over the time evolution. If the deterministic system is ergodic, then, asympotically, this time average over one realization is the same as the ensemble average.

For the logistic map, it can be calculated that $\lambda = \log 2$. In other words, two logistic trajectories with infinitesimally different initial conditions will diverge, on the average, at the rate of 2^m after m steps, with the averaging over a large number of iterates. The trajectory becomes unpredictable very quickly. The constant λ is called the **Lyapunov exponent**. If it is positive, then the difference in the initial condition is amplified exponentially at the rate of $\exp(\lambda m)$ after m iterates, in which case we say that the dynamical system is **sensitive to initial conditions**. On the other hand, a negative Lyapunov exponent means that the effect of the initial condition is short lived and the dynamical system is not sensitive to initial conditions. As a working definition for **deterministic chaos**, a *dynamical system is said to be chaotic if and only if it admits a positive Lyapunov exponent.* The above discussion can be generalized to the case of higher dimensional maps. In particular, λ will be generalized to an array of Lyapunov exponents.

It is almost impossible to give a precise mathematical definition of deterministic chaos that encapsulates all the term implies in the diverse literature. We have seen, however, that it is widely accepted that sensitivity to initial conditions is a typical feature of a deterministic chaotic system. Similarly, there is no universally accepted definition of chaos for a stochastic system. Some authors define stochastic chaos base on ideas motivated by deterministic chaos, for example, initial-value sensitivities. We will defer this discussion to later sections. And as for how to characterize a stochastic system, we will hold off this discussion until the next chapter.

3.1.2 Kolmogorov-Sinai entropies

Now we give a brief discussion of the **Kolmogorov-Sinai (KS) entropy**, which is a characteristic property of those deterministic dynamical systems with "randomness" properties. The KS entropy is essential for formulating the escape-rate expressions for transport coefficients. As explained by Eckmann and Ruelle [70], dynamical systems with positive Lyapunov exponents produce information. The KS entropy is an information theoretic measure to characterize the production of this information. Consider a phase space, Ω, which we can partition into a non-trivial collection of non-overlapping sets $\{W_i\}$ such that

$$\{W_i : \Omega = \cup_i W_i,\ W_i \cap W_j = \emptyset \text{ for } i \neq j,\ \mu(W_i) > 0\} . \tag{3.4}$$

Kolmogorov [160] and Sinai [261] define the entropy of a partition in terms of a normalized invariant measure on the phase space as

$$H(\{W_i\}) = -\sum_i \mu(W_i)\log(\mu(W_i)), \tag{3.5}$$

with normalization $\sum_i \mu(W_i) = 1$. When the partition is the trivial partition, $W = \Omega$, then $H = 0$. To get a number that indicates how much information is gained per step, define

$$h = \lim_{n\to\infty} \frac{1}{n} H_n. \tag{3.6}$$

This quantity can be interpreted as a measure for the rate at which information is produced. Numerically, it is found that a more useful, but equivalent, definition of h is given by

$$h = \lim_{n\to\infty} [H_{n+1} - H_n]. \tag{3.7}$$

These definitions depend on the choice of the partition, but the Kolmogorov-Sinai entropy, h_{KS}, is defined as the supremum of the above expression over all possible finite partitions of the space at $t = 0$:

$$h_{\mathrm{KS}} = \sup_{W_i} h. \tag{3.8}$$

In bounded systems, the KS entropy per unit time is known to be given by **Pesin's theorem** [221] as the sum of positive Lyapunov exponents:

$$h_{\mathrm{KS}} = \sum_i \lambda_i^+. \tag{3.9}$$

3.1.3 The escape-rate formalism

One of the most interesting developments in the theory of irreversible processes is a connection between the dynamical properties of open systems and their hydrodynamic or transport properties. Consider a Brownian particle diffusing in a fluid inside an interval with absorbing boundaries. The motion of the particle is really deterministic and can be described, microscopically, but Newton's laws of motion. If we were to describe the motion macroscopically, we would solve the diffusion equation (the Fokker-Planck equation) for the probability density of the Brownian particle. The probability of finding the particle inside the interval is an exponentially decreasing function of time with decay rate depending on the mass and size of the particle, the viscosity of the fluid, and the geometry of the boundaries.

Transport processes like diffusion can be conceived in terms of chaotic scattering, which leads to the so-called escape-rate formulas relating the diffusion coefficient to the Lyapunov exponents and the KS entropy or the fractal dimensions, characterizing chaos in dynamical

systems [99]:

$$\gamma = \sum_{\lambda_i > 0} \lambda_i(F_{\mathrm{R}}) - h_{\mathrm{KS}}(F_{\mathrm{R}}). \tag{3.10}$$

The general derivations of the escape-rate formula are too involved to be presented here in any detail, there is a nice simple argument that captures the general features of the proofs. Let us imagine that we want to determine the rate of information production by trajectories. The quantity we want is then $\exp[h_{\mathrm{KS}}(F_{\mathrm{R}})t]$. By the definition of the KS entropy, we would expect that this information is generated, exponentially in time according to $\exp\{t\sum_{\lambda_i>0}\lambda_i(F_{\mathrm{R}})\}$, by the stretching of phase-space regions. But at the same time, it is reduced, exponentially in time according to $\exp(-\gamma t)$, by the fact that most of the trajectories escape from the system at the absorbing boundaries. We would then write

$$\mathrm{e}^{h_{\mathrm{KS}}(F_{\mathrm{R}})t} = \mathrm{e}^{-\gamma t} \exp\left\{ t \sum_{\lambda_i > 0} \lambda_i(F_{\mathrm{R}}) \right\}. \tag{3.11}$$

The escape-rate formula follows immediately. More rigorous derivations of this formula are given in Refs. [23] [98] [115] [147].

3.1.4 Thermodynamic formalism of chaos

The results for the escape rate, the Lyapunov exponent, and the KS entropy can be combined conveniently into one expression from which all of the others can be derived. This expression is called a **dynamic partition function** [16] [97]. We start with the set of intervals that survive the first n iterations of the map, the dynamic partition function is defined by

$$Z_n(\beta) = \sum_{i=1}^{2^n} (\ell_i)^\beta, \tag{3.12}$$

where the ℓ_i are the lengths of the intervals, and β is a parameter. We can make this expression look like a canonical partition function if we associate an "energy", ϵ_i, with each interval ℓ_i, by $\ell_i = \exp(-\epsilon_i)$. In the same way as one defines a free energy from the equilibrium canonical partition function, we define a quantity called the **topological pressure** from the dynamical partition function. We write $Z_n(\beta)$ as

$$Z_n(\beta) = \mathrm{e}^{n\Xi(\beta)}. \tag{3.13}$$

For large n we can formally compare n with the number of particles N in a canonical ensemble. Then $\Xi(\beta)$ can be thought of as a kind of negative "free energy per particle times β" and Ξ is called the topological pressure, defined properly by the limit

$$\Xi(\beta) = \lim_{n\to\infty} \frac{1}{n} \log(Z_n(\beta)). \tag{3.14}$$

Now, all of the quantities discussed earlier can be expressed in terms of $\Xi(\beta)$. Specifically, the Lyapunov exponent can be related to $\Xi(\beta)$ as:

$$\lambda_i(F_{\mathrm{R}}) = \frac{\mathrm{d}\Xi(\beta)}{\mathrm{d}\beta}|_{\beta=1}. \tag{3.15}$$

The escape rate, γ, is

$$\gamma = -\Xi(1). \tag{3.16}$$

The Kolmogorov-Sinai entropy can be written as

$$h_{\mathrm{KS}}(F_{\mathrm{R}}) = -\Xi'(1) + \Xi(1). \tag{3.17}$$

The thermodynamic formalism, created by Sinai, Ruelle, and Bowen, provides very deep insights into the structure of chaos and into the connections of dynamical systems theory with statistical mechanics.

3.2 Typical routes to chaos

Deterministic chaos is the behavior of a physical system whose time dependence is purely deterministic. As a prerequisite for the occurrence of deterministic chaos in physical systems, the mathematical description of the time evolution has to be given by nonlinear equations of motion. They need to be very sensitive to small variations of the initial conditions. Then originally close phase space trajectories separate exponentially fast. Usually, chaotic solutions are not found for the complete region of control parameters. The system approaches the chaotic regime in a very characteristic manner. The mechanism of the transition to chaos is of fundamental importance for understanding the phenomenon of chaotic behavior. In the following we briefly describe three main routes to chaos which can be observed in nonlinear oscillators. These are Feigenbaum's period-doubling route [74], the intermittency route, and the Ruelle-Takens-Newhouse route. A detailed discussion of these mechanisms would go far beyond the scope of this thesis.

In the context of optical bistability, deterministic chaos has been predicted first by Ikeda [141] [142] for a ring resonator with a very long round-trip time. The obtained solutions have been investigated in some detail, e.g., in Ref. [86]. They could be linked to a multimodal instability of the Maxwell-Bloch equations for this system. In the meantime, the Ikeda instability has been demonstrated experimentally in several devices [77] [300]. The study of the influence of noise in the route to chaos has direct practical implications, for instance, it is useful in the interpretation of the dynamical behavior of semiconductor laser diode [42] and optically injected semiconductor laser [93]; the unexpected dynamics in epidemic outbreaks [21]; and the development of bursting neuron models [39].

3.2.1 Period doubling

In 1974, Princeton biologist Robert May published the first analysis of chaos in the logistic equation, and introduced the term "chaos" for the first time [192]. Since then, the logistic map became a classic example of the know routes to chaos: period-doubling, intermittency and introduction of a new fundamental frequency [74] [193]. The bifurcation diagram of the noise-free logistic map is shown in Fig. 3-1 along with the Lyapunov exponents as the parameter μ varies. The bifurcation of the randomly perturbed logistic map has been widely studied, see for example, Ref. [227].

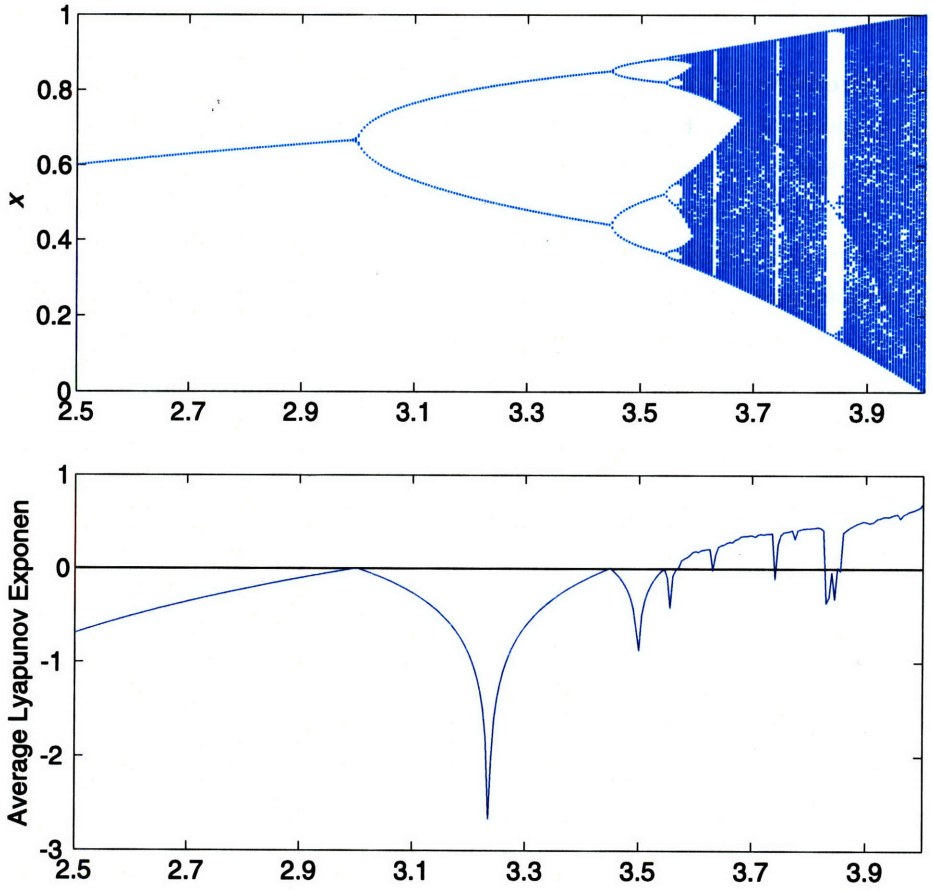


Figure 3-1: Top: Bifurcation diagram and, Bottom: Lyapunov exponent plot for the noise-free logistic map as the μ varies from 2.5 to 4. We can see the period-doubling cascades of periodic states for $1 < \mu < \mu_\infty$, where the Lyapunov exponent is negative. At bifurcation points $\mu_1 = 3$, $\mu_2 \approx 3.4$, the Lyapunov exponent is zero. For chaotic domain, for $\mu_\infty < \mu < 4$, the Lyapunov exponent is mostly positive. We can also observe windows in the chaotic domain where the Lyapunov exponent becomes negative.

As shown in the bifurcation diagram, the system undergoes a sequence of period-doubling cascades until above a critical parameter value a chaotic regime begins. For $1 < \mu < \mu_\infty$, where the Lyapunov exponent is negative. At bifurcation points $\mu_1 = 3$,

$\mu_2 \approx 3.4$, the Lyapunov exponent is zero. In 1978, Feigenbaum proved that the "period doubling" pattern found in the logistic equation is not unique, and is in fact found in every nonlinear system [74]; this had the effect of showing that chaos was *universal*, and not limited to some small class of equations. For chaotic domain, the region $\mu_\infty < \mu < 4$, the Lyapunov exponent is mostly positive. We can also observe windows in the chaotic domain where the Lyapunov exponent becomes negative. These "windows" have been seen not only in one-dimensional models but also, e.g., in the two-dimensional discrete Hénon attractor [75] and even in the three-dimensional Lorenz attractor [256].

3.2.2 Intermittency

By **intermittency** we mean the occurrence of fluctuations that alternate "randomly" between long periods of regular behavior and relatively short irregular bursts. It has been found that the density of chaotic bursts increases with an external control parameter, which shows that intermittency presents a continuous route from periodic to chaotic behavior. Fig. 3-2 shows a typical orbit for $\mu = 3.8282$. Part of the orbit looks like a stable 3-cycle, but this is spooky since the 3-cycle no longer exists! We're seeing the *ghost* of the 3-cycle. The theory of intermittency has bee established in a pioneering study of Pomeau and Manneville [224]. They connected intermittency with tangent bifurcations and distinguished three types of intermittency basing on the mechanism of the loss of stability.

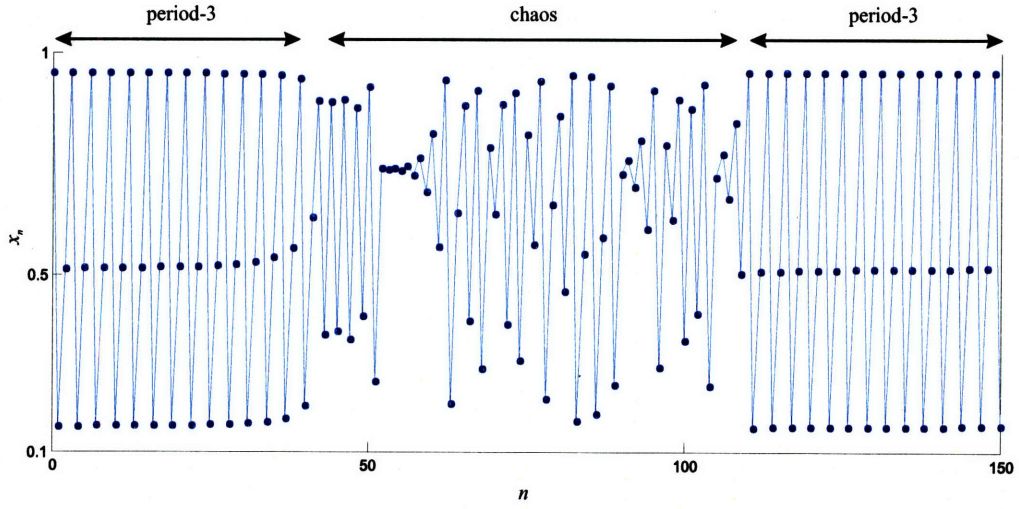


Figure 3-2: Typical orbit for logistic map ($\mu = 3.8282$) with intermittent chaos.

3.2.3 Break of torus

The third route to chaos is connected with Hopf bifurcation, which generates a limit cycle starting from a fixed point. The Landau-Hopf theory of turbulence was until the mid 1970s the accepted theory of how a fluid flow becomes turbulent [169]. The theory states that as

a fluid flows faster, it develops more and more Fourier modes, as shown in Fig. 3-3. In this hypothesis the chaotic state is approached by an infinite sequence of Hopf bifurcations.

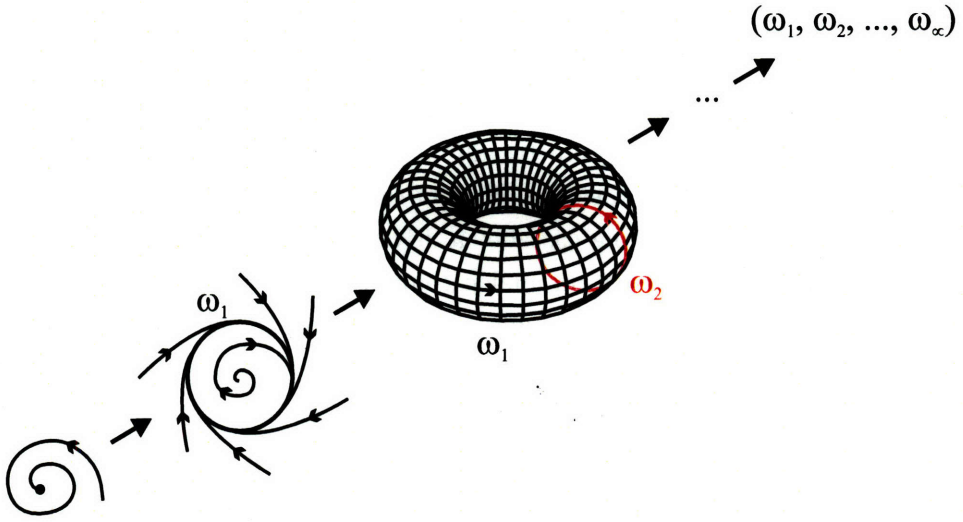


Figure 3-3: Landau's model of turbulence.

The modification of this model which better describes the route observed in many experiments was proposed by Newhouse, Ruelle and Takens [242], in which a much shorter route was suggested. They showed that after three Hopf bifurcations regular motion becomes highly unstable in favor of motion on a strange attractor as illustrated in Fig. 3-4. Ruelle and Takens also coined the term "strange attractor" in their paper.

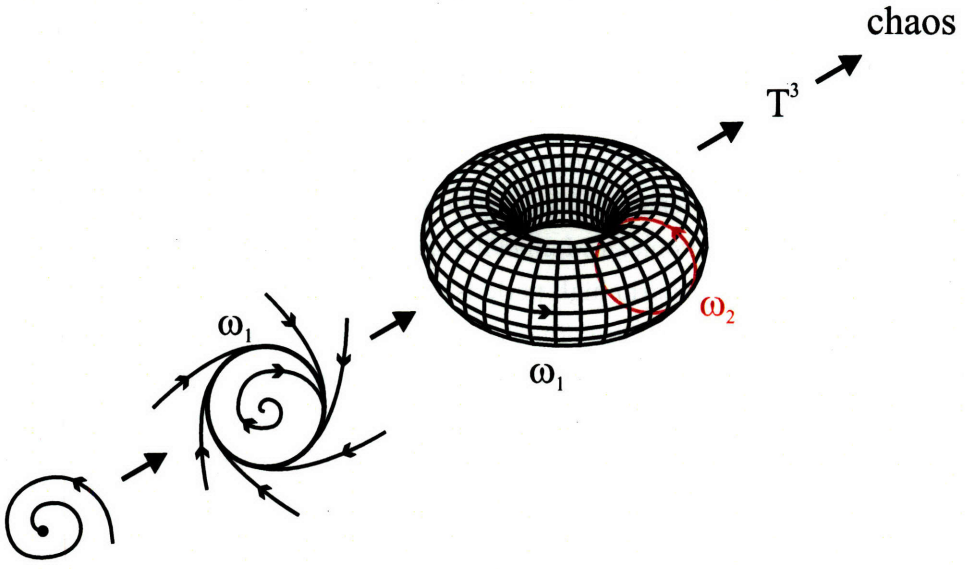


Figure 3-4: Ruelle-Takens-Newhouse model of turbulence.

Experimental results suggest that direct transitions from quasiperiodicity to high-dimensional chaos may be reached without the need of noise, but it's a particular type of global bifur-

cation. This type of transition is not related by hyperchaos [217].

3.3 Noisy route to chaos

In the absence of noise the period-doubling route to chaos consists of an infinite number of bifurcations which accumulate geometrically at some critical parameter value. Crutchfield and Huberman [56] first studied the effect of random noise on a differential equation known to possess a period-doubling cascade: a driven, damped oscillator in a quartic potential. That study, carried out on an analog computer, showed that the noise truncates the observable sequence after a finite number of bifurcations. To describe this phenomenon, the authors introduced the notion of a "bifurcation gap" to describe the manner in which noise washed out the fine structure of the deterministic bifurcation diagram. One way to understanding the gap is to average the deterministic dynamical behavior over a window of parameter values—this picture leads to good quantitative results. Fig. 3-5 showed evidence that the size of the bifurcation gap on the logistic map scaled with the input noise level. The windows in the intermittent chaotic regime, present in the noise-free case, disappear. The bifurcation diagram is thus "smeared" horizontally and the narrow windows disappear. In contrast, measurement noise corresponds a vertical smearing in the bifurcation diagram rather than a horizontal one.

The literature reports that, of the three common scenarios, only the Ruelle-Takens scenario is not affect by small noise. This surprising result is demonstrated by Kifer [155] Kifer proves, roughly speaking, that for systems with an Axiom A attractor the stationary probability of the system perturbed by colored external noise converges weakly to the invariant measure of the attractor as the intensity of the noise goes to zero. The influence of noise on the Lorenz strange attractor has been studied numerically. The result indicate that noise has the intuitively expected effect, namely it lowers the threshold of the transition to turbulence. The road to chaos via period-doubling, the Feigenbaum scenario, and the road via intermittency, the Pomeau-Manneville scenario, are both influenced even by small noise. In the Feigenbaum scenario high periods are wiped out leading to a bifurcation gap. For a detailed study of the influence of noise on a system with a period-doubling sequence, namely the Verhulst or logistic model, see [194]. These authors also find that the behavior of the system in a deterministic periodic environment differs markedly from that in a noisy environment. For systems following the Feigenbaum scenario, scaling laws can be obtained for the influence of external noise at the onset of chaos [54] [258]. In systems which display intermittent behavior, the effect of noise was investigated by Mayer-Kress and Hakens [119], Eckmann et al. [70], and Hirsch et al. [134]. We can summarize this by saying that in the three commonly used scenarios for roads to chaos noise has either no influence or the expected effect of making the transition to turbulence easier. These scenarios offer therefore not even a qualitative understanding of our experimental findings presented below.

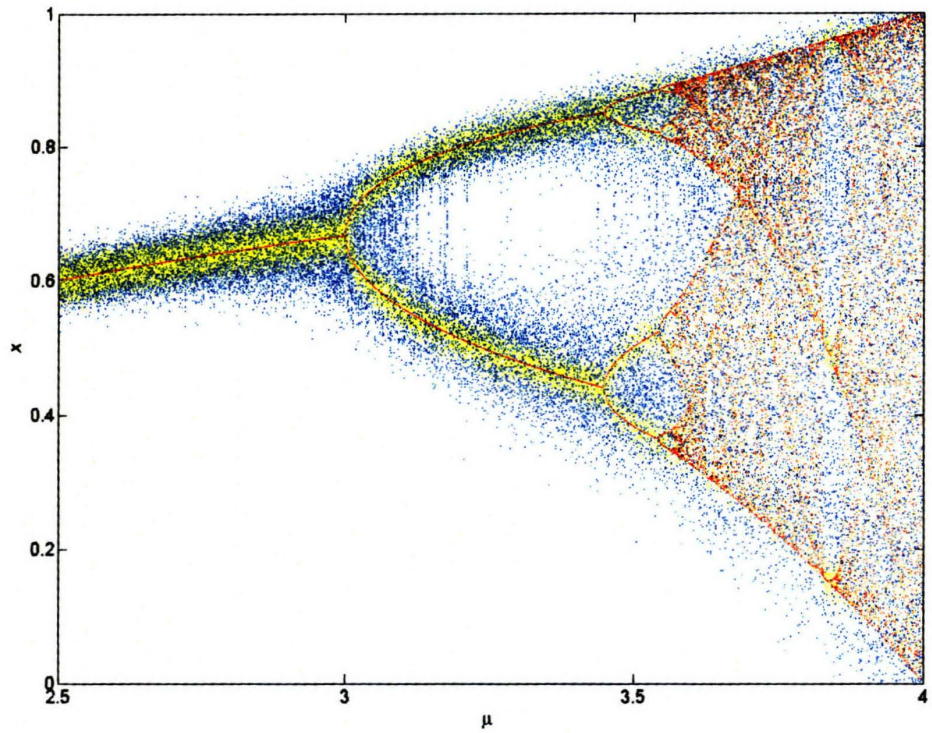


Figure 3-5: Bifurcation diagrams for logistic map. Bifurcation of the noise-free model is plotted in with red points; the same map in the presence of measurement noise is plotted with yellow points; and in the presence of dynamic noise is plotted with blue points. Clearly, there is no interaction between measurement noise and the nonlinear system dynamics and the combined effect is the simple superposition of the intrinsic nonlinear dynamics and the extrinsic noise. In contract, dynamic noise introduces far greater variability and complexity in the resulting dynamics.

3.3.1 Arrhenius formula and the logistic map

In this section we investigate the interference of noise and chaos in the simple logistic map to demonstrate the connection between chaotic dynamics and kinetic rate theory. We shall be concerned with a particle in a double-well potential and to understand its behavior in time and to try to make quantitative statements about the motion of the trajectories in the phase space that describe the particle. We consider the range of noise level where the system develops fluctuations of sufficient size to overcome the potential barrier barring it from equilibrium. As noted in Chapter 2, the studies of thermally activated escape from metastable states or the rate of transition of a particle over a potential barrier are known as the Kramers problem [167]. Our approach may be considered as a discrete-time variant of the celebrated Kramers escape problem.

Consider the logistic map, including process noise, given by the difference equation:

$$x_{k+1} = \mu x_k(1 - x_k) + \sigma\eta_{n+1} \quad \forall x_k \in [0, 1], \tag{3.18}$$

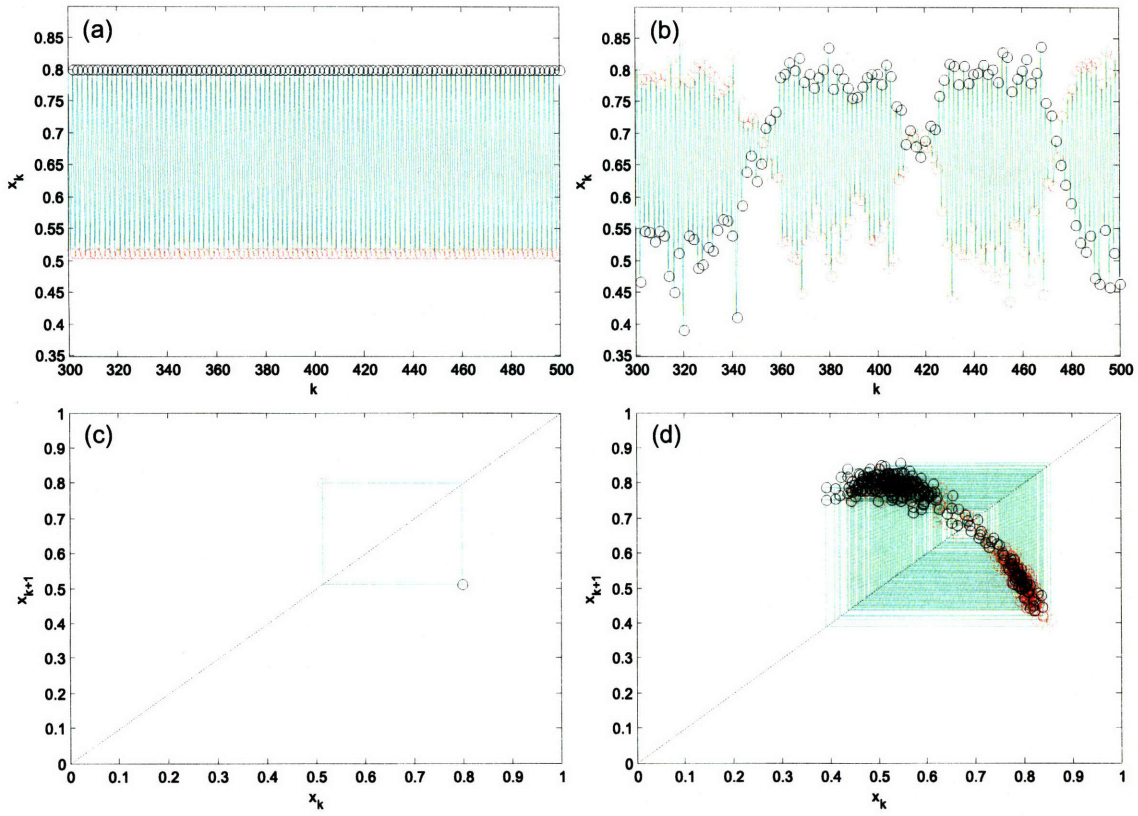


Figure 3-6: (a) Noise-free periodic logistic time series. (b) Periodic logistic time series with dynamic noise. (c) Noise-free return map corresponding to the time series in (a). (d) Return map correspond to the time series in (b).

where μ is the bifurcation parameter, σ is the standard deviation of the noisy and η is white, Gaussian-distributed random error with vanishing mean and unit variance. A noise-free time series in the period-2 regime ($\mu = 3.2$) is plotted in Fig. 3-6 (a), the corresponding return map is plotted in Fig. 3-6 (c). In the plots, we make the even samples with red circles and the odd samples with black circles. We proceed to add dynamic noise to the time series; we can begin to observe a "crossover" phenomenon of the red and black paths in Fig. 3-6 (b), specifically, in the vicinity of $k = 350$, 420, and 475. We can interpret these crossovers as the particle escaping from one well into the other (or in phase-space terms, escaping from one stable fixed point to the other). The rate of crossovers should depend on the intensity of the dynamic noise according to the Arrhenius factor we derived.

In order to verify this relationship, we wrote a simple program to count the number of crossing in the logistic time series with varying levels of dynamic noise. Fig. 3-7 shows the transition rate, k, plotted again the inverse of the standard deviation of the noise. In kinetic rate theory, such a plotting scheme is known as an **Arrhenius-plot**. The idea is that many thermally excited reactions are described by

$$y = y_o \mathrm{e}^{-E_\mathrm{A}/k_\mathrm{B}T}, \tag{3.19}$$

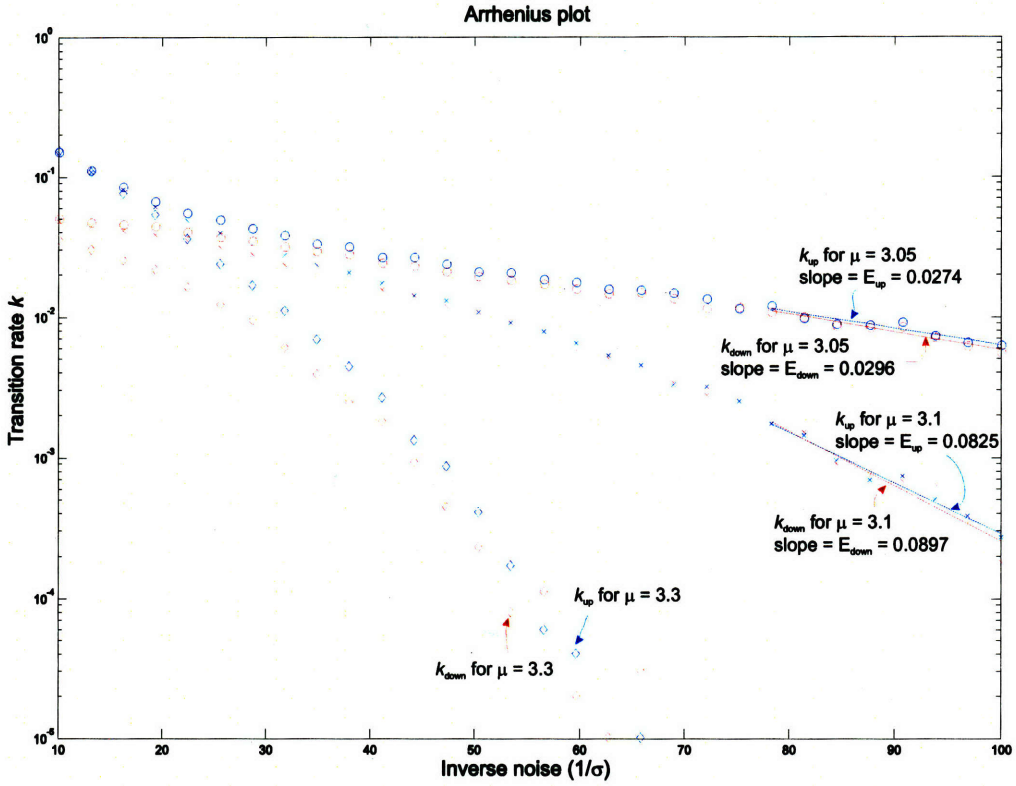


Figure 3-7: Arrhenius plot for the logistic map with bifurcation values 3.05, 3.1, and 3.3. The slope of the lines fitted at the low noise end indicate the value of the activation energy. Note that for $\mu = 3.3$, the curve breaks down at before the inverse noise level gets to 70. This is because the potential barrier is higher than the other two cases, therefore, at the low noise levels, zero crossover was observed and thus from a calculation standpoint, the potential barrier is essential infinitely high at those noise levels. Note that the probability of a reaction occurring in this case is still finite but extremely small.

where E_{A} is the **activation energy** (or **enthalpy**) of the process, and $k_{\mathrm{B}}T$ is the noise variance as given by statistical mechanics. An Arrhenius plot of this equation is simply a plot of $\log y$ over $1/T$ (or $1/k_{\mathrm{B}}T$). This produces a straight line:

$$\log y = \log y_o - \frac{E_{\mathrm{A}}}{k_{\mathrm{B}}T}. \tag{3.20}$$

The extrapolation of the straight line with the $\log y$-axis gives directly the value of the pre-exponential factor y_o, and the slope of the straight line gives the activation energy. An Arrhenius plot is extremely useful if data are determined experimentally. As seen in Fig. 3-7, the transition rates fit on a straight line in the low noise limit, indicating an Arrhenius relation. As we have seen in the previous chapter, the escape rate of such particle in a potential well interacting with a heat bath of equilibrium fluctuations has the form of an Arrhenius equation, but with a prefactor that depends on whether the interaction is moderately to strongly coupled (strong friction regime) or weakly coupled to the heat bath.

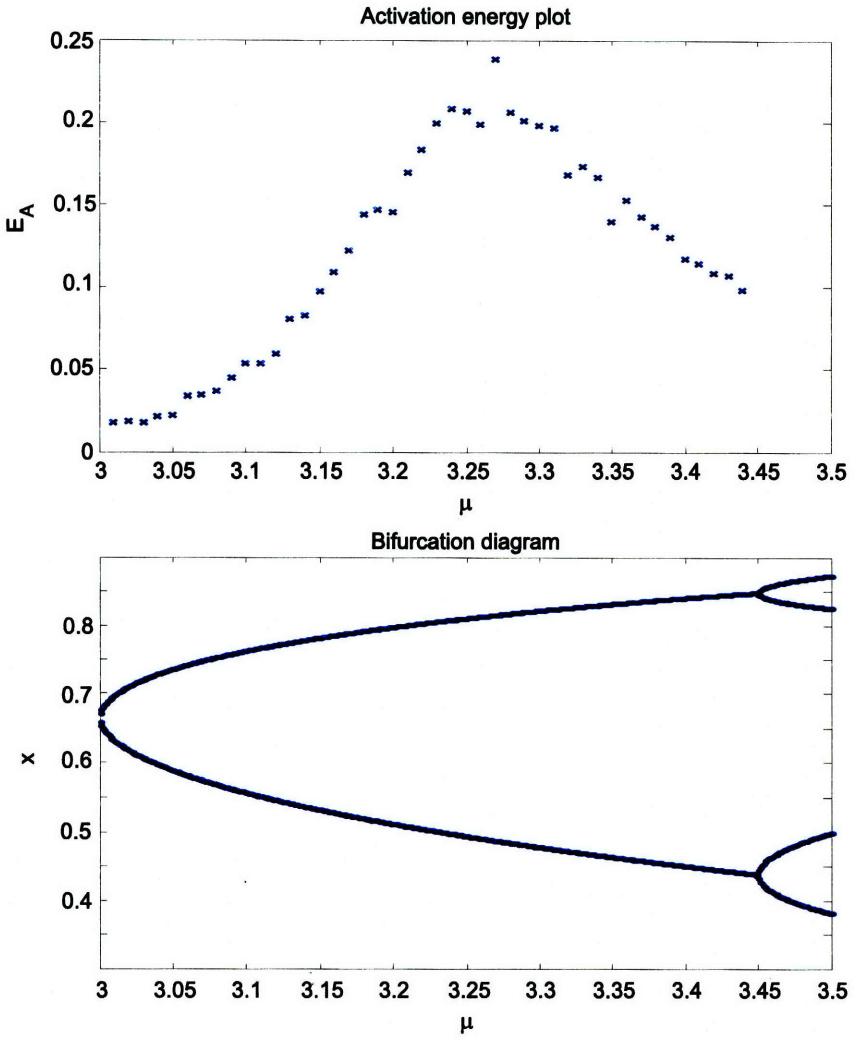


Figure 3-8: Top panel: Activation energy as a function of the bifurcation parameter μ. The activation energy peaks near $\mu = 3.28$, then falls for higher values of μ. Bottom panel: Bifurcation diagram of the logistic map, the distance between two branches is increasing as a function of μ. The height of the potential barrier is getting higher.

Note that the activation energy should increase as the bifurcation parameter increases. This is because the two stable fixed point moves farther apart for increasing values of the bifurcation parameter in the period-2 regime. The activation energy as a function of the bifurcation parameter is plotted in Fig. 3-8. Notice that the activation energy reaches a peak at around $\mu = 3.28$ before it falls in a linear fashion for the rest of the period-2 regime. We suspect that the decrease in activation energy is a shadow effect from higher periods.

Fogedby and Jensen [76], who also treated Kramers' problem on the noise-driven discrete logistic map, showed that the Arrhenius factor can be obtained from the dynamic partition function. In chemical reactors, Caroli et al. [40] indicated that for additive stochastic processes in a bistable potential, two main physical behaviors occur. The first is the exchange of population between the two wells, which leads to a switch from local to global

equilibrium. If the barrier is very high with respect to the diffusion coefficient, the Kramers' relaxation time becomes extremely long. From the point of view of a chemist, a very high barrier means that the corresponding chemical reaction does not, for practical purposes, take place. The second effect is known as the activation process or the **Suzuki enhancement of fluctuation** [273]. Starting from the top of the barrier, the time required to reach equilibrium is called the decay time, and it is after this that the activation process begins.

A one-dimensional bistable map in the presence of multiplicative Gaussian white noise was considered by Reimann and Talkner [235]. Recently, the additive noise case was explored by Fogedby and Jensen [76], who analyzed the structure of the noisy fixed points of the logistic map. An exact expression for the escape rate for asymptotically small noise strengths is derived consisting of an exponentially leading Arrhenius factor and a pre-exponential factor that shows a nontrivial dependence on the noise strength. A similar study was conducted by Reimann and Lootens [234] in which they considered the noise-assisted escape from a metastable point attractor across an unstable fixed point in one-dimensional dynamics, and found that the escape rate is dominated by an exponentially leading Arrhenius-like factor in the weak-noise limit.

In addition to finding the Arrhenius-like dependence of escape rate on the noise strength, what is more interesting is the observation that the dynamics of the stochastically perturbed system is highly irregular and resembles chaotic behavior. It is useful to analyze the return map in order to see if it verifies the existence of chaos. Return maps are plotted for the system with dynamic noise in Fig. 3-6 (d). The return map for the dynamic noise case appears more like the chaotic region of the noise-free map, in that the points spread out along all the points accessible in the logistic equation. It appears that the noise causes the system to undergo a bifurcation to chaos.

3.3.2 Nature of stochastic chaos

It is possible that there exists a different route to chaos: chaos induced by dynamic noise. This type of chaos may be termed **stochastic chaos**. Note that since measurement noise does not alter the system, and thus cannot be expected to induce chaotic behavior, it is useful to compare the effects of the two types of noise on the system in order to determine what part (if any) of the effect produced by the dynamic noise is indeed due to fundamentally altering the system dynamics, and what part is of stochastic origin. From a physical point of view it seems natural to ask how much of the qualitative behavior of these model systems survives in nature, where small perturbations are permanently present [56] [306]. So one can introduce noise, not in order to produce chaos but to check how stable a deterministic dynamical system is against small perturbations.

First, from the bifurcation diagram of the logistic map with dynamic noise in Fig. 3-5, vertical lines may be observed forming between the branches of the period-2 region. These lines are reminiscent of the intermittent chaotic region of the noise-free bifurcation diagram.

In other words, dynamic noise truly interacts with the original nonlinear system to produce a different system. Formation of the lines would seem to be related to the onset of chaos due to dynamic noise. It was found that for the logistic system the stable periodic orbit with periods not equal to 2^n disappear much faster than orbits of stable period 2^n and in a way which is qualitatively different [194]. The threshold for the size of the fluctuations for which, e.g., the Lyapunov exponents become positive is an order of magnitude smaller in the first case. The transition to chaos manifests itself in a characteristic change of the probability density, the Lyapunov exponents, and the power spectrum.

A stochastic model can also display behavior considerably different from its deterministic counterpart, the central tendency measures such as means or modes can bear little resemblance to the deterministic dynamics. Differences in the dynamic behaviors of deterministic models and their stochastic versions, termed "noise-induced transitions," were studied intensively in physics [138]. Deterministic chaos and stochastic chaos differ in their statistical resemblance. Specifically, Fig. 3-9 (c) shows the histogram of the deterministic trajectory of the chaotic logistic map ($\mu = 3.7$), which is noticeably different from Fig. 3-9 (d), which shows stochastic perturbed limit cycle trajectory ($\mu = 3.2$, $\sigma^2 = 0.05$). Their difference is clearly revealed by the non-straight-line-like appearance of the Q-Q plot shown in Fig. 3-9 (e). The Q-Q plot is obtained by first sorting, separately, the data of each times series in ascending order, and then plotting the scatter diagram of the two sets of ordered data. In the ideal case of infinite sample size, the scatter diagram will fall on a straight line if the two data sets are generated by the same marginal distribution.

The period-merging data was obtained by analyzing probability distributions of x values as (dynamic) noise was increased [194]. For the noise-free case one obtains delta functions for the probability distribution, with one delta function corresponding to each branch (in the periodic region). As the noise level is increased the probability distribution functions broaden, and eventually merge. This is illustrated in Fig. 3-10.

3.3.3 Existing definitions of stochastic chaos

An important component of the mathematical definition of deterministic chaos is sensitivity to initial conditions. Some authors define stochastic chaos base on ideas motivated by deterministic chaos, for example, the sensitivity measure has been extended to stochastic models [55] [194]. It was noted that for simple models such as the logistic map perturbed by noise, positive **stochastic Lyapunov exponents (SLE)** occurred at far lower values of the control parameter than for positive Lyapunov exponents. Such dynamic phenomena were referred to as "noise-induced chaos." Later, a formal mathematical definition of chaos for stochastic systems was given by Eckmann and Ruelle [70]. They considered a system to be chaotic if it exhibits sensitivity to initial condition and bounded fluctuations. However, it is possible for the SLE to be positive when the Lyapunov exponents of the underlying deterministic model is negative in the absence of noise, and vice versa [63]. This occurs

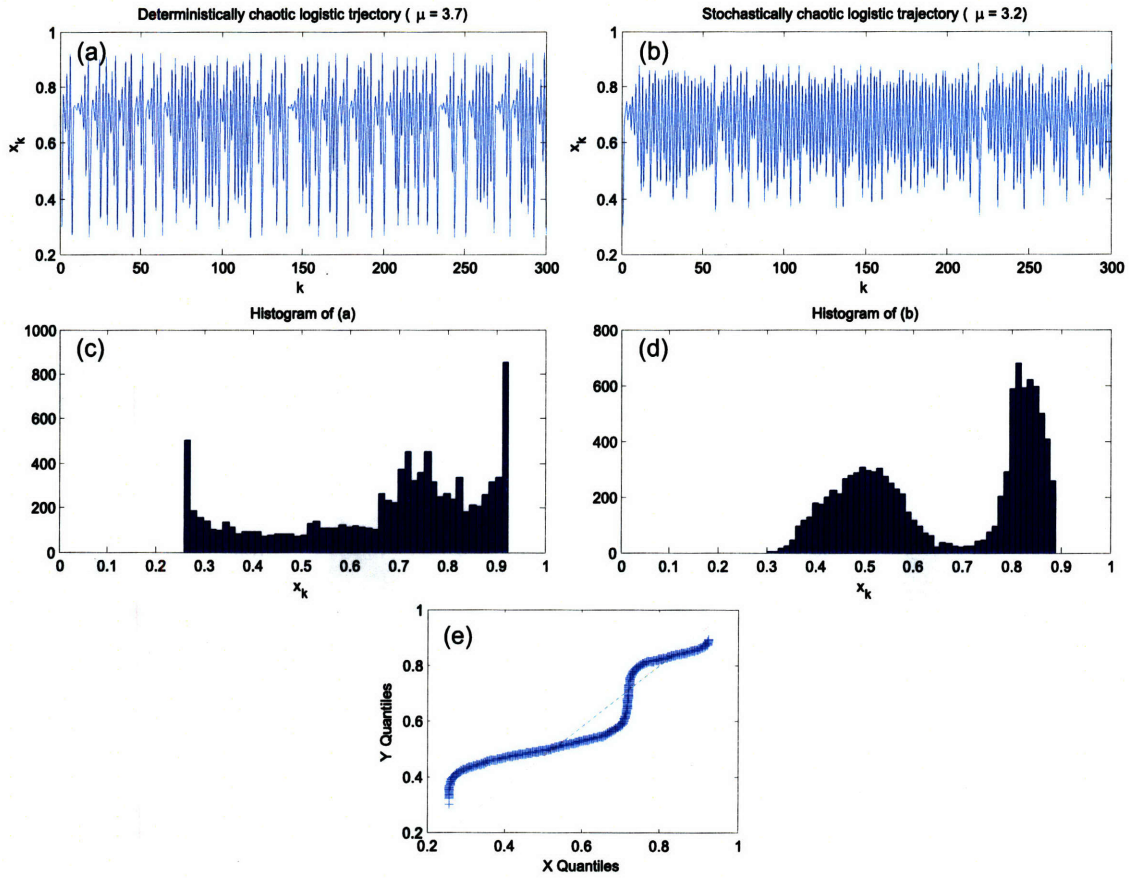


Figure 3-9: (a) Time series plot of a deterministically chaotic trajectory of the logistic map with $\mu = 3.7$. (b) Time series plot of a stochastically perturbed limit cycle trajectory with $\mu = 3.74$ and noise variance of 0.05. (c) Histogram of the deterministic trajectory. (d) Histogram of the stochastically perturbed trajectory. (e) Q-Q plot of the two trajectories.

because the LE is a long-term average over the deterministic attractor while the SLE is the long-term average over the stationary probability distribution.

Gao et al. [92] [94] taking issue with Crutchfield et al. suggested that the term "noise-induced chaos" should refer only to such situations involving chaotic invariants. The idea is that in a noisy system, a trajectory would often leave the deterministic attractor and come under the influence of the manifold of the chaotic set. The trajectory would display a "fly by" of the unstable chaotic set, and thus would sometimes appear to be under the influence of chaotic dynamics [57]. However, the SLE by itself cannot distinguish such chaotic invariant set from other types of initial condition sensitivity. The property of sensitivity to initial conditions, uniquely associated with chaotic dynamics in deterministic systems, is widespread in stochastic systems because the time spent near repelling invariant sets. Such sensitivity is due to a mechanism fundamentally different from deterministic chaos. Positive SLEs should therefore not be viewed as a hallmark of chaos [229].

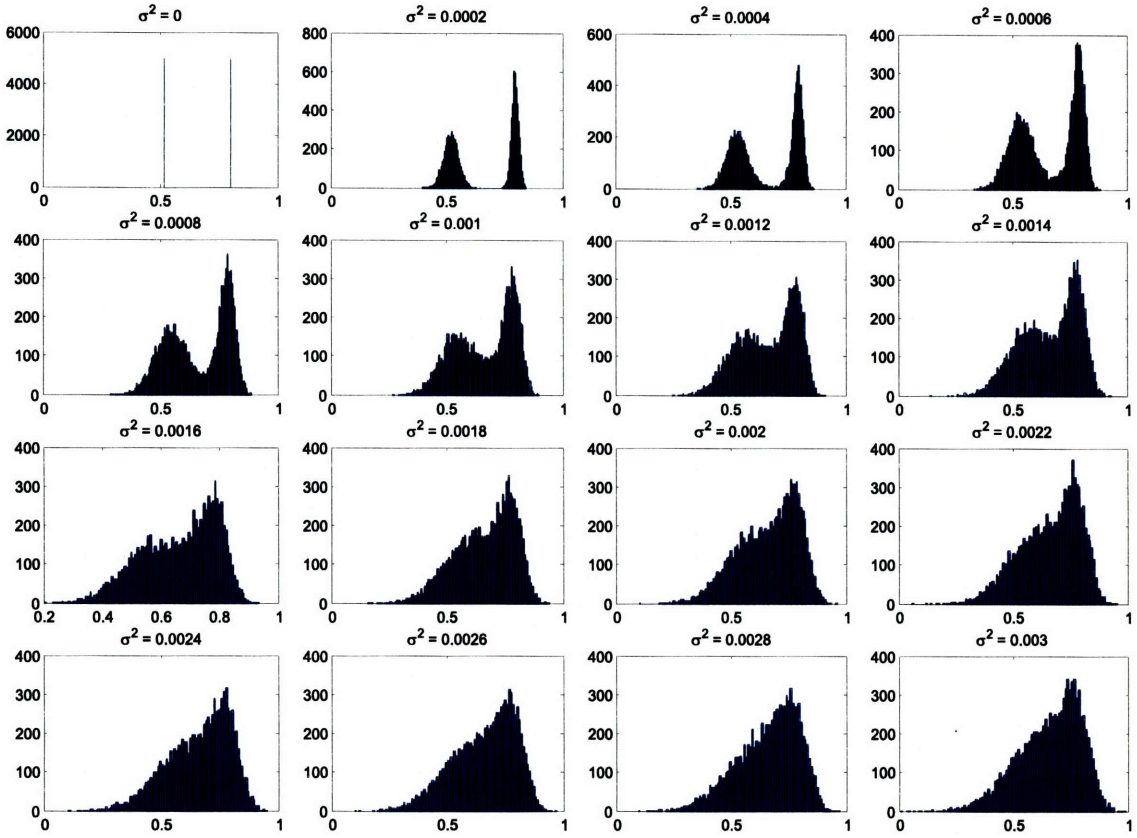


Figure 3-10: For the noise-free case one obtains delta functions for the probability distribution, with one delta function corresponding to each branch (in the periodic region). As the noise level is increased the probability distribution functions broaden, and eventually merge.

3.3.4 Novel definition for stochastic chaos

In our simulations of the logistic map in the period-2 regime, even though the bifurcation diagram and power spectra analysis can show that the system is not periodic, no features have yet arisen that would distinguish them as chaotic versus merely noisy. In order to show the existence of chaos, we employ the numerical-titration method [225]. Numerical titration produces a positive noise limit value for the period-2 logistic map perturbed by dynamic noise, indicating the presence of chaotic dynamics. The origin of this chaos can be thought of as follows. The perturbed system is given by sequences $\{\tilde{x}_k\}$, where

$$\begin{aligned} \tilde{x}_0 &= x_0, \\ \tilde{x}_k &= f(\tilde{x}_{k-1}) + \eta_k, \quad \text{for } n > 0. \end{aligned} \tag{3.21}$$

The noise term is assumed to be independent of the state, x, independent identically distributed, which we tacitly presume to be relatively small, so that the deterministic part has primary influence. The domain of the logistic map is restricted to the unit interval because the iterations of starting values outside of $[0, 1]$ diverge to minus infinity. That is also the

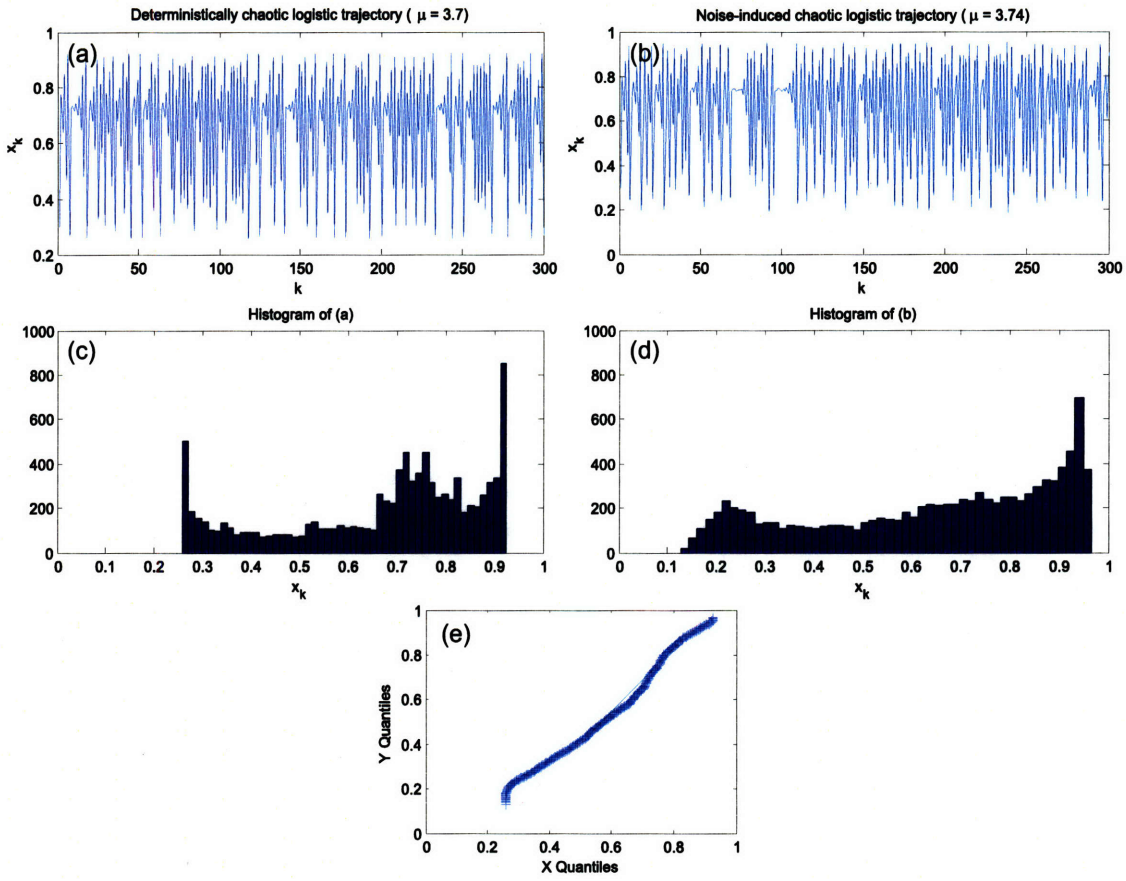


Figure 3-11: (a) Time series plot of a deterministically chaotic trajectory of the logistic map with $\mu = 3.7$. (b) Time series plot of a stochastically perturbed limit cycle trajectory with $\mu = 3.74$ and noise variance of 0.05. (c) Histogram of the deterministic trajectory. (d) Histogram of the stochastically perturbed trajectory. (e) Q-Q plot of the two trajectories.

reason why the fluctuations must have an upper bound, if the system should be confined to a finite interval. The fact that we are not allowed to admit fluctuations of arbitrary largeness is not in contradiction with intuition since we are dealing with a system which corresponds, e.g., in ecology to a finite population which of course will die out if there is a large enough fluctuation, i.e., a fatal catastrophe. The noise can also be interpreted as a *high-dimensional dynamical system* with some deterministic generating function

$$\begin{aligned} g : I_\beta &\mapsto I_\beta, \\ \eta_k &\mapsto \eta_{k+1}. \end{aligned} \tag{3.22}$$

In this way we are in fact considering a two-dimensional dynamical system

$$\begin{aligned} F : I_B(\beta) \times I_\beta &\mapsto I_B(\beta) \times I_\beta, \\ (\tilde{x}_k, \eta_k) &\mapsto (\tilde{x}_{k+1}, \eta_{k+1}). \end{aligned} \tag{3.23}$$

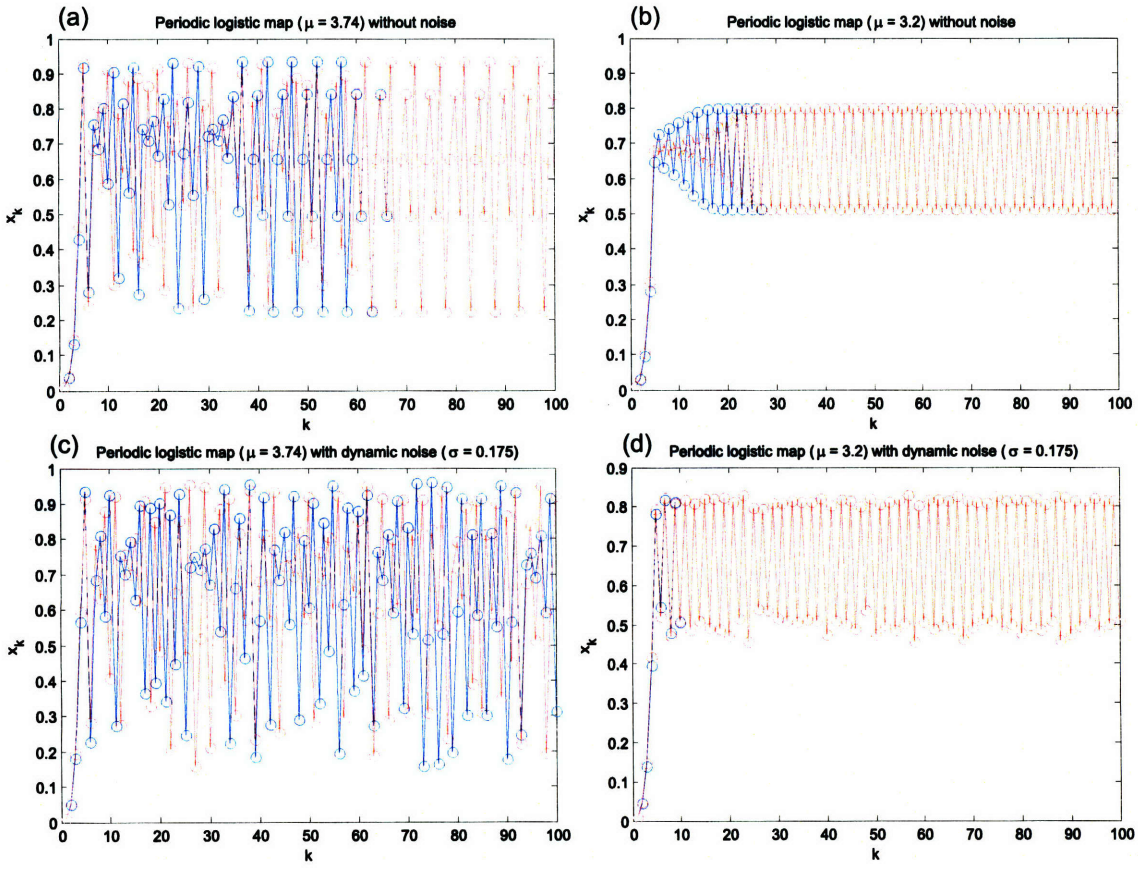


Figure 3-12: (a) The periodic logistic map trajectory for $\mu = 3.74$. (b) Periodic logistic map trajectory for $\mu = 3.2$. (c) The periodic logistic map trajectory for $\mu = 3.74$ with dynamic noise, this is the noise-induced chaos case described by Gao et al. [94], who demonstrated initial value sensitivity in this model. (d) The stochastic chaos case in the period-2 does not have sensitivity to initial conditions, but yet titratable.

where $I_B(\beta) \subset [0, 1]$ is the basin of the system. This can also be interpreted as the interaction of the two systems f and g. We speculate that the filtering of this high-dimensionally chaotic noise by the low-dimensional nonlinear (non-chaotic) deterministic skeleton model produces a low-dimensional chaotic projection which is detectable by the numerical-titration algorithm. This comment is by no means mathematically rigorous.

The systems in the previous studies are chosen so that they are close to the parameter where deterministic chaos occurs naturally [54] [55] [94]. Therefore, it is not surprising that small noise induces chaos, effectively simulating the dynamics of a larger but nearby parameter value (see Fig. 3-11). In our logistic map example presented above, we observed irregular oscillations well into the period-2 region, away from the deterministic chaos parameters. Initial value sensitivity may not hold in the case of stochastic chaos, as shown in Fig. 3-12. The chaotic motion generated is truly the effect of the dynamic noise and can only be detectable by numerical titration. Therefore, we propose the use of another quantity to characterize stochastic chaos, namely, the noise limit value from the numerical-titration

technique, which we will derive in the next chapter. Our simulation results (we will also show this in the next chapter) show that when dynamic is recycled through the nonlinear deterministic evolution law, numerical titration yields a nonzero value for the noise limit, which indicates the presence of chaotic dynamics. In contrast, when measurement noise is added to the system, the noise tends to obscure the chaotic dynamics and the noise limit is decreased. Hence, we propose a new definition for stochastic chaos: *a dynamical system is said to be stochastically chaotic if and only if it admits a positive noise limit*, as produced by the numerical-titration algorithm.

3.3.5 Connection between kinetic rate theory and stochastic chaos

In statistical physics, the popular view is that the evolution is given by the familiar deterministic macroscopic equations (such as Newton's equation of motion) and that if we are interested in fluctuations, such as the systematic interaction force due to the theintramolecular and intermolecular interactions of the Brownian particles, we should tag them on somehow in the form of a Langevin term. Many interesting results can often be obtained without solving the Langevin equation, from the fluctuation dissipation theorem. The main method of solution if a solution is required is by use of the Fokker-Planck equation, which provides a deterministic equation satisfied by the time dependent probability density. The Fokker-Planck-Langevin method places heavy emphasis on the role of fluctuations. This turns out to be very useful in dealing with diffusion and phase transition processes near critical points, where fluctuations are of overwhelming importance.

We adopt a discrete form of the Langevin's equation that has a double-well potential (the logistic map) so that we can derive the Arrhenius factor analytically and show the interdependency between the nonlinear deterministic evolution law and the stochastic perturbation. Specifically, the geometry of the potential barrier is expressed as the activation energy, which is determined deterministically; while the $k_{\mathrm{B}}T$ dependence of the transition rate is effect of the thermal fluctuation. In this way, the Arrhenius formula expresses the interplay between deterministic and stochastic properties of the system in one elegant expression. This is precisely our concept of stochastic chaos: noise generated from a high-dimensional chaotic system recycled through a low-dimensional deterministic "skeleton" evolutionary law.

We did not derive an explicit function relationship between the escape rate and noise limit in this thesis. However, we show that there is a positive correlation between noise limit and transition rate, as shown in Fig. 3-13 (c). Upon closer inspection, we see that at low noise levels, noise limit can still be positive when no transition occurs (transition rate = 0). This implies that the motion of the particle can become complex way before it gains enough energy to transition to another well. We believe the transition can contribute to additional stochastic chaotic dynamics, but very slightly.

We have simulated a few trials for the logistic map, injecting the same level of dynamic

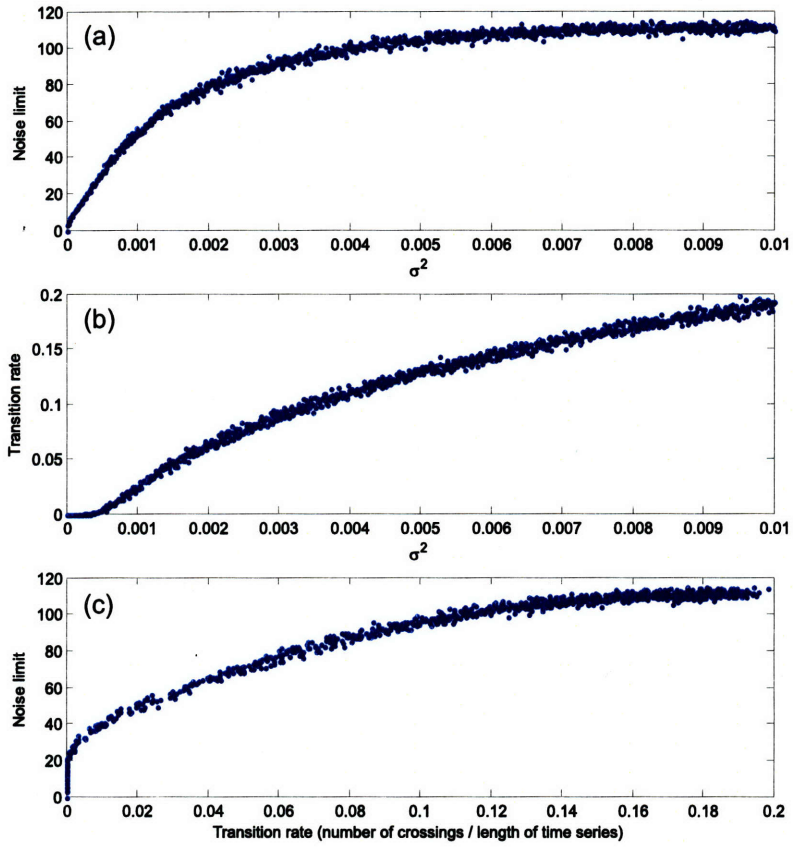


Figure 3-13: (a) Noise limit vs. noise intensities σ^2. (b) Transition rate vs. noise intensities. (c) Correlation between noise limit and transition rate.

noise, then we counted the number of transitions in the series and finally computed the noise limit. Table 3.1 shows that, *at the same noise level*, the transitions have only a marginally effect on noise limit. Since these transitions are such rare events (at low noise levels anyway), the effect on stochastic chaos is barely observable.

Table 3.1: Contribution of phase transition to stochastic chaos

Noise intensity σ	Length of time series	Number of transitions	Noise limit
0.28	10000	2	46.6
0.28	10000	3	48.2
0.3	10000	11	61.0
0.3	10000	17	62.4
0.3	10000	25	62.8

Since noise limit begins to become positive before the thermal agitation is large enough to excite the particle to escape the local potential well, we conclude that motion of the particle can be chaotic even when it's in a single-well potential. Indeed, numerical titration outputs a positive noise limit for the logistic map in the stable fixed point regime, with

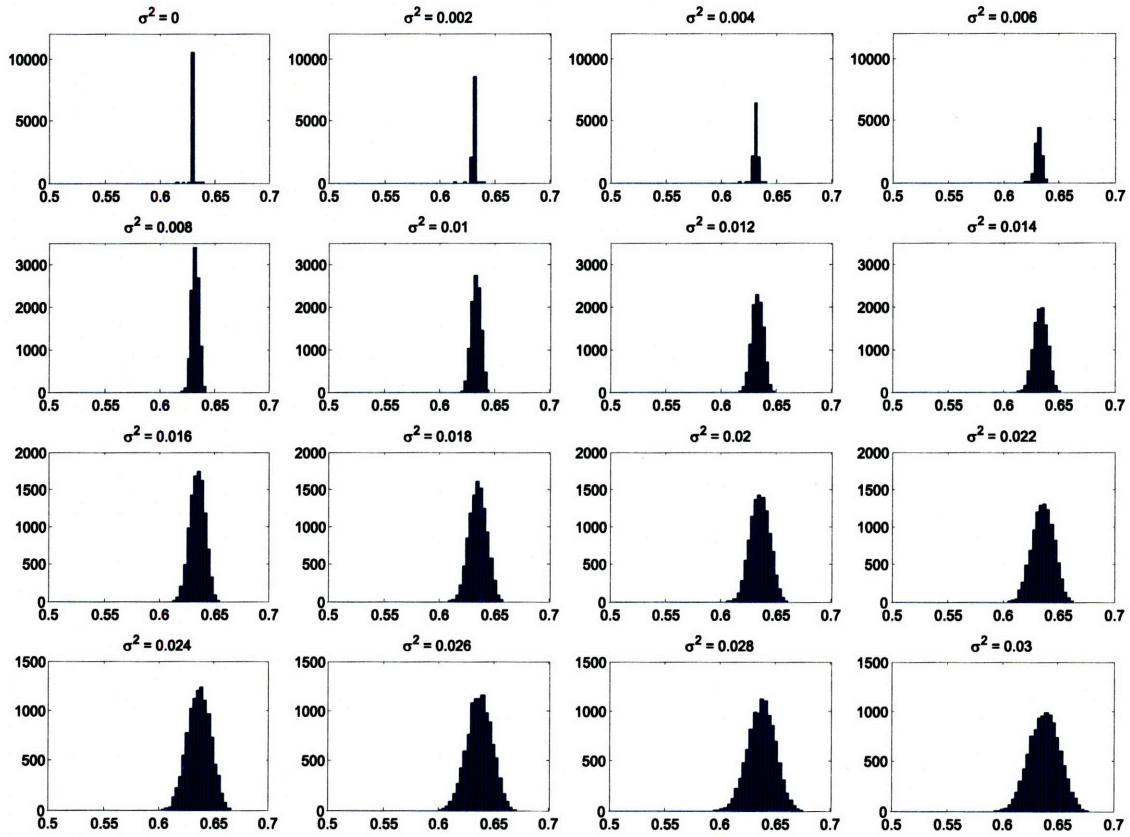


Figure 3-14: Probability distribution for the fixed-point logistic map $\mu = 2.7$ for increasing noise level. Again, the effect of the noise is to smear the delta function.

bifurcation parameter value of, say, $\mu = 2.7$. The evolution of the distribution of the fixed-point logistic map is shown in Fig. 3-14. The effect of the noise is to smear the single delta function; yet the time series is still titratable. This phenomenon is to be investigated further.

3.4 Summary

This chapter is an introduction to the applications in statistical mechanics of chaotic dynamics, and also to the use of techniques in statistical mechanics important for an understanding of the chaotic behavior of a dynamical system. We discussed the connections between kinetic rate theory, needed to describe macroscopic properties, and quantities, such as Lyapunov exponents and Kolmogorov-Sinai entropies, which describe the microscopic, chaotic behavior of the system. Moreover, such chaos-like behavior could be analyzed in detail for simple systems with just a few degrees of freedom. These simple model systems provide a great deal of insight into how more complicated dynamical systems might actually behave, and how one could then provide a more substantial foundation for the validity of the Boltzmann equation and related chaos-like equations commonly used to treat nonequilibrium systems.

With the logistic model, not only can we predict the Arrhenius-like exponential dependence of the escape rate on temperature, we also see that the transitions play a critical role in the chaotic dynamics of the time series. We have also discussed some differences between our definition of stochastic chaos and the exist definition of noise-induced chaos, which is really "near-deterministic" chaos. We showed that the phase transition phenomenon has a minimal effect on the chaotic behavior of the system. It is a future task to quantify to what extend do the transition affect stochastic chaotic dynamics at high noise levels.

Finally, one would also be interested in the quantum effects on our analysis of the particle's transition from one potential well to another. Quantum mechanical effects such as tunneling through a classically impenetrable potential energy barrier have long been known to occur in chemical reactions [17]. A detail analysis of this effect is beyond the scope of this thesis.

Chapter 4

Simulation Studies I

> I have had my results for a long time: but I do not yet know how I am to arrive at them.
>
> —Carl Friedrich Gauss [11]

In the previous chapter, we have expounded on the interdependence between deterministic and probabilistic properties of dynamical systems under the influence of stochastic perturbations. Randomness may have several sources, including chaotic dynamics, temperature fluctuations that are internal to the system, and random external perturbations. The aim of this chapter is present theoretical methods which allow one to characterize such systems.

In deterministic dynamical systems there exist well-established ways to define the complexity of a temporal evolution in terms of **Lyapunov exponents** [157] and **Kolmogorov-Sinai entropy** [286]; see Ref. [255] for a review. However, the situation becomes much more ambiguous in presence of a random perturbation, which are always present in physical systems as a consequence of thermal fluctuations, or uncontrollable changes of control parameters, or in numerical experiments simply because of the roundoff errors. In the small noise limit, the perturbation changes the stationary behavior by introducing the possibility of jumps among different attractors (stable fixed points, stable limit cycles, or tori). An example is the Langevin equation describing the motion of an overdamped particle in a double well, as presented in Chapter 2. As we have seen, the combined effects of the noise and of the deterministic part of the evolution law can produce highly intriguing behavior.

In Sec. 4.1, we present one approach to characterize stochastically perturbed systems by calculating the Lyapunov exponent for the rate of divergence of two initially nearby trajectories evolving under the same realization of the noise, in order to avoid the devastating impact of noise on the calculation. Although the Lyapunov exponent is a well-defined quantity, it is neither unique nor the most useful characterization of complexity. Furthermore, it is practically impossible to extract the Lyapunov exponent from an analysis of experiment

data [181]. This chapter presents several other measures of complexity for stochastically perturbed dynamical systems. In Sec. 4.2, we use the **numerical-titration** technique to detect for the existence of chaos and give a quantitative measure of its intensity in the stochastic system. In Sec. 4.3, we present a popular statistical index that quantifies the unpredictability of fluctuations in a time series, a method known as **approximate entropy**. The fractal scaling properties are quantified with the use of **detrended fluctuation analysis** in Sec. 4.4. Finally, we give a brief note on power spectral analysis and multifractal analysis in Sec. 4.5 and Sec. 4.6, respectively.

4.1 Lyapunov exponent: identical-noise-realization approach

It is perhaps quite natural to try to retain the Lyapunov exponent even after we have clothed the deterministic skeleton with dynamic noise. However, in the presence of dynamic noise, the problem arises how to quantify the divergence of two realizations starting from two nearby initial values. One proposed way is to use a device analogous to the well-known coupling technique in probability theory by assuming that the two realizations *share the same noise sequence.* This device was first used by Herzel et al. [132] in 1987 and later adopted by Nychka et al. [209] and others [46] [152]. Clearly, if two nearby initial states share the same realization of the dynamic noise, then the calculation and interpretation of the Lyapunov exponent remain almost unchanged. Specifically, consider the model

$$x_{k+1} = f(x_k) + \eta_{k+1} \tag{4.1}$$

Let x_0 and $\tilde{x}_0$ denote two initial values such that $x_0 - \tilde{x}_0 = \delta$. After one iteration of the system dynamics we have

$$|x_1 - \tilde{x}_1| = |(f(x_0) + \eta_1) - (f(\tilde{x}_0) + \tilde{\eta}_1)|. \tag{4.2}$$

The assumption of identical dynamic noise realizations implies that $\eta_1 = \tilde{\eta}_1$, and this gives

$$|x_1 - \tilde{x}_1| = |f(x_0) - f(\tilde{x}_0)|. \tag{4.3}$$

It is now clear that the assumption of identical dynamic noise realization allows us to obtain the Lyapunov exponents in a similar way as in the deterministic case.

The Lyapunov exponent of a dynamical system can be determined without actually knowing and solving the underlying evolution law explicitly. This occurs when we obtain a chaotic time series from a dynamical system, reconstruct its strange attractor in the corresponding pseudo-phase space, and then compute the Lyapunov exponent from the reconstructed strange attractor directly, without its explicit mathematical model. Some common methods for estimating the Lyapunov exponent are presented by Wolf et al. [298],

Rosenstein et al. [241], and Brown et al. [30]. Gottwald and Melbourne [111] proposed a modification to the above methods for the detection of determinism in a noisy time series. However, the problem remains when the analyzed signal contains a nontrivial noise component, dimension and Lyapunov measure are unsuitable, since the Lyapunov exponent of noise is infinite. Paladin et al. [212] devised an alternative characterization of complexity in dynamical systems with a random perturbation by considering the rate K of divergence of nearby orbits evolving under two *different* noise realizations. The meaning of K has some physical relevance within the framework of information theory, and very different from the Lyapunov exponent.

4.2 Numerical titration

Though the return maps and other forms of analysis seem to imply that dynamic noise can in fact produce chaos, it is desirable to have a more quantitative measure of the existence of chaos and its intensity. This is provided by the numerical-titration method [225]. The numerical-titration technique offers a highly sensitive litmus test for chaotic dynamics and a relative measure for tracking chaoticity in a noise-contaminated experimental time series. Rather than filtering the noise before chaos analysis, which could risk "bleaching" the chaotic dynamics [277], the numerical-titration technique is analogous to a chemical titration: it measures chaos (acidity) by a control neutralization with added noise (base), thus allowing a "litmus test" for chaos. Numerical titration has been applied to heart-rate variability analysis in patients with congestive heart failure [226], in speech signal analysis [275], and more recently, in pediatric sleep state analysis [60], and in the detection of obstructive sleep apnea syndrome in children [61].

4.2.1 Volterra autoregressive model

The heart of the numerical-titration algorithm is the **Volterra autoregressive series (VAR)** nonlinear detection (NLD) algorithm. The development of Volterra models relies on the mathematical notion of the Volterra series (a functional power series expansion) introduced by the distinguished Italian mathematician Vito Volterra about a century ago [6]. The term "Volterra series" is used to denote the functional expansion of an analytic functional representing the input-output relation of an continuous and stable nonlinear dynamic system with *finite memory*. The requirement of finite memory is necessary for the convergence of the Volterra functional series expansion. Even non-analytic, but continuous, functionals (corresponding to systems with non-differentiable continuous input-output relations) can be approximated to any desired degree of accuracy by an analytic functional (and therefore a Volterra series) in a matter akin to the Weierstrass polynomial approximation theorem for non-analytic continuous functions. Thus, the applicability of the Volterra series expansion to system modeling is very broad and requires a minimum of prior assumptions.

We can state that the output $y(t)$ of a stationary stable causal system can be expressed in terms of its input signal $x(t)$ by means of the Volterra series expansion as [28] [243]:

$$y(t) = h_0 + \int_0^\infty \mathrm{d}\tau\, h_1(\tau)x(t-\tau) + \int_0^\infty \mathrm{d}\tau_1 \int_0^\infty \mathrm{d}\tau_2\, h_2(\tau_1,\tau_2)x(t-\tau_2) + \cdots$$
$$+ \int_0^\infty \mathrm{d}\tau_1 \cdots \int_0^\infty \mathrm{d}\tau_r\, h_r(\tau_1,\ldots,\tau_r)x(t-\tau_1)\cdots x(t-\tau_r) + \cdots \tag{4.4}$$

where the range of integration (τ_i from 0 to ∞) indicates that past and present values of the input affect the output present value in a manner determined by the Volterra kernels (i.e., the kernels should be viewed as weighting functions in this integration).

In discrete time, the Volterra expansion formulates the discrete time-invariant and causal nonlinear input/output relationship as a polynomial expansion of the output y_k in terms of the input delays x_k, x_{k-1}, ..., $x_{k-\kappa+1}$, where κ is the memory of the system:

$$y_k = h_0 + \sum_{m_1=0}^{\kappa-1} h_1(m_1)x_{n-m_1} + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} h_2(m_1,m_2)x_{n-m_1}x_{n-m_2} + \cdots \tag{4.5}$$

where $\{h_r(m_1,m_2,\ldots,m_r)\}$ is the rth Volterra kernel, which are analogous to the impulse response of linear systems theory.

The Volterra autoregressive method recasts the moving-average Volterra expansion into an autoregressive form to partially capture the dynamic behavior of the system [14]:

$$\hat{x}_{n+1} = h_0 + \sum_{m_1}^{\kappa-1} h_1(m_1)x_{n-m_1} + \sum_{m_1=0}^{\kappa-1}\sum_{m_2=0}^{\kappa-1} h_2(m_1,m_2)x_{n-m_1}x_{n-m_2} + \cdots \tag{4.6}$$

To convert the functional forms of the Volterra kernels into a set of scalar coefficients, we reformulate the expansion as a weighted sum of functionals:

$$\begin{aligned}\hat{x}_k(\kappa,d) &= a_0 + a_1x_{k-1} + a_2x_{k-2} + \cdots + a_\kappa x_{k-\kappa} + a_{\kappa+1}x_{k-1}^2 + \\ &\quad a_{\kappa+2}x_{k-1}x_{k-2} + \cdots + a_{M-1}x_{k-\kappa}^d \\ &\equiv \sum_{m=0}^{M-1} a_m z_m(k)\end{aligned} \tag{4.7}$$

where the functional basis $\{z_m(k)\}$ is composed of all the distinct polynomial combinations of the embedding space coordinates (x_{k-1}, x_{k-2}, ..., $x_{k-\kappa}$) up to a maximum delay κ and maximum order d, with total dimension

$$M = \frac{(\kappa+d)!}{\kappa!d!}. \tag{4.8}$$

The corresponding set of scalar coefficients $\{a_m\}$ represent the set of Volterra kernel functions, characterizing the autoregressive behavior of the dynamical system. Each model is

parameterized by κ, the embedding dimension, and d, the degree of the nonlinearity of the model (i.e., $d = 1$ for linear model and $d > 1$ for nonlinear model). The coefficients a_m are recursively estimated from eq. (4.7) from the entire time series using the Korenberg algorithm [162].

The conversion from the original representation to the modified representation is

$$z_m(k) = \begin{cases} 1, & m = 0; \\ x_{k-m+1}, & 1 \le m \le \kappa; \\ x_{k-j_1} x_{k-j_2}, & \kappa + 1 \le m \le \kappa^2/2. \end{cases} \tag{4.9}$$

where $j_1 = \lfloor m/\kappa \rfloor$ and $j_2 = \mathrm{mod}(m, \kappa)$.

The kernels of the original representation can be recovered as follow:

$$\begin{aligned} h_0 &= a_0 \\ h_1(m_1) &= a_{m_1+1} \\ h_2(m_1, m_2) = h_2(m_2, m_1) &= \begin{cases} a_{m_1\kappa+m_2}, & m_1 = m_2; \\ a_{m_1\kappa+m_2}/2, & m_1 \neq m_2. \end{cases} \end{aligned} \tag{4.10}$$

4.2.2 Goodness of fit

In the VAR algorithm the coefficient $\hat{a}_m$ for $m = 1, 2, \ldots$ (estimated by using the Korenberg algorithm) are used to generate a predicted time series $\hat{x}_k$ as follows:

$$\hat{x}_k(\kappa, d) = \sum_{m=0}^{M-1} \hat{a}_m \hat{z}_m(n) \tag{4.11}$$

The goodness of fit of a model is measured by the normalized residual sum-of-square errors:

$$\epsilon^2(\kappa, d) \equiv \frac{\sum_{n=1}^{N} (\hat{x}_n(\kappa, d) - x_k)^2}{\sum_{n=1}^{N} (x_k - \langle x_n \rangle)^2} \tag{4.12}$$

where $\langle x_n \rangle = (1/N) \sum_{i=1}^{N} x_i$ and $\epsilon^2(\kappa, d)$ is in effect a normalized variance of the residual error. An optimal model is one that minimized the following information criterion [5]:

$$C(r) = \log \epsilon(r) + \frac{r}{N} \tag{4.13}$$

where $r \in [1, M]$ is the number of polynomial terms of a truncated Volterra expansion.

Increasing the number of free parameters to be estimated improves the goodness of fit, regardless of the number of free parameters in the data generating process. The **Akaike**

information criterion (AIC) not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages overfitting. The preferred model is the one with the lowest AIC value. The AIC methodology attempts to find the model that best explains the data with a minimum of free parameters (parsimony principle). Many researchers extended Akaike's original work to include bias correction to adjust for AIC's propensity to favor high-dimensional models when the sample size is small relative to the maximum order of the models in the candidate class. Such a variant, called the "improved" Akaike information criterion or AICi, is based on an idea advanced by Hurvich et al. [139] in the context of univariate Gaussian autoregressive models. The criterion involves the same goodness-of-fit term as AIC, yet features a penalty term that arises via a simulated bias correction [18]. The resulting AIC variant has been shown to be less bias than traditional AIC, that AICi does not exhibit a propensity to favor grossly over-parameterized models, and that AICi generally outperforms its competitors in terms of correct order selections.

4.2.3 Detection of chaotic dynamics in measurement noise

One may determine whether the underlying dynamics of the system under study are linear ($d_{\text{optimal}} = 1$) or nonlinear ($d_{\text{optimal}} > 1$). For the closed-loop Korenberg series it is found that the nonlinear Korenberg model is only superior when the system is not only nonlinear but also chaotic. This is reasonable since a periodic (but nonlinear) system returns to the same values and thus it should be possible to represent a given value as linear combination of previous values. For example, in the period-two region the model is just $x_{n+1} = x_{n-1}$ and $d_{\text{optimal}} = 1$. In contrast, a chaotic system never returns to the same value; thus, it cannot be represented in such a simple fashion and $d_{\text{optimal}} > 1$.

In addition to linear vs. nonlinear hypothesis testing, the VAR method provides a sufficient test for chaotic dynamics when used in conjunction with a "noise titration" procedure: the dynamics of nonlinearity are tested on the time series. If linearity is detected, then the noise titration method rejects the null hypothesis of no chaotic behavior. If nonlinearity is detected, small amounts of Gaussian noise are successively added until nonlinearity is no longer detected (within a prescribed level of statistical confidence). Since pure noise is best modeled by a linear model, this method can be used not only to determine the presence of chaos, but its intensity. If enough measurement noise is added to a chaotic system, its nonlinearity will no longer be apparent using the method described above. This is the basis of the numerical-titration method: measurement noise "base" is added to the chaotic time series ("acid") until it is neutralized. The point at which neutralization occurs is signified by the transition of d_{optimal} from a value great than one to unity. The maximum noise added before nonlinearity goes undetected is called the noise limit. Under this scheme, a noise limit greater than zero represents the detection of chaotic dynamics. In addition, the noise limit mirrors the maximal Lyapunov exponent of the system dynamics [225].

4.2.4 Numerical titration results

The titration method was applied to the periodic limit cycles as well as the chaotic logistic map. In Fig. 4-1, the noise-limit values are plotted against noise intensities for different values of the bifurcation parameter, μ.

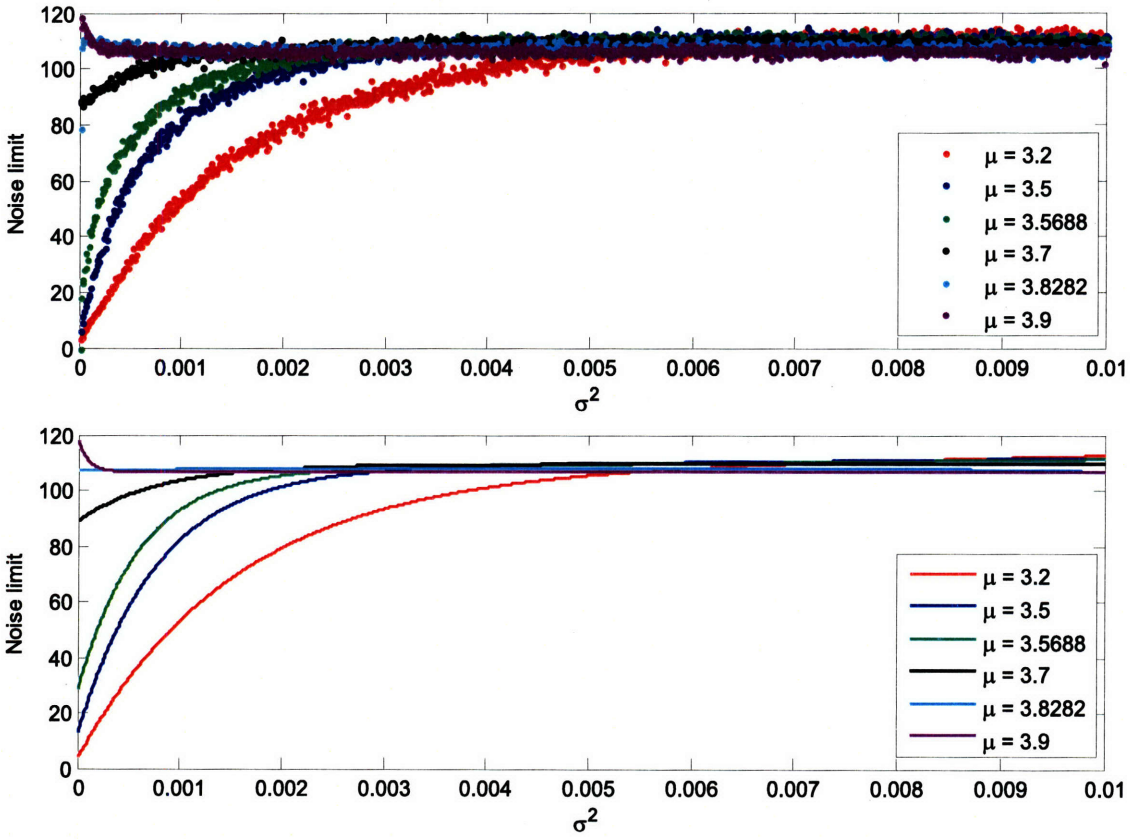


Figure 4-1: Titration of the logistic map. Top panel: Noise-limit values vs. noise intensities, plotted for the different values of the bifurcation parameter, μ. Bottom panel: Exponential fittings of the data in the top panel. Evidence for stochastic chaos is clearly observed as the increasing noise intensities drive the system to higher chaos levels.

From Fig. 4-1, we can clear observe the evidence for stochastic chaos as we have defined it in Chapter 3. For the periodic trajectories driven by dynamic noise (μ = 3.2, 3.5, and 3.5688), the noise-limit values increase in a convex exponential fashion as the noise intensities increase. Even the "shallow" chaotic orbit (μ = 3.7 and 3.8282) was driven to become slightly more chaotic. For the trajectories well into the chaotic regime ($\mu = 3.9$), the chaos level actually decreases slightly from it's intrinsic chaos level. This could correspond to the case of "noise-induced order," as the noise kicks the system slightly out of the deep chaos regime. Or it is more likely an artifact of our numerical procedure to keep the logistic map out of bound when we add dynamic noise to it.

Curiously, the noise-limit curves all converge to approximately the same value at high noise levels regardless of the value for the bifurcation parameter. This observation suggest

that perhaps there is a natural, preferred induced-chaos level that is universal for the system.

4.3 Approximate entropy

Rooted in the work of Grassberger and Procaccia [113] and Eckman and Ruelle [70], **approximate entropy (ApEn)** is a "regularity statistic" closely related to the Kolmogorov entropy that estimates the level of complexity and predictability of fluctuations in dynamical systems [222]. ApEn provides quantitative information about the complexity of experimental data that are often short and noise contaminated, and in many cases, have inherent dynamics that exhibit both deterministic and stochastic behaviors. ApEn reflects the logarithmic likelihood that templates in the sequence that are similar (within a tolerance r) will remain similar on next incremental comparisons. A time series that exhibits frequent and similar epochs has a relatively small ApEn value, reflecting strong regularity. Alternatively, large values of ApEn imply the presence of substantial fluctuation in the time series. This method has been applied in physiological time-series analysis [2] [135] [223] [238].

4.3.1 Approximate entropy algorithm

The algorithm for computing ApEn is as follows:

1. Given an N-point time series , we form m-vectors, $X(1)$ to $X(N-m+1)$ defined by: $X(i) = [x(i), x(i+1), \ldots, x(i+m-1)]$, $i = 1, \ldots, N-m+1$.

2. Define the distance $d[X(i), X(j)]$ between vectors $X(i)$ and $X(j)$ as the maximum absolute difference between their respective scalar components:

$$d[X(i), X(j)] = \max_{k=0\ldots m-1} [|x(i+k) - x(j+k)|]. \tag{4.14}$$

3. Define for each i, for $i = 1, \ldots, N-m+1$, let $C_r^{(m)}(i) = V^{(m)}(i)/(N-m+1)$, where $V^{(m)}(i)$ = number of $d[X(i), X(j)] \leq r$, where r defines the criterion of similarity. The quantity $C(r)$ expresses the prevalence of repetitive patterns of length m in $x(n)$.

4. Take the natural logarithm of each $C_r^{(m)}(i)$, and average it over i as defined in Step 3:

$$\phi^{(m)}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_r^{(m)}(i). \tag{4.15}$$

5. Increase the pattern dimension to $m+1$ and repeat Steps 1 to 4.

6. Calculate ApEn value for a patterns of length m, and similarity criterion r:

$$\text{ApEn}(m, r, N) = \phi^{(m)}(r) - \phi^{(m+1)}(r). \tag{4.16}$$

Pincus, who developed the ApEn method, suggested that r be 0.1 to 0.25 times the standard deviation of the data, and that m be 1 or 2 for data lengths (N) ranging from 100 to 5000 data points. We chose $m = 2$ and $r = 20\%$ of the standard deviation of the time series.

4.3.2 Approximation entropy results

In Fig. 4-2, the approximate entropy values are plotted against noise intensities for different values of the bifurcation parameter, μ.

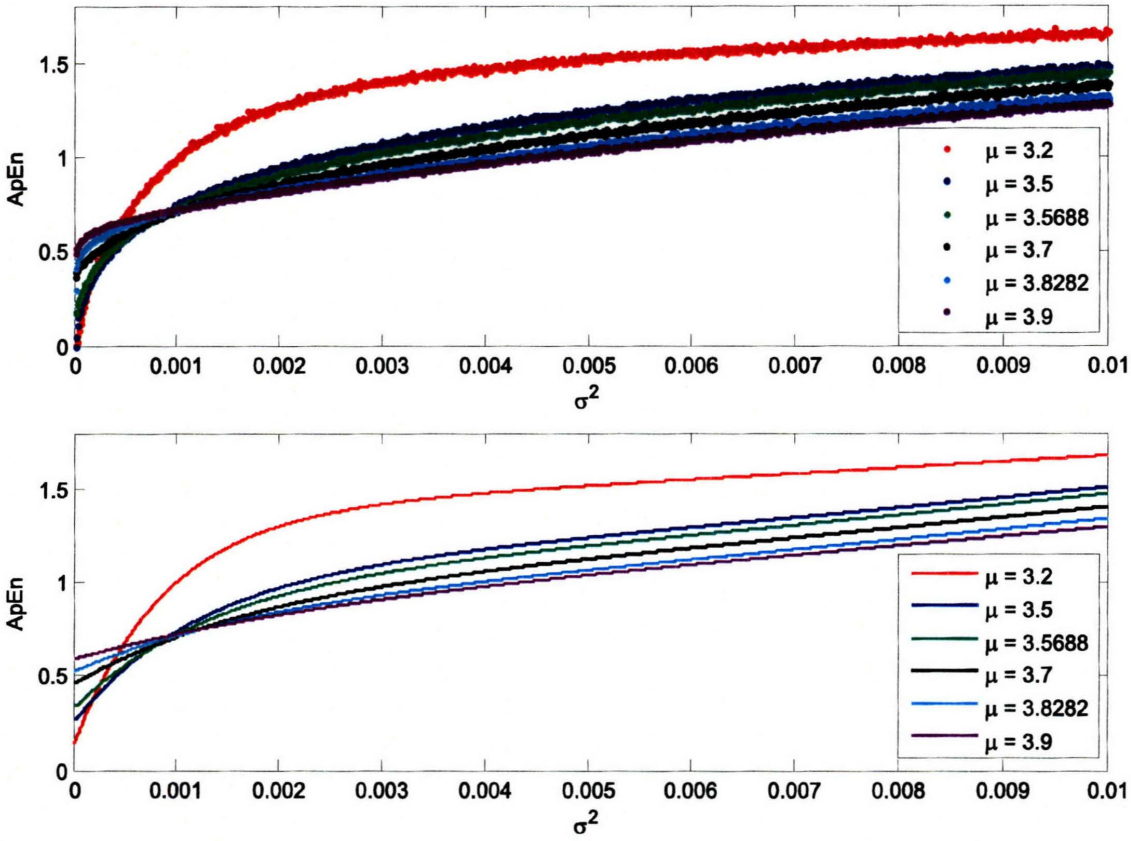


Figure 4-2: Approximate entropy for the logistic map. Top panel: ApEn values vs. noise intensities, plotted for the different values of the bifurcation parameter, μ. Bottom panel: Exponential fittings of the data in the top panel. The approximate entropy curves are increasing functions of noise intensities in all cases, as expected.

Since approximate entropy measures the level of randomness in the time series, we naturally expect the entropy values to rise with as level of noise increase. In this way, we have little to infer about noise-induced complexity from approximate entropy calculations. We make one more observation before we go on to another approach. And that is, noise has a greater effect for the trajectories with smaller bifurcation parameter. Specifically, the approximate entropy values change the most for the logistic map with $\mu = 3.2$ and least for the map with $\mu = 3.9$. This result suggest that initially chaotic trajectories have some "immunity" to noise, in the sense that their statistical behavior is not greatly affected by

the addition of noise.

4.4 Detrended fluctuation analysis

Detrended fluctuation analysis (DFA) was developed by Peng et al. [218] for the analysis of DNA and soon applied to many dynamical phenomena including a wide range of physiologic signals [127], particularly in the analysis of cardiac rhythm fluctuations [1] [2] [219], and the identification of coding regions in DNA sequences [211]. DFA quantifies the fractal scaling properties (the long-range power-law correlation exponents, or the **Hurst exponent**) of a signal by computing the root-mean-square fluctuation of the integrated and subtractive detrended series. This technique is a modification of root-mean-square analysis of random walks applied to non-stationary data.

4.4.1 Detrended fluctuation analysis algorithm

The DFA algorithm is briefly described below:

1. The N-point time series $x(n) = [x(1), x(2), \ldots, x(N)]$ is integrated,

$$y(k) = \sum_{i=1}^{k} [x(i) - \mu_x], \tag{4.17}$$

   where $\mu_x = \sum_{i=1}^{N} x(i)/N$ is the mean of the time series.

2. The integrated time series is divided into windows of equal length, n.

3. In each window, a least-squares line is fit to the data to represent the local trend and is denoted $y_n(k)$. We detrend the integrated time series, $y(k)$, by subtracting the local trend, $y_n(k)$, in each box.

4. The root-mean-square fluctuation of this integrated and detrended time series is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [y(k) - y_n(k)]^2}. \tag{4.18}$$

5. The above computation is repeated over all time scales (window lengths) to characterize the relationship between $F(n)$, the average fluctuation, and the window length n. Typically, $F(n)$ will increase with n. A power-law relation between $F(n)$ and n indicates the presence of scaling $F(n) \sim n^{\alpha}$. A linear relationship on a log-log plot indicates the presence of power law (fractal) scaling. Under such conditions the fluctuations can be characterized by the scaling exponent α, the slope of the line relating

$\log F(n)$ to $\log n$. Note that since for $n = \mathcal{O}(N)$, only very few time windows constitute the average in the $F(n)$ equation, the results for large n become biased, therefore, only $n \leq N/4$ was used. The trend removal also introduces bias when n is too small.

The scaling exponent can also be found in the power spectrum [185]. However, with the detrending done in the time domain on time scales matching exactly the time scale on which the fluctuations are analyzed, the scaling laws are typically better expressed by DFA than by the power spectrum.

4.4.2 Detrended fluctuation analysis results

In Fig. 4-3, the scaling exponent α is plotted against noise intensities for different values of the bifurcation parameter, μ.

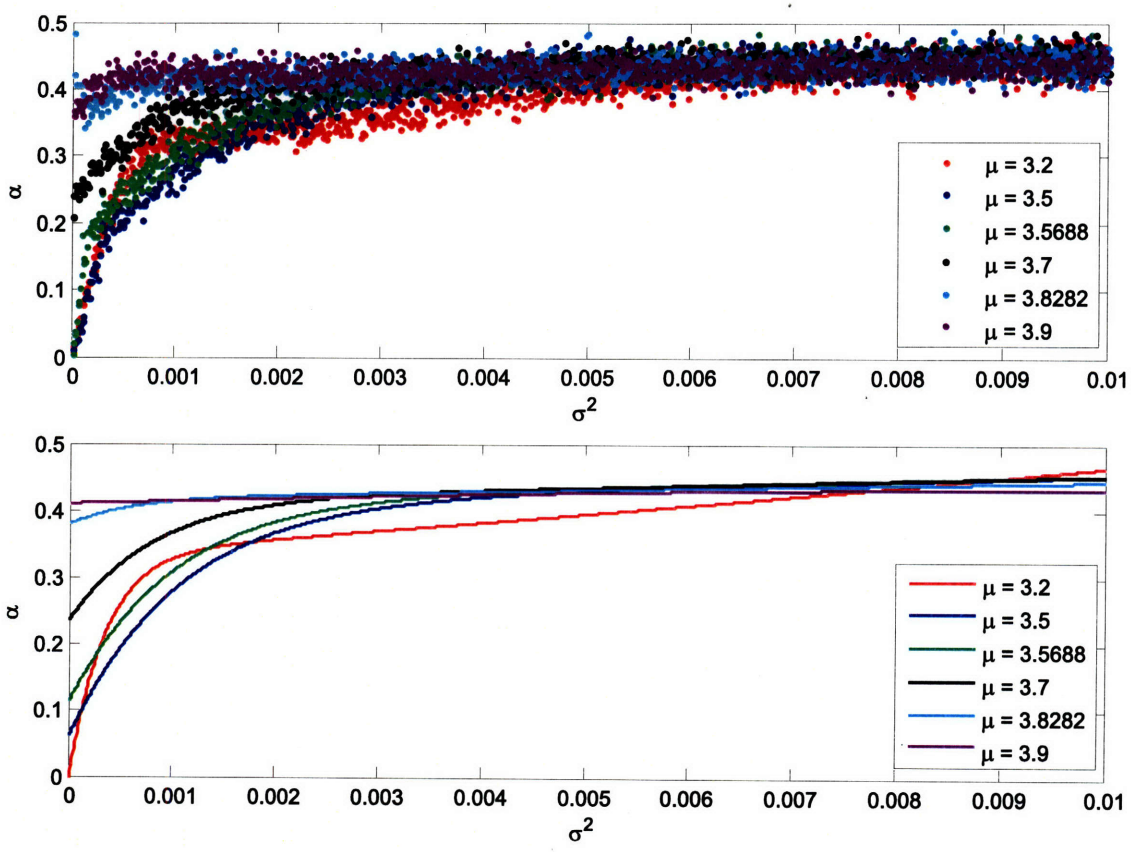


Figure 4-3: Detrended fluctuation analysis of the logistic map. Top panel: Scaling exponent α vs. noise intensities, plotted for the different values of the bifurcation parameter, μ. Bottom panel: Exponential fittings of the data in the top panel.

As shown in Fig. 4-4, when the calculated scaling exponent is equal to 0.5, it is indicative of white noise. If the value of α lies between 0.5 and 1 then the time series exhibits long-range correlations; $\alpha \approx 1.5$ for Brownian noise. We can see from Fig. 4-3 that the scaling exponent as a function of the noise intensity is approaching 0.5 for all values of

the bifurcation parameter. Again, as in the case of approximate entropy, calculation of the scaling exponent using DFA merely reflects the presence of noise.

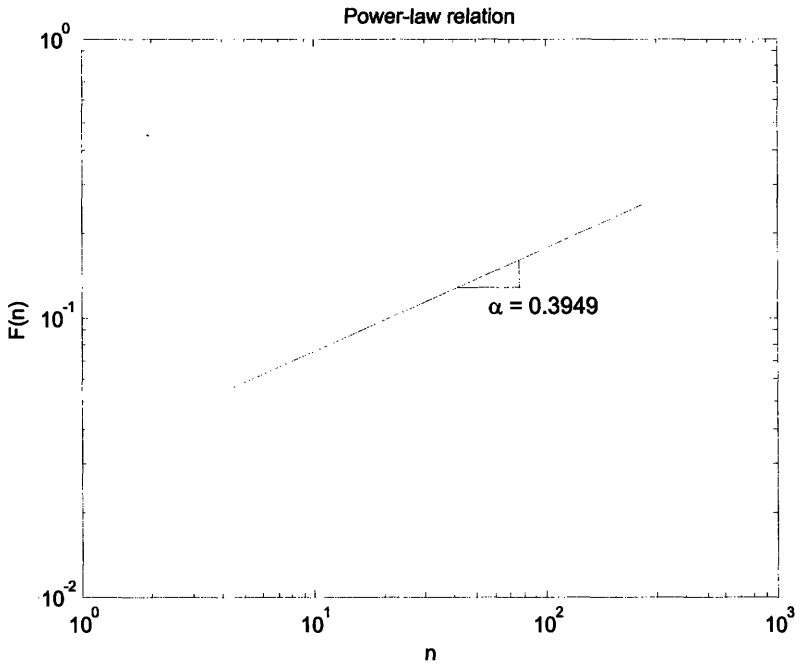


Figure 4-4: Power law relation of white noise. The calculated scaling exponent α is approximately equal to 0.5.

4.5 A note on power spectral analysis

Perez and Jeffries [220] analyzed the power spectral density for the logistic map and found that the additive noise produces sudden aperiodicity at bifurcation windows; their experimental findings are in agreement with the computed predicted behavior and are further detailed evidence for universal chaotic behavior. Similar results were obtained by Testa et al. [276] for a driven nonlinear oscillator. In this study, we compared the power spectra of white noise, deterministic chaos, and stochastic chaos case, as shown in Fig. 4-5. The power spectra for the noisy systems are broadband, as is the case for chaotic systems, in which time correlations decay; however, such spectra are also obtained due to noise alone so these results do not verify the existence of chaos.

4.6 A note on multifractal analysis

Scaling invariance plays a fundamental role in many natural phenomena and is often related to the appearance of irregular forms which cannot be described by the usual differential geometry. A classical example is again given by Brownian motion. The necessity of introducing a new class of geometrical objects, namely, **fractals**, has subsequently arisen in

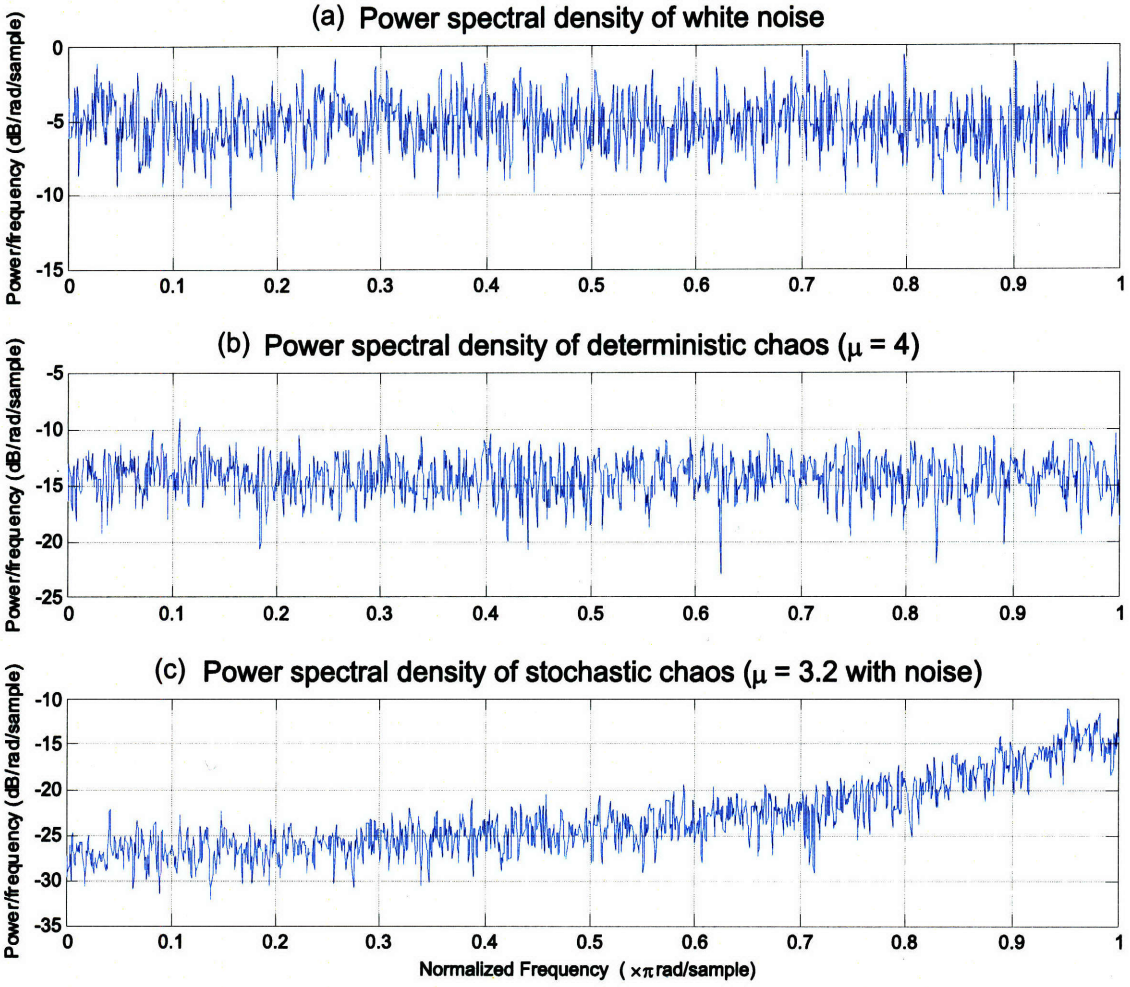


Figure 4-5: (a) Power spectral density of white noise; (b) Power spectral density of deterministic chaos (logistic map with $\mu = 4$), which shows the same characteristic as the white noise spectrum; (c) Power spectral density of stochastic chaos (logistic map with $\mu = 3.2$ with dynamic noise), which also exhibits a broadband spectrum.

various different problems. The concept of "fractal object" was explicitly formulated and popularized by Mandelbrot [184]. For a comprehensive review the multifractal formalism, see, for example, Ref. [213].

The central idea consists in the characterization of the scaling structure of an object by means of an index, the **fractal dimension**, D_{F}, which coincides for "ordinary" shapes with the usual topological dimension D_{T}. We can define the fractal dimension as a "capacity" measure by considering the number $N(\ell)$ of hypercubes of edge ℓ necessary to cover an object embedded in a D-dimensional space in the limit $\ell \to 0$:

$$N(\ell) \propto \ell^{-D_{\mathrm{F}}}. \tag{4.19}$$

This pointwise dimension allows us to quantify the local variations in scaling. The object

is called fractal if $D_F \geq D_T$. For chaotic time series analysis, the trajectory in the D-dimensional phase space does not fill the space even when we have infinite number of points. This is because the trajectory lies on a strange attractor of dimension $D_F < D$.

Since we have divided the object with N spatially isolated regions of characteristic length ℓ, we ask how many times the time series visits the ith region. The number of times α takes on a value between α' and $\alpha' + d\alpha'$ will be of the form

$$d\alpha' \rho(\alpha') \ell^{-f(\alpha')}, \tag{4.20}$$

where $f(\alpha')$ is a continuous function called the **multifractal spectrum**. The exponent $f(\alpha')$ reflects the differing dimensions of the sets upon which the singularities of strength α' may lie. This expression is roughly equivalent to eq. (4.19), except that now, instead of the dimension D_F, we have a fractal dimension $f(\alpha)$, which varies with α. It therefore offers the global scaling information of the fractal object. In this way, the fractal measures is modeled by interwoven sets of fractals of different dimensions α, where $f(\alpha)$ measures their relative weights [120].

In order to determine the function $f(\alpha)$ for a given measure, we must relate it to observable properties of the measure. We relate $f(\alpha)$ to a set of dimensions which have been introduced by Hentschel and Procaccia [130], the set of "generalized dimensions" of multifractility, is defined as:

$$D_q = \lim_{\ell \to 0} \left[\frac{1}{q-1} \frac{\log \chi(q)}{\log \ell} \right], \tag{4.21}$$

where $\chi(q)$ is the **generalized correlation sum**

$$\chi(q) = \sum_i p_i^q. \tag{4.22}$$

In the above equation, $p_i(\ell)$ is the weight associated with the number of points the region i contains:

$$p_i(\ell) = \lim_{N \to \infty} \frac{N_i}{N}, \tag{4.23}$$

where N_i is the number of points in the ith region when we restrict to a randomly chosen subset of N points.

As q varies, different subsets, which are associated with different scaling indices, become dominant. Therefore, we get

$$\chi(q) = \int d\alpha' \, \rho(\alpha') \ell^{-f(\alpha')} \ell^{q\alpha'}. \tag{4.24}$$

Since ℓ is very small, the integral in eq. (4.24) will be dominated by the value of α' which makes $q\alpha' - f(\alpha')$ smallest, provided that $\rho(\alpha')$ is nonzero. Thus, we replace α' by $\alpha(q)$,

which is defined by the first order condition

$$\frac{\mathrm{d}}{\mathrm{d}\alpha'}\left[q\alpha' - f(\alpha')\right]\Big|_{\alpha'=\alpha(q)} = 0. \tag{4.25}$$

We also have

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha'^2}\left[q\alpha' - f(\alpha')\right]\Big|_{\alpha'=\alpha(q)} > 0. \tag{4.26}$$

so that

$$\begin{aligned} \frac{\mathrm{d}f(\alpha')}{\mathrm{d}\alpha'} &= q, \\ \frac{\mathrm{d}^2 f(\alpha')}{\mathrm{d}\alpha'^2} &< 0. \end{aligned} \tag{4.27}$$

It then follows from eq. (4.21) that

$$D_q = \frac{1}{q-1}\left[q\alpha(q) - f\left(\alpha(q)\right)\right]. \tag{4.28}$$

Thus, if we know $f(\alpha)$, and the spectrum of *alpha* values, we can find D_q. Alternatively, given D_q, we can find $\alpha(q)$ since

$$\alpha(q) = \frac{\mathrm{d}}{\mathrm{d}q}\left[(q-1)D_q\right]. \tag{4.29}$$

For monofractals (such as white noise) both α and $f(\alpha)$ are constant and identical; thus the $D_q - q$ plot is a horizontal straight line (with constant $D_q \equiv \alpha \equiv f(\alpha)$). For multifractals (such as the Poincaré plot of deterministic or stochastic chaos) the $D_q - q$ plot is typically a downward-sigmoid function, while $f(\alpha)$ is typically a concave-down parabolic function over a wide range of α values. The range of spread of the D_q and α values indicate the degree of complexity of the system. The maximum value of $f(\alpha)$ turns out to be the box dimension. Equivalently, a $\tau - q$ plot should be a straight line with constant slope D_q if the system is monofractal, and a nonlinear curve with decreasing slope for multifractals. The degree of nonlinearity of the graph indicates the degree of complexity of the system. The theory's predictions have been checked for a variety of experimental systems at the onset of chaos, with great success [106].

Multifractality has been useful to describe the natural irregularity of physiological systems because their fluctuation is not truly random and can be demonstrated to have spatial or temporal correlation [107]. It was reported that heart rate fluctuations of healthy individuals are multifractal [145], and that a major life-threatening condition, congestive heart failure, leads to loss of multifractality [109]. Hence it would be beneficial to uncover how multifractality in the healthy heart dynamics arises, perhaps it is primarily a consequence of the response of neuroautonomic control mechanisms to activity-related fractal stimuli. An

alternative possibility could be that the neuroautonomic control mechanisms endogenously generate multifractal dynamics.

4.7 Summary

In this chapter, we have delineated several methods from nonlinear dynamics for characterizing stochastically perturbed nonlinear systems. For approximate entropy, detrended fluctuation analysis, and power spectral analysis, we could not extract information about the induced effects of noise on the nonlinear system. Numerical titration proved to be the only one that can provide a quantitative measure of the intensity of (deterministic or stochastic) chaos. In the future, we should also employ multifractal analysis as a complementary procedure to gain some insight into the geometry of noise-induced complexity in simulated data and empirical data.

Part II

Nonlinear State Estimation

Chapter 5

Nonlinear State Estimation

> What are Nonlinear problems? It has already been mentioned that practically every problem in theoretical physics is governed by nonlinear mathematics except quantum theory, and even in quantum theory it is a rather controversial question whether it will finally be a linear or nonlinear theory... It has been argued that every nonlinear problem is really individual; that is, it requires individual methods, usually very complicated and difficult methods, and it is rather improbable that one can learn from one nonlinear problem to solve another nonlinear problem. I have to emphasize that I am certainly not an expert in this field, neither in nonlinear problems in mathematics, nor in physics; but I have... actually come across a few nonlinear problem on my way through physics, so I at least know some of the horrible difficulties and troubles which one meets in these problems.
>
> — Werner Karl Heisenberg [301]

Nonlinear filtering has been the focus of interest in the engineering community for more than 30 years [102] [146]. The problem is to estimate sequentially the state of a dynamic system using a sequence of noisy measurements made on the system. This chapter is devoted to the review the Kalman filter and the extended Kalman filter. Beyond the Kalman filter framework, we focus on the tools of **sequential Monte Carlo estimation**, collectively referred to as **particle filters**.

We adopt the state-space approach to modeling dynamic systems and we focus on the discrete-time formulation of the problem. Thus, difference equations will be used to model the evolution of the system over time, and measurements are assumed to be available at discrete times. The state-space approach to time-series modeling focuses attention on the state vector of a system. The state vector contains all relevant information required to describe the system. For example, in the cardiovascular system this information could be

related to the characteristics of heart rate dynamics. The measurement vector represents (noisy) observations that are related to the state vector. The state-space approach is convenient for handling multivariate data and nonlinear/non-Gaussian processes and it provides a significant advantage over traditional time-series techniques for these problems.

In order to analyze the make inferences about a dynamic system, at least two models are required: first, a model describing the evolution of the state with time (the system model), and second, a model relating the noisy measurements to the state (the measurement model). We shall assume that these models are available in a probabilistic form. The probabilistic state-space formulation and the requirement for the updating of information on receipt of new measurements are ideally suited for the Bayesian approach. This provides a rigorous general framework for dynamic state estimate problems.

In the Bayesian approach to dynamic state estimation one attempts to construct the *posterior* probability density function (pdf) of the state, based on all available information, including the sequence of received measurements. If either the system or measurement model is nonlinear, the posterior pdf will be non-Gaussian. Since this pdf embodies all available statistical information, it may be regarded to be the complete solution to the estimation problem. In principle, an optimal estimate of the state may be obtained from the posterior pdf. A measure of the accuracy of the estimate may also be obtained. For many problems an estimate is required every time a measurement is received. In this case a recursive filter is a convenient solution. A recursive filtering approach means that received data can be processed sequentially rather than as a batch, so that it is not necessary to store the complete data set nor to reprocess existing data if a new measurement becomes available. Such a filter consists of essentially two stage: prediction and update. The prediction stage uses the system model to predict the state pdf forward from one measurement time to the next. Since the state is usually subject to unknown disturbances, prediction generally translates, deforms, and broadens the state pdf. The update operation uses Bayes theorem, which is the mechanism for updating knowledge about the state in the light extra information from new data.

5.1 The Kalman filter

The Kalman filter is the best known filter, a simple and elegant algorithm formulated more than 40 years ago [150], as an optimal recursive Bayesian estimator for a somewhat restricted class of linear Gaussian problems. The filter is named after Rudolf E. Kalman. Stanley F. Schmidt is generally credited with developing the first implementation of a Kalman filter. It was during a visit to the NASA Ames Research Center that Kalman saw the applicability of his ideas to the problem of trajectory estimation for the Apollo program. The Kalman filter in its various forms is clearly established as a fundamental tool for analyzing and solving a broad class of estimation problems.

5.1.1 The discrete time Kalman filter

Suppose we have a linear discrete-time system given as follows:

$$\begin{aligned} \mathbf{z}_k &= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \eta_{k-1}, \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k. \end{aligned} \tag{5.1}$$

Here $\mathbf{z}_k$ is the state, $\mathbf{y}_k$ is the output. The vector $\mathbf{u}_k$ is a known, deterministic input. For simplicity, we will only examine the case when $\mathbf{u}_k \equiv \mathbf{0}$. If a nonzero input is present, we can simply add in its effect by superposition.

The noise processes $\{\eta_k\}$ and $\{\nu_k\}$ are white, zero-mean, uncorrelated, and have known covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively:

$$\begin{aligned} \langle \eta_k \eta_j^T \rangle &= \mathbf{Q}_k \delta_{kj}, \\ \langle \nu_k \nu_j^T \rangle &= \mathbf{R}_k \delta_{kj}, \\ \langle \eta_k \nu_j^T \rangle &= \mathbf{0}, \end{aligned} \tag{5.2}$$

where δ_{kj} is the Kronecker delta function.

Our goal is to estimate the state $\mathbf{z}_k$ based on our knowledge of the system dynamics and the availability of the noisy measurements $\{\mathbf{y}_k\}$. If we have all of the measurements up to and including time k available for use in our estimate of $\mathbf{z}_k$, then we can form an a posteriori estimate, which we denote as $\hat{\mathbf{z}}_k^+$. One way to form the a posteriori state estimate is to compute the expected value of $\mathbf{z}_k$ conditioned on all of the measurements up to and including time k:

$$\hat{\mathbf{z}}_k^+ = \langle \mathbf{z}_k | \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k \rangle = \text{a posteriori estimate.} \tag{5.3}$$

If we have all of the measurement before (but not including) time k available for use in our estimate of $\mathbf{z}_k$, then we can form an a priori estimate, which we denote as $\hat{\mathbf{z}}_k^-$. One way to form the a priori state estimate is to compute the expected value of $\mathbf{z}_k$ conditioned on all of the measurements before (but not including) time k:

$$\hat{\mathbf{z}}_k^- = \langle \mathbf{z}_k | \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{k-1} \rangle = \text{a priori estimate.} \tag{5.4}$$

Both $\hat{\mathbf{z}}_k^-$ and $\hat{\mathbf{z}}_k^+$ are estimates of $\mathbf{z}_k$. However, $\hat{\mathbf{z}}_k^-$ is our estimate of $\mathbf{z}_k$ *before* the measurement $\mathbf{y}_k$ is taken into account, and $\hat{\mathbf{z}}_k^+$ is our estimate of $\mathbf{z}_k$ *after* we process the measurement $\mathbf{y}_k$ at time k. We naturally expect $\hat{\mathbf{z}}_k^+$ to be a better estimate than $\hat{\mathbf{z}}_k^-$, because we use more information to compute $\hat{\mathbf{z}}_k^+$.

We define $\mathbf{\Lambda}_k$ to be the covariance of the estimation error. $\mathbf{\Lambda}_k^-$ denotes the covariance

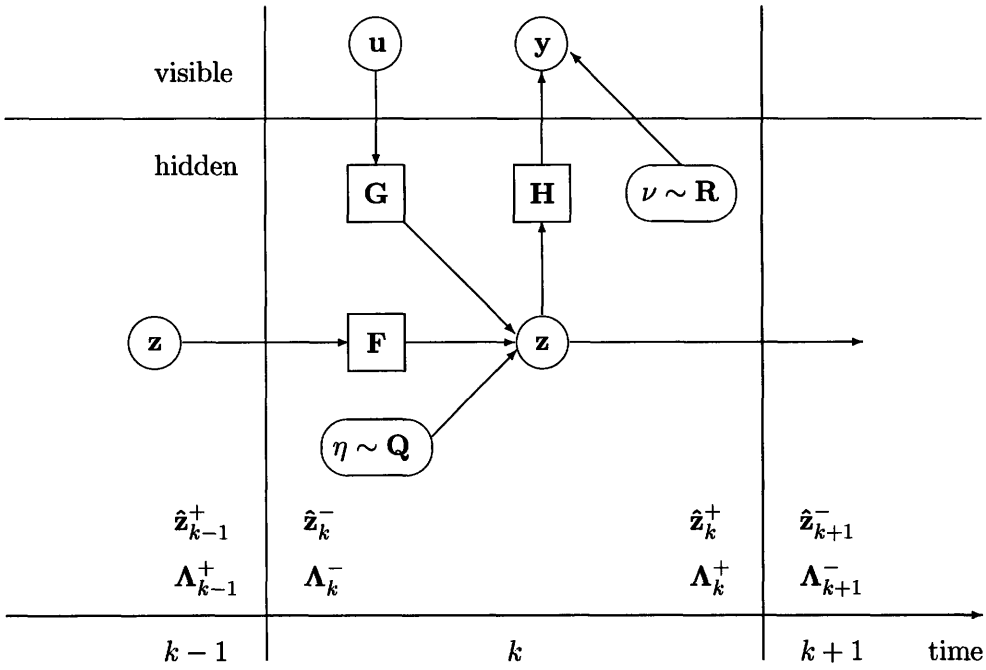


Figure 5-1: Diagram of the model underlying the Kalman filter. The Kalman filter model assumes the true state at time k is evolved from the state at time $k-1$ according to $\mathbf{z}_k = \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \eta_{k-1}$, where $\mathbf{F}_{k-1}$ is the state transition model which is applied to the previous state $\mathbf{z}_{k-1}$. The matrix $\mathbf{G}_{k-1}$ is the control-input model which is applied to the control vector $\mathbf{u}_{k-1}$. The vector η_{k-1} is the process noise which is assumed to be drawn from a zero mean Gaussian distribution with covariance $\mathbf{Q}_k$. At time k an observation $\mathbf{y}_k$ of the true state $\mathbf{z}_k$ is made according to $\mathbf{y}_k = \mathbf{H}_k\mathbf{z}_k + \nu_k$, where $\mathbf{H}_k$ is the observation model and ν_k is the observation noise which is assumed to be zero mean Gaussian white noise with covariance $\mathbf{R}_k$. The time line also shows the a priori and a posteriori state estimates and estimation error covariances.

of the estimation error of $\hat{\mathbf{z}}^-_k$, and $\mathbf{\Lambda}^+_k$ denotes the covariance of the estimation error of $\hat{\mathbf{z}}^+_k$:

$$\begin{aligned}\mathbf{\Lambda}^-_k &= \left\langle (\mathbf{z}_k - \hat{\mathbf{z}}^-_k)(\mathbf{z}_k - \hat{\mathbf{z}}^-_k)^T \right\rangle, \\ \mathbf{\Lambda}^+_k &= \left\langle (\mathbf{z}_k - \hat{\mathbf{z}}^+_k)(\mathbf{z}_k - \hat{\mathbf{z}}^+_k)^T \right\rangle. \end{aligned} \tag{5.5}$$

These relationships are depicted in Fig. 5-1. The figure shows that after we process the measurement at time $(k-1)$, we have an estimate of $\mathbf{z}_{k-1}$ (denote $\hat{\mathbf{z}}^+_{k-1}$) and the covariance of that estimate (denoted $\mathbf{\Lambda}^+_{k-1}$). When time k arrives, before we process the measurement at time k we compute an estimate of $\mathbf{z}_k$ (denote $\hat{\mathbf{z}}^-_k$) and the covariance of that estimate (denote $\mathbf{\Lambda}^-_k$). Then we process the measurement at time k to refine our estimate of $\mathbf{z}_k$. The resulting estimate of $\mathbf{z}_k$ is denoted $\hat{\mathbf{z}}^+_k$, and its covariance is denoted $\mathbf{\Lambda}^+_k$.

We begin the estimation process with $\hat{\mathbf{z}}^+_0$, our best estimate of the initial state $\mathbf{z}_0$ before any measurements are available. The first measurement is taken at time $k = 1$. Since we do not have any measurements available to estimate $\mathbf{z}_0$, it is reasonable to form $\hat{\mathbf{z}}^+_0$ as the

expected value of the initial state $\mathbf{z}_0$:

$$\hat{\mathbf{z}}_0^+ = \langle \mathbf{z}_0 \rangle . \tag{5.6}$$

The mean of $\mathbf{z}$ propagates with time as:

$$\hat{\mathbf{z}}_k^- = \mathbf{F}_{k-1}\hat{\mathbf{z}}_{k-1}^+. \tag{5.7}$$

This is called the time-update equation for $\hat{\mathbf{z}}$. From time $(k-1)^+$ to time k^-, the state estimate propagates the same way that the mean of the state propagates. This makes sense intuitively. We do not have any additional measurements available to help us update our state estimate between time $(k-1)^+$ and time k^-, so we should just update the state estimate based on our knowledge of the system dynamics.

Next we need to compute the time update equation for $\mathbf{\Lambda}$, the covariance of the state estimate error. We begin with $\mathbf{\Lambda}_0^+$, which is the covariance of our initial estimate of $\mathbf{z}_0$. If we know the initial state perfectly, then $\mathbf{\Lambda}_0^+ = \mathbf{0}$. If we have no idea of the value of $\mathbf{z}_0$, then $\mathbf{\Lambda}_0^+ = \infty\mathbf{I}$. In general, $\mathbf{\Lambda}_0^+$ represents the uncertainty in the initial estimate of $\mathbf{z}_0$:

$$\begin{aligned}\mathbf{\Lambda}_0^+ &= \left\langle (\mathbf{z}_0 - \bar{\mathbf{z}}_0)(\mathbf{z}_0 - \bar{\mathbf{z}}_0)^T \right\rangle \\ &= \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)(\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)^T \right\rangle .\end{aligned}$$

The covariance of the state of a linear discrete-time system propagates with time:

$$\mathbf{\Lambda}_k^- = \mathbf{F}_{k-1}\mathbf{\Lambda}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}. \tag{5.8}$$

This is called the time-update equation for $\mathbf{\Lambda}$.

We have derived the time-update equations for $\hat{\mathbf{z}}$ and $\mathbf{\Lambda}$. Now we need to derive the measurement-update equations for $\hat{\mathbf{z}}$ and $\mathbf{\Lambda}$. Given $\hat{\mathbf{z}}_k^-$, how should we compute $\hat{\mathbf{z}}_k^+$? Both $\hat{\mathbf{z}}_k^-$ and $\hat{\mathbf{z}}_k^+$ are estimates of $\mathbf{z}_k$, the only difference between them is that $\hat{\mathbf{z}}_k^+$ takes the measurement $\mathbf{y}_k$ into account. The availability of the measurement $\mathbf{y}_k$ changes the estimate of a constant $\mathbf{z}$ as follows:

$$\begin{aligned}\mathbf{K}_k &= \mathbf{\Lambda}_{k-1}\mathbf{H}_k^T \left(\mathbf{H}_k\mathbf{\Lambda}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \\ &= \mathbf{\Lambda}_k\mathbf{H}_k^T\mathbf{R}_k^{-1}, \\ \hat{\mathbf{z}}_k &= \hat{\mathbf{z}}_{k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_{k-1}\right), \\ \mathbf{\Lambda}_k &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Lambda}_{k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\ &= \left(\mathbf{\Lambda}_{k-1}^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\right)^{-1} \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Lambda}_{k-1},\end{aligned} \tag{5.9}$$

where $\hat{\mathbf{z}}_{k-1}$ and $\mathbf{\Lambda}_{k-1}$ are the estimate and its covariance *before* the measurement $\mathbf{y}_k$ is

processed, and $\hat{\mathbf{z}}_k$ and $\mathbf{\Lambda}_k$ are the estimate and its covariance *after* the measurement $\mathbf{y}_k$ is processed. These are the measurement-update equations for $\hat{\mathbf{z}}_k$ and $\mathbf{\Lambda}_k$.

The matrix $\mathbf{K}_k$ in the above equation is called the **Kalman filter gain**. The first expression for $\mathbf{\Lambda}_k^+$ above is called the Joseph stabilized version of the covariance measurement update equation. It was formulated by Peter Joseph in the 1960s and can be shown to be more stable and robust than the third expression for $\mathbf{\Lambda}_k^+$ [32] [52]. The first expression for $\mathbf{\Lambda}_k^+$ guarantees that $\mathbf{\Lambda}_k^+$ will always be symmetric positive definite, as long as $\mathbf{\Lambda}_k^-$ is symmetric positive definite. The third expression for $\mathbf{\Lambda}_k^+$ is computationally simpler than the first expression, but its form does not guarantee symmetric or positive definiteness for $\mathbf{\Lambda}_k^+$. The second form for $\mathbf{\Lambda}_k^+$ is rarely implemented as written above but will be useful in our derivation of the information filter in the next section.

The discrete-time Kalman filter

Here we summarize the discrete-time Kalman filter by combining the above equations.

1. The dynamic system is given by the following equations:

$$\begin{aligned}
\mathbf{z}_k &= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \eta_{k-1}, \\
\mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\
\eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\
\nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).
\end{aligned} \tag{5.10}$$

2. The Kalman filter is initialized as follows:

$$\begin{aligned}
\hat{\mathbf{z}}_0^+ &= \langle \mathbf{z}_0 \rangle, \\
\mathbf{\Lambda}_0^+ &= \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)(\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)^T \right\rangle.
\end{aligned} \tag{5.11}$$

3. For each time step $k = 1, 2, \ldots$, compute:

$$\begin{aligned}
\mathbf{\Lambda}_k^- &= \mathbf{F}_{k-1}\mathbf{\Lambda}_{k-1}^+\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \\
\mathbf{K}_k &= \mathbf{\Lambda}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{\Lambda}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\
&= \mathbf{\Lambda}_k^+\mathbf{H}_k^T\mathbf{R}_k^{-1}, \\
\hat{\mathbf{z}}_k^- &= \mathbf{F}_{k-1}\hat{\mathbf{z}}_{k-1}^+ = \text{a priori state estimate}, \\
\hat{\mathbf{z}}_k^+ &= \hat{\mathbf{z}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-) = \text{a posteriori state estimate}, \\
\mathbf{\Lambda}_k^+ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Lambda}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\
&= \left[(\mathbf{\Lambda}_k^-)^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\right]^{-1} \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Lambda}_k^-.
\end{aligned} \tag{5.12}$$

5.1.2 Information filtering

In this section, we discuss information filtering. This is an implementation of the Kalman filter that propagates the *inverse* of $\mathbf{\Lambda}$ rather than propagating $\mathbf{\Lambda}$; that is, information filtering propagates the information matrix of the system. The $\mathbf{\Lambda}$ matrix represents the uncertainty in the state estimate. If $\mathbf{\Lambda}$ is "large" then we have a lot of uncertainty in our state estimate. In the limit as $\mathbf{\Lambda} \to \mathbf{0}$ we have perfect knowledge of $\mathbf{z}$, and as $\mathbf{\Lambda} \to \infty\mathbf{I}$ we have zero knowledge. The information matrix is defined as

$$\mathcal{I} = \mathbf{\Lambda}^{-1}. \tag{5.13}$$

That is, $\mathcal{I}$ represents the certainty in the state estimate. If $\mathcal{I}$ is "large" then we have a lot of confidence in our state estimate. In the limit as $\mathcal{I} \to \mathbf{0}$ we have zero knowledge of $\mathbf{z}$, and as $\mathcal{I} \to \infty\mathbf{I}$ we have perfect knowledge of $\mathbf{z}$. If the initial uncertainty is infinite, we cannot numerically set $\mathbf{\Lambda}_0^+ = \infty\mathbf{I}$, but we can numerically set $\mathcal{I}_0^+ = \mathbf{0}$. This makes the information filter more mathematically precise for the zero initial certainty case. However, if the initial uncertainty is zero, we can numerically set $\mathbf{\Lambda}_0^+ = \mathbf{0}$, but we cannot numerically set $\mathcal{I}_0^+ = \infty\mathbf{I}$. This makes the standard Kalman filter more mathematically precise for the zero initial uncertainty case.

From eq. (5.12), that the measurement update equation for $\mathbf{\Lambda}$ can be written as

$$(\mathbf{\Lambda}_k^+)^{-1} = (\mathbf{\Lambda}_k^-)^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k. \tag{5.14}$$

Substituting the definition of $\mathcal{I}$ into this equation gives the measurement-update equation for the information matrix:

$$\mathcal{I}_k^+ = \mathcal{I}_k^- + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k. \tag{5.15}$$

Recall from eq. (5.12) the time-update for $\mathbf{\Lambda}$:

$$\mathbf{\Lambda}_k^- = \mathbf{F}_{k-1} \mathbf{\Lambda}_{k-1}^+ \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}. \tag{5.16}$$

This implies that

$$\mathcal{I}_k^- = \left[\mathbf{F}_{k-1} (\mathcal{I}_{k-1}^+)^{-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}\right]^{-1}. \tag{5.17}$$

Now we can apply the matrix inversion lemma [see eq. (A.4)] to eq. (5.17) to obtain

$$\mathcal{I}_k^- = \mathbf{Q}_{k-1}^- - \mathbf{Q}_{k-1}^- \mathbf{F}_{k-1} (\mathcal{I}_{k-1}^+ + \mathbf{F}_{k-1}^T \mathbf{Q}_{k-1}^- \mathbf{F}_{k-1})^{-1} \mathbf{F}_{k-1}^T \mathbf{Q}_{k-1}^-. \tag{5.18}$$

This gives the time-update equation for the information matrix. The information filter can be summarized as follows.

The information filter

1. The dynamic system is given by the following equations:

$$\begin{aligned}
\mathbf{z}_k &= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \eta_{k-1}, \\
\mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\
\eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\
\nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).
\end{aligned} \tag{5.19}$$

2. The information filter is initialized as follows:

$$\begin{aligned}
\hat{\mathbf{z}}_0^+ &= \langle \mathbf{z}_0 \rangle, \\
\mathcal{I}_0^+ &= \left\{ \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)(\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)^T \right\rangle \right\}^{-1}.
\end{aligned} \tag{5.20}$$

3. The information filter is given by the following equations, which are computed for each time step $k = 1, 2, \ldots$:

$$\begin{aligned}
\mathcal{I}_k^- &= \mathbf{Q}_{k-1}^- - \mathbf{Q}_{k-1}^-\mathbf{F}_{k-1}\left(\mathcal{I}_{k-1}^+ + \mathbf{F}_{k-1}^T\mathbf{Q}_{k-1}^-\mathbf{F}_{k-1}\right)^{-1}\mathbf{F}_{k-1}^T\mathbf{Q}_{k-1}^-, \\
\mathcal{I}_k^+ &= \mathcal{I}_k^- + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k, \\
\mathbf{K}_k &= (\mathcal{I}_k^+)^{-1}\mathbf{H}_k^T\mathbf{R}_k^{-1}, \\
\hat{\mathbf{z}}_k^- &= \mathbf{F}_{k-1}\hat{\mathbf{z}}_{k-1}^+, \\
\hat{\mathbf{z}}_k^+ &= \hat{\mathbf{z}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-).
\end{aligned} \tag{5.21}$$

5.2 The extended Kalman filter

The extended Kalman filter (EKF) was originally proposed by Stanley Schmidt in order to adopt the Kalman filter to nonlinear spacecraft navigation problems [246]. The EKF has become one of the most widely used nonlinear filtering algorithms.

Consider the evolution of a discrete nonlinear system in the presence of noise according to the following equations:

$$\begin{aligned}
\mathbf{z}_k &= \mathcal{F}(\mathbf{z}_{k-1}, \eta_{k-1}), \\
\mathbf{y}_k &= \mathcal{H}_k(\mathbf{z}_k, \nu_k), \\
\eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\
\nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).
\end{aligned} \tag{5.22}$$

We linearize the nonlinear system around the Kalman filter estimate by performing a Taylor series expansion of the state equation around $\mathbf{z}_{k-1} = \hat{\mathbf{z}}_{k-1}^+$ and $\eta_{k-1} = \mathbf{0}$ to obtain the

following:

$$\begin{aligned}
\mathbf{z}_k &= \mathcal{F}_{k-1}(\hat{\mathbf{z}}^+_{k-1}, \mathbf{0}) + \left[\nabla_{\mathbf{z}_k} \mathcal{F}^T_{k-1}\right]^T \Big|_{\hat{\mathbf{z}}^+_{k-1}} (\mathbf{z}_{k-1} - \hat{\mathbf{z}}^+_{k-1}) + \left[\nabla_{\eta_k} \mathcal{F}^T_{k-1}\right]^T \Big|_{\hat{\mathbf{z}}^+_{k-1}} \eta_{k-1} \\
&= \mathcal{F}_{k-1}(\hat{\mathbf{z}}^+_{k-1}, \mathbf{0}) + \mathbf{F}_{k-1}(\mathbf{z}_{k-1} - \hat{\mathbf{z}}^+_{k-1}) + \mathbf{L}_{k-1}\eta_{k-1} \\
&= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \left[\mathcal{F}_{k-1}(\hat{\mathbf{z}}^+_{k-1}, \mathbf{0}) - \mathbf{F}_{k-1}\hat{\mathbf{z}}^+_{k-1}\right] + \mathbf{L}_{k-1}\eta_{k-1} \\
&= \mathbf{F}_{k-1}x_{k-1} + \tilde{\mu}_{k-1} + \tilde{\eta}_{k-1},
\end{aligned} \tag{5.23}$$

where $\mathbf{F}_{k-1} = \left[\nabla_{\mathbf{z}_k} \mathcal{F}^T_{k-1}\right]^T \Big|_{\hat{\mathbf{z}}^+_{k-1}}$ and $\mathbf{L}_{k-1} = \left[\nabla_{\eta_k} \mathcal{F}^T_{k-1}\right]^T \Big|_{\hat{\mathbf{z}}^+_{k-1}}$ and $\nabla_{\mathbf{z}_k} = \left[\frac{\partial}{\partial \mathbf{z}_{ki}} \cdots \frac{\partial}{\partial \mathbf{z}_{kN}}\right]^T$, with $\mathbf{z}_{ki}$, being the ith component of vector $\mathbf{z}_k$. The signals $\tilde{\mu}_k$ and $\tilde{\eta}_k$ are defined as

$$\begin{aligned}
\tilde{\mu}_k &= \mathcal{F}_k(\hat{\mathbf{z}}^+_k, \mathbf{0}) - \mathbf{F}_k\hat{\mathbf{z}}^+_k, \\
\tilde{\eta}_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{L}_k\mathbf{Q}_k\mathbf{L}^T_k).
\end{aligned} \tag{5.24}$$

We linearize the measurement equation around $\mathbf{z}_k = \hat{\mathbf{z}}^-_k$ and $\nu_k = \mathbf{0}$ to obtain

$$\begin{aligned}
\mathbf{y}_k &= \mathcal{H}_k(\hat{\mathbf{z}}^-_{k-1}, \mathbf{0}) + \left[\nabla_{\mathbf{z}_k} \mathcal{H}^T_k\right]^T \Big|_{\hat{\mathbf{z}}^-_k} (\mathbf{z}_k - \hat{\mathbf{z}}^-_k) + \left[\nabla_{\nu_k} \mathcal{H}^T_k\right]^T \Big|_{\hat{\mathbf{z}}^-_k} \nu_k \\
&= \mathcal{H}_k(\hat{\mathbf{z}}^-_{k-1}, \mathbf{0}) + \mathbf{H}_k(\mathbf{z}_k - \hat{\mathbf{z}}^-_k) + \mathbf{M}_k\nu_k \\
&= \mathbf{H}_k\mathbf{z}_k + \left[\mathcal{H}_k(\hat{\mathbf{z}}^-_k, \mathbf{0}) - \mathbf{H}_k\hat{\mathbf{z}}^-_k\right] + \mathbf{M}_k\nu_k \\
&= \mathbf{H}_k\mathbf{z}_k + \tilde{\kappa}_k + \tilde{\nu}_k,
\end{aligned} \tag{5.25}$$

where $\mathbf{H}_k = \left[\nabla_{\mathbf{z}_k} \mathcal{H}^T_k\right]^T \Big|_{\hat{\mathbf{z}}^-_k}$ and $\mathbf{M}_k = \left[\nabla_{\nu_k} \mathcal{H}^T_k\right]^T \Big|_{\hat{\mathbf{z}}^-_k}$. The signals $\tilde{\kappa}_k$ and $\tilde{\nu}_k$ are defined as

$$\begin{aligned}
\tilde{\kappa}_k &= \mathcal{H}_k(\hat{\mathbf{z}}^-_k, \mathbf{0}) - \mathbf{H}_k\hat{\mathbf{z}}^-_k, \\
\tilde{\nu}_k &\sim \mathcal{N}\left(\mathbf{0}, \mathbf{M}_k\mathbf{R}_k\mathbf{M}^T_k\right).
\end{aligned} \tag{5.26}$$

We have a linear state-space system in eq. (5.23) and a linear measurement in eq. (5.25). That means we can use the standard Kalman filter equations to estimate the state. This results in the following equations for the discrete-time extended Kalman filter.

$$\begin{aligned}
\mathbf{\Lambda}^-_k &= \mathbf{F}_{k-1}\mathbf{\Lambda}^+_{k-1}\mathbf{F}^T_{k-1} + \mathbf{L}_{k-1}\mathbf{Q}_{k-1}\mathbf{L}^T_{k-1}, \\
\mathbf{K}_k &= \mathbf{\Lambda}^-_k\mathbf{H}^T_k\left(\mathbf{H}_k\mathbf{\Lambda}^-_k\mathbf{H}^T_k + \mathbf{M}_k\mathbf{R}_k\mathbf{M}^T_k\right)^{-1}, \\
\hat{\mathbf{z}}^-_k &= \mathcal{F}_{k-1}(\hat{\mathbf{z}}^+_{k-1}, \mathbf{0}), \\
\mathbf{z}_k &= \mathcal{H}_k(\hat{\mathbf{z}}^-_k, \mathbf{0}) - \mathbf{H}_k\hat{\mathbf{z}}^-_k, \\
\hat{\mathbf{z}}^+_k &= \hat{\mathbf{z}}^-_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}^-_k - \mathbf{z}_k) \\
&= \hat{\mathbf{z}}^-_k + \mathbf{K}_k\left[\mathbf{y}_k - \mathcal{H}_k(\hat{\mathbf{z}}^-_k, \mathbf{0})\right], \\
\mathbf{\Lambda}^+_k &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{\Lambda}^-_k.
\end{aligned} \tag{5.27}$$

The discrete-time extended Kalman filter

1. The system and measurement equations are given as follows:

$$\begin{aligned}
\mathbf{z}_k &= \mathcal{F}(\mathbf{z}_{k-1}, \eta_{k-1}), \\
\mathbf{y}_k &= \mathcal{H}_k(\mathbf{z}_k, \nu_k), \\
\eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\
\nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).
\end{aligned} \tag{5.28}$$

2. Initialize the filter as follows:

$$\begin{aligned}
\hat{\mathbf{z}}_0^+ &= \langle \mathbf{z}_0 \rangle, \\
\mathbf{\Lambda}_0^+ &= \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)(\mathbf{z}_0 - \hat{\mathbf{z}}_0^+)^T \right\rangle.
\end{aligned} \tag{5.29}$$

3. For $k = 1, 2, \ldots$, perform the following.

   (a) Compute the following partial derivative matrices:

$$\begin{aligned}
\mathbf{F}_{k-1} &= \left[\nabla_{\mathbf{z}_k} \mathcal{F}_{k-1}^T \right]^T \Big|_{\hat{\mathbf{z}}_{k-1}^+}, \\
\mathbf{L}_{k-1} &= \left[\nabla_{\eta_k} \mathcal{F}_{k-1}^T \right]^T \Big|_{\hat{\mathbf{z}}_{k-1}^+}.
\end{aligned} \tag{5.30}$$

   (b) Perform the time update of the state estimate and error covariance as follows:

$$\begin{aligned}
\mathbf{\Lambda}_k^- &= \mathbf{F}_{k-1} \mathbf{\Lambda}_{k-1}^+ \mathbf{F}_{k-1}^T + \mathbf{L}_{k-1} \mathbf{Q}_{k-1} \mathbf{L}_{k-1}^T, \\
\hat{\mathbf{z}}_k^- &= \mathcal{F}_{k-1}(\hat{\mathbf{z}}_{k-1}^+, \mathbf{0}).
\end{aligned} \tag{5.31}$$

   (c) Compute the following partial derivative matrices:

$$\begin{aligned}
\mathbf{H}_k &= \left[\nabla_{\mathbf{z}_k} \mathcal{H}_k^T \right]^T \Big|_{\hat{\mathbf{z}}_k^-}, \\
\mathbf{M}_k &= \left[\nabla_{\nu_k} \mathcal{H}_k^T \right]^T \Big|_{\hat{\mathbf{z}}_k^-}.
\end{aligned} \tag{5.32}$$

   (d) Perform the measurement update of the state estimate and error covariance:

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{\Lambda}_k^- \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{\Lambda}_k^- \mathbf{H}_k^T + \mathbf{M}_k \mathbf{R}_k \mathbf{M}_k^T \right)^{-1}, \\
\hat{\mathbf{z}}_k^+ &= \hat{\mathbf{z}}_k^- + \mathbf{K}_k \left[\mathbf{y}_k - \mathcal{H}_k(\hat{\mathbf{z}}_k^-, \mathbf{0}) \right], \\
\mathbf{\Lambda}_k^+ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{\Lambda}_k^-.
\end{aligned} \tag{5.33}$$

   Note that other equivalent expressions can be used for $\mathbf{K}_k$ and $\mathbf{\Lambda}_k^+$, as is apparent from eq. (5.12).

Taylor series approximation leading to the EKF makes gross simplification of the probabilistic specification of the model. With nonlinear models, the probability distributions of interest tend to be multimodal. Linear approximation in such cases could miss the rich structure. We address this issue in the following section.

5.3 The particle filter

Particle filters had their origin in the work of Metropolis and Ulam in 1949 [199][1], in which they proposed studying systems by investigating the properties of sets of particles (point mass representation of probability densities) rather than the properties of the probability densities. He used the analogy of the card game of solitaire. The probability of success in a game of solitaire may be analytically intractable. Obviously a more practical approach would be to produce a large number of experiments and then to examine the relative proportion of successes.

In the automatic control community the ideas were introduced in the late 1960s by Handschin and Mayne [122]; Handschin [121], but they were largely overlooked and ignored; only since the 1980s has computational power been adequate in their implementation. Even now it is the computational burden of the particle filter that is its primary obstacle to more widespread use. The particle filter is a statistical, brute-force approach to estimation that often works for problems that are difficult for the conventional Kalman filter (i.e., systems that are highly nonlinear and non-Gaussian). Particle filtering goes by many other names, including sequential importance sampling [66], bootstrap filtering [110], the condensation algorithm [143] [144] [182], interacting particle approximation [201], Monte Carlo filtering [158], sequential Monte Carlo (SMC) filtering [8] [53], and survival of the fittest [151].

The major contribution to the development of sequential Monte Carlo method was the inclusion of the resampling step [110], which, coupled with ever faster computers, made the particle filters useful in practice for the first time. Since then, the particle filters have become one of the most popular methods for stochastic dynamic estimation problems, including tracking problems [240], demodulation of communication signals [7], estimation of ecological parameters and populations [25], image processing [22], neural network training [81], fault detection [80], speech recognition [292], pattern recognition [267], system identification [148] [249], and separation of chaotic signals [116] [177]. Particle filtering is a growing area of research with many unexplored avenues and applications. Some of the more important areas of open research include the avoidance of sample impoverishment, methods for determining how many particles are required for a given problem, convergence results [53], application to control and parameter estimation [8] [201], connections with genetic algorithms [133], real-time implementation issues [168], and hardware implementations of parallel particle filters.

[1]Norbert Wiener suggested something much like particle filtering as early as 1940.

This chapter presents the particle filter, which is a completely nonlinear state estimator. Of course, there is no free lunch. The price that must be paid for the high performance of the particle filter is an increased level of computational effort. The improved performance of the particle filter may or may not be worth the increased computational effort. The trade-offs are problem dependent. The particle filter is a probability-based estimator. Therefore, in Sec. 5.3.1, we will discuss the Bayesian approach to state estimation. This will provide a foundation for the derivation of the particle filter in Sec. 5.3.2.

5.3.1 Bayesian state estimation

In this section, we will briefly discuss the Bayesian approach to state estimation. This is based on Bayes' Rule. This section is based on the presentation given in Ref. [110], which is similar to many other books and papers on the subject of Bayesian estimation [66] [240]. A general parametric nonlinear state space model, with additive noise processes is given by:

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathcal{F}_k(\mathbf{z}_k, \eta_k), \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{z}_k, \nu_k). \end{aligned} \tag{5.34}$$

The functions $\mathcal{F}_k(\cdot)$ and $\mathcal{H}_k(\cdot)$ are time-varying nonlinear system and measurement equations. The noise sequences $\{\eta_k\}$ and $\{\nu_k\}$ are assumed to be independent and white with known pdf's. The goal of the estimator is to approximate the conditional pdf of $\mathbf{z}_k$ based on measurements $\{\mathbf{y}_k\}$. This conditional pdf is denoted as

$$p(\mathbf{z}_k|\mathbf{Y}_k) = \text{pdf of} \, \mathbf{z}_k \text{ conditioned on measurements } \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k. \tag{5.35}$$

The first measurement is obtained at $k = 1$, so the initial condition of the estimator is the pdf of $\mathbf{z}_0$, which can be written as

$$p(\mathbf{z}_0) = p(\mathbf{z}_0|\mathbf{Y}_0), \tag{5.36}$$

since $\mathbf{Y}_0$ is defined as the set of no measurements. Once we compute $p(\mathbf{z}_k|\mathbf{Y}_k)$ then we can estimate $\mathbf{z}_k$ in whatever way we think is most appropriate. The conditional pdf $p(\mathbf{z}_k|\mathbf{Y}_k)$ may be multimodal, in which case we may not want to use the mean of $\mathbf{z}_k$ as our estimate. For example, suppose that the conditional pdf is computed as shown in Fig. 5-2. In this case, the mean of $\mathbf{z}$ is zero, but there is zero probability that $\mathbf{z}$ is equal to zero, so we may not want to use zero as our estimate of $\mathbf{z}$. Instead we might want to use fuzzy logic and say that $\hat{\mathbf{z}} = \pm 2$, each with a level of membership equal to 0.5 [175].

Our goal is to find a recursive way to compute the conditional pdf $p(\mathbf{z}_k|\mathbf{Y}_k)$. Before we find this conditional pdf, we will find the conditional pdf $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$. This is the pdf of $\mathbf{z}_k$

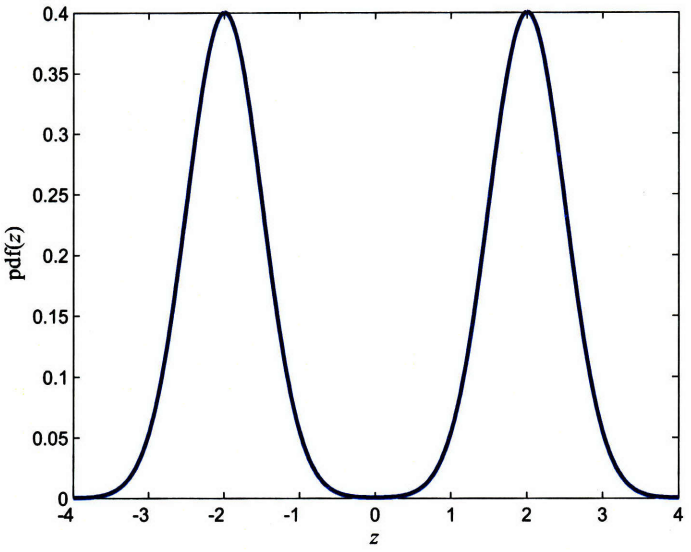


Figure 5-2: An example of a multimodal probability density function. What single number should be used as an estimate of $\mathbf{z}$?

given all measurements *prior to* time k. We can write the pdf as

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{Y}_{k-1}) &= \int d\mathbf{z}_{k-1}\, p(\mathbf{z}_k, \mathbf{z}_{k-1}|\mathbf{Y}_{k-1}) \\ &= \int d\mathbf{z}_{k-1}\, p(\mathbf{z}_k|\mathbf{z}_{k-1}, \mathbf{Y}_{k-1}) p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1}). \end{aligned} \tag{5.37}$$

But notice from our system description in eq. (5.34) that $\mathbf{z}_k$ is entirely determined by $\mathbf{z}_{k-1}$ and η_{k-1}; therefore $p(\mathbf{z}_k, \mathbf{z}_{k-1}|\mathbf{Y}_{k-1}) = p(\mathbf{z}_k|\mathbf{z}_{k-1})$ and we that

$$p(\mathbf{z}_k|\mathbf{Y}_{k-1}) = \int d\mathbf{z}_{k-1}\, p(\mathbf{z}_k|\mathbf{z}_{k-1}) p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1}). \tag{5.38}$$

The second pdf on the right side of the above equation is not available yet, but it is available at the initial time [see eq. (5.36)]. Later in this section we will see how to compute it recursively. The first pdf on the right side of the above equation is available. The pdf $p(\mathbf{z}_k|\mathbf{z}_{k-1})$ is simply the pdf of the state a time k given a specific state at time $(k-1)$. We know this pdf because we know the system equation $\mathcal{F}_k(\cdot)$ and we know the pdf of the noise η_k. For example, suppose that the system equation is given as $\mathbf{z}_{k+1} = \mathbf{z}_k + \eta_k$ and suppose that $\mathbf{z}_{k-1} = 1$ and η_{k-1} is uniformly distributed on $[-1, 1]$. Then the pdf $p(\mathbf{z}_k|\mathbf{z}_{k-1})$ is uniformly distributed on $[0, 2]$.

Now consider the a posteriori conditional pdf of $\mathbf{z}_k$. We can write this pdf as

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{Y}_k) &= \frac{p(\mathbf{Y}_k|\mathbf{z}_k)}{p(\mathbf{Y}_k)} p(\mathbf{z}_k) \\ &= \frac{p(\mathbf{y}_k, \mathbf{Y}_{k-1}|\mathbf{z}_k)}{p(\mathbf{y}_k, \mathbf{Y}_{k-1})} \underbrace{\frac{p(\mathbf{z}_k|\mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1})}{p(\mathbf{Y}_{k-1}|\mathbf{z}_k)}}_{p(\mathbf{z}_k)} \\ &= \frac{p(\mathbf{z}_k, \mathbf{y}_k, \mathbf{Y}_{k-1})}{p(\mathbf{z}_k)p(\mathbf{y}_k, \mathbf{Y}_{k-1})} \frac{p(\mathbf{z}_k, \mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1})}{p(\mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1}|\mathbf{z}_k)}. \end{aligned} \tag{5.39}$$

We can multiply both the numerator and denominator of this equation by $p(\mathbf{z}_k, \mathbf{y}_k)$ to obtain

$$p(\mathbf{z}_k|\mathbf{Y}_k) = \frac{p(\mathbf{z}_k, \mathbf{y}_k, \mathbf{Y}_{k-1})p(\mathbf{z}_k, \mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1})}{p(\mathbf{z}_k)p(\mathbf{y}_k, \mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1}|\mathbf{z}_k)} \frac{p(\mathbf{z}_k, \mathbf{y}_k)}{p(\mathbf{z}_k, \mathbf{y}_k)}. \tag{5.40}$$

Now we use the ratios of various joint pdfs to marginal pdfs in the above equation to obtain conditional pdfs. This gives

$$p(\mathbf{z}_k|\mathbf{Y}_k) = \frac{p(\mathbf{Y}_{k-1}|\mathbf{z}_k, \mathbf{y}_k)p(\mathbf{y}_k|\mathbf{z}_k)p(\mathbf{z}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})p(\mathbf{Y}_{k-1}|x_k)}. \tag{5.41}$$

Note that $\mathbf{y}_k$ is a function of $\mathbf{z}_k$, so $p(\mathbf{Y}_{k-1}|\mathbf{z}_k, \mathbf{y}_k) = p(\mathbf{Y}_{k-1}|\mathbf{z}_k)$. These two terms cancel in the above equation and we obtain

$$p(\mathbf{z}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{z}_k)p(\mathbf{z}_k|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k|\mathbf{Y}_{k-1})}. \tag{5.42}$$

All of the pdf's on the right side of the above equation are available. The pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ is available from our knowledge of the measurement equation $\mathcal{H}_k(\cdot)$ and our knowledge of the pdf of the measurement noise ν_k. The pdf $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$ is available from eq. (5.38). Finally, the pdf $p(\mathbf{y}_k|\mathbf{Y}_{k-1})$ is obtained as follows:

$$\begin{aligned} p(\mathbf{y}_k|\mathbf{Y}_{k-1}) &= \int d\mathbf{z}_k \, p(\mathbf{y}_k, \mathbf{z}_k|\mathbf{Y}_{k-1}) \\ &= \int d\mathbf{z}_k \, p(\mathbf{y}_k|\mathbf{z}_k, \mathbf{Y}_{k-1})p(\mathbf{z}_k|\mathbf{Y}_{k-1}). \end{aligned} \tag{5.43}$$

But $\mathbf{y}_k$ is completely determined by $\mathbf{z}_k$ and ν_k, so $p(\mathbf{y}_k|\mathbf{z}_k, \mathbf{Y}_{k-1}) = p(\mathbf{y}_k|\mathbf{z}_k)$ and

$$p(\mathbf{y}_k|\mathbf{Y}_{k-1}) = \int d\mathbf{z}_k \, p(\mathbf{y}_k|\mathbf{z}_k)p(\mathbf{z}_k|\mathbf{Y}_{k-1}). \tag{5.44}$$

Both of the pdf's on the right side of the above equation are available as discussed above. $p(\mathbf{y}_k|\mathbf{z}_k)$ is available from our knowledge of the measurement equation $\mathcal{H}(\cdot)$ and the pdf of ν_k, and $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$ is available from eq. (5.38).

Summarizing the development above, the recursive equations of the Bayesian state estimation filter can be summarized as follows.

The recursive Bayesian state estimator

1. The system and measurement equations are given as follows:

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathcal{F}_k(\mathbf{z}_k, \eta_k), \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{z}_k, \eta_k). \end{aligned} \tag{5.45}$$

2. Assuming that the pdf of the initial state $p(\mathbf{z}_0)$ is known, initialize the estimator follows:

$$p(\mathbf{z}_0|\mathbf{Y}_0) = p(\mathbf{z}_0). \tag{5.46}$$

3. For $k = 1, 2, \ldots$, perform the following.

   (a) The a priori pdf is obtained from eq. (5.38).

$$p(\mathbf{z}_k|\mathbf{Y}_{k-1}) = \int \mathrm{d}\mathbf{z}_{k-1}\, p(\mathbf{z}_k|\mathbf{z}_{k-1}) p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1}). \tag{5.47}$$

   (b) The a posteriori pdf is obtained from eq. (5.42) and (5.44).

$$p(\mathbf{z}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{z}_k) p(\mathbf{z}_k|\mathbf{Y}_{k-1})}{\int \mathrm{d}\mathbf{z}_k\, p(\mathbf{y}_k|\mathbf{z}_k) p(\mathbf{z}_k|\mathbf{Y}_{k-1})}. \tag{5.48}$$

Analytical solutions to these equations are available only for a few special cases. In particular, if $\mathcal{F}(\cdot)$ and $\mathcal{H}(\cdot)$ are linear, and $\mathbf{z}_0$, $\{\eta_k\}$, and $\{\nu_k\}$ are additive, independent, and Gaussian, then the solution is the Kalman filter discussed in Sec. 5.1. The Bayesian derivation of the Kalman filter can be found in many references [136] [237] [266]; they are more complicated than the least squares approach that we used in Sec. 5.1. When the Kalman filter is derived this way, then no conclusions can be drawn about the optimality of the filter when the noise is not Gaussian. In fact, other optimal (non-Kalman) filters have been derived for other noise distributions [252]. Nevertheless, the Bayesian derivation proves that when the noise is Gaussian, the Kalman filter is the optimal filter. However, the least squares derivation that we used in Sec. 5.1 shows that the Kalman filter is the optimal *linear* filter, regardless of the pdf of the noise.

5.3.2 Particle filtering

The basic ideas behind particle filtering are intuitive and straightforward. Specifically, since in general we can't carry around complete probability densities, instead we carry around point masses, or *particles*, drawn from these distributions. The key, then, is to develop methods for generating these particles and for propagating them through the two general filtering equations: we use eq. (5.47) to simulate the Markov dynamics to generate samples of the a priori pdf, $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$ at time k, given samples of the updated density $p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1})$

at time $k-1$; and we use eq. (5.48) to provide weights on the samples of the predicted density $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$ at time k, to produce a weighted set of samples for the updated density $p(\mathbf{z}_k|\mathbf{Y}_k)$ at time k.

To begin, we generate a given number, N, of particles, $\mathbf{z}_0^{(1)}, \mathbf{z}_0^{(2)}, \ldots, \mathbf{z}_0^{(N)}$, based on the initial pdf $p(\mathbf{z}_0)$ before any measurements are taken (here we assume that the first observation is at time 1). This provides an approximate distribution $p(\mathbf{z}_0)$, with equally weighted particles. At each time step $k = 1, 2, \ldots$, we propagate the particles to the next time step using the process equation $\mathcal{F}(\cdot)$:

$$\mathbf{z}_k^{(i)} = \mathcal{F}_{k-1}\left(\mathbf{z}_{k-1}^{(i)}, \eta_{k-1}^{(i)}\right) \qquad \text{for } i = 1, 2, \ldots, N, \tag{5.49}$$

where each $\eta_{k-1}^{(i)}$ noise vector is randomly generated on the basis of the know pdf of η_{k-1}.

The observations $\mathbf{y}_k$ at time k tells us something about how likely each of our predicted particles is. Therefore, we use these likelihoods to put weights on these particles. Specifically, suppose the at time k we have N particles $\mathbf{z}_k^{(1)}, \mathbf{z}_k^{(2)}, \ldots, \mathbf{z}_k^{(N)}$, with associated weights, $w_k^{(1)}, w_k^{(2)}, \ldots, w_k^{(N)}$. Using this approximation for the a priori pdf, $p(\mathbf{z}_k|\mathbf{Y}_{k-1})$ in eq. (5.47), we find that we directly obtain an approximation for the updated density at time k with the same particles but with updated weights:

$$\hat{p}(\mathbf{z}_k|\mathbf{Y}_k) = \sum_{i=1}^{N} w_k^{(i)} \delta\left(\mathbf{z} - \mathbf{z}_k^{(i)}\right), \tag{5.50}$$

$$w_k^{(i)} \propto w_{k-1}^{(i)} p\left(\mathbf{y}_k|\mathbf{z}_k^{(i)}\right). \tag{5.51}$$

As discussed in the previous section, this can be done if we know the nonlinear measurement equation and the pdf of the measurement noise. For example, if an m-dimensional measurement equation is given as $\mathbf{y}_k = \mathcal{H}(\mathbf{z}_k) + \nu_k$ and $\nu_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ then a relative likelihood $w_k^{(i)}$ that the measurement is equal to a specific measurement $\mathbf{y}^*$, given the premise that $\mathbf{z}_k$ is equal to the particle $\mathbf{z}_k^{(i)}$, can be computed as follows:

$$\begin{aligned} w_k^{(i)} &= \Pr\left[(\mathbf{y}_k = \mathbf{y}^*)|\left(\mathbf{z}_k = \mathbf{z}_k^{(i)}\right)\right] = \Pr\left[\nu_k = \mathbf{y}^* - \mathcal{H}\left(\mathbf{z}_k^{(i)}\right)\right] \\ &\propto \frac{1}{(2\pi)^{m/2}|\mathbf{R}|^{1/2}} \exp\left(\frac{-\left[\mathbf{y}^* - \mathcal{H}\left(\mathbf{z}_k^{(i)}\right)\right]^T \mathbf{R}^{-1}\left[\mathbf{y}^* - \mathcal{H}\left(\mathbf{z}_k^{(i)}\right)\right]}{2}\right). \end{aligned} \tag{5.52}$$

Now we normalize the relative likelihoods obtained in eq. (5.52) as follows.

$$\bar{w}^{(i)} = \frac{w^{(i)}}{\sum_{j=1}^{N} w^{(j)}}. \tag{5.53}$$

This ensures that the likelihoods sum to unity.

The problem encountered by this approach is that, as the filtering process proceeds,

the distribution of the importance weights $w^{(i)}$ becomes more and more skewed, that is the variance of the importance weights increases over time and thus the estimate will finally diverge. Practically, after a few time steps, only one particle has a non-zero importance weight. The algorithm, consequently, fails to represent the posterior distributions of interest adequately. To avoid this degeneracy, one needs to introduce an additional selection step, namely one that can be thought of as resampling from the approximation to the updated density, thus obtaining a new set of N, equally weighted particles. If we use the point approximation in eq. (5.50), this corresponds to making a set of N independent choices from the set of particles $\mathbf{z}_k^{(1)}, \mathbf{z}_k^{(2)}, \ldots, \mathbf{z}_k^{(N)}$, where the ith of these is chosen with probability $w_k^{(i)}$, which in turn implies that we will, mostly likely, *replicate* some of the particles and not pick some of the others. One straightforward (but not necessarily efficient) way to do this is the following [240]. For $i = 1, \ldots, N$, perform the following two steps.

1. Generate a random number r that is uniformly distributed on $[0, 1]$.
2. Accumulate the likelihoods $w^{(i)}$ into a sum, one at a time, until the accumulated sum is greater than r. That is, $\sum_{m=1}^{j-1} w^{(m)} < r$ but $\sum_{m=1}^{j} w^{(m)} \geq r$. The new particle $\mathbf{z}_k^{(i)}$ is then set equal to the old particle $\mathbf{z}_k^{(j)}$.

This resampling idea is formally justified in [263], where it is shown that the ensemble pdf of the new particles $\mathbf{z}_k^{(i)}$ tends to the pdf $p(\mathbf{z}_k|\mathbf{y}_k)$ as the number of samples N approaches ∞. The resampling step can be summarized as follows:

$$\mathbf{z}_k^{(i)} = \mathbf{z}_k^{(j)} \text{ with probability } w^{(j)} \qquad \text{for } i, j = 1, \ldots, N. \tag{5.54}$$

This is illustrated in Fig. 5-3.

The drawback of this approach is **sample impoverishment**, that is, if we don't choose some of the particles, we don't have the same amount of diversity right after resampling. This occurs when the region of state space in which the pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ has significant values does not overlap with the pdf $p(\mathbf{z}_k|\mathbf{y}_{k-1})$. This means, that if all of our a priori particles are distributed according to $p(\mathbf{z}_k|\mathbf{y}_{k-1})$, and we then use the computed pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ to resample the particles, only a few particles will be resampled to become a posteriori particles. This is because only a few of the a priori particles will be in a region of state space where the computed pdf $p(\mathbf{y}_k|\mathbf{z}_k)$ has a significant value. Eventually, all of the particles will collapse to the same value. Thus, we want to do this *only* when we have to, namely when the weights begin to deviate drastically from the best case in which they are all equal (to $1/N$). There is a very simple measure of the degeneracy of the weights in the density approximation, which can be thought of as the "effective" number of degrees of freedom in our samples:

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} \left(w^{(i)}\right)^2}. \tag{5.55}$$

Note that $N_{\text{eff}} = N$ in the best case scenario, and $N_{\text{eff}} = 1$ in the worst case in which one of

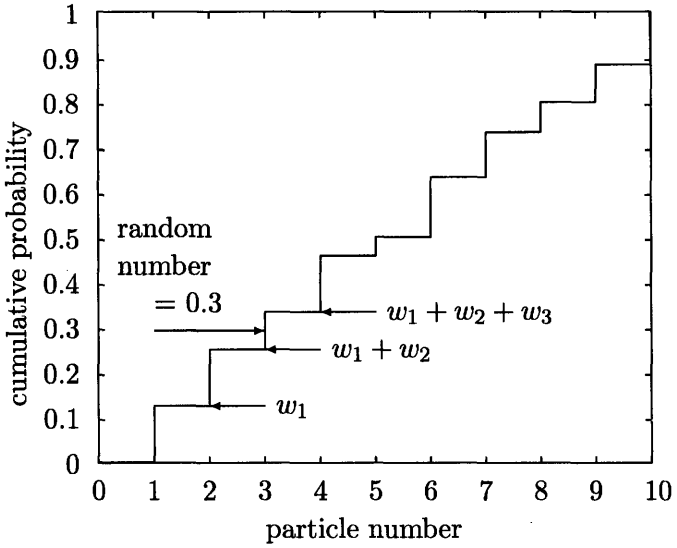


Figure 5-3: Illustration of resampling in the particle filter. For example, if a random number $r = 0.3$ is generated (from a distribution that is uniform on $[0, 1]$), the smallest value of j for which $\sum_{m=1}^{j} w^{(m)} \geq r$ is $j = 3$. Therefore the resampled particle is set equal to $\mathbf{z}_k^{(3)}$.

the weights has value 1 and all others are zero. Thus we would like to keep $1/\sum_{i=1}^{N} \left(w^{(i)}\right)^2$ as small as possible. Consequently, a widely used method is to set a threshold on this quantity, to keep going without resampling as long as this quantity remains below the threshold, and to resample (reset weights to $1/N$) only when this quantity exceeds the threshold. The particle-based recursive approximation of the pdf and resampling step is illustrated in Fig. 5-4.

The computational effort of the particle filter is often a bottleneck to its implementation. With this in mind, more efficient resampling methods can be implemented, such as order statistics [41], stratified sampling and residual sampling [176], and systematic resampling [158]. Other ways of resampling have also been proposed [204]. In some cases, additional logic must be incorporated to maintain a constant sample size.

Now we have a set of particles $\mathbf{z}_k^{(i)}$ that are distributed according to

$$\hat{p}(\mathbf{z}_k|\mathbf{y}_k) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(\mathbf{z}_k - \mathbf{z}_k^{(i)}\right). \tag{5.56}$$

We can compute any desired statistical measure of this pdf. For example, if we want to compute the expected value $\langle f(\mathbf{z}_k)|\mathbf{y}_k\rangle$ then we can approximate it as the algebraic mean of the particles:

$$\langle (f(\mathbf{z}_k)|\mathbf{y}_k)\rangle \approx \int \mathrm{d}\xi\, f(\xi)\hat{p}(\xi) = \frac{1}{N}\sum_{i=1}^{N} f\left(\mathbf{z}_k^{(i)}\right), \tag{5.57}$$

which, under fairly general conditions, converges to the true expectation as $N \to \infty$.

The particle filter

1. The system and measurement equations are given as follows:

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathcal{F}_k(\mathbf{z}_k, \eta_k), \\ \mathbf{y}_k &= \mathcal{H}_k(\mathbf{z}_k, \nu_k), \end{aligned} \tag{5.58}$$

   where $\{\eta_k\}$ and $\{\nu_k\}$ are independent white noise processes with known pdf's.

2. Assume that the pdf of the initial state $p(\mathbf{z}_0)$ is known, randomly generate N initial particles on the basis of the pdf $p(\mathbf{z}_0)$. These particles are denoted $\mathbf{z}_0^{(i)}$. The parameter N is chosen as a trade-off between computational effort and estimation accuracy.

3. For $k = 1, 2, \ldots$, do the following.

   (a) Perform the time propagation step to obtain a priori particles $\mathbf{z}_k^{(i)}$ using the known process equation and the known pdf of the process noise:

$$\mathbf{z}_k^{(i)} = \mathcal{F}_{k-1}\left(\mathbf{z}_{k-1}^{(i)}, \eta_{k-1}^{(i)}\right) \qquad \text{for } i = 1, \ldots, N, \tag{5.59}$$

   where each $\eta_{k-1}^{(i)}$ noise vector is randomly generated from the known pdf of η_{k-1}.

   (b) Compute the relative likelihood $w^{(i)}$ of each particle $\mathbf{z}_k^{(i)}$ conditioned on the measurement $\mathbf{y}_k$. This is done by evaluating the pdf $p\left(\mathbf{y}_k|\mathbf{z}_k^{(i)}\right)$ on the basis of the nonlinear measurement equation and the pdf of the measurement noise.

   (c) Scale the relative likelihoods obtained in the previous step as follows:

$$w^{(i)} = \frac{w^{(i)}}{\sum_{j=1}^{N} w^{(j)}}. \tag{5.60}$$

   Now the sum of all the likelihoods is equal to one.

   (d) Generate a set of a posteriori particles $\mathbf{z}_{k,i}^{+}$ on the basis of the relative likelihoods $w^{(i)}$. Then reset all the weights to $1/N$. This is called **Sampling Importance Resampling (SIR)**. However, to prevent sample impoverishment, only resample when the effective number of samples is less than a threshold N_{th},

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} \left(w^{(i)}\right)^2} < N_{\text{th}}. \tag{5.61}$$

   The threshold can be chosen as $N_{\text{th}} = 2N/3$.

   (e) Now that we have a set of particles $\mathbf{z}_k^{(i)}$ that are distributed according to the pdf $p(\mathbf{z}_k|\mathbf{y}_k)$, we can compute any desired statistical measure of this pdf. We typically are most interested in computing the mean and the covariance.

5.4 Summary

In this chapter, we gave a brief account of some common approaches to nonlinear filtering. In a linear system with Gaussian noise, the Kalman filter is optimal. The extended Kalman filter operate in the framework of Gaussian approximation for the posterior density. While this makes it simple to implement and fast to execute, it suffers from an inherent inability to model higher-order moments of truly non-Gaussian posterior densities. The particle filter overcomes this limitation by approximating the entire posterior density. The Monte Carlo estimation approach has a serious drawback: it is characterized by an exponentially growing computational complexity that prevents their widespread use in practice.

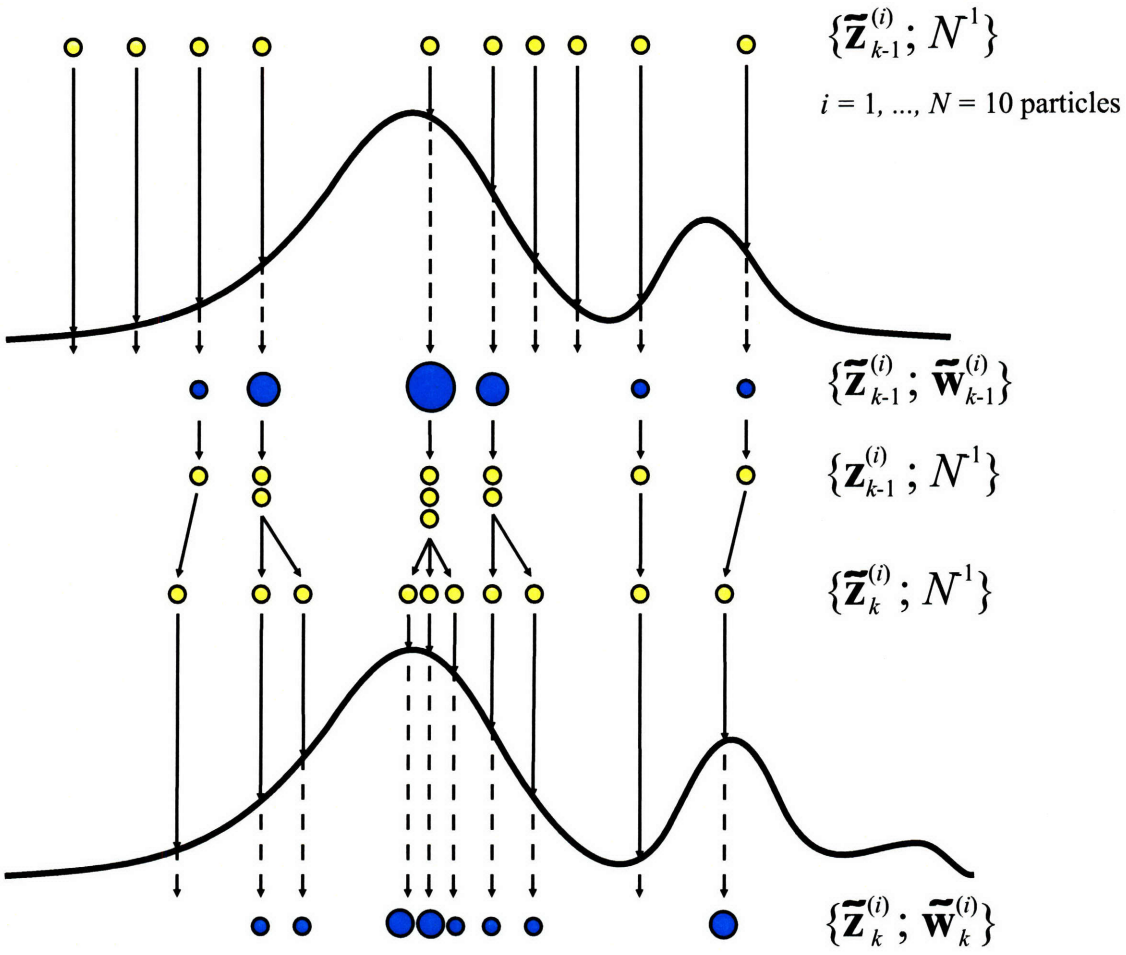


Figure 5-4: A particle filter starts at time $k-1$ with an unweighted measure $\left\{\tilde{\mathbf{z}}_{k-1}^{(i)}, N^{-1}\right\}$, which provides an approximation of $p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-2})$. For each particle we compute the importance weights using the information at time $k-1$. This results in the weighted measure $\left\{\tilde{\mathbf{z}}_{k-1}^{(i)}, \tilde{w}_{k-1}^{(i)}\right\}$, which yields an approximation of $p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1})$. Subsequently, a resampling step selects only the "fittest" particles to obtain the unweighted measure $\left\{\mathbf{z}_{k-1}^{(i)}, N^{-1}\right\}$. This yields an approximation of $p(\mathbf{z}_{k-1}|\mathbf{Y}_{k-1})$ that is "concentrated" on the most likely hypothesis, thereby allowing for nonstationary tracking. Finally, a prediction step introduces variety, resulting in the measure $\left\{\tilde{\mathbf{z}}_{k}^{(i)}, N^{-1}\right\}$, which is an approximation of $p(\mathbf{z}_{k}|\mathbf{Y}_{k-1})$. Figure adopted from Ref. [80].

Chapter 6

Additional Topics in Nonlinear Filtering

> The use of wrong a priori statistics in the design of a Kalman filter can lead to large estimation errors or even to a divergence of errors.
>
> — Raman Mehra [197]

The previous chapter covered the essentials of nonlinear filtering. This chapter discusses some additional important topics related to nonlinear state estimation. Sec. 6.1 talks about how to improve our state estimate via optimal smoothing. In fixed-interval smoothing we seek an estimate of the state at some of the interior points of the time interval. Sec. 6.1.1 discusses the forward-backward approach to smoothing, which is perhaps the most straightforward smoothing algorithm. Sec. 6.1.2 discusses the RTS smoother, which is conceptually more difficult but is computationally cheaper than forward-backward smoothing. Sec. 6.1.3 discusses the generalized fixed interval smoother.

Sec. 6.2 covers an important problem in system identification—parameter estimation. System parameters are considered to be constant and a nonlinear state estimator can be adapted to estimate the system state and parameters simultaneously. Sec. 6.3 discusses how to verify that a filter is operating reliably. Sec. 6.4 discusses multiple-model estimation, which is a way of estimating system states when we are not sure of which model is governing the dynamics of the system.

6.1 Optimal smoothing

In the previous chapter, we discussed how to obtain the optimal a priori and a posteriori state estimates. If we have measurements after time k available for use in our estimate of $\mathbf{z}_k$,

then we can form a *smoothed* estimate. We compute the expected value of $\mathbf{z}_k$ conditioned on all of the measurements that are available:

$$\hat{\mathbf{z}}_{k|k+N} = \langle \mathbf{z}_k | \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k, \ldots, \mathbf{y}_{k+N} \rangle = \text{smoothed estimate}, \tag{6.1}$$

where N is some positive integer. Here we defined $\hat{\mathbf{z}}_{k|j}$ as the estimate of $\mathbf{z}_k$ given all measurements up to and including time j. If we have more measurements, it stands to reason that we should be able to get an even better estimate of $\mathbf{z}_k$.

6.1.1 Forward-backward smoothing

The forward-backward approach obtains two estimates of $\mathbf{z}_m$. The first estimate, $\hat{\mathbf{z}}_f$, is based on the standard Kalman filter that operates from $k = 1$ to $k = m$. The second estimate, $\hat{\mathbf{z}}_b$, is based on a Kalman filter that runs backward in time from $k = N$ back to $k = m$. The forward-backward approach to smoothing combines the two estimates to form an optimal smoothed estimate. This approach was first suggested by Fraser and Potter [79].

Suppose that we combine a forward estimate $\hat{\mathbf{z}}_f$ of the state and a backward estimate $\hat{\mathbf{z}}_b$ of the state to get a smoothed estimate of $\mathbf{z}$ as follows:

$$\hat{\mathbf{z}} = \mathbf{K}_f \hat{\mathbf{z}}_f + \mathbf{K}_b \hat{\mathbf{z}}_b, \tag{6.2}$$

where $\mathbf{K}_f$ and $\mathbf{K}_b$ are constant matrix coefficients to be determined. Note that $\hat{\mathbf{z}}_f$ and $\hat{\mathbf{z}}_b$ are both unbiased since they are both outputs from Kalman filters. Therefore, if $\hat{\mathbf{z}}$ is to be unbiased, we require $\mathbf{K}_f + \mathbf{K}_b = \mathbf{I}$. This gives

$$\hat{\mathbf{z}} = \mathbf{K}_f \hat{\mathbf{z}}_f + (\mathbf{I} - \mathbf{K}_f)\hat{\mathbf{z}}_b. \tag{6.3}$$

The covariance of the estimate can then be found as

$$\begin{aligned}
\mathbf{\Lambda} &= \left\langle (\mathbf{z} - \hat{\mathbf{z}})(\mathbf{z} - \hat{\mathbf{z}})^T \right\rangle \\
&= \left\langle [\mathbf{z} - \mathbf{K}_f \hat{\mathbf{z}}_f - (\mathbf{I} - \mathbf{K}_f)\hat{x}_b] [\mathbf{z} - \mathbf{K}_f \hat{\mathbf{z}}_f - (\mathbf{I} - \mathbf{K}_f)\hat{x}_b]^T \right\rangle \\
&= \left\langle [\mathbf{K}_f(\varepsilon_f - \varepsilon_b) + \varepsilon_b] [\mathbf{K}_f(\varepsilon_f - \varepsilon_b) + \varepsilon_b]^T \right\rangle \\
&= \left\langle \mathbf{K}_f(\varepsilon_f \varepsilon_f^T + \varepsilon_b \varepsilon_b^T)\mathbf{K}_f^T + \varepsilon_b \varepsilon_b^T - \mathbf{K}_f \varepsilon_b \varepsilon_b^T - \varepsilon_b \varepsilon_b^T \mathbf{K}_f^T \right\rangle,
\end{aligned} \tag{6.4}$$

where $\varepsilon_f = \mathbf{z} - \mathbf{z}_f, \varepsilon_b = \mathbf{z} - \mathbf{z}_b$, and $\left\langle \varepsilon_f \varepsilon_b^T \right\rangle = 0$. The estimates $\hat{\mathbf{z}}_f$ and $\hat{\mathbf{z}}_b$ are both unbiased, and ε_f and ε_b are independent (since they depend on separate sets of measurements). We can minimize the trace of $\mathbf{\Lambda}$ with respect to $\mathbf{K}_f$:

$$\begin{aligned}
\frac{\partial \mathrm{Tr}(\mathbf{\Lambda})}{\partial \mathbf{K}_f} &= 2 \left\langle \mathbf{K}_f \left(\varepsilon_f \varepsilon_f^T + \varepsilon_b \varepsilon_b^T \right) - \varepsilon_b \varepsilon_b^T \right\rangle \\
&= 2[\mathbf{K}_f(\mathbf{\Lambda}_f + \mathbf{\Lambda}_b) - \mathbf{\Lambda}_b],
\end{aligned} \tag{6.5}$$

where $\mathbf{\Lambda}_f = \left\langle \varepsilon_f \varepsilon_f^T \right\rangle$ is the covariance of the forward estimate, and $\mathbf{\Lambda}_b = \left\langle \varepsilon_b \varepsilon_b^T \right\rangle$ is the covariance of the backward estimate. The optimality condition gives:

$$\begin{aligned} \mathbf{K}_f &= \mathbf{\Lambda}_b (\mathbf{\Lambda}_f + \mathbf{\Lambda}_b)^{-1}, \\ \mathbf{K}_b &= \mathbf{\Lambda}_f (\mathbf{\Lambda}_f + \mathbf{\Lambda}_b)^{-1}. \end{aligned} \tag{6.6}$$

We can substitute this result into eq. (6.4) to find the covariance of the fixed-interval smoother (after some simplification) as follows:

$$\mathbf{\Lambda} = (\mathbf{\Lambda}_f^{-1} + \mathbf{\Lambda}_b^{-1})^{-1}. \tag{6.7}$$

These results form the basis for the fixed-interval smoothing problem. Suppose we want a smoothed estimate at time index m. First we run the forward Kalman filter normally, using the measurements up to and including time m.

1. Initialize the forward filter as follows:

$$\begin{aligned} \hat{\mathbf{z}}_{f0}^+ &= \langle \mathbf{z}_0 \rangle, \\ \mathbf{\Lambda}_{f0}^+ &= \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_{f0}^+)(\mathbf{z}_0 - \hat{\mathbf{z}}_{f0}^+)^T \right\rangle. \end{aligned} \tag{6.8}$$

2. For $k = 1, \ldots, m$, perform the following:

$$\begin{aligned} \mathbf{\Lambda}_{fk}^- &= \mathbf{F}_{k-1} \mathbf{\Lambda}_{f,k-1}^+ \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \\ \mathbf{K}_{fk} &= \mathbf{\Lambda}_{fk}^- \mathbf{H}_k^T (\mathbf{H}_k \mathbf{\Lambda}_{fk}^- \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ &= \mathbf{\Lambda}_{fk}^+ \mathbf{H}_k^T \mathbf{R}_k^{-1}, \\ \hat{\mathbf{z}}_{fk}^- &= \mathbf{F}_{k-1} \hat{\mathbf{z}}_{f,k-1}^+, \\ \hat{\mathbf{z}}_{fk}^+ &= \hat{\mathbf{z}}_{fk}^- + \mathbf{K}_{fk} (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{z}}_{fk}^-), \\ \mathbf{\Lambda}_{fk}^+ &= (\mathbf{I} - \mathbf{K}_{fk} \mathbf{H}_k) \mathbf{\Lambda}_{fk}^-. \end{aligned} \tag{6.9}$$

At this point we have a forward estimate for $\mathbf{z}_m$, along with its covariance.

The backward filter needs to run backwards in time, starting at the final time index N. Since the forward and backward estimates must be independent, none of the information that was used in the forward filter is allowed to be used in the backward filter. Therefore, $\mathbf{\Lambda}_{bN}^-$ must be infinite:

$$\mathbf{\Lambda}_{bN}^- = \infty \mathbf{I}. \tag{6.10}$$

We are using the minus superscript on $\mathbf{\Lambda}_{bN}^-$ to indicate the backward covariance at time N before the measurement at time N is processed. (Recall that the filtering is performed backward in time.) So $\mathbf{\Lambda}_{bN}^-$ will be updated to obtain $\mathbf{\Lambda}_{bN}^+$ after the measurement at time N is processed. Then it will be extrapolated backward in time to obtain $\mathbf{\Lambda}_{b,N-1}^-$, and so on.

Now the question arises how to initialize the backward state estimate $\hat{\mathbf{z}}_{bk}^-$ at the final time $k = N$. We can solve this problem by introducing the new variable

$$\varrho_k = \mathbf{\Lambda}_{bk}^{-1} \hat{\mathbf{z}}_{bk}. \tag{6.11}$$

A minus or plus superscript can be added on all the quantities in the above equation to indicate values before or after the measurement at time k is taken into account. Since $\mathbf{\Lambda}_{bN}^- = \infty$ it follows that

$$\varrho_k^- = \mathbf{0}. \tag{6.12}$$

The infinite boundary condition on $\mathbf{\Lambda}_{bk}^-$ means that we cannot run the standard Kalman filter backward in time because we have to begin with an infinite covariance. Instead we run the information filter backward in time. This can be done by writing the system as

$$\begin{aligned}
\mathbf{z}_{k-1} &= \mathbf{F}_{k-1}^{-1}\mathbf{z}_k + \mathbf{F}_{k-1}^{-1}\eta_{k-1} \\
&= \mathbf{F}_{k-1}^{-1}\mathbf{z}_k + \eta_{b,k-1}, \\
\mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\
\eta_{bk} &\sim \mathcal{N}(\mathbf{0}, \mathbf{F}_k^{-1}\mathbf{Q}_k\mathbf{F}_k^{-T}), \\
\nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).
\end{aligned} \tag{6.13}$$

Note that $\mathbf{F}_k^{-1}$ should always exist for a real system since $\mathbf{F}_k$ comes from a matrix exponential that is always invertible. The backward information filter can be written as follows.

1. Initialize the filter as follows:

$$\begin{aligned}
\varrho_N^- &= \mathbf{0}, \\
\mathcal{I}_{bN}^- &= \mathbf{0}.
\end{aligned} \tag{6.14}$$

2. For $k = N, N-1, \ldots, m+1$, perform the following:

$$\begin{aligned}
\mathcal{I}_{bk}^+ &= \mathcal{I}_{bk}^- + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k, \\
\varrho_{bk}^+ &= \mathbf{z}_{bk}^- + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{y}_k, \\
\mathcal{I}_{b,k-1}^- &= \left[\mathbf{F}_{k-1}^{-1}(\mathcal{I}_{bk}^+)^{-1}\mathbf{F}_{k-1}^{-T} + \mathbf{F}_{k-1}^{-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T}\right]^{-1} \\
&= \mathbf{F}_{k-1}^T\left[(\mathcal{I}_{bk}^+)^{-1} + \mathbf{Q}_{k-1}\right]^{-1}\mathbf{F}_{k-1} \\
&= \mathbf{F}_{k-1}^T\left[\mathbf{Q}_{k-1}^{-1} - \mathbf{Q}_{k-1}^{-1}(\mathcal{I}_{bk}^+ + \mathbf{Q}_{k-1}^{-1})^{-1}\mathbf{Q}_{k-1}^{-1}\right]\mathbf{F}_{k-1}, \\
\varrho_{k-1}^- &= \mathcal{I}_{b,k-1}^-\mathbf{F}_{k-1}^{-1}(\mathcal{I}_{bk}^+)^{-1}\varrho_k^+.
\end{aligned} \tag{6.15}$$

3. Perform one final time update to obtain the backward estimate of $\mathbf{z}_m$:

$$\begin{aligned}
\mathcal{I}_{bm}^{-} &= \mathbf{Q}_m^{-1} - \mathbf{Q}_m^{-1}\mathbf{F}_m^{-1}(\mathcal{I}_{b,m+1}^{+} + \mathbf{F}_m^{-T}\mathbf{Q}_m^{-1}\mathbf{F}_m^{-1})^{-1}\mathbf{F}_m^{T}\mathbf{Q}_m^{-1}, \\
\mathbf{\Lambda}_{bm}^{-} &= (\mathcal{I}_{bm}^{-})^{-1}, \\
\varrho_m^{-} &= \mathcal{I}_{bm}^{-}\mathbf{F}_m^{-1}(\mathcal{I}_{b,m+1}^{+})^{-1}\varrho_{m+1}^{+}, \\
\hat{\mathbf{z}}_{bm}^{-} &= (\mathcal{I}_{bm}^{-})^{-1}\varrho_m^{-}.
\end{aligned} \tag{6.16}$$

Now we have the backward estimate $\hat{\mathbf{z}}_{bm}^{-}$ and its covariance $\mathbf{\Lambda}_{bm}^{-}$. These quantities are obtained from measurements $m+1, m+2, \ldots, N$.

The first form for $\mathcal{I}_{b,k-1}^{-}$ above requires the inversion of $\mathcal{I}_{b,k}^{+}$. Consider the first time step for the backward filter (i.e., at $k = N$). The information matrix $\mathcal{I}_{bN}^{-}$ is initialize to zero, and then the first time through the loop we set $\mathcal{I}_{bk}^{+} = \mathcal{I}_{bk}^{-} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k$. If there are fewer measurements than states, $\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k$ will always be singular. Therefore, the first form given above for $\mathcal{I}_{b,k-1}^{-}$ will not be computable. In practice we can initialize $\mathcal{I}_{bN}^{-}$ to a small nonzero matrix. The third form for $\mathcal{I}_{b,k-1}^{-}$ above has its own problems. It does not require the inversion of $\mathcal{I}_{bk}^{+}$, but it does require the inversion of $\mathbf{Q}_{k-1}$. Again, in practice we can make a small modification to $\mathbf{Q}_{k-1}$ so that it is numerically nonsingular.

After we obtain the backward quantities as outlined above, we combine them with the forward quantities from eq. (6.9) to obtain the final state estimate and covariance:

$$\begin{aligned}
\mathbf{K}_f &= \mathbf{\Lambda}_{bm}^{-}(\mathbf{\Lambda}_{fm}^{+} + \mathbf{\Lambda}_{bm}^{-})^{-1}, \\
\hat{\mathbf{z}}_m &= \mathbf{K}_f\hat{\mathbf{z}}_{fm}^{+} + (\mathbf{I} - \mathbf{K}_f)\hat{\mathbf{z}}_{bm}^{-}, \\
\mathbf{\Lambda}_m &= \left[(\mathbf{\Lambda}_{fm}^{+})^{-1} + (\mathbf{\Lambda}_{bm}^{-})^{-1}\right]^{-1}.
\end{aligned} \tag{6.17}$$

We can obtain an alternative equation for $\hat{\mathbf{z}}_m$ by manipulating the above equations. If we substitute for $\mathbf{K}_f$ in the above expression, apply the matrix inversion lemma [eq. (A.4)], and perform some other manipulation for $\hat{\mathbf{z}}_m$ then we obtain

$$\begin{aligned}
\hat{\mathbf{z}}_m &= \mathbf{\Lambda}_m\mathcal{I}_{fm}^{+}\hat{\mathbf{z}}_{fm}^{+} + \mathbf{\Lambda}_m\mathcal{I}_{bm}^{-}\hat{\mathbf{z}}_{bm}^{-} \\
&= \mathbf{\Lambda}_m(\mathcal{I}_{fm}^{+}\hat{\mathbf{z}}_{fm}^{+} + \mathcal{I}_{bm}^{-}\hat{\mathbf{z}}_{bm}^{-}).
\end{aligned} \tag{6.18}$$

Fig. 6-1 illustrates how the forward-backward smoother works.

6.1.2 RTS smoothing

The other common smoother was presented by Rauch, Tung, and Striebel, usually called the RTS smoother [230]. the RTS smoother is more computationally efficient than the smoother presented in the previous section because we do not need to directly compute the backward estimate or covariance in order to get the smoothed estimate and covariance. In

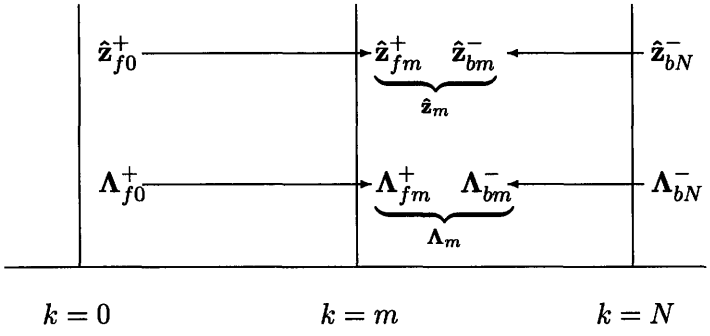


Figure 6-1: This figure illustrates the concept of the forward-backward smoother. The forward filter is run to obtain a posteriori estimates and covariances up to time m. Then the backward filter is run to obtain a priori estimates and covariances back to time m (i.e., a priori from a reversed time perspective). Then the forward and backward estimates and covariances at time m are combined to obtain the final estimate $\hat{\mathbf{z}}_m$ and covariance $\mathbf{\Lambda}_m$.

order to obtain the RTS smoother, we will first look at the smoothed covariance given in eq. (6.17) and obtain an equivalent expression that does not use $\mathbf{\Lambda}_{bm}$. Then we will look at the smoothed estimate given in eq. (6.17), which uses the gain $\mathbf{K}_f$, which depends on $\mathbf{\Lambda}_{bm}$, and obtain an equivalent expression that does not use $\mathbf{\Lambda}_{bm}$ or $\hat{\mathbf{z}}_{bm}$.

RTS covariance update

First consider the smoothed covariance given in eq. (6.17). This can be written as

$$\begin{aligned}\mathbf{\Lambda}_m &= \left[(\mathbf{\Lambda}_{fm}^{+})^{-1} + (\mathbf{\Lambda}_{bm}^{-})^{-1}\right]^{-1} \\ &= \mathbf{\Lambda}_{fm}^{+} - \mathbf{\Lambda}_{fm}^{+}(\mathbf{\Lambda}_{fm}^{+} + \mathbf{\Lambda}_{bm}^{-})^{-1}\mathbf{\Lambda}_{fm}^{+},\end{aligned} \tag{6.19}$$

where the second expression comes from an application of the matrix inversion lemma to the first expression. From eq. (6.15) we see that

$$\mathbf{\Lambda}_{bm}^{-} = \mathbf{F}_m^{-1}\left[\mathbf{\Lambda}_{b,m+1}^{+} + \mathbf{Q}_m\right]\mathbf{F}_m^{-T}. \tag{6.20}$$

Substituting this into the expression $(\mathbf{\Lambda}_{fm}^{+} + \mathbf{\Lambda}_{bm}^{-})^{-1}$ gives the following:

$$\begin{aligned}(\mathbf{\Lambda}_{fm}^{+} + \mathbf{\Lambda}_{bm}^{-})^{-1} &= \left[\mathbf{\Lambda}_{fm}^{+} + \mathbf{F}_m^{-1}(\mathbf{\Lambda}_{b,m+1}^{+} + \mathbf{Q}_m)\mathbf{F}_m^{-T}\right]^{-1} \\ &= \left[\mathbf{F}_m^{-1}\mathbf{F}_m\mathbf{\Lambda}_{fm}^{+}\mathbf{F}_m^{T}\mathbf{F}_m^{-T} + \mathbf{F}_m^{-1}(\mathbf{\Lambda}_{b,m+1}^{+} + \mathbf{Q}_m)\mathbf{F}_m^{-T}\right]^{-1} \\ &= \left[\mathbf{F}_m^{-1}(\mathbf{F}_m\mathbf{\Lambda}_{fm}^{+}\mathbf{F}_m^{T} + \mathbf{\Lambda}_{b,m+1}^{+} + \mathbf{Q}_m)\mathbf{F}_m^{-T}\right]^{-1} \\ &= \mathbf{F}_m^{T}(\mathbf{F}_m\mathbf{\Lambda}_{fm}^{+}\mathbf{F}_m^{T} + \mathbf{\Lambda}_{b,m+1}^{+} + \mathbf{Q}_m)^{-1}\mathbf{F}_m \\ &= \mathbf{F}_m^{T}(\mathbf{\Lambda}_{f,m+1}^{-} + \mathbf{\Lambda}_{b,m+1}^{+})^{-1}\mathbf{F}_m.\end{aligned} \tag{6.21}$$

From eq. (5.15) and (6.15) recall that

$$\begin{aligned}\mathcal{I}_{fm}^{+} &= \mathcal{I}_{fm}^{-} + \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m, \\ \mathcal{I}_{bm}^{+} &= \mathcal{I}_{bm}^{-} + \mathbf{H}_m^T \mathbf{R}_m^{-1} \mathbf{H}_m. \end{aligned} \tag{6.22}$$

We can combine these two equations to obtain

$$\mathcal{I}_{b,m+1}^{+} = \mathcal{I}_{b,m+1}^{-} + \mathcal{I}_{f,m+1}^{+} - \mathcal{I}_{f,m+1}^{-}. \tag{6.23}$$

Substituting this into eq. (6.17) gives

$$\begin{aligned}\mathbf{\Lambda}_{m+1} &= \left[\mathcal{I}_{f,m+1}^{+} + \mathcal{I}_{b,m+1}^{-}\right]^{-1} \\ &= \left[\mathcal{I}_{b,m+1}^{+} + \mathcal{I}_{f,m+1}^{-}\right]^{-1}, \\ \mathbf{\Lambda}_{m+1}^{-1} &= \mathcal{I}_{b,m+1}^{+} + \mathcal{I}_{f,m+1}^{-}, \\ \mathbf{\Lambda}_{b,m+1}^{-1} &= \left[\mathbf{\Lambda}_{m+1}^{-1} - \mathcal{I}_{f,m+1}^{-}\right]^{-1}. \end{aligned} \tag{6.24}$$

Substituting this into eq. (6.21) and applying the matrix inversion lemma [eq. (A.4)] gives

$$(\mathbf{\Lambda}_{fm}^{+} + \mathbf{\Lambda}_{bm}^{-})^{-1} = \mathbf{F}_m^T \mathcal{I}_{f,m+1}^{-} (\mathbf{\Lambda}_{f,m+1}^{-} - \mathbf{\Lambda}_{m+1}) \mathcal{I}_{f,m+1}^{-} \mathbf{F}_m. \tag{6.25}$$

Substituting this expression into eq. (6.19) gives

$$\mathbf{\Lambda}_m = \mathbf{\Lambda}_{fm}^{+} - \mathbf{K}_m (\mathbf{\Lambda}_{f,m+1}^{-} - \mathbf{\Lambda}_{m+1}) \mathbf{K}_m^T, \tag{6.26}$$

where the smoother gain $\mathbf{K}_m$ is given as

$$\mathbf{K}_m = \mathbf{\Lambda}_{fm}^{+} \mathbf{F}_m^T \mathcal{I}_{f,m+1}^{-}. \tag{6.27}$$

The covariance update equation for $\mathbf{\Lambda}_m$ is not a function of the backward covariance. The smoother covariance $\mathbf{\Lambda}_m$ can be solved by using only the forward covariance $\mathbf{\Lambda}_{fm}$, which reduces the computational effort (compared to the algorithm presented in Sec. 6.1.1).

RTS state estimate update

Next we consider the smoothed estimate $\hat{\mathbf{z}}_m$ given in eq. (6.17). We will find an equivalent expression that does not use $\mathbf{\Lambda}_{bm}$ or $\hat{\mathbf{z}}_{bm}$. Starting with the expression for ϱ_{k-1}^{-} in eq. (6.15),

and substituting the expression for $\mathcal{I}_{b,k-1}^{-}$ in eq. (6.15) gives

$$\begin{aligned}
\varrho_{k-1}^{-} &= \mathcal{I}_{bk}^{-}\mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
&= \mathbf{F}_{k-1}^{T}\left[\mathbf{Q}_{k-1}^{-1} - \mathbf{Q}_{k-1}^{-1}(\mathcal{I}_{bk}^{+} + \mathbf{Q}_{k-1}^{-1})^{-1}\mathbf{Q}_{k-1}^{-1}\right]\mathbf{F}_{k-1}\mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
&= \mathbf{F}_{k-1}^{T}\mathbf{Q}_{k-1}^{-1}\left[\mathbf{I} - (\mathcal{I}_{bk}^{+} + \mathbf{Q}_{k-1}^{-1})^{-1}\mathbf{Q}_{k-1}^{-1}\right]\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
&= \mathbf{F}_{k-1}^{T}\mathbf{Q}_{k-1}^{-1}(\mathcal{I}_{bk}^{+} + \mathbf{Q}_{k-1}^{-1})(\mathcal{I}_{bk}^{+} + \mathbf{Q}_{k-1}^{-1} - \mathbf{Q}_{k-1}^{-1})\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
&= \mathbf{F}_{k-1}^{T}\mathbf{Q}_{k-1}^{-1}(\mathcal{I}_{bk}^{+} + \mathbf{Q}_{k-1}^{-1})^{-1}\varrho_{k}^{+} \\
&= \mathbf{F}_{k-1}^{T}(\mathbf{I} + \mathcal{I}_{bk}^{+}\mathbf{Q}_{k-1})^{-1}\varrho_{k}^{+}.
\end{aligned} \tag{6.28}$$

Rearranging this equation, and premultiplying both sides by $\mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}$, gives

$$\mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\mathbf{F}_{k-1}^{-T}\varrho_{k-1}^{-} + \mathbf{F}_{k-1}^{-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T}\varrho_{k-1}^{-} = \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+}. \tag{6.29}$$

Substituting for $\mathbf{F}_{k-1}^{-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T}$ from eq. (B.1) gives

$$\begin{aligned}
\mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\mathbf{F}_{k-1}^{-T}\varrho_{k-1}^{-} + \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{fk}^{-}\mathbf{F}_{k-1}^{-T}\varrho_{k-1}^{-} - \mathbf{\Lambda}_{f,k-1}^{+}\varrho_{k-1}^{-} &= \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
\left[\mathbf{F}_{k-1}^{-1}(\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})\mathbf{F}_{k-1}^{-T} - \mathbf{\Lambda}_{f,k-1}^{+}\right]\varrho_{k-1}^{-} &= \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+} \\
\left[(\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})\mathbf{F}_{k-1}^{-T} - \mathbf{F}_{k-1}\mathbf{\Lambda}_{f,k-1}^{+}\right]\varrho_{k-1}^{-} &= \mathbf{\Lambda}_{bk}^{+}\varrho_{k}^{+}.
\end{aligned} \tag{6.30}$$

Substituting in this expression for $\mathbf{\Lambda}_{bk}^{+}$ from eq. (B.4) gives

$$\left[(\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})\mathbf{F}_{k-1}^{-T} - \mathbf{F}_{k-1}\mathbf{\Lambda}_{f,k-1}^{+}\right]\varrho_{k-1}^{-} = (\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})\mathcal{I}_{fk}^{-}\mathbf{\Lambda}_{k}\varrho_{k}^{+}. \tag{6.31}$$

Substituting for $(\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})$ from eq. (B.6) on both sides of this expression gives

$$\left[\mathbf{F}_{k-1}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-}) - \mathbf{F}_{k-1}\mathbf{\Lambda}_{f,k-1}^{+}\right]\varrho_{k-1}^{-} = \mathbf{F}_{k-1}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})\mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\mathbf{\Lambda}_{k}\varrho_{k}^{+}. \tag{6.32}$$

Premultiplying both sides by $(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1}\mathbf{F}_{k-1}^{-1}$, and then substituting eq. (B.14) for $(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1}$ gives

$$\varrho_{k-1}^{-} - \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}(\mathbf{\Lambda}_{fk}^{-} - \mathbf{\Lambda}_{k})\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}\mathbf{\Lambda}_{f,k-1}^{+}\varrho_{k-1}^{-} = \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\mathbf{\Lambda}_{k}\varrho_{k}^{+}. \tag{6.33}$$

Now from eq. (B.14) we see that

$$-(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1}\mathbf{F}_{k-1}^{-1}\hat{\mathbf{z}}_{fk}^{-} = \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}(\mathbf{\Lambda}_{k} - \mathbf{\Lambda}_{fk}^{-})\mathcal{I}_{fk}^{-}\hat{\mathbf{z}}_{fk}^{-}. \tag{6.34}$$

So we can add the two sides of this equation to the two sides of eq. (6.33) to get

$$\varrho_{k-1}^- - \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_k) \mathcal{I}_{fk}^- \mathbf{F}_{k-1} \mathbf{\Lambda}_{f,k-1}^+ \varrho_{k-1}^- - (\mathbf{\Lambda}_{f,k-1}^+ + \mathbf{\Lambda}_{b,k-1}^-)^{-1} \mathbf{F}_{k-1}^{-1} \hat{\mathbf{z}}_{fk}^-$$
$$= \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- \mathbf{\Lambda}_k \varrho_k^+ + \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\mathbf{\Lambda}_k - \mathbf{\Lambda}_{fk}^-) \mathcal{I}_{fk}^- \hat{\mathbf{z}}_{fk}^-. \tag{6.35}$$

Now use eq. (B.9) to substitute for $\mathbf{\Lambda}_k \varrho_k^+$ in the above equation (after some rearranging) and obtain

$$\varrho_{k-1}^- - \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_k) \mathcal{I}_{fk}^- \mathbf{F}_{k-1} \mathbf{\Lambda}_{f,k-1}^+ \varrho_{k-1}^- +$$
$$\big[- (\mathbf{\Lambda}_{f,k-1}^+ + \mathbf{\Lambda}_{b,k-1}^-)^{-1} \mathbf{F}_{k-1}^{-1} + \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- \mathbf{\Lambda}_k \mathcal{I}_{fk}^+ - \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- \mathbf{\Lambda}_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k -$$
$$\mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- \mathbf{\Lambda}_k \mathcal{I}_{fk}^- \big] \hat{\mathbf{z}}_{fk}^- = \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{fk}^-). \tag{6.36}$$

From eq. (B.15) we see that $\mathcal{I}_{fk}^+ - \mathcal{I}_{fk}^- = \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k$. Also note that part of the coefficient of $\hat{\mathbf{z}}_{fk}^-$ on the left side of the above equation can be expressed as

$$(\mathbf{\Lambda}_{f,k-1}^+ + \mathbf{\Lambda}_{b,k-1}^-)^{-1} \mathbf{F}_{k-1}^{-1} = \mathcal{I}_{b,k-1}^- (\mathbf{I} + \mathbf{\Lambda}_{f,k-1}^+ \mathcal{I}_{b,k-1}^-)^{-1} \mathbf{F}_{k-1}^{-1}. \tag{6.37}$$

From eq. (6.9) we see that $\mathbf{F}_{k-1}^{-1} \hat{\mathbf{z}}_{fk}^- = \hat{\mathbf{z}}_{f,k-1}^+$. Therefore eq. (6.36) can be written as

$$\varrho_{k-1}^- - \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_k) \mathcal{I}_{fk}^- \mathbf{F}_{k-1} \mathbf{\Lambda}_{f,k-1}^+ \varrho_{k-1}^- -$$
$$\mathcal{I}_{b,k-1}^- (\mathbf{I} + \mathbf{\Lambda}_{f,k-1}^+ \mathcal{I}_{b,k-1}^-)^{-1} \hat{\mathbf{z}}_{f,k-1}^+ = \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{fk}^-). \tag{6.38}$$

Now substitute for $\mathbf{\Lambda}_k$ from eq. (6.26) and use eq. (6.27) in the above equation, and premultiplying both sides by $\mathbf{\Lambda}_{f,k-1}^+$ gives

$$\Big[\mathbf{\Lambda}_{f,k-1}^+ - \mathbf{\Lambda}_{f,k-1}^+ \mathbf{F}_{k-1}^T \mathcal{I}_{fk}^- (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_{fk}^+ + \mathbf{K}_k \mathbf{\Lambda}_{f,k+1}^- \mathbf{K}_k^T - \mathbf{K}_k \mathbf{\Lambda}_{k+1} \mathbf{K}_k^T) \mathbf{K}_{k-1}^T \Big] \varrho_{k-1}^-$$
$$- \mathbf{\Lambda}_{f,k-1}^+ \mathcal{I}_{b,k-1}^- (\mathbf{I} + \mathbf{\Lambda}_{f,k-1}^+ \mathcal{I}_{b,k-1}^-)^{-1} \hat{\mathbf{z}}_{f,k-1}^+ = \mathbf{K}_{k-1} (\hat{\mathbf{z}} - \hat{\mathbf{z}}_{fk}^-). \tag{6.39}$$

Now use eq. (6.27) to notice that the coefficient of ϱ_{k-1}^- on the left side of the above equation can be written as

$$\mathbf{\Lambda}_{f,k-1}^+ - \mathbf{K}_{k-1} (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_{fk}^+ + \mathbf{K}_k \mathbf{\Lambda}_{f,k+1}^- \mathbf{K}_k^T - \mathbf{K}_k \mathbf{\Lambda}_{k+1} \mathbf{K}_k^T) \mathbf{K}_{k-1}^T. \tag{6.40}$$

Using eq. (6.26) to substitute for $\mathbf{K}_k \mathbf{\Lambda}_{k+1} \mathbf{K}_k^T$, we get

$$\mathbf{\Lambda}_{f,k-1}^+ - \mathbf{K}_{k-1} (\mathbf{\Lambda}_{fk}^- - \mathbf{\Lambda}_{fk}^+ + \mathbf{K}_k \mathbf{\Lambda}_{f,k+1}^- \mathbf{K}_k^T - \mathbf{\Lambda}_k + \mathbf{\Lambda}_{fk}^+ - \mathbf{K}_k \mathbf{\Lambda}_{f,k+1}^- \mathbf{K}_k^T) \mathbf{K}_{k-1}^T$$
$$= \mathbf{\Lambda}_{f,k-1}^+ - \mathbf{K}_{k-1} \mathbf{\Lambda}_{fk}^- \mathbf{K}_{k-1}^T + \mathbf{K}_{k-1} \mathbf{\Lambda}_k \mathbf{K}_{k-1}^T$$
$$= \mathbf{\Lambda}_{f,k-1}^+ - \mathbf{K}_{k-1} \mathbf{\Lambda}_{fk}^- \mathbf{K}_{k-1}^T + \mathbf{\Lambda}_{k-1} - \mathbf{\Lambda}_{f,k-1}^+ + \mathbf{K}_{k-1} \mathbf{\Lambda}_{fk}^- \mathbf{K}_{k-1}^T$$
$$= \mathbf{\Lambda}_{k-1}. \tag{6.41}$$

Since this is the coefficient of ϱ_{k-1}^- in eq. (6.39), we can write that equation as

$$\mathbf{\Lambda}_{k-1}\varrho_{k-1}^- - \mathbf{\Lambda}_{f,k-1}^+\mathcal{I}_{b,k-1}^-(\mathbf{I}+\mathbf{\Lambda}_{f,k-1}^+\mathcal{I}_{b,k-1}^-)^{-1}\hat{\mathbf{z}}_{f,k-1}^+ = \mathbf{K}_{k-1}(\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{fk}^-). \tag{6.42}$$

Now from eqs. (6.17) and (6.18) we see that

$$\begin{aligned}\hat{\mathbf{z}}_k &= (\mathcal{I}_{fk}^+ + \mathcal{I}_{bk}^-)^{-1}\mathcal{I}_{fk}^+\hat{\mathbf{z}}_{fk}^+ + \mathbf{\Lambda}_k\mathcal{I}_{bk}^-\hat{\mathbf{z}}_{bk}^- \\ &= (\mathbf{I}+\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-)^{-1}\hat{\mathbf{z}}_{fk}^+ + \mathbf{\Lambda}_k\varrho_k^-.\end{aligned} \tag{6.43}$$

From this we see that

$$\begin{aligned}\hat{\mathbf{z}}_k - \hat{\mathbf{z}}_{fk} &= \left[(\mathbf{I}+\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-)^{-1} - \mathbf{I}\right]\hat{\mathbf{z}}_{fk}^+ + \mathbf{\Lambda}_k\varrho_k^- \\ &= \left[\mathbf{I} - (\mathbf{I}+\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-)\right](\mathbf{I}+\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-)^{-1}\hat{\mathbf{z}}_{fk}^+ + \mathbf{\Lambda}_k\varrho_k^- \\ &= -\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-(\mathbf{I}+\mathbf{\Lambda}_{fk}^+\mathcal{I}_{bk}^-)^{-1}\hat{\mathbf{z}}_{fk}^+ + \mathbf{\Lambda}_k\varrho_k^-.\end{aligned} \tag{6.44}$$

Rewriting the above equation with the time subscripts $(k-1)$ and then substituting for the left side of eq. (6.42) gives

$$\hat{\mathbf{z}}_k = \hat{\mathbf{z}}_{fk}^+\mathbf{K}_k(\hat{\mathbf{z}}_{k+1} - \hat{\mathbf{z}}_{f,k+1}^-). \tag{6.45}$$

This gives the smoothed estimate $\hat{\mathbf{z}}_k$ without needing to explicitly calculate the backward estimate. The RTS smoother is implemented by first running the standard Kalman filter of eq. (6.9) forward in time to the final time, and then implementing eqs. (6.26), (6.27), and (6.45) backward in time. The RTS smoother can be summarized as follows.

The RTS smoother

1. The system model is given as follows:

$$\begin{aligned}\mathbf{z}_k &= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \eta_{k-1}, \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\ \eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \\ \nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).\end{aligned} \tag{6.46}$$

2. Initialize the forward filter as follows:

$$\begin{aligned}\hat{\mathbf{z}}_{f0} &= \langle \mathbf{z}_0 \rangle, \\ \mathbf{\Lambda}_{f0}^+ &= \left\langle (\mathbf{z}_0 - \hat{\mathbf{z}}_{f0})(\mathbf{z}_0 - \hat{\mathbf{z}}_{f0})^T \right\rangle.\end{aligned} \tag{6.47}$$

3. For $k = 1, \ldots, N$, execute the standard forward Kalman filter:

$$\begin{aligned}
\mathbf{\Lambda}_{fk}^{-} &= \mathbf{F}_{k-1}\mathbf{\Lambda}_{f,k-1}^{+}\mathbf{F}_{k-1}^{T} + \mathbf{Q}_{k-1}, \\
\mathbf{K}_{fk} &= \mathbf{\Lambda}_{fk}^{-}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{\Lambda}_{fk}^{-}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \\
&= \mathbf{\Lambda}_{fk}^{+}\mathbf{H}_k^T\mathbf{R}_k^{-1}, \\
\hat{\mathbf{z}}_{fk}^{-} &= \mathbf{F}_{k-1}\hat{\mathbf{z}}_{f,k-1}^{+}, \\
\hat{\mathbf{z}}_{fk}^{+} &= \hat{\mathbf{z}}_{fk}^{-} + \mathbf{K}_{fk}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_{fk}^{-}), \\
\mathbf{\Lambda}_{fk}^{+} &= (\mathbf{I} - \mathbf{K}_{fk}\mathbf{H}_k)\mathbf{\Lambda}_{fk}^{-}(\mathbf{I} - \mathbf{K}_{fk}\mathbf{H}_k)^T + \mathbf{K}_{fk}\mathbf{R}_k\mathbf{K}_{fk}^T \\
&= \left[(\mathbf{\Lambda}_{fk}^{-})^{-1} + \mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\right]^{-1} \\
&= (\mathbf{I} - \mathbf{K}_{fk}\mathbf{H}_k)\mathbf{\Lambda}_{fk}^{-}.
\end{aligned} \tag{6.48}$$

4. Initialize the RST smoother as follows:

$$\begin{aligned}
\hat{\mathbf{z}}_N &= \hat{\mathbf{z}}_{fN}^{+}, \\
\mathbf{\Lambda}_N &= \mathbf{\Lambda}_{fN}^{+}.
\end{aligned} \tag{6.49}$$

5. For $k = N-1, \ldots, 1, 0$, execute the following RTS smoother equations:

$$\begin{aligned}
\mathcal{I}_{f,k+1}^{-} &= \left(\mathbf{\Lambda}_{f,k+1}^{-}\right)^{-1}, \\
\mathbf{K}_k &= \mathbf{\Lambda}_{fk}^{+}\mathbf{F}_k^T\mathcal{I}_{f,k+1}^{-}, \\
\mathbf{\Lambda}_k &= \mathbf{\Lambda}_{fk}^{+} - \mathbf{K}_k(\mathbf{\Lambda}_{f,k+1}^{-} - \mathbf{\Lambda}_{k+1})\mathbf{K}_k^T \\
\hat{\mathbf{z}} &= \hat{\mathbf{z}}_{fk}^{+} + \mathbf{K}_k(\hat{\mathbf{z}}_{k+1} - \hat{\mathbf{z}}_{f,k+1}).
\end{aligned} \tag{6.50}$$

6.1.3 Generalized fixed interval smoothing

Let us first take a look at the generalization of the RTS algorithm, i.e., a procedure that involves a first filtering pass followed by a second, backward sweep. Specifically, we've already described the filtering procedure, i.e., for the recursive computation of

$$p_{k|k}(\mathbf{z}) = p_{\mathbf{z}_k|\mathbf{Y}_{0:k}}(\mathbf{z}|\mathbf{Y}_{0:M}), \tag{6.51}$$

$$p_{k+1|k}(\mathbf{z}) = p_{\mathbf{z}_{k+1}|\mathbf{Y}_{0:k}}(\mathbf{z}|\mathbf{Y}_{0:M}). \tag{6.52}$$

In particular, eqs. (5.47) and (5.48) indicate how these are computed recursively. As in RTS, assume that during this forward sweep that we have stored these quantities.

The quantity we wish to compute at each point in time in the smoothing problem is the smoothed density:

$$p_k^{\mathrm{S}}(\mathbf{z}) \equiv p_{\mathbf{z}_k|\mathbf{Y}_{0:M}}(\mathbf{z}_k|\mathbf{Y}_{0:M}). \tag{6.53}$$

As in RTS, at the completion of the last step of the forward, filtering step we have this quantity at the end point M. That is:

$$p_M^{\mathrm{S}}(\mathbf{z}) = p_{M|M}(\mathbf{z}), \tag{6.54}$$

and this quantity is used to initialize a backward sweep. The general step of this backward sweep, going from time $k+1$ to time k computes the smoothed density at that time, i.e., $p_k^{\mathrm{S}}(\mathbf{z})$, by making used of the quantities (6.51) (6.52) that were stored in the forward pass as well as the just-computed value of the smoothed density at time $k+1$, i.e., $p_{k+1}^{\mathrm{S}}(\mathbf{z})$.

By breaking up the full set of observations to those up to and after time k (i.e., by writing $\mathbf{Y}_{0:M} = \{\mathbf{Y}_{0:k}, \mathbf{Y}_{k+1:M}\}$) and then applying Bayes' Rule to the conditioning on $\mathbf{Y}_{k+1:M}$, we can write the smoothed density that we already have as follows:

$$p_{k+1}^{\mathrm{S}}(\mathbf{z}) = \frac{p_{k+1|k}(\mathbf{z}) p_{\mathbf{Y}_{k+1:M}|\mathbf{Y}_{0,k},\mathbf{z}_{k+1}}(\mathbf{Y}_{k+1:M}|\mathbf{Y}_{0:k}, \mathbf{z})}{\kappa_k}, \tag{6.55}$$

where κ_k is a normalizing constant equal to:

$$\kappa_k = p_{\mathbf{Y}_{k+1:M}|\mathbf{Y}_{0:k}}(\mathbf{Y}_{k+1:M}|\mathbf{Y}_{0:k}). \tag{6.56}$$

Using the Markovianity of $\mathbf{z}_k$ and the conditional independence of the measurements, we can simplify eq. (6.55), yielding

$$p_{k+1}^{\mathrm{S}}(\mathbf{z}) = \frac{p_{k+1|k}(\mathbf{z}) p_{\mathbf{Y}_{k+1:M}|\mathbf{z}_{k+1}}(\mathbf{Y}_{k+1:M}|\mathbf{z})}{\kappa_k}. \tag{6.57}$$

Remember that we already have the quantity on the left-hand side and $p_{k+1|k}(\mathbf{z})$, in the numerator on the right hand side of eq. (6.57), so that we also can compute the ratio of the second term in the numerator on the right-hand side divided by κ_k.

In a similar manner we can write the smoothed density that we *now want to compute* as follows:

$$p_k^{\mathrm{S}}(\mathbf{z}) = \frac{p_{k|k}(\mathbf{z}) p_{\mathbf{Y}_{k+1:M}|\mathbf{z}_k}(\mathbf{Y}_{k+1:M}|\mathbf{z})}{\kappa_k}. \tag{6.58}$$

Now, we *do* have the filtered density $p_{k+1|k}(\mathbf{z})$, so the remaining problem is computing the ratio of the second term in the numerator on the right-hand side divided by κ_k, something we will compute and what we can compute from eq. (6.57) and what we already know. the key to this is the exploitation of Markovianity as follows:

$$\begin{aligned} p_{\mathbf{Y}_{k+1:M}|\mathbf{z}_k}(\mathbf{Y}_{k+1}|\mathbf{z}) &= \int \mathrm{d}\xi \, p_{\mathbf{Y}_{k+1:M},\mathbf{z}_{k+1}|\mathbf{z}_k}(\mathbf{Y}_{k+1}, \xi|\mathbf{z}) \\ &= \int \mathrm{d}\xi \, p_{\mathbf{Y}_{k+1:M}|\mathbf{z}_k}(\mathbf{Y}_{k+1}|\xi) p_{\mathrm{T}}(\xi|\mathbf{z}, k+1). \end{aligned} \tag{6.59}$$

Combining this with eqs. (6.57) and (6.58), we can now write the backward step of the

generalized RTS recursion:

$$p_k^{\mathrm{S}}(\mathbf{z}) = p_{k|k}(\mathbf{z}) \int \mathrm{d}\xi \, \frac{p_{k+1}^{\mathrm{S}}(\xi)}{p_{k+1|k}(\xi)} p_{\mathrm{T}}(\xi|\mathbf{z}, k+1). \tag{6.60}$$

The generalization of the two-filter approach to fixed-interval smoothing computes the unknown quantity in relationships analogous to eqs. (6.57), (6.58) but does it in a more direct manner. Specifically, let's look at eq. (6.57) but evaluated at time k rather than $k+1$:

$$p_k^{\mathrm{S}}(\mathbf{z}) = \frac{p_{k|k-1}(\mathbf{z}) p_{\mathbf{Y}_{k:M}|\mathbf{z}_k}(\mathbf{Y}_{k:M}|\mathbf{z})}{\kappa_{k-1}} \equiv \frac{p_{k|k-1}(\mathbf{z}) \ell_k(\mathbf{z})}{\kappa_{k-1}}. \tag{6.61}$$

Since the denominator in this equation is simply a normalizing constant, the keys are computing the two terms in the numerator. The first is obtained via the usual forward filtering recursion, so we need to develop a *backward* recursion for $\ell_k(\mathbf{z})$. The measurement density gives us the starting point for this recursion at time M:

$$\ell_M(\mathbf{z}) = p_m(\mathbf{y}_k|\mathbf{z}, M). \tag{6.62}$$

Suppose that we have computed $\ell_{k+1}(\mathbf{z})$. Then, using Markovianity, conditional independence of the measurements, as well as eq. (6.61), we obtain the following backward recursion:

$$\begin{aligned} \ell_k(\mathbf{z}) &= p_{\mathbf{Y}_{k,M}|\mathbf{z}_k}(\mathbf{Y}_{k,M}|\mathbf{z}) = p_{\mathbf{y}_k|\mathbf{Y}_{k+1,M}}(\mathbf{y}_k, \mathbf{Y}_{k+1,M}|\mathbf{z}) \\ &= p_m(\mathbf{y}_k|\mathbf{z}, k) p_{\mathbf{Y}_{k+1,M}|\mathbf{z}_k}(\mathbf{Y}_{k+1,M}|\mathbf{z}) \\ &= p_m(\mathbf{y}_k|\mathbf{z}, k) \int \mathrm{d}\xi \, \ell_{k+1}(\xi) p_{\mathrm{T}}(\xi|\mathbf{z}, k+1). \end{aligned} \tag{6.63}$$

It is worth noting that, as in the linear case, we can interpret this backward recursion as an anticausal filter, in this case starting with a non-informative (i.e., flat) prior — as reflected in the fact that the initialization of the recursion in eq. (6.62) uses the measurement at time M but no weighting for a prior distribution on $\mathbf{z}_M$.

6.2 Parameter estimation

State estimation theory can be used to not only estimate the states of a system, but also to estimate the unknown parameters of a system. This may have first been suggested by Kopp in 1963 [161]. Suppose that we have a discrete-time system model, but the system matrices depend in a nonlinear way on an unknown parameter vector $\boldsymbol{\Theta}$:

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathbf{F}_k(\Theta)\mathbf{z}_k + \mathbf{L}_k(\Theta)\eta_k, \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\ \boldsymbol{\Theta} &= \text{unknown parameter vector.} \end{aligned} \tag{6.64}$$

In this model, we are assuming that the measurement is independent of $\boldsymbol{\Theta}$. We further assume that $\boldsymbol{\Theta}$ is a constant parameter vector. In order to estimate the parameter $\boldsymbol{\Theta}$, we first augment the state with the parameter to obtain an augmented state vector ζ:

$$\zeta_k = \begin{bmatrix} \mathbf{z}_k \\ \boldsymbol{\Theta}_k \end{bmatrix} \tag{6.65}$$

If $\boldsymbol{\Theta}_k$ is constant then we model $\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_k + \eta_{\boldsymbol{\Theta}k}$, where $\eta_{\boldsymbol{\Theta}k}$ is a small artificial noise term that allows the Kalman filter to change its estimate of $\boldsymbol{\Theta}_k$. Our augmented system model can be written as

$$\begin{aligned} \zeta_{k+1} &= \begin{bmatrix} \mathbf{F}_k(\Theta_k)\mathbf{z}_k + \mathbf{L}_k(\Theta_k)\eta_k \\ \boldsymbol{\Theta}_k + \eta_{\boldsymbol{\Theta}k} \end{bmatrix} \\ &= \mathcal{F}(\mathbf{z}_k, \eta_k, \eta_{\boldsymbol{\Theta}k}). \\ \mathbf{y}_k &= \begin{bmatrix} \mathbf{H}_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{z}_k \\ \boldsymbol{\Theta}_k \end{bmatrix} + \nu_k. \end{aligned} \tag{6.66}$$

Note that $\mathcal{F}(\mathbf{z}_k, \eta_k, \eta_{\boldsymbol{\Theta}k})$ is a nonlinear function of the augmented state ζ_k. We can, for example, use the particle filter to estimate simultaneously both the parameters and the conditional distribution for the signal state [45].

6.3 Verifying filter performance

We can verify Kalman filter performance, or adjust the gain of the Kalman filter, using our knowledge of the statistics of the *innovations*. The innovations is defined as $(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-)$, and in this section we will show that it is a zero mean white stochastic process with a covariance of $(\mathbf{H}_k\boldsymbol{\Lambda}_k^-\mathbf{H}_k^T + \mathbf{R}_k)$.

Recall our original system model, along with the one-step a priori update equation for the state estimate:

$$\begin{aligned} \mathbf{z}_k &= \mathbf{F}_{k-1}\mathbf{z}_{k-1} + \eta_{k-1}, \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{z}_k + \nu_k, \\ \hat{\mathbf{z}}_{k+1}^- &= \mathbf{F}_k\hat{\mathbf{z}}_k^- + \mathbf{F}_k\mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-). \end{aligned} \tag{6.67}$$

The innovations is defined as the quantity in parenthesis in the update equation. The innovations can be thought of as the part of the measurement that contain new information and that is therefore used to update the state estimate (apart from our knowledge of the state transition matrix). If the innovations was zero then the state estimate would simply by updated according to the state transition matrix. A nonzero innovations allows the

measurement to affect the state estimate. The innovations ϑ_k can be written as

$$\begin{aligned}\vartheta_k &= \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^- \\ &= (\mathbf{H}_k\mathbf{z}_k + \nu_k) - \mathbf{H}_k\hat{\mathbf{z}}_k^- \\ &= \mathbf{H}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k^-) + \nu_k \\ &= \mathbf{H}_k\epsilon_k + \nu_k, \end{aligned} \tag{6.68}$$

where ϵ_k, the a priori estimation error, is defined as $\mathbf{z}_k - \hat{\mathbf{z}}_k^-$. The covariance of the innovations is given as

$$\left\langle \vartheta_k\vartheta_i^T \right\rangle = \left\langle (\mathbf{H}_k\epsilon_k + \nu_k)(\mathbf{H}_i\epsilon_i + \nu_i)^T \right\rangle. \tag{6.69}$$

Let us see what the covariance is when $k \neq i$. We can assume without loss of generality that $k > i$. We then obtain

$$\left\langle \vartheta_k\vartheta_i^T \right\rangle = \mathbf{H}_k \left\langle \epsilon_k\epsilon_i^T \right\rangle \mathbf{H}_i^T + \mathbf{H}_k \left\langle \epsilon_k\nu_i^T \right\rangle. \tag{6.70}$$

Note that two of the cross terms reduced to zero because of the whiteness of ν_k, and the fact that the estimation error ϵ_i is independent of ν_k for $k > i$. In order to evaluate this covariance, we need to evaluate $\left\langle \epsilon_k\epsilon_i^T \right\rangle$ and $\left\langle \epsilon_k\nu_i^T \right\rangle$. First we will evaluate $\left\langle \epsilon_k\epsilon_i^T \right\rangle$. In order to evaluate this term, notice that the a priori state estimate can be written as follows:

$$\begin{aligned}\hat{\mathbf{z}}_{k+1}^- &= \mathbf{F}_k\hat{\mathbf{z}}_k^- + \mathbf{F}_k\mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-) \\ &= \mathbf{F}_k\hat{\mathbf{z}}_k^- + \mathbf{F}_k\mathbf{K}_k(\mathbf{H}_k\mathbf{z}_k + \nu_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-) \\ &= \mathbf{F}_k\hat{\mathbf{z}}_k^- + \mathbf{F}_k\mathbf{K}_k\mathbf{H}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k^-) + \mathbf{F}_k\mathbf{H}_k\nu_k. \end{aligned} \tag{6.71}$$

The a priori estimation error can be written as

$$\begin{aligned}\epsilon_{k+1} &= \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1}^- \\ &= \mathbf{F}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k^-) - \mathbf{F}_k\mathbf{K}_k\mathbf{H}_k(\mathbf{z}_k - \hat{\mathbf{z}}_k^-) + \eta_k - \mathbf{F}_k\mathbf{K}_k\nu_k \\ &= \mathbf{F}_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\epsilon_k + (\eta_k - \mathbf{F}_k\mathbf{K}_k\nu_k) \\ &= \tilde{\psi}_k\epsilon_k + \nu_k', \end{aligned} \tag{6.72}$$

where $\tilde{\psi}_k$ and ν_k' are defined by the above equation. This is a linear discrete-time system for ϵ_k with the state transition matrix

$$\tilde{\psi}_{k,i} = \begin{cases} \tilde{\psi}_{k-1}\tilde{\psi}_{k-1}\cdots\tilde{\psi}_i, & k > i; \\ \mathbf{I}, & k = i. \end{cases} \tag{6.73}$$

The error ϵ_k can be solved from the initial condition ϵ_i as follows:

$$\epsilon_k = \tilde{\psi}_{k,i}\epsilon_i + \sum_{j=i}^{k-1} \tilde{\psi}_{k,j+1}\nu_j'. \tag{6.74}$$

The covariance of $\epsilon_k\epsilon_i^T$ can be written as

$$\left\langle \epsilon_k\epsilon_i^T \right\rangle = \left\langle \left(\tilde{\psi}_{k,i}\epsilon_i + \sum_{j=i}^{k-1} \tilde{\psi}_{k,j+1}\nu_j' \right) \epsilon_i^T \right\rangle. \tag{6.75}$$

We see that all of the $\nu_j'\epsilon_i^T$ terms in the above expression are zero-mean. This is because all of the ν_j' noise terms occur at time i or later and so do not affect ϵ_i. Note from eq. (6.72) that ϵ_i is affected only by the noise terms at time $(i-1)$ or earlier. Therefore,

$$\left\langle \nu_j'\epsilon_i^T \right\rangle = \mathbf{0} \qquad \text{for } j \geq i. \tag{6.76}$$

We therefore see that eq. (6.75) can be written as

$$\begin{aligned} \left\langle \epsilon_k\epsilon_i^T \right\rangle &= \tilde{\psi}_{k,i} \left\langle \epsilon_i\epsilon_i^T \right\rangle \\ &= \tilde{\psi}_{k,i}\mathbf{\Lambda}_i^-. \end{aligned} \tag{6.77}$$

Now that we have computed $\left\langle \epsilon_k\epsilon_i^T \right\rangle$, we need to solve for $\left\langle \epsilon_i\epsilon_i^T \right\rangle$ in order to arrive at our goal, which is the evaluation of eq. (6.70). $\left\langle \epsilon_k\nu_i^T \right\rangle$ can be written as

$$\left\langle \epsilon_k\nu_i^T \right\rangle = \left\langle \left(\tilde{\psi}_{k,i}\epsilon_i + \sum_{j=i}^{k-1} \tilde{\psi}_{k,j+1}\nu_j' \right) \nu_i^T \right\rangle. \tag{6.78}$$

The $\epsilon_i\nu_i^T$ term in the above expression is zero-mean, and the $\nu_j'\nu_i^T$ terms are zero-mean for $j > i$. The above covariance can therefore be written as

$$\begin{aligned} \left\langle \epsilon_k\nu_i^T \right\rangle &= \left\langle \tilde{\psi}_{k,i+1}\nu_i'\nu_i^T \right\rangle \\ &= \left\langle \tilde{\psi}_{k,i+1}(\eta_i - \mathbf{F}_i\mathbf{K}_i\nu_i)\nu_i^T \right\rangle \\ &= -\tilde{\psi}_{k,i+1}\mathbf{F}_i\mathbf{K}_i\mathbf{R}_i. \end{aligned} \tag{6.79}$$

Substituting this equation, along with eq. (6.77), into eq. (6.70) gives

$$\begin{aligned} \left\langle \vartheta_k\vartheta_i^T \right\rangle &= \mathbf{H}_k \left\langle \epsilon_k\epsilon_i^T \right\rangle \mathbf{H}_i^T + \mathbf{H}_k \left\langle \epsilon_k\nu_i^T \right\rangle \\ &= \mathbf{H}_k\tilde{\psi}_{k,i}\mathbf{\Lambda}_i^-\mathbf{H}_i^T - \mathbf{H}_k\tilde{\psi}_{k,i+1}\mathbf{F}_i\mathbf{K}_i\mathbf{R}_i \\ &= \mathbf{H}_k\tilde{\psi}_{k,i+1}\left(\tilde{\psi}_i\mathbf{\Lambda}_i^-\mathbf{H}_i^T - \mathbf{F}_i\mathbf{K}_i\mathbf{R}_i \right). \end{aligned} \tag{6.80}$$

Now use the fact from eq. (6.72) that $\tilde{\psi}_i = \mathbf{F}_i(\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)$ to obtain

$$\begin{aligned}\left\langle \vartheta_k \vartheta_i^T \right\rangle &= \mathbf{H}_k \tilde{\psi}_{k,i+1} \left(\mathbf{F}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T - \mathbf{F}_i \mathbf{K}_i \mathbf{H}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T - \mathbf{F}_i \mathbf{K}_i \mathbf{R}_i \right) \\ &= \mathbf{H}_k \tilde{\psi}_{k,i+1} \left[\mathbf{F}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T - \mathbf{F}_i \mathbf{K}_i \left(\mathbf{H}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T + \mathbf{R}_i \right) \right]. \end{aligned} \tag{6.81}$$

Now use the fact that $\mathbf{K}_i = \mathbf{\Lambda}_i^- \mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T + \mathbf{R}_i \right)^{-1}$ (the standard Kalman filter gain equation) to obtain

$$\begin{aligned}\left\langle \vartheta_k \vartheta_i^T \right\rangle &= \mathbf{H}_k \tilde{\psi}_{k,i+1} \left(\mathbf{F}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T - \mathbf{F}_i \mathbf{\Lambda}_i^- \mathbf{H}_i^T \right) \\ &= \mathbf{0} \qquad k > i. \end{aligned} \tag{6.82}$$

So we see that the innovations ν_k is white noise. Our next task is to determine its covariance. In order to do this we write the covariance as

$$\begin{aligned}\left\langle \vartheta_k \vartheta_k^T \right\rangle &= \left\langle (\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{z}}_k^-)(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{z}}_k^-)^T \right\rangle \\ &= \left\langle \left[\mathbf{H}_k (\mathbf{z}_k - \hat{\mathbf{z}}^-) + \nu_k \right] \left[\mathbf{H}_k (\mathbf{z}_k - \hat{\mathbf{z}}^-) + \nu_k \right]^T \right\rangle \\ &= \mathbf{H}_k \left\langle \epsilon_k \epsilon_k^T \right\rangle \mathbf{H}_k^T + \left\langle \nu_k \nu_k^T \right\rangle \\ &= \mathbf{H}_k \mathbf{\Lambda}_k^- \mathbf{H}_k^T + \mathbf{R}_k. \end{aligned} \tag{6.83}$$

We therefore see that the innovations is a white noise process with zero mean and a covariance of $\left(\mathbf{H}_k \mathbf{\Lambda}_k^- \mathbf{H}_k^T + \mathbf{R}_k \right)$. While the Kalman filter is operating, we can process the innovations, compute its mean and covariance, and verify that it is white with the expected mean and covariance. If it is colored, nonzero-mean, or has the wrong covariance, then there is something wrong with the filter. The most likely reason for such a discrepancy is a modeling error. In particular, an incorrect value of $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, or $\mathbf{R}$ could cause the innovations to statistically deviate from its theoretically expected behavior. Statistical methods can then be used to tune $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ in order to force the innovations to be white zero-mean noise with a covariance of $\left(\mathbf{H}_k \mathbf{\Lambda}_k^- \mathbf{H}_k^T + \mathbf{R}_k \right)$ [196] [197].

Alternatively, if we are uncertain of the correct values of $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$, we can either use an adaptive approach to estimate the state noise statistic [206], or run a bank of Kalman filters in parallel, each Kalman filter with a value of $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ that we think may be likely. Then the innovations can be inspected in each filter (see Appendix C), and the one that matches theory is assumed to have the correct $\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$, so the state estimate that comes out of that filter is probably the most correct. The prediction errors whiteness criterion is a common practice in time series modeling [124] [279] [280].

6.4 Multiple-model estimation

Suppose our system model is not known, or the system model changes depending on unknown factors. We can use multiple filters (one for each possible system model) and combine

the state estimates to obtain a refined state estimate. Suppose that a random vector $\mathbf{z}$ can take on one of N mutually exclusive values $\mathbf{z}_1, \ldots, \mathbf{z}_N$. We can use Bayes' rule to write

$$\begin{aligned} \Pr(\mathbf{y}) &= \Pr(\mathbf{y}|\mathbf{z}_1)\Pr(\mathbf{z}_1) + \cdots + \Pr(\mathbf{y}|\mathbf{z}_N)\Pr(\mathbf{z}_N), \\ \Pr(\mathbf{z}|\mathbf{y}) &= \frac{p(\mathbf{y}|\mathbf{z})\Pr(\mathbf{z})}{\sum_{i=1}^{N} p(\mathbf{y}|\mathbf{z}_i)\Pr(\mathbf{z}_i)}. \end{aligned} \tag{6.84}$$

where we have used the fact that the probability of an event occurring is directly proportional to the value of its pdf. Now suppose that we have the time-invariant system

$$\begin{aligned} \mathbf{z}_k &= \mathbf{F}\mathbf{z}_{k-1} + \eta_{k-1}, \\ \mathbf{y}_k &= \mathbf{H}\mathbf{z}_k + \nu_k, \\ \eta_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \\ \nu_k &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \end{aligned} \tag{6.85}$$

The parameter set Ω is defined as the set ($\mathbf{F}$, $\mathbf{H}$, $\mathbf{Q}$, $\mathbf{R}$). Suppose that Ω can take one of N possible values $\Omega_1, \ldots, \Omega_N$. The question that we want to answer in this section is as follows: Given the measurements $\mathbf{y}_k$, what is the probability that $\Omega = \Omega_j$? From eq. (6.84) this probability can be written as

$$\Pr(\Omega_j|\mathbf{y}_k) = \frac{p(\mathbf{y}_k|\Omega_j)\Pr(\Omega_j)}{\sum_{i=1}^{N} p(\mathbf{y}_k|\Omega_i)\Pr(\Omega_i)}. \tag{6.86}$$

Now think about the probability that measurement $\mathbf{y}_k$ is observed given the facts that $\Omega = \Omega_j$. If $\Omega = \Omega_j$ then the state will take on some value $\mathbf{z}_k$ that is determined by the parameter set Ω_j. We therefore see that

$$\begin{aligned} \Pr(\mathbf{y}_k|\Omega_j) &= \Pr(\mathbf{y}_k|\mathbf{z}_k), \\ p(\mathbf{y}_k|\Omega_j) &= p(\mathbf{y}_k|\mathbf{z}_k). \end{aligned} \tag{6.87}$$

However, if our state estimate is accurate, then we know that $\mathbf{z}_k \approx \hat{\mathbf{z}}_k^-$. Therefore, the above equation can be written as

$$p(\mathbf{y}_k|\Omega_j) \approx p(\mathbf{y}_k|\hat{\mathbf{z}}_k^-). \tag{6.88}$$

The right side of the equation is the pdf of the measurement $\mathbf{y}_k$ given the fact that the state is $\hat{\mathbf{z}}_k^-$. But since $\mathbf{y}_k \approx \mathbf{H}\hat{\mathbf{z}}_k^- + \nu_k$, this pdf is approximately equal to the pdf of $(\mathbf{y}_k - \mathbf{H}\hat{\mathbf{z}}_k^-)$. We therefore have

$$p(\mathbf{y}_k|\Omega_j) \approx p(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_k^-) = p(\nu_k). \tag{6.89}$$

where ν_k is the residual defined in Sec. 6.3. From Sec. 6.3 we see that if η_k, ν_k, and $\mathbf{z}_0$ are Gaussian, then the residual ν_k is a linear combination of Gaussian random variables. In

Sec. 6.3 we found the mean and variance of ν_k. The pdf of ν_k, which is approximated by the pdf of $\mathbf{y}_k$ given Ω_j, can therefore be approximated as

$$p(\mathbf{y}_k|\Omega_j) \approx \frac{1}{(2\pi)^{m/2}|\mathbf{S}_k|^{1/2}} \exp\left[-\frac{\nu_k^T \mathbf{S}_k^{-1} \nu_k}{2}\right], \tag{6.90}$$

where $\nu_k = \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{z}}_k^-$, $\mathbf{S}_k = \mathbf{H}_k \mathbf{\Lambda}_k^- \mathbf{H}_k^T + \mathbf{R}_k$, and m is the number of measurements.

Now from Bayes' rule we can write the following equation for the probability that $\Omega = \Omega_j$ given the fact that the measurements $\mathbf{y}_{k-1}$ is observed.

$$\Pr(\Omega_j|\mathbf{y}_{k-1}) = \frac{\Pr(\mathbf{y}_{k-1}|\Omega_j)\Pr(\Omega_j)}{\Pr(\mathbf{y}_{k-1})}. \tag{6.91}$$

If we are presently at time k, then the measurement at time $(k-1)$ is a given. The value of the measurement at time $(k-1)$ is a certain event with a probability equal to one. Therefore, $\Pr(\mathbf{y}_{k-1}|\Omega_j) = \Pr(\mathbf{y}_{k-1}) = 1$ and the above equation becomes

$$\Pr(\Omega_j|\mathbf{y}_{k-1}) = \Pr(\Omega_j). \tag{6.92}$$

Now in eq. (6.86) we can substitute this equation for $\Pr(\Omega_j)$, and we substitute eq. (6.90) for $p(\mathbf{y}_k|\Omega_j)$. This gives a time-recursive equation for evaluating the probability that $\Omega = \Omega_j$ given the fact that the measurement was equal to $\mathbf{y}_k$. The multiple-model estimator can be summarized as follows.

The multiple-model estimator

1. For $j = 1, \ldots, N$, initialize the probabilities of each parameter set before any measurements are obtained. These probabilities are denoted as $\Pr(\Omega_j|\mathbf{y}_0)$.

2. At each time step k we perform the following steps.

   (a) Run N (extended) Kalman filters, one for each parameter set Ω_j for $j = 1, \ldots, N$. The a priori state estimate and covariance of the jth filter are denoted as $\hat{\mathbf{z}}_{kj}^-$ and $\mathbf{\Lambda}_{kj}^-$.

   (b) After the measurement at time k is received, for each parameter set approximate the pdf of $\mathbf{y}_k$ given Ω_j as follows:

$$p(\mathbf{y}_k|\Omega_j) \approx \frac{1}{(2\pi)^{m/2}|\mathbf{S}_k|^{1/2}} \exp\left[-\frac{\nu_k^T \mathbf{S}_k^{-1} \nu_k}{2}\right], \tag{6.93}$$

   where $\nu_k = \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{z}}_{kj}^-$, $\mathbf{S}_k = \mathbf{H}\mathbf{\Lambda}_{kj}^- \mathbf{H}^T + \mathbf{R}_k$, m is the number of observations.

(c) Estimate the probability that $\Omega = \Omega_j$ as follows.

$$\Pr(\Omega_j|\mathbf{y}_k) = \frac{p(\mathbf{y}_k|\Omega_j)\Pr(\Omega_j|\mathbf{y}_{k-1})}{\sum_{i=1}^{N} p(\mathbf{y}_k|\Omega_i)\Pr(\Omega_i|\mathbf{y}_{k-1})}. \tag{6.94}$$

(d) Now that each parameter set Ω_j has an associated probability, we can weight each $\hat{\mathbf{z}}_{kj}^-$ and $\mathbf{\Lambda}_{kj}^-$ accordingly to obtain

$$\begin{aligned} \hat{\mathbf{z}}_k^- &= \sum_{j=1}^{N} \Pr(\Omega_j|\mathbf{y}_k)\hat{\mathbf{z}}_{kj}^-, \\ \mathbf{\Lambda}_k^- &= \sum_{j=1}^{N} \Pr(\Omega_j|\mathbf{y}_k)\mathbf{\Lambda}_{kj}^-. \end{aligned} \tag{6.95}$$

(e) We can estimate the true parameter set in one of several ways, depending on the application. For example, we can use the parameter set with the highest conditional probability as our parameter estimate, or we can estimate the parameter set as a weighted average of the parameter sets:

$$\hat{\Omega} = \begin{cases} \underset{p_j}{\operatorname{argmax}} \Pr(\Omega_j|\mathbf{y}_k), & \text{max-probability method;} \\ \sum_{j=1}^{N} \Pr(\Omega_j|\mathbf{y}_k)\Omega_j, & \text{weighted-average method.} \end{cases} \tag{6.96}$$

6.5 Summary

In this chapter we discussed some important topics related to nonlinear state estimation. We discussed how to obtain a smoothed estimate given all the available measurements. We have seen how to use nonlinear filters to solve the parameter estimation problem. We have also learned how to verify the performance of a filter, which gives us quantifiable confidence in the accuracy of our filter estimates. Finally we discussed multiple-model estimation, which is a way of estimating system states when we are not sure of which model is governing the dynamics.

Chapter 7

Simulation Studies II

> Classifying the world into linear and nonlinear systems is like classifying the world into bananas and nonbananas.
>
> — Anonymous [188]

The quest to distinguish deterministic chaos from noise has resulted in a number of sophisticated algorithms. Sugihara and May [272] proposed a simplex time series forecasting technique in which a flat distribution of the correlation coefficient between the original and the predicted series indicates noise, whereas in case of chaos it would gradually fall. Such approach may yield false results if a small amount of noise is added to the chaotic time series [44]. Theiler et al. [278] proposed a test for detecting nonlinearity by comparing particular characteristic—such as the correlation dimension, the forecasting error, or the largest Lyapunov exponent—of the original series with those of the surrogate series, a phase randomized version of the original time series. Kaplan and Glass [153] proposed a search for the particular orientation of the phase-space trajectory and coarse-grained d-dimensional embedding as tools for indicating underlying determinism. Schouten et al. [250] developed a method based on the correlation integral to estimate the correlation dimension belonging to the underlying dynamic phenomenon for a noisy attractor. Bhattacharya and Kanjilal [19] proposed that the distribution profiles of the scaled and normalized singular values for the data series can provide distinctive information concerning the deterministic and stochastic dynamics. Finally, Poon and Barahona [225] developed the numerical-titration method for the detection of chaos in short, noisy time series.

With these available techniques, a clear differentiation between chaotic and stochastic processes seems possible. It does not seem to be an easy task, however, to distinguish between the type of chaos that is purely deterministic and the type of chaos that is induced by noise. Especially when the source of the dynamic noise is an intrinsic feature of the system, one can even raise the question whether such a separation would be of great use

in trying the understand the system. The point is, of course, whether the "skeleton," the noiseless deterministic part is intrinsically chaotic, whether the dynamic noise has destroyed the deterministic chaos, or whether the dynamic noise has induced chaotic dynamics in an otherwise nonchaotic system. From a statistical mechanics view point, the more irregular the motion, the more likely the system will be to wander about the phase-space and to sample a large enough region. Furthermore, it is impossible to isolate any mechanical system. Our laboratory system is constantly subjected to outside perturbations. These perturbations may themselves make the system visit all regions of phase space, even if the system, considered in isolation, is unable to do so. It is one of the tasks of modern statistical mechanics to try to sort out for a given system, how much of its apparent ergodic behavior can be attributed to its deterministic dynamical properties, how much is due to the inevitable external perturbations, and to address the question of whether global ergodic behavior is necessary at all for the foundations of statistical mechanics.

Algorithms exist to estimate the level of noise in a chaotic time series [284]. However, such algorithm is not capable of distinguishing measurement noise from dynamic noise. Strumik et al. [271] proposed a method to simultaneously estimate noise level and distinguish between measurement and dynamic noise for chaotic time series. They showed that dynamic noise corrupting these deterministic systems can be considered effectively as an additive "pseudonoise" with the Cauchy distribution. That method has only been demonstrated for the Hénon and the Ikeda maps.

In this chapter, we will propose a method based on Volterra series modeling and nonlinear filtering for the purpose of differentiating between dynamic noise and measurement noise (ultimately between deterministic chaos and stochastic chaos). Initially, we will use relatively simple simulation models with well-known bifurcation routes (such as the logistic map and the Hénon map) in order to establish the feasibility of the proposed techniques. Next, we will examine the effects of measurement noise and dynamic noise on the intrinsic and noise-induced complexity of a variety of canonical models such as the Mackay-Glass model of respiratory instability, and the ecological map. These simulation models have been used for algorithmic testing in previous studies [225] and represent a wide range of complex behaviors including all standard routes to chaos (period-doubling, intermittency, subcritical, quasiperiodicity, and non-chaotic fractal), high-dimensional and high-order nonlinear systems, loosely coupled systems, and chaotic-periodic systems. By use of these simulation models we plan to systematically evaluate the performance of the nonlinear filtering algorithm. Note that the noise levels encountered in biomedical studies are generally orders of magnitude higher than the finite precision of numerical computation, and hence the latter source of error is neglected in these simulations. The proposed comprehensive agenda of model simulations and algorithmic testing will allow us to delineate the capability of these analytical tools and their limitations.

7.1 Simulation setup

Following the same notation in Chapter 5, the plant model is given by

$$\begin{aligned} x_k &= \sum_{m=0}^{M-1} a_m z_m(k) + \eta_k \\ y_k &= x_k + \nu_k \end{aligned} \tag{7.1}$$

where x_k is the system state and y_k, η_k and ν_k are Gaussian white noise uncorrelated with each other and the system state.

Model I (Dynamic noise model, $\eta_k \neq 0$, $\nu_k = 0$)—variability is due to stochastic chaos induced by dynamic noise η_k:

$$\begin{aligned} x_k &= a_0 + a_1 x_{k-1} + a_2 x_{k-2} + \cdots + a_\kappa x_{k-\kappa} + a_{\kappa+1} x_{k-1}^2 + \\ &\quad a_{\kappa+2} x_{k-1} x_{k-2} + \cdots + a_{M-1} x_{k-\kappa}^d + \eta_k \\ &\equiv \sum_{m=0}^{M-1} a_m z_m(k) + \eta_k \end{aligned} \tag{7.2}$$

$$y_k = x_k \tag{7.3}$$

Model II (measurement noise model, $\eta_n = 0$, $\nu_n \neq 0$)—variability is due to deterministic chaos obscured by measurement noise ν_k:

$$\begin{aligned} x_k &= a_0 + a_1 x_{k-1} + a_2 x_{k-2} + \cdots + a_\kappa x_{k-\kappa} + a_{\kappa+1} x_{k-1}^2 + \\ &\quad a_{\kappa+2} x_{k-1} x_{k-2} + \cdots + a_{M-1} x_{k-\kappa}^d \\ &\equiv \sum_{m=0}^{M-1} a_m z_m(n) \end{aligned} \tag{7.4}$$

$$y_k = x_k + \nu_k \tag{7.5}$$

Suppose the time series y_n has been found positive for chaotic dynamics by using the numerical-titration algorithm described in Chapter 4. We wish to test whether the chaos is stochastic or deterministic by determining whether Model I or Model II provides a better statistical fit to the data. The Volterra autoregressive (VAR) algorithm previously used to detect nonlinear dynamics cannot readily distinguish between these two models because both of them conform to the Volterra series expansion. As shown previously, both deterministic chaos (with measurement noise) and stochastic chaos (with dynamic noise) would test positive under the VAR algorithm and thus would be classified non-discriminatively as "chaotic dynamics" under the NL algorithm.

Despite this, careful examination of the VAR algorithm shows that it is in closer conformity with model I than Model II. Thus, although the VAR algorithm is responsive to both models, it should demonstrate preferential selectivity for Model I compared to Model II. By comparing the goodness-by-fit by the VAR algorithm with an alternative method that shows preferential selectively for model II, one should be able to statistically discriminate between these two models. To compare the goodness-of-fits, Model I is first *exactly* identified by application of the VAR algorithm. Because the coefficients a_m of Model II cannot be directly estimated by the VAR algorithm (since the system state x_k is unknown), exact identification of Model II calls for a different approach than the VAR algorithm. We propose two algorithms for this purpose.

7.1.1 Successive VAR algorithm

The identification of Model II can be divided into two sequential steps: estimation of the system state x_k, $k = 1, \ldots, N$ and estimation of the model coefficients $\hat{a}_m$, $m = 1, \ldots, M$. Each step is dependent on the other and thus the entire identification procedure cannot be completed in one iteration. We propose the following iterative algorithm for successive estimation of $\hat{x}_k$ and $\hat{a}_m$:

Successive VAR algorithm

1. Given a time series y_k, set $\hat{x}_k = y_k$, $k = 1, \ldots, N$.
2. Update $\hat{a}_m$ by applying the VAR algorithm to the time series $\hat{x}_k$.
3. Update $\hat{x}_k$ recursively ($k = \kappa + 1, \kappa + 2, \ldots, N$) by substitute $\hat{a}_m$ into eq. (7.4).
4. Repeat Steps 2-3 until the estimates $\hat{x}_k$, and $\hat{a}_m$ converge.

The above algorithm involves successive VAR estimation and polynomial expansion, which are relatively straightforward computations. The algorithm will converge provided η_k is sufficiently small, i.e., the observed time series y_k is a reasonably good estimate of x_k. In general, one may expect that the algorithm may diverge for large η_k, particularly if the system is chaotic. Therefore, this approach may not be suitable for demanding applications where the signal-to-noise ratio is low. We propose an alternative approach—nonlinear filtering—which is computationally demanding but in theory should have better convergence properties.

7.1.2 Nonlinear filtering

We start with a fixed order, fixed lag Volterra model. We augment the system state with the parameter to obtain an augmented state vector $\zeta_k = (x_k, x_{k-1}, \ldots, x_{k-\kappa+1}, a_0, a_1, \ldots, a_M)^T$.

The state space form now becomes

$$\zeta_{k+1} = \begin{bmatrix} \sum_{m=0}^{M} a_m z_m(k) + \eta_{k+1} \\ x_k \\ \vdots \\ x_{k-\kappa} \\ a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} \tag{7.6}$$

with the observation

$$y_k = \underbrace{[1, \overbrace{0, \cdots, 0}^{M+\kappa \text{ zeros}}]}_{\mathbf{H}_k} \begin{bmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-\kappa+1} \\ a_0 \\ a_1 \\ \vdots \\ a_M \end{bmatrix} + \nu_k. \tag{7.7}$$

Since the augmented state vector ζ_k evolves nonlinearly. We can, for example, use the particle filter to estimate simultaneously both the parameters and the conditional distribution for the signal state. The parameter estimation techniques may not take the measurement errors explicitly into account, like regression approaches, noisy measurements can produce inaccurate parameter estimates [293].

The goal is to recover the system state x given the noisy signal y. Fig. 7-1 shows an example of nonlinear state estimation of a dynamic-noise perturbed logistic map in the period-doubling region. In the figure, the true state (the "skeleton" evolution law) displays a period-2 oscillation. The noisy observations are plotted with green stars. Numerical titration of the noisy observation gives a positive value for the noise limit. We can see that the nonlinear filter state estimate gives a much closer representation to the unperturbed state. Numerical titration of the state estimate gives a *zero noise limit*, indicating that the underlying time transition law is not chaotic by itself, but that the chaos is induced by dynamic noise.

Fig. 7-2 shows another example of a chaotic logistic map ($\mu = 3.9$) contaminated with measurement noise. We can again see that the nonlinear filter was able to track the true state quite well. In this case, the noise limit of the state estimate is *higher* than that of the noisy observations. This suggests that the underlying dynamics is intrinsically chaotic to begin with, but *measurement noise* has masked some of the chaos. This explains why

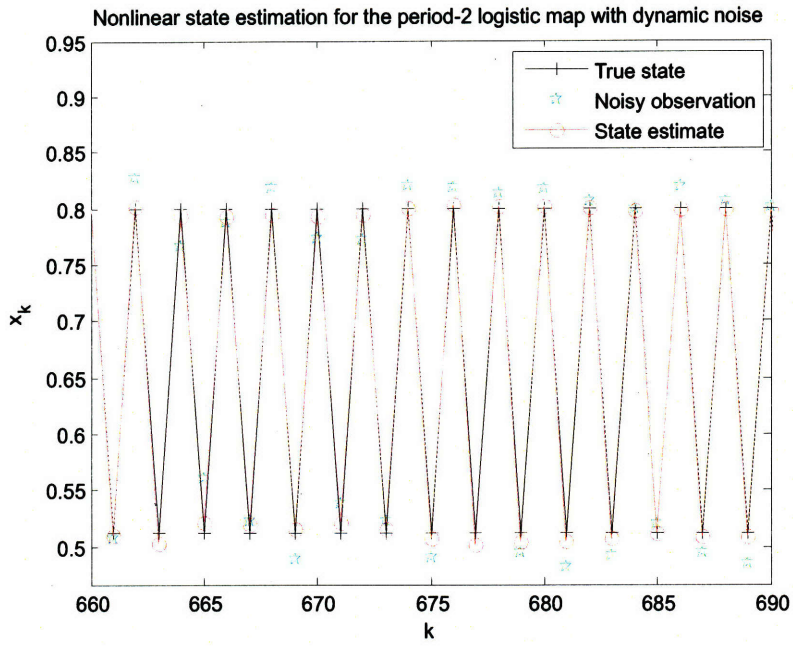


Figure 7-1: State estimation of the logistic map with $\mu = 3.2$ perturbed with dynamic noise. The true state displays a period-2 oscillation. The noisy observations are plotted with green stars and would give rise to a positive noise limit. The state estimate recovers the system dynamics of the unperturbed, nonchaotic state.

the recovered state estimate has a higher chaos level than the noisy observations. The estimates of the system parameters (Volterra series coefficients) are shown in Fig. 7-3. The expected behavior is clearly visible as the nonlinear filter correctly estimates the mean and bifurcation parameter of the logistic map.

7.2 Results

We will present the results of the noise-discrimination method using the nonlinear filtering technique. Results with the successive VAR approach will not be presented in this thesis.

7.2.1 The logistic map

We have already examined the logistic map in detail in Chapter 3 and Chapter 4. Recall that the discrete-time population grow model is described by the following equation:

$$x_{k+1} = \mu x_k(1 - x_k) \quad \forall x_k \in [0, 1], \tag{7.8}$$

In the simulation of the time series, the error term may cause x_{n+1} to be less than zero or greater than one... reset procedure. This procedure is necessary to prevent the system from diverging towards infinity. The first 300 points of the time series were discarded in order to allow for transients to decay. The results do not appear to depend on the initial value

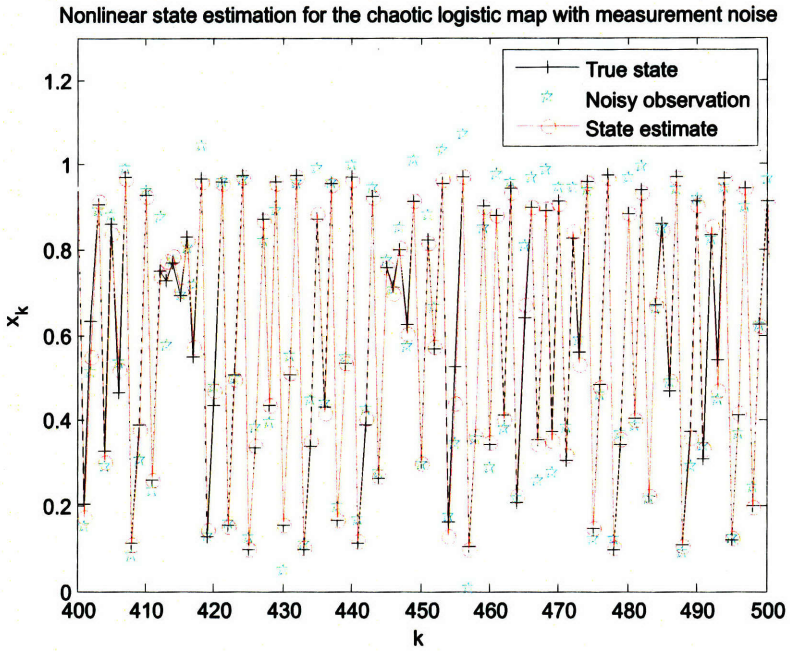


Figure 7-2: State estimation of the chaotic logistic map ($\mu = 3.9$) with measurement noise. The state estimate tracks the true state quite well.

chosen for x or the seed value of the pseudorandom number generator used to produce the error term. Times series of 1000 points were collected; no significant differences were noted in the length of the generated series. The choice of parameter for the bifurcation parameter μ in eq. (7.8) is shown in Table 7.1:

Table 7.1: Choice of parameter for the logistic map

μ	Dynamic behavior
3.2	period-2 limit cycle
3.5	period-4 limit cycle
3.568759	period-16 limit cycle
3.7	low chaos
3.8282	intermittency
3.9	high chaos

Results for the logistic map

The results for the period-2 logistic map in the presence of dynamic noise are shown Fig. 7-5[1]. As we have discussed earlier, the state estimate should always show a lower noise limit (chaos level) than the noisy time series.

[1]Figures attach at the back of the chapter for convenience.

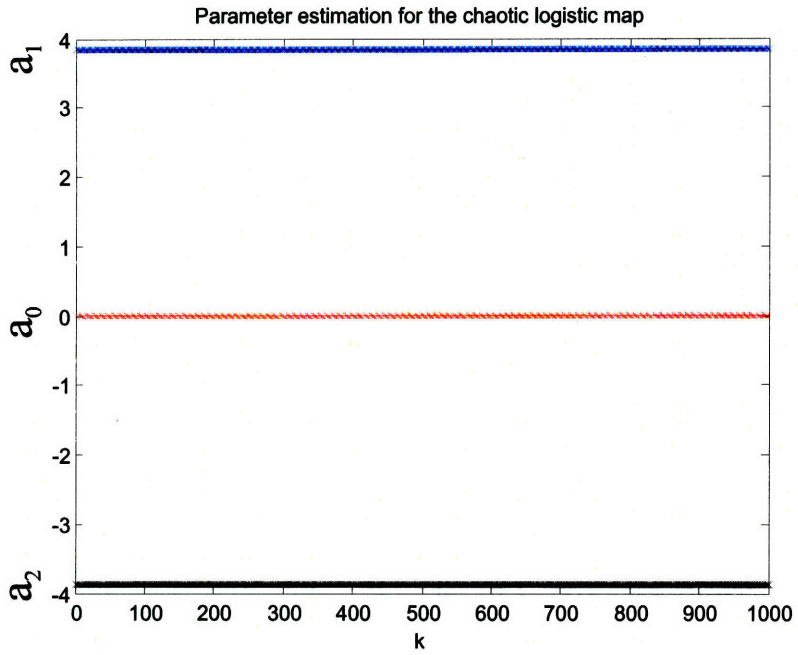


Figure 7-3: Parameter estimation of the chaotic logistic map ($\mu = 3.9$) with measurement noise. The red line is the estimate of the mean of the time series, which should be zero since we subtracted the mean off before estimation. The blue line is the estimate of the bifurcation parameter $\alpha_1 = \mu = 3.9$.

The results for the chaotic logistic map in the presence of measurement noise are shown in in Fig. 7-6 ($\mu = 3.7$), Fig. 7-7 ($\mu = 3.8282$), and Fig. 7-8 ($\mu = 3.9$). The blue dots are the noise limit values for the noisy time series. Ultimately the graph of the noise limit against noise intensities must fall to zero since for very large values of the measurement noise the original system dynamics will be completely obscured and there will just be noise. The noise limit values of the state estimate, indicated by the red dots, are always higher than the blue. This shows that our nonlinear filtering technique is capable of recovering at least some chaotic dynamics.

7.2.2 The Hénon map

The Hénon map is a discrete-time dynamical system. The map was introduced by Michel Hénon [129] as a simplified model of the Poincaré section of the Lorenz model [180]. It is one of the most studied examples of dynamical systems that exhibit chaotic behavior. The Hénon map takes a point (x, y) in the plane and maps it to a new point according to

$$\begin{aligned} x_{k+1} &= y_k + 1 - ax_k^2, \\ y_{k+1} &= bx_k. \end{aligned} \tag{7.9}$$

The map depends on two adjustable parameters: a, which controls the type of the attractor and determines whether the trajectories diverge to infinity; and b, which changes the visi-

bility of fine structure of the attractor by controlling the level of dissipation of the system. For the **canonical Hénon map**, a and b take on values of 1.4 and 0.3, respectively. For these parameter values, the map is well into the chaotic region; an initial point of the plane will either approach a set of points known as the Hénon strange attractor, or diverge to infinity [270]. The Hénon attractor is a fractal, smooth in one direction and a Cantor set in another. Numerical estimates yield a correlation dimension of 1.42 ± 0.02 [114], and a Hausdorff dimension of 1.261 ± 0.003 for the attractor of the canonical map [244]. For other values of a and b the map may converge to a periodic orbit, exhibit intermittency, or chaotic behavior. In Table 7.2, we show the parameters for a and b used in this study.

Table 7.2: Choice of parameters for the Hénon map

a	b	Dynamic behavior
1	0.3	limit cycle
1.08	0.3	low chaos
1.25	0.3	nonchaotic
1.4	0.3	high chaos

Results for the Hénon map

The results for the dynamically perturbed cases are shown in Fig. 7-9 ($a = 1$, $b = 0.3$), Fig. 7-10 ($a = 1.08$, $b = 0.3$), and Fig. 7-11 ($a = 1.25$, $b = 0.3$). The results for the measurement noise cases are shown in Fig. 7-12 ($a = 1.08$, $b = 0.3$), and Fig. 7-13 ($a = 1.4$, $b = 0.3$). The interpretation of the results for the Hénon map is the same as for the logistic model. Specifically, after nonlinear filtering of a dynamic-noise-perturbed time series, the noise limit of the state estimate should be lower than that of the noisy time series, indicating the presence of chaotic dynamics induced by stochasticity. On the other hand, the noise limit of the state estimate of a measurement noise perturbed time series show be higher than that of the noisy time series. That is, the nonlinear filter has successfully countered the blurring effect of measurement noise and has recovered the chaotic dynamic that was obscured by measurement noise.

7.2.3 The Mackey-Glass equation

The Mackey-Glass equation models dynamics of white blood cell production in the human body [183]. Because rates of stem cell proliferation entail a time delay, periodic dynamics and chaos can obtain. Indeed, Mackey and Glass have suggested that long-term fluctuations in cell counts observed in certain forms of leukemia are evidence for these behaviors in vivo. It has been shown that the equation displays a broad diversity of dynamical behavior including limit cycle oscillations, and chaotic solutions.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{ax(t-\tau)}{1+x^c(t-\tau)} - bx(t). \tag{7.10}$$

For $a = 0.2$, $b = 0.1$ $c = 8$, and $\tau = 100$, eq. (7.10) has an attractor whose estimated dimension is 7.5 [113].

Results for the Mackey-Glass equation

Fig. 7-14 shows the noise limits of the noisy time series and of the state estimate as a function of noise intensities. We see that noise limit values are not significantly different for the two cases. In this case, the fixed-order Volterra structure was inadequate in capturing the highly nonlinear behavior of the Mackey-Glass series.

7.2.4 The ecological map

The ecological map is a model of species dispersal in evolutionary ecology. It addresses the important question of how does spatial movement of species in the environment affect the population. The model exhibits bifurcation towards high-dimensional chaos [125] [126]. Strange attractors are split into several disconnected pieces. Such bifurcations are common in ecological models.

In patch j at generation t, $N_{ij}(t)$ represents the density of population of clone i, where $i, j = 1, 2$. Clones differ only in a fixed dispersal rate, e_i, which is defined as the fraction of individuals dispersing from their natal patch at each generation. Assume that the realized fitness in patch j, $W_j\left[N_{Tj}(t)\right]$, is identical for individuals of both clones and depends functionally on the summed abundances, $N_{Tj}(t) = N_{1j}(t) + N_{2j}(t)$, of all clones in patch j. The dynamics of clone i in patch j are governed by the following model [137]:

$$\begin{aligned} N_{11}(t+1) &= (1-e_1)W_1\left[N_{T1}(t)\right]N_{11}(t) + me_1W_2\left[N_{T2}(t)\right]N_{12}(t), \\ N_{12}(t+1) &= (1-e_1)W_2\left[N_{T2}(t)\right]N_{12}(t) + me_1W_1\left[N_{T1}(t)\right]N_{11}(t), \end{aligned} \tag{7.11}$$

where m is the fraction of dispersers that survive to enter their nonnatal patch. Similar equations describe clone 2. The fitness is given by the following relation:

$$W_j = \exp\left[r_j\left(1 - \sum_i \frac{N_{ij}}{K_{ij}}\right)\right], \tag{7.12}$$

where K_j is the carrying capacity of patch j. Regarding the two interacting subsystems $\{N_{11}(t), N_{12}(t)\}$ and $\{N_{21}(t), N_{22}(t)\}$, where subsystem A is the low-dispersal clone and subsystem B is the high-dispersal clone, the full system, which is a four-dimensional map, becomes high-dimensionally chaotic. The ecological significance is that chaotic dynamics in fact favor the evolution of dispersal. When the dynamics of clone 2 are not chaotic, the population of clone 1 becomes zero asymptotically, indicating extinction of the species with

high-dispersal rate. However, when the dynamics of clone 2 becomes chaotic, the dynamics of clone 1 becomes chaotic as well, with nonzero population densities in both patches.

Physically, a subset of dynamical variables becomes chaotic via some known route to low-dimensional chaos such as the period-doubling bifurcation route, after which the complementary subset also becomes chaotic. As a consequence, the latter subset of dynamical variables become chaotic in a relatively abrupt fashion.

Choice of parameters for the ecological map

Consider the following equations for the noisy density-dependent two age-class model:

$$\begin{aligned} x_k &= \lambda y_{k-1} e^{-0.001(y_{k-1}+x_{k-1})} + \eta_{x,k}, \\ y_k &= 0.2x_{k-1} e^{-0.07(y_{k-1}+x_{k-1})} + 0.8y_{k-1} e^{-0.05(y_{k-1}+0.5x_{k-1})} + \eta_{y,k} \end{aligned} \tag{7.13}$$

where λ is a parameter and η_x, η_y are independent white Gaussian noise with zero mean and variance σ^2. Let us first examine the noise-free model. For $\lambda = 117$, the motion is periodic, as shown in Fig. 7-4. For $\lambda = 118$, the trajectory evolves in an attractor comprised of fractals in eight disconnected domains and visits each of them in a periodic manner as indicated by the arrows in Fig. 7-4. This strong global periodicity makes detection of the chaotic component difficult. In fact, if one tries to identify chaotic dynamics with the method Sugihara and May [272], one finds that the chaos is hidden in the overall periodicity of the attractor and it would be classified as "residual noise" [44].

Results for the ecological map

The results of the nonlinear filtering algorithm is shown in Fig. 7-15. The eight "islands" of the ecological attractor are of very different size. In fact, the attractor located at the lower left corner is very much like a single point. Magnification shows this is not really the case, as shown in Fig. 7-4 (d), this nevertheless constitutes another dimension of difficulty in successfully characterizing the global attractor [281].

7.3 Limitations and implementation considerations

We have seen that Volterra series expansion represent an important model for the representation, analysis and synthesis of weakly nonlinear dynamic systems. However, a significant difficulty we have encountered with this approach to kernel identification is the complexity due to the large number of coefficients that characterize such models. The number of Volterra coefficients increases geometrically as the memory or order increases, which also makes cumbersome the design and implementation based upon such models. Consequently, practical implementations are limited to severely truncated Volterra models consisting of, at most, second degree terms only [73]. It is therefore desired to reduce the filter scale

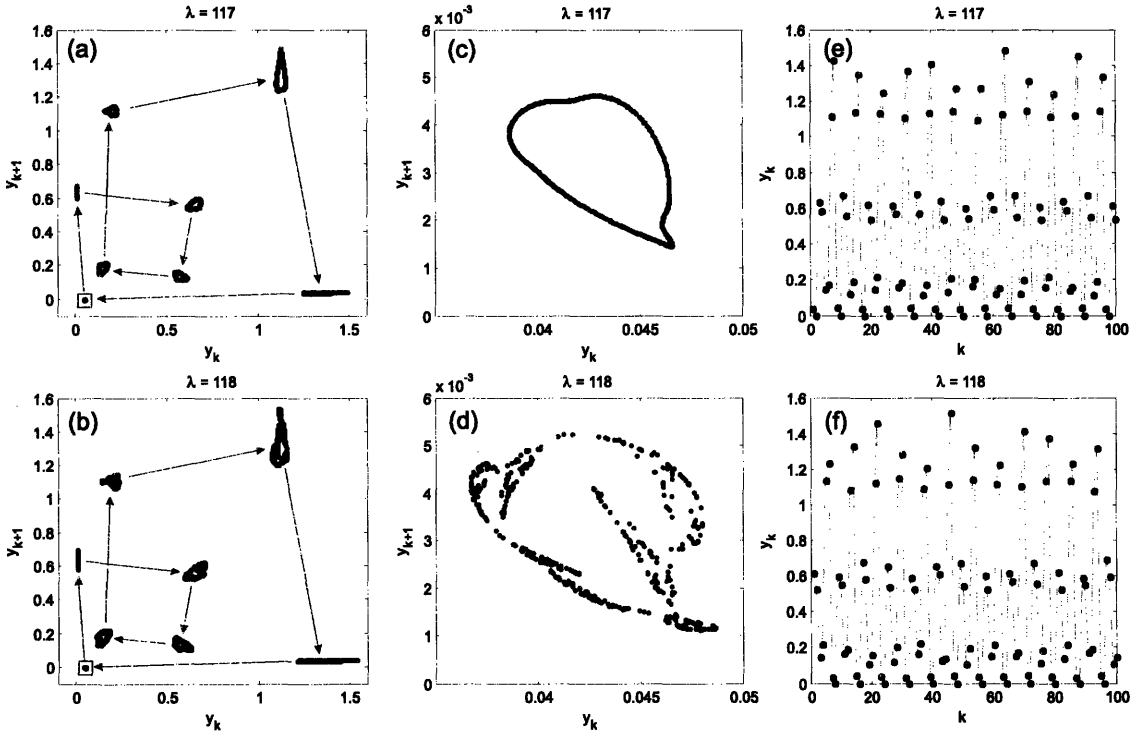


Figure 7-4: Phase diagrams for the ecological map, (a) $\lambda = 117$, (b) $\lambda = 118$. The "island" located at the lower left corner of the figure are replotted in (c) and (d). The map shows period motion when $\lambda = 117$ and fractal features when $\lambda = 118$. (e) and (f) show the typical orbits associated with the parameters above. The trajectory for $\lambda = 118$ shows that even when the map is chaotic, there is still a predominant periodic behavior.

without degrading the performance. The first approach transforms the Volterra model to parallel-cascade structures by matrix decomposition techniques [163] [165]. Parallel-cascade realizations implement higher order Volterra systems as a parallel connection of multiplicative combinations of lower order truncated Volterra systems.

The method begins by approximating the system with first cascade of a dynamic linear element followed by a static nonlinearity. The residual is treated as the output of a new nonlinear system, driven by the same input, and a second cascade is found to approximate the latter system. Each time, the static nonlinearities are best-fit to the latest residual. Nonlinear dynamic systems with fading memory can be approximated to an arbitrary accuracy, in the mean-square sense, by such truncated Volterra models. Furthermore, the parallel-cascade model can readily be converted into a corresponding Volterra series [163]. The PCI is generally faster, particularly when longer memory lengths, or higher-degree nonlinearities, are required. This approach has illustrated its utility for protein family prediction [164] [166]. The parallel-cascade structure has been extended to adaptive structure where the eigenvalues and vectors are adjusted adaptively [215].

The second approach exploits the idea of the **reproducing kernel Hilbert spaces (RKHS)** [12], that is, to use a particular reproducing kernel to summarize the complete

Volterra series in a linearized feature space, which leads to a more parsimonious estimation problem [64]. The solution to approximation in the Volterra RKHS with respect to a large class of loss functions was shown to be simply a linear combination of a set of kernel functions. The main reason for using the RKHS approach is that the estimation complexity of the implicit representation is linear in the input dimensionality and independent of the degree of nonlinearity. Experiments show performance advantages in terms of convergence, interpretability, and system sizes that can be handled [78]. This can therefore represent a significant reduction over the standard Volterra series case. This computationally efficient approach is related to support vector machine approach to nonlinear regression (see Appendix D).

Recursive filter techniques can also enable great scale reduction. However, the main problem associated with recursive nonlinear Volterra systems is that their stability is not guaranteed. Moreover, the stability condition depends not only on their coefficients but also on their input signals [149]. The filter design is consequently very difficult because the stability condition varies. In this regard, some stabilization techniques of recursive nonlinear filters have been reported [205].

7.4 Summary

With the nonlinear filtering method, we can answer the question: can the deterministic skeleton of a complex nonlinear system be experimentally identified as equilibrium, periodic, or chaotic? The main advantage of using nonlinear filters compared to the successive VAR method is a better quantification of the underlying chaos level of the original unperturbed system, where as successive VAR can only produce a binary result (dynamic noise or measurement noise). For example, in the measurement noise case, nonlinear filtering improves the sensitivity for the numerical-titration technique. The nonlinear filters also allows for the presence of both types of noise in the system specification. That means we can readily apply this method to experimental data, since in the real-world setting, the collected data is likely to contain a mixture of both dynamic noise and measurement artifacts. But often we deal with one kind of predominant noise and the influence of the other kind of noise on dynamics remains relatively small. In addition, nonlinear filtering has better convergence properties than the successive VAR approach.

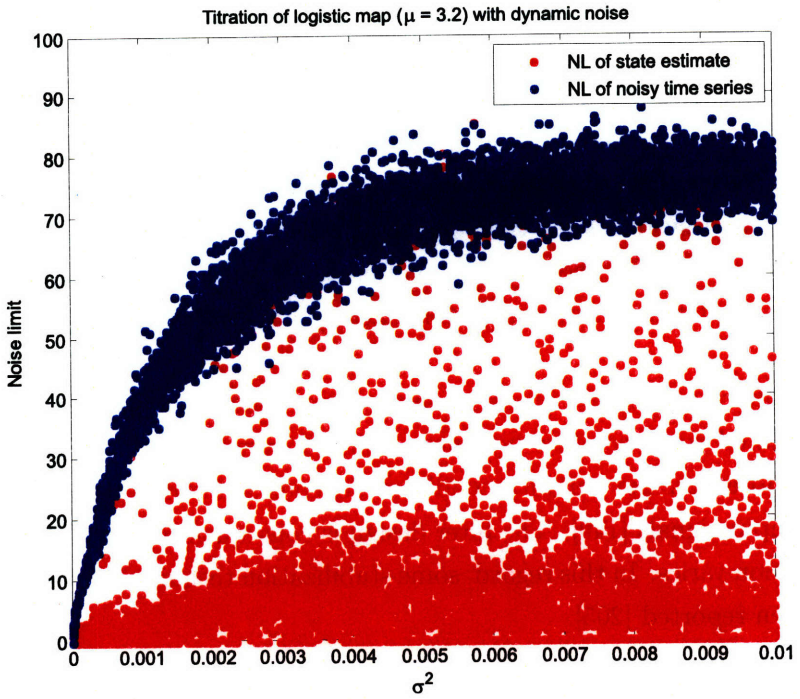


Figure 7-5: Titration and filtering of the periodic logistic map ($\mu = 3.2$) with dynamic noise.

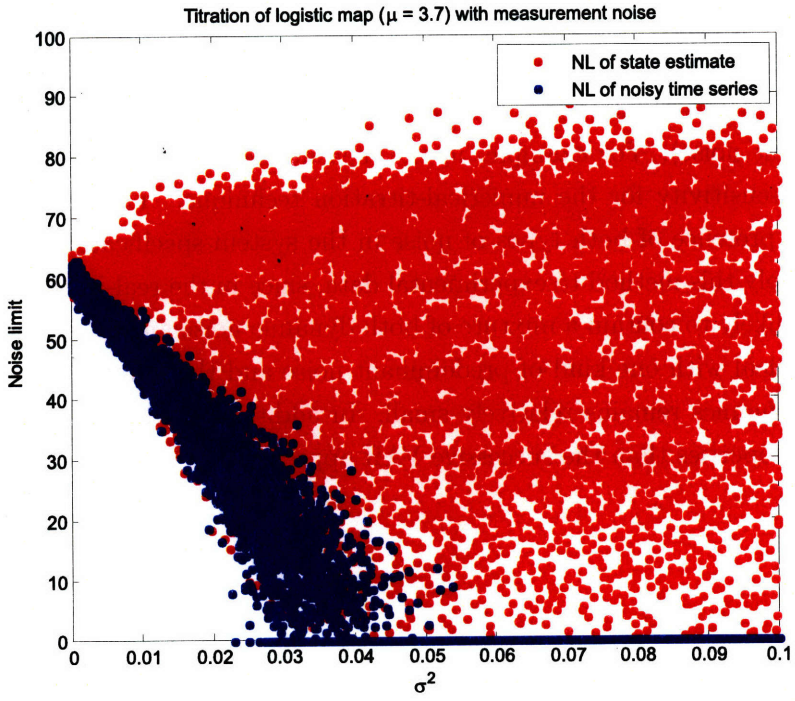


Figure 7-6: Titration and filtering of the weakly chaotic logistic map ($\mu = 3.7$) with measurement noise.

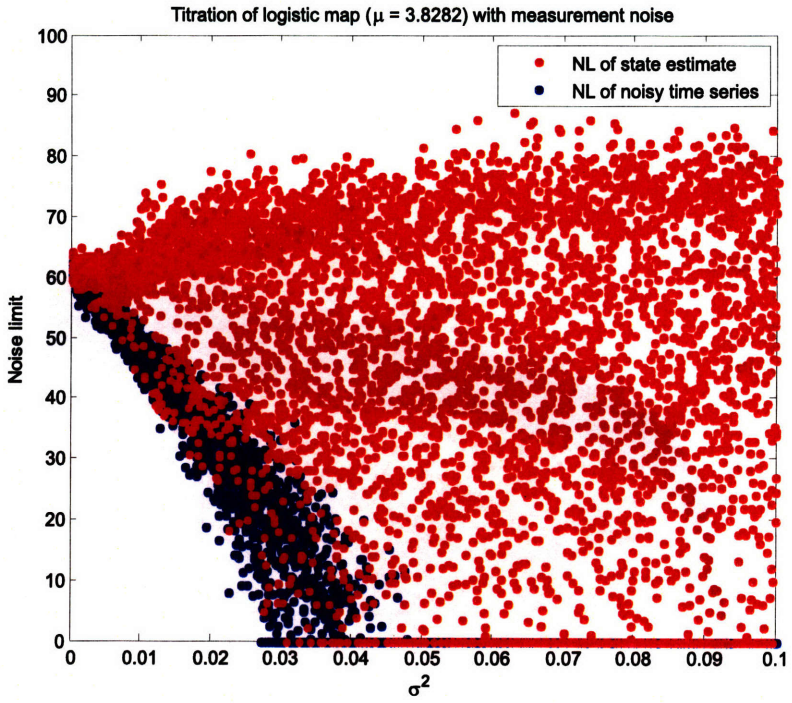


Figure 7-7: Titration and filtering of the intermittently chaotic logistic map ($\mu = 3.8282$) with measurement noise.

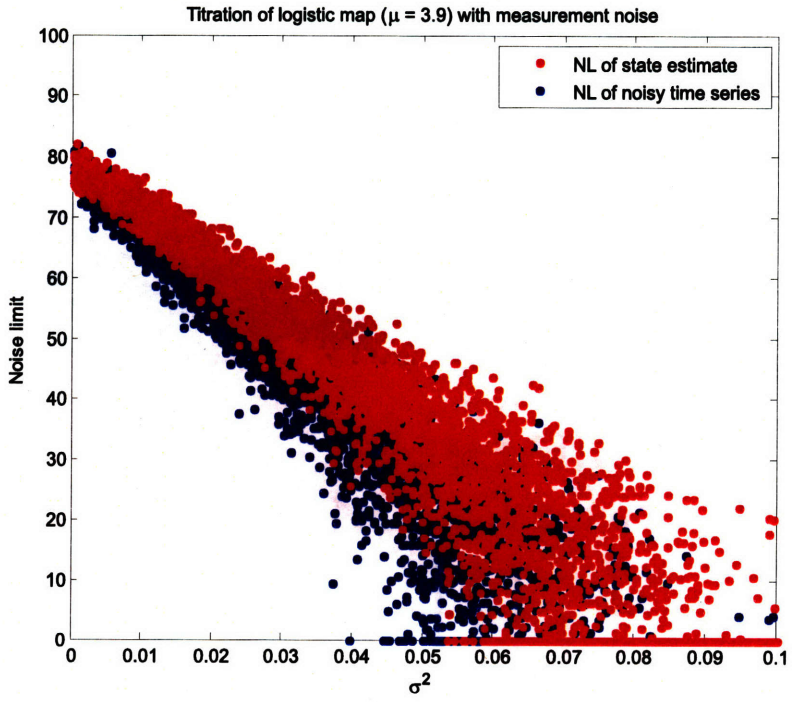


Figure 7-8: Titration and filtering of the deeply chaotic logistic map ($\mu = 3.9$) with measurement noise.

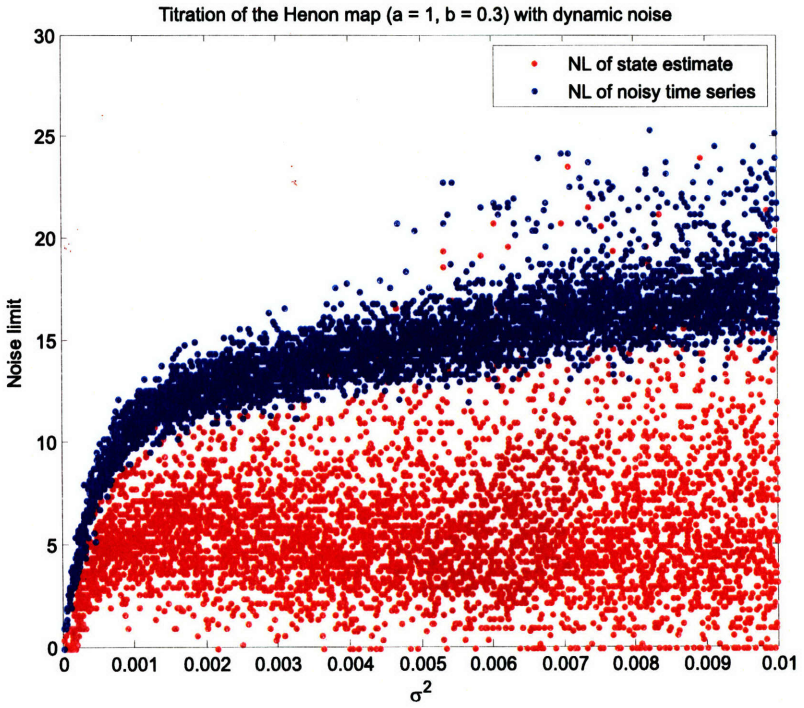


Figure 7-9: Titration and filtering of the periodic Hénon map ($a = 1$) with dynamic noise.

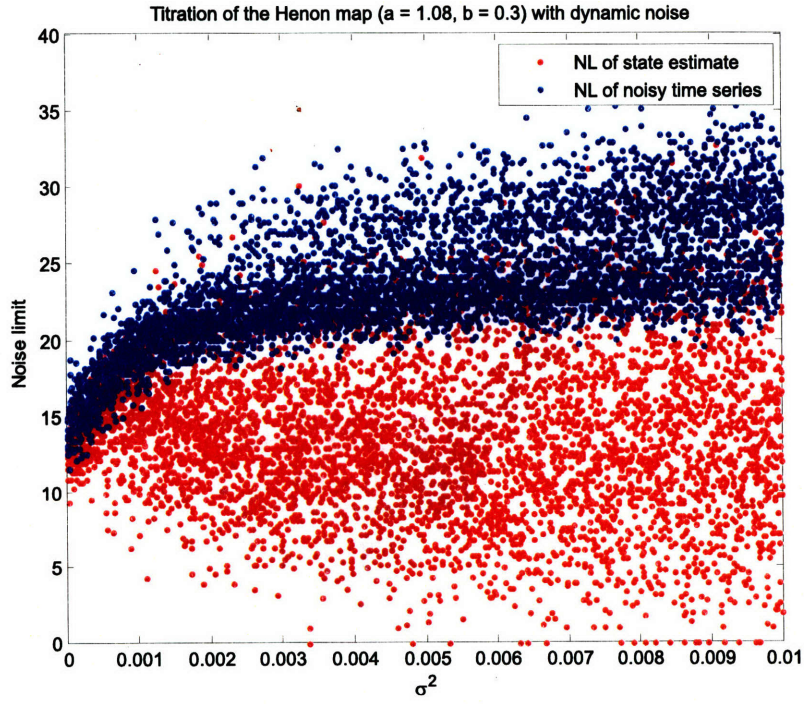


Figure 7-10: Titration and filtering of the weakly chaotic Hénon map ($a = 1.08$) with dynamic noise.

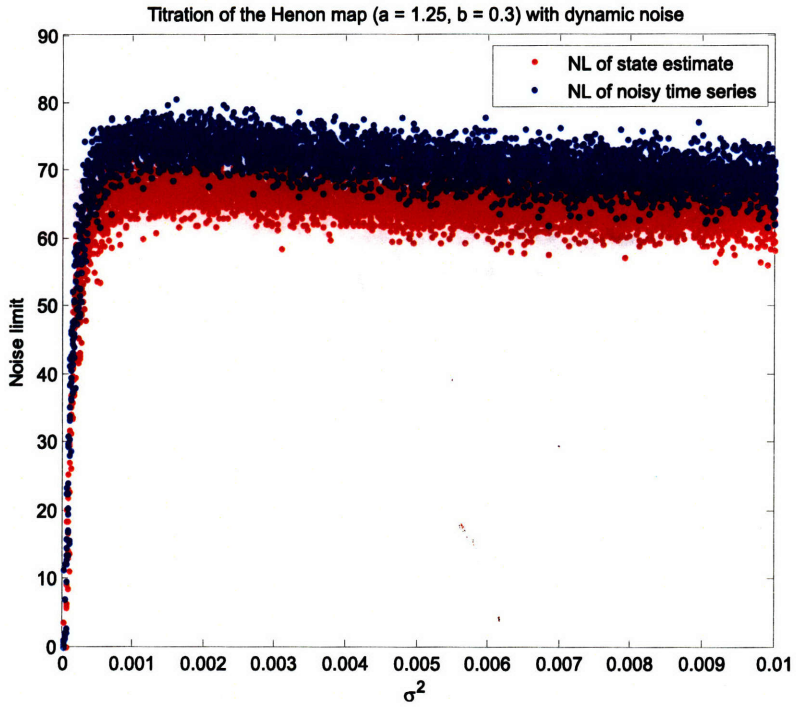


Figure 7-11: Titration and filtering of the nonchatic Hénon map ($a = 1.25$) with dynamic noise.

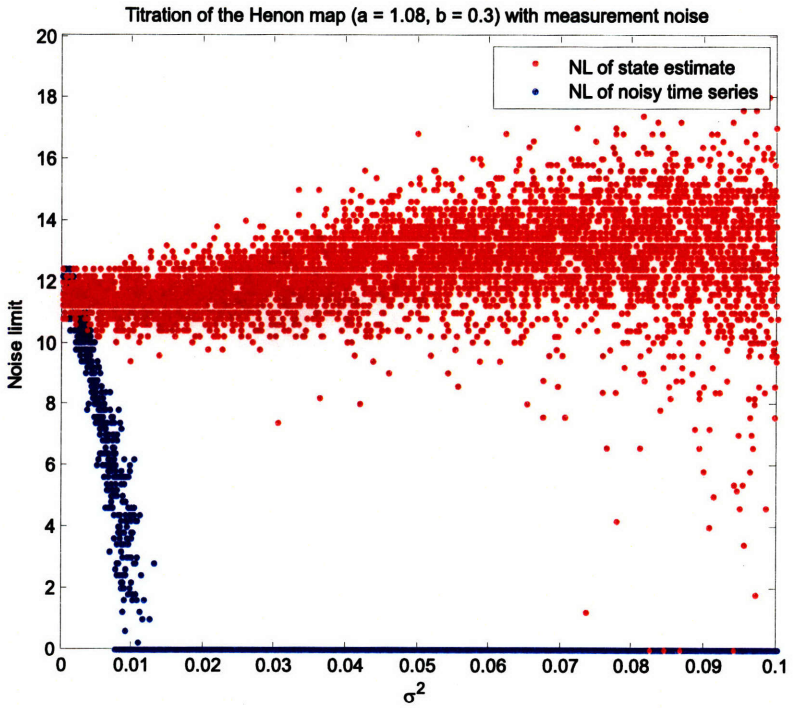


Figure 7-12: Titration and filtering of the weakly chaotic Hénon map ($a = 1.08$) with measurement noise.

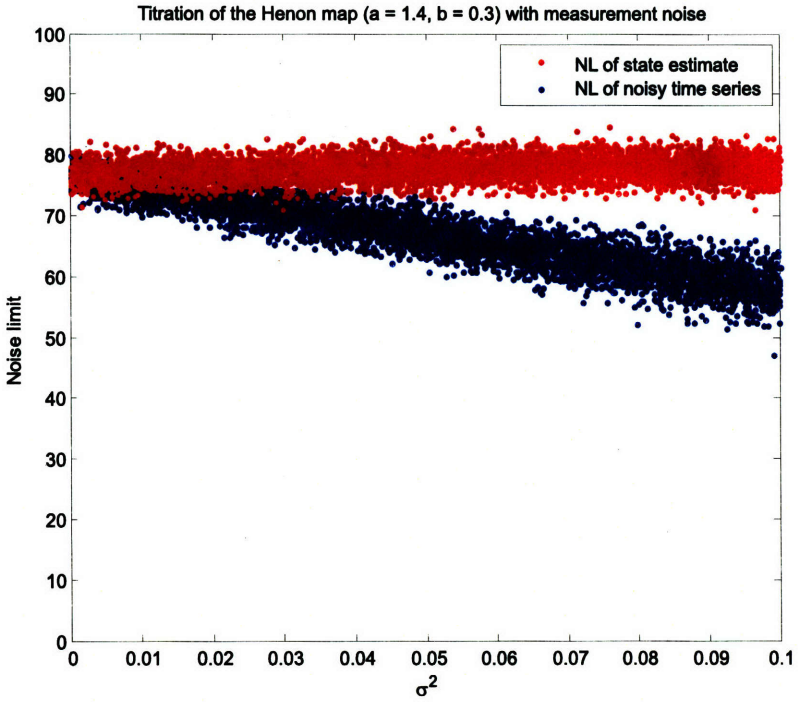


Figure 7-13: Titration and filtering of the highly chaotic Hénon map ($a = 1.4$) with measurement noise.

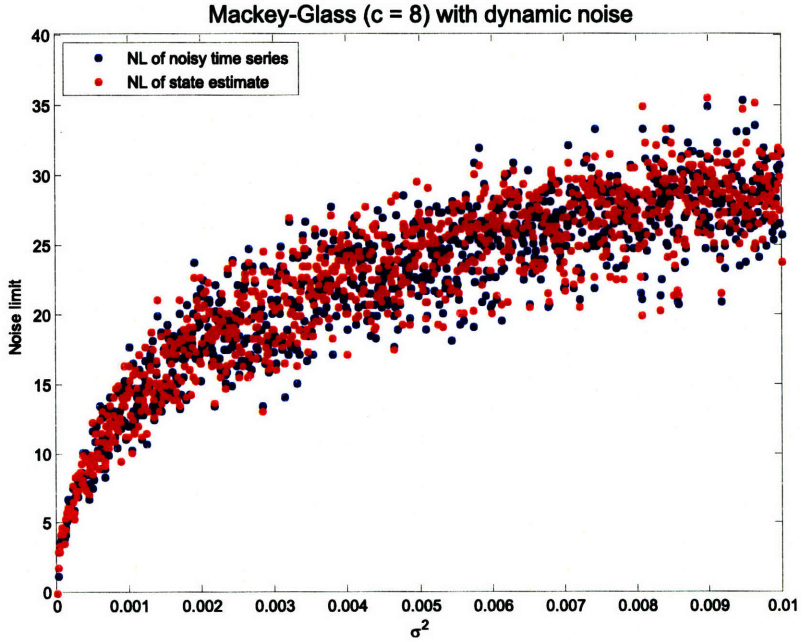


Figure 7-14: Titration and filtering of the Mackey-Glass (c = 8) equation.

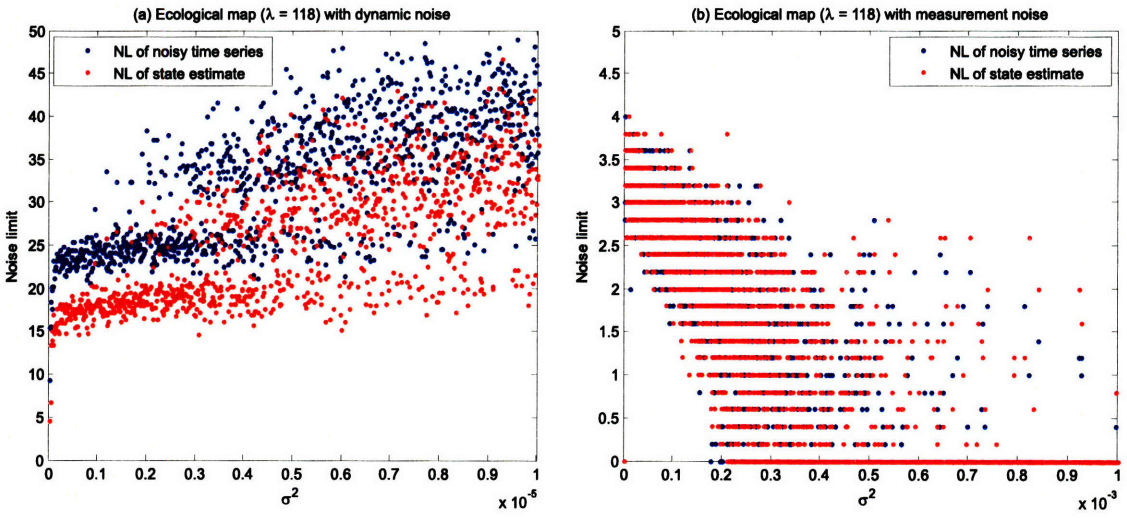


Figure 7-15: (a) Titration and filtering of the ecological map with dynamic noise, (b) Titration and filtering of the ecological map with measurement noise.

Chapter 8

Concluding Remarks

> The Answer to the Ultimate Question of Life, the Universe and Everything is.....42.
>
> —Deep Thought (after 7.5 million years of calculation) [3]

In this final chapter, we will highlight some key ideas presented in this thesis. We will also discuss some of the challenges and outline some of the future prospects of this exciting research area.

8.1 Conclusions

We close this thesis by summarizing the importance of the approach of using simple models and discussing more generally why and to what extent they can give useful information about real systems whose behavior is determined by complicated many-body interactions. Frequently we encounter deterministic systems which we cannot fully describe, one could make simplifying assumption and then try to solve the approximated system, or one could start from much simpler model systems and generalize to more complicated systems. The prime advantage of using model systems is that they can be chosen to be tractable theoretically and therefore the details of their behavior can be understood with some confidence. In particular, the aim is to extract a clear understanding of the physics leading to the behavior which will be mirrored in the real system. Once the basic principle have been established, various refinements can be included in the models. Often it is possible to go further than this and obtain a quantitative fit to experimental data, rather than just a qualitative understanding of it.

In this thesis, we have shown that a simple model such as the logistic map in the period-2 limit cycle regime can be used to predict the correct temperature dependence of the kinetic crossover rate of bistable systems in a fluctuating environment. By simulating

the dynamics of the one-dimensional map, we found that the escape rate scales exponentially as the Arrhenius factor, which shows a nontrivial dependence on the noise strength. This result is consistent with first principal calculations based on the Fokker-Planck theory. It goes beyond stability analysis, we have demonstrated that some of the properties of thermodynamical systems can be understood by studying noise-induced complexity and nonlinear dynamics. We have also given a new definition of **stochastic chaos** and developed nonlinear-filtering techniques for its detection in physical nonlinear systems. Our analysis and theoretical viewpoint are not only philosophically appealing, but also practical and undoubtedly will become more applicable in the future.

8.2 Future prospects

Each kinetic equation necessarily involves a stochastic assumption such as Boltzmann's **Stosszahlansatz** (molecular chaos), which is introduced by some truncation of the evolution equation either in the low density limit or in the weak coupling limit. Recently, adequate constructions have been proposed in the theory of chaotic dynamics, which overcome the previous difficulties caused by the stochastic assumptions. Since 1990s, stochastic Melnikov method has been applied to study the effect of noise on homoclinic or heteroclinic bifurcation and noise-induced chaos [35] [82] [83]. A simple criterion guaranteeing the nonoccurrence of chaos was derived for continuous time systems using the stochastic Melnikov approach [260]. Gan [89] characterized noise-induced chaotic response in the quadratically nonlinear oscillator by demonstrating that the boundary of the safe basin of the system can be incursively fractal when Gaussian white noise excitation is added to the system [90] [91].

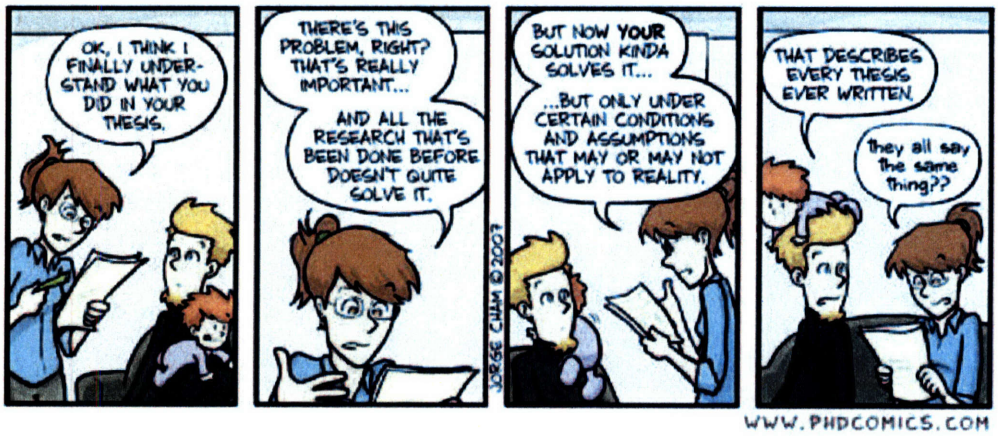
Systems driven by chaos are still relatively unexplored, especially with respect to the Kramers problem, as most research on particle escape from stationary metastable nonequilibrium systems are carried out based on stochastic noise. Recently, Chew et al. [49] [50] studied the resonant effects of chaotic fluctuations on a strongly damped particle in a bistable potential drive by weak sinusoidal perturbation. Their first-order analysis reveals that the transition rate has the form of the Kramers escape rate except for a perturbed prefactor. This modification to the prefactor is found to arise from the statistical asymmetry of the chaos-generated noise.

We have no touch on quantum version of the classical chaotic dynamics discussed in this thesis. However, nature is fundamentally quantum mechanical. So in so far as this material has any relevance for natural processes, a quantum version is certainly necessary. Fundamental equation of quantum mechanics, the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{\mathcal{H}}\psi, \tag{8.1}$$

can be written as a system of hydrodynamic equations without additional assumptions.

Other techniques of solving the Langevin's equation, such as path integration have also been used, drawing on the analogy between statistical physics and quantum mechanics, for example, the Fokker-Planck equation can be transformed into the Schrödinger equation by rescaling a few variables. One would very much like to know something about the quantum mechanical quantities that play a role similar to that of Lyapunov exponents, Kolmogorov-Sinai entropies, etc., in classical physics. The stochastic resonance behavior in a classical system has also found to manifest itself in its quantum counterpart [156] [232]. It has been shown that the dynamics of the Fokker-Planck equation governing the system is equivalent to that of a quantum system governed by a Hamiltonian that can be obtained from the FPE through an appropriate transformation. When the Hamiltonian system exhibits the transition to chaos, the decay rates of the FPE show level repulsion [200]. This is all open for investigation.



Appendix A

Matrix Inversion Lemma

We will now derive the **matrix inversion lemma**, which we will use many times.

Suppose we have the partitioned matrix $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$ where $\mathbf{A}$ and $\mathbf{D}$ are invertible square matrices (the $\mathbf{B}$ and $\mathbf{C}$ matrices may or may not be square). We define $\mathbf{E}$ and $\mathbf{F}$ matrices as follows:

$$\begin{aligned} \mathbf{E} &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}, \\ \mathbf{F} &= \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}. \end{aligned} \tag{A.1}$$

Assume that $\mathbf{E}$ is invertible. Then we can show that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{E}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{A.2}$$

Now assume that $\mathbf{F}$ is invertible. Then we can show that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{F}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{F}^{-1} & \mathbf{E}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{A.3}$$

Since the two expressions in eqs. (A.2) and (A.3) are inverses of the same matrix, they must be equal. We therefore conclude that the upper-left partitions of the matrices are equal. Substituting for the definition of $\mathbf{F}$, we obtain

$$\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1}\mathbf{C}\mathbf{A}^{-1}. \tag{A.4}$$

This is the matrix inversion lemma.

Appendix B

RTS Smoother Lemmas

Here we prove several lemmas to aid the development of the RTS smoother in Sec. 6.1.2.

Lemma 1.

$$\mathbf{F}_{k-1}^{-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T} = \mathbf{F}_{k-1}^{-1}\boldsymbol{\Lambda}_{fk}^{-}\mathbf{F}_{k-1}^{-T} - \boldsymbol{\Lambda}_{f,k-1}^{+}. \tag{B.1}$$

Proof. From eq. (6.9) we see that

$$\boldsymbol{\Lambda}_{fk}^{-} = \mathbf{F}_{k-1}\boldsymbol{\Lambda}_{f,k-1}^{-}\mathbf{F}_{k-1}^{T} + \mathbf{Q}_{k-1}. \tag{B.2}$$

Rearranging this equation gives

$$\mathbf{Q}_{k-1} = \boldsymbol{\Lambda}_{fk}^{-} - \mathbf{F}_{k-1}\boldsymbol{\Lambda}_{f,k-1}^{-}\mathbf{F}_{k-1}^{T}. \tag{B.3}$$

Premultiplying both sides by $\mathbf{F}_{k-1}^{-1}$ and postmultiplying both sides by $\mathbf{F}_{k-1}^{-T}$ gives the desired results. □

Lemma 2. *The a posteriori covariance $\boldsymbol{\Lambda}_{bk}^{+}$ of the backward filter satisfies the equation*

$$\boldsymbol{\Lambda}_{bk}^{+} = (\boldsymbol{\Lambda}_{fk}^{-} + \boldsymbol{\Lambda}_{bk}^{+})\mathcal{I}_{fk}^{-}\boldsymbol{\Lambda}_{k}. \tag{B.4}$$

Proof. From eq. (6.17) we obtain

$$\begin{aligned}
\mathbf{I} &= (\mathcal{I}_{bk}^{+} + \mathcal{I}_{fk}^{-})\boldsymbol{\Lambda}_{k}, \\
\boldsymbol{\Lambda}_{bk}^{+} &= (\mathbf{I} + \boldsymbol{\Lambda}_{bk}^{+}\mathcal{I}_{fk}^{-})\boldsymbol{\Lambda}_{k} \\
&= \boldsymbol{\Lambda}_{k} + \boldsymbol{\Lambda}_{bk}^{+}\mathcal{I}_{fk}^{-}\boldsymbol{\Lambda}_{k} \\
&= \boldsymbol{\Lambda}_{fk}^{-}\mathcal{I}_{fk}^{-}\boldsymbol{\Lambda}_{k} + \boldsymbol{\Lambda}_{bk}^{+}\mathcal{I}_{fk}^{-}\boldsymbol{\Lambda}_{k} \\
&= (\boldsymbol{\Lambda}_{fk}^{-} + \boldsymbol{\Lambda}_{bk}^{+})\mathcal{I}_{fk}^{-}\boldsymbol{\Lambda}_{k}.
\end{aligned} \tag{B.5}$$

□

Lemma 3. *The covariances of the forward and backward filters satisfy the equation*

$$\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+} = \mathbf{F}_{k-1}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})\mathbf{F}_{k-1}^{T}. \tag{B.6}$$

Proof. From eqs. (6.9) and (6.15) we see that

$$\begin{aligned}
\mathbf{\Lambda}_{f,k-1}^{+} &= \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{fk}^{-}\mathbf{F}_{k-1}^{-T} - \mathbf{F}_{k-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T}, \\
\mathbf{\Lambda}_{b,k-1}^{-} &= \mathbf{F}_{k-1}^{-1}\mathbf{\Lambda}_{bk}^{+}\mathbf{F}_{k-1}^{-T} + \mathbf{F}_{k-1}\mathbf{Q}_{k-1}\mathbf{F}_{k-1}^{-T}.
\end{aligned} \tag{B.7}$$

Adding these two equations and rearranging, after the smoke clears, we get

$$\begin{aligned}
\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-} &= \mathbf{F}_{k-1}^{-1}(\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+})\mathbf{F}_{k-1}^{-T}, \\
\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+} &= \mathbf{F}_{k-1}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})\mathbf{F}_{k-1}^{T}.
\end{aligned} \tag{B.8}$$

□

Lemma 4. *The smoothed estimate $\hat{\mathbf{z}}_k$ can be written as*

$$\hat{\mathbf{z}}_k = \mathbf{\Lambda}_k\mathcal{I}_{fk}^{+}\hat{\mathbf{z}}_{fk}^{-} - \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\varrho_k^{+}. \tag{B.9}$$

Proof. From eqs. (6.11) and (6.18) we have

$$\begin{aligned}
\hat{\mathbf{z}}_k &= \mathbf{\Lambda}_k\mathcal{I}_{fk}^{+}\hat{\mathbf{z}}_{fk}^{+} + \mathbf{\Lambda}_k\mathcal{I}_{bk}^{-}\hat{\mathbf{z}}_{bk}^{-} \\
&= \mathbf{\Lambda}_k\mathcal{I}_{fk}^{+}\hat{\mathbf{z}}_{fk}^{+} + \mathbf{\Lambda}_k\mathcal{I}_{bk}^{-}\varrho_k^{-}.
\end{aligned} \tag{B.10}$$

From eq. (6.15) we see that

$$\varrho_k^{-} = \varrho_k^{+} - \mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{y}_k. \tag{B.11}$$

Substitute this expression for ϱ_k^{-}, and the expression for $\hat{\mathbf{z}}_{fk}^{+}$ from eq. (6.9), into eq. (B.10) to obtain

$$\hat{\mathbf{z}}_k = \mathbf{\Lambda}_k\mathcal{I}^{+}\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\mathcal{I}_{fk}^{+}\mathbf{K}_{fk}(\mathbf{y}_k - \mathbf{H}_k\hat{x}_{fk}^{-}) + \mathbf{\Lambda}_k\varrho_k^{+} - \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{y}_k. \tag{B.12}$$

Now substitute $\mathbf{\Lambda}_{fk}^{+}\mathbf{H}_k^{T}\mathbf{R}_k^{-1}$ for $\mathbf{K}_{fk}$ (from eq. (6.9)), turn the page upside down and squint, you obtain

$$\begin{aligned}
\hat{\mathbf{z}}_k &= \mathbf{\Lambda}_k\mathcal{I}^{+}\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\mathcal{I}_{fk}^{+}\mathbf{\Lambda}_{fk}^{+}\mathbf{H}_k^{T}\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_{fk}^{-}) + \mathbf{\Lambda}_k\varrho_k^{+} - \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{y}_k \\
&= \mathbf{\Lambda}_k\mathcal{I}^{+}\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{z}}_{fk}^{-}) + \mathbf{\Lambda}_k\varrho_k^{+} - \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{y}_k \\
&= \mathbf{\Lambda}_k\mathcal{I}^{+}\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\mathbf{H}_k^{T}\mathbf{R}_k^{-1}\mathbf{H}_k\hat{\mathbf{z}}_{fk}^{-} + \mathbf{\Lambda}_k\varrho_k^{+}.
\end{aligned} \tag{B.13}$$

□

Lemma 5.

$$(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1} = \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}(\mathbf{\Lambda}_{fk}^{-} - \mathbf{\Lambda}_{k})\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}. \tag{B.14}$$

Proof. Recall from eqs. (5.15) and (6.15) that

$$\begin{aligned}\mathcal{I}_{fk}^{+} &= \mathcal{I}_{fk}^{-} + \mathbf{H}_{k}^{T}\mathbf{R}_{k}^{-1}\mathbf{H}_{k},\\ \mathcal{I}_{bk}^{-} &= \mathcal{I}_{bk}^{-} + \mathbf{H}_{k}^{T}\mathbf{R}_{k}^{-1}\mathbf{H}_{k}.\end{aligned} \tag{B.15}$$

Combining these two equations gives

$$\begin{aligned}\mathcal{I}_{bk}^{+} &= \mathcal{I}_{bk}^{-} + \mathcal{I}_{fk}^{+} - \mathcal{I}_{fk}^{-}\\ &= \left[(\mathcal{I}_{bk}^{-} + \mathcal{I}_{fk}^{+})^{-1}\right]^{-1} - \mathcal{I}_{fk}^{-}\\ &= \mathbf{\Lambda}_{k}^{-1} - \mathcal{I}_{fk}^{-},\\ \mathbf{\Lambda}_{bk}^{+} &= (\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}.\end{aligned} \tag{B.16}$$

where we have used eq. (6.17) in the above derivation. Substitute this expression for $\mathbf{\Lambda}_{bk}^{+}$ into eq. (B.6) to obtain

$$\begin{aligned}\mathbf{F}_{k-1}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})\mathbf{F}_{k-1}^{T} &= \mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{bk}^{+}\\ &= \mathbf{\Lambda}_{fk}^{-} + (\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}.\end{aligned} \tag{B.17}$$

Invert both sides to obtain

$$\begin{aligned}\mathbf{F}_{k-1}^{-T}(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1}\mathbf{F}_{k-1}^{-1} &= \left[\mathbf{\Lambda}_{fk}^{-} + (\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}\right]^{-1},\\ (\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1} &= \mathbf{F}_{k-1}^{T}\left[\mathbf{\Lambda}_{fk}^{-} + (\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}\right]^{-1}\mathbf{F}_{k-1}\\ &= \mathbf{F}_{k-1}^{T}\left[\mathbf{\Lambda}_{fk}^{-}\mathcal{I}_{fk}^{-}\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{fk}^{-}\mathcal{I}_{fk}^{-}(\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}\mathcal{I}_{fk}^{-}\mathbf{\Lambda}_{fk}^{-}\right]^{-1}\mathbf{F}_{k-1}\\ &= \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\left[\mathcal{I}_{fk}^{-} + \mathcal{I}_{fk}^{-}(\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}\mathcal{I}_{fk}^{-}\right]^{-1}\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}.\end{aligned} \tag{B.18}$$

Apply the matrix inversion lemma to the term $(\mathcal{I}_{k} - \mathcal{I}_{fk}^{-})^{-1}$ in the above equation

$$\begin{aligned}&(\mathbf{\Lambda}_{f,k-1}^{+} + \mathbf{\Lambda}_{b,k-1}^{-})^{-1}\\ &= \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\left[\mathcal{I}_{fk}^{-} + \mathcal{I}_{fk}^{-}\left(-\mathbf{\Lambda}_{fk}^{-} - \mathbf{\Lambda}_{fk}^{-}(-\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{k})^{-1}\mathbf{\Lambda}_{fk}^{-}\right)\mathcal{I}_{fk}^{-}\right]^{-1}\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}\\ &= \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\left[\mathcal{I}_{fk}^{-} + \left(-\mathbf{I} - (-\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{k})^{-1}\mathbf{\Lambda}_{fk}^{-}\right)\mathcal{I}_{fk}^{-}\right]^{-1}\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}\\ &= \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}\left[\mathcal{I}_{fk}^{-} - \mathcal{I}_{fk}^{-} - (\mathbf{\Lambda}_{fk}^{-} + \mathbf{\Lambda}_{k})^{-1}\right]^{-1}\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}\\ &= \mathbf{F}_{k-1}^{T}\mathcal{I}_{fk}^{-}(\mathbf{\Lambda}_{fk}^{-} - \mathbf{\Lambda}_{k})\mathcal{I}_{fk}^{-}\mathbf{F}_{k-1}.\end{aligned} \tag{B.19}$$

□

Appendix C

Test for White Noise

The residuals of the filter output are expected to be a sequence of serially uncorrelated random variables with zero mean and constant variance. One standard procedure to test for white noise is to apply the Kolmogorov-Smirnov test to the modified cumulative periodogram [69] [84]. The Bartlett's Kolmogorov-Smirnov test (BKS test) statistic [15] addresses the series of observations for a coded variable in the time domain and tests the "goodness of fit" of an observed error series to a normally distributed error series.

The time series x_n is decomposed by the Fourier transform

$$x_n = \frac{a_0}{2} + \sum_{k=1}^{m} \left[a_k \cos(\omega_k n) + b_k \sin(\omega_k n) \right]. \tag{C.1}$$

The amplitude periodogram I_k is defined as follows [210]:

$$I_k = \frac{n}{2} \left(a_k^2 + b_k^2 \right). \tag{C.2}$$

The normalized cumulative periodogram, L_j, for $j = 1, 2, \ldots, m-1$, is

$$L_j = \frac{\sum_{k=1}^{j} I_k}{\sum_{k=1}^{m} I_k}, \tag{C.3}$$

where $m = n/2$ if n is even or $m = (n-1)/2$ if n is odd.

Under the null hypothesis of Gaussian white noise, $I_0, \ldots, I_n$ have the same distribution function as a set of $n+1$ spacings determined by n independent uniform random variables $U_1, \ldots, U_n$. The spacings d_{kn}, for $0 \leq k \leq n$, are simply the differences between the successive order statistics $U_{k:n}$ and $U_{k+1:n}$, with $U_{0:n} = 0$, $U_{n+1:n} = 1$. Thus testing for white noise may be reduced to testing whether a random sample comes from a uniform distribution over the unit interval. A natural test statistic for the latter problem is the Kolmogorov-Smirnov statistic given by the largest deviation of the empirical distribution function from the true distribution function.

Appendix D

Support Vector Regression

Support vector machines (SVM) have become a subject of intensive study [26] [264] [287]. The SV algorithm is a nonlinear generalization of the **Generalized Portrait** technique developed in Russia in the 1960s [289] [291]. Support vector machines are grounded in the framework of statistical learning theory, or **Vapnik-Chervonenkis theory**, which has been developed over the last three decades by Vapnik and Chervonenkis [287] [288]. Similar optimization approach using linear instead of quadratic programming was used in pattern recognition by Mangasarian [186]. However, it was not until 1992 that the algorithm evolved to form the maximal margin classifier, the basic support vector machine, and not until 1995 that the soft margin version was introduced by Cortes and Vapnik [51]. After introudction of the basic SVM, an increasing number of researchers have worked on both the algorithmic and theoretical analysis of these systems, creating in just a few years a new research direction, merging concepts from disciplines as distant as statistics, functional analysis, optimization, as well as machine learning. These techniques have been recently applied successfully on a number of practical applications—in optical character recognition classification tasks [248] (for a comprehensive tutorial on SV pattern recognition, see [37]), in fields as diverse of bioinformatics [29], computational linguistics [208] and computer vision [128]. In addition, support vector machines have been applied to regression and time series prediction with outstanding performance [68] [202] [203].

We review the basic concepts underlying support vector machines in the context of function estimation. The goal is to explore the potential of SVM as a general regression method for detecting the presence of nonlinearity in time series. For pedagogical reasons, our treatment begin with the linear regression case in Sec. D.1. Then in Sec. D.2, we introduce nonlinear mapping functions that transform the data into a high-dimensional feature space, thus resulting in a nonlinear regression in the input space.

D.1 Linear regression

Suppose we are given training data

$$\mathcal{D} = \{(\mathbf{z}_1, y_1), \ldots, (\mathbf{z}_N, y_N)\} \subset \mathcal{Z} \times \mathbb{R}, \tag{D.1}$$

where $\mathcal{Z}$ denotes the space of the input patterns. In ε-SV regression [287], our goal is to find a function $f(\mathbf{z})$ that has at most ε deviation from the targets y_i for all the training data, and at the same time be as flat as possible. In other words, we do not care about errors as long as they are less than ε, but will not accept any deviation larger than this.

We begin by describing the case of linear function f, taking the form

$$f(x) = \mathbf{v}^T \mathbf{z}_i + \beta, \tag{D.2}$$

with $\mathbf{v} \in \mathcal{Z}$ and $\beta \in \mathbb{R}$. We minimize the norm of the complexity term, $\mathbf{v}$, to enforce flatness in f. We can write this problem as a convex optimization problem:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{v}\|^2, \\ &\text{subject to } \begin{cases} y_i - \mathbf{v}^T \mathbf{z}_i - \beta \leq \varepsilon, \\ \mathbf{v}^T \mathbf{z}_i + \beta - y_i \leq \varepsilon. \end{cases} \end{aligned} \tag{D.3}$$

The tacit assumption in eq. (D.3) was that the convex optimization problem is feasible, that is, a function f actually exists that approximates all pairs $(\mathbf{z}_i, y_i)$ with ε precision. However, this may not always be the case. Therefore, "soft margin" loss functions were adopted by Cortes and Vapnik [51], which introduce slack variables ξ_i^+, ξ_i^- to cope with the otherwise infeasible constraints of the optimization problem. Hence we arrive at the formulation stated in [287]:

$$\begin{aligned} &\text{minimize } \frac{1}{2}\|\mathbf{v}\|^2 + \lambda \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-), \\ &\text{subject to } \begin{cases} y_i - \mathbf{v}^T \mathbf{z}_i - \beta \leq \varepsilon + \xi_i^+, \\ \mathbf{v}^T \mathbf{z}_i + \beta - y_i \leq \varepsilon + \xi_i^-, \\ \xi_i^+, \xi_i^- \leq 0. \end{cases} \end{aligned} \tag{D.4}$$

The positive, prespecified constant λ determines the trade-off between the flatness of f and the amount up to which deviation larger than ε are tolerated. This corresponds to using the ε-insensitive loss function $|\xi|_\varepsilon$ described by

$$|\xi|_\varepsilon = \begin{cases} 0, & \text{if } |\xi| \leq \varepsilon; \\ |\xi| - \varepsilon, & \text{otherwise.} \end{cases} \tag{D.5}$$

Note that this cost function introduces a systematic bias, since we tend to underfit if ε is too large, for example, in the extreme case of very large ε the resulting regression will be a constant.

Fig. D-1 depicts the situation graphically. Only the points outside the shaded region contribute to the cost, as the deviation are penalized in a linear fashion. It turns out that in most cases the optimization problem can be solved more easily in its dual formulation. Moreover, the dual formulation provides the key for extending SVM to nonlinear functions. Hence we will use a standard dualization method utilizing Lagrange multipliers.

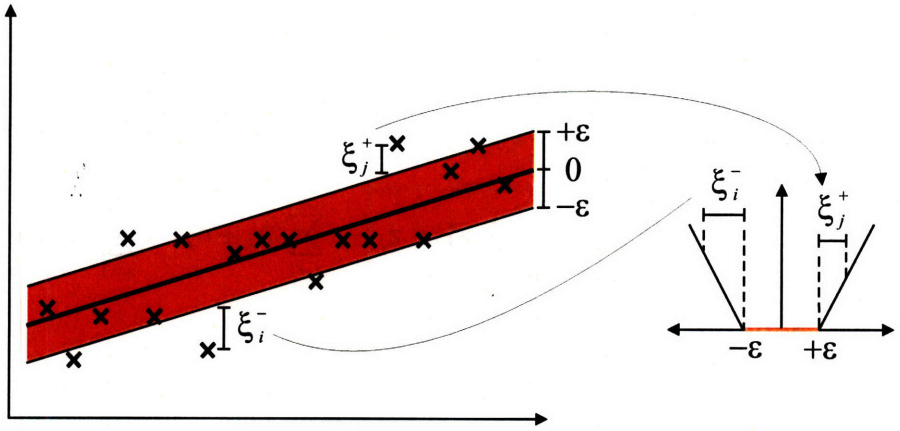


Figure D-1: The soft margin loss setting for a linear support vector machine. Only the points outside the shaded region contribute to the cost, as the deviation are penalized in a linear fashion.

D.1.1 Dual problem and quadratic programming

The key idea is to construct a Lagrange function form the primal objective function and the corresponding constraints, by introducing a dual set of variables. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution. We proceed as follows:

$$\mathcal{L} = \frac{1}{2}\|\mathbf{v}\|^2 + \lambda \sum_{i=1}^{N}(\xi_i^+ + \xi_i^-) - \sum_{i=1}^{N}(\eta_i \xi_i^+ + \eta_i^* \xi_i^-) \\ - \sum_{i=1}^{N} \alpha_i(\varepsilon + \xi_i^+ - y_i + \mathbf{v}^T\mathbf{z}_i + \beta) - \sum_{i=1}^{N} \alpha_i^*(\varepsilon + \xi_i^- + y_i - \mathbf{v}^T\mathbf{z}_i - \beta), \quad \text{(D.6)}$$

where $\mathcal{L}$ is the Lagrangian and η_i, η_i^*, α_i, α_i^* are Lagrange multipliers. Hence the dual variables in eq. (D.6) have to satisfy positivity constraints, that is

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0. \quad \text{(D.7)}$$

It follows from the saddle point condition that the partial derivatives of $\mathcal{L}$ with respect

to the primal variables ($\mathbf{v}$, β, ξ_i^+, ξ_i^-) have to vanish for optimality.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = \mathbf{v} - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{z}_i = 0, \tag{D.8}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \tag{D.9}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^+} = \lambda - \alpha_i - \eta_i = 0, \tag{D.10}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^-} = \lambda - \alpha_i^* - \eta_i^* = 0. \tag{D.11}$$

Substituting the conditions above into eq. (D.6) yields the dual optimization problem.

$$\begin{aligned} &\text{maximize } -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\mathbf{z}_i^T\mathbf{z}_j - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)y_i, \\ &\text{subject to } \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, \lambda]. \end{aligned} \tag{D.12}$$

In deriving eq. (D.12) we already eliminated the dual variables η_i, η_i^* through conditions (D.10) and (D.11). Eq. (D.8) can be rewritten as follows

$$\mathbf{v} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{z}_i, \tag{D.13}$$

$$f(\mathbf{z}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{z}_i^T\mathbf{z} + \beta. \tag{D.14}$$

This is the so-called **support vector expansion**; $\mathbf{v}$ can be completely described as a linear combination of the training patterns $\mathbf{z}_i$. The Lagrange multipliers α_i, α_i^* have an intuitive interpretation as forces pushing and pulling the estimate $f(\mathbf{z}_i)$ towards the measurements y_i. In a sense, the complexity of a function's representation by SVs is independent of the dimensionality of the input space $\mathcal{Z}$, and depends only on the number of SVs.

Moreover, note that the complete algorithm can be described in terms of inner products between the data. Even when evaluating $f(\mathbf{z})$ we need not compute $\mathbf{v}$ explicitly. These observations will come in handy for the formulation of a nonlinear extension.

D.1.2 Computing the threshold

The constant β can be calculated by exploiting the **Karush-Kuhn-Tucker (KKT) conditions**. These state that at the point of the solution the product between dual variables

and constraints has to vanish.

$$\alpha_i(\varepsilon + \xi_i^+ - y_i + \mathbf{v}^T\mathbf{z} + \beta) = 0, \tag{D.15}$$

$$\alpha_i^*(\varepsilon + \xi_i^- - y_i + \mathbf{v}^T\mathbf{z} - \beta) = 0, \tag{D.16}$$

$$(\lambda - \alpha_i)\xi_i^+ = 0, \tag{D.17}$$

$$(\lambda - \alpha_i^*)\xi_i^- = 0. \tag{D.18}$$

This allows us to make several useful conclusions. First, only samples $(\mathbf{z}_i, y_i)$ with corresponding α_i, $\alpha_i^* = \lambda$ lie outside the ε-insensitive tube. Second, $\alpha_i\alpha_i^* = 0$, that is, there can never be a set of dual variables α_i, α_i^* which are simultaneously nonzero. This allows us to conclude that

$$\begin{aligned} \varepsilon - y_i + \mathbf{v}^T\mathbf{z}_i + \beta \geq 0 \text{ and } \xi_i = 0 \text{ if } \alpha_i < \lambda, \\ \varepsilon - y_i + \mathbf{v}^T\mathbf{z}_i + \beta \leq 0 \text{ if } \alpha_i > 0. \end{aligned} \tag{D.19}$$

In conjunction with an analogous analysis on α_i^* we have

$$\max\{-\varepsilon + y_i - \mathbf{v}^T\mathbf{z}_i | \alpha_i < \lambda \text{ or } \alpha_i^* > 0\} \leq \beta \leq \min\{-\varepsilon + y_i - \mathbf{v}^T\mathbf{z}_i | \alpha_i > 0 \text{ or } \alpha_i^* < \lambda\}. \tag{D.20}$$

If some α_i, $\alpha_i^* \in (0, \lambda)$ the inequalities become equalities.

In other words, we pick those values α_i, α_i^* for which the prediction error $\delta_i = f(\mathbf{z}_i) - y_i$ can be determined uniquely. In the ε-insensitive case this means picking points $\mathbf{z}_i$ on the margin, by requiring that one of the corresponding α_i or α_i^* be in the open interval $(0, \lambda)$. In that case we know the exact value

$$\delta_i = \varepsilon \operatorname{sign}(\alpha_i - \alpha_i^*), \tag{D.21}$$

of the prediction error. For stability purposes it is recommended to take the average over all points on the margin with

$$\beta = \text{average}\left\{\delta_i + y_i - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{z}_i^T\mathbf{z}\right\}. \tag{D.22}$$

We make a note regarding the *sparsity* of the SV expansion. From eqs. (D.15) and (D.16) it follows that only for $|f(\mathbf{z}_i - y_i)| \geq \varepsilon$ the Lagrange multipliers may be nonzero, or in other words, for all samples inside the ε-tube (the shaded region in Fig. D-1) the α_i, α_i^* vanish; for $|f(\mathbf{z}_i - y_i)| < \varepsilon$ the second factor in eqs. (D.15) and (D.16) is nonzero, hence α_i, α_i^* has to be zero such that the KKT conditions are satisfied. Therefore we do no need all $\mathbf{z}_i$ in the expansion of $\mathbf{v}$, hence a sparse expansion. The samples that come with nonvanishing coefficients are called **support vectors**.

D.1.3 Huber's loss function

Other cost functions such as Huber's robust loss function can also be used. This cost function has the advantage of not introducing additional bias (like the ε-insensitive one does), at the expense of sacrificing sparsity in the coefficients α_i, α_i^*. The Huber's loss function takes the form

$$L_{\text{Huber}}(f(\mathbf{z}) - y) = \begin{cases} \frac{1}{2}(f(\mathbf{z}) - y)^2, & \text{for } |f(\mathbf{z}) - y| < \mu; \\ \mu|f(\mathbf{z}) - y| - \frac{\mu^2}{2}, & \text{otherwise.} \end{cases} \tag{D.23}$$

The corresponding quadratic programming problem takes the following form

$$\begin{aligned} &\text{minimize } \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\mathbf{z}_i^T\mathbf{z}_j + \frac{1}{2\lambda}(\alpha_i^2 + \alpha_i^{*2}) + \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)y_i, \\ &\text{subject to } \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in \left[0, \frac{\mu}{\lambda}\right]. \end{aligned} \tag{D.24}$$

All patterns become support vectors.

D.2 Nonlinear regression by implicit kernel mapping

This section discusses the method that can be used to construct a mapping of the training patterns in the input space $\mathcal{Z}$ into a high dimensional feature space $\mathcal{F}$ via a nonlinear mapping Ψ, and do linear regression in this space

$$f(\mathbf{z}) = \mathbf{v}^T\Psi(\mathbf{z}) + \beta, \tag{D.25}$$

with $\Psi : \mathcal{Z} \to \mathcal{F}$, $\mathbf{v} \in \mathcal{F}$ and $\beta \in \mathbb{R}$. Thus, *linear* regression in a *high* dimensional feature space corresponds to *nonlinear* regression in the *low* dimensional input space. The concept of the nonlinear preprocessing of the data is depicted in Fig. D-2.

Note that the inner product $\mathbf{v}^T\Psi(\mathbf{z})$ would have to be computed in this high dimension space, which is usually intractable. Therefore, we introduce the use of kernel functions that leaves us with inner products that can be implicitly expressed in the low dimensional input space $\mathcal{Z}$.

D.2.1 Implicit mapping via kernels

As noted in the previous section, the SV algorithm only depends on inner products between patterns $\mathbf{z}_i$. Hence it suffices to know $K(\mathbf{z}, \mathbf{z}') = \Psi(\mathbf{z})^T\Psi(\mathbf{z}')$ rather than Ψ explicitly which

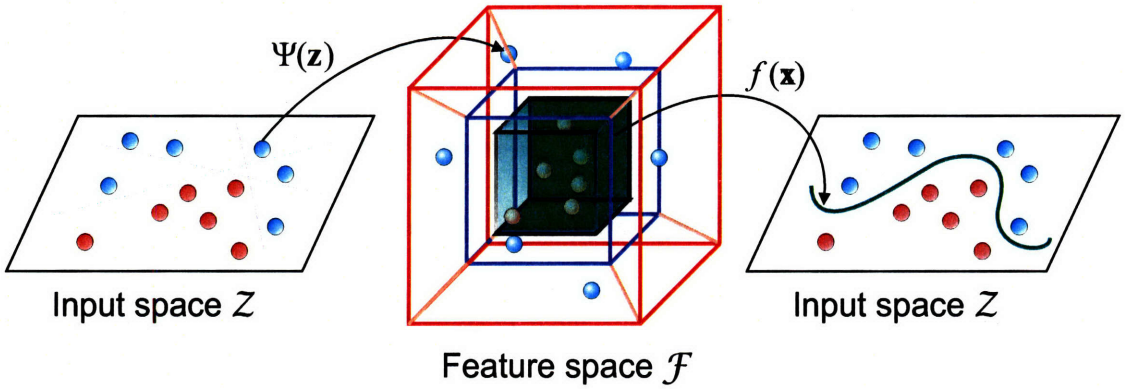


Figure D-2: Nonlinear regression is achieved by simply preprocessing the training patterns by a map $\Psi : \mathcal{Z} \rightarrow \mathcal{F}$ into some high dimensional feature space $\mathcal{F}$, and then applying the standard SV regression algorithm. The solution of the linear regression in the feature space corresponds to nonlinear regression in the input space $f(\mathbf{z})$. By the use of a nonlinear kernel function, it is possible to compute a hyperplane with maximum margin in a feature space without explicitly mapping into that space.

allows us to restate the SV optimization problem:

$$\text{maximize} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{z}_i, \mathbf{z}_j) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)y_i,$$

$$\text{subject to} \quad \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, \lambda]. \tag{D.26}$$

Likewise the expansion of $\mathbf{v}$ and f in eqs. (D.13) and (D.14) may be written as

$$\mathbf{v} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\Psi(\mathbf{z}_i), \tag{D.27}$$

$$f(\mathbf{z}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(\mathbf{z}_i, \mathbf{z}) + \beta; \tag{D.28}$$

the difference to the linear case is that $\mathbf{v}$ is no longer given explicitly. Also note that in the nonlinear case, the optimization problem corresponds to finding the flattest function in the *feature* space, not in the input space.

D.2.2 Conditions for kernels

The functions $K(\mathbf{z}, \mathbf{z}')$ corresponding to an inner product in some feature space $\mathcal{F}$ must satisfy the following conditions:

1. If K is a symmetric positive definite function, which satisfies Mercer's condition [198].

$$K(\mathbf{z}, \mathbf{z}') = \sum_{m=0}^{\infty} a_m \phi_m(\mathbf{z}) \phi_m(\mathbf{z}'), \quad a_m \geq 0;$$
$$\iint_{\mathcal{Z} \times \mathcal{Z}} \mathrm{d}\mathbf{z}\, \mathrm{d}\mathbf{z}'\, K(\mathbf{z}, \mathbf{z}') f(\mathbf{z}) f(\mathbf{z}') \geq 0 \text{ for all } f \in L_2(\mathcal{Z}), \tag{D.29}$$

   then $K(\mathbf{z}, \mathbf{z}')$ represents a legitimate inner product in some feature space, $\mathcal{F}$.

2. Let K_1 and K_2 be admissible SV kernels and $c_1, c_2 \geq 0$, then

$$K(\mathbf{z}, \mathbf{z}') = c_1 K_1(\mathbf{z}, \mathbf{z}') + c_2 K_2(\mathbf{z}, \mathbf{z}') \tag{D.30}$$

   is also an admissible kernel, by the virtue of the linearity of integrals.

3. Let $\psi(\mathbf{z}, \mathbf{z}')$ be a symmetric function on $\mathcal{Z} \times \mathcal{Z}$ such that

$$K(\mathbf{z}, \mathbf{z}') = \int_{\mathcal{Z}} \mathrm{d}\zeta\, \psi(\mathbf{z}, \zeta) \psi(\mathbf{z}', \zeta) \tag{D.31}$$

   exists. Then K is an admissible SV kernel.

4. Let K_1 and K_2 be admissible SV kernels, then

$$K(\mathbf{z}, \mathbf{z}') = K_1(\mathbf{z}, \mathbf{z}') K_2(\mathbf{z}, \mathbf{z}') \tag{D.32}$$

   is an admissible kernel.

5. A translation invariant kernel $K(\mathbf{z}, \mathbf{z}') = K(\mathbf{z} - \mathbf{z}')$ is an admissible SV kernel if and only if the Fourier transform

$$F(\omega) = (2\pi)^{-\frac{d}{2}} \int_{\mathcal{Z}} \mathrm{d}\mathbf{z}\, K(\mathbf{z}) \mathrm{e}^{-\mathrm{i}\mathbf{v}^T \mathbf{z}} \tag{D.33}$$

   is nonnegative [265].

6. Any kernel of dot-product type $K(\mathbf{z}, \mathbf{z}') = K(\mathbf{z}^T \mathbf{z}')$ has to satisfy

$$\begin{aligned} K(\zeta) &\geq 0, \\ \frac{\partial}{\partial \zeta} K(\zeta) &\geq 0, \\ \frac{\partial}{\partial \zeta} K(\zeta) + \zeta \frac{\partial^2}{\partial \zeta^2} K(\zeta) &\geq 0 \end{aligned} \tag{D.34}$$

   for any $\zeta \geq 0$, is a necessary (but not sufficient) condition for K to be an admissible SV kernel [38].

7. A kernel of dot-product type $K(\mathbf{z}, \mathbf{z}') = K\left(\mathbf{z}^T\mathbf{z}'\right)$ defined on an infinite dimensional Hilbert space, with a power series expansion

$$K(t) = \sum_{n=0}^{\infty} a_n t^n \tag{D.35}$$

is admissible if and only if all $a_n \geq 0$ [247].

D.2.3 Examples of kernel functions

Below are examples of some common kernel functions:

Polynomial A homogeneous polynomial mapping, with $p \in \mathbb{N}$, is a popular method for nonlinear modeling

$$K(\mathbf{z}, \mathbf{z}') = (\mathbf{z}^T\mathbf{z}')^d. \tag{D.36}$$

From this observation one can conclude that inhomogeneous polynomial kernels of the type

$$K(\mathbf{z}, \mathbf{z}') = (\mathbf{z}^T\mathbf{z}' + c)^d \tag{D.37}$$

are also admissible. The second kernel is usually preferred as it avoids problems with the Hessian becoming zero.

Gaussian radial basis function Radial basis functions have received significant attention. The most common one—the translationally invariant Gaussian,

$$K(\mathbf{z}, \mathbf{z}') = \exp\left(-\frac{\|\mathbf{z} - \mathbf{z}'\|^2}{2\sigma^2}\right) \tag{D.38}$$

has shown to be an admissible SV kernel [4, 26, 290].

Exponential radial basis function A radial basis function of the form

$$K(\mathbf{z}, \mathbf{z}') = \exp\left(-\frac{\|\mathbf{z} - \mathbf{z}'\|}{2\sigma^2}\right) \tag{D.39}$$

produces a piecewise linear solution which can be attractive when discontinuities are acceptable.

Multi-layer perception The long established MLP, with a single hidden layer, also has a valid kernel representation,

$$K(\mathbf{z}, \mathbf{z}') = \tanh(\varrho + \rho\, \mathbf{z}^T\mathbf{z}'), \tag{D.40}$$

for certain values of the scale ρ, and offset, ϱ, parameters. Here the SV correspond to the first layer and the Lagrange multipliers to the weights. This hyperbolic tangent

kernel is appealing due to its resemblance to neural networks. This kernel does not actually satisfy Mercer's condition, yet it has been successfully applied in practice.

B-splines B-splines are a popular spline formulation. The kernel is defined on the interval $[-1, 1]$, and has an attractive closed form,

$$K(\mathbf{z}, \mathbf{z}') = B_{2n+1}(\|\mathbf{z} - \mathbf{z}'\|) \text{ with } B_k = \bigotimes_{i=1}^{k} \mathbf{1}_{[-\frac{1}{2}, \frac{1}{2}]}, \tag{D.41}$$

where $\bigotimes$ denote the convolution operation, and the $\mathbf{1}_X$ denotes the indicator function on the set X.

Tensor product Kernels multidimensional kernels can be obtained by forming tensor products of kernels [12]

$$K(\mathbf{z}, \mathbf{z}') = \prod_i K_i(\mathbf{z}_i, \mathbf{z}_i'). \tag{D.42}$$

This is particularly useful in the construction of multidimensional spline kernels, which are simply obtained from the product of the univariate kernels.

While there exists some experimental results and discussions about which kernel to choose for a particular problem, the fact remains that kernel selection remains more of an art than a precise procedure, often requiring trial and error to find one that works.

D.3 Summary

In this appendix we have reviewed some basic concepts of support vector machines. They are a class of algorithms characterized by the use of kernels as inner products in a feature space, the absence of local minima, the sparsity of the solution and the capacity control. SVM also generalize well in high dimensions. Extensions of the SVM concept have been made by several authors, for example Mangasarian's generalized support vector machines [187]. Particularly interesting is the development of the Bayes Point Machine [131], that enforces another inductive principle, given by Bayesian generalization theory. The SVM algorithm showed excellent performances on the database of chaotic time series, outperforming other techniques including polynomial and rational approximation, local polynomial techniques, and neural networks [202]. Support vector machines are indeed an attractive approach to data modeling.

Bibliography

[1] P.-A. Absil, R. Sepulchre, A. Bilge, and P. Gérard. Nonlinear analysis of cardiac rhythm fluctuations using DFA method. *Physica A: Statistical Mechanics and its Applications*, 272(1-2):235–244, Oct. 1999.

[2] U. R. Acharya, N. Kannathal, O. W. Sing, L. Y. Ping, and T. Chua. Heart rate analysis in normal subjects of various age groups. *BioMedical Engineering OnLine*, 3(1):24, Jul. 2004.

[3] D. N. Adams. *The Hitchhiker's Guide to the Galaxy.* Pan Books, UK, 1979.

[4] M. A. Aizerman, É. M. Braverman, and L. I. Rozonoér. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

[5] H Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(16):716–723, 1974.

[6] E. S Allen. The scientific work of Vito Volterra. *The American Mathematical Monthly*, 48(8):516–519, Oct. 1941.

[7] C. Andrieu, A. Doucet, and E. Punskaya. Sequential Monte Carlo methods for optimal filtering. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 4, pages 79–95. Springer-Verlag, New York, 2001.

[8] C. Andrieu, A. Doucet, S. S. Sigh, and V. B. Tadic. Particle methods for change detection, system identification, and control. *Proceedings of the IEEE*, 92(3):423–438, Mar. 2004.

[9] D. V. Anosov and Ya. G. Sinai. Some smooth ergodic systems. *Russian Mathematical Surveys*, 22(5):103–167, 1967.

[10] D. M. Apalkov and P. B. Visscher. Spin-torque switching: Fokker-Planck rate calculation. *Physical Review B: Condensed Matter and Material Physics*, 72(180405):1–4, Nov. 2005.

[11] A. Arber. *The Mind and the Eye.* Cambridge University Press, London, 1954.

[12] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[13] S. A. Arrhenius. Uber die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch siiuren. *Zeitschrift f ur Physikalische Chemie*, 4:226–248, 1889.

[14] M. Barahona and C.-S. Poon. Detection of nonlinear dynamics in short, noisy time series. *Nature*, 381(6579):215–217, May 1996.

[15] M. S. Bartlett. *An Introduction to Stochastic Processes: With Special Reference to Methods and Applications.* Cambridge University Press, London, 2nd edition, 1966.

[16] C. Beck and F. Sch ogl. *Thermodynamics of Chaotic Systems: An Introduction*, volume 4 of *Cambridge Nonlinear Science Series.* Cambridge University Press, Cambridge, UK, 1993.

[17] R. P. Bell. *The Tunnel Effect in Chemistry.* Chapman and Hall, London, 1980.

[18] T. Bengtsson and J. E. Cavanaugh. An improved Akaike information criterion for state-space model selection. *Computational Statistics & Data Analysis*, 50(10):2635–2654, 2006.

[19] J. Bhattacharya and P. P. Kanjilal. On the detection of determinism in a time series. *Physica D: Nonlinear Phenomena*, 132(1-2):100–110, Jul. 1999.

[20] L. Billings, E. M. Bollt, and I. B. Schwartz. Phase-space transport of stochastic chaos in population dynamics of virus spread. *Physical Review Letters*, 88(23):234101, Jun. 2002.

[21] L. Billings and I. B. Schwartz. Exciting chaos with noise: unexpected dynamics in epidemic outbreaks. *Journal of Mathematical Biology*, 44(1):31–48, 2002.

[22] A. Blake, M. Isard, and J. MacCormick. Statistical models of visual shape and motion. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 16, pages 339–357. Springer-Verlag, New York, 2001.

[23] T. Bohr and D. Rand. The entropy function for characteristic exponents. *Physica D: Nonlinear Phenomena*, 25(1-3):387–398, Mar.–Apr. 1987.

[24] L. Boltzmann. *Vorlesungen üeber Gastheorie.* Berkeley, CA. University of California Press, Leipzig: Barth, 1896. Translated by Stephen G. Brush in 1964. Reprinted (1995). New York: Dover Publications.

[25] E. Bølviken and G. Storvik. Deterministic and stochastic particle filters in state-space models. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 5, pages 97–116. Springer-Verlag, New York, 2001.

[26] B. E. Boser, I. M. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, Pittsburgh, PA, Jul. 1992. ACM Press.

[27] R. Bowen and D. Ruelle. The ergodic theory of Axion A flows. *Inventiones Mathematicae*, 29(3):181–202, Oct. 1975.

[28] S. Boyd, L. O. Chua, and C. A. Desoer. Analytical foundations of Volterra series. *IMA Journal of Mathematical Control and Information*, 1:243–282, 1984.

[29] M. P. S. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, and D. Haussler. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*, 97(1):262–267, 2000.

[30] R. Brown, P. Bryant, and H. D. I. Abarbanel. Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Physical Review A: Atomic, Molecular, and Optical Physics*, 43(6):2787–2806, Mar. 1991.

[31] W. F. Brown. Thermal fluctuations of a single-domain particle. *Physical Review*, 130(5):1677–1686, Jun. 1963.

[32] R. Bucy and P. Joseph. *Filtering for Stochastic Processes with Applications to Guidance.* John Wiley and Sons, New York, 1968.

[33] A. R. Bulsara. Noise and nonlinearity in neuron modeling. In M. Millonas, editor, *Fluctuations and Order: The New Synthesis*, Institute for Nonlinear Science, chapter 23, pages 359–396. Springer-Verlag, New York, 1996.

[34] A. R. Bulsara, E. W. Jacobs, and W. C. Schieve. Noise effects in a nonlinear dynamic system: The rf superconducting quantum interference device. *Physical Review A: Atomic, Molecular, and Optical Physics*, 42(8):4614–4621, Oct. 1990.

[35] A. R. Bulsara, W. C. Schieve, and E. W. Jacobs. Homoclinic chaos in systems perturbed by weak Langevin noise. *Physical Review A: Atomic, Molecular, and Optical Physics*, 41(2):668–681, Jan. 1990.

[36] F. V. Bunkin, N. A. Kirichenko, and B. S. Luk'yanchuk. Thermochemical bistability and chemical phase transitions stimulated by laser radiation. *Soviet Journal of Quantum Electronics*, 14(6):798–807, Jun. 1984.

[37] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[38] C. J. C. Burges. Geometry and invariance in kernel based methods. In B. Schölkopf, C. J. C. Burges, and J. A. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 89–116, Cambridge, MA, 1999. MIT Press.

[39] C. C. Canavier, J. W. Clark, and J. H. Byrne. Routes to chaos in a model of bursting neuron. *Biophysical Journal*, 57(6):1245–1251, Jun. 1990.

[40] B. Caroli, C. Caroli, B. Roulet, and D. Saint-James. On fluctuations and relaxation in systems described by a one-dimensional Fokker-Planck equation with a time-dependent potential. *Physica A: Statistical and Theoretical Physics*, 108(1):233–256, Aug. 1981.

[41] J. Carpenter, P. Clifford, and P. Fearnheard. Improved particle filter for nonlinear problems. *IEE Proceedings Part F: Radar and Sonar Navigation*, 146(1):2–7, Feb. 1999.

[42] G. Carpintero and H. Lamela. Influence of noise in the route to chaos of directly modulated semiconductor lasers. *Journal of Applied Physics*, 82(6):2766–2772, Sep. 1997.

[43] J. Casado-Pascual, J. Gómez-Ordóñez, and M. Morillo. Stochastic resonance: Theory and numerics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15:026115, Jun. 2005.

[44] B. Cazelles. How predictable is chaos? *Nature*, 355(6355):25–26, Jan. 1992.

[45] H. Y. Chan and M. A. Kouritzin. Particle filters for combined state and parameter estimation. In I. Kadar, editor, *Signal Processing, Sensor Fusion, and Target Recognition X*, pages 244–252, Orlando, FL, Aug. 2001. The International Society for Optical Engineering.

[46] K. S. Chan and H. Tong. *Chaos: A Statistical Perspective.* Springer Series in Statistics. Springer-Verlag, New York, 2001.

[47] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15(1):1–89, Jan. 1943.

[48] Z.-Y. Chen. Noise-induced instability. *Physical Review A: Atomic, Molecular, and Optical Physics*, 42(10):5837–5843, Nov. 1990.

[49] L. Y. Chew, C. Ting, and C. H. Lai. Chaos-induced escape over a potential barrier. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 70(4):045203, Oct. 2004.

[50] L. Y. Chew, C. Ting, and C. H. Lai. Chaotic resonance: two-state model with chaos-induced escape over potential barrier. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 72(3):036222, Sep. 2005.

[51] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20(3):273–297, Sep. 1995.

[52] J. Crassidis and J. Junkins. *Optimal Estimation of Dynamic Systems.* CRC Press, New York, 2004.

[53] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, Mar. 2002.

[54] J. Crutchfield, M. Nauenberg, and J. Rudnick. Scaling for external noise at the onset of chaos. *Physical Review Letters*, 46(14):933–935, Apr. 1981.

[55] J. P. Crutchfield, J. D. Farmer, and B. A. Huberman. Fluctuations and simple chaotic dynamics. *Physics Reports*, 92(2):45–82, 1982.

[56] J. P. Crutchfield and B. A. Huberman. Fluctuations and the onset of chaos. *Physics Letters A*, 77(6):407–410, Jun. 1980.

[57] J. M. Cushing, R. F. Costantino, B. Dennis, R. A. Desharnais, and S. M. Hensona. Nonlinear population dynamics: models, experiments and data. *Journal of Theoretical Biology*, 194(1):1–9, Sep. 1998.

[58] I. Dayan, M. Gitterman, and G. H. Weiss. Stochastic resonance in transient dynamics. *Physical Review A: Atomic, Molecular, and Optical Physics*, 46(2):757–761, Jul. 1992.

[59] R. J. Deissler and J. D. Farmer. Deterministic noise amplifiers. *Physica D: Nonlinear Phenomena*, 55:155–165, 1992.

[60] Z.-D. Deng, N. M. Arzeno, E. Katz, and C.-S. Poon. Heart rate variability is more chaotic in rem than nrem sleep in children. In *The 38th Annual Neuroscience Conference*, Atlanta, GA, Oct. 2006. The Society for Neuroscience.

[61] Z.-D. Deng, C.-S. Poon, N. M. Arzeno, and E. S. Katz. Heart rate variability in pediatric obstructive sleep apnea. In *The 28th Annual International Conference of the IEEE*, pages 3565–3568, New York, NY, Aug. 2006. IEEE Engineering in Medicine and Biology Society.

[62] S. I. Denisov, K. Sakmann, P. Talkner, and P. H anggi. Mean first-passage times for an ac-driven magnetic moment of a nanoparticle. *Europhysics Letters*, 76(6):1001–1007, Dec. 2006.

[63] B. Dennis, R. A. Desharnais, J. M. Cushing, S. M. Henson, and R. F. Costantino. Can noise induce chaos? *OIKOS*, 102(2):329–339, Aug. 2003.

[64] T. J. Dodd and R. F. Harrison. A new solution to Volterra series estimation. In *Proceedings of the 15th IFAC World Congress on Automatic Control*, Barcelona, Spain, Jul. 2002.

[65] J. R. Dorfman. Deterministic chaos and the foundations of the kinetic theory of gases. *Physics Reports*, 301(1):151–185, Jul. 1998.

[66] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice.* Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.

[67] W. Drost-Hansen. Phase transitions in biological systems: manifestations of cooperative processes in vicinal water. *Annals of the New York Academy of Sciences*, 204:100–112, Mar. 1973.

[68] H. Druker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. Support vector regression machines. In M. C. Mozer, M. I. Jordan, and Petsche T., editors, *Advances in Neural Information Processing Systems 9*, pages 375–381, Cambridge, MA, 1997. MIT Press.

[69] J. Durbin. Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika*, 56(1):1–15, Mar. 1969.

[70] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57(3):617–656, Jul. 1985.

[71] D. A. Egolf. Equilibrium regained: from nonequilibrium chaos to statistical mechanics. *Science*, 287(5450):101–104, Jan. 2000.

[72] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flssigkeiten suspendierten teilchen. *Annalen der Physik*, 17:549–560, 1905.

[73] G. Favier, A. Y. Kibangou, and A. Khouaja. Nonlinear system modeling by means of Volterra models. Approaches for parametric complexity reduction. In *Invited paper in Symposium Techniques Avancées et Stratégies Innovantes en Modélisation et Commande Robuste des Processus Industriels*, Martigues, France, Sep. 2004. The Instrumentation, Systems, and Automation Society.

[74] M. J. Feigenbaum. Quantitative universality for a class of nonlinear transformations. *Journal of Statistical Physics*, 19(1):25–52, Jul. 1978.

[75] D. S. Feit. Characteristic exponents and strange attractors. *Communications in Mathematical Physics*, 61(3):249–260, Aug. 1978.

[76] H. C. Fogedby and M. H. Jensen. Weak noise approach to the logistic map. *Journal of Statistical Physics*, 121(516):759–778, Dec. 2005.

[77] M. Fortunato, G. Kurizki, and W. P. Schleich. Stabilization of deterministically chaotic systems by interference and quantum measurements: the Ikeda map case. *Physical Review Letter*, 80(26):5730–5733, Jun. 1998.

[78] M. O. Franz and B. Schölkopf. A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Computation*, 18(12):3097–3118, Dec. 2006.

[79] D. Fraser and J. Potter. The optimum linear smoother as a combination of two optimum linear filters. *IEEE Transactions on Automatic Control*, 14(4):387–390, Aug. 1969.

[80] N. de Freitas. Rao-Blackwellised particle filtering for fault diagnosis. In *Proceedings of the IEEE Aerospace Conference*, pages 1767–1772, 2002.

[81] N. de Freitas, C. Andrieu, P. Højen-Sørensen, M. Niranjan, and A. Gee. Sequential Monte Carlo methods for neural networks. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 17, pages 359–379. Springer-Verlag, New York, 2001.

[82] M. Frey and E. Simiu. Noise-induced chaos and phase space flux. *Physica D: Nonlinear Phenomena*, 63(3-4):321–340, Mar. 1993.

[83] M. Frey and E. Simiu. Noise-induced transitions to chaos. In P. E. Cladis and P. Palffy-Muhoray, editors, *Proceedings of the NATO Advanced Research Workshop: Spatio-Temporal Patterns in Nonequilibrium Complex Systems*, pages 529–544, Santa Fe, New Mexico, Apr. 1995. Santa Fe Institute, Addison-Wesley Publishing.

[84] W. A. Fuller. *Introduction to Statistical Time Series.* John Wiley and Sons, New York, 2nd edition, 1996.

[85] Debenedetti P. G. and F. H. Stillinger. Supercooled liquids and the glass transition. *Nature*, 410:259–267, Mar. 2001.

[86] Z. Galias. Rigorous investigation of the Ikeda map by means of interval arithmetic. *Nonlinearity*, 15(6):1759–1779, Nov. 2002.

[87] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Reviews of Modern Physics*, 70(1):223–287, Jan. 1998.

[88] G. Gamow. *One, Two, Three, ... Infinity: Facts and Speculations of Science.* Dover Publications, New York, 1988.

[89] C. Gan. Noise-induced chaos and basin erosion in softening Duffing oscillator. *Chaos, Solitons, and Fractals*, 25(5):1069–1081, Sep. 2005.

[90] C. Gan. Noise-induced chaos in Duffing oscillator with double wells. *Nonlinear Dynamics*, 45(3-4):305–317, Aug. 2005.

[91] C. Gan. Noise-induced chaos in a quadratically nonlinear oscillator. *Chaos, Solitons, and Fractals*, 30(4):920–929, Nov. 2006.

[92] J. B. Gao, C. C. Chen, S. K. Hwang, and J. M. Liu. Noise-induced chaos. *International Journal of Modern Physics B: Condensed Matter Physics; Statistical Physics; Applied Physics*, 13(28):3283–3305, Nov. 1999.

[93] J. B. Gao, S. K. Hwang, and J. M. Liu. Effects of intrinsic spontaneous-emission noise on the nonlinear dynamics of an optically injected semiconductor laser. *Physical Review A: Atomic, Molecular, and Optical Physics*, 59(2):1582–1585, Feb. 1999.

[94] J. B. Gao, S. K. Hwang, and J. M. Liu. When can noise induce chaos? *Physical Review Letters*, 82(6):1132–1135, Feb. 1999.

[95] J. García-Ojalvo and J. M. Sancho. *Noise in Spatially Extended Systems.* Springer-Verlag, New York, 1999.

[96] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, volume 13 of *Springer Series in Synergetics.* Springer-Verlag, Berlin, 1983.

[97] P. Gaspard and F. Baras. Chaotic scattering and diffusion in the Lorentz gas. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 51(6):5332–5352, 1995 1995.

[98] P. Gaspard and J. R. Dorfman. Chaotic scattering theory, thermodynamic formalism, and transport coefficients. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 52(4):3525–3552, Oct. 1995.

[99] P. Gaspard and G. Nicolis. Transport properties, Lyapunov exponents, and entropy per unit time. *Physical Review Letters*, 65(14):1693—1696, Oct. 1990.

[100] F. Gassmann. Noise-induced chaos-order transitions. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 55(3):2215–2221, Mar. 1997.

[101] N. Gedik, D.-S. Yang, G. Logvenov, I. Bozovic, and A. H. Zewail. Nonequilibrium phase transitions in cuprates observed by ultrafast electron crystallography. *Science*, 316(5823):425–429, Apr. 2007.

[102] A. Gelb, editor. *Applied Optimal Estimation.* MIT Press, Cambridge, MA, 1974.

[103] J. W. Gibbs. *Elementary Principles in Statistical Mechanics*, volume 2 of *The Collected Works of J. W. Gibbs*. Yale University Press, New Haven, 1902.

[104] V. Giorno, P. Lánsky, A. P. Noblie, and L. M. Ricciardi. Diffusion approximation and first-passage-time problem for a model neuron III. A birth-and-death process approach. *Biological Cybernetics*, 58(6):387–404, 1988.

[105] L. Glass. Synchronization and rhythmic processes in physiology. *Nature*, 410:277–284, Mar. 2001.

[106] J. A. Glazier and A. Libchaber. Quasi-periodicity and dynamical systems: An experimentalist's view. *IEEE Transactions on Circuits and Systems*, 35(7):790–809, Jul. 1988.

[107] R. W. Glenny, H. T. Robertson, S. Yamashiro, and J. B. Bassingthwaighte. Applications of fractal analysis to physiology. *Journal of Applied Physiology*, 70(6):2351–2367, 1991.

[108] A. L. Goldberger. Is the normal heartbeat chaotic or homeostatic? *News in Physiological Sciences*, 6(2):87–91, Apr. 1991.

[109] A. L. Goldberger, L. A. N. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley. Fractal dynamics in physiology: alterations with disease and aging. *Proceedings of the National Academy of Sciences*, 99(Suppl 1):2466–2472, Feb. 2002.

[110] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F: Radar and Signal Processing*, 140(2):107–113, 1993.

[111] G. A. Gottwald and I. Melbourne. Testing for chaos in deterministic systems with noise. *Physica D: NOnlinear Phenomena*, 212(1-2):100–110, Dec. 2005.

[112] I. Goychuk and P. H anggi. Ion channel gating: a first passage time analysis of the Kramers type. *Proceedings of the National Academy of Sciences*, 99(6):3552–3556, Mar. 2002.

[113] P. Grassberger and I. Procaccia. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A: Atomic, Molecular, and Optical Physics*, 28(4):2591–2593, Oct. 1983.

[114] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9(1-2):189–208, Oct. 1983.

[115] C. Grebogi, E. Ott, and J. A. Yorke. Unstable periodic orbits and the dimensions of multifractal chaotic attractors. *Physical Review A: Atomic, Molecular, and Optical Physics*, 37(5):1711–1724, Mar. 1988.

[116] F. Gustafsson and P. Hriljac. Particle filters for system identification with application to chaos prediction. In *Proceedings of the 13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, Sep. 2003.

[117] L. Györgyi and R. J. Field. Simple models of deterministic chaos in the belousov-zhabotinsky reaction. *Journal of Physical Chemistry*, 95:6594–6602, 1991.

[118] L. Györgyi and R. J. Field. A three-variable model of deterministic chaos in the belousov-zhabotinsky reaction. *Nature*, 355:808–810, 1992.

[119] H. Haken. Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Reviews of Modern Physics*, 47(1):67–121, Jan. 1975.

[120] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman. Fractal measures and their singularities: The characterization of strange sets. *Physical Review A: Atomic, Molecular, and Optical Physics*, 33(2):1141–1151, Feb. 1986.

[121] J. E. Handschin. Monte Carlo techniques for prediction and filtering of non-linear stochastic processes. *Automatica*, 6:555–563, 1970.

[122] J. E. Handschin and Mayne D. Q. Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering. *International Journal of Control*, 9:547–559, 1969.

[123] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Reviews of Modern Physics*, 62(2):251–341, Apr. 1990.

[124] P. D. Hanlon and P. S. Maybeck. Characterization of Kalman filter residuals in the presence of mismodeling. *IEEE Transactions on Aerospace and Electronic Systems*, 36(1):114–131, Jan. 2000.

[125] M. A. Harrison and Y.-C. Lai. Route to high-dimensional chaos. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 59(4):R3799–R3802, Apr. 1999.

[126] M. A. Harrison and Y.-C. Lai. Bifurcation to high-dimensional chaos. *International Journal of Bifurcation and Chaos*, 10(6):1471–1483, 2000.

[127] J. M. Hausdorff and C.-K. Peng. Multiscaled randomness: A possible source of 1/f noise in biology. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 54(2):2154–2157, Aug. 1996.

[128] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Proceedings of the 8th IEEE International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, 2001.

[129] M. Hénon. A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics*, 50(1):69–77, Feb. 1976.

[130] H. G. E. Hentschel and I. Procaccia. The infinite number of generalized dimensions of fractals and strange attractors. *Physica D: Nonlinear Phenomena*, 8(3):435–444, Sep. 1983.

[131] R. Herbrich, T. Graepel, and C. Campbell. Bayes point machines. *Journal of Machine Learning Research*, 1:245–279, 2001.

[132] H. Herzel, W. Ebeling, and Th. Schulmeister. Nonuniform chaotic dynamics and effects of noise in biochemical systems. *Zeitschr. Naturforsch.*, 42A:136–142, 1987.

[133] T. Higuchi. Self-organizing time series model. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 20, pages 429–444. Springer-Verlag, New York, 2001.

[134] J. E. Hirsch, B. A. Huberman, and D. J. Scalapino. Theory of intermittency. *Physical Review A: Atomic, Molecular, and Optical Physics*, 25(1):519–532, Jan. 1982.

[135] K. K. L. Ho, G. B. Moody, C.-K. Peng, J. E. Mietus, M. G. Larson, D. Levy, and A. L. Goldberger. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation*, 96(3):842–848, Aug. 1997.

[136] Y. Ho and R. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Transactions on Automatic Control*, AC-9:333–339, 1964.

[137] R. D. Holt and M. A. McPeek. Chaotic population dynamics favors the evolution of dispersal. *The American Naturalist*, 148(4):709–718, Oct. 1996.

[138] W. Horsthemke and R. Lefever. *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, volume 15 of *Springer Series in Synergetics*. Spring-Verlag, Berlin, 1984.

[139] C. M. Hurvich, R. Shumway, and C.-L. Tsai. Improved estimators of Kullback-Leibler information for autoregressive model selection in small samples. *Biometrika*, 77(4):709–719, Dec. 1990.

[140] R. A. Ibrahim. Excitation-induced stability and phase transition: a review. *Journal of Vibration and Control*, 12(10):1093–1170, 2006.

[141] K. Ikeda. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Optics Communications*, 30(2):257–261, Aug. 1979.

[142] K. Ikeda, H. Daido, and O. Akimoto. Optical turbulence: chaotic behavior of transmitted light from a right cavity. *Physical Review Letters*, 45(9):709–712, Sep. 1980.

[143] M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *Proceedings of the 4th European Conference on Computer Vision*, pages 343–356, Cambridge, UK, 1996.

[144] M. Isard and A. Blake. Condensation – Conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[145] P. C. Ivanov, L. A. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399(6735):461–465, Jun. 1999.

[146] A. H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[147] L. P. Kadanoff and C. Tang. Escape from strange repellers. *Proceedings of the National Academy of Sciences*, 81(4):1276–1279, Feb. 1984.

[148] V. Kadirkamanathan, M. H. Jaward, S. G. Fabri, and M. Kadirkamanathan. Particle filters for recursive model selection in linear and nonlinear system identification. In *Proceedings of the 39th IEEE Conference on Decision and Control*, volume 3, pages 2391–2396, Sydney, Australia, Dec. 2000.

[149] Y. Kajikawa. The adaptive Volterra filter: its present and future. *Electronics and Communications in Japan*, 83(12):51–60, 2000. Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J82-A, No. 6, June 1999, pp. 759–768.

[150] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions of the ASME*, 82:35–45, 1960.

[151] K. Kanazawa, D. Koller, and S. Russell. Stochastic simulation algorithms for dynamics probabilistic networks. In *Proceedings of the 11th Annual Conference on Uncertainty in AI*, pages 346–351, 1995.

[152] T. Kapitaniak. *Chaos in Systems with Noise.* World Scientific Publishing, Singapore, 2nd edition, 1990.

[153] D. T. Kaplan and L. Glass. Direct test for determinism in a time series. *Physical Review Letters*, 68(4):427–430, Jan. 1992.

[154] J. Kepler. *Mysterium Cosmographicum.* 1609.

[155] J. I. Kifer. On small random perturbations of some smooth dynamical systems. *Mathematics of the USSR*, 8(5):1083–1107, 1974.

[156] S. Kim and L. E. Reichl. Stochastic chaos and resonance in a bistable stochastic system. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 53(4):3088–3095, Apr. 1996.

[157] W. Kinsner. Characterizing chaos through Lyapunov metrics. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Review*, 36(2):141–151, Mar. 2006.

[158] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[159] C. Koch. *Biophysics of Computation: Information Processing in Single Neurons.* Oxford University Press, New York, 1999.

[160] A. N. Kolmogorov. A new metric invariant of transitive dynamical systems and Lebesgue space automorphisms. *Doklady Academy of Sciences USSR*, 124(4):861–864, 1958.

[161] R. Kopp and R. Orford. Linear regression applied to system identification for adaptive control systems. *American Institute of Aeronautics and Astronautics Journal*, 1:2300–2306, Oct. 1963.

[162] M. Korenberg. Identifying nonlinear difference equation and functional expansion representations. *Annals of Biomedical Engineering*, 16(1):123–142, Jan. 1988.

[163] M. J. Korenberg. Parallel cascade identification and kernel estimation for nonlinear systems. *Annals of Biomedical Engineering*, 19(4):429–455, Jul. 1991.

[164] M. J. Korenberg, R. David, I. W. Hunter, and J. E. Solomon. Parallel cascade identification and its application to protein family prediction. *Journal of Biotechnology*, 91(1):35–47, Sep. 2001.

[165] M. J. Korenberg and L. D. Paarmann. Orthogonal approaches to time-series analysis and system identification. *IEEE Signal Processing Magazine*, 8(3):29–43, Jul. 1991.

[166] M. J. Korenberg, J. E. Solomon, and M. E. Regelson. Parallel cascade identification as a means for automatically classifying protein sequences into structure/function groups. *Biological Cybernetics*, 82(1):14–21, Jan. 2000.

[167] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[168] C. Kwok, D. Fox, and M. Meila. Real-time particle filters. *Proceedings of the IEEE*, 92(3):469–484, Mar. 2004.

[169] L. D. Landau. On the problem of turbulence. *Doklady Akademii Nauk SSSR (Proceedings of the Russian Academy of Sciences)*, 44:339–342, 1944.

[170] L. J. Lapidus, D. Enzer, and G. Gabrielse. Stochastic phase switching of a parametrically driven electron in a Penning trap. *Physical Review Letters*, 83(5):899–902, Aug. 1999.

[171] M. Lax. Fluctuations from the nonequilibrium steady state. *Reviews of Modern Physics*, 32(1):25–64, Jan. 1960.

[172] M. Lax. Classical noise III: Nonlinear Markoff processes. *Reviews of Modern Physics*, 38(2):359–379, Apr. 1966.

[173] M. Lax. Classical noise IV: Langevin methods. *Reviews of Modern Physics*, 38(3):541–566, Jul. 1966.

[174] P. Lett, W. Christian, S. Singh, and L. Mandel. Macroscopic quantum fluctuations and first-order phase transition in a laser. *Physical Review Letters*, 47(26):1892–1895, Dec. 1981.

[175] H. Lewis. *Foundations of Fuzzy Control.* Plenum, New York, 1997.

[176] J. Liu and S. Marcus. Sequential Monte Carlo methods for dynamical systems. *Journal of American Statistical Assocation*, 93:1032–1044, 1998.

[177] K. Liu, H. Li, X. Dai, and P. Xu. Particle filtering based separation of chaotic signals. *Journal of Information & Computational Science*, 2(2):283–287, Jun. 2005.

[178] E. S. Lobanova, E. E. Shnol, and F. I. Ataullakhanov. Complex dynamics of the formation of spatially localized standing structures in the vicinity of saddle-node bifurcations of waves in the reaction-diffusion model of blood clotting. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 70(3):032903, Sep. 2004.

[179] K.-K. Loh, A. Saxena, T. Lookman, and A. N. Parikh. Phase transition induced hydrodynamic instability and Langmuir-Blodgett deposition. *ArXiv Condensed Matter e-prints*, Nov. 2001.

[180] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, Jan. 1963.

[181] V. Loreto, G. Paladin, and A. Vulpiani. Concept of complexity in random dynamical systems. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 53(3):2087–2098, Mar. 1996.

[182] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, pages 572–578, Kerkyra, Greece, Sep. 1999.

[183] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, Jul. 1977.

[184] B. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman and Company, New York, 1982.

[185] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *Society for Industrial and Applied Mathematics Review*, 10(4):422–437, 1968.

[186] O. L. Mangasarian. Linear and nonlinear separation of patterns by linear programming. *Operations Research*, 13:444–452, 1965.

[187] O. L. Mangasarian. Generalized support vector machines. In P. Bartlett, B. Schölkopf, D. Schuurmans, and A. J. Smola, editors, *Advances in Large Margin Classifiers*, pages 135–146, Cambridge, MA, 2000. MIT Press. Mathematical Programming Technical Report 98-14, October 1998.

[188] V. Z. Marmarelis. *Nonlinear Dynamic Modeling of Physiological Systems.* IEEE Press Series in Biomedical Engineering. Wiley Interscience, New Jersey, 2004.

[189] K. Matsumoto and I. Tsuda. Noise-induced order. *Journal of Statistical Physics*, 31(1):87–106, Apr. 1983.

[190] J. C. Maxwell. Molecules. *Nature*, 8:437–441, Sep. 1873.

[191] J. C. Maxwell. *The Scientific Papers of James Clerk Maxwell*, volume 1. The University Press, Cambridge, 1890. Reprinted (1952). New York: Dover Publications.

[192] R. May. Biological populations with non-overlapping generations: stable points, stable cycles, and chaos. *Science*, 186(4164):645–647, Nov. 1974.

[193] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, Jun. 1976.

[194] G. Mayer-Kress and H. Haken. The influence of noise on the logistic model. *Journal of Statistical Physics*, 26(1):149–171, Sep. 1981.

[195] B. McNamara and K. Wiesenfeld. Theory of stochastic resonance. *Physical Review A: Atomic, Molecular, and Optical Physics*, 39(9):4854–4869, 1989.

[196] R. Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(2):175–184, Apr. 1970.

[197] R. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17(5):693–698, Oct. 1972.

[198] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London*, 209:415–446, 1909.

[199] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, Sep. 1949.

[200] M. M. Millonas and L. E. Reichl. Stochastic chaos in a class or fokker-planck equations. *Physical Review Letters*, 68(21):3125–3128, May 1992.

[201] R. Morales-Menendez, N. de Freitas, and D. Poole. Estimation and control of industrial processes with particle filters. In *Proceedings of the American Control Conference*, pages 579–584, Denver, CO, Jun. 2003.

[202] S. Mukherjee, E. Osuna, and F. Girosi. Nonlinear prediction of chaotic time series using support vector machines. In J. Principe, L. Giles, N. Morgan, and E. Wilson, editors, *IEEE Workshop on Neural Networks for Signal Processing VII*, pages 511–520. IEEE Press, 1997.

[203] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik. Using support vector machines for time series prediction. In B. Schölkopf, C. J. C. Burges, and J. A. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 243–253, Cambridge, MA, 1999. MIT Press. Short version appeared in Proceedings of the 7th International Conference on Artificial Neural Networks, 1997.

[204] P. Muller. Monte Carlo integration in general dynamic models. *Contemporary Mathematics*, 115:145–163, 1991.

[205] E. Mumolo and A. Carini. A stability condition for adaptive recursive second-order polynomial filters. *Signal Processing*, 54(1):85–90, Oct. 1996.

[206] K. A. Myers and B. D. Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 21(4):520–523, Aug. 1976.

[207] T. Nagatani. Hydrodynamic instability and the structural phase transition in diffusion-limited aggregation with drift. *Physical Review A: Atomic, Molecular, and Optical Physics*, 40(9):5351–5355, Nov. 1989.

[208] T. Nakagawa and Y. Matsumoto. Detecting errors in corpora using support vector machines. In *Proceeding of the 19th International Conference on Computational Linguistics*, pages 1–7, Morristown, NJ, 2002. Association for Computational Linguistics.

[209] D. Nychka, S. Ellner, A. R. Gallant, and D. McCaffrey. Finding chaos in noisy systems. *Journal of the Royal Statistical Society B*, 54(2):399–426, 1992.

[210] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing.* Prentice Hall, New Jersey, 2nd edition, 1999.

[211] S. M. Ossadnik, S. V. Buldyrev, A. L. Goldberger, S. Havlin, R. N. Mantegna, C. K. Peng, M. Simons, and H. E. Stanley. Correlation approach to identify coding regions in DNA sequences. *Biophyiscal Journal*, 67(1):64–70, Jul. 1994.

[212] G. Paladin, M. Serva, and A. Vulpiani. Complexity in dynamical systems with noise. *Physical Review Letters*, 74(1):66–69, Jan. 1995.

[213] G. Paladin and A. Vulpiani. Anomalous scaling laws in multifractal objects. *Physics Reports*, 156(4):147–225, Dec. 1987.

[214] A. V. Panfilov, Stefan C. Müller, Vladimir S. Zykov, and James P. Keener. Elimination of spiral waves in cardiac tissue by multiple electrical shocks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 61(4):4644–4647, Apr. 2000.

[215] T. M. Panicker, V. J. Mathews, and G. L. Sicuranza. Adaptive parallel-cascade truncated Volterra filters. *IEEE Transactions on Signal Processing*, 46(10):2664–2673, Oct. 1998.

[216] R. F. Pawula. Approximation of the linear Boltzmann equation by the Fokker-Planck equation. *Physical Review*, 162(1):186–188, Oct. 1967.

[217] D. Pazó and M. A. Matías. Direct transition to high-dimensional chaos through a global bifurcation. *Europhysics Letters*, 72(2):176–182, Sep. 2005.

[218] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of DNA nucleotides. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 49(2):1685–1689, Feb. 1994.

[219] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):82–87, Mar. 1995.

[220] J. Pérez and C. Jeffries. Effects of additive noise on a nonlinear oscillator exhibiting period doubling and chaotic behavior. *Physical Review B: Condensed Matter and Material Physics*, 26(6):3460–3462, Sep. 1982.

[221] Ya. B. Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys*, 32(4):55–114, 1977.

[222] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, Mar. 1991.

[223] S. M. Pincus and A. L. Goldberger. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology—Heart and Circulatory Physiology*, 266(4):H1643–H1656, 1994.

[224] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74(2):189–197, Jun. 1980.

[225] C.-S. Poon and M. Barahona. Titration of chaos with added noise. *Proceedings of the National Academy of Science*, 98(13):7107–7112, Jun. 2001.

[226] C.-S. Poon and C. K. Merrill. Decrease of cardiac chaos in congestive heart failure. *Nature*, 389(6650):492–495, Oct. 1997.

[227] Schenk-Hoppé. K. R. Bifurcations of the randomly perturbed logistic map. Discussion paper 353, Department of Business Administration and Economics, University of Bielefeld, Bielefeld, Germany, 1997.

[228] D. A. Rand and H. B. Wilson. Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proceedings: Biological Sciences*, 246(1316):179–184, Nov. 1991.

[229] D. A. Rand and H. B. Wilson. Detecting chaos: a critique of the Sugihara-May approach to time series analysis. Technical Report 49, University of Warwick, 1991.

[230] H. Rauch, F. Tung, and C. Striebel. Maximum likelihood estimates of linear dynamic systems. *American Institute of Aeronautics and Astronautics Journal*, 3(8):1445–1450, Aug. 1965.

[231] S. Redner. *A Guide to First-Passage Processes.* Cambridge University Press, London, 2001.

[232] L. E. Reichl. *The Transition to Chaos in Conservative Systems: Quantum Manifestations.* Springer-Verlag, New York, 2nd edition, 2004.

[233] F. Reif. *Fundamentals of Statistical and Thermal Physics.* McGraw-Hill Series in Fundamentals of Physics. McGraw-Hill Science Engineering, New York, 1965.

[234] P. Reimann and E. Lootens. Escape rates for noisy maps with anomalous prefactors. *Europhysics Letters*, 34(1):1–6, Apr. 1996.

[235] P. Reimann and P. Talkner. Escape rates for noisy maps. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 51(5):4105–4113, May 1995.

[236] P. M. V. Résibois and M. De Leneer. *Classical Kinetic Theory of Fluids.* John Wiley and Sons, New York, 1977.

[237] I. Rhodes. A tutorial introduction to estimation and filtering. *IEEE Transactions on Automatic Control*, AC-16(6):688–706, Dec. 1971.

[238] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology—Heart and Circulatory Physiology*, 278(6):H2039–H2049, Jun. 2000.

[239] H. Risken. *The Fokker-Planck Equation: Methods of Solutions and Applications*, volume 18 of *Springer Series in Synergetics.* Springer-Verlag, Berlin, 1984.

[240] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, Norwell, MA, 2004.

[241] M. T. Rosenstein, J. J. Collins, and C. J. De Luca. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2):117–134, May 1993.

[242] D. Ruelle and F. Takens. On the nature of turbulence. *Communications in Mathematical Physics*, 20(3):167–192, 1971.

[243] W. J. Rugh. *Nonlinear System Theory: The Volterra/Wiener Approach.* The Johns Hopkins University Press, Baltimore, MD, 1981.

[244] D. A. Russell, J. D. Hanson, and E. Ott. Dimension of strange attractors. *Physical Review Letters*, 45(14):1175–1178, Oct. 1980.

[245] J. M. Sancho and J. García-Ojalvo. Noise-induced order in extended systems: a tutorial. In J. A. Freund and Th. Pöschel, editors, *Stochastic Processes in Physics, Chemistry, and Biology*, volume 557 of *Lecture Notes in Physics*, page 235. Springer-Verlag, New York, 2000.

[246] S. F. Schmidt. Computational techniques in Kalman filtering. In *AGARDograph 139*, London, Feb. 1970. NATO Advisory Group for Aerospace Research and Development.

[247] I. Schoenberg. Positive definite functions on spheres. *Duke Mathematical Journal*, 9:96–108, 1942.

[248] B. Schölkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In U. M. Fayyad and R. Uthurusamy, editors, *Proceedings of the 1st International Conference on Knowledge Discovery & Data Mining*, Menlo Park, CA, 1995. AAAI Press.

[249] T. Schön and F. Gustafsson. Particle filters for system identification of state-space models linear in either parameters or states. In *Proceedings of the 13th IFAC Symposium on System Identification*, pages 1287–1292, Rotterdam, The Netherlands, Sep. 2003.

[250] J. C. Schouten, F. Takens, and C. M. van den Bleek. Estimation of the dimension of a noisy attractor. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 50(3):1851–1861, Sep. 1994.

[251] E. Schrödinger. *Statistical Mechanics.* Cambridge University Press, Cambrdige, UK, 1952.

[252] L. Servi and Y. Ho. Recursive estimation in the presence of uniformly distributed measurement noise. *IEEE Transactions on Automatic Control*, 26(2):563–564, Apr. 1981.

[253] J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, 410:242–250, Mar. 2001.

[254] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana-Champaign, 1949.

[255] A. Shashidhar. Generalized Volterra-Wiener and surrogate data methods for complex time series analysis. Master's thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, Aug. 2006.

[256] I. Shimada and T. Nagashima. The iterative transition phenomenon between periodic and turbulent states in a dissipative dynamical system. *Progress of Theoretical Physics*, 59(3):1033–1036, Mar. 1978.

[257] T. Shinbrot and F. J. Muzzio. Noise to order. *Nature*, 410:251–258, Mar. 2001.

[258] B. Shraiman, C. E. Wayne, and P. C. Martin. Scaling theory for noisy period-doubling transitions to chaos. *Physical Review Letters*, 46(14):935–939, Apr. 1981.

[259] A. J. F. Siegert. On the first passage time probability problem. *Physical Review*, 81(4):617–623, Feb. 1951.

[260] E. Simiu and M. Frey. Melnikov processes and noise-induced exits from a well. *Journal of Engineering Mechanics*, 122(3):263–270, Mar. 1996.

[261] Ya. G. Sinai. On the concept of entropy for a dynamical system. *Comptes Rendus (Doklady) Academy of Sciences USSR*, 124(4):768–771, 1959.

[262] J. E. Skinner. Low-dimensional chaos in biological systems. *Nature Biotechnology*, 12(6):596–600, Jun. 1994.

[263] A. Smith and A. Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88, May 1992.

[264] A. J. Smola and B. Schölkopf. On a kernel-based method for pattern recognition, regression, approximation, and operator inversion. *Algorithmica*, 22:211–231, 1998.

[265] J. A. Smola, N. Murata, B. Schölkopf, and K.-R. Müller. General cost functions for support vector regression. In T. Downs, M. Frean, and M. Gallagher, editors, *Proceedings of the 9th Australian Conference on Neural Networks*, pages 79–83, Brisbane, Australia, 1998.

[266] J. Spall, editor. *Bayesian Analysis of Times Series and Dynamic Models.* Marcel Dekker, New York, 1988.

[267] A. Srivastava, A. D. Lanterman, U. Grenander, M. Loizeaux, and M. I. Miller. Monte Carlo techniques for automated target recognition. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 26, pages 533–552. Springer-Verlag, New York, 2001.

[268] J. Stark. Delay embeddings for forced systems I: deterministic forcing. *Journal of Nonlinear Science*, 9(3):255–332, Jun. 1999.

[269] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke. Delay embeddings for forced systems II: stochastic forcing. *Journal of Nonlinear Science*, 13(6):519–577, Dec. 2003.

[270] S. H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering.* Addison-Wesley, Reading, MA, 1994.

[271] M. Strumik, W. M. Macek, and S. Redaelli. Discriminating additive from dynamical noise for chaotic time series. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 72(3):036219, Sep. 2005.

[272] G. Sugihara and R. M. May. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734–741, 1990.

[273] M. Suzuki. Scaling theory of transient nonlinear fluctuations and formation of macroscopic order. *Progress of Theoretical Physics Supplement*, 64:402–424, 1978.

[274] F. Takens. *Detecting strange attractors in turbulence*, volume 898 of *Lecture notes in mathematics.* Springer-Verlag, Berlin, 1981.

[275] C. Tao, J. Mu, X. Xu, and G. Du. Chaotic characteristics of speech signal and its LPC residual. *Acoustical Science and Technology*, 25(1):50–53, 2004.

[276] K. Testa, J. Péerez, and C. Jeffries. Evidence for universal chaotic behavior of a driven nonlinear oscillator. *Physical Review Letters*, 48(11):714–717, Mar. 1982.

[277] J. Theiler and S. Eubank. Don't bleach chaotic data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 3(4):771–782, Oct. 1993.

[278] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, Sep. 1992.

[279] J. Timmer. Modeling noisy time series physiological tremor. *International Journal of Bifurcation and Chaos*, 8(7):1505–1516, May 1998.

[280] J. Timmer and A. S Weigend. Modeling volatility using state space models. *International Journal of Neural Systems*, 8(4):385–398, Aug. 1997.

[281] W.-W. Tung, Y. Qi, Y. Cao, and L. Billings. Direct characterization of chaotic and stochastic dynamics in a population model with strong periodicity. *Chaos, Solitons, and Fractals*, 24:645–652, 2005.

[282] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Physical Review*, 36(5):823–841, Sep. 1930.

[283] S. Ulam and J. von Neumann. On combination of stochastic and deterministic processes. *Bulletin of the American Mathematical Society*, 53:1120, Nov. 1947.

[284] K. Urbanowicz and J. A. Hołyst. Noise-level estimation of time series using coarse-grained entropy. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 67(4):046218, Apr. 2003.

[285] C. Van den Broech, J. M. R. Parrondo, R. Toral, and R. Kawai. Nonequilibrium phase transitions induced by multiplicative noise. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 55(4):4084–4094, Apr. 1997.

[286] R. van Zon, H. van Beijeren, and J. R. Dorfman. Kinetic theory of dynamical systems. In J. Karkheck, editor, *Proceedings of the 1998 NATO-ASI on Dynamics: Models and Kinetic Methods for Non-equilibrium Many-Body Systems*, volume 317 of *NATO Science Series E*, pages 131–168, Leiden, The Netherlands, 2000. The NATO Advanced Study Institute, Kluwer Academic Publishers.

[287] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.

[288] V. Vapnik. *Statistical Learning Theory.* John Wiley and Sons, New York, 1998.

[289] V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, 25, 1964.

[290] V. Vapnik, S. Golowich, and A. Smola. Support vector method for function approximation, regression estimation, and signal processing. In M. C. Mozer, M. I. Jordan, and Petsche T., editors, *Advances in Neural Information Processing Systems 9*, pages 281–287, Cambridge, MA, 1997. MIT Press.

[291] V. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24, 1963.

[292] J. Vermaak, C. Andrieu, A. Doucet, and S. Godsill. Particle methods for Bayesian modeling and enhancement of speech signals. *IEEE Transactions on Speech and Audio Processing*, 10:173–185, Mar. 2002.

[293] H. U. Voss, J. Timmer, and J. Kurths. Nonlinear dynamical system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos*, 14(6):1905–1933, 2004.

[294] H. Wang and Q. Ouyang. Effect of noise on defect chaos in a reaction-diffusion model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):023702, Apr. 2005.

[295] M. C. Wang and G. E. Uhlenbeck. On the theory of the Brownian motion II. *Reviews of Modern Physics*, 17(2-3):323–342, Apr. 1945.

[296] J. N. Weiss, A. Garfinkel, M. L. Spano, and W. L. Ditto. Chaos and chaos control in biology. *Journal of Clinical Investigation*, 93(4):1355–1360, Apr. 1994.

[297] N. Wiener. *Cybernetics.* The MIT Press, Cambridge, MA, 1948.

[298] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, 16(3):1985, Jul. 1985.

[299] M. Woltering and M. Markus. Riddled-like basins of transient chaos. *Physical Review Letters*, 84(4):630–633, Jan. 2000.

[300] W. Yang, A. Joshi, and M. Xiao. Controlling dynamic instability of three-level atoms inside an optical ring cavity. *Physical Review A: Atomic, Molecular, and Optical Physics*, 70(3):033807, Sep. 2004.

[301] N. J. Zabusky, editor. *Topics in Nonlinear Physics.* Springer-Verlag, New York, 1968.

[302] A. Zaikin and J. Kurths. Additive noise in noise-induced nonequilibrium transitions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 11(3):570–580, Sep. 2001.

[303] M. A. Zaks, X. Sailer, L. Schimansky-Geier, and A. B. Neiman. Noise induced complexity: from subthreshold oscillations to spiking in coupled excitable systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026117, Jun. 2005.

[304] D. Zhang, L. Györgyi, and R. Peltier. Deterministic chaos in the Belousov-Zhabotinsky reaction: experiments and simulations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 3(4):723–745, 1993.

[305] K. Ziemelis. Nature insight: Complex systems. *Nature*, 410:241, Mar. 2001.

[306] A. Zippelius and M. Lücke. The effect of external noise in the Lorenz model of the Bénard problem. *Journal of Statistical Physics*, 24(2):345–358, Feb. 1981.